Title

Material properties by design: controlling the self-assembly, dynamics, and emergent attributes of reconfigurable two-dimensional protein crystals

Author

Alberstein, Robert Gregory

Publication Date

2020

UNIVERSITY OF CALIFORNIA SAN DIEGO

Material properties by design: controlling the self-assembly, dynamics, and emergent attributes of reconfigurable two-dimensional protein crystals

A dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy

in

Chemistry

by

Robert Gregory Alberstein

Committee in charge:

Professor F. Akif Tezcan, Chair
Professor Galia Debelouchina
Professor Olga Dudko
Professor Francesco Paesani
Professor Wei Xiong

2020

The Dissertation of Robert Gregory Alberstein is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California San Diego

2020

DEDICATION

To my friends and family, near and far, for your eternal support and encouragement,
without which this thesis could not have been completed.

EPIGRAPH

“In Physicis nil desperandum est disciplina chimica duce”

(“In physics one need not despair when guided by chemistry”)

Boerhaave

“If you want to know the secrets of the universe, think in terms of energy, frequency, and vibration”

Nikola Tesla

“Facts are of not much use, considered as facts. They bewilder by their number and their apparent incoherency. Let them be digested into theory, however, and brought into mutual harmony, and it is another matter”

Oliver Heaviside

“A mathematician may say anything he pleases, but a physicist must be at least partially sane”

Josiah Willard Gibbs

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

1D	one dimensional
2D	two dimensional
3D	three dimensional
AFM	atomic force microscopy
AMBER	Assisted Model Building with Energy Refinement
APBS	adaptive Poisson-Boltzmann solver
Asp	aspartate
ATP	adenine triphosphate
Cbi	cobinamide
CHARMM	Chemistry at HARvard Molecular Mechanics
CID	chemically induced dimerization
Cryo-EM	cryogenic electron microscopy
Cterm	C-terminus
Cys	cysteine
DFT	density functional theory
E	energy
EDTA	ethylenediaminetetraacetic acid
ES	electrostatics
F-phlogopite	fluorophlogopite
FF	force field
FPLC	fast protein liquid chromatography
ΔG	Gibbs free energy
Glu	glutamate
GSH	reduced glutathione

H	enthalpy
HCN	hydrogen cyanide
His	histidine
H_{water}	water hydrogen atom(s)
IFF	INTERFACE force field
IPTG	isopropyl β-D-1- thiogalactopyranoside
k_B	Boltzmann constant
KCl	potassium chloride
KCN	potassium cyanide
LJ	Lennard-Jones
MBPC1	metal-binding protein construct 1
MD	molecular dynamics
MDPSA	metal-directed protein self-assembly
MES	2-(N-morpholine)ethanesulfonic acid
MeTIR	metal-templated interface redesign
micro-ED	microcrystal electron diffraction
m-mica	muscovite mica
MT	microtubule
NAMD	NAnoscale Molecular Dynamics
NMR	nuclear magnetic resonance
NP	nanoparticle
NPT	constant pressure/temperature simulation
ns-TEM	negative-strain transmission electron microscopy
Nterm	N-terminus
NVT	constant volume/temperature simulation

OH	hydroxyl
OH-phlogopite	phlogopite
O_{water}	water oxygen atom
PBC	periodic boundary conditions
PME	particle mesh ewald
PMF	potential of mean force
PPI	protein-protein interaction
pSi	porous silicon
RDF	radial distribution function
RhuA	L-rhamnulose-1-phosphate aldolase
RI	refractive index
RIDC	Rosetta interface designed construct
S	entropy
SEM	scanning electron microscopy
Ser	serine
S-layer	surface layer
SPC(/E)	(extended) simple point charge
T	temperature
TEM	transmission electron microscopy
TIP3P	transferable intermolecular potential 3 point
TMZ	Tris, β-Mercaptoethanol, Zinc
Tris	tris(hydroxymethyl)aminomethane hydrochloride
U	internal energy
UV-Vis	ultraviolet-visible spectroscopy
vdW	van der Waals

VMD	Visual Molecular Dynamics
xy	in-plane or planar
z	normal to the plane
ZDF	Z-density distribution function
$ZnCl_2$	zinc chloride
βME	beta-mercaptoethanol

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I of course must thank Akif for taking a chance on me and my aspirations, allowing me to charge forth on my own in pursuit of the science I dreamed of doing, while providing crucial guidance and motivation along the way. In many ways this was essential for my overwhelmingly positive experience in graduate school. I truly believe it was the experience I needed to lay the foundation for my long-term research goals, and I will do my best to carry the creativity and rigor impressed upon us here into the future.

The requisite primer for this undertaking was certainly my time in the lab of Mark Mayer (now FRS!) at the NIH. As a post-baccalaureate researcher, that one year single-handedly transformed me from a relatively inexperienced undergraduate scientist into one with detailed experimental technique (and strong opinions about how best to do them) and a dedicated work ethos. I forever appreciate the patience he exhibited during the more frustrating moments of closely mentoring me at my most inexperienced.

I'd like to thank Francesco Paesani for taking me under his wing as my "unofficial second advisor" and helping me get established in the world of computational science. Similarly, I'd like to acknowledge Mike Gilson for meeting with me on several occasions over the years for fun and stimulating discussions. I thank both of them for always being available to listen, discuss, and advise me on all manners of topics, as well as introducing me to many others along the way.

For the past several years I have had the pleasure of working with several tremendous scientists at PNNL. Jim De Yoreo is an exemplary leader who somehow carries (what appears to be) a huge chunk of the materials science world on his shoulders, which has been an awe to behold and a constant source of inspiration. A special shout-out to Shuai Zhang for being an excellent collaborator, as well as a good friend. It was truly enjoyable working together and hope we will continue to cross paths in the future. I was introduced to Chris Mundy about a year and a half ago and have loved every minute. It's too bad we couldn't pull off a short stint at PNNL but I'm sure we'll figure something out for the future. To all of you, it is my sincere hope that we manage to remain in contact and collaboration for years to come.

On that note, I would like to acknowledge the CSSAS crew for really providing a space where I felt at home and got to experience the most fun I've had in a grant yet. Special shout outs to Harley Pyles, Julia Boese, Laura Pflum, Maxim Ziatdinov, Shuai Zhang, Jim De Yoreo, and the secret theory crew: Sarah Alamdari, Orion Dollar, Andrew Ferguson, Kacper Lachowski, Chris Mundy, Jim Pfaendtner, Lilo Pozzo, Jesse Prelesnik, Xin Qi, Janani Sampath, and Mingfei Zhao.

At home I have been fortunate to work with a talented group of scientists in the Tezcan lab. Truly this is the place when good ideas are made great, the unusual ideas made extraordinary, and colleagues become family (it turns out I have overlapped with quite a number of individuals throughout the evolution of the lab). Special shouts to Lewis Churchfield for our enduring friendship: it has been fun partying together and quite enjoyable seeing you become involved in the community you have found, as well as keeping me a part of it. To Jake Bailey: for whom I have tremendous respect, whose intense intelligence and work ethic are inspirational, who always has an incisive take on information, and isn't afraid to get crazy with his hypotheses. We've had a lot of great, heated discussions which were lots of fun and kept things fresh. And to the "hard working" materials subgroup in all of its iterations: it was an honor and joy to work with you. Yuta Suzuki was the inventor of ^{C98}RhuA, and my first working mentor at UCSD. Now as a professor at Kyoto University, I feel we see each other as true colleagues, and it has been fun to catch up and keep in touch. Hope we can bring our proposed collaboration to fruition in the future. Zhiyin (Iris) Zhang, who I have mentored for the past two years, was thrust into a sea of unknowns just as I was as a young grad student in this lab. Keep your head up and work hard, the physics will come! I know you will make it. Nicole Avakyan's gentle demeanor and boisterous laughter has been wonderful to be around, and she lends a great element of coziness to our group. Tomoyuki Yao's time with us was too brief. His quirky sense of humor, happiness to swap Japanese and English phrases, and unbelievable ability to pull off the look of a professional model in photos are missed dearly. I thank him and his wife Saori for all of the good times and for hosting my brother and I during our first expedition to Japan. And of course last but not least, Rohit Subramanian. Following a chance overlap in niche music, he has become my closest friend in lab and a trusted advisor. Few things go out, scientific or otherwise, without us acting as mutual sounding boards beforehand. I'm quite certain that our ability to produce an endless stream of banter has both shocked and appalled the rest of the lab, but it was fun, and a hell of a lot of good science came out of it. Finally, I of course need to acknowledge Ling Zhang (and Miumiu) for all of her support, especially in these past few months. I honestly don't know how I would have done it without you. It's been great getting to explore our complementary viewpoints on the world and see us grow stronger together. Your toughness and intelligence are enough to keep even the biggest personalities at bay, yet you are also kind and resolute. Thanks for lending me your strength.

To my mom: what can I even say here that would capture what I wish to articulate? You are my eternal role model, my guiding light, my most supportive cheerleader, a source of strength, warmth, compassion, and intelligence. Our family has endured so many ups and downs and yet you've never wavered as our steadfast captain. Truly I have no idea how you do it all. From the life lessons you instilled in me as a child all the way until now, there's no way I could have done this without you. There is no better mom out there, and you've shown me the kind of human (and parent) I should strive to emulate. I love you so very much.

To my brother Scott: you are the real OG. For all the twists and turns, separations and reunions, you’ve been by my side (literally and figuratively) the entirety of our journey to become the people we are today. Thanks for all the memories and for being the best brother, roommate, teammate, partner, and Japan-companion a guy could ask for. I know we will continue to have a lot of fun in our new chapter in SF and beyond, so let’s have a blast. I love you dude.

To my sister Ainsley: watching you grow into the strong young lady you are today has been a magical experience, though you will remain eternally 6 years old in my heart. There is still so much left for you to learn and see and become and I can’t wait to see where your journey takes you. Keep it up and you know we have your back! I love you too.

To the rest of my found family: I can’t believe how many years it’s been already but you are all a real part of my home now. I have been fortunate to get to live close enough to return time and time again, and don’t know how I would have done it without all of your support. Keep up the gaming, computer, dancing, baking, rocking and K-popping you do. And to Jim for all of the fatherly support and laughs he has provided along the way. Love you guys.

At last, to my grandparents: the other real OGs. I am so happy you were able to see me defend and become a Ph.D; a silver lining in the necessity of a remote defense. I’m so proud of how far our family has come in just a few generations, and it is in no small part due to your influence. Staying smart, “not doing anything dumb”, “saving our money”, having a blast, crocheting afghans, providing taxi service & donut runs, and giving me some of my most cherished memories will forever be a part of your legacy. I can’t imagine life without you. Love you all.

This dissertation was funded in part by UCSD's Molecular Biophysics Training Grant (NIH), Bruno Zimm Award in Biophysics, and Distinguished Graduate Student Fellowship.

Chapter 2 is reproduced, in part, with permission, from: Alberstein, R.G., Suzuki, Y., Paesani, F., Tezcan, F.A. "Engineering the entropy-driven free-energy landscape of a dynamic nanoporous protein assembly", *Nature Chemistry* **10**, 732–739 (2018). The dissertation author was the primary author on all reprinted materials.

Chapter 3, in part, is currently being prepared for submission for publication: Alberstein, R.G.*, Vijayakumar, S.*, Lu, Y.-S., Chaix, A., Wahl, C.E., Ha, J.S., Hunka, D.E., Boss, G.R., Tezcan, F.A., Sailor, M.J. "A photonic porous silicon sensor with integrated biological gatekeeper for remote detection of HCN vapor". The dissertation author was the co-primary investigator and author of this manuscript.

Chapter 4, in part, has been submitted for publication: Zhang, S.*, Alberstein, R.G.*, De Yoreo, J.J., Tezcan, F.A. "Assembly of a patchy protein into variable 2D lattices via tunable, multiscale interactions" (2020). The dissertation author was the co-primary investigator and author of this manuscript.

Chapter 5, in part, is currently being prepared for submission for publication: Alberstein, R.G.*, Prelesnik, J.L.*, Mundy, C.J., Tezcan, F.A. "Molecular basis for surface patterning of biomolecules at aqueous mineral interfaces". The dissertation author was the co-primary investigator and author of this manuscript.

VITA

EDUCATION

2013	B.S., Biological Sciences, Carnegie Mellon University
2016	M.S., Chemistry, University of California, San Diego
2020	Ph.D., Chemistry, University of California, San Diego

HONORS AND AWARDS

2015	San Diego Fellowship Recipient
2015–2016	Molecular Biophysics Training Grant
2018	Bruno Zimm Award in Biophysics
2019	UCSD Distinguished Graduate Student Fellowship

PUBLICATIONS

Alberstein, R.G.*, Prelesnik, J.L.*, Mundy, C.J., Tezcan, F.A. “Investigating the molecular basis of surface patterning of biomolecules at aqueous mineral interfaces” (*in preparation*)

Alberstein, R.G.*, Vijayakumar, S.*, Liu, Y.-S., Chaix, A., Wahl, C.E., Ha, J.S., Hunka, D.E., Boss, G.R., Tezcan, F.A., Sailor, M.J. “A photonic porous silicon sensor with integrated biological gatekeeper for remote detection of HCN vapor” (*in preparation*)

Zhang, S.*, **Alberstein, R.G.***, De Yoreo, J.J., Tezcan, F.A. “Assembly of a patchy protein into variable 2D lattices via tunable, multiscale interactions” (*under review*)

Golub, E.G., Subramanian, R.H., Esselborn, J., **Alberstein, R.G.**, *et al*. “Constructing Protein Polyhedra from Asymmetric Monomers via Chemically Orthogonal Interactions.” *Nature* **578**, 172-176 (2020)

Subramanian, R.H.*, Smith, S.J.*, **Alberstein, R.G.**, *et al.* “Self-assembly of a designed nucleoprotein architecture through multimodal interactions.” *ACS Central Science* **4**, 1578-1586 (2018)

Churchfield, L.A., **Alberstein, R.G.**, Williamson, L.M., Tezcan, F.A. "Determining the structural and energetic basis of allostery in a de novo designed metalloprotein assembly." *Journal of the American Chemical Society* **140**, 10043-10053 (2018)

Alberstein, R., Suzuki, Y., Paesani, F., Tezcan, F.A. "Engineering the entropy-driven free-energy landscape of a dynamic nanoporous protein assembly." *Nature Chemistry* **10**, 732-739 (2018)

Yu, A., **Alberstein, R.**, Thomas, A., Zimmet, A., Grey, R., Mayer, M.L., & Lau, A.Y. "Molecular lock regulates binding of glycine to a primitive NMDA receptor." *Proceedings of the National Academy of Sciences* **113**, E6786-E6795 (2016)

Alberstein, R., *et al.* "Glycine activated ion channel subunits encoded by ctenophore glutamate receptor genes." *Proceedings of the National Academy of Sciences* **112**, E6048-E6057 (2015)

* These authors contributed equally to this work

FIELDS OF STUDY

Major Field: Chemistry

Studies in Theoretical and Computational Chemistry, Biophysics, Bioinorganic Chemistry
Professor F. Akif Tezcan

ABSTRACT OF THE DISSERTATION

Material properties by design: controlling the self-assembly, dynamics, and emergent attributes of reconfigurable two-dimensional protein crystals

by

Robert Gregory Alberstein

Doctor of Philosophy in Chemistry

University of California San Diego, 2020

Professor F. Akif Tezcan, Chair

Proteins, perhaps more than any other class of molecule, make life as we know it possible. The unrivaled champions of functional diversity, the 20+ different amino acids which comprise these polypeptide chains enable them to fold into sequence-specified structures (and structural ensembles) that physically encode their (structure-specified) functions. Yet, the vast majority of cellular proteins operate exclusively as self-assembled multi-subunit complexes, which combinatorially broadens the scope of their (shapes, structures, sizes, and thus) capabilities to encompass roles as molecular-scale catalysts, supramolecular nonequilibrium nanomachines, and

cellular-scale motors and intercommunications. This makes proteins highly desirable for the bottom-up construction of new materials: they are perfectly monodisperse nanoscale building blocks, possessing well-defined molecular structure(s) (and thus function(s)) that can be facilely/robustly fine-tuned via routine genetic engineering procedures. As exemplified throughout biology, (synthetic) proteins possess the intrinsic potential to propagate and amplify functional structure (structural function) over multiple scales, promising tailor-made self-assembled macroscale architectures and devices with emergent properties. Towards this goal, this dissertation describes several specific examples of this general principle, whereby new physical phenomena spontaneously manifest within flexible 2D materials self-assembled from an engineered protein.

We first report a theoretical and experimental dissection (followed by exploitation) of the bulk-scale behavior of crystalline lattices constructed from a square-shaped tetrameric aldolase with a designed cysteine at its corners (^{C98}RhuA). Upon oxidative self-assembly, the topology and flexibility of the intermolecular disulfide bonds enable the resulting architecture to undergo "breathing" motions that allow it to switch between porous and close-packed conformations without losing crystallinity. We characterized the free energy landscape associated with these dynamic motions, validated its accuracy against experiment, and determined that it is dominated by solvent reorganization entropy. We then exploited the lack of native intermolecular interactions to rationally engineer this landscape through the introduction of negatively charged glutamate residues at prescribed locations, which was then analogously characterized both experimentally and computationally. We further determined that this repulsion could be toggled via the chelation of Ca^{2+} ions, enabling controllable access to discrete conformational states through environmental conditions. This variant was subsequently utilized as a selective "gatekeeper" coating for a highly specific and sensitive chemical sensor for HCN.

Simultaneously, we characterized the self-assembly of ^{C98}RhuA at mineral interfaces via *in situ* AFM. We found that the use of a charged surface as a template for crystallization enabled the formation of three new crystal morphologies. We validated the observed crystal patterns against atomic structural models and used free-energy calculations to rationalize the formation of each crystal morphology. We also determined that ^{C98}RhuA possesses a sizeable macrodipole moment (~1200 D), which calculations indicated explained the preference for antiparallel packing of protein units ($p42_12$ symmetry); this is the first time dipole-dipole interactions have been shown to control protein self-assembly. Finally, the alignment of dipole moments in surface-grown crystals was predicted to make them "electrets". Indeed, MD simulations of both $p4$ symmetry (parallel packing) and $p42_12$ protein crystals revealed that the former exhibits a conformation-dependent membrane voltage of ≤ -100 mV (in excellent agreement with analytical calculations), thus delineating the connection between the individual building block and macroscale properties when the patterning of component orientations is controlled. Finally, preliminary computational investigations into the epitaxial growth of ^{C98}RhuA crystals on these surfaces are described.

Chapter 1: The hierarchical organization of macromolecular structure and function

1.1 Introduction

Life as we know it arises from non-living organic matter through the spontaneous emergence of physical properties that originate from their particular organization[1]. This relation between form (structure) and function (properties) is one of the most fundamental aspects of the physical and chemical processes occurring in biological systems. Most notable among them, living organisms are dependent on three specific macromolecules: the biopolymers DNA, RNA, and proteins, whose distinct structural characteristics make them suitable for their individual functions (**Figure 1.1**). DNA, being double-stranded, provides a level of redundancy that makes it ideal for information storage. RNA, similarly composed of only four nucleotide units, is single-stranded, enabling it to fold up into a variety of three-dimensional shapes (depending on their sequence) to serve in an ever-expanding range of genetic, catalytic, and regulatory cellular functions. By far, however, the greatest share of biochemical processes is undertaken by proteins, whose 20+ distinct monomeric units endow them with the sequential (and thus structural) diversity required to perform the innumerable processes which underpin all of biology. In short, proteins are unparalleled as specialist tools, making them desirable targets for performing chemistry.

Key to this specialization is the direct relationship between each protein sequence and its unique 3D structure. The chemical and physical specificity of the particular order of amino acids—its primary structure—defines (in the vast majority of cases) the particular landscape over which this chain adopts common structural motifs (secondary structure) and collapses into its final folded state (tertiary structure). This hierarchy enables the precise placement of chemical groups to bind ligands and/or perform catalysis, as well as the generation of a tremendous diversity of shapes that exhibit different mechanical properties, with rigid structures serving as structural scaffolds while

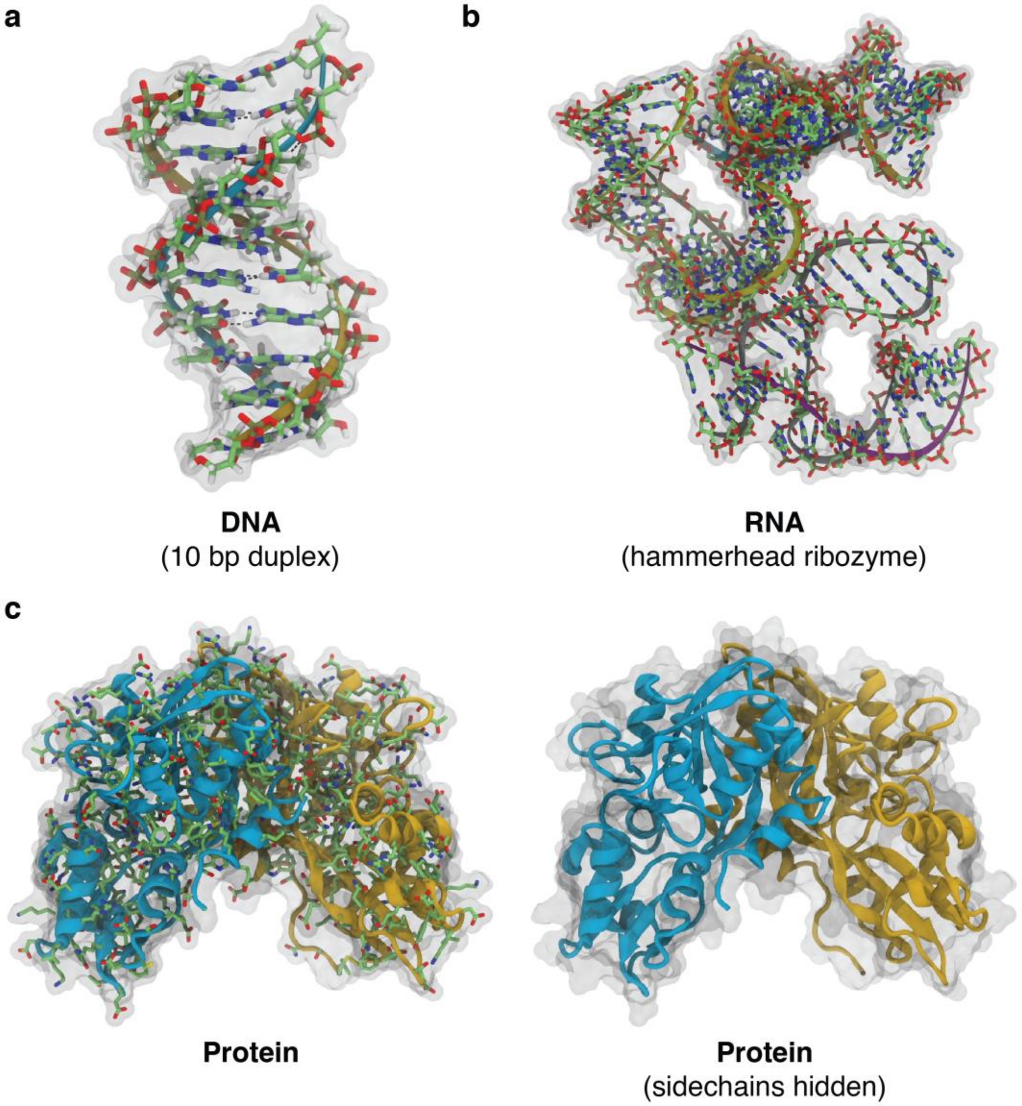
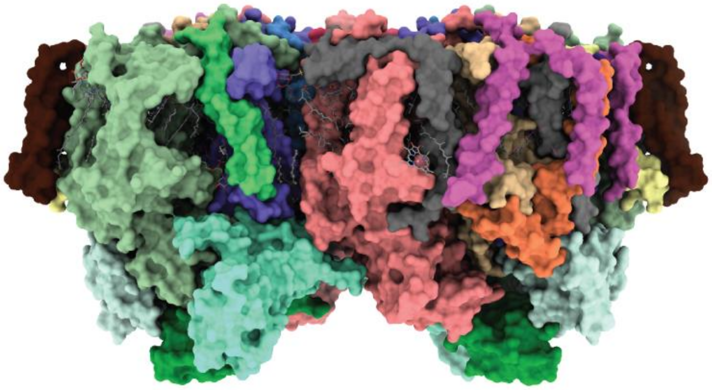


Figure 1.1 | Biomacromolecules and their structures. a, A simple DNA double helix formed from 10 base pairs, with specificity encoded by hydrogen bonding interactions between nucleotides. **b,** A folded RNA complex (PDB ID: 1HMH) with catalytic ability (a ribozyme). **c,** A folded protein (PDB ID: 4YKI), with a standard cartoon-only representation at right. For all sections, non-backbone atoms are shown as sticks, backbones are shown as cartoon and colored by chain, and each complex is encased within its transparent solvent-excluded surface. Note the greater diversity of protein sidechains compared to nucleic acids.

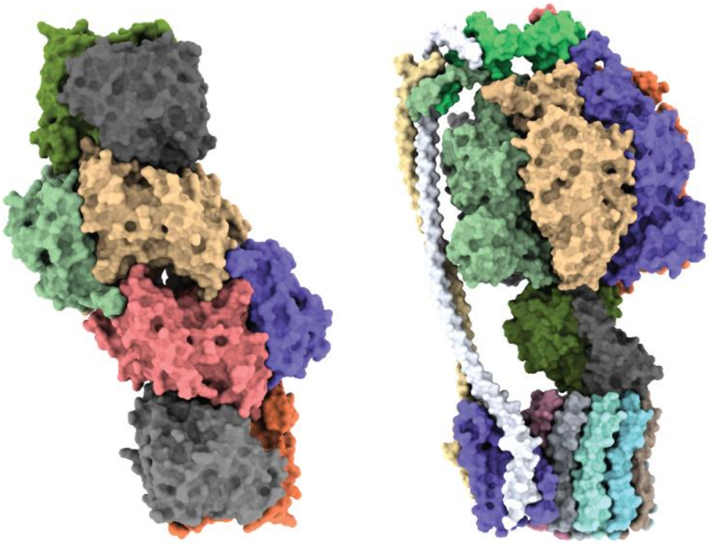
flexible structures serve as gates, switches, keys, and motors. Many of these functions arise from the highest level of protein organization, quaternary structure, whereby several protein units spontaneously form multimeric complexes[1]. In fact, it is estimated that as many as 80% of proteins form oligomers *in vivo* to carry out their roles[2]. Indeed, highly complex behaviors including the electrochemical transformation of photons into oxygen (*i.e.*, photosynthesis) by photosystem II[3], the generation of the metabolic "currency" adenine triphosphate (via the exploitation of chemical gradients) by ATP synthase[4], and the fixation of ambient nitrogen into ammonia (via multistep proton/electron injections) by nitrogenase[5], are all premier examples of near-inaccessible (to the laboratory scientist) chemical processes made possible through the association of multiple protein subunits (**Figure 1.2**). Moreover, spontaneous self-organization ("self-assembly") can be employed to generate periodic 1D, 2D, and 3D assemblies which serve other vitally important macroscale roles in cellular structure[6], mobility[7], replication[8], and defense[9]. Such examples underscore the possibilities inherent to the bottom-up self-assembly of nanoscale structures. They have also provided tremendous inspiration for the design of new biological materials, which is the primary subject of this dissertation.

1.2 Inverse design of biological materials: from function to sequence

An important phenomenon which may be inferred from the above examples is the concept of emergent behavior, by which new properties arise through the particular composition and arrangement of smaller, less-functional, building blocks[1]. This topic is well-appreciated within materials science research (most notably the field of nanoparticle self-assembly)[10,11], but is particularly challenging to extend to proteins. This is because the collective behavior exhibited by an assemblage of protein units is determined hierarchically: the quaternary (self-assembled) structure is dictated by the identity and positioning of interaction patches on the protein surface (*i.e.*, where multiple proteins will come together), meaning it is a function of the protein's tertiary structure, which (as discussed above) is itself unintuitively linked to its primary sequence.



Photosystem II



Nitrogenase

ATP synthase

Figure 1.2 | Biological machines constructed from many protein units. Natural examples of self-assembled proteins complexes which carry out challenging chemical reactions. Each colored surface is a separate protein chain within each complex. PDB IDs: 2AXT, 4WZA, 6N2Y.

Consequently, the design of new protein-based materials with specified properties not only requires an *a priori* understanding of how those properties arise from the specific self-assembled structure, but also necessitates a "top-down" deconstruction from that architecture all the way to a primary sequence. This is the "inverse design" problem in materials science, and is an open question for the vast majority of materials. To begin to address this problem with the goal of making it tractable to solve, we can attempt to cast this hierarchy into a common framework (as discussed in the next section) based on energy landscapes, which can be utilized to understand the folding, self-assembly (structure), and emergent behavior (properties) of protein building blocks. While even getting from primary to tertiary structure in proteins (protein folding) is itself an only partially solved problem, and the mechanisms of self-assembly are still being fully elucidated, we shall see that there is still a tremendous amount of information to be gleaned from what we can currently compute and experimentally test.

1.3 Statistical thermodynamics of biological complexity

1.3.1 Energy landscape theory

Let us first start by introducing the concept of the energy landscape. The “energy landscape” is a way of picturing the distribution of different possible states of a system, where the energy is a function of one or more parameters (“coordinates”) which can be used to define these states. In the context of the chemical systems pertaining to this work, these landscapes will represent the thermodynamic free energy (typically the Gibbs free energy), providing a statistical description of distribution of the system states at equilibrium. We can see this connection by considering the Boltzmann distribution, which links the probability of a macroscopic system being in a particular microstate to the energy of that state (**Equation 1.1**):

$$P(X = x) = \frac{1}{Z(\beta)} e^{-\beta E(x)} \tag{1.1}$$

Here, β is taken to be the inverse temperature $(k_B T)^{-1}$, E(x) is the (free) energy associated with the state x, and Z is the partition function (the sum of all possible microstates) for that temperature, which serves to normalize the probability distribution. What can be seen is that there is an inverse relation between the energy of a state and its probability (*i.e.*, the lower the energy, the higher the probability) (**Figure 1.3**). For this reason, it is not uncommon for me to say "(free) energies are just a convenient way of thinking about probabilities", which hopefully reduces the inherent abstractness of the concept somewhat for the reader.

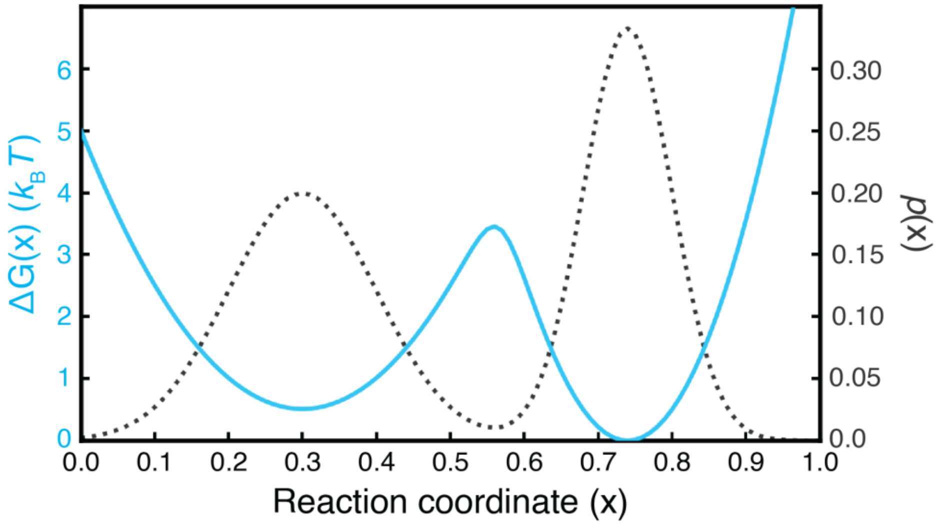


Figure 1.3 | Relationship between probability and free energy. The probability distribution (and corresponding free energy) for a hypothetical double-well potential is plotted as a function of an arbitrary reaction coordinate. Note the inverse relationship between the two functions.

As a final topic before we dive in any further, we should discuss the actual definition of the Gibbs free energy (ΔG). Most formally, ΔG is the maximum reversible work that can be extracted from a closed thermodynamic system at constant temperature and pressure, where the delta refers to moving from one system state to the next. More practically, we will refer to ΔG as the thermodynamic potential which is minimized for a system at equilibrium under the conditions of constant temperature and pressure (*i.e.*, conditions most resembling a typical biochemical experiment). This is apparent by consideration of the Gibbs free energy equation[12] (**Equation 1.2**):

$$\Delta G = \Delta H - T\Delta S \tag{1.2}$$

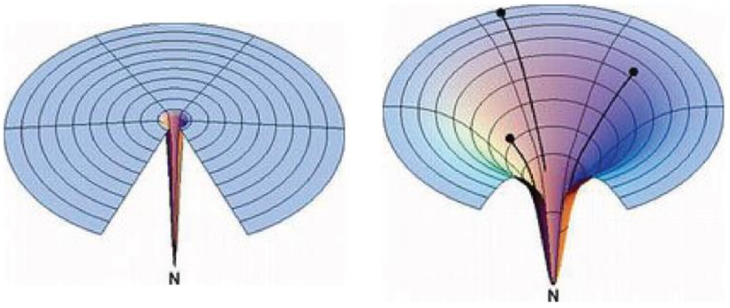
There are up to two ways by which the ΔG can be minimized at equilibrium: the change in enthalpy (ΔH), which is the sum of the internal energy (U) and the product of the pressure and volume (pV) at each state, can be minimized, and/or the entropy (S) can be maximized. Should the ΔG from the initial state to the final state be <0, the reaction (or more generally progress along that reaction coordinate) is said to occur "spontaneously" (*i.e.*, can be achieved without additional energy input). By convention, "favorable" interactions are negatively signed (*i.e.*, they always reduce the internal energy), while the ΔpV work should be approximately 0 for a constant pressure environment. In this way, the ΔG can be made negative through the use of either many weakly favorable interactions or a few very favorable interactions, creating spontaneous organization, as we will discuss further below. In complementary fashion, we may also reduce ΔG by having the entropy in the final state be higher than the initial state ($\Delta S > 0$). This is a natural consequence (or rather corollary of) the 2nd law of thermodynamics[13], which states that systems tend to increase their entropy over time, and that the equilibrium configuration is that which has maximized the entropy[12]. It may be the case that the final state has lower entropy than the initial one ($\Delta S < 0$), but within the confines of that state, the entropy is maximized at equilibrium. Needless to say, for a system to *spontaneously* evolve to a lower entropy state, this must be significantly offset by a significant enthalpic driving force, which we will see immediately in the subsequent section.

In summary, we can project chemical transformations along one or more different "reaction coordinates" and quantify the change in Gibbs free energy (ΔG)—a thermodynamic potential which itself is based largely on energetic and entropic terms—as a function of the coordinate value(s), which in many ways is simply a convenient way to think about probabilities. Now equipped with a theoretical basis for connecting microstate configurations to macroscale observables, we can begin to deconstruct the hierarchical organization required to obtain extended biological materials. To achieve our goal, it is instructive to start at the foundation: protein folding.

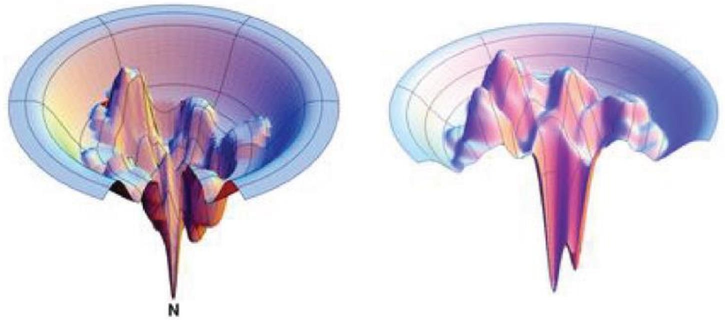
1.3.2 Protein folding

How can a protein "know" how to fold up into its final 3D structure based purely on its sequence, particularly when we (now) know that proteins are not simple repeating structures (as seen for other biopolymers like DNA)? Ignoring known exceptions such as chaperone-mediated folding and intrinsically disordered proteins, cellular polypeptides—as synthesized by ribosomes—fold spontaneously (*i.e.*, they must have a $\Delta G < 0$), rapidly, and very consistently, suggesting that the folding process is likely to not be just well-defined, but inevitable[14]. Such behaviour was well-established by early experiments into protein folding, which showed that unfolded proteins could regain their functional state even in the absence of additional factors[15], culminating in Anfinsen's dogma (the "thermodynamic hypothesis")[16], which proposed that the *correct* folded state of a protein corresponds to its free energy minimum. Simple enough. However, this thermodynamic explanation seems at odds with an intuitive expectation about folding kinetics from polymer physics, where we might consider that (in the worst possible case) all possible configurations must be explored before reaching the final folded structure[17]. Cyrus Levinthal posited this thought experiment in the late 1960s, pointing out that an average sized polypeptide of 150 amino acids (possessing then, say, three stable phi/psi angle conformations for each of its 149 peptide bonds) would have on the order of 3^{298} (10^{142}) possible configurations, an astronomical number that could hardly be sampled even at ultrafast rates[18]. This is "Levinthal's paradox". The resolution, as we shall presently see, is that the search from unfolded chain to final structure is not completely random, but rather guided through the formation of local favorable interactions[17,19], which serve as nucleation points during the folding process. These nucleation points are secondary structure elements, the next level in our hierarchy.

As we have established, proteins are not random polymers, however they are subject to the randomness associated with nanoscale fluctuations. The result of this is that there is rarely (if ever) a singular well-defined folding pathway associated with a particular protein. Instead, we can adopt a statistical treatment of the problem and instead consider the folding process as occurring over a landscape representing the protein's potential energy surface. The result, taking into account Anfinsen's dogma, is a "folding funnel", where the native structure resides within a deep well, separated from all other states. This perspective has been studied extensively by Ken Dill and coworkers[14,20], and several representative versions of the folding funnel are shown below (**Figure 1.4**). The landscape corresponding to Levinthal's paradox is shown at upper-left (aptly named the "golf funnel"): a totally flat surface over which the protein would randomly explore until landing in a single potential minimum corresponding to the native state. The idealized funnel for protein folding (the "smooth funnel") is shown as having a negative gradient everywhere leading to the global minimum, such that no matter where you are on this landscape, you're heading to the bottom. Most realistic, however, are the myriad other funnel shapes possessing numerous small pits, grooves, and traps, reflecting locations where partially folded or off-pathway intermediates accumulate before (hopefully) eventually making it to the native state(s). Increasing evidence from MD simulations[21], spectroscopy[22,23], and ultrafast laser experiments[24,25] have shown that these intermediates often include elements of secondary structure that are found in the final fold, but form only transiently up until the ultimate collapse into the final structure. Such concepts have been recently incorporated into folding funnel theory, with Rollins and Dill proposing the "foldon funnel model", where the instability of secondary structures leads to an uphill "volcano" shape,



Golf funnel **Smooth funnel**



Bumpy funnel **Bistable funnel**

Figure 1.4 | Selected illustrative protein folding funnels. These surfaces represent the potential energy surfaces of proteins during folding. The “golf funnel” corresponds to Levinthal’s paradox, while the “smooth funnel” is the idealized version for rapid folding. The remaining two are more representative of real proteins, which can have one or more minima associated with two interconverting native states. The bistable minimum resembles **Figure 1.3**. Images adapted from the Dill research group website.

except where stabilized by native-like tertiary interactions which help to ultimate guide the polymer chain to its destined conformation[26]. This partially explains phenomena such as the sickling of hemoglobin through a single amino acid mutation: the primary sequence defines the potential energy surface comprising the folding funnel, and if sufficiently perturbed, can lead to a radically different native state. Considering this phenomenon, one can see with greater clarity the connection between the primary structure (protein sequence) and tertiary structure (final globular fold), which is guided in part by the secondary structure elements ubiquitous in evolved proteins.

This is the relation between thermodynamics and hierarchical organization of individual protein structures.

A final note about the shape of folding funnels: these funnels are part-quantitative and part-qualitative, mostly serving to provide a descriptor for the overall process. However, if the depth axis is free energy, what about the radial shape of the funnel? This is taken to represent the entropy: unfolded polypeptides have significantly more conformational freedom, and thus much higher entropy, so there is a large “spread” of possible conformations[27]. Consequently, there are almost no intra-chain interactions to provide an enthalpic term to the free energy, so all of these comparably entropic conformations have similar free energies, leading to a relatively flat surface. As the chain collapses, new interactions form, reducing the internal energy and lowering the free energy. Thus, the enthalpy shares the depth axis with the free energy, being the negative driving force for the overall reduction. As this occurs though, there is of course a corresponding loss of entropy, which is considered minimized in the native state (for most proteins), where the structure typically only exhibits modest fluctuations about its equilibrium configuration. The takeaway is that the folding process, which typically only exhibits ΔG on the order of 10 kcal/mol, is in fact the sum of a huge trade-off between large entropic loss against a (slightly larger) enthalpic gain[27]. This enthalpy-entropy compensation is another ubiquitous biological phenomenon which we shall see reappear again and again in our investigations into protein biophysics.

1.3.3 Conformational dynamics of biomolecules

Despite all this talk of proteins adopting specified structures, do not be fooled into thinking proteins are intricate rocks. In addition to their ability to fold and unfold, a tremendous amount of functionality is the result of conformational changes (structural rearrangements) of proteins, usually involving relatively localized stretching, bending, twisting, or looping motions while generally retaining their well-packed hydrophobic interiors[28,29]. As we might expect, the nature of these motions and the residues which undertake them are necessarily dictated by the overall folded structure of the protein, which defines the energy landscape over which this structure can explore. Just as discussed for protein folding, one can imagine the energetic cost of breaking various secondary or tertiary interactions in order to obtain more mobility (higher entropy)[27]. These costs are typically lower for individual sidechains moving between alternative conformations (rotamers) or for already unstructured loops. Through this mechanism, proteins continuously perform a statistical fluctuating exploration of near-native structures, with occasional excursions (depending on the protein) to strikingly different configurations, which commonly encompass ones of functional relevance. Correspondingly, these are often coupled to ligand binding or posttranslational modification, which perturb the energy landscape to facilitate important rearrangements.

One manifestation of this effect is the origin of allosteric regulation in proteins, whereby local structural changes (usually in response to a stimulus) are transmitted through the body of the protein itself through coupled rearrangements of packed sidechains and/or mechanical forces such as strain. The oldest and most well-studied examples of this is hemoglobin[30], the tetrameric protein responsible for O_2 and CO_2 transport within your blood. Each subunit of this protein contains an iron-porphyrin (heme) group which is responsible for gas binding. The protein can exist in two conformations: T (tense) and R (relaxed), with the latter exhibiting much higher affinity for O_2 (**Figure 1.5**). Upon the binding of O_2 to one of the subunits, movement of the iron atom within the heme tugs on a histidine sidechain, adding additional tension to the attached helix and nudging the already strained structure towards the relaxed state. The result of this tug-of-war is communicated mechanically to the other subunits, bringing them closer to the R state (even without O_2 bound)

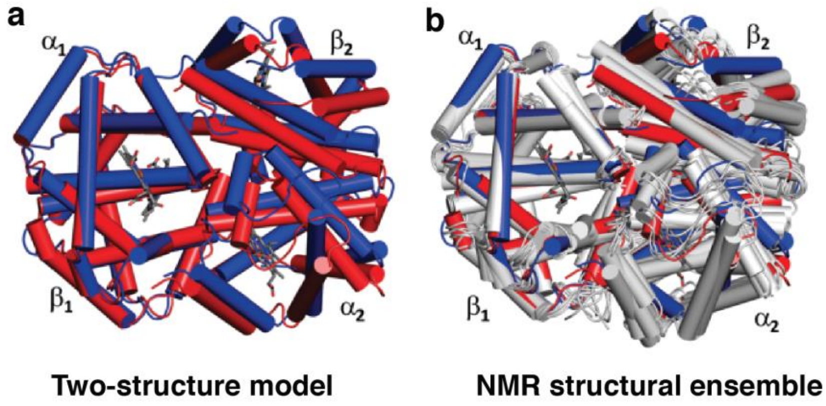


Figure 1.5 | Allosteric structural rearrangements in hemoglobin. a, Crystal structures of the unliganded T state (blue, PDB ID: 4HHB) and O_2-bound R state (red, PDB ID: 2DN3), where structures of the α_1/β_1 chains have been superimposed, highlighting the structural rearrangement. **b,** 10 lowest-energy structures as determined by NMR, similarly superimposed, revealing an ensemble of states connecting the two extremes. Adapted from ref. 30.

and increasing their O_2 affinity. This is an example of positive cooperativity, where single events make subsequent events more likely. What makes this feature so functionally elegant is that when O_2 is abundant (such as in the lungs), hemoglobin can rapidly saturate with its cargo, yet when O_2 is depleted (such as in the muscles), this cooperativity works in reverse to rapidly deposit O_2 (and exchange it for CO_2) to maintain cellular respiration. The evolution of cooperativity in this protein not only ensures that binding/unbinding kinetics are fast and occur where they're needed, it also ensures that hemoglobin works at maximum efficiency by exchanging the maximum number of gas molecules at each step. This highlights not only the ability of protein dynamics to provide important functional capabilities, but also how the particular structure enables this behavior to occur through long-range mechanical coupling between subunits, a hallmark of evolved proteins.

Excitingly, recent theoretical investigations have shown that the ability to generate functional motions through nonequilibrium switching between conformational free-energy surfaces (once thought to be the exclusive purview of molecular motors) is likely to be, in fact, a general phenomenon of enzymatic proteins[31]. What this means is that the free-energy landscape associated with functional motions is mutable (as depicted in **Figure 1.6**), typically switching during ligand binding/unbinding and/or catalytic steps. Within this formalism, one may now picture how the various landscapes of hemoglobin (structural and ligand-binding) might change during sequential O_2 binding events. Under nonequilibrium conditions, such as in the presence of a concentration gradient, these proteins can generate (mechanically or chemically) directional

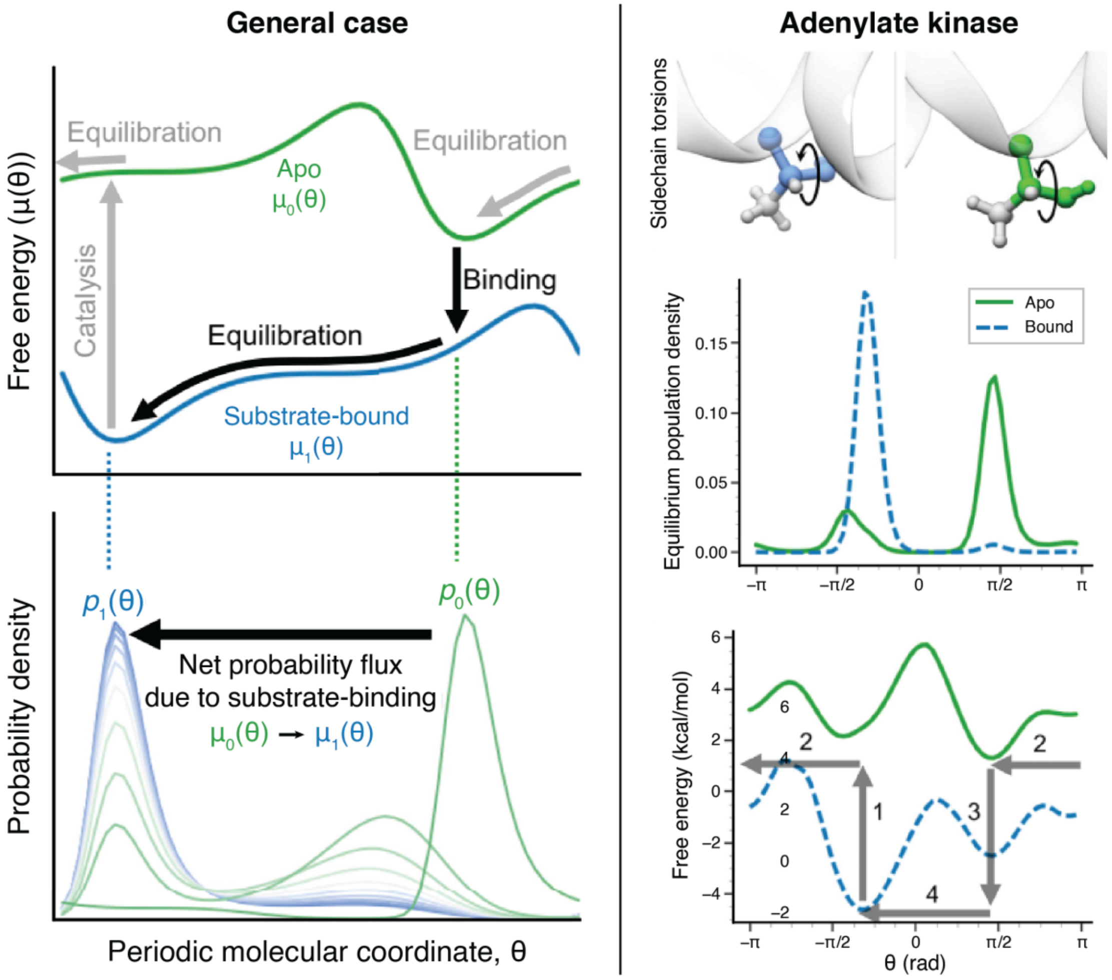


Figure 1.6 | Motor-like properties of nonmotor enzymes. Illustrative free energy surfaces for an enzyme in apo and holo forms (green and blue lines), depicting how ligand binding can reshape these surfaces such that equilibration enables continuous travel between them (**left**). Results of all-atom MD studies of adenylate kinase, which reveal a system-specific manifestation of the general case (**right**). Adapted from ref. 31.

motion through a continuous cycling of steps (driven by the forward process of substrate binding) which is continuously energetically downhill. This principle underscores the utility in employing a quantitative thermodynamic description of protein dynamics, which the author believes will come to represent a fundamental principle for the design of new proteins and protein-based materials which are dynamic/adaptable and stimuli-responsive.

1.3.4 Macromolecular self-assembly

Finally, we are nearing the top of the hierarchy. Before we can get to materials, we need to briefly discuss a topic which has been largely glossed over so far: self-assembly. As discussed above, the vast majority of proteins operate through the coming-together of multiple units (subsequently termed "subunits") into a larger complex, which typically exhibit new properties difficult to access by single chains. These complexes are obligate oligomers, meaning their assembly is not only spontaneous, but so energetically favorable that they do not come apart (unless it is functionally relevant) under non-denaturing conditions. Impressively, these oligomers form successfully in the crowded environment of the cell, highlighting the highly specific nature of the interactions; competition due to off-target pathways would be a significant burden on organismal resources. This affinity and specificity is typically achieved through the formation of complementarily shaped hydrophobic interfaces spanning several hundred square angstroms. The shape complementarity provides specificity, while the hydrophobic effect (the expulsion of water away from nonpolar surfaces, the same phenomenon that separates oil from water and produces the densely packed hydrophobic core of folded proteins) provides the strong driving forces required to form and retain the interaction. As a consequence of evolutionary constraints, most of the structures are symmetric, enabling the formation of homooligomeric complexes through the simple optimization one self-interacting interface. Among these structures we can find a tremendous diversity of geometries and oligomer size (**Figure 1.2**), enabling the formation of some of the most complex macromolecular machines (*e.g.*, ATP synthase), where the interacting of (mis)matched symmetric components forms an integral part of the chemomechanical mechanism. We shall see presently that the advantages of symmetry (so extensively exploited by evolution) also makes it one of the most powerful tools for the construction of extended crystalline materials.

1.3.5 Crystal nucleation and growth

As we stated towards the beginning of this dissertation, our goal is to understand how to manufacture new protein-based materials, ideally with unique emergent properties which we can understand from physical principles. To bridge from atomic/molecular to the micro/macro scales required for many materials to be useful, we need to co-organize a great number of oligomeric units. Unlike plastics, which are composed of flexible polymer chains interacting through weak noncovalent interactions, we seek to preserve the protein structure (and its corresponding function) in the aggregate material, and thus we turn to crystalline materials, which include common materials such as ice, metals, clays, gemstones, and the liquid crystals in flat-panel displays.

Crystalline protein assemblies have long been exploited for structure determination by X-ray crystallography (and more recently by microcrystal electron diffraction), where the periodic arrangement of atoms causes the high-energy waves to diffract into well-defined spot patterns that can be deconvoluted to yield the atomic coordinates of the repeating unit cell. Crystalline materials are defined by the translational periodicity of this unit cell, which typically contains symmetric arrangements of atoms or molecules within its volume. In total, all possible combinations of symmetry and translation operations produces 230 different space groups, which describe that specific arrangement of particles in 3D space (though periodic arrangements are also possible in lower dimensions). For proteins, which are chiral and thus restricted to non-inversion and non-mirror symmetry operations, this number is greatly reduced to only 65 possible space groups. What is important for our purposes is that the principle which is exploited in the generation of symmetric oligomers (that few interfaces have to be evolved/optimized simultaneously) can be readily generalized to the formation of space-filling extended materials (*i.e.*, appropriately placed interaction patches at crystallographic interfaces will promote the formation of a specific crystal lattice with minimal design). How these crystals actually form, however, is both an old subject and one that remains an area of active research. Let us briefly discuss the topic (which will be a significant component of Chapter 4) and some natural examples of crystalline protein assemblies.

Crystallization is most fundamentally described as a phase change from a dense, metastable, disordered liquid phase to a dense, more stable, ordered (and usually solid) phase[32]. Picturing this process in your mind's eye, imagine how the wiggling and jiggling of free molecules must come together and arrange themselves in a highly ordered manner. One may surmise that this should exhibit some interesting thermodynamics, as we must achieve the spontaneous formation of a (presumably) lower-entropy phase, which must be offset by some other driving forces. It is also likely clear that the earliest formation events ("nucleation") of the new phase are both rare and stochastic in nature[33]. This rarity is explained by the dynamic equilibrium experienced by the growing nucleus: not only do multiple units have to come together nearly simultaneously, the units on the outside of the nucleus can be dissolved back into solution, defining a critical nucleus size that must be reached before the new phase indeed becomes more stable than the preceding one (we shall see the details of this below). The oldest and most common theory describing this phenomenon is called "classical nucleation theory", where the kinetic rate of nucleation (R) is predicted by **Equation 1.3**:

$$R = N_s Z k \exp(-\beta \Delta G^*) \tag{1.3}$$

Here, ΔG^* is the free energy cost at the top of the nucleation barrier, N_s is the number of nucleation sites, Z is the "Zeldovich factor" (the forward probability of forming a new phase at the top of the nucleation barrier), k is the rate of particle attachment to the nucleus, and β is the inverse temperature. What we can see is that the rate most strongly (to a fault, it has been shown) depends on the barrier free-energy height, with large energies predictably leading to reduced rates, and the particle radius at which $\Delta G(\mathrm{r}) = \Delta G^*$ is the critical nucleus size. The origin of this effect is shown in **Figure 1.7 (left)** where the favorable (usually enthalpic) contributions to the free energies arise from the formation of interatomic/intermolecular interactions in the bulk phase, while there is an energetic penalty associated with the preservation of the phase-separating interface (the interfacial free energy). Together, these define the crystallization free-energy landscape associated with a classical crystallization mechanism. In solution, both terms depend on the particle radius, with the bulk phase ultimately winning out due to its cubic volume dependence vs quadratic surface area interfacial free energy. This relation for a spherical nucleus is provided in **Equation 1.4** (Δg is the free energy per unit volume and σ is the surface tension). By contrast, *heterogeneous nucleation*, where nucleation events are promoted by adsorption onto a surface or local defect, does so by reducing the exposed surface area (and thus the interfacial free energy) of the nucleus, greatly reducing ΔG^*. This effect is shown in **Figure 1.7 (right)**.

$$\Delta G(r) = \frac{4}{3}\pi r^3 \Delta g + 4\pi r^2 \sigma \tag{1.4}$$

A final point about classical nucleation theory is that it assumes that crystals nucleate and mature through monomer-by-monomer addition to the growing edges of the crystal (nucleus). It turns out that while this is one viable mechanism, a large number of alternative *nonclassical* pathways have been discovered in recent years, which can involve the initial formation of dense

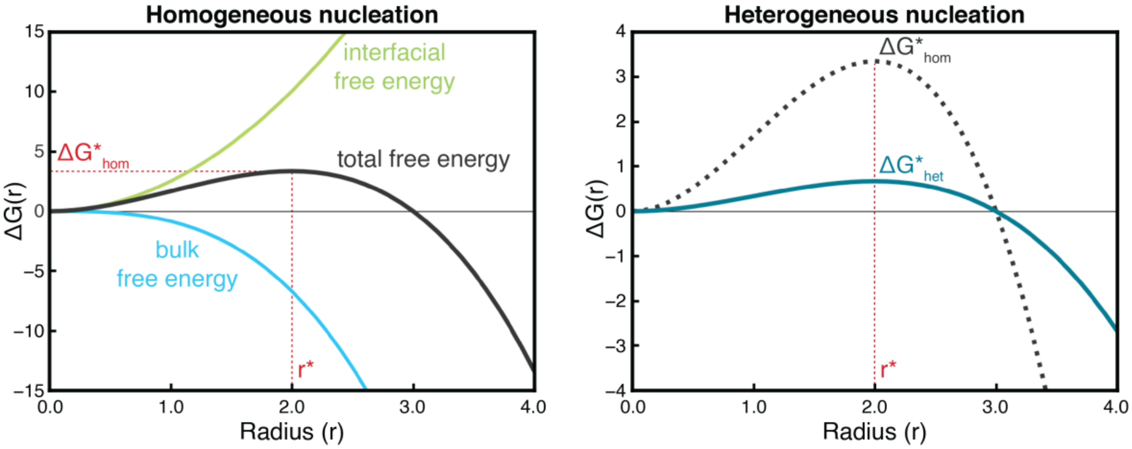


Figure 1.7 | Classical nucleation theory. Illustrative free-energy landscapes for crystallization processes as predicted by classical nucleation theory. The competition of bulk and interfacial terms gives rise to a peak of height ΔG* at the critical radius r* for the total energy (**left**). Heterogeneous nucleation greatly reduces the magnitude of the interfacial term, lowering the energy barrier to formation of the critical nucleus (**right**).

liquid phases (of varying degrees of order) that then spontaneously transform into crystalline domains, or the oriented attachment of small crystallites towards the formation of larger single crystals. The diversity of nonclassical crystallization mechanisms[33] and representations of their assembly are summarized in **Figure 1.8**, which will suffice for the purposes of this dissertation.

One example of classical crystal growth in nature are microtubules (MTs): 1D hollow cylinders composed of helical repeating units of the heterodimeric protein tubulin. Microtubules play an essential role as structural cytoskeletal elements and (among many other functions) double as highways over which transport proteins move cargo. As regulators of cellular structure and movement, these filaments need to grow and shrink in specified manners to meet essential organismal needs[34]. This is achieved through a classical (monomer-by-monomer) rapid growth/collapse mechanism—termed “dynamic instability”—that is synchronized through GTP hydrolysis by tubulin dimers, thus enabling precise chemical control over the self-assembly state (and thus cellular-scale consequences) of entire MTs[34]. The periodicity of the helical units means that microtubules are true 1D crystals, yet exhibit the remarkable ability to condense or shed 100's to 1000's of proteins per MT in a stimuli-responsive capacity.

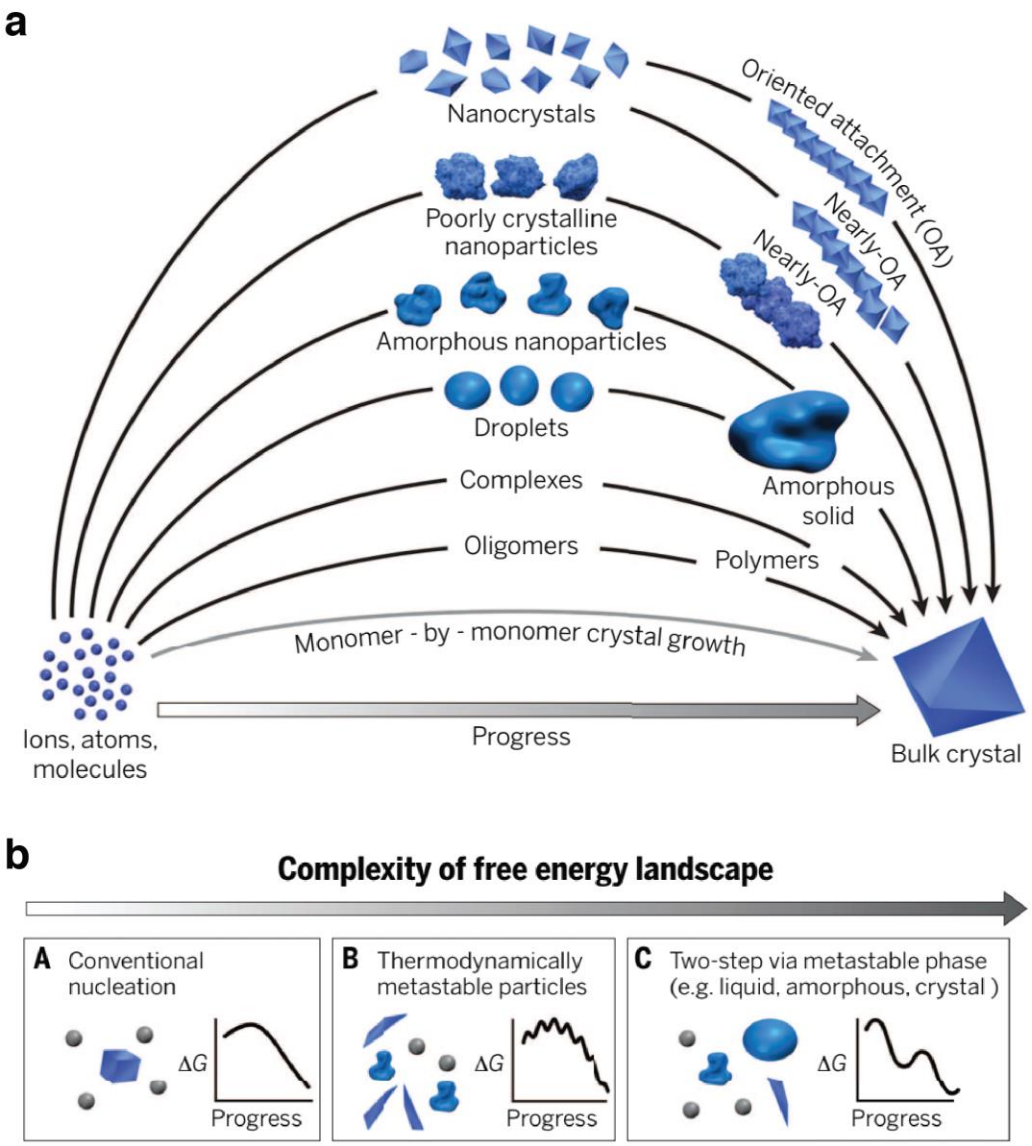


Figure 1.8 | Crystal nucleation and growth. a, Scheme depicting various modes of crystallization (classical is monomer-by-monomer, the remainder are nonclassical). **b,** Illustrative free energy landscapes for different crystallization pathways, which become more complex as the process becomes more dynamic. Adapted from ref. 32.

In contrast, there is a growing body of evidence that S-layer proteins, 2D protein arrays secreted to the outer surfaces of many bacteria and archaea, crystallize to form a protective shell around the entire cell through nonclassical growth mechanisms. These outer layers come in a variety of plane symmetry ("wallpaper") groups, and serve to both protect the organism and participate in intercellular recognition[9]. The functional need to maximize surface coverage while not interfering with cellular motility and division requires that these crystals be adaptable to both flexible cell shapes and total surface area. This need has been met by the emergence of a "continuous crystallization" mechanism that occurs as proteins as migrated to the cell surface *in vivo*[35]. Furthermore, *in vitro* characterization of S-layer crystallization on mineral or graphene surfaces has shown that they first adsorb to the surface as disordered clusters, before undergoing a concerted rearrangement into the final crystal lattice, sometimes concurrently with a conformational change in the subunits[36,37]. These capabilities are in stark opposition to the prototypical stiff, brittle, crystal (*e.g.*, clays, graphite), and highlights the materials properties enabled by the flexibility and tunability of protein structures.

Yet perhaps there is no better exemplar of a hierarchically organized semi-crystalline protein material than spider silk. Long touted for its high tensile strength, ductility, and toughness, this material owes both its flexibility and its strength to protein-protein interactions[38]. "Spinning" of silk fibers involves pultrusion of silk peptides through a pH gradient (decreasing from 8 to 5) across the length of the spider silk gland. The low pH and high peptide concentrations leads to amyloidation, forming β-sheet secondary structure motifs that then stack upon each other to form 3D crystals. These ultra-stiff crystals are linked together by unstructured polypeptide loops, which provides flexibility to the aggregate material. Together, these two components comprise individual silk fibrils, which are then bundled to form the individual spider web strands that ultimately perform many famous functions for our eight-legged friends. In purely physical terms, this material is strong, yet flexible, functionally and structurally diverse, and forms on-demand in a chemical-responsive fashion: truly a wonder material, and only possible with proteins.

In the next and penultimate section, we will overview methods developed to construct higher-order assemblies (from discrete oligomers to extended crystals) and the properties which are endowed by their specific structures.

1.4 Design methods for constructing supramolecular assemblies

1.4.1 “Biomimetic” self-assembly

Artificial design methods for enacting self-assembly span a spectrum of philosophies, ranging from emulating natural interfaces to more “abstracted” interactions; we will start with the former. As discussed above, the interfaces between evolved protein interaction partners are typically complementarily shaped hydrophobic surfaces, comprising a large number of relatively weak van der Waals interactions that collectively exhibit the hydrophobic effect. The plurality of individual interactions betrays their evolved nature, as the precise placement of all sidechains requires an understanding of the full protein folding hierarchy. While this places this design philosophy largely out of the reach of human designers, this can be sidestepped through the creative manipulation of native protein oligomers, or directly addressed by computational protein design methods, which have undergone massive development and refinement in recent years and now regularly produce well-predicted protein structures, folds, and interfaces, totally *de novo*.

Starting from the straightforward repurposing of natural proteins, small one- and two-dimensional extended materials (“crysalins”) were shown to be achievable by Sinclair, *et al.* through the generation of fusion proteins with matching rotational symmetry[39]. To do so, extant proteins with D_2 or D_4 point group symmetries were connected via C_2 or D_2-symmetric linkers to form 1D (D_2 + D_2) or 2D (D_4 + D_2 or D_4 + C_2) structures (**Figure 1.9a**). These linkages could either be formed via genetic fusion of self-associating coiled-coil motifs, or through tagging of the building block with high-affinity Strep tags (which are unstructured until complex formation with D_2-symmetric streptavidin). A similar strategy was demonstrated by Padilla, Colovos, and Yeates, where natively symmetric oligomeric proteins were fused together via rigid α-helical spacers to generate protein-based nanomaterials[40], including helical bundles and cages, though they noted that the strategy would also be readily extendible to 2D/3D structures (**Figure 1.9b**). Moving away from extant proteins, computational protein design efforts were successfully employed (more than a decade later) to generate several novel proteins that self-assembled into 2D protein arrays of varying plane group symmetries[41] (**Figure 1.9c**). This not only further demonstrated the principle of exploiting symmetry to form protein oligomers and extended materials, but represented a significant advance in *de novo* protein interface design, though only 4 out of 62 designs successfully formed. This, compounded with the limited size of the crysalins described above, highlights the inherent difficulty associated with using large, hydrophobic interfaces to drive the self-assembly of extended materials: the association process is largely uncontrollable, and the stability of the final complexes are such that there is no opportunity for error correction, an essential process for the successful construction of crystalline assemblies from stochastic nucleation and growth events. Computational protein design is, however, amenable to the formation of discrete oligomeric complexes[42,43] which are not restricted by the requirement to form correct nearest-neighbor interactions with near-infinite repetition. To highlight one example, the concept of symmetry matching has been extended to icosahedral cages using *de novo* designed dimeric, trimeric, and pentameric blocks to form large protein enclosures up to 40 nm in diameter

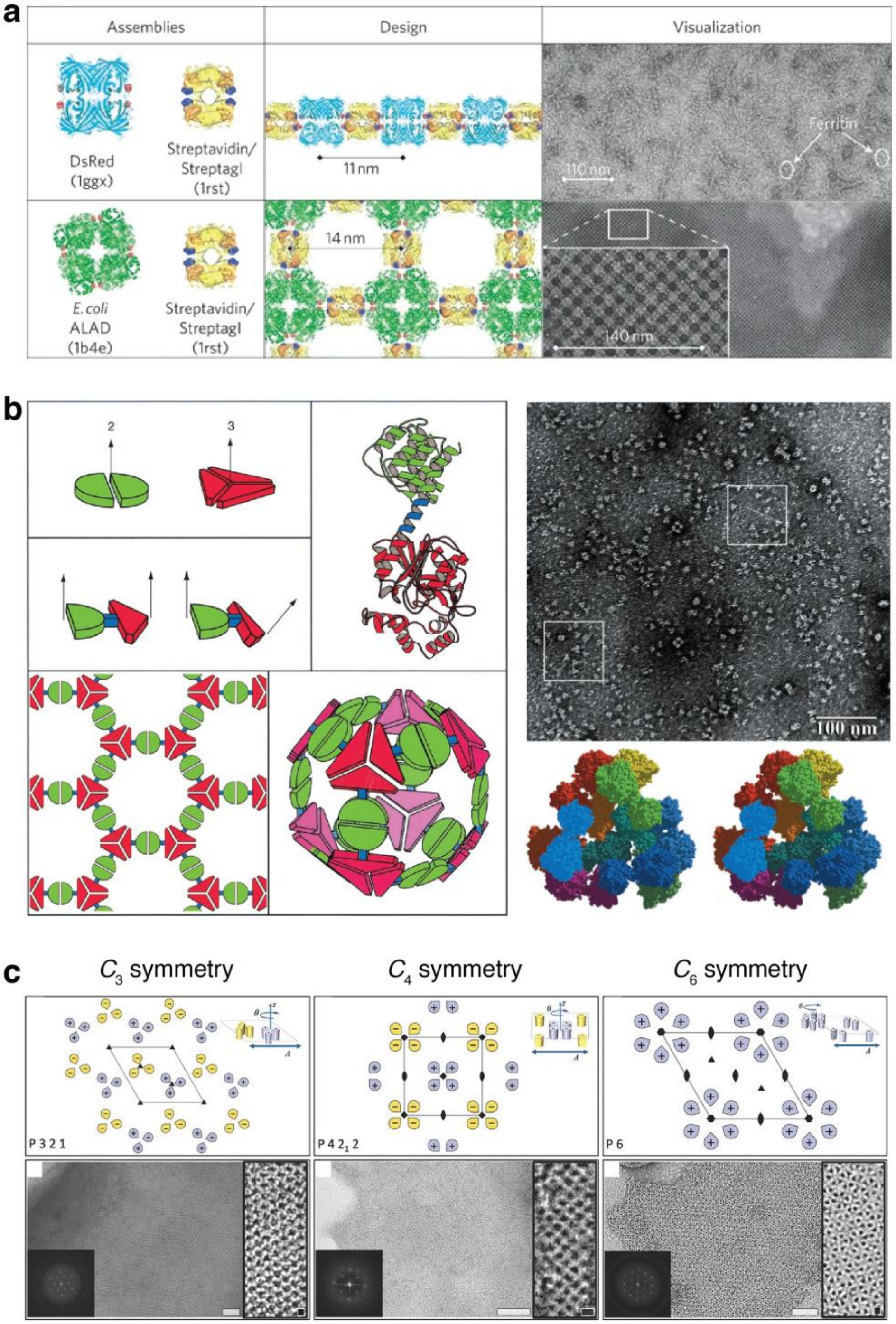


Figure 1.9 | "Biomimetic" self-assembly of proteins. a, Strategies for generating 1D/2D "crysalin" assemblies from protein fusions. Adapted from ref. 39. **b,** Geometric principles of symmetry matching to generate extended arrays or discrete cages, depending on hinge angle of the linker. TEM image and models of cage structures shown at right. Adapted from ref. 40. **c,** *De novo* designed 2D lattices which form via hydrophobic interactions, demonstrating the use of symmetry to generate different lattice structures. Adapted from ref. 41.

reproducing three distinct architectures[44], highlighting the substantial advances being continuously made in the field of computational protein design[45].

While the above designs are exceptionally stable to high temperature and chemical denaturants (owing to the well-packed hydrophobic cores which are output by the design software), this also makes their structures static and “rock-like”, in stark contrast to evolved proteins. However, there have been efforts to generate functional protein dynamics computationally as well. In a seminal paper by DeGrado and coworkers, a transmembrane four-helix bundle dubbed “Rocker” was computationally designed to adopt two different conformations which position the flared geometry to each side of the membrane, as seen in natural ion channels[46]. This bundle, which features hydrophobic interactions to stabilize its structure in the membrane, also contains ionizable glutamate and histidine sidechains towards its interior, which can exhibit coupled (de)protonation and metal-ion binding configurations. Through efforts to design bistability into this structure, the resulting construct was shown to be a Zn^{2+}/H^{+} antiporter channel, where concentration gradients of one class of ion can be used to drive transport of the other, analogous to evolved antiporters. More recently, new protein backbones have been reported which mimic the viral protein hemagluttinin, and exhibit comparable pH-responsive behavior due to extensive H-bond networks involving histidine residues[47]. These examples promise a bright future ahead for computational design techniques, though there is plenty of room to be complemented or augmented by the other philosophies.

1.4.2 “Bioinspired” self-assembly

One of the most distinguishing features of proteins are their ability to recognize and bind molecules with very high specificity. For the purposes of producing oligomers or materials, this means that molecular binding partners can be similarly exploited to bridge multiple proteins, as described for the all-protein systems in the previous section. Following the current theme, computational protein design was recently used to produce a chemically induced dimerization (CID) biological switch from two previously non-interacting protein partners[48]. To do so, a new binding site was designed *de novo* into a docked heterodimeric protein model such that the dimeric complex would only form in the presence of the small-molecule ligand. The result was a protein sense-response system which selectively reported on the presence of farnesyl pyrophosphate, an important metabolic intermediate found within many biosynthetic pathways for terpenoid compounds. This promises to be a generalizable strategy for the formation of other responsive oligomerization systems that can link biological output to detectable signals.

For materials, an early demonstration using ligand binding was reported by Freeman and coworkers, who used concanavalin A (conA), a tetrahedral lectin protein which natively binds sugar molecules, and a ditopic bismannopyranoside ligand to produce lattices which extend in all three dimensions[49] (**Figure 1.10a**). Analogously, 1D chains have been produced by Hayashi and coworkers using heme-globin interactions, where apo-hemoproteins (cytochrome b_{562}) were covalently labeled with free heme moieties, enabling them to form continuous head-to-tail interactions[50] (**Figure 1.10b**). Rounding out the dimensionalities, 2D lattices have also been generated using the C_4-symmetric protein L-Rhamnulose-1-phosphate aldolase (RhuA; the protagonist of this dissertation), which are labeled with the small molecule biotin and bridged using the D_2-symmetric streptavidin[51] (**Figure 1.10c**). However, the ultrahigh affinity of streptavidin for biotin ($k_d \sim 10^{-14}$) means these lattices are extremely limited in size, suffering from the same restrictions of near-irreversibility as described for hydrophobic interfaces.

We can also turn to DNA, a biological macromolecule coveted for the high tunability of its assembly owing to the complementarity of Watson-Crick base pairing, as a driver of protein

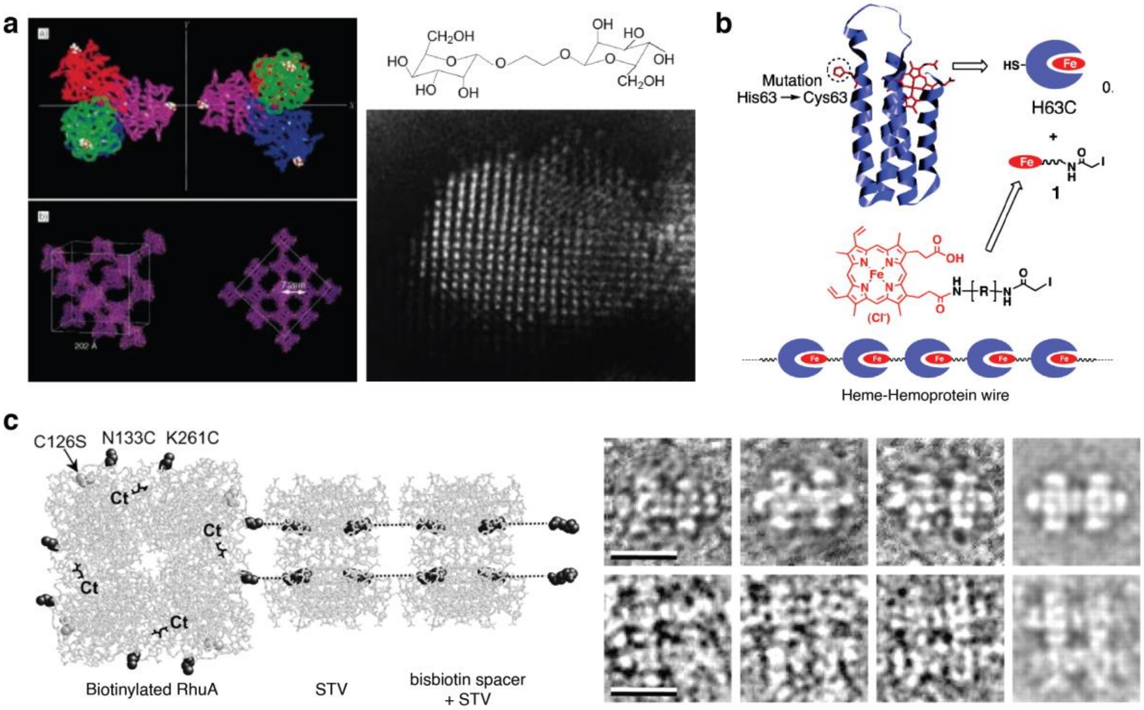


Figure 1.10 | “Bioinspired” self-assembly of proteins. a, Modeled assembly (left) of conA into diamond-like lattices (bottom right) via interaction with ditopic sugar linkers (top right). Adapted from ref. 49. **b,** Assembly of 1D nanowires via continuous head-to-tail binding of apo-hemoprotein to heme moieties covalently attached to the protein scaffold. Adapted from ref. 50. **c,** Assembly of small 2D arrays using biotin-labeled tetrameric RhuA, which are then organized by high-affinity strepatividin-biotin interactions. Adapted from ref. 51.

self-assembly. Nucleic acids are integral components of nucleosomes (responsible for chromosome condensation within cells), and are the targets of transcription factors, proteins which bind to DNA with high sequence specificity to regulate the production of mRNA. Building on the principles underpinning spherical nucleic acids[52,53], efforts from the Mirkin lab have shown that globular proteins can be substituted for nanoparticles, enabling the organization of these proteins into 3D lattices whose lattice parameters and thermostability are dictated by the length and sequence of DNA linkers[54,55]. However, in this case the material is almost entirely dictated by the nucleic acid interactions, with proteins merely serving as structural nodes that offer no chemical specificity. A recent report from our group has brought the capability of utilizing multimodal interactions to generate a protein-DNA material to the fore. Here, a designed variant of the protein cytochrome *cb*$_{562}$ (to be expanded upon in the following subsection), which possesses numerous metal-binding residues on its surface, was covalently modified with complementary single-stranded DNA such that they could dimerize[56]. At low pH and temperature, large 2D crystals spanning multiple square microns could be obtained within a few hours. Extensive structural modeling, biochemical characterization, and MD simulations revealed that the resulting material was the product of neatly balanced energetic drivers. This material could be assembled and disassembled in response to numerous stimuli, as would be expected of a natural protein-NA assembly, where processes such as transcription regulation must be precisely tuned. This was only possible through the presence of orthogonal interactions (metal binding, H-bonding, salt bridges) which are of comparable collective interaction strength to DNA hybridization. The careful utility of strong, reversible interactions such as these will be the final topic of this chapter.

1.4.3 "Chemical" self-assembly

While the computational design of noncovalent protein interfaces has advanced tremendously, there remain significant hurdles to its utility for the construction of protein assemblies (as described above). To address this difficulty, "chemical" design strategies utilizing strong, reversible interactions to achieve varying degrees of organization have been developed. Most prominently among these are the use of metal-ion chelation and disulfide bonding to drive protein association. While the examples of these are numerous (particularly for the latter), we will briefly discuss several examples from our group which exemplify these principles.

Transition metal ions serve essential roles as catalytic cofactors and structural elements for many folded proteins. These ions are typically chelated by nitrogen- or oxygen-based ligands, exhibiting different preferences for their local chemical environment and geometry depending on ion identity and oxidation state. Taking inspiration from coordination chemistry, one can imagine entire protein units as giant ligands for these ions, where contributions of >1 sidechain contribute to the chelate effect, which lends considerable strength to the interaction. This philosophy was formalized in our lab as metal-directed protein self-assembly (MDPSA)[57] (**Figure 1.11a**). As a proof-of-principle, the four-helix bundle protein cytochrome *cb*$_{562}$ was modified with histidine residues at two pairs of *i* and *i*+4 positions along one helix to yield a construct dubbed metal-binding protein construct 1 (MPBC1)[58,59]. Upon exposure to Zn^{2+} ions, this protein was found to oligomerize into an antiparallel-packed homotetramer, with each Zn^{2+} ion being coordinated by 4 histidine sidechains. Substitution of Zn^{2+} by metal ions with different coordination preferences (Cu^{2+} and Ni^{2+}) were shown to yield alternative oligomeric states[60] (dimers and trimers, respectively), underscoring the modularity and tunability of the strategy while requiring minimal modifications to an existing protein scaffold (**Figure 1.11b**). This technique can also be combined with computational design in a process known as metal-templated interface redesign (MeTIR)[57]. This process simplifies the computational design problem by providing a well-defined interface (based on metal binding) that can be optimized computationally, rather than attempting to jump to the final structure directly. MeTIR was performed on the MBPC1 scaffold to yield several variants termed RIDC*[61]. RIDC1 possesses a noncovalent interface which can form without the metal (producing a dimer), which can then undergo subsequent metal binding to recreate the original tetrameric structure. Additional variants of RIDC constructs containing a diversity of metal binding motifs have been shown to form 1D nanotubes[62], 2D arrays[63,64] (**Figure 1.11c**), and discrete oligomers[65] which can exhibit catalytic activity *in vivo*[66] (**Figure 1.11d**), highlighting the utility of MDPSA and MeTIR for designing self-assembling proteins.

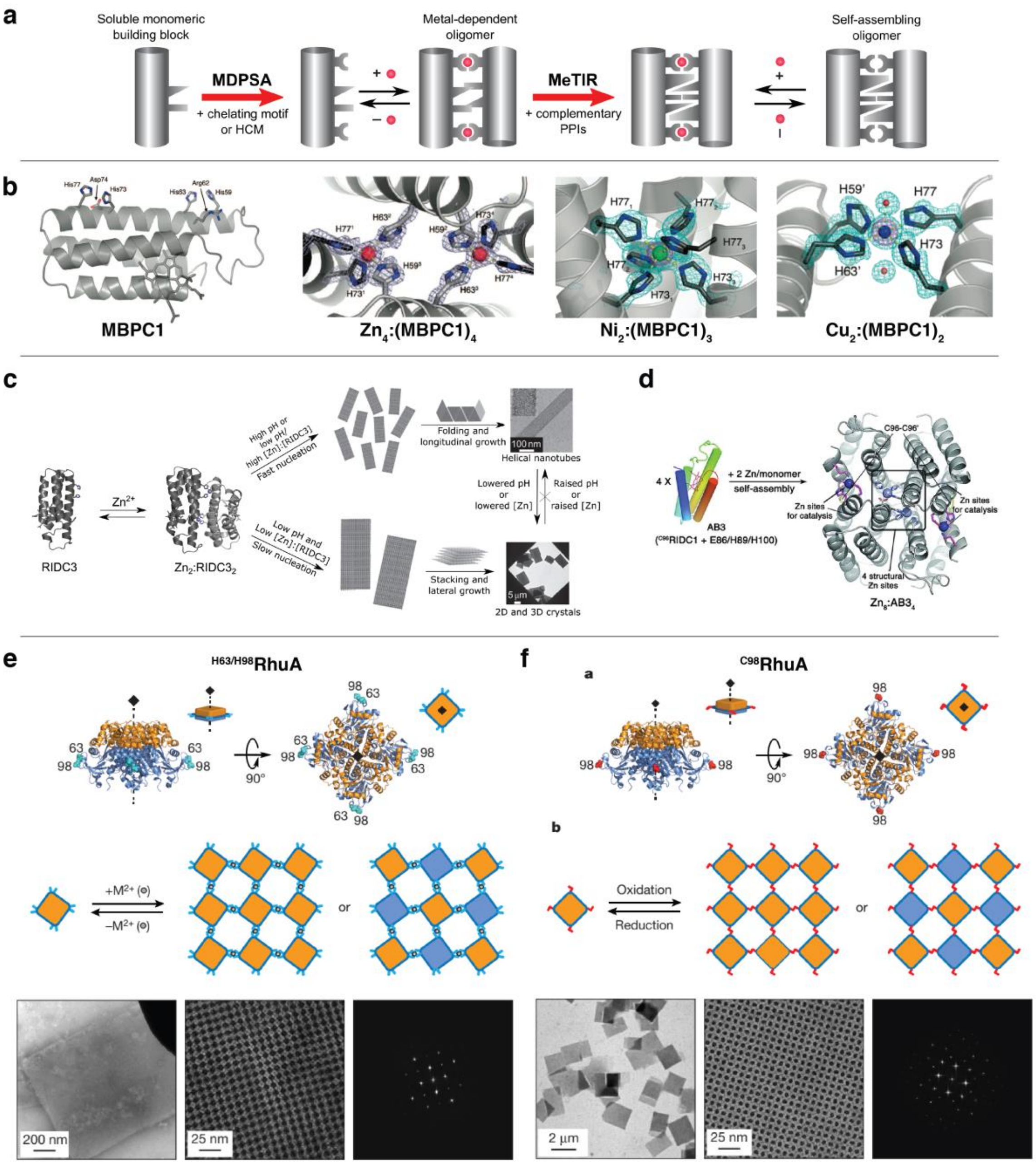


Figure 1.11 | "Chemical" self-assembly of proteins. a, MDPSA and MeTIR metal-binding-based methods for self-assembly. Adapted from ref. 57. **b,** Assembly of MBPC1 into oligomers of different symmetry based on metal identity. Adapted from refs 58 and 60. **c,** Assembly of RIDC3 into 2D crystals or 1D nanotubes, controlled by solution conditions. Adapted from ref. 63. **d,** Overview of AB3, an RIDC1-based artificial Zn-hydrolase with *in vivo* activity possessing structural and catalytic Zn sites. Adapted from ref. 66. Self-assembly of RhuA into square lattices using **e,** metal-binding, and **f,** disulfide bonding. Adapted from ref. 69.

Despite metal-binding being a relatively abstracted method of oligomerization, these complexes can still exhibit some interesting emergent behavior of evolved proteins. In one example, a triple-cysteine variant of RIDC1 was shown to form the canonical tetrameric structure in solution, positioning each pair of cysteines such that they formed disulfide bonds in a specified arrangement[67] (not dissimilar to structural disulfide bonds commonly found in natural proteins). In this configuration, with all metal-binding sites saturated with Zn^{2+}, the protein structure was in many ways unremarkable: there was a well-packed hydrophobic core, standard disulfide bond angles and lengths, and normal metal coordination. However, upon removal of Zn^{2+} ions from the complex, the structure became enormously strained, ultimately resulting the selective rupture of one particular disulfide bond, leading to a hinging "butterfly" motion of the two dimer pairs about the greasy hydrophobic core[67]. Subsequent MD simulations showed that this was ultimately an enthalpy/entropy trade-off between the conformational entropy of the whole structure and the enthalpic gain from repacking of the hydrophobic core[68]. Importantly, the coupling of metal-binding to covalent-bond hydrolysis and global structural rearrangement indicated that the resulting complex was, in fact, an allosteric supramolecular assembly, a true rarity for designed proteins.

These chemical design methods are not limited to the cytochrome scaffold so extensively discussed so far. We shall end with the protein RhuA, which will be our topic of focus for the remainder of this dissertation. RhuA, which is square-shaped, can easily be equipped with either metal-binding motifs or cysteine residues at its four corners, enabling the formation of large. porous 2D crystals free in solution[69]. The bis-His variant of RhuA can self-assemble in the presence of either Zn^{2+} or Cu^{2+} ions (**Figure 1.11e**), though the latter is prone to aggregation. The double mutant $^{F88/C98}$RhuA, in which the secondary mutation causes RhuA to become a D_4-

symmetric octamer, forms slightly larger crystals with a significant number of grain boundaries owing to the large number of cysteines, which increases the frequency of forming mismatched pairs. In contrast, the single-cysteine variant ^{C98}RhuA assembles into crystals on the order of 4–5 μm^2 in area with an extremely low defect frequency (**Figure 1.11f**). Upon self-assembly, the topology and flexibility of the intermolecular disulfide bonds enable the resulting architecture to undergo “breathing” motions that allow it to switch between porous and close-packed conformations without losing crystallinity.

Unlike common materials (*e.g.*, rubber), which shrink in one dimension as they are stretched along another, the changes in individual RhuA crystal dimensions are equal along both X and Y axes, rendering these crystal auxetic[69]. This property makes these materials intriguing candidates for a variety of applications as molecular displays with tunable densities, nanomembranes with variable porosity, or as mechanical shock absorbers. Consequently, understanding and manipulating the thermodynamics underlying self-assembly and conformational dynamics of these protein lattices (and the achievement of new applications through this knowledge) represent the overarching goals of the research described in this dissertation.

1.5 Dissertation objectives

The principal objectives of this dissertation are three-fold:

1. Characterize the thermodynamic landscapes underpinning the self-assembly (both in solution and on surfaces) and conformational dynamics of 2D protein-based materials
2. Develop and carry out rational engineering strategies which can be used to modify these behaviors in a predictable manner

3. Leverage this understanding to enable the development of new applications for our primary material of interest: a covalently bonded, porous 2D protein assembly

Proteins are highly tunable and monodisperse building blocks for the construction of new materials. Their self-assembly into higher-order structures greatly enhances their capabilities, a consequence of "emergent behavior" whereby the particular structure, topology, and connectivity of assembled proteins leads to the spontaneous appearance of new physics. Understanding how such properties emerge at the macroscale from microscopic chemistry, as well as how we can construct these desired architectures in a bottom-up fashion, are important and unresolved questions of great importance for chemistry, biophysics, and materials science. It is also clear that achieving this understanding will require a multi-faceted philosophy which is still nebulous at this time.

Thus, here we aim to demonstrate that applying the principles of thermodynamic energy landscapes provides a common theoretical framework for comprehending complex biophysical problems which subsequently enable the rational (re-)design of the moving parts to achieve new ends. We show that this redesign can be used to identify the thermodynamic driving forces underlying the conformational dynamics of our 2D protein material, which can then be sidestepped to provide chemical control over this process. We report significant progress towards adapting this redesigned material for new applications as an HCN-selective sensor upon integration with an inorganic photonic-crystal substrate. We then describe comprehensive theoretical and experimental investigations into the self-assembly of this protein into multiple distinct 2D architectures, both in solution and on mineral interfaces, which exhibit novel physicochemical properties. Finally, we summarize initial efforts to understand the physical patterning of this protein on different mineral surfaces, providing fresh insight into epitaxial growth on solid substrates between two highly complex (but of great biotechnological relevance) targets.

1.6 References

1. Ali, M. H. & Imperiali, B. Protein oligomerization: How and why. *Bioorg. Med. Chem.* **13**, 5013-5020, (2005).

2. Goodsell, D. S. & Olson, A. J. Structural symmetry and protein function. *Annu. Rev. Biophys. Struct.* **29**, 105–153, (2000).

3. McEvoy, J. P. & Brudvig, G. W. Water-Splitting Chemistry of Photosystem II. *Chem. Rev.* **106**, 4455-4483, (2006).

4. Junge, W. & Nelson, N. ATP Synthase. *Annu. Rev. Biochem.* **84**, 631-657, (2015).

5. Rutledge, H. L. & Tezcan, F. A. Electron Transfer in Nitrogenase. *Chem. Rev.*, (2020).

6. Hall, A. Rho GTPases and the Actin Cytoskeleton. *Science* **279**, 509-514, (1998).

7. Janke, C. & Chloë Bulinski, J. Post-translational regulation of the microtubule cytoskeleton: mechanisms and functions. *Nat. Rev. Mol. Cell Biol.* **12**, 773-786, (2011).

8. Freeman, J. A., Cheshire, L. B. & MacRae, T. H. Epithelial morphogenesis in developing Artemia: The role of cell replication, cell shape change, and the cytoskeleton. *Dev. Biol.* **152**, 279-292, (1992).

9. Sleytr, U. B., Schuster, B., Egelseer, E.-M. & Pum, D. S-layers: principles and applications. *FEMS Microbiol. Rev.* **38**, 823-864, (2014).

10. Glotzer, S. C. & Solomon, M. J. Anisotropy of building blocks and their assembly into complex structures. *Nature* **6**, 557-562, (2007).

11. Thorkelsson, K., Bai, P. & Xu, T. Self-assembly and applications of anisotropic nanomaterials: A review. *Nano Today* **10**, 48-66, (2015).

12. Gibbs, J. W. On the Equilibrium of Heterogeneous Substances. *Trans. Conn. Acad. Arts Sci.* **3**, 108-248, (1876).

13. Clausius, R. *The Mechanical Theory of Heat: with its Applications to the Steam Engine and to Physical Properties of Bodies*. (London: John van Voorst, 1 Paternoster Row. MDCCCLXVII, 1867).

14. Dill, K. A. & MacCallum, J. L. The Protein-Folding Problem, 50 Years On. *Science* **338**, 1042-1046, (2012).

15. Anfinsen, C. B., Haber, E., Sela, M. & White, F. H. The Kinetics of Formation of Native Ribonuclease During Oxidation of the Reduced Polypeptide Chain. *Proc. Natl. Acad. Sci. U.S.A.* **47**, 1309-1314, (1961).

16. Anfinsen, C. B. Principles that Govern the Folding of Protein Chains. *Science* **181**, 223-230, (1973).

17. Levinthal, C. Are there pathways for protein folding? *J. Chim. Phys.* **65**, 44-45, (1968).

18. Levinthal, C. *How to Fold Graciously.* In: "Mossbauer Spectroscopy in Biological Systems." Proceedings of a meeting held at Allerton house, Monticello, Illinois, (1969).

19. Kauzmann, W. Some Factors in the Interpretation of Protein. *Adv. Protein Chem.* **14**, 1-63, (1959).

20. Dill, K. A. & Chan, H. S. From Levinthal to pathways to funnels. *Nat. Struct. Mol. Biol.* **4**, 10-19, (1997).

21. Piana, S., Lindorff-Larsen, K. & Shaw, D. E. Atomic-level description of ubiquitin folding. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 5915-5920, (2013).

22. Sivaraman, T., Arrington, C. B. & Robertson, A. D. Kinetics of unfolding and folding from amide hydrogen exchange in native ubiquitin. *Nat. Struct. Mol. Biol.* **8**, 331-333, (2001).

23. Schanda, P., Forge, V. & Brutscher, B. Protein folding and unfolding studied at atomic resolution by fast two-dimensional NMR spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 11257-11262, (2007).

24. Ballew, R. M., Sabelko, J. & Gruebele, M. Direct observation of fast protein folding: the initial collapse of apomyoglobin. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 5759-5764, (1996).

25. Ballew, R. M., Sabelko, J. & Gruebele, M. Observation of distinct nanosecond and microsecond protein folding events. *Nat. Struct. Mol. Biol.* **3**, 923-926, (1996).

26. Rollins, G. C. & Dill, K. A. General Mechanism of Two-State Protein Folding Kinetics. *J. Am. Chem. Soc.* **136**, 11420-11427, (2014).

27. Onuchic, J. N., and, Z. L.-S. & Wolynes, P. G. Theory of Protein Folding: The Energy Landscape Perspective. *Annu. Rev. Phys. Chem.* **48**, 545-600, (1997).

28. McCammon, J. A. & Harvey, S. C. *Dynamics of Proteins and Nucleic Acids*. (Cambridge University Press, 1988).

29. McCammon, J. A., Gelin, B. R. & Karplus, M. Dynamics of folded proteins. *Nature* **267**, 585-590, (1977).

30. Yuan, Y., Tam, M. F., Simplaceanu, V. & Ho, C. New Look at Hemoglobin Allostery. *Chem. Rev.* **115**, 1702-1724, (2015).

31. Slochower, D. R. & Gilson, M. K. Motor-like Properties of Nonmotor Enzymes. *Biophys. J.* **114**, 2174-2179, (2018).

32. De Yoreo, J. J. & Vekilov, P. G. Principles of Crystal Nucleation and Growth. *Rev. Mineral Geochem.* **54**, 57-93, (2003).

33. De Yoreo, J. J., Gilbert, P. U. P. A., Sommerdijk, N. A. J. M., Penn, R. L., Whitelam, S., Joester, D., Zhang, H., Rimer, J. D., Navrotsky, A., Banfield, J. F., Wallace, A. F., Michel, F. M., Meldrum, F. C., Cölfen, H. & Dove, P. M. Crystallization by particle attachment in synthetic, biogenic, and geologic environments. *Science* **349**, aaa6760, (2015).

34. Hess, H. & Ross, J. L. Non-equilibrium assembly of microtubules: from molecules to autonomous chemical robots. *Chem. Soc. Rev.* **46**, 5570-5587, (2017).

35. Comerci, C. J., Herrmann, J., Yoon, J., Jabbarpour, F., Zhou, X., Nomellini, J. F., Smit, J., Shapiro, L., Wakatsuki, S. & Moerner, W. E. Topologically-guided continuous protein crystallization controls bacterial surface layer self-assembly. *Nat. Commun.* **10**, 2731, (2019).

36. Chung, S., Shin, S.-H., Bertozzi, C. R. & De Yoreo, J. J. Self-catalyzed growth of S layers via an amorphous-to-crystalline transition limited by folding kinetics. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 16536-16541, (2010).

37. Shin, S.-H., Chung, S., Sanii, B., Comolli, L. R., Bertozzi, C. R. & De Yoreo, J. J. 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).

38. Vollrath, F. & Knight, D. P. Liquid crystalline spinning of spider silk. *Nature* **410**, 541-548, (2001).

39. Sinclair, J. C., Davies, K. M., Venien-Bryan, C. & Noble, M. E. M. Generation of protein lattices by fusing proteins with matching rotational symmetry. *Nat. Nanotechnol.* **6**, 558–562, (2011).

40. Padilla, J. E., Colovos, C. & Yeates, T. O. Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2217–2221, (2001).

41. Gonen, S., DiMaio, F., Gonen, T. & Baker, D. Design of ordered two-dimensional arrays mediated by noncovalent protein-protein interfaces. *Science* **348**, 1365-1368, (2015).

42. King, N. P., Sheffler, W., Sawaya, M. R., Vollmar, B. S., Sumida, J. P., André, I., Gonen, T., Yeates, T. O. & Baker, D. Computational Design of Self-Assembling Protein Nanomaterials with Atomic Level Accuracy. *Science* **336**, 1171-1174, (2012).

43. King, N. P., Bale, J. B., Sheffler, W., McNamara, D. E., Gonen, S., Gonen, T., Yeates, T. O. & Baker, D. Accurate design of co-assembling multi-component protein nanomaterials. *Nature* **510**, 103-108, (2014).

44. Bale, J. B., Gonen, S., Liu, Y., Sheffler, W., Ellis, D., Thomas, C., Cascio, D., Yeates, T. O., Gonen, T., King, N. P. & Baker, D. Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science* **353**, 389-394, (2016).

45. Korendovych, I. V. & DeGrado, W. F. De novo protein design, a retrospective. *Q. Rev. Biophys.* **53**, e3, (2020).

46. Joh, N. H., Wang, T., Bhate, M. P., Acharya, R., Wu, Y., Grabe, M., Hong, M., Grigoryan, G. & DeGrado, W. F. De novo design of a transmembrane Zn^{2+}-transporting four-helix bundle. *Science* **346**, 1520-1524, (2014).

47. Boyken, S. E., Benhaim, M. A., Busch, F., Jia, M., Bick, M. J., Choi, H., Klima, J. C., Chen, Z., Walkey, C., Mileant, A., Sahasrabuddhe, A., Wei, K. Y., Hodge, E. A., Byron, S., Quijano-Rubio, A., Sankaran, B., King, N. P., Lippincott-Schwartz, J., Wysocki, V. H., Lee, K. K. & Baker, D. De novo design of tunable, pH-driven conformational changes. *Science* **364**, 658-664, (2019).

48. Glasgow, A. A., Huang, Y.-M., Mandell, D. J., Thompson, M., Ritterson, R., Loshbaugh, A. L., Pellegrino, J., Krivacic, C., Pache, R. A., Barlow, K. A., Ollikainen, N., Jeon, D., Kelly, M. J. S., Fraser, J. S. & Kortemme, T. Computational design of a modular protein sense-response system. *Science* **366**, 1024-1028, (2019).

49. Dotan, N., Arad, D., Frolow, F. & Freeman, A. Self-Assembly of a Tetrahedral Lectin into Predesigned Diamondlike Protein Crystals. *Angew. Chem. Int. Ed. Engl.* **38**, 2363-2366, (1999).

50. Kitagishi, H., Oohora, K., Yamaguchi, H., Sato, H., Matsuo, T., Harada, A. & Hayashi, T. Supramolecular Hemoprotein Linear Assembly by Successive Interprotein Heme−Heme Pocket Interactions. *J. Am. Chem. Soc.* **129**, 10326-10327, (2007).

51. Ringler, P. & Schulz, G. E. Self-assembly of proteins into designed networks. *Science* **302**, 106–109, (2003).

52. Mirkin, C. A., Letsinger, R. L., Mucic, R. C. & Storhoff, J. J. A DNA-based method for rationally assembling nanoparticles into macroscopic materials. *Nature* **382**, 607-609, (1996).

53. Macfarlane, R. J., Lee, B., Jones, M. R., Harris, N., Schatz, G. C. & Mirkin, C. A. Nanoparticle superlattice engineering with DNA. *Science* **334**, 204-208, (2011).

54. Brodin, J. D., Auyeung, E. & Mirkin, C. A. DNA-mediated engineering of multicomponent enzyme crystals. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 4564-4569, (2015).

55. McMillan, J. R., Brodin, J. D., Millan, J. A., Lee, B., Olvera de la Cruz, M. & Mirkin, C. A. Modulating Nanoparticle Superlattice Structure Using Proteins with Tunable Bond Distributions. *J. Am. Chem. Soc.* **139**, 1754-1757, (2017).

56. Subramanian, R. H., Smith, S. J., Alberstein, R. G., Bailey, J. B., Zhang, L., Cardone, G., Suominen, L., Chami, M., Stahlberg, H., Baker, T. S. & Tezcan, F. A. Self-Assembly of a Designed Nucleoprotein Architecture through Multimodal Interactions. *ACS Cent. Sci.* **4**, 1578-1586, (2018).

57. Bailey, J. B., Subramanian, R. H., Churchfield, L. A. & Tezcan, F. A. Metal-Directed Design of Supramolecular Protein Assemblies. *Methods Enzymol.* **580**, 223-250, (2016).

58. Salgado, E. N., Faraone-Mennella, J. & Tezcan, F. A. Controlling protein-protein interactions through metal coordination: assembly of a 16-helix bundle protein. *J. Am. Chem. Soc.* **129**, 13374–13375, (2007).

59. Salgado, E. N., Lewis, R. A., Faraone-Mennella, J. & Tezcan, F. A. Metal-mediated self-assembly of protein superstructures: Influence of secondary interactions on protein oligomerization and aggregation. *J. Am. Chem. Soc.* **130**, 6082-6084, (2008).

60. Salgado, E. N., Lewis, R. A., Mossin, S., Rheingold, A. L. & Tezcan, F. A. Control of protein oligomerization symmetry by metal coordination: C_2 and C_3 symmetrical assemblies through Cu^{II} and Ni^{II} coordination. *Inorg. Chem.* **48**, 2726–2728, (2009).

61. Salgado, E. N., Ambroggio, X. I., Brodin, J. D., Lewis, R. A., Kuhlman, B. & Tezcan, F. A. Metal-templated design of protein interfaces. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 1827–1832, (2010).

62. Brodin, J. D., Smith, S. J., Carr, J. R. & Tezcan, F. A. Designed, helical protein nanotubes with variable diameters from a single building block. *J. Am. Chem. Soc.* **137**, 10468–10471, (2015).

63. Brodin, J. D., Ambroggio, X. I., Tang, C. Y., Parent, K. N., Baker, T. S. & Tezcan, F. A. Metal-directed, chemically tunable assembly of one-, two- and three-dimensional crystalline protein arrays. *Nat. Chem.* **4**, 375–382, (2012).

64. Brodin, J. D., Carr, J. R., Sontz, P. A. & Tezcan, F. A. Exceptionally stable, redox-active supramolecular protein assemblies with emergent properties. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2897–2902, (2014).

65. Brodin, J. D., Medina-Morales, A., Ni, T., Salgado, E. N., Ambroggio, X. I. & Tezcan, F. A. Evolution of metal selectivity in templated protein interfaces. *J. Am. Chem. Soc.* **132**, 8610–8617, (2010).

66. Song, W. J. & Tezcan, F. A. A designed supramolecular protein assembly with *in vivo* enzymatic activity. *Science* **346**, 1525–1528, (2014).

67. Churchfield, L. A., Medina-Morales, A., Brodin, J. D., Perez, A. & Tezcan, F. A. *De novo* design of an allosteric metalloprotein assembly with strained disulfide bonds. *J. Am. Chem. Soc.* **138**, 13163–13166, (2016).

68. Churchfield, L. A. & Tezcan, F. A. Design and Construction of Functional Supramolecular Metalloprotein Assemblies. *Acc. Chem. Res.* **52**, 345-355, (2019).

69. Suzuki, Y., Cardone, G., Restrepo, D., Zavattieri, P. D., Baker, T. S. & Tezcan, F. A. Self-assembly of coherently dynamic, auxetic, two-dimensional protein crystals. *Nature* **533**, 369–373, (2016).

Chapter 2: Engineering the entropy-driven free-energy landscape of a dynamic, nanoporous protein assembly

2.1 Abstract

De novo design and construction of stimuli-responsive protein assemblies that predictably switch between discrete conformational states remains an essential but highly challenging goal in biomolecular design. We previously reported synthetic, 2D protein lattices self-assembled via disulfide bonding interactions, which endows them with a unique capacity to undergo coherent conformational changes without losing crystalline order. Here we have carried out all-atom molecular dynamics simulations to map the free-energy landscape of these lattices, validated this landscape through extensive structural characterization by electron microscopy, and established that it is predominantly governed by solvent reorganization entropy. Subsequent redesign of the protein surface with conditionally repulsive electrostatic interactions enabled us to predictably perturb the free-energy landscape, and obtain a new protein lattice whose conformational dynamics can be chemically and mechanically toggled between three different states with varying porosities and molecular densities.

2.2 Introduction

Protein self-assembly is the predominant means of forming functional materials and machines in living systems[1] and has been increasingly exploited in the laboratory for the bottom-up construction of synthetic materials[2-5]. A fundamental physical understanding of the molecular interactions which underpin protein self-assembly is critical for the rational design of targeted supramolecular architectures, as well as for engineering their structural dynamics and their response to environmental cues. To date, much research on protein or biomolecular design has centered on the optimization of energetically–or more accurately, enthalpically–favorable bonding interactions between the molecular components to achieve the thermodynamically preferred self-assembly products[6,7]. Such focus on enthalpy minimization has placed considerable emphasis on obtaining singular architectures that assemble predominantly via quasi-irreversible, high-affinity interactions. These approaches frequently lead to a propensity for forming aggregates or misfolded products, and even when successful, yield assemblies with no or a low degree of structural adaptiveness and stimuli-responsiveness[5,8,9]. Despite their critical role in every aspect of protein structure and function (*e.g.*, folding[10], recognition[11], catalysis[12]), entropy and water-protein interactions are not explicitly considered in most design efforts, owing largely to the difficulty in their rational implementation by experiment or computation. Clearly, the design of protein assemblies with sophisticated functional and physical properties necessitates the attainment of not only structures corresponding to deep free-energy minima, but also of entire free-energy landscapes that such structures can traverse in a programmable manner.[13-16] This, in turn, requires a detailed understanding and mastery of enthalpic and entropic factors that govern protein-protein and protein-solvent interactions, which we set out to achieve in this work for the first time for a synthetic protein assembly.

In order to streamline the protein design process and gain access to protein complexes whose structures can be modulated through chemical stimuli, our laboratory has employed strong but reversible bonding interactions (*e.g.*, metal coordination[17-19], disulfide bonds) to mediate protein self-assembly. In a recent example, a C_4 symmetric tetrameric protein, L-rhamnulose-1-phosphate aldolase or RhuA, was functionalized with cysteines in its four corners at positions 98 (^{C98}RhuA), which directed its self-assembly into 2-D crystalline lattices of μm dimensions via disulfide bond formation[20] (**Figure 2.1a**). Whereas the reversibility and redox-tunability of the disulfide bonds enabled the formation of defect-free crystals, the flexibility of these bonds allowed the ^{C98}RhuA lattices to isotropically expand and contract without losing crystallinity (**Figure 2.1b**).

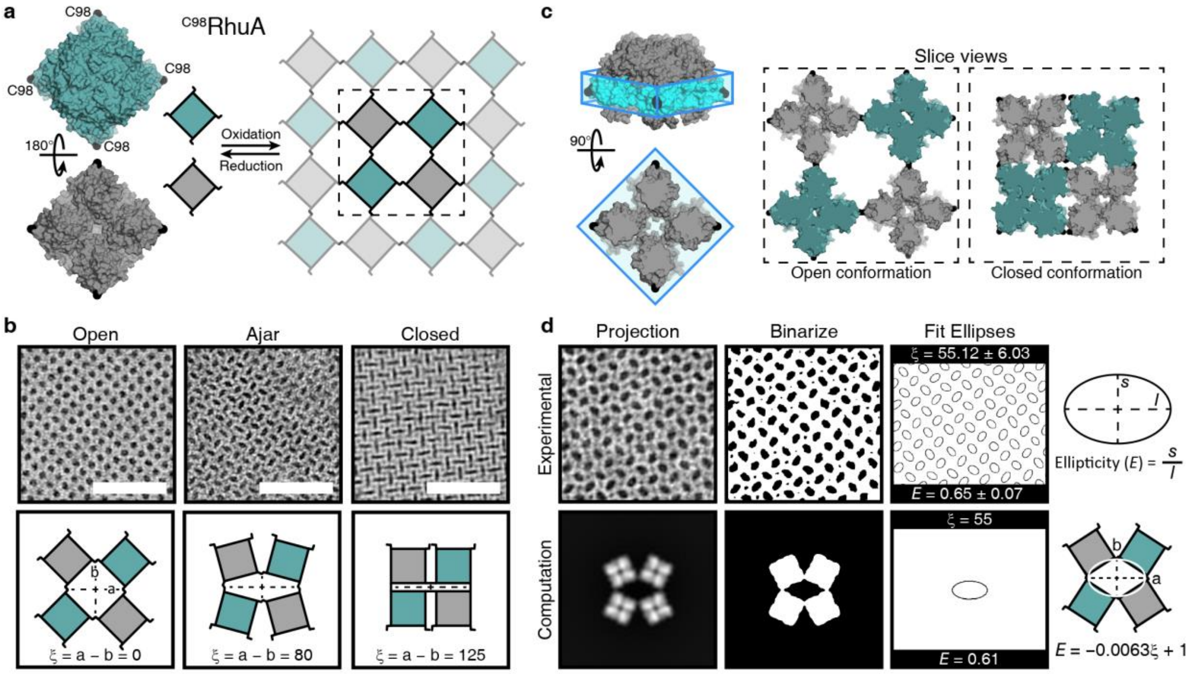


Figure 2.1 | Structural features of ^{C98}RhuA crystals. a, Surface representation of a ^{C98}RhuA tetramer, with positions 98 highlighted in black, and a schematic depiction of its oxidative self-assembly into porous 2D crystals. **b,** TEM and cartoon representations of the conformational states accessible by ^{C98}RhuA lattices and their respective ξ values. Scale bars are 50 nm. **c,** Description of the "interaction belt" (±10 Å of residues 98), which protrudes out from the protein. Slice views are of this belt region as viewed normal to the crystal. **d,** Overview of the calculation of the parameter E from experiment and its generalization to the coordinate ξ. The measured parameters and their converted values are reported for each example

This coherent dynamic behavior has two unique and potentially useful physical consequences. First, owing to the rotary motion of the ^{C98}RhuA building blocks with respect to one another and the resultant retention of overall $p42_12$ symmetry, these 2D lattices possess the lowest thermodynamically allowed Poisson's ratio of -1[20,21] (*i.e.*, they expand and contract by the same amount in both lateral dimensions). Materials with negative Poisson's ratios (auxetic materials) are expected to display increased toughness and resistance to fracture and have been proposed for numerous applications[22,23]. Second, the ^{C98}RhuA lattices essentially are 2D molecular membranes whose porosity can be coherently changed from a pore diameter of ~4 nm in the fully open state to ~1 nm in the fully closed state (**Figure 2.1b**). The transition from the open to closed state occurs spontaneously in aqueous suspension via a continuous population of many observable intermediate states, and is reversible upon mechanical agitation of the suspension. These unique physical properties of the 2D ^{C98}RhuA lattices have prompted the following questions: a) what are the determinants of the free-energy landscape that governs 2D ^{C98}RhuA lattice dynamics and b) can this landscape be engineered at the molecular level to modulate the collective structural and physical properties of the lattice? To address these questions, we carried out a combination of computational and experimental studies, which revealed the substantial role that entropy and water-protein interactions can play in protein structural dynamics, and allowed us to rationally modulate the lattice conformations and create mechanically and chemically switchable systems.

2.3 Results

2.3.1 Distinctive structural features of ^{C98}RhuA lattices

There are fundamental challenges in the computational and experimental mapping of the free energy landscape of a dynamic protein assembly. First, any protein-protein interface and the associated solvent molecules present an immense chemical complexity. This complexity, coupled with the many degrees of freedom of motion (*i.e.*, interaction trajectories) that a dynamic interface can possess, gives rise to a multidimensional energy landscape which would be computationally intractable to model with atomistic accuracy. Second, the experimental evaluation of such a landscape or any of the trajectories it contains would be a considerable undertaking with any physical tool for studying structural dynamics. The ^{C98}RhuA lattices possess distinctive features that provide advantages in both regards:

1) *^{C98}RhuA lattice dynamics can be described by a simple free energy landscape with a single reaction coordinate*, corresponding to the "openness" of the lattice pores. Due to the coherent dynamics of the lattice and the maintenance of its flatness during opening and closing at all times, a single parameter ξ is sufficient to define the reaction coordinate and the extent of interactions between neighboring protein units (**Figure 2.1b**). We define ξ as the difference in the lengths of the two principal axes, a and b, of the pores, ranging from a minimum of 0 Å in the fully open state to a maximum of 125 Å in the fully closed state.

2) *^{C98}RhuA lattice energy landscape is further simplified by the topology and composition of RhuA surfaces*. As viewed along its principal C_4 symmetry axis (**Figure 2.1c**), each ^{C98}RhuA building block is wider in its middle portion where the C98 residues are located compared to its top and bottom. This reduces the possible interaction surface between neighboring units to a small belt region (colored in cyan) as they hinge about the disulfide linkages. Additionally, the ^{C98}RhuA lateral surfaces including the belt regions are characterized by a diffuse distribution of hydrophilic residues and are expected to be non-self-interacting (**Figure 2.2b**), thus giving rise to a landscape devoid of deep energy wells.

3) *The reaction coordinate parameter ξ can be directly determined by transmission electron microscopy* (TEM). Owing to the coherent dynamics of ^{C98}RhuA crystals, each lattice possesses a single value for pore ellipticity (E) that can be obtained accurately by TEM image processing (**Figure 2.1d**). ξ can then be calculated through its linear relationship to E, where $E = -0.0063\xi + 1$. This relationship was determined by applying the same image processing procedure to calculated 2D projection maps of model structures with known ξ values (**Figure 2.3b**). Furthermore, the fact that each TEM sample contains multiple individual ^{C98}RhuA lattices

(prepared under the same condition) enables the statistical analysis of multiple "single-lattice" measurements.

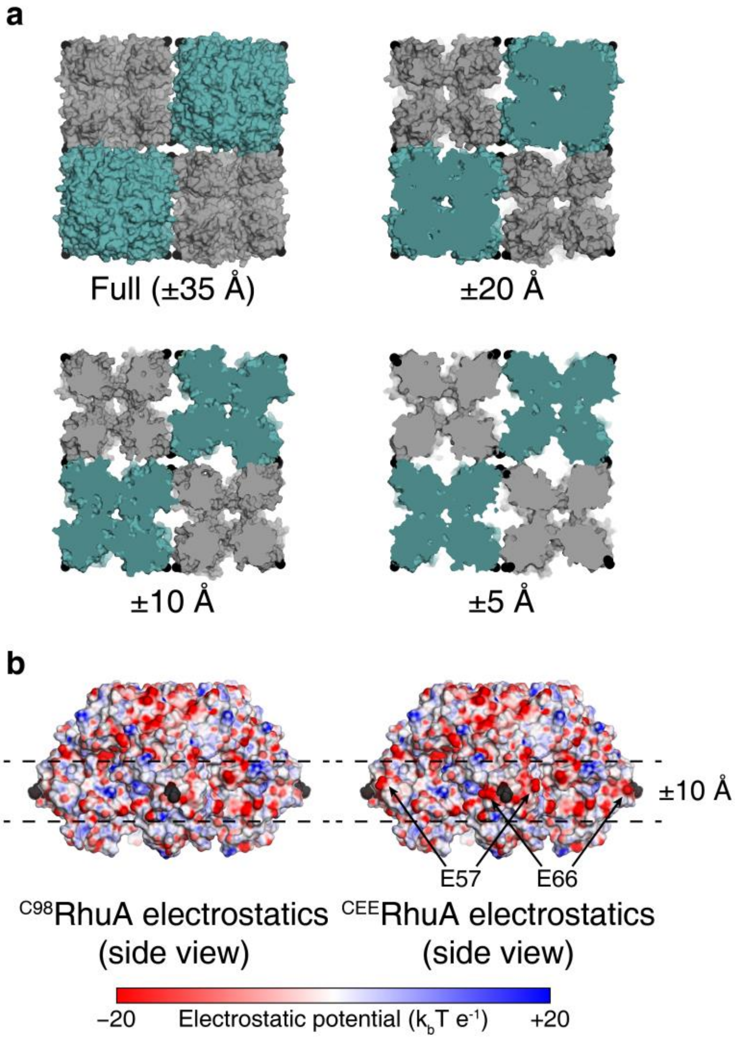


Figure 2.2 | Features of the ^{C98}RhuA interface. a, Slice views of the ^{C98}RhuA 2x2 unit cell in the fully closed (ξ = 125 Å) state at different thicknesses. All slices are centered on the average height of the C98 residues. Despite the appearance of close contacts in the full structure, the true contact area is limited to the "interaction belt", consistent with the relatively small total (all 4 interfaces) buried surface area (361 Å^2 at ξ = 125; 115 Å^2 at the free energy minimum ξ = 106 Å). Except where otherwise noted, all water structure calculations were performed with a slice thickness of ±10 Å. **b,** Electrostatic maps of both RhuA constructs as calculated using APBS 1.5.[52,53] In this view, the protrusion of the "belt" region away from the protein, which geometrically limits the total interfacial area formed upon adoption of the closed state by the crystal, is clearly visible.

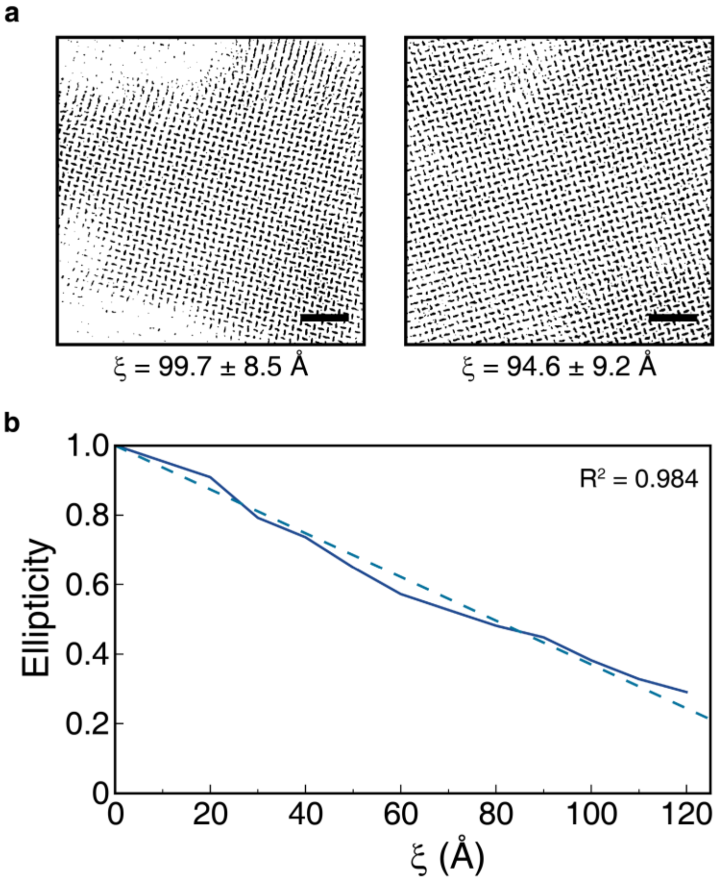


Figure 2.3 | TEM analysis and fitting of ξ to ellipticity. a, Examples of binarized TEM images of ^{C98}RhuA crystals utilized for the statistical analysis of crystal conformations. The average fitted ξ values and their standard deviations are listed below each images. Scale bars are 50 nm. **b,** Plot of ellipticity values, as calculated from calculated projection maps of GIST reference structures for ^{C98}RhuA, versus their known ξ values (solid blue line). The dashed line is the linear fit of $E = -0.0063\xi+1.0$. The R^2 value of the linear fit is 0.984.

2.3.2 Solvent entropy dominates the free-energy landscape of ^{C98}RhuA crystals

In order to evaluate the energy landscape of ^{C98}RhuA crystals, we carried out all-atom umbrella sampling (US) calculations on a set of four ^{C98}RhuA tetramers connected via intermolecular disulfide bonds (**Figure 2.1a**). This 2×2 unit, containing 67,512 atoms, represents the minimum number of protein subunits required to explore the relevant conformational changes for ^{C98}RhuA crystals while still preserving the native lattice topology and ensuring computational tractability. As explained above, ξ was chosen as the reaction coordinate, which was divided into 26 windows from 0 to 125 Å centered at 5-Å intervals. Mild boundary conditions were employed to maintain lattice planarity imposed in the bulk crystal (**Figure 2.4b**), and sampling was carried out at 300 K for 30 ns. Sampling was periodically tested for equilibration via the block averaging method[24], and the system was judged to have equilibrated after 15 ns. Sampling statistics from the remaining 15 ns (**Figure 2.4a**) were combined and reweighted using the WHAM algorithm[25] to generate a one-dimensional potential of mean force (PMF) (**Figure 2.5a**, top).

The ^{C98}RhuA PMF reveals a smooth landscape with a monotonic 12 kcal/mol decrease in free energy from the fully open state to a close-packed state centered at $\xi = 106$ Å, followed by an increase upon further compaction. This landscape with a distinct minimum is consistent with the spontaneous closure of the ^{C98}RhuA lattices in aqueous suspensions, and their conversion into open forms upon gentle mechanical agitation. Importantly, the Boltzmann population distribution of ^{C98}RhuA lattice states calculated from the PMF reveals a close match to the experimental distributions determined by TEM imaging ($\xi_{min} = 102.9 \pm 7.3$ Å) (**Figure 2.5b**; see **Figure 2.3a** for representative binarized TEM images). This observation provides strong validation for the efficacy of both the one-dimensional PMF along ξ to describe the ^{C98}RhuA lattice dynamics and the TEM experiments to sample them.

An examination of the trajectories generated in US simulations reveals that the interfaces between ^{C98}RhuA units remain separated by a modest distance even upon full closure (ξ = 120–125 Å) (**Figure 2.2a**). In fact, the minimum projected protein-protein distance (excluding C98) is ca. 4 Å between D39 and L64' side chains at $\xi = 125$ Å, with the vast majority of all contacts >5 Å. The buried interface area between the protein units was calculated to be 361 Å^2 at $\xi = 125$ Å, and only 115 Å^2 at $\xi = 106$ Å. These findings led us to consider the possibility that the free energy changes observed in the PMF were not due to enthalpically favorable protein-protein interactions, but were rather dominated by entropy. Generally, the entropic contributions to the free energy can be intuitively broken up into two components, protein configurational

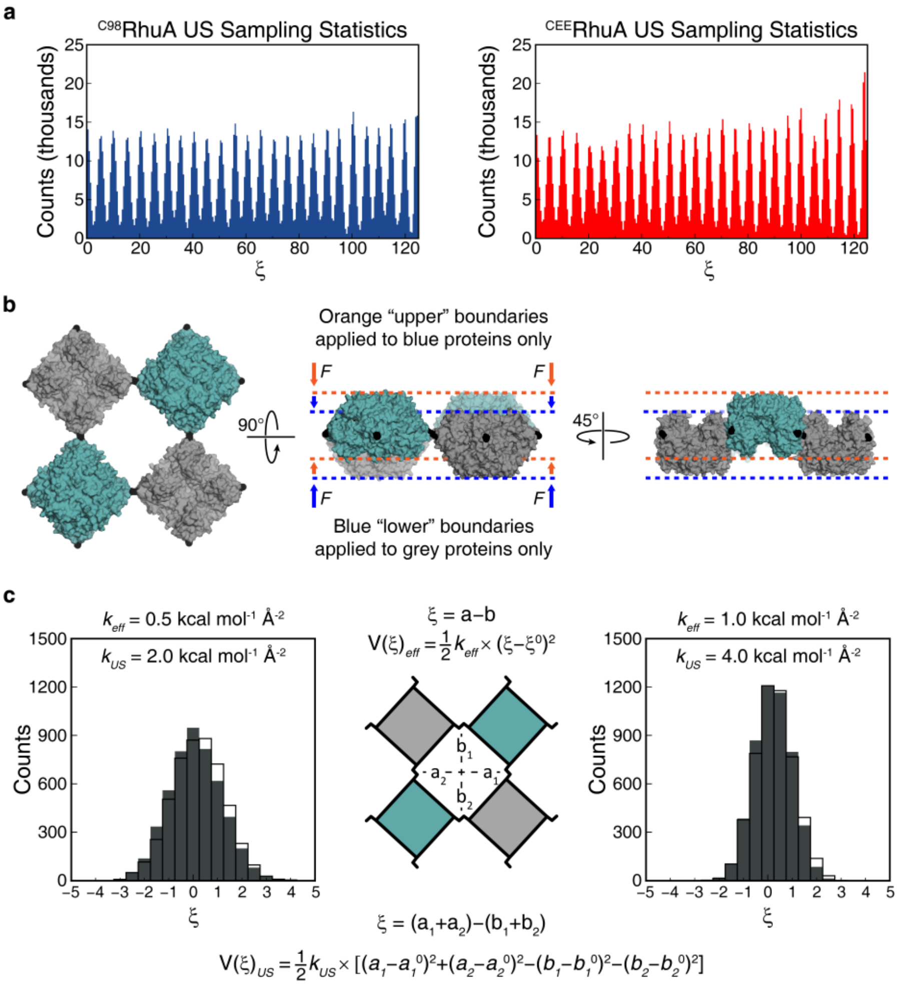


Figure 2.4 | Umbrella sampling statistics and calculation parameters. **a,** Complete histograms of all window sampling statistics used for calculation of the PMFs for ^{C98}RhuA and ^{CEE}RhuA. **b,** Schematic of the artificial boundary constraints imposed to maintain planarity during sampling. The side view of the crystal is shown from two different angles to highlight the vertical offset between neighboring subunits which necessitated the use of two pairs of boundary restraints. **c,** Comparison between distributions of ξ numerically obtained by two distinct harmonic restraint configurations: 1) a single restraint enforced on the full reaction coordinate (top ξ equation; empty bars), and 2), four pore-center-to-disulfide distance restraints (bottom ξ equation; filled bars). These distributions are equivalent when $k_{US} = 4 \times k_{eff}$, allowing the PMFs to be formally calculated from 1D $V(\xi)_{eff}$ potentials, despite the use of the $V(\xi)_{US}$ restraint for more efficient (see Methods) collection of US window statistics. This relationship likely arises because a and b can be written as a linear combination of a_1, a_2, b_1, and b_2.

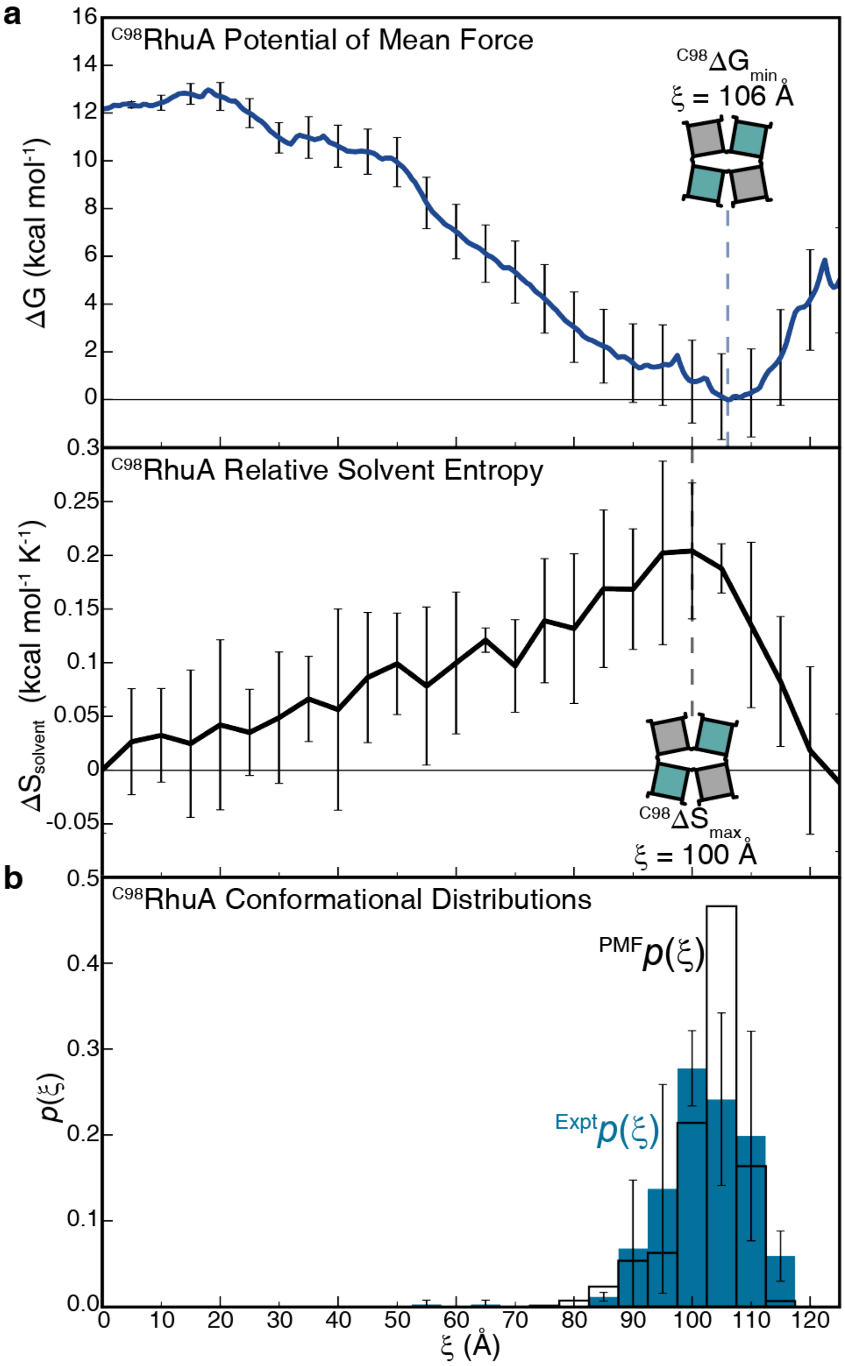


Figure 2.5 | Thermodynamic analysis of ^{C98}RhuA lattice structural dynamics. a, The calculated free energy landscape and concomitant changes in solvation entropy over a continuous range of ^{C98}RhuA lattice conformations. **b,** Direct comparison of the equilibrium conformations of ^{C98}RhuA crystals from experimental measurement (blue) and as calculated from the PMF (colorless). Experimental error bars are determined from three statistical analyses, each comprising >100 TEM images of ^{C98}RhuA crystals.

entropy ($\Delta S_{protein}$) and solvent entropy ($\Delta S_{solvent}$), both of which can be critical determinants of protein-protein interactions and protein crystallization[26,27]. Given that the compaction of the lattice cannot be expected to increase $\Delta S_{protein}$ and indeed was not found to (**Figure 2.6**), we investigated the role of $\Delta S_{solvent}$ through Grid Inhomogeneous Solvation Theory (GIST) calculations[28]. Sampling for GIST calculations was carried out on representative protein structures selected from

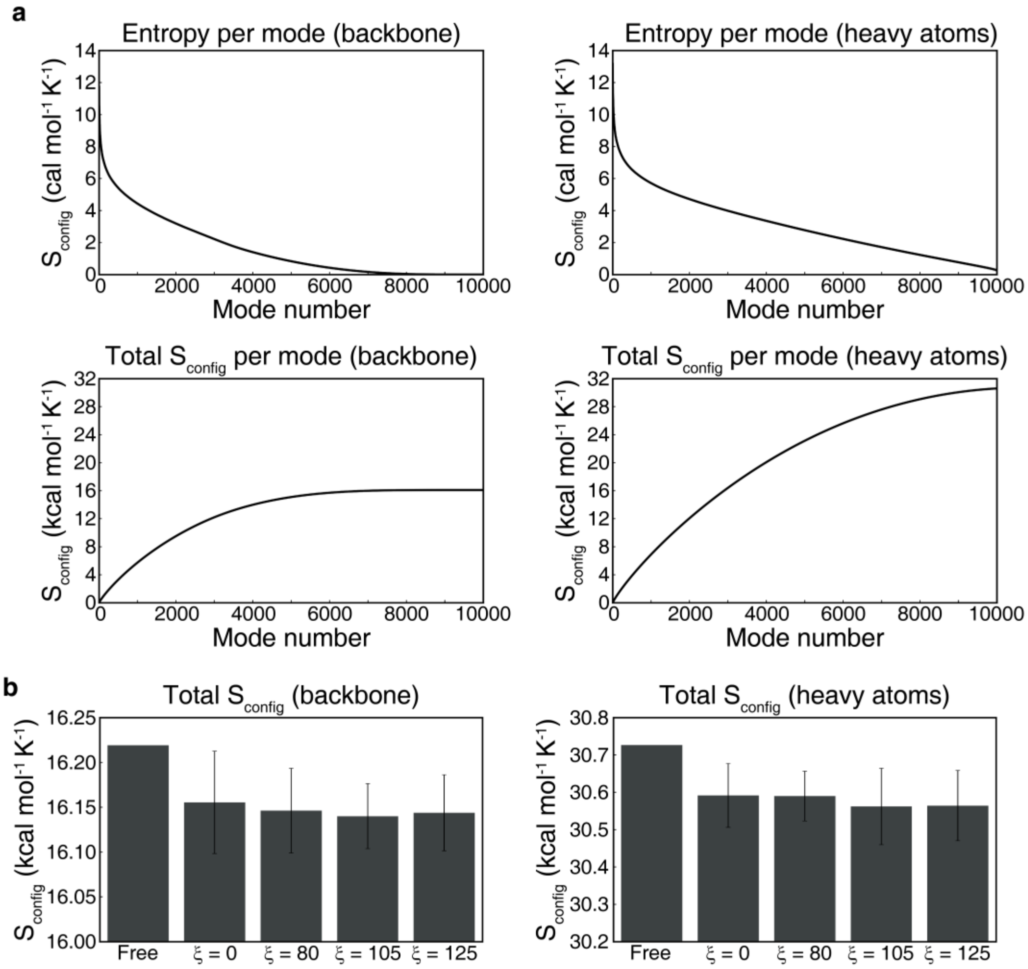


Figure 2.6 | Configurational entropy calculations. a, Convergence plots for configurational entropy (S_{config}) of ^{C98}RhuA over the first 10,000 vibrational modes. Top: S_{config} per mode, Bottom: total S_{config} vs mode number, Left: backbone atoms only, Right: all protein heavy atoms. **b,** Total configurational entropy for backbone atoms (left) and all heavy atoms (right). In both cases there is a small entropic penalty for constraining RhuA within a lattice (relative to a free protein), but there is effectively no change in the configurational entropy as a function of ξ (i.e., changes in average S_{config} are all within error of each other). Values and error bars are calculated from averaging of all four proteins in the system.

each US window for 20 ns each, keeping the protein coordinates maintained at their original positions with a 10 kcal/mol restraint, and saving configurations every 0.5 ps. To provide an estimation for the error, we took advantage of the inherent 4-fold symmetry of the simulated system by splitting it into four symmetrically equivalent quadrants, each containing one protein, and applying GIST to each quadrant individually. To avoid a systematic overestimation of the entropy by the nearest-neighbor algorithm employed by GIST[29], the full trajectories were split into even and odd frames (1 frame/ps) and analyzed separately, thus permitting an additional degree of averaging for error estimation. The GIST results reveal a monotonic increase in $\Delta S_{solvent}$ from ξ = 0 Å to 100 Å and a rapid decrease after ξ = 105 Å (**Figure 2.5a**, bottom), a profile that is almost exactly the mirror image of the PMF (**Figure 2.5a**, top). This striking correspondence between solvent entropy and lattice free energy, coupled with the lack of appreciable protein-protein contacts, strongly suggests that ^{C98}RhuA lattice dynamics are driven by solvent entropy.

2.3.3 Water structure reorganization during ^{C98}RhuA lattice motions

As the microscopic detail afforded by molecular dynamics simulations allows solvent reorganization to be rigorously interrogated, we sought to rationalize the entropic changes accompanying ^{C98}RhuA lattice dynamics in the context of molecular hydration theory. We initially characterized the density profiles as a function of ξ, which is conveniently projected into a 2D plot due to the low dimensionality of the system (**Figure 2.7a**, top; **Figure 2.8**). Consistent with previous computational[30,31] and experimental[32,33] studies, the anomalous structure of water within the hydration shells of biomolecules are clearly visible as density variations which line the perimeter of the pore and propagate ~10 Å away from the protein surface. Strong protein-water interactions give rise to the high density primary hydration layer, which in turn induces additional structure in the form of concentric density peaks. As neighboring RhuA subunits draw closer with

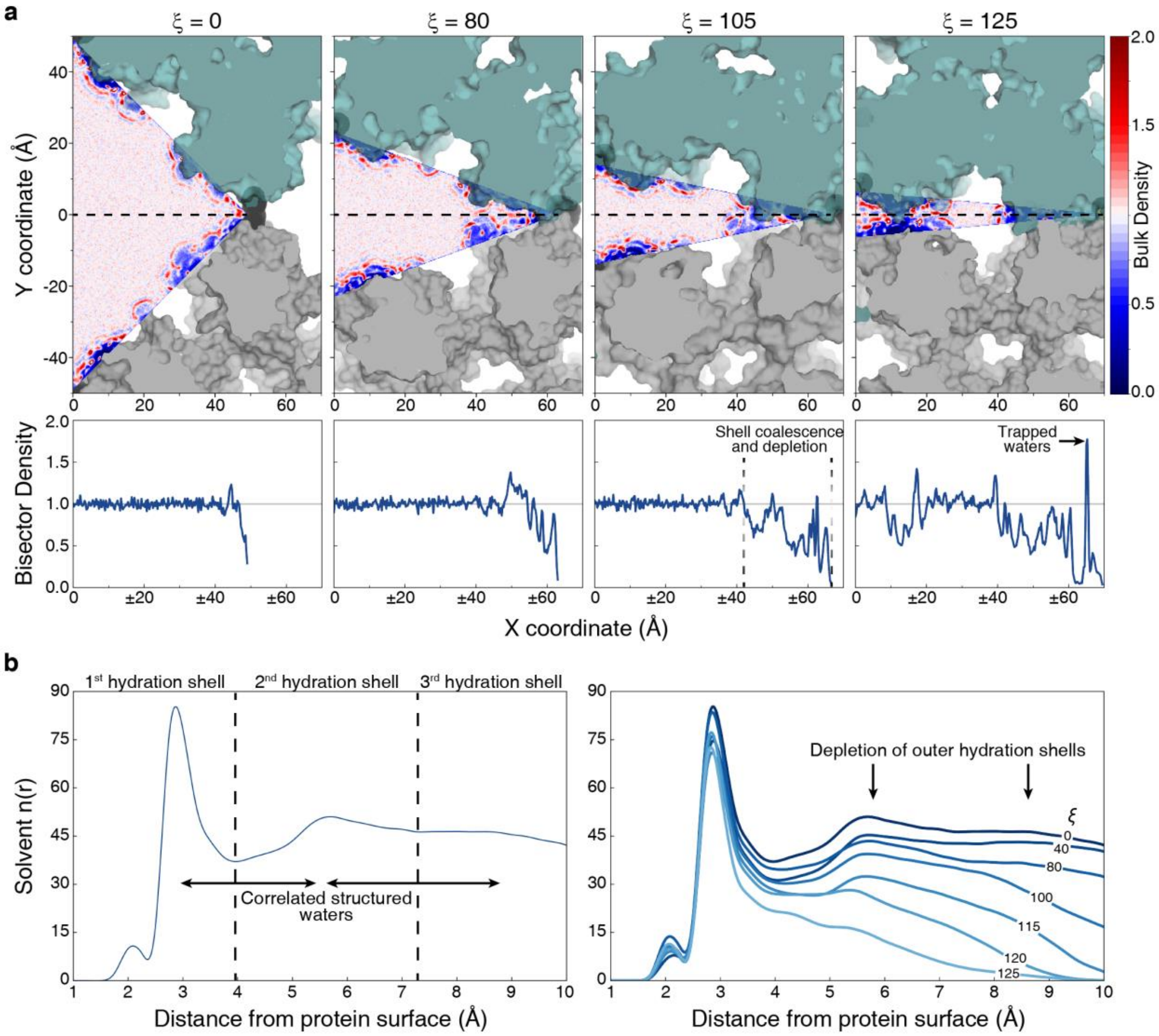


Figure 2.7 | Consequences of lattice compaction on solvent structure within the pore. a, Two-dimensional plots of the normalized water density within the pore. 1D plots of the density along the pore bisector are shown below to facilitate quantitative interpretation of the data. Labels here and in **b** identify hydration effects of interest. **b,** Number distributions of solvent molecules proximal to the protein surface. Concentric hydration shells around the protein are identified as peaks in the distribution for ξ = 0 (left).

increasing ξ, their hydration shells coalesce, leading to an ejection of these semi-structured water molecules into the bulk solvent (**Figure 2.7a**, bottom). Inspection of the number distribution of water molecules as a function of distance from the protein surface reveals that it is primarily the secondary and tertiary hydration shells which are released into the bulk (**Figure 2.7b,** right). Upon further narrowing of the pore, isolated pockets of high density manifest at the acute hinge regions (**Figure 2.7a**, bottom), which reflects a trapping of water molecules by the protein surfaces. This trapping effect may contribute to the sharp decrease in solvent entropy[34] for $\xi > 105$ Å.

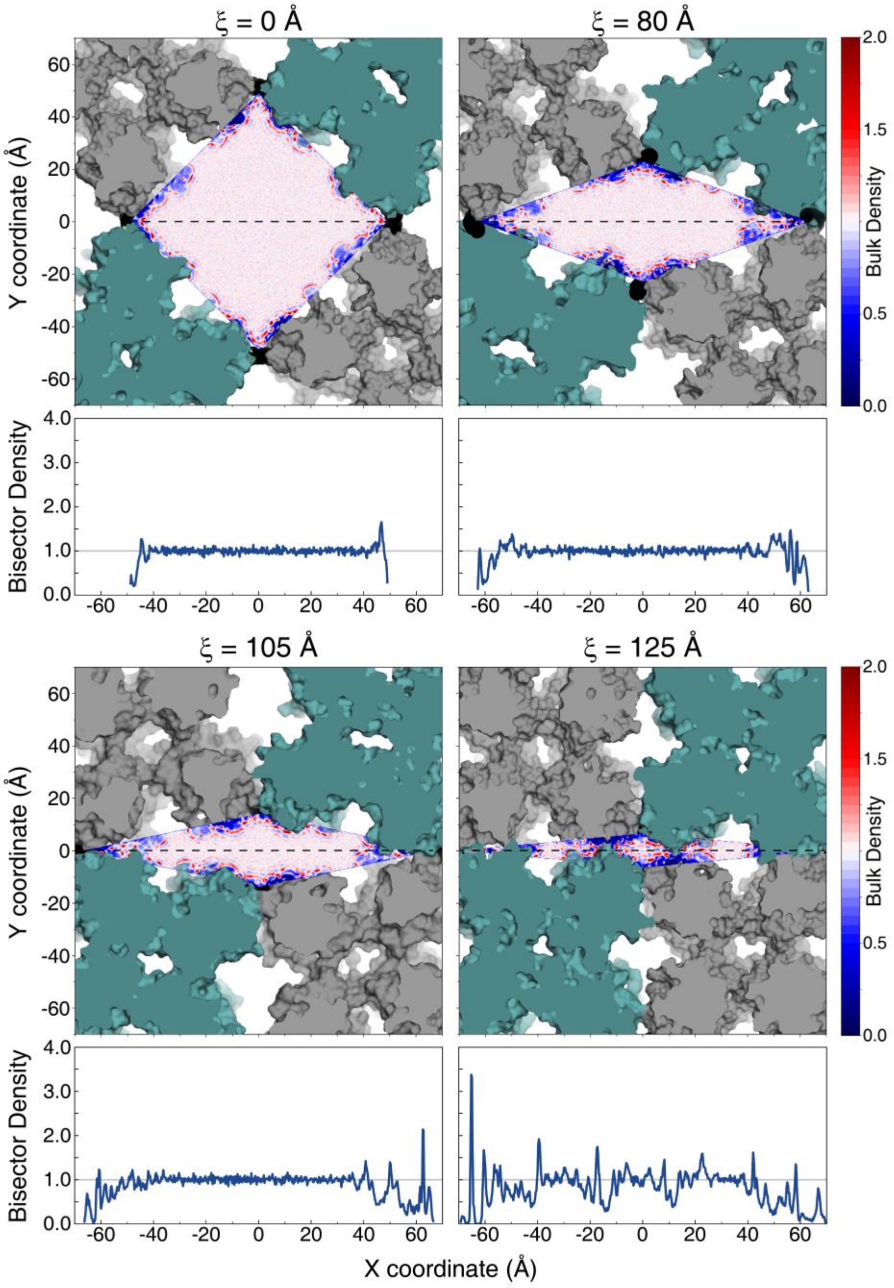


Figure 2.8 | Full 2D water density plots for ^{C98}RhuA. Complete (non-symmetrically averaged) maps of the 2D water density plots presented in Figure 3. Here, the hydration effects induced by the protein surfaces (and minor asymmetry of the pore) on the local solvent structure can be visualized in greater detail.

The computational tractability afforded by the protein structure provides an opportunity to examine the structure of water under confinement for a non-idealized system. We calculated two additional order parameters for water within the pore (see Methods): q_{tetra}, a measure of local tetrahedral structure, and coordination number (CN), the number of neighboring solvent molecules within 3.5 Å of a given water. We find that q_{tetra} decreases on average as ξ increases, although at very high values of ξ there is also an increase in the fraction of ice-like (tetrahedral) water molecules, with q_{tetra} values close to 1 (**Figure 2.9**, top/bottom). Similarly, we find that the average CN decreases with increasing ξ (**Figure 2.9**, middle/bottom), which is indicative of the isolation

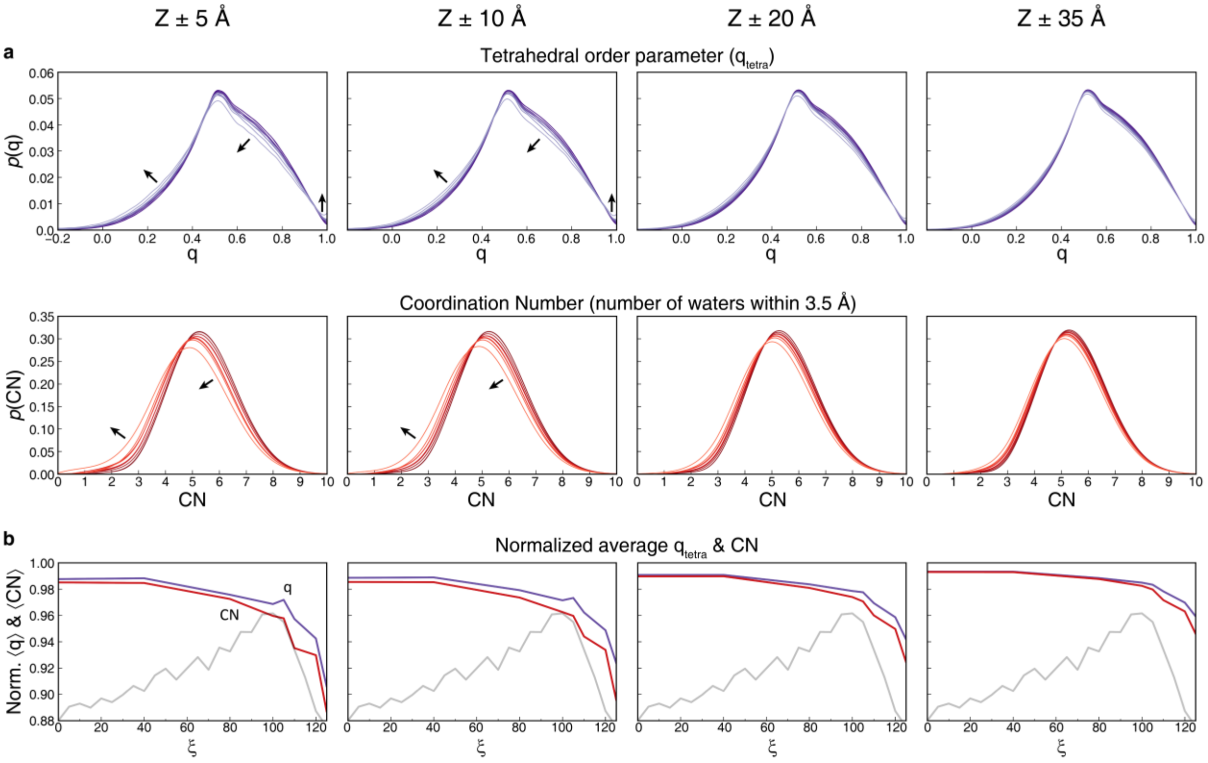


Figure 2.9 | Structural parameters for water confined within the ^{C98}RhuA pore. a, Distributions of q_{tetra} and CN at different slice thicknesses. The windows presented are (from darkest to lightest): ξ = 0, 40, 80, 100, 105, 110, 115, 120, and 125 Å. Arrows indicate trends with increasing values of ξ, which are most pronounced for water localized within the "interaction belt" of the pore. **b,** Average values of q_{tetra} and CN plotted as a function of ξ (with ^{C98}RhuA GIST results underlay for qualitative comparison). The values are normalized to a neat water box, so they are always <1 due to solvation effects.

of water molecules from other solvent molecules by the proteins. When these parameters are averaged over increasingly thick slabs within the pores, water molecules become progressively more bulk-like, further emphasizing that most of the perturbation to the water structure occurs within the interaction belt and thus directly reflects the effect of nearby protein surfaces.

2.3.4 Rational perturbation of the free-energy landscape of ^{C98}RhuA lattices

We next set out to rationally modulate the free energy landscape of ^{C98}RhuA lattices through site-specific modification of protein interaction surfaces. Given the non-trivial nature of engineering entropic factors, we envisioned that electrostatically repulsive (*i.e.*, enthalpically unfavorable) interactions placed near the C98-C98 hinges could hinder the closure of the lattices independently of solvent entropy, thus yielding an altered equilibrium state compared to ^{C98}RhuA. To achieve this effect, we chose to replace the residues P57 and A66 with negatively charged glutamates; these positions lie at comparable heights within the "interaction belt" and are immediately adjacent to the C98-C98 hinges to maximize the range over which their electrostatic repulsion can influence the lattice dynamics (**Figure 2.10a, b**).

We first evaluated the consequences of the designed mutations *in silico*. The US calculations for the E57/E66/C98-RhuA mutant (^{CEE}RhuA) were carried out in an identical fashion to those for ^{C98}RhuA, with the exception that stronger force constants were required at ξ >110 Å (**Table 2.1**), an early indication that the mutations disfavored the closed conformations. The PMF of the ^{CEE}RhuA lattice is shown in **Figure 2.10d** (top) overlaid with that of ^{C98}RhuA to highlight their differences. The driving force for ^{CEE}RhuA lattice closure (–6 kcal/mol) is halved

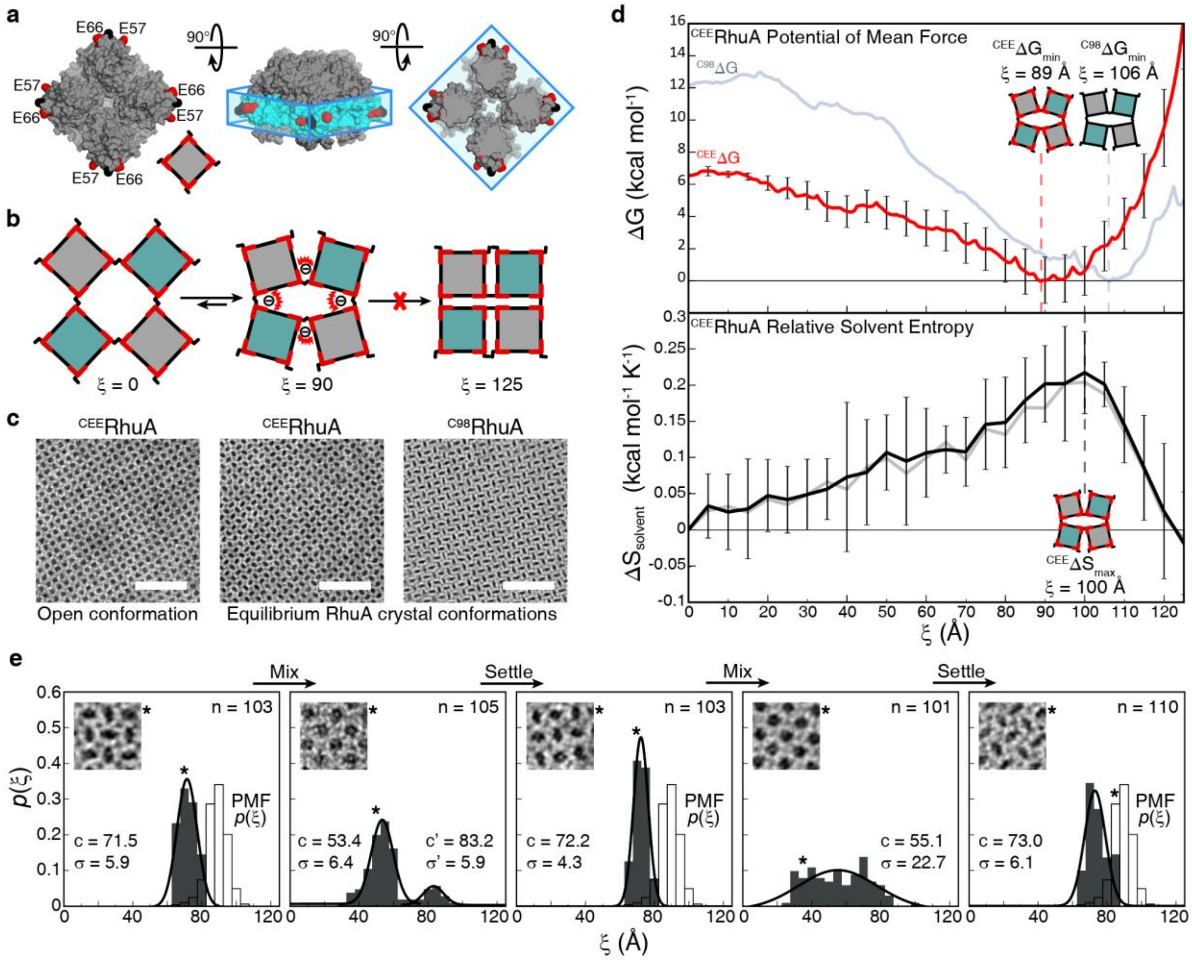


Figure 2.10 | Design and analysis of the designed construct $^{\mathrm{CEE}}$RhuA. a, Surface representation of a $^{\mathrm{CEE}}$RhuA tetramer, with residues 98 and 57/66 colored in black and red, respectively. The installed residues are coplanar with C98 and lie within the “interaction belt”. **b,** Cartoon of $^{\mathrm{CEE}}$RhuA lattice dynamics and the anticipated effects of the design. **c,** TEM images of RhuA lattices, showing the open and equilibrium states for $^{\mathrm{CEE}}$RhuA alongside that for a fully closed $^{\mathrm{C98}}$RhuA crystal. Scale bars are 50 nm. **d,** Relative to $^{\mathrm{C98}}$RhuA, the thermodynamic analysis of $^{\mathrm{CEE}}$RhuA dynamics reveals a shifted free energy minimum towards the open state, but retention of a nearly identical solvent entropy profile, demonstrating the purely enthalpic consequences of the design. The free energy and solvent entropy profiles for $^{\mathrm{CEE}}$RhuA are shown as red and black lines; those for $^{\mathrm{C98}}$RhuA are depicted as faint blue and grey lines. **e,** Experimental distributions of $^{\mathrm{CEE}}$RhuA conformations over multiple cycles of sedimentation and resuspension. The equilibrium distributions are significantly more open than $^{\mathrm{C98}}$RhuA, and slightly more open than predicted by the PMF. All distributions are fit to Gaussian distributions and labeled with their center (c) and standard deviation (σ). *n* is the number of crystals analyzed. The lattice conformation of each inset is marked with an asterisk.

compared to the ^{C98}RhuA lattice due to repulsive electrostatic interactions, which give rise to a shallower landscape at low ξ values and a steep increase at $\xi > 100$ Å due the proximity of E57 and E66 sidechains upon lattice closure. GIST calculations (**Figure 2.10d,** bottom) revealed that the solvent entropy profile for the ^{CEE}RhuA lattice is nearly identical to that of ^{C98}RhuA (faint underlay), indicating that the observed changes in the ^{CEE}RhuA energy landscape are entirely enthalpic in origin.

An important outcome of E57 and E66 mutations is that the PMF energy minimum is shifted from $\xi_{min} = 106$ Å for ^{C98}RhuA to $\xi_{min} = 89$ Å for ^{CEE}RhuA (**Figure 2.10d,** top), which predicts a considerably less compact lattice conformation at equilibrium for the latter. To experimentally probe this prediction, we prepared the ^{CEE}RhuA mutant and optimized the solution conditions for its self-assembly into μm-sized, 2D lattices in high yields (**Figure 2.11**, top). To attain the equilibrium state for these lattices, the crystal suspensions were left unperturbed for three days during which they sedimented to the bottom of the container, as was also observed for the ^{C98}RhuA crystals. In accordance with the predictions, TEM analysis of equilibrated ^{CEE}RhuA lattices revealed a significantly more relaxed conformation (termed “ajar”) compared to ^{C98}RhuA (**Figure 2.10c**), with a tight distribution of states centered at $\xi_{min} = 72.1 \pm 5.2$ Å (**Figure 2.10e,** “settled” states). This value is smaller than that predicted by the PMF, but corresponds to deviation of only ~2 kcal/mol, within the error of experimental measurements and the precision of US calculations. The reduced accuracy of the predicted minimum may be exacerbated by the shallowness of the ^{CEE}RhuA landscape. A statistical analysis of the distribution of ^{CEE}RhuA lattice conformations was carried out over multiple rounds of vigorous mixing followed by sedimentation (**Figure 2.10e**), establishing that the opening-closing dynamics of this variant were also reversible and mechanically actuated.

Table 2.1 | Umbrella sampling force constants. This table shows the actual force constants (k_{US}) used for each harmonic restraint during umbrella sampling (a1, a2, b1, b2), as well as the effective force constant (k_{eff}) that yields the identical distribution when applied to the full reaction coordinate ξ = a−b, that is: $k_{US}\,[(a_1 + a_2) - (b_1 + b_2)] \equiv k_{eff}\,(a-b)$. For calculation of the PMF via WHAM, all sampling statistics were converted to the form ξ = a−b and scaled using the corresponding effective force constant for each window. See Figure 2.4 for additional details.

ξ_0	^{C98}RhuA k_{US}	^{C98}RhuA k_{eff}	^{CEE}RhuA k_{US}	^{CEE}RhuA k_{eff}
0	2	0.5	2	0.5
5	2	0.5	2	0.5
10	2	0.5	2	0.5
15	2	0.5	2	0.5
20	2	0.5	2	0.5
25	2	0.5	2	0.5
30	2	0.5	2	0.5
35	2	0.5	2	0.5
40	2	0.5	2	0.5
45	2	0.5	2	0.5
50	2	0.5	2	0.5
55	2	0.5	2	0.5
60	2	0.5	2	0.5
65	2	0.5	2	0.5
70	2	0.5	2	0.5
75	2	0.5	2	0.5
80	2	0.5	2	0.5
85	2	0.5	2	0.5
90	2	0.5	2	0.5
95	2	0.5	2	0.5
100	2	0.5	2	0.5
105	2	0.5	2	0.5
110	2	0.5	2	0.5
115	2	0.5	2.5	0.625
120	2	0.5	3	0.75
125	2	0.5	4	1

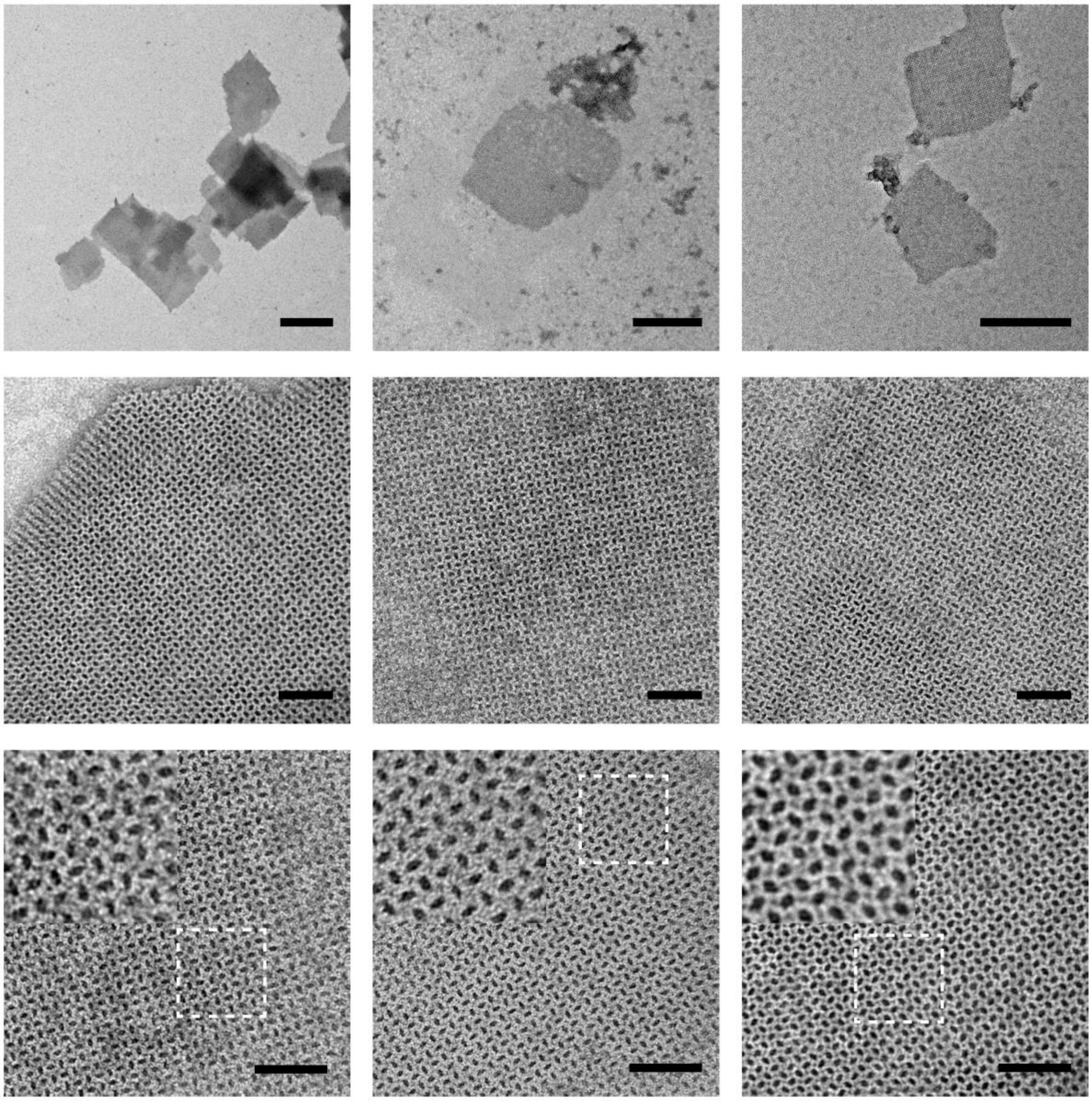


Figure 2.11 | TEM characterization of ^{CEE}RhuA lattices. Negatively-stained micrographs of equilibrated ^{CEE}RhuA crystals at different magnifications. Matured crystals grown from 100-125 μM protein in a buffer of 20 mM Tris (pH 7.5), 5-10 mM βME, and 1 mM $ZnCl_2$, were left to stand for 3 days prior to imaging. Scale bars are 500 nm for the top row and 50 nm elsewhere. Insets: additional 2x magnification of the white boxed regions.

2.3.5 Selective, metal-mediated switching of ^{CEE}RhuA lattices

An inspection of the ^{CEE}RhuA conformations from US simulations indicates that E57 and E66 carboxylate groups could come within sufficient proximity (<5 Å) upon lattice compaction to enable metal coordination across protein-protein interfaces (**Figure 2.12**). Thus, we reasoned that the interfacial E57 and E66 dicarboxylate motifs may engender a chemically switchable system whereby the full closure of the lattice may be induced by metal binding (**Figure 2.13a**). Indeed, we found that the addition of Ca^{2+} to ^{CEE}RhuA lattices in the equilibrium "ajar" state led to a shift in the population of lattices to higher ξ values in a concentration-dependent manner (**Figure 2.14**), reaching a distribution that is similar to the closed form of ^{C98}RhuA lattices at 20 mM Ca^{2+} (**Figure 2.13a**). In contrast, similar amounts of Mg^{2+} or up to 150 mM monovalent cations (Na^+ or K^+) did not cause any noticeable deviation from the "ajar" state (**Figure 2.15**), but the addition of 20 mM Ca^{2+} to ^{CEE}RhuA lattices pre-treated with 20 mM Mg^{2+} induced their closure (**Figure 2.14**). These results indicate that observed effect of Ca^{2+} on the lattice conformational state is selective and not due to a non-specific electrostatic screening effect. Like ^{C98}RhuA lattices, the Ca^{2+}-incubated ^{CEE}RhuA lattices could be reverted into a fully open state upon mechanical mixing (**Figure 2.13a**). Importantly, these lattices could be switched to the "ajar" state upon dialysis or treatment with ethylenediaminetetraacetate (EDTA) to remove Ca^{2+}, and this process could be reversed upon re-incubation with Ca^{2+}.

The conformational properties of ^{C98}RhuA and ^{CEE}RhuA lattices are summarized in **Figure 2.13b**. Whereas the ^{C98}RhuA system could only be mechanically switched between open and closed states, ^{CEE}RhuA lattices can adapt two different mechanical switching modes (open-closed or open-ajar) or a chemical switching mode (ajar-closed), depending on the presence or the absence of Ca^{2+} in the system. Thus, ^{CEE}RhuA lattices provide a unique 2D membrane system (or molecular display) whose porosity (or display density) can be chemically and mechanically toggled between three different states in a fully coherent fashion.

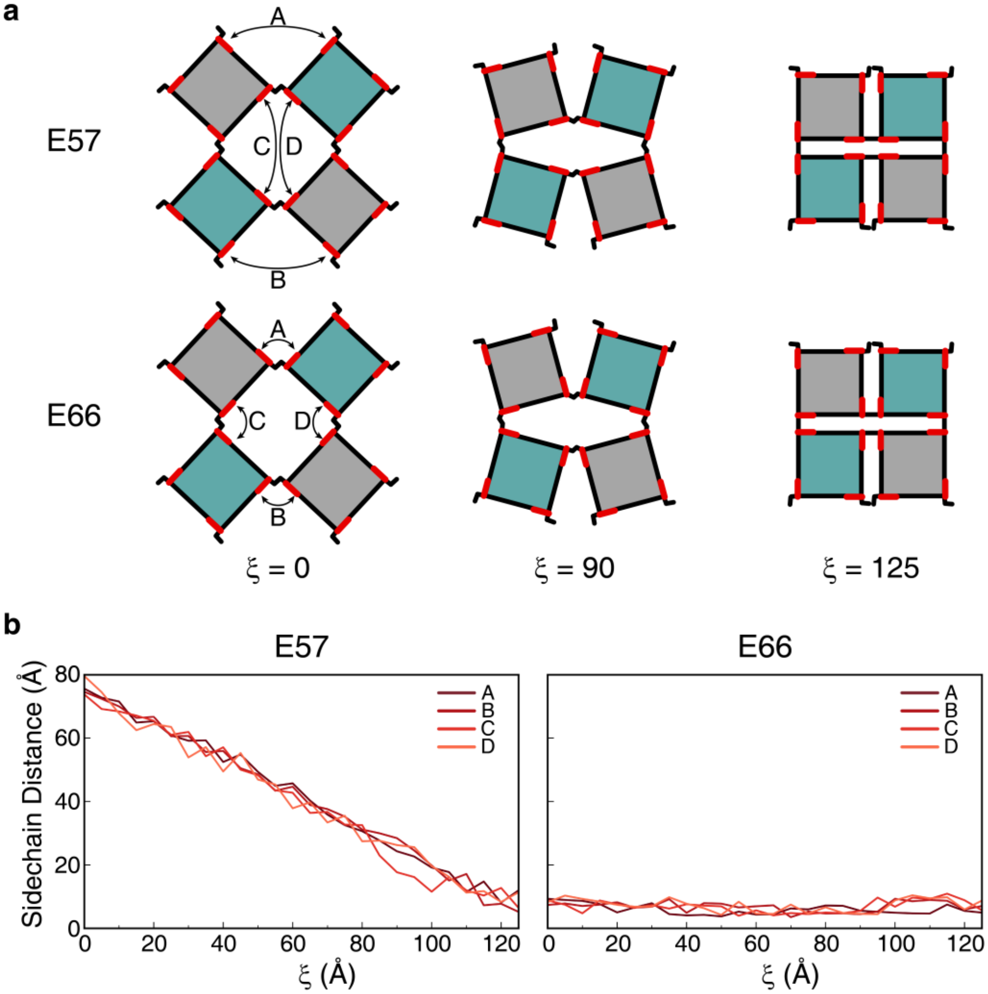


Figure 2.12 | Pairwise sidechain distances for ^{CEE}RhuA glutamates. a, Cartoons depicting pairs of glutamate sidechains expected to decrease in distance upon closure of the lattice, highlighting only one mutation at a time. **b,** Plots of the sidechain carboxylate-carboxylate distances illustrated above. These distances are calculated directly from the GIST reference structures (which were subject to repulsive electrostatic effects), and could often be reduced by substitution of alternative sidechain rotamers. While it may appear that the single mutation P57E would be sufficient to achieve the desired behavior of ^{CEE}RhuA crystals, its distance profile would resemble that of E66 (in the present example) were the pore stretched along its other principal axis. Although an advantage of the crystal symmetry is the facile design of self-interacting point mutations, its topology also dictates that neighboring pores close in opposite directions. Consequently, both mutations were deemed necessary to retain isotropic conformational changes across entire crystals.

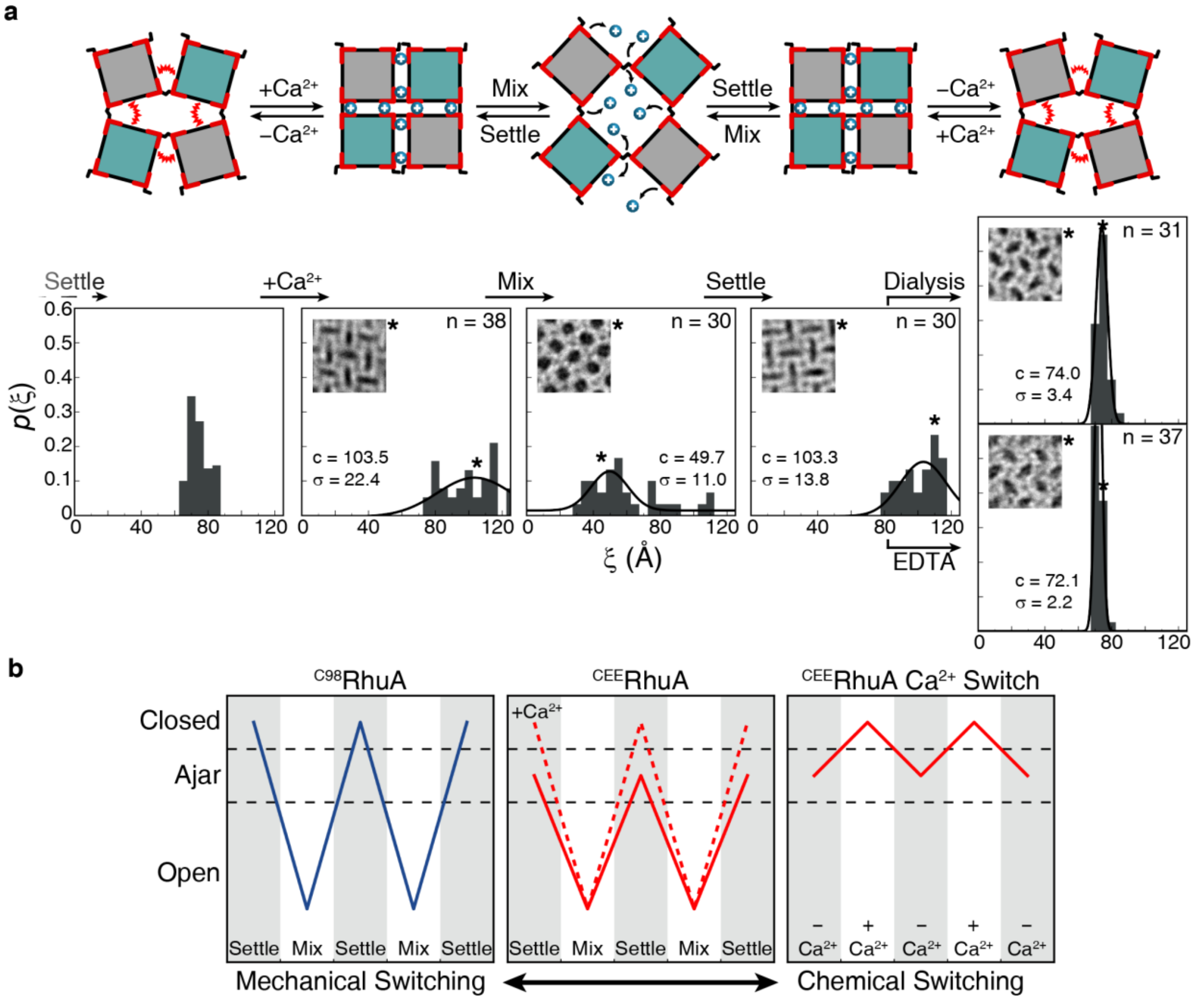


Figure 2.13 | Chemical and mechanical switching behavior of $^{\mathrm{CEE}}$RhuA crystals. a, Cartoon depicting all possible switching modes of $^{\mathrm{CEE}}$RhuA lattices, with each cartoon state directly corresponding to the experimental distribution(s) below it. Addition of 20 mM Ca^{2+} to the equilibrium "ajar" population of $^{\mathrm{CEE}}$RhuA crystals induces a shift towards more closed conformations, from which $^{\mathrm{C98}}$RhuA-like mechanical switching is possible. The ajar conformation is fully recoverable upon removal of Ca^{2+} via dialysis or EDTA, thus providing three distinct switching modes. **b,** Summary of switching modes for RhuA crystals. In contrast with $^{\mathrm{C98}}$RhuA, $^{\mathrm{CEE}}$RhuA has two mechanical modes dictated by the presence of Ca^{2+}, as well as a purely chemical mode via the addition/removal of Ca^{2+}.

2.4 Discussion

An essential feature of natural protein assemblies is their ability to adapt their structures and functions with high fidelity and coherence in specific response to environmental cues. Perhaps one of the best-known examples is the cooperative O_2 binding behavior of hemoglobin that is differentially regulated by pH or CO_2 concentration, as well as by specific metabolites like 2,3-biphosphoglyceric acid[35], although essentially every cellular protein is subject to some type of regulation by external stimuli. Accordingly, a major goal in the *de novo* design of proteins is to construct dynamic systems whose structural adaptability can be predicted and programmed at a molecular level. Toward this end, we have reported here the first example of a synthetic protein assembly whose free energy landscape was fully delineated and whose mechano- and chemo-responsive structural switching behavior was predictably engineered. In this regard, a key design element was the use of strong but reversible and flexible bonding interactions to assemble the 2D RhuA lattices. The small footprint of the disulfide bonds on the protein surfaces not only endowed the RhuA lattices with an immense range of conformational flexibility, but it also allowed other types of interactions (electrostatic/metal coordination) to be "dialed in" to predictably alter its free-energy landscape and control its structural dynamics. In an alternative strategy, one can envision the *en-masse* design of more extensive, non-covalent interfaces between protein subunits that could accommodate multiple discrete conformations which interchange upon application of stimuli of interest, as found in natural functional complexes. This remains an exciting but challenging goal for the future that will require considerable advances in the understanding and control of non-covalent interactions, as well as of protein-solvent interactions and entropy, as our study indicates. In addition to these general protein design considerations, the dynamic two-dimensionality of RhuA lattices also has specific functional implications. Owing to their high surface-to-volume ratios and uniform porosities, 2D nanomaterials are particularly well-suited for applications in separation/filtration[36], templating[37], sensing[38], catalysis[39], coatings[40], among many others[41]. The unique ability to control and tune the porosity or molecular density of RhuA lattices could offer important advantages in the application and fabrication of adaptive molecular devices.

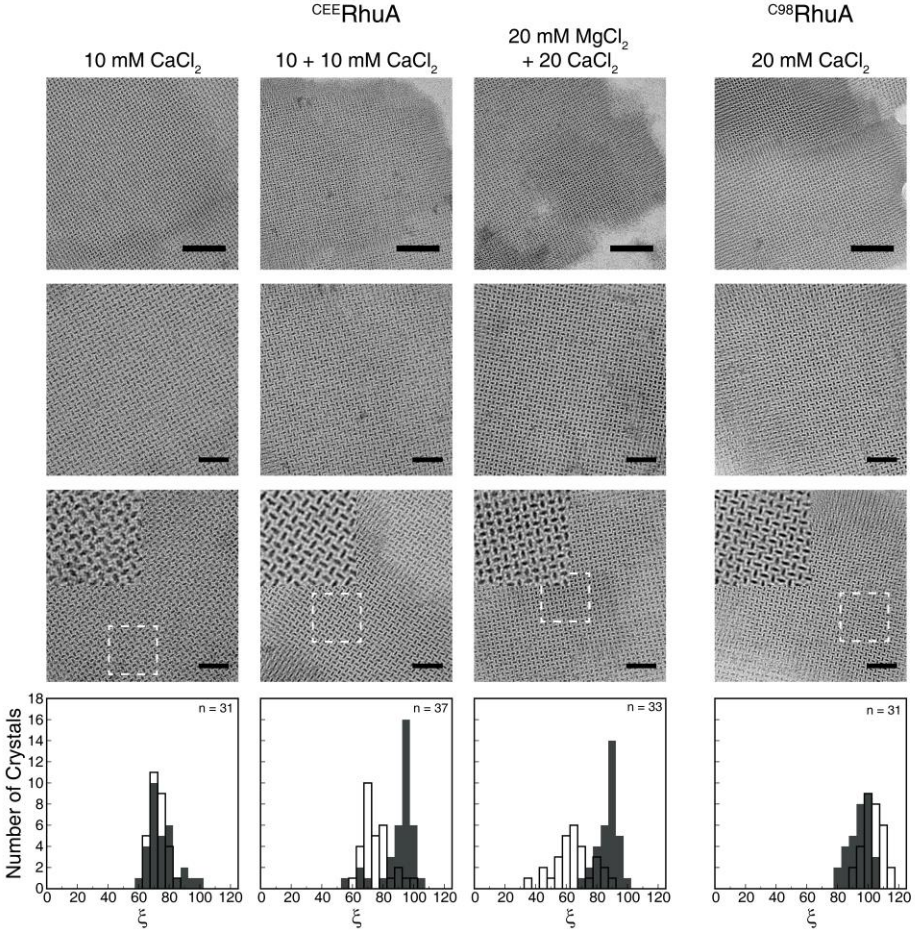


Figure 2.14 | TEM characterization of calcium-dependent switching behavior in ^{CEE}RhuA lattices. Selective and concentration-dependent effects of Ca^{2+} on ^{CEE}RhuA crystals. 10 mM $CaCl_2$ yields a slight rightward shift relative to the equilibrium ^{CEE}RhuA conformations (**left**, unfilled underlay). Subsequent treatment of these crystals (underlay) with an additional 10 mM $CaCl_2$ recovered a closed distribution (**middle**) resembling ^{CEE}RhuA crystals incubated with 20 mM $CaCl_2$ (**Figure 2.13a**). Although 20 mM $MgCl_2$ was not found to induce closure of ^{CEE}RhuA crystals (underlay), subsequent addition of 20 mM $CaCl_2$ to this sample yielded closed state crystals (**right**). As a control, we verified that the equilibrium state of ^{C98}RhuA crystals (underlay) was unchanged in the presence of 20 mM $CaCl_2$. Scale bars are 100 nm for the top row and 50 nm elsewhere. Insets: additional 2x magnifications of the white boxed regions.

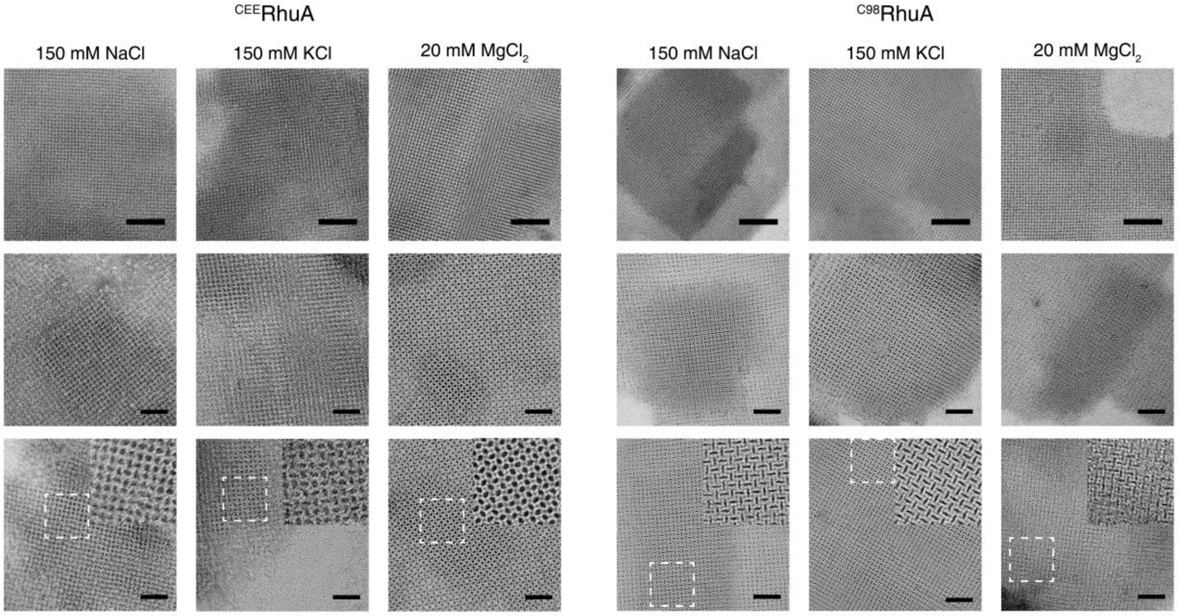


Figure 2.15 | TEM characterization of non-calcium additives on ^{CEE}RhuA lattices. Negative-stain micrographs of other additives screened for possible screening effects on the designed glutamate mutations. In all cases, the additive resulted in a negligible effect on the equilibrium conformations of RhuA crystals after a 3-day incubation. A statistical analysis of ^{CEE}RhuA crystals in the presence of 20 mM MgCl$_2$ is included in the results of **Figure 2.14**. Scale bars are 100 nm for the top row and 50 nm elsewhere. Insets: additional 2x magnification of the white boxed regions

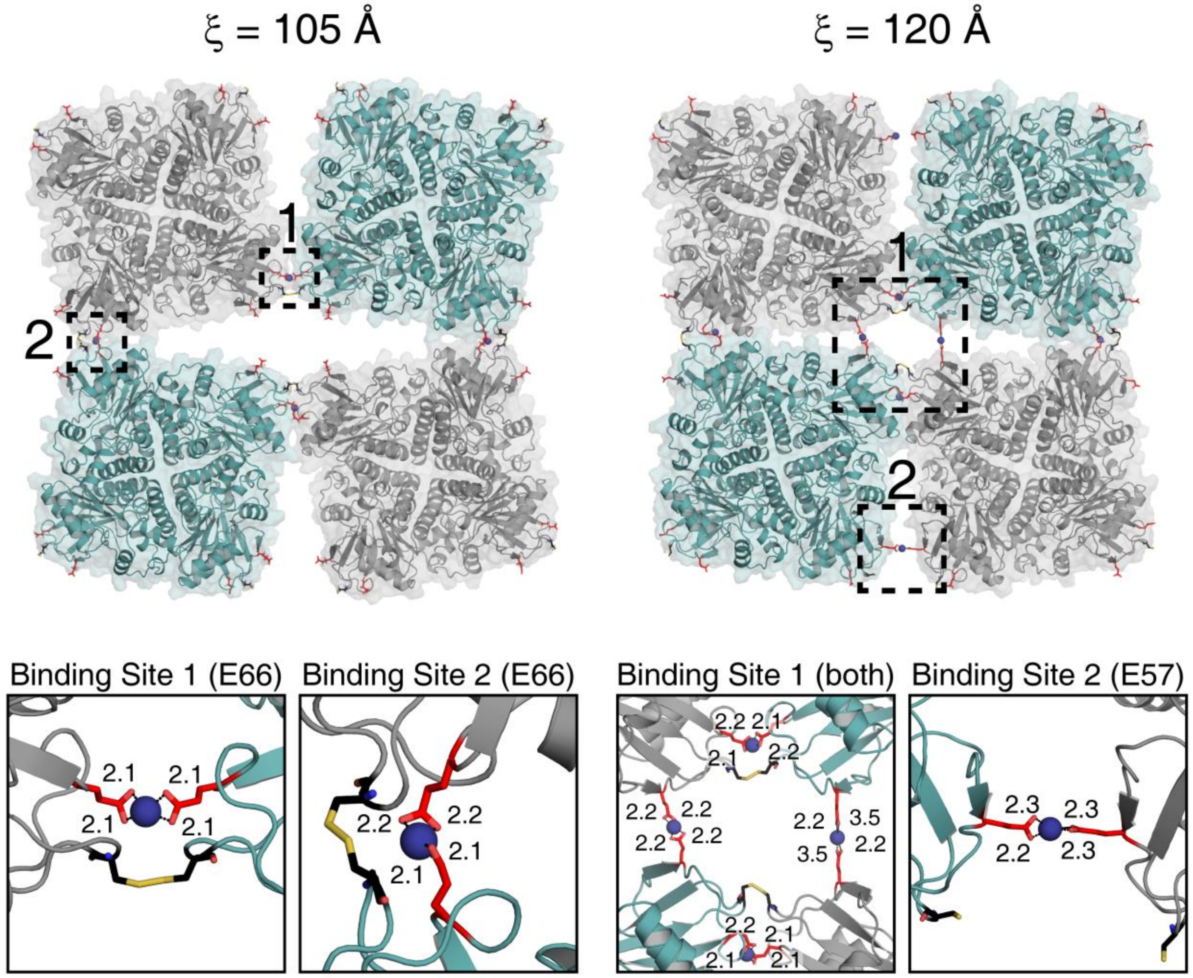


Figure 2.16 | Ca^{2+} binding models of $^{\mathrm{CEE}}$RhuA lattices. Model structures of putative Ca^{2+} binding modes accessible by $^{\mathrm{CEE}}$RhuA crystals at ξ = 105 and 120 Å (see Methods for full model building procedure). From the ajar state, E66 sidechains are poised for metal binding, and further compaction of the pore permits metal coordination by all glutamate pairs. The numbers in the lower panels are oxygen-calcium distances, reported in Å. Ca^{2+} ions are scaled to their divalent ionic radius of 1.14 Å.

2.5 Methods

2.5.1 Molecular modeling and simulations

The ^{C98}RhuA tetramer was initially constructed using a previously published crystal structure of RhuA (PDB ID: 1OJR), which harbors the E192A residue present in ^{C98}RhuA. Mutation of residue 98 to cysteine, repair of unmodeled backbone, and generation of the other 3 symmetry-related monomers was performed in PyMOL[42]. This complete ^{C98}RhuA tetramer was then oriented with its diagonals along the x and y axes, and copies were translated to generate an initial 2×2 starting structure. The ^{CEE}RhuA structure was generated from this same structure, whereby the mutation of P57 and A66 to glutamates was performed with PyMOL. The PSFGEN plugin of VMD[43] was used to add missing hydrogens, assign atom types, and join disulfide-linked C98 residues, yielding open-state starting structures for each RhuA crystal. Subsequently, the ^{C98}RhuA (^{CEE}RhuA) structure was inserted into a pre-equilibrated water box using the plugin SOLVATE, leaving 8 Å of solvent on all sides, and was neutralized by the addition of 64 (96) sodium ions. The final system consisted of 342,322 (342,267) atoms, of which 67,512 (67,608) were protein and Zn atoms. Periodic boundary conditions were imposed for all simulated systems.

Energy minimization and MD simulations were carried out using NAMD 2.10 with multi-core and CUDA support[44]. CHARMM36[45] parameters were used for all simulations. Temperature was regulated by Langevin dynamics with a damping coefficient of 1 ps^{-1}, and pressure was maintained at 1 atm via the Nosé-Hoover method with a period of 200 ps and decay of 100 ps. Non-bonded interactions were limited to 12.0 Å, with a switching function enforced at 10.0 Å, and long-range electrostatics were evaluated using the Particle mesh Ewald method. Hydrogen bond lengths were kept constant using the SHAKE[46] algorithm, and the trajectory was propagated using the velocity Verlet integration method using a timestep of 2 fs.

2.5.2 Umbrella sampling calculations

Each starting structure was minimized for 10,000 steps with all protein atoms held fixed, then subjected to a further 10,000 step minimization without restraints. The systems were thermalized at 300 K and subjected to a brief (1 ns) equilibration in the isobaric-isothermal (NPT) ensemble, at which time the protein structure was observed to have relaxed. For calculation of the reaction coordinate ξ, four vectors were calculated as the xy projection of the distances from the average position of the disulfide β carbons to the center of the pore (a_1, a_2, b_1, b_2), from which ξ was calculated as $(a_1+a_2) - (b_1+b_2)$. Initial configurations for each US window were generated via a 10 kcal/mol/Å^2 moving harmonic restraint on each distance, which smoothly adjusted the restraint centers to the target values (final $\xi = 130$ Å) over 5 ns. During this pulling trajectory, the z coordinates of all protein atoms were harmonically constrained at their initial positions using a 10 kcal/mol/Å^2 force constant. Initial coordinates for the US simulations were obtained by selection of frames from the pulling simulations based on ξ value. To loosely enforce planarity, two pairs of artificial boundary conditions were applied using the TclBC module of NAMD which served to restrict the Z coordinates of the protein alpha carbons to within ±1 Å of their equilibrated positions, using an applied force of 5 kcal/mol/Å to all alpha carbons outside of the boundaries (**Figure 2.4b**). Two sets were required because the symmetry of the crystal causes tetramers along each diagonal of the 2×2 structure to reside at offset z positions, and a single set of boundaries about the whole structure was found to be insufficient to effectively maintain planarity. Windows were maintained at their target positions using the 4 distances from the pulling simulations, which were found to significantly improve the overall symmetry of the pore (as expected within the bulk crystal) relative to application of a single restraint directly on the difference of principal axes *a* and *b*. Numerical calculations demonstrated that use of a 2 kcal/mol force constant on each distance

(a_1, a_2, b_1, b_2) yielded a distribution identical to a 0.5 kcal/mol force constant on the difference of the full distances ($\xi = a - b$; $a - b \equiv a_1 + a_2 - b_1 - b_2$) (**Figure 2.4c**). Due to the geometric advantages afforded by the former configuration, as well as its straightforward conversion to the desired final distributions, all US statistics were collected in this manner. A total of 30 ns of sampling was carried out for each window in order to permit full equilibration of the system and accumulation of sufficient statistics for a well-converged PMF.

2.5.3 Calculation of PMF and error bars

For each window of both RhuA variants (5 Å width; 26 windows each), sampling statistics were collected (as described above) every 200 timesteps from the colvars module of NAMD for a total of 75,000 values, of which the latter half were used for calculation of the respective PMFs. Final distributions of ξ were directly calculated from the raw statistics for each window (**Figure 2.4a**). The block averaging method (using 5 ns blocks) was employed to confirm that the system was fully equilibrated and thus suitable for generation of the PMF. Statistics were combined and reweighted using the WHAM algorithm in order to obtain the unbiased PMF, using a bin width of 0.5 Å[25]. Bootstrap error analysis yielded error values on the order of 0.05 kcal/mol for both protein constructs, so additional error analysis was performed based on the block averaging method published in ref. 24.

2.5.4 GIST simulations and calculations

To facilitate the direct comparison of results from GIST and US simulations, solute conformations for GIST analysis were selected directly from the latter. Each conformation was selected as the frame which minimized both the RMSD of the disulfide bonds to their idealized positions used for harmonic restraints and the asymmetry of the pore, which was evaluated as the ratio of each pair of vectors used for restraints (*e.g.*, minimize $a_1/a_2 - 1$ and $b_1/b_2 - 1$). This procedure maximized the symmetry of the system for more efficient averaging of the GIST results. The resulting protein-only structures were inserted into a slightly larger pre-equilibrated water box as done for the US simulations, with the final number of water molecules kept constant for all windows prior to neutralization (final counts: 104,711 waters for ^{C98}RhuA; 104,581 waters for ^{CEE}RhuA). As this discrepancy in number of waters between constructs is <0.1%, and bulk-like water does not contribute to the solvation entropy estimated by GIST, this is not expected to complicate interpretation of the GIST results for both constructs. For each window, minimization and equilibration was carried out in several steps as described previously[47]. First a 5,000-step minimization was carried out with all protein heavy atoms fixed, and this minimized structure was subsequently minimized again for 5,000 steps with all protein atoms fixed. For all subsequent runs, all protein atoms were constrained to these positions with a 10 kcal/mol/Å^2 force constant. The system was heated from 100 to 300 K in 5 steps over 100 ps at constant volume, followed by a 100 ps equilibration at constant pressure, after which the volume was found to have stabilized. The system was further equilibrated at constant pressure for an additional 1 ns, after which 20 ns of sampling was carried out for analysis.

GIST calculations were performed using the GIST analysis method as implemented within the cpptraj[48] program of AmberTools 16[49]. The trajectories were stripped of sodium counterions, and the CHARMM PSF files were converted using the chamber functionality of ParmEd. GIST analysis was carried out with a reference density of 0.0334 waters/Å^3 (calculated from a neat water box under identical simulation conditions), which ensured that the solvation entropy of pure water appropriately evaluates to 0. The system was discretized into cubic voxels of size 0.125 Å^3, on four equal quadrants of 202×202×182 voxels (each encompassing a single RhuA protein), for a total volume of 200×200×90 Å. The implementation of GIST in cpptraj ignores the outermost perimeter of voxels as part of the nearest-neighbor algorithm, so these dimensions were chosen to continuously cover the system. The trajectories were split into even and odd frames (1 ps/frame) to avoid a systematic overestimation of the entropy by the aforementioned algorithm, giving 8 equivalent quadrants for each window, from which the standard error of mean was calculated.

2.5.5 Configurational entropy calculations

The configurational entropy of RhuA proteins was estimated via the quasiharmonic (QH) approximation using the method described in ref. 50, as implemented within the cpptraj program of AmberTools 16[49]. Protein coordinates from 10 ns (10,000 frames) of the US simulations for windows 0, 80, 105, and 125 were extracted for analysis. Each protein was evaluated individually, both reducing computational expense and allowing the calculation of averages and errors. A trajectory of a single ^{C98}RhuA protein (with planarity boundary conditions as used for US) was simulated in the NPT ensemble to provide a "free protein" reference. The backbone (or heavy atom) coordinates of each protein were aligned to the backbone atoms of the first frame of their respective trajectory, from which an average structure was calculated. All coordinates were then realigned to this average structure and the mass-weighted covariance matrix was calculated and diagonalized to obtain the QH mode frequencies. The entropy of the first 10,000 modes was then evaluated using the analytical expression for the entropy of a quantum harmonic oscillator and summed to yield the final value.

2.5.6 Water structure calculations

Results for the average pore solvent density, radial number distribution, tetrahedral order parameter (q_{tetra}), and coordination number were obtained with *ad hoc* algorithms implemented in Tcl (for use with VMD) and Python from 4000 frames evenly distributed across the 20 ns GIST simulations. Maps of the solvent density within the pore were generated by binning the positions of water oxygens into 0.25×0.25 Å regions, from which the average density of water molecules was calculated and subsequently normalized to bulk density (as obtained from simulation of the neat water box used for solvation under identical NPT conditions). The solvent radial number distributions n(r) were obtained by determining of the number of waters between (r – Δr) and r from any atom on the protein surface, where Δr is 0.25 Å. The order parameter q_{tetra}, which describes the overall tetrahedrality of a given water molecule to its nearest 2-4 neighbors, ranges from 1 for a perfect tetrahedral arrangement and 0 for a random network. q_{tetra} was calculated using the modified form defined by Godec and coworkers[51]: $q_i = 1 - \frac{1}{N(N-1)}\frac{9}{8}\sum_{j=1}^{N-1}\sum_{k=j+1}^{N}(\cos\varphi_{jik} + \frac{1}{3})^2$, for $2 \leq N \leq 4$. The distributions for $p(q_{tetra})$ were then binned into increments of width 0.1. Coordination number was calculated as the number of other water molecules within 3.5 Å of a given solvent molecule and serves as an indicator for the trapping of water molecules as well as a measure of relative solvent density.

2.5.7 Protein expression and self-assembly

Purification of ^{C98}RhuA was carried out following the method published previously[20], while ^{CEE}RhuA required minor modification of the procedure. All purification steps were carried out at 4 °C. Briefly, expression was carried out in BL21 (DE3) *Escherichia coli* cells, and induced with 1mM IPTG for 12 hours before harvesting by centrifugation. Cells were resuspended in a buffer of 20 mM Tris (pH 7.5), 1 mM $ZnCl_2$, and 10 mM βME and subjected to sonication. Treatment of the crude lysate with Polymin-P was found to aggregate the more negatively charged ^{CEE}RhuA, so this step was not included in the purification. Instead, the pH of the lysate was slowly reduced to 5 using HCl, gently stirred for 30 min at 4 °C, then centrifuged at 10,500 RPM for 45 min at 4 °C. After the supernatant was decanted, the pH was brought back to 7.5 with NaOH, and the clarified supernatant was run on a DEAE column. Peak fractions were pooled and subjected to ammonium sulfate precipitation, then repeatedly dialyzed into a buffer of 20 mM sodium acetate (pH 5), 10 mM βME, and purified over an S resin via FPLC, which yielded pure protein. The purified protein was dialyzed into a solution of 20 mM Tris (pH 7.5) and 10 mM βME, concentrated to 100-150 µM, aliquoted, and flash-frozen in liquid nitrogen for storage at −80 °C.

Self-assembly of ^{CEE}RhuA crystals was carried out under solution conditions very similar to those published for ^{C98}RhuA[20]. Purified ^{CEE}RhuA was exchanged into a freshly prepared solution of 20 mM Tris (pH 7.5), 1 mM $ZnCl_2$, and 5-10 mM βME (final [^{CEE}RhuA] = 100-125 µM), and subjected to continuous gentle shaking at 4 °C. Nucleation of ^{CEE}RhuA crystals typically occurred over a period of a few days, and the suspensions were allowed to mature up to several weeks. Mature ^{CEE}RhuA crystals were commonly observed to be square or rectangular, with edge lengths in the range of 1-1.5 µm, although they possessed slightly less well-defined edges relative to ^{C98}RhuA.

2.5.8 TEM imaging and image processing

For analysis by negative-stain TEM, 3.5 µL aliquots of ^{CEE}RhuA crystals were applied to negative mode glow-discharged carbon-coated copper grids (Ted Pella, Inc), washed with MilliQ water, and stained with filtered 2% uranyl acetate solution. Data were collected using an FEI Sphera transmission electron microscope operating at 200 keV, equipped with a LaB_6 filament and Gatan 4K CCD. Images for the statistical analysis of crystal conformations were captured at a nominal magnification of 150,000x in numbers of ≥100 and ≥30 for the investigation of mechanical reversibility and metal-binding, respectively. Conformational state analysis was performed using the same macro described previously[20] in the program Fiji (http://fiji.sc/Fiji), which generates a binary image via thresholding from which pores in the lattice can be identified. Pores with areas of 1200-4800 pixels were kept for analysis, which comprised fitting an ellipse to each pore, and calculating the ellipticity E as the ratio of the short axis to the long axis. Due to the uniform conformation of each crystal, the average ellipticity of all measured pores in the crystal represents a single-crystal measurement for statistical analysis. The resulting values were then converted to units of ξ via the linear relation $E = -0.0063\xi + 1.0$ (**Figure 2.3**) to permit direct comparison with computational results. Histograms of the conformational distributions were generated from the converted values binned into 5-Å intervals. Gaussian fits to the distributions were calculated using OriginPro 2017.

2.5.9 Additive screening with ^{CEE}RhuA crystals

For all solution conditions tested for interaction with ^{CEE}RhuA crystals, stocks of various additives (NaCl, KCl, $MgCl_2$, $CaCl_2$) were dissolved in a solution of 20 mM Tris (pH 7.5) at double the final concentration desired for screening and passed through 0.2 μm cellulose filters. These stocks were then mixed with a suspension of ^{CEE}RhuA crystals (also in 20 mM Tris at pH 7.5) at a 1:1 v/v ratio. Tubes of ^{CEE}RhuA crystals + additives were allowed to stand for 3 days before beginning TEM analysis. For the treatment of ^{CEE}RhuA crystals with EDTA, an equimolar amount of EDTA was added via a Tris-buffered solution of EDTA (pH 7.5). For Ca^{2+} removal by dialysis, 25-50 μL of suspended crystals was dialyzed against 50 mL of 20 mM Tris pH 7.5, for a 1000-2000 fold exchange of solution.

2.5.10 Construction of Ca^{2+} binding model for ^{CEE}RhuA

The structural models were built from the ^{CEE}RhuA GIST reference structures for $\xi = 105$ and 120 Å by placement of 4 and 8 Ca^{2+} ions at the centers of the E66 and E57/E66 carboxylate oxygens. The sidechain atoms and Ca^{2+} ions were then subjected to two 5000 step minimizations *in vacuo* (with all other atoms held fixed): first with all O-Ca^{2+} distances harmonically restrained to 2.1 Å using a 10 kcal/mol/$Å^2$ force constant, which established initial coordination geometry, then again without restraints to allow adoption of preferred sidechain conformations.

2.6 Acknowledgments

We thank M. Gilson, T. Kurtzman, and S. Ramsey for helpful discussions regarding GIST, R. Subramanian for assistance with generation of projection maps from computational models, and T. Baker for use of EM facilities. This work was primarily supported by the US Department of Energy (DOE) (Division of Materials Sciences, Office of Basic Energy Sciences, Award DE-SC0003844 to F.A.T.). F.P. was supported by the National Science Foundation through grant CHE-1453204 (computation). R.A. was supported in part by the UCSD NIH Molecular Biophysics Training Grant (T32-GM08326). All computer simulations were performed on the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation through grant ACI-1053575.

Chapter 2 is reproduced, in part, with permission, from: Alberstein, R.G., Suzuki, Y., Paesani, F., Tezcan, F.A. "Engineering the entropy-driven free-energy landscape of a dynamic nanoporous protein assembly", *Nature Chemistry* **10**, 732–739 (2018). The dissertation author was the primary author on all reprinted materials.

2.7 References

1. Marsh, J. A. & Teichmann, S. A. Structure, Dynamics, Assembly, and Evolution of Protein Complexes. *Annu. Rev. Biochem.* **84**, 551-575, (2015).

2. Salgado, E. N., Radford, R. J. & Tezcan, F. A. Metal-directed protein self-assembly. *Acc. Chem. Res.* **43**, 661–672, (2010).

3. Song, W. J. & Tezcan, F. A. A designed supramolecular protein assembly with *in vivo* enzymatic activity. *Science* **346**, 1525–1528, (2014).

4. Yeates, T. O. Geometric Principles for Designing Highly Symmetric Self-Assembling Protein Nanomaterials. *Annu. Rev. Biophys.* **46**, 23-42, (2017).

5. Bale, J. B., Gonen, S., Liu, Y., Sheffler, W., Ellis, D., Thomas, C., Cascio, D., Yeates, T. O., Gonen, T., King, N. P. & Baker, D. Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science* **353**, 389-394, (2016).

6. Koga, N., Tatsumi-Koga, R., Liu, G., Xiao, R., Acton, T. B., Montelione, G. T. & Baker, D. Principles for designing ideal protein structures. *Nature* **491**, 222-227, (2012).

7. Norn, C. H. & André, I. Computational design of protein self-assembly. *Curr. Opin. Struct. Biol.* **39**, 39-45, (2016).

8. Ringler, P. & Schulz, G. E. Self-Assembly of Proteins into Designed Networks. *Science* **302**, 106-109, (2003).

9. Sinclair, J. C., Davies, K. M., Venien-Bryan, C. & Noble, M. E. M. Generation of protein lattices by fusing proteins with matching rotational symmetry. *Nat. Nanotechnol.* **6**, 558–562, (2011).

10. Levy, Y. & Onuchic, J. N. Water Mediation in Protein Folding and Molecular Recognition. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 389-415, (2006).

11. Young, T., Abel, R., Kim, B., Berne, B. J. & Friesner, R. A. Motifs for molecular recognition exploiting hydrophobic enclosure in protein–ligand binding. *Proc. Natl. Acad. Sci.* **104**, 808-813, (2007).

12. Hedstrom, L. Serine Protease Mechanism and Specificity. *Chem. Rev.* **102**, 4501-4524, (2002).

13. Davey, J. A., Damry, A. M., Goto, N. K. & Chica, R. A. Rational design of proteins that exchange on functional timescales. *Nat. Chem. Biol.* **13**, 1280, (2017).

14. Joh, N. H., Wang, T., Bhate, M. P., Acharya, R., Wu, Y., Grabe, M., Hong, M., Grigoryan, G. & DeGrado, W. F. De novo design of a transmembrane Zn^{2+}-transporting four-helix bundle. *Science* **346**, 1520-1524, (2014).

15. Cerasoli, E., Sharpe, B. K. & Woolfson, D. N. ZiCo: A Peptide Designed to Switch Folded State upon Binding Zinc. *J. Am. Chem. Soc.* **127**, 15008-15009, (2005).

16. Ambroggio, X. I. & Kuhlman, B. Computational design of a single amino acid sequence that can switch between two distinct protein folds. **128**, 1154-1161, (2006).

17. Brodin, J. D., Ambroggio, X. I., Tang, C. Y., Parent, K. N., Baker, T. S. & Tezcan, F. A. Metal-directed, chemically tunable assembly of one-, two- and three-dimensional crystalline protein arrays. *Nat. Chem.* **4**, 375–382, (2012).

18. Sontz, P. A., Bailey, J. B., Ahn, S. & Tezcan, F. A. A metal organic framework with spherical protein nodes: rational chemical design of 3D protein crystals. **137**, 11598–11601, (2015).

19. Bailey, J. B., Subramanian, R. H., Churchfield, L. A. & Tezcan, F. A. Metal-Directed Design of Supramolecular Protein Assemblies. *Methods Enzymol.* **580**, 223-250, (2016).

20. Suzuki, Y., Cardone, G., Restrepo, D., Zavattieri, P. D., Baker, T. S. & Tezcan, F. A. Self-assembly of coherently dynamic, auxetic, two-dimensional protein crystals. *Nature* **533**, 369–373, (2016).

21. Grima, J. N. & Evans, K. E. Auxetic behavior from rotating squares. *J. Mater. Sci. Lett.* **19**, 1563-1565, (2000).

22. Yang, W., Li, Z.-M., Shi, W., Xie, B.-H. & Yang, M.-B. Review on auxetic materials. *J. Mater. Sci.* **39**, 3269-3279, (2004).

23. Evans, K. E. & Alderson, A. Auxetic Materials: Functional Materials and Structures from Lateral Thinking! *Adv. Mater.* **12**, 617-628, (2000).

24. 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).

25. Kumar, S., Rosenberg, 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).

26. De Yoreo, J. J. & Vekilov, P. G. Principles of Crystal Nucleation and Growth. *Rev. Mineral Geochem.* **54**, 57-93, (2003).

27. Derewenda, Z. S. & Vekilov, P. G. Entropy and surface engineering in protein crystallization. *Acta Crystallogr. D* **62**, 116-124, (2006).

28. Nguyen, C. N., Young, T. K. & Gilson, M. K. Grid inhomogeneous solvation theory: Hydration structure and thermodynamics of the miniature receptor cucurbit[7]uril. *J. Chem. Phys.* **137**, 044101, (2012).

29. Huggins, D. J. Comparing distance metrics for rotation using the k-nearest neighbors algorithm for entropy estimation. *J. Comput. Chem.* **35**, 377-385, (2014).

30. Giovambattista, N., Rossky, P. J. & Debenedetti, P. G. Effect of Temperature on the Structure and Phase Behavior of Water Confined by Hydrophobic, Hydrophilic, and Heterogeneous Surfaces. *J. Phys. Chem. B.* **113**, 13723-13734, (2009).

31. Tarek, M. & Tobias, D. J. The Dynamics of Protein Hydration Water: A Quantitative Comparison of Molecular Dynamics Simulations and Neutron-scattering Experiments. *Biophys. J.* **79**, 3244-3257, (2000).

32. Ebbinghaus, S., Kim, S. J., Heyden, M., Yu, X., Heugen, U., Gruebele, M., Leitner, D. M. & Havenith, M. An extended dynamical hydration shell around proteins. *Proc. Natl. Acad. Sci.* **104**, 20749-20752, (2007).

33. Heyden, M., Sun, J., Funkner, S., Mathias, G., Forbert, H., Havenith, M. & Marx, D. Dissecting the THz spectrum of liquid water from first principles via correlations in time and space. *Proc. Natl. Acad. Sci.* **107**, 12068-12073, (2010).

34. Dunitz, J. D. The Entropic Cost of Bound Water in Crystals and Biomolecules. *Science* **264**, 670-670, (1994).

35. Yuan, Y., Tam, M. F., Simplaceanu, V. & Ho, C. New Look at Hemoglobin Allostery. *Chem. Rev.* **115**, 1702-1724, (2015).

36. Joshi, R. K., Carbone, P., Wang, F. C., Kravets, V. G., Su, Y., Grigorieva, I. V., Wu, H. A., Geim, A. K. & Nair, R. R. Precise and Ultrafast Molecular Sieving Through Graphene Oxide Membranes. *Science* **343**, 752-754, (2014).

37. Tahara, K., Nakatani, K., Iritani, K., De Feyter, S. & Tobe, Y. Periodic Functionalization of Surface-Confined Pores in a Two-Dimensional Porous Network Using a Tailored Molecular Building Block. *ACS Nano* **10**, 2113-2120, (2016).

38. Huber, C., Liu, J., Egelseer, E. M., Moll, D., Knoll, W., Sleytr, U. B. & Sára, M. Heterotetramers Formed by an S-Layer–Streptavidin Fusion Protein and Core-Streptavidin as a Nanoarrayed Template for Biochip Development. *Small* **2**, 142-150, (2006).

39. Li, Y., Wang, H., Xie, L., Liang, Y., Hong, G. & Dai, H. MoS2 Nanoparticles Grown on Graphene: An Advanced Catalyst for the Hydrogen Evolution Reaction. *J. Am. Chem. Soc.* **133**, 7296-7299, (2011).

40. Mader, C., Küpcü, S., Sleytr, U. B. & Sára, M. S-layer-coated liposomes as a versatile system for entrapping and binding target molecules. *Biochim. Biophys. Acta. - Biomembranes* **1463**, 142-150, (2000).

41. Butler, S. Z., Hollen, S. M., Cao, L., Cui, Y., Gupta, J. A., Gutiérrez, H. R., Heinz, T. F., Hong, S. S., Huang, J., Ismach, A. F., Johnston-Halperin, E., Kuno, M., Plashnitsa, V. V., Robinson, R. D., Ruoff, R. S., Salahuddin, S., Shan, J., Shi, L., Spencer, M. G., Terrones, M., Windl, W. & Goldberger, J. E. Progress, Challenges, and Opportunities in Two-Dimensional Materials Beyond Graphene. *ACS Nano* **7**, 2898-2926, (2013).

42. Schrodinger, LLC. *The PyMOL Molecular Graphics System, Version 1.3* (2010).

43. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33-38, (1996).

44. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L. & Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781-1802, (2005).

45. 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).

46. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327-341, (1977).

47. Nguyen, C. N., Cruz, A., Gilson, M. K. & Kurtzman, T. Thermodynamics of Water in an Enzyme Active Site: Grid-Based Hydration Analysis of Coagulation Factor Xa. *J. Chem. Theory. Comput.* **10**, 2769-2780, (2014).

48. Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory. Comput.* **9**, 3084-3095, (2013).

49. D.A. Case, R. M. B., D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, L. Xiao and P.A. Kollman (2016). AMBER 2016. University of California, San Francisco.

50. Andricioaei, I. & Karplus, M. On the calculation of entropy from covariance matrices of the atomic fluctuations. *J. Chem. Phys.* **115**, 6289-6292, (2001).

51. Godec, A., Smith, J. C. & Merzel, F. Increase of both Order and Disorder in the First Hydration Shell with Increasing Solute Polarity. *Phys. Rev. Lett.* **107**, 267801, (2011).

52. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci.* **98**, 10037-10041, (2001).

53. Dolinsky, T. J., Czodrowski, P., Li, H., Nielsen, J. E., Jensen, J. H., Klebe, G. & Baker, N. A. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* **35**, W522-W525, (2007).

Chapter 3: A photonic porous silicon sensor with integrated biological gatekeeper for remote detection of HCN vapor

3.1 Abstract

The utility of chemical sensors in real-world environments (*i.e.*, outside of controlled laboratory settings) is fundamentally tied to their ability to distinguish specific (or classes of) target analytes. The stringency of this criterion becomes pronounced, for example, in industrial settings which require careful monitoring for the presence of noxious agents among chemically complex multi-component atmospheres, where sensing fidelity can become a determinant of life-or-death situations. Here we report the design and characterization of an HCN-selective optical sensor reinforced with a stimuli-responsive biological framework, which serves as a protective coating against a broad range of chemical species, that is rendered semi-permeable only in the presence of the prescribed target. Reliable detection was achieved for minute quantities (≥5 ppm) of HCN, even in the presence of significant quantities of (up to 43 unique) non-target species, exhibiting a rapid temporal response that could be remotely quantified by optical measurement. These capabilities were achieved through the careful exploitation of both the highly tunable spectral properties of the porous silicon sensor substrate and the precision afforded by biomolecular re-engineering of protein-based materials, suggesting that this strategy should be broadly applicable for the rational design of new sensors capable of detecting arbitrary analytes with high precision.

3.2 Introduction

The real-time detection and collection/passivation of small molecule toxicants represents an essential topic of research with general application in industrial environments[1]. However, a significant limitation of current technologies is the saturation of the absorbent by non-target analytes ("clutter") present in the ambient atmosphere[1-3]. Take, for example, activated charcoal (one of the most commonly appreciated filtration mediums), where one might imagine the nonspecific sorption of chemical species to be quite beneficial. This is certainly the case in personal respirators or water filtration devices, where such technologies serve to remove as many organic contaminants as possible in order to protect the user. However, the corresponding lack of specificity results in the rapid nonspecific saturation of the absorbent bed with *e.g.*, water (vapor), leading to a substantial decrease in performance over relatively short time periods[4]. While these limitations may be tolerable in such use cases, where the sorbent material is replaced at predefined intervals or on as-needed basis when coupled to monitoring systems, such restrictions significantly undercut their utility as passive sensors or environmental monitors[1,5-8].

Spurred on by these limitations, a burgeoning area of research involves the development of mesoporous nanostructures capable of undergoing stimuli-responsive changes to their chemical or physical properties[9-17]. Many of these capabilities involve the selective sorption of target molecules, which can be viewed as the functional complement to drug-delivery systems, where therapeutic agents are sealed within porous structures that localize to targeted regions (with high specificity) before releasing their payload into the surrounding tissue, thus minimizing off-target effects associated with simple internalization of the drug[17-20]. By applying this inverted perspective to absorbent materials, it should be possible to generate robust passive sensors consisting of high-sensitivity detection modalities enveloped by physical, chemical, or biological “gatekeeper” functionalities that toggle their structural porosity in response to a specified (or class of) analyte. In this way, sensing is initiated by the target, directly “activating” the device for molecular uptake by permitting selective infiltration into the absorbent bed while eschewing ambient “clutter”.

The selective gating of analytes through molecular membranes is a hallmark of biological systems (*e.g.*, ion transport across lipid bilayers via protein channels and pumps[21,22]), where the precise placement of chemical groups within well-defined structures can result in selectivity sufficiently stringent to differentiate monovalent and divalent ions with near-elemental accuracy (*e.g.*, Na^+/Li^+ vs K^+, Ca^{2+} vs Mg^{2+})[23-25] based on small differences in size[23,26] and hydration structure[27,28]. Indeed, synthetic membranes constructed using proteins have been demonstrated to exhibit high-specificity separation capabilities[29-31]. However, these separation processes all occur within aqueous environments, which may be sufficient for solution-sensing purposes, but may fail in dehydrated conditions for gating of vapor-phase molecules. The latter role is commonly filled by metal organic frameworks (MOFs), which are widely used for gas separation but also typically exhibit more modest selectivity[32]. We thus sought to construct an integrated chemical sensor where an engineered biological membrane (whose porosity can be toggled by a specific analyte) serves as a gatekeeper to form a bio-inorganic heterojunction.

3.3 Results and Discussion

3.3.1 Design of a pSi-based sensor with HCN-selective gatekeeper coating

As our detection modality (and physical substrate) we elected to utilize photonic porous silicon (pSi) wafers, pristine semiconductor-grade silicon chips that have been electrochemically etched using sinusoidal current amplitudes to generate a surface layer of optically active porous microstructures along their (100) face. This technology has been successfully employed for the spectroscopic detection of numerous gaseous toxins with high sensitivity, including volatile organic compounds[33,34], chlorine[35], and nerve agents[36]. Porous silicon substrates offer numerous advantages as passive sensors[37]: they produce high intensity photonic signals which can be precisely tuned by etching conditions (permitting broad-spectrum label-free detection capabilities)[33,38,39], their porous microstructure provides very high surface areas (≥200 m^2/g) for chemical sorption and sequestration, they tolerate chemical functionalization with a wide variety of reagents[34,40], can be physically converted into alternative form factors such as free-standing thin films[39] or particles[41,42], and have been employed for drug delivery applications *in vivo*[41].

Key to this broad utility is the ability to control the properties of pSi structures during manufacture. To generate the porous surface layer(s), chemical etching of the wafer surface (accomplished via a 3:1 (v/v) solution of 48% aqueous hydrogen fluoride (HF) and ethanol) is precisely modulated via computer-controlled electrical current to yield nano-to-mesoporous photonic crystal thin films, where pore depth, size, and spectral properties are tuned via the current amplitude waveform[37]. Sinusoidal current waveforms yield “rugate reflectors” that exhibit high optical reflectance at a specific wavelength (a “stop band”) (**Figure 3.1**). This behavior is characteristic of photonic crystals, which possess periodic (“crystalline”) microstructures on the order of visible light to near-infrared wavelengths, causing them to interact strongly with incident photons (making them “photonic”)[43]. Such optical nanostructures are not uncommon in nature, ranging from butterflies (whose wings are decorated with nanometer-sized scales) and the characteristic iridescence of nacre (making lustrous both pearls and the oysters who produce them alike), to opals (which are not solid rock, but rather densely packed spheres of silica on the order of a few hundred nm each). Importantly, this renders pSi rugate filters exceptionally sensitive to the ambient environment within the pores (which affect *e.g.*, the refractive index through which the light travels), and is the fundamental principle upon which optical gas-phase sensing with pSi is realized. However, it is also inherently nonspecific for any particular analyte[33,34,39,44].

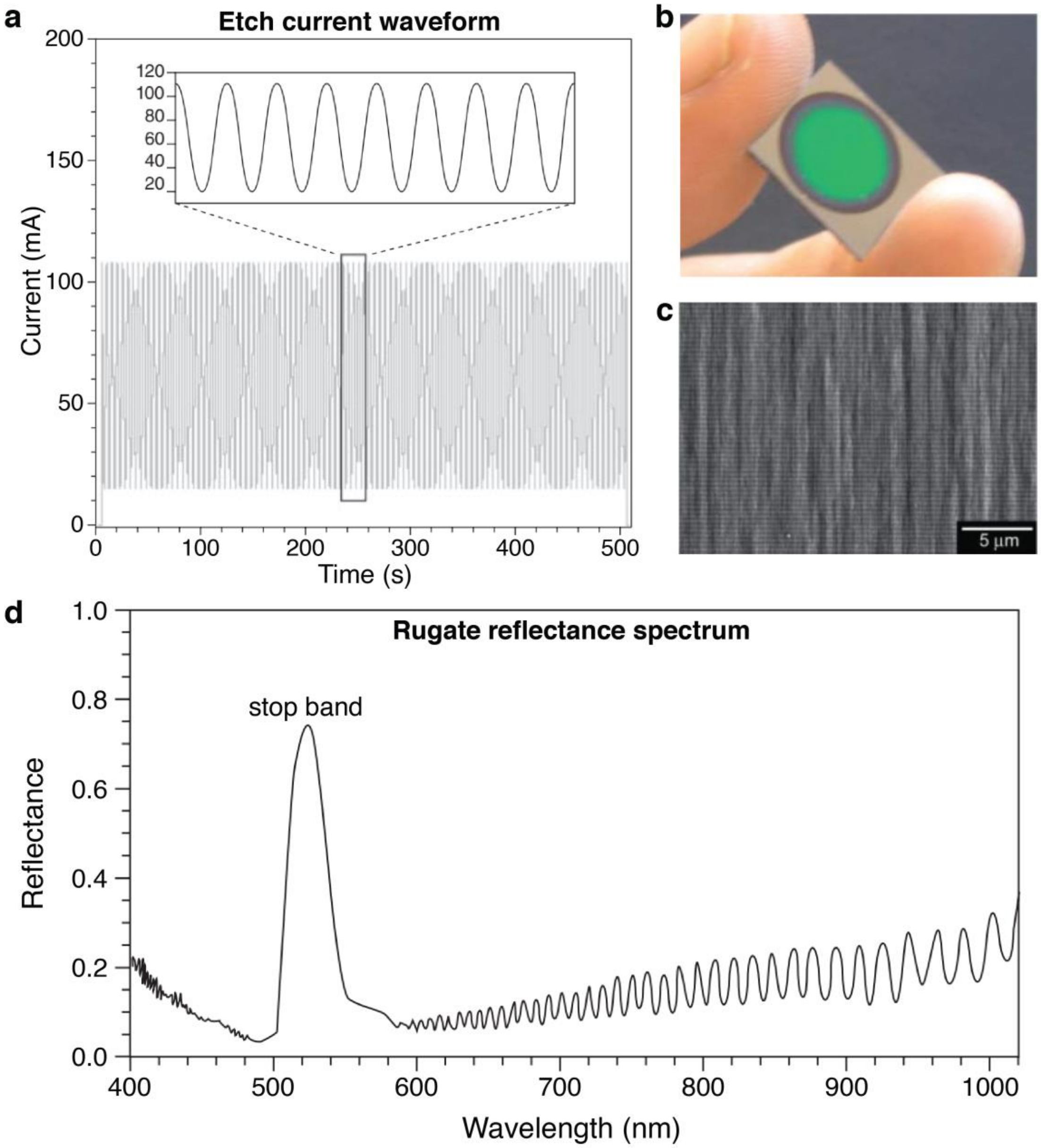


Figure 3.1 | Construction principles of porous silicon rugate filters. a, Representative current vs time waveform used to electrochemically etch a pSi photonic crystal. **b,** Representative image of such a microstructure etched into a silicon wafer. **c,** Cross-sectional image of such a wafer. **d,** Representative optical reflectance spectrum of a pSi rugate filter, featuring a single stop band and Fabry-Pérot fringing. Note that the spectral peak from 500–550 nm corresponds to green visible light, as seen in **b**. Adapted from ref. 37.

We adopted a two-pronged design strategy to augment our sensors with high chemical specificity, using hydrogen cyanide (HCN) gas as a representative target toxin that is commonly produced and used in large quantities in industrial environments. First, selective detection would be achieved following a previously demonstrated colorimetric method[45]. Here we use cobinamide (Cbi), a vitamin B_{12} analog with very high affinity and rapid binding kinetics for CN^- (so much so that Cbi is used medically to treat cyanide poisoning[46]) as an indicator dye[47-49]. Upon complexation with CN^-, Cbi undergoes a colorimetric shift from red to purple, with a major absorption peak emerging at ~583 nm[47-49]. A composite etch waveform yields a rugate filter with two reflectance peaks, a “signal” peak at ~580 nm which overlaps with the $(CN)_2$Cbi absorption spectrum and a near-infrared reference “fiducial” peak at ~700 nm. The resulting decrease in optical intensity at 583 nm (normalized to the reference peak) serves to enable selective detection of HCN[35,45] (**Figure 3.2**). This methodology presents an additional advantage in that the optical readout renders it amenable to remote detection, thus protecting the user from possible exposure to toxic vapor.

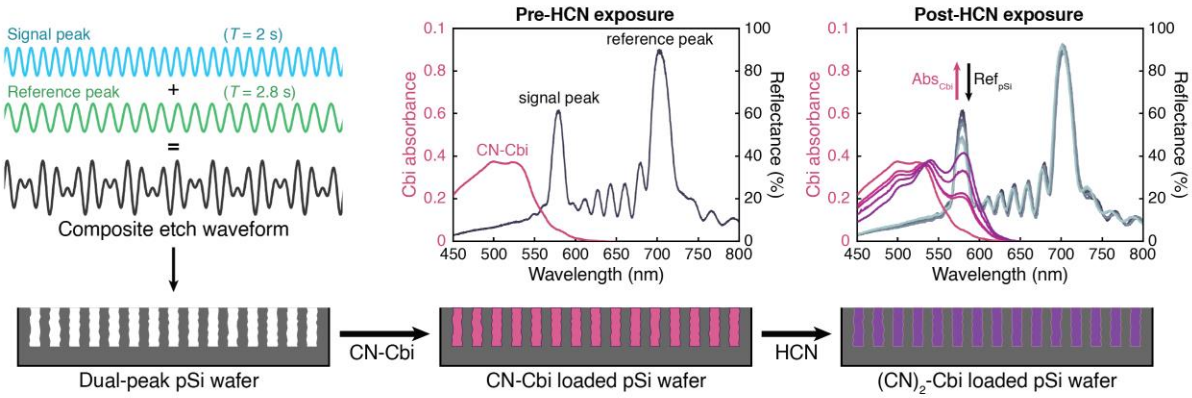


Figure 3.2 | Mechanism of HCN detection via dye-loaded pSi wafers. Electrochemical etching using a composite waveform yields a dual-peak pSi rugate filter displaying two stop bands. The pores are filled with monocyano-cobinamide (CN-Cbi), whose absorbance spectrum does not overlap with either stop band. Upon binding another CN^-, the $(CN)_2$Cbi absorption spectrum competes with the signal stop band, enabling selective detection of HCN.

Next, we sought to protect the sensor against non-target gases by coating the wafer with a stimuli-responsive "gatekeeper" functionality. This structure would perform broad-spectrum rejection of chemical species by default, remaining in a nonporous state until exposure to HCN triggered "gate opening", thus only permitting access to the substrate when our target analyte is present (**Figure 3.3**). In addition to possessing stimuli-responsive behavior, this coating must also achieve high surface coverage while minimizing the total quantity of material, because: 1) photons must be able to pass through the coating to enable detection and preserve optical intensity, and 2) excessively thick coatings would hinder gas permeation and may sacrificially uptake free analyte (gate opening events) without a corresponding color change, reducing detection sensitivity. These requirements suggested that the optimal gatekeeper would be a 2D material (ideally of monolayer thickness) that is capable of switching between porous and nonporous states in a selective manner.

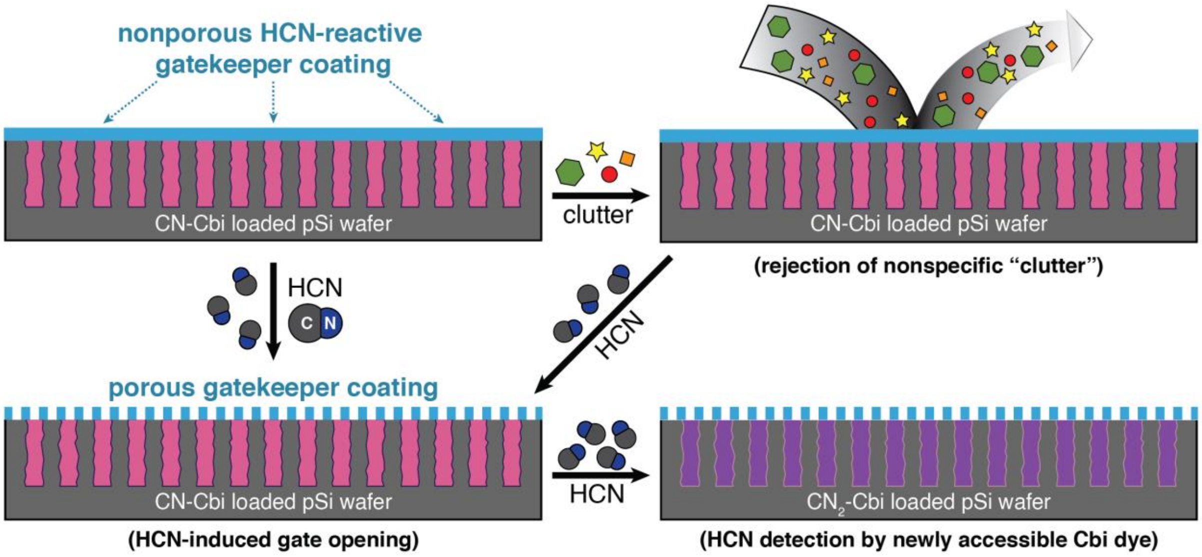


Figure 3.3 | Selective sorption and detection via gatekeeper technology. Illustration of the gatekeeping principle as it applies to the pSi sensor described here. A reactive coating provides passive, continuous, broad-spectrum rejection of arbitrary substances in the local environment until the target analyte triggers "opening" of the gate via chemical or mechanical means.

To meet these stringent criteria, we turned to a variant of L-Rhamnulose-1-phosphate aldolase (RhuA), a square-shaped C_4-symmetric protein bearing cysteine residues at its four corners (the 98 position) and pairs of glutamate residues on each side (at positions 57 and 66), $^{C98/E57/E66}$RhuA or ^{CEE}RhuA[11]. Under mildly reducing conditions, ^{CEE}RhuA (as well as the parent construct ^{C98}RhuA[10]) undergoes solution self-assembly into checkerboard-like µm-scale porous 2D crystals via intermolecular disulfide bonding at the corners. Once tethered together, the flexibility of the disulfide bonds permits them to act as hinges, enabling RhuA crystals to undergo coherent structural dynamics that induce uniform changes in pore size (ranging from ~6 to <1 nm without losing crystallinity[10]. Subsequent studies demonstrated that the introduction of glutamate residues at suitable locations on the protein surface endowed these crystals with a porosity switch toggled by specific ions. Close-packing of the protein subunits (the nonporous conformation) was disfavored at thermodynamic equilibrium due to electrostatic repulsion, but could be achieved upon the introduction of Ca^{2+} ions, which bind to and bridge neighboring carboxylate sidechains, neutralizing the charge[11]. Thus, the 2D morphology of ^{CEE}RhuA lattices, coupled with the ability to modulate their porosity through metal (un)binding, suggested that these crystals should fulfill all of the requirements of the optimal gatekeeper coating (**Figure 3.4**).

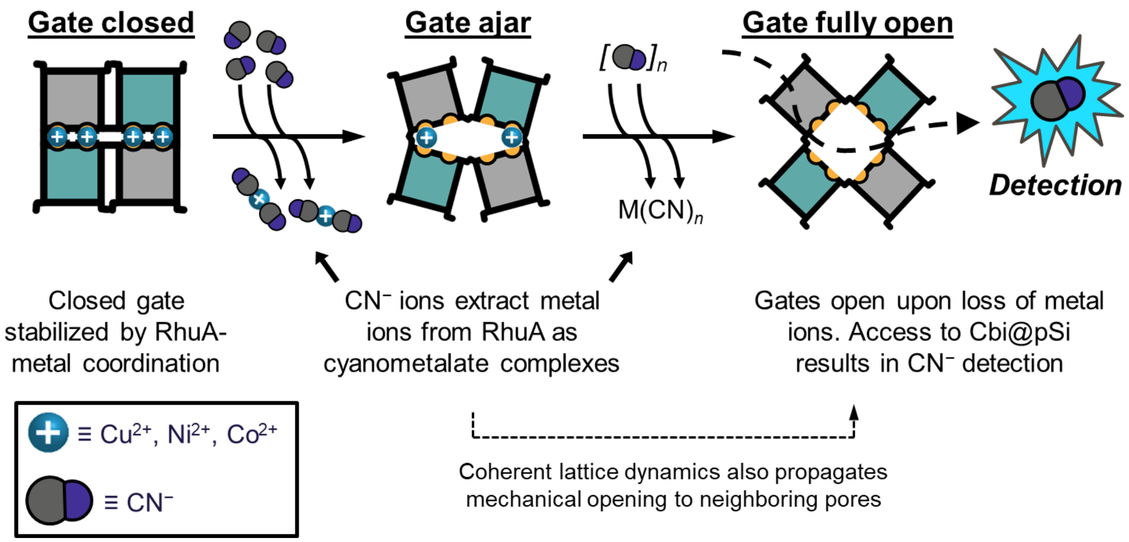


Figure 3.4 | HCN-selective gatekeeping mechanism of ^{CEE}RhuA crystals. Metal-bound ^{CEE}RhuA crystals adopt a spring-loaded nonporous "gate closed" conformation, which is held under tension by metal ions chelated by geometrically opposed Glu sidechains, which neutralizes like-charge repulsion and structurally connects neighboring proteins. In the presence of HCN, these ions will be stripped from the lattices through the formation of cyano complexes, releasing the spring to open the gate, allowing access to the underlying sensor.

3.3.2 Detection of HCN gas using uncoated, Cbi-loaded pSi photonic crystals

We first carried out proof-of-principle experiments to demonstrate HCN detection using the envisaged colorimetric sensing mechanism. pSi wafers were etched using a composite waveform to produce a 1D photonic nanostructure exhibiting two high-intensity stop bands at ~583 and ~700 nm (**Figure 3.2**). After rinsing and drying steps, 5.175 µg of monocyanocobinamide (mono-Cbi) was loaded into the surface pores (1.2 cm^2 area) via drop casting in 100 µl MeOH (50 µM mono-Cbi). At this point, the absorption spectrum of mono-Cbi (peak at ~504 nm) does not overlap with either stop band in the reflectance spectrum (**Figure 3.2**); this is the zero-exposure reference signal. Here we used monocyano-substituted Cbi because the majority of the absorption increase at 583 nm occurs during the transition from mono- to di-substituted states[49]. As a result, nearly all of the optical intensity change in response to HCN occurs upon individual (rather than double) binding events[49], reducing the detection threshold just under twofold.

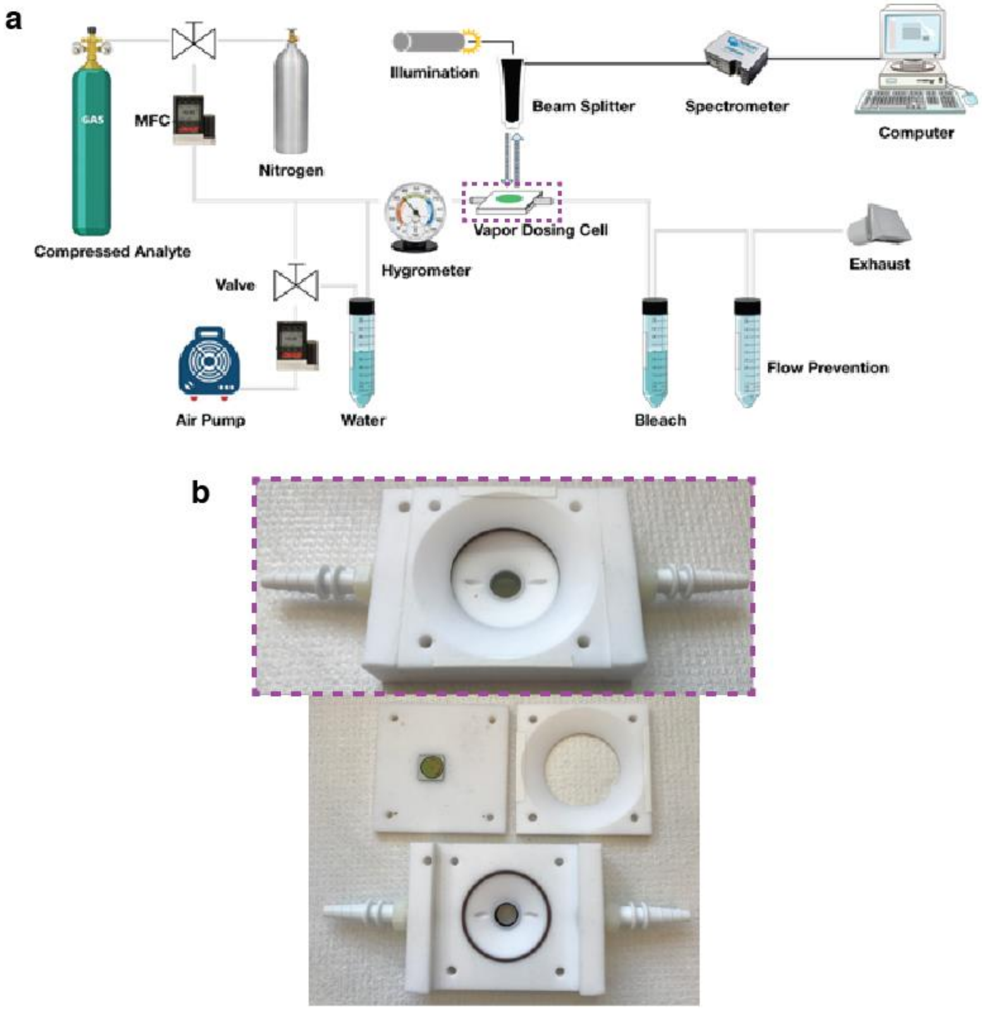


Figure 3.5 | Overview of the dosing system used to perform vapor-phase experiments with pSi wafers. a, Schematic of all dosing system components and their configuration. Analyte concentration, flow rate, and relative humidity are regulated via mass flow controllers, then pumped through the flow cell containing the wafer. **b,** The flow cell is composed of Teflon, with a sealed chamber for controlled vapor exposure, and a window through which optical measurements can be taken. The inlet and outlet ports are visible on the sides of the cell.

Cbi-loaded pSi chips were placed into a Teflon gas flow chamber fitted with a window (parallel to the chip) through which the reflectance spectrum could be measured[37] (**Figure 3.5**). To negate any systematic artifacts in the optical measurement that might arise from temperature fluctuations and nonspecific gas adsorption, we calculate the relative change at 583 nm (signal peak) normalized against the change at 700 nm (reference peak) (**Equation 3.1**):

$$\Delta I/I_0 = \frac{reflectance_{583\,nm}^{initial} - reflectance_{583\,nm}^{final}}{reflectance_{583\,nm}^{initial}} \Big/ \frac{reflectance_{700\,nm}^{initial} - reflectance_{700\,nm}^{final}}{reflectance_{700\,nm}^{initial}} \tag{3.1}$$

Upon exposure to 20 ppm HCN gas, we observed a monotonic decrease in the normalized reflectance intensity (I/I_0) by nearly 30% over 10 min (**Figure 3.6**). This suggested that the *in situ* formation of dicyano-Cbi could indeed be monitored in real-time via colorimetric change, leading to visible-light absorption at wavelengths competitive with the "signal" spectral peak. Subsequent experiments demonstrated that HCN was readily detected at concentrations as low as 5 ppm, which still resulted in a 5–10% in I/I_0 over 10 min (**Figure 3.6**).

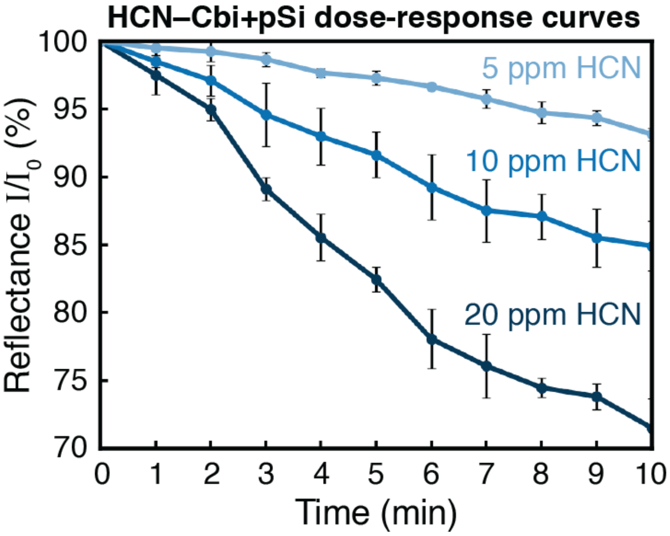


Figure 3.6 | HCN detection by Cbi-loaded pSi wafers. Time-resolved dose-response curves for Cbi-loaded pSi wafers exposed to different concentrations of HCN. Error bars are the s.d. for triplicate measurements.

3.3.3 Sensitivity of uncoated pSi photonic crystals to non-target analytes

The production of specified stop bands in the reflectance spectra of rugate pSi thin films arises from the constructive interference of reflected light which has interacted with the underlying photonic nanostructure. This renders pSi photonic crystals extremely sensitive to changes in local environment (*e.g.*, refractive index, film thickness), which has been exploited to detect analytes at $\leq 10^{-12}$ M concentrations via observable wavelength shifts in the interferometric reflectance pattern resulting from their presence[38,45]. The inherently general physics of this phenomenon means that nearly anything outside of the standard reference environment for that sensor can result in movement of its spectral features. This raises two challenges for the present system: 1) shifts in the stop bands are not only non-diagnostic for any particular analyte, 2) excessive shifting of the signal peak away from 583 nm will reduce overlap with the dicyano-Cbi absorption band, attenuating/preventing the decrease in signal band intensity required for HCN sensing.

Indeed, open-air etched pSi wafers exhibited dramatic shifts in their reflectance spectra upon exposure to several arbitrary species, including liquid EtOH (**Figure 3.7a**), vapor-phase water, hexanes, ammonia, and the Supelco EPA TO-14 (43 component) chemical standard. Somewhat surprisingly, Cbi-loaded pSi sensors were not immune to this effect either (**Figure 3.7b**), despite the porous microstructure being (at least partially) filled with dye. This indicated that dye loading alone is insufficient to protect the sensor from changes to its spectral properties which arise from the nonspecific adsorption of arbitrary species. Consequently, we sought to functionalize the sensor surface with a selectively permeable protective coating (a “gatekeeper”) which would occlude the underlying pSi structure until the presence of HCN rendered it porous, opening the gate to enable detection.

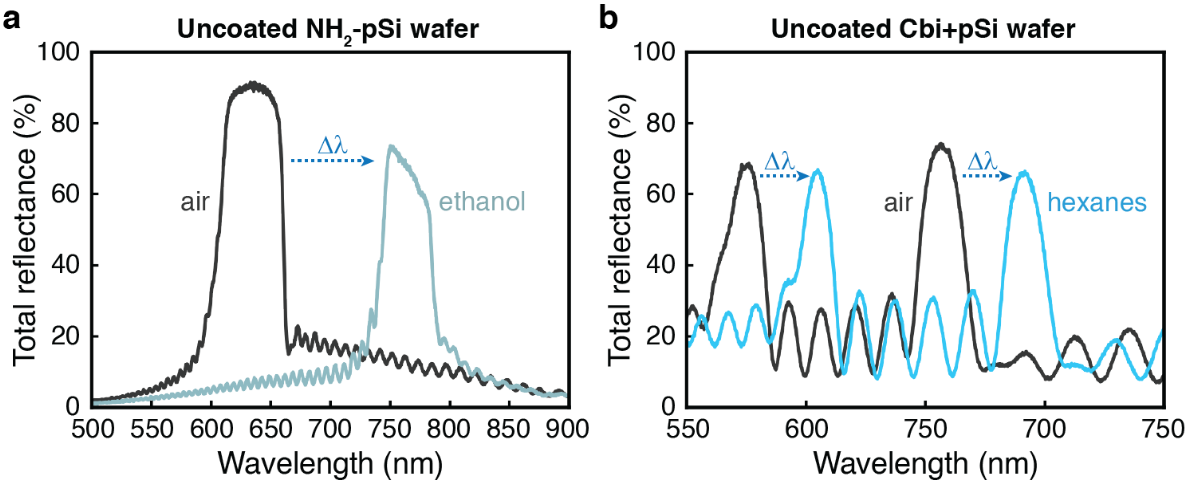


Figure 3.7 | Representative optical responses in the reflectance spectra of photonic pSi wafers upon exposure to generic substances. a, Infiltration of ethanol into the surface pores of an uncoated, unfilled (open-air) pSi wafer changes the refractive index of the optical film (here increasing from RI ≈ 1.00 (air) to 1.36), leading to a >100 nm redshift of the spectral peak position (indicated as Δλ). **b,** An uncoated, Cbi-loaded wafer exhibits similar behavior, producing a ~40 nm redshift in the optical spectrum after a 10 min exposure to vapor-phase hexanes. Despite its pores being partially occupied by dye molecules, this sensor remains susceptible to optical artifacts arising from ambient species without a protective coating.

3.3.4 Selective switching of gatekeeper porosity by CN$^-$ in solution

As outlined in the initial design, we explored the possibility of using the protein variant $^{\mathrm{CEE}}$RhuA as our HCN-selective gatekeeper. This protein self-assembles into 2D frameworks capable of undergoing cooperative changes between porous and nonporous conformations upon binding/release of metal ions. Previous investigation of $^{\mathrm{CEE}}$RhuA crystals found that this conformational switching behavior occurred when incubated with Ca^{2+} ions, but not Mg^{2+} or monovalent salts (indicating that metal-ion binding, not electrostatic screening effects, was responsible for lattice closure[11]). This is in accordance with calculated affinities of carboxylate groups for divalent ions, which place Ca^{2+} at the top of the list[50], while the inability of Mg^{2+} (also commonly found coordinated by hard oxygen ligands) to substitute for Ca^{2+} suggested inherently weak coordination by the simple dicarboxylate motif[11].

As the nonporous state of ^{CEE}RhuA crystals is only accessible when bound to Ca^{2+}, we reasoned that displacement of the bridging ion by HCN (or CN^{-}) would provide the necessary gating mechanism (**Figure 3.4**), as ^{CEE}RhuA crystals spontaneously open when Ca^{2+} is removed by dialysis or chelation with EDTA. However, $Ca(CN)_2$ is highly susceptible to hydrolysis, producing HCN gas even in moist air[51], suggesting that chemical driving forces would render HCN sequestration thermodynamically impossible. In contrast, Co^{2+}—which exhibits the 2nd highest affinity for dicarboxylate coordinating ligands[50]—readily forms water-stable inorganic salts with CN^{-}, namely $Co(CN)_2$ trihydrate and insoluble $K_4Co(CN)_6$ (in the presence of excess KCN)[51]. Animated by this finding, we first set out to establish whether Co^{2+} could substitute for Ca^{2+} to induce lattice closure in solution.

Following the methodology used to screen other ionic species, clarified suspensions (with unincorporated protein monomers removed) of ^{CEE}RhuA crystals were mixed in 1:1 (v/v) ratios with Tris-buffered $CoCl_2$ solutions at twice the desired final Co^{2+} concentration, and were allowed to stand for three days at 4 °C to ensure sufficient time for binding to occur. Direct visualization of the crystals by negative-stain transmission electron microscopy (ns-TEM) revealed that 25 mM $CoCl_2$ was sufficient to induce closure of the ^{CEE}RhuA crystal pores (**Figure 3.8**). As seen previously with Ca^{2+}, vigorous mixing of the solution or dialysis to remove free ions both resulted in open-state crystals, suggesting that Co^{2+} indeed served as a surrogate bridging ion.

Next, we investigated whether Co^{2+} could be displaced from closed lattices through competitive binding by CN^{-}. When Co^{2+}-bound ^{CEE}RhuA crystals were gently mixed with Tris-buffered KCN solutions to a Co:KCN stoichiometric ratio of 1:2, the ^{CEE}RhuA crystals remained nonporous after 3 days. However, when KCN was added to a final 1:6 stoichiometry, open-state ^{CEE}RhuA crystals were observed after 1 day (**Figure 3.8**). Subsequent time-course experiments established that pore opening occurred within 2–5 min after KCN addition (**Figure 3.9**). This provided compelling evidence that not only could the nonporous conformation be stabilized by Co^{2+} ions, but also that Co^{2+}-binding could be rapidly reversed upon exposure to KCN, restoring crystal permeability on the timescale of HCN detection by Cbi absorption.

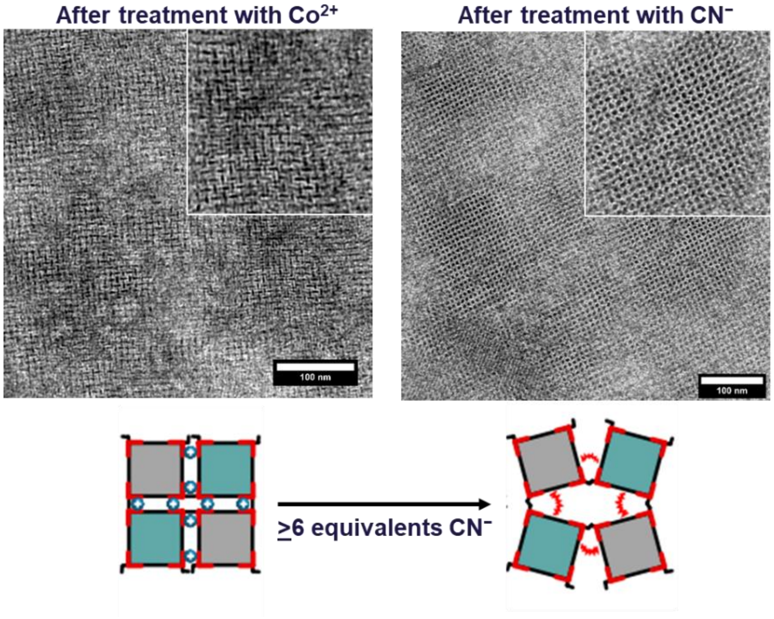


Figure 3.8 | KCN modulates the conformation of Co^{2+}-bound ^{CEE}RhuA lattices. TEM images of ^{CEE}RhuA lattices after binding to 25 mM $CoCl_2$ for 3 days (**left**), and after incubation with 6 equivalents of KCN for 24 hours (**right**).

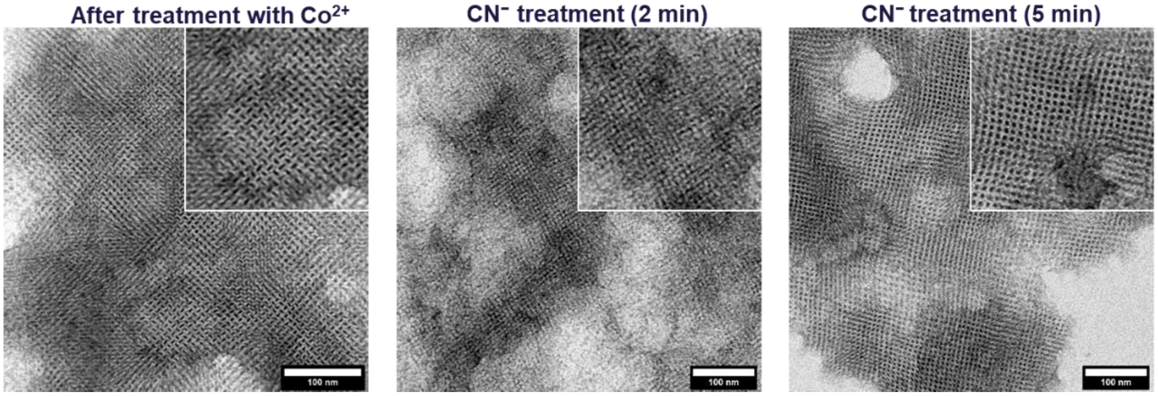


Figure 3.9 | Temporal response of ^{CEE}RhuA+Co^{2+} lattices to KCN. ns-TEM images from a time-course experiment following the evolution of ^{CEE}RhuA+Co^{2+} crystal porosity. The porous structure of the lattices is visible after only 5 min exposure to 6 equivalents of KCN.

3.3.5 Clutter rejection by a biological gatekeeper-coated pSi photonic crystal sensor

In order to structurally integrate the gatekeeper onto the sensor surface while still allowing for conformational motions of the protein subunits, we employed noncovalent electrostatic interactions as our attachment chemistry. It has previously been determined that the faces of 2D RhuA crystals are negatively charged, owing to the up-down orientation of the protein's C-terminal face, leading to preferential binding of crystals onto positively charged surfaces[52]. To render the substrate surface charge positive, we prepared pSi wafers functionalized with amine-containing heterocyclic-silanes, which undergo "ring-opening click" reactions with pSi surface hydroxyl groups[40], yielding "NH_2-pSi". By utilizing RhuA crystal suspensions buffered at pH 6, which is below the surface amine pKa (~7.2) and above the ^{CEE}RhuA isoelectric point (~5.5), opposite-charge interactions were maximized. Crystal suspensions (~5 μM protein concentration) were drop-cast onto both NH_2-pSi and untreated ("bare") pSi wafers and allowed to dry overnight, followed by several rinses with diH_2O before being wicked dry. Scanning electron microscopy (SEM) imaging of the wafer surfaces clearly revealed extensive binding of well-resolved gatekeeper crystals distributed across the surface of NH_2-pSi wafers; this is in stark contrast to untreated pSi chips, which were barren (**Figure 3.10**).

Seeking now to establish the capacity of ^{CEE}RhuA lattices to serve as selective gatekeepers, we first deposited Ca^{2+}-bound ^{CEE}RhuA crystals onto NH_2-pSi wafers without embedded Cbi (**Figure 3.11a**). Omission of the dye allowed direct evaluation of the ability of the gatekeeper coating to reject vapor-phase species (without concern for any unexpected effects that could arise from the dye) by monitoring the stability of the reflectance spectrum. No shifting of the spectral peaks was observed over serial 5 min exposures to humid air (20% RH) and hexane vapor (**Figure 3.11b**). However, treatment of the surface with a buffered EDTA (large excess) solution produced

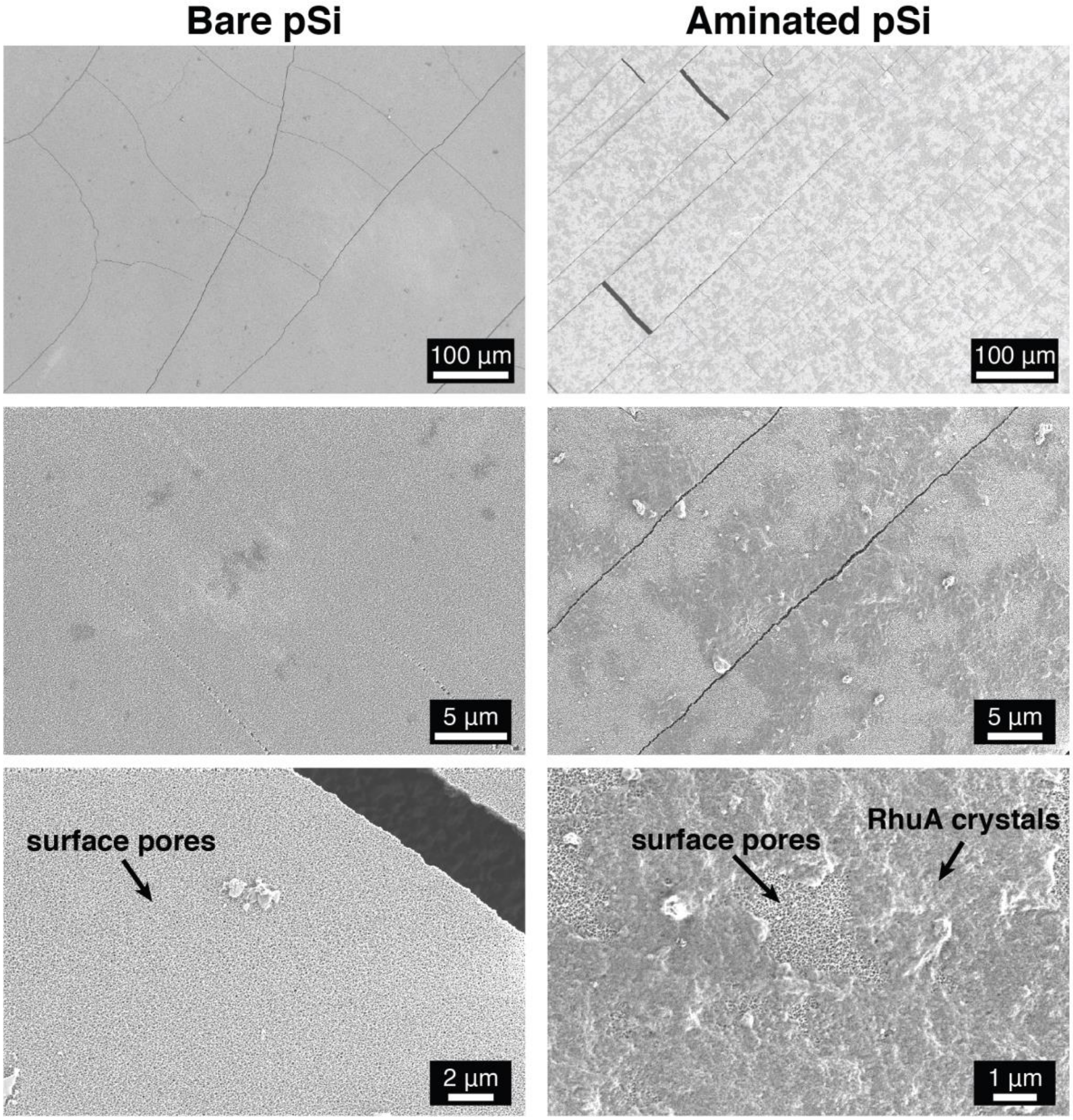


Figure 3.10 | Coating of pSi wafers with RhuA lattices. SEM images of pSi wafers after overnight incubation in pH 6 buffer, followed by a wash step with diH_2O. Unfunctionalized wafers (**left**) retained almost no crystals, while aminated surfaces (**right**) bound a large quantity, consistent with the negative charge on both faces of RhuA crystals.

a slight (*ca.* 3 nm) optical redshift (likely due to infiltration of water into the pores), which became exaggerated upon exposure to ethanol (which has a greater RI), suggesting that the coating indeed switched from a nonporous state to a porous one in response to Ca^{2+} chelation (**Figure 3.11c**). We then characterized the optical spectrum and porosity of a replicate sensor at each step of production (**Figure 3.11d**). Amination of the pSi surface (including within the pores) unsurprisingly leads to a slight reduction in total porosity, however the sensor is rendered totally nonporous after deposition of the gatekeeper coating. The concomitant blueshift of the stop band is likely due to mild oxidation of the pSi surface due to the aqueous buffer. Consistent with the previous results, EDTA treatment produces a −15 nm spectral shift, which we attribute to the release of water trapped within the porous structure during crystal deposition (reducing the RI from 1.33 to 1.00), and restores just under 70% of the pre-coating porosity. Now vacated of solvent, ethanol permeation into the pores produced a pronounced redshift of +31 nm, consistent with a re-raising of the sensor RI back to 1.36 (**Figure 3.11d**).

Encouraged by the apparent ability of the gatekeeper to undergo metal-mediated structural switching on the pSi surface, we carried out analogous experiments with Co^{2+}-bound lattices, which similarly demonstrated continuous impermeability of the coating to sequential independent 25 min exposures of the same non-target agents (**Figure 3.12a**). With this capability now established, these lattices were deposited onto a Cbi-loaded pSi photonic crystal substrate, rendering it similarly nonporous. In stark contrast to all previous trials, a brief 5 min dosing of 20 ppm HCN vapor induced a *ca.* −16 nm spectral shift (again attributed to the release of water vapor), with most of the shift occurring within 2 min **Figure 3.12b**). Thus, these preliminary tests strongly indicated that the Co^{2+}-gatekeeper is not only impermeable to small molecules, but can indeed undergo the HCN-selective switching observed in solution while attached to the pSi sensor surface.

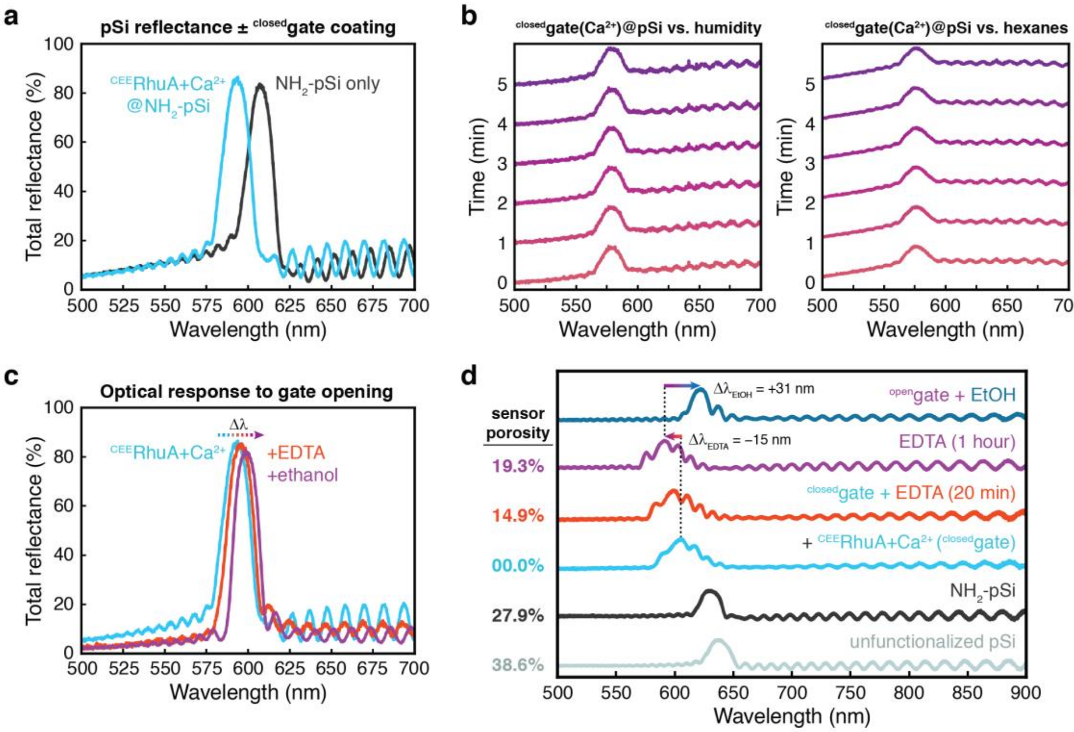


Figure 3.11 | Ca^{2+}-^{CEE}RhuA lattices are effective gatekeepers for NH_2-pSi sensors. a, Coating of the sensor leads to a slight blueshift in the reflectance peak, but does not attenuate optical intensity. **b,** Ca^{2+}-bound ^{CEE}RhuA crystals fully reject both humidity and hexanes as representative non-target agents. **c,** Upon EDTA treatment, a slight redshift is observed, indicating the infiltration of ambient chemical species. This shift is exacerbated upon exposure to EtOH (owing to its higher refractive index) demonstrating greater gate permeability. **d,** Optical characterization at each step of sensor construction, with porosities indicated at left.

We next began testing the final version of the fully integrated HCN sensor: Cbi-loaded NH_2-pSi rugate filters coated with a layer of nonporous Co^{2+}-bound ^{CEE}RhuA crystals, yielding Co^{2+}-^{CEE}RhuA@NH_2-pSi+Cbi (^{closed}gate@pSi+Cbi). In line with previous results, serial 10 min exposures of ^{closed}gate@pSi+Cbi wafers to 50% RH water vapor, nitrogen, and hexanes (as representative challenge agents) all failed to induce any change in the I/I_0 signal reflectance intensity (**Figure 3.13**), indicating full rejection of the non-target species by the gatekeeper. Subsequent introduction of 10 ppm HCN resulted in a steady decrease of the I/I_0 signal over the full duration of exposure, with the I/I_0 metric reaching ~90% by 3 min and 85% at 10 min (**Figure 3.13**). This $\Delta I/I_0$ magnitude (*ca.* −15%) is somewhat lower than non-coated pSi+Cbi sensors ($\Delta I/I_0$

< −20%), consistent with partial occlusion of the surface by the protein lattices, which are only 50% porous even in the maximally open configuration. Regardless, this does not significantly impair HCN sensing ability, as the detection threshold is $I/I_0 \approx 98\%$. Importantly, when ^{CEE}RhuA lattices were deposited in the absence of Co^{2+} (**Figure 3.14**), gaseous hexanes readily permeated the surface coating (as evidenced by shifting of the reflectance spectra), consistent with an ^{open}gate@pSi+Cbi composition. Co^{2+}-bound ^{closed}gate@pSi+Cbi structures (prepared in parallel) fully rejected these short-chain alkanes for 10 min under identical conditions (**Figures 3.14**).

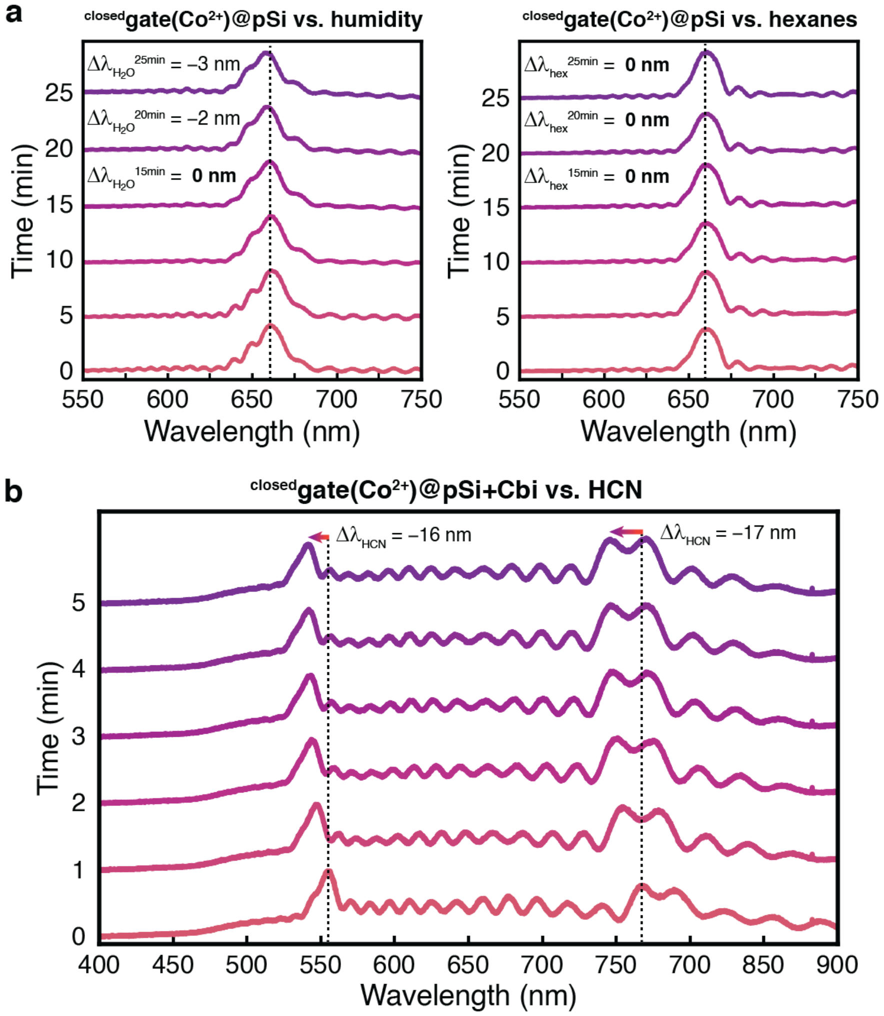


Figure 3.12 | Selective clutter rejection by Co^{2+}-^{CEE}RhuA-coated NH_2-pSi wafers. a, Co^{2+}-^{CEE}RhuA crystals coatings (on no-Cbi wafers) fully and continuously reject water and hexane vapor streams for at least 15 min. **b,** When the same crystals are deposited onto Cbi-loaded wafers, HCN exposure leads to an immediate 16–17 nm peak shift within 2 min.

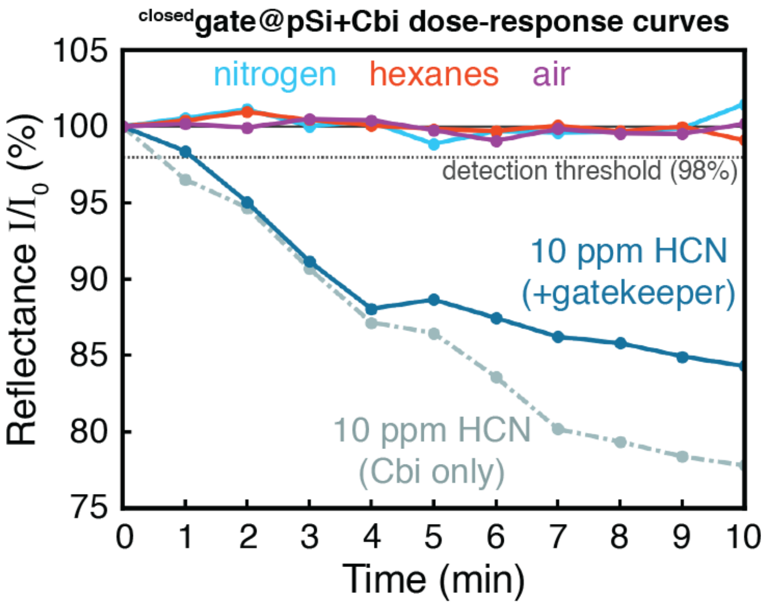


Figure 3.13 | Selectivity of ^{CEE}RhuA@pSi+Cbi sensors for HCN. Temporal dose response curves for several challenge agents exposed to ^{closed}gate@pSi+Cbi sensors. All non-HCN compounds fail to break the detection threshold, while HCN produces a clear change in optical intensity, both with and without the gatekeeper.

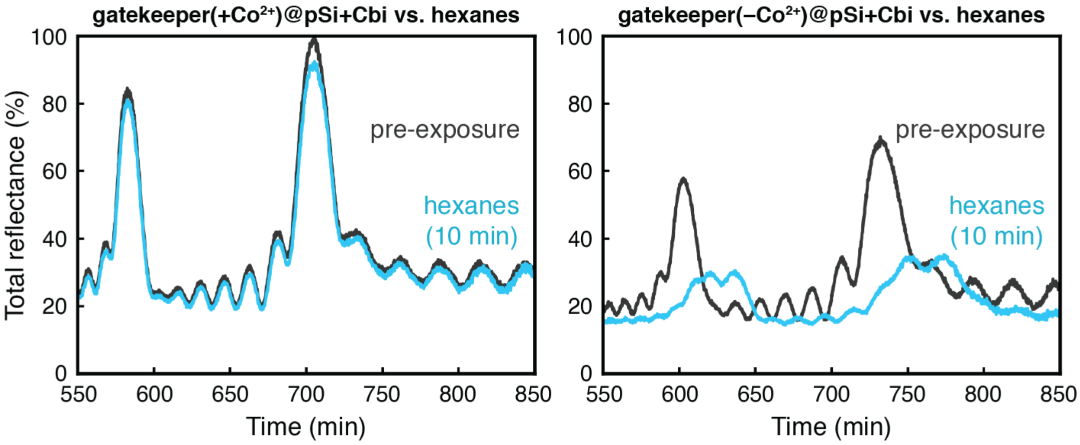


Figure 3.14 | Closed-gate conformations are required for sensor protection. Reflectance spectra for psi+Cbi wafers coated with gatekeeper crystals (± Co^{2+} ions) reveal distinct optical signals in response to hexane vapor. Co^{2+}-bound crystals fully protect the sensor (as indicated by the lack of spectral shift, **left**), while crystals deposited in the absence of the Co^{2+} fail to prevent infiltration of these alkanes into the pSi substrate (resulting in a 20–30 nm redshift in the optical spectrum, **right**). This indicates that the latter coating must be in the "gate open" conformation prior to attachment (the thermodynamically preferred state in the absence of Ca^{2+} or Co^{2+} ions), rendering it unable to occlude the underlying sensor substrate.

3.4 Conclusions

The reliable, sensitive, and specific detection (or sequestration) of predefined chemical species is a demanding task for synthetic indicators, and represents a principal objective of active research into the development of novel sensing materials[53-55]. Yet, these characteristics are a universal feature of biological systems (predominantly among proteins), where chemical precision is an absolute necessity to maintain life. Here, we have reported a fully integrated chemical sensor that addresses all three criteria by simultaneously exploiting the high sensitivity and tunable optical properties of photonic pSi crystals coupled with the chemical specificity and conformational flexibility afforded by biological macromolecules. The resulting material serves as a passive indicator for HCN gas (a significant common industrial toxin) that can be assessed remotely from a safe distance through optical interferometric methods. The rejection of non-HCN chemical species present in the ambient environment was achieved through the use of a crystalline biological framework (the "gatekeeper"), a nonporous coating which only becomes permeable upon loss of specific chemical interactions (metal coordination) that are directly severed by HCN molecules. As a result, this sensor possesses a long shelf-life (>3 weeks) and is resistant to erroneous readouts, demonstrating its promise as an unattended sentinel for the detection of noxious gas. Moreover, the ability to tailor both the indicator dye and protein "gatekeeper" coating to target alternative species of interest should make this an attractive general design strategy for the construction of next-generation chemical sensors.

3.5 Methods

3.5.1 Electrochemical synthesis of porous silicon (pSi) photonic crystals

Dual-peak pSi photonic crystals were prepared by electrochemical etching of the (100) plane of highly doped, single-crystal p-type Si wafers (Boron doped, 0.8–1.0 mΩ-cm resistivity; Siltronix Corp.) using a 3:1 (v/v) 48% aqueous hydrofluoric acid:ethanol solution[37]. Wafers were mounted inside a custom Teflon etching cell with a 1.2 cm^2 circular surface area, contacted on the backside with aluminum foil, and anodized using a platinum coil counter-electrode. The wafer surface was first polished in the 3:1 HF:EtOH solution using a sacrificial etch of 400 mA/cm2 for 50 s, rinsed with ethanol, and the resultant porous film was removed by dissolution with a 2 M KOH solution. The polished wafer was rinsed with water followed by ethanol prior to the final etch. A composite of two sinusoidal waveforms (periods of 2.0 and 2.8s, corresponding to the signal and reference reflectance peaks) was applied for 300 s using a variable current density between 4.167 and 60 mA/cm^2. Current amplitude waveforms were generated using the LabView software (National Instruments, Inc.) and were interfaced with a Keithley 2651A SourceMeter power supply.

3.5.2 Surface functionalization of pSi photonic crystals

Amine functionalization of pSi photonic crystals was performed by sequential triplicate rinses of the wafer with 100% EtOH, 50:50 EtOH:DCM, and 100% DCM, before submerging it in a glass vial containing 1.5 mL of cyclic silane (either 2,2-dimethoxy-1,6-diaza-2-silacyclooctane [DMDASCP] or N-n-butyl-aza-2,2-dimethoxy-silacyclopentane [BADMSCP])[40], with gentle mixing for 3 hours. Excess silane was removed, and the wafer was again subjected to serial triplicate rinses in the reverse order (100% DCM, 50:50 EtOH:DCM, 100% EtOH) and allowed to dry.

3.5.3 Preparation of Cobinamide (Cbi) dye

Pure aquohydroxocobinamide (OH(H_2O)Cbi) was synthesized by base hydroloysis from hydroxocobalamin (Sigma-Aldrich) as described previously[49]. A stock solution of monocyanocobinamide (CN(H_2O)Cbi) was then prepared by reacting equimolar quantities of aquohydroxocobinamide (dissolved in aqueous 1 M NaOH) with KCN (Fisher Scientific) for 6h under mild agitation at room temperature, which produced a dark red solution. Successful synthesis of monocyanocobinamide was verified by identification of the characteristic absorbance peak spanning 475–520 nm using a benchtop UV-Vis spectrometer (Molecular Devices Spectramax 340Pc384 Microplate Spectrophotometer), a distinct absorbance peak from 475-520 nm.

3.5.4 Preparation of Cbi-loaded pSi photonic crystals

Mono-Cbi stock solutions were diluted to 50 μM in ethanol, and 100 μL aliquots were drop-cast onto the etched pSi wafer surface and the solvent was allowed to dry to completeness at room temperature (approximately 5 hours).

3.5.5 Testing of Cbi-loaded pSi photonic crystals (with or without gatekeeper)

Spectroscopic measurements of Cbi-loaded pSi photonic crystals were perfomed by placing the wafer into a Teflon cell fitted with an optical window (Figure 3.5). Analyte vapors were generated at room temperature either directly from certified gas cylinders (Gasco, Inc.) or by bubbling inert N_2 through the analyte solution to produce saturated vapor. Mass flow controllers were used to control the ratio of N_2 carrier gas and analyte vapors to produce specified values of %RH (monitored by probe hygrometer) and HCN concentration at a constant flow rate of 300 sccm. All gases were mixed prior to entering the sample cell, and concentrations were calculated from their partial pressures. The white light (tungsten-halogen light source) reflectance spectrum from the pSi photonic crystal sample was collected using an Ocean Optics S4000 CCD spectrometer connected to a bifurcated fiber optic cable. Both the signal and reference peaks were measured continuously for the entire duration of the vapor exposure period, enabling collection of time-resolved (Signal/Reference) intensity changes.

3.5.6 Expression and solution self-assembly of the gatekeeper protein

^{CEE}RhuA protein was purified by published methods[11]. Briefly, ^{CEE}RhuA was overexpressed in *E. coli* cells which were grown to high density and lysed by sonication. The resulting solution was clarified by centrifugation, and subjected to multiple chromatographic steps, ultimately yielding approximately 10 mg pure protein per liter of cell culture. The purified protein was concentrated to 150 μM (~18 mg/mL) in the presence of 10 mM β-mercaptoethanol and 1 mM $ZnCl_2$, conditions suitable for crystallization. 0.5 mL aliquots were flash frozen in liquid nitrogen and stored at −80 °C. Crystallization was induced by thawing the aliquots and leaving them on a shaking platform in the cold room at 4 °C for several days (up to several weeks).

3.5.7 Preparation and deposition of ^{CEE}RhuA crystals onto pSi photonic crystals

^{CEE}RhuA crystal suspensions were clarified (unincorporated protein monomers were removed) by repeated cycles of gentle centrifugation (~3000 rpm) in a benchtop centrifuge for 30-60s, followed by replacement of the supernatant with fresh buffer solution containing 20 mM 2-(N-morpholino)ethanesulfonic acid (MES) pH 6 and 50 mM $CoCl_2$. 80-90% of the solution was replaced at each wash step, leading to complete exchange of the ^{CEE}RhuA crystals into the new buffer solution. After washing, the crystal suspensions typically contained protein concentrations of ~100 μM. Experimental screening determined that a protein concentration of 5 μM provided the functionally sufficient surface coverage of the sensor while preserving its optical reflectance spectrum. Crystal solutions were gently resuspended prior to deposition, using 50 μL to drop-cast onto 1.2 cm^2 square area chips, and a proportional 400 μL of crystal suspension for 9.6 cm^2 chips. Amine functionalization of the pSi surface (described above) was found to be an essential prerequisite for effective crystal binding to the substrate. The pKas of the amines in these chemical groups are estimated to be ~7.2, while the isoelectric point of the ^{CEE}RhuA protein is ~5.5, so drop-casting of the protein crystals was carried out using in the MES (pH 6) buffer introduced during the previous wash step, serving to promote binding by ensuring that the surface was positively charged while the protein remained negatively charged.

3.5.8 Evaluation of conformational switching of gatekeeper lattices in solution

To induce closure of the gate structure, crystal suspensions in 20 mM MES buffer were mixed 1:1 with 50 mM $CoCl_2$/20 mM MES stock solutions and allowed to stand at 4 °C for at least 3 days. Evaluation of CN^--mediated switching was performed by gently mixing this suspension with a 50 mM KCN MES-buffered solution to a 1:3 volume ratio ($[Co^{2+}]_{final}$ = 6.25 mm, $[KCN]_{final}$ = 37.5 mM), yielding a final 1:6 stoichiometric ratio.

All solution-state characterization of the lattice conformations was performed using ns-TEM. Samples were prepared by deposition of 3.5 μL volumes of crystal suspensions onto glow-discharged formvar/carbon-coated copper grids (Ted Pella, Inc.) and were allowed to bind for 5 min. The grids were washed with 50 μL filtered MilliQ water, wicked dry, and stained with 3.5 μL of 2% uranyl acetate solution. Samples were imaged using an FEI Sphera (LaB_6 filament) transmission electron microscope operating at a high tension of 200 keV and captured with a Gatan 2K CCD camera.

3.5.9 SEM imaging of pSi photonic crystals

pSi photonic crystal wafers were attached onto 6 mm pin SEM stubs (12.7 mm diameter; Ted Pella, Inc.) using double-sided carbon tape and gentle applied pressure. Samples were mounted into the standard carousel of a Zeiss Sigma 500 scanning electron microscope operating at accelerating voltages between 1.0–1.5 kV, and images were captured as secondary electron events (SE2 detector) using a 30 μm aperture. Samples were not sputter-coated in order to preserve the integrity of the protein framework, which proved unproblematic for low-voltage imaging.

3.6 Acknowledgments

We thank R. Subramanian and L. Zhang for constructive discussions. This research is based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via contract number 2018-18071700005. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein. The authors acknowledge the facilities, along with the scientific and technical assistance of the staff of the cryo-EM facility at UC San Diego.

Chapter 3, in part, is currently being prepared for submission for publication: Alberstein, R.G.*, Vijayakumar, S.*, Lu, Y.-S., Chaix, A., Wahl, C.E., Ha, J.S., Hunka, D.E., Boss, G.R., Tezcan, F.A., Sailor, M.J. “A photonic porous silicon sensor with integrated biological gatekeeper for remote detection of HCN vapor”. The dissertation author was the co-primary investigator and author of this manuscript.

3.7 References

1 Wolfbeis, O. S. Fiber-Optic Chemical Sensors and Biosensors. *Anal. Chem.* **78**, 3859-3874, (2006).

2 Tschmelak, J., Kumpf, M., Proll, G. & Gauglitz, G. Biosensor for Seven Sulphonamides in Drinking, Ground, and Surface Water with Difficult Matrices. *Anal. Lett.* **37**, 1701-1718, (2004).

3 Marinaro, G., Etiope, G., Gasparoni, F., Calore, D., Cenedese, S., Furlan, F., Masson, M., Favali, F. & Blandin, J. GMM—a gas monitoring module for long-term detection of methane leakage from the seafloor. *Env. Geol.* **46**, 1053-1058, (2004).

4 Favas, G. End of Service Life Indicator (ESLI) for Respirator Cartridges. Part I: Literature Review. Australian Government Department of Defense, DSTO-TN-0657: 2005.

5 Mayer, K. M. & Hafner, J. H. Localized Surface Plasmon Resonance Sensors. *Chem. Rev.* **111**, 3828-3857, (2011).

6 Sailor, M. J. & Wu, E. C. Photoluminescence-Based Sensing With Porous Silicon Films, Microparticles, and Nanoparticles. *Adv. Funct. Mater.* **19**, 3195-3208, (2009).

7 Lieberzeit, P. A. & Dickert, F. L. Rapid bioanalysis with chemical sensors: novel strategies for devices and artificial recognition membranes. *Anal. Bioanal. Chem.* **391**, 1629-1639, (2008).

8 Gauglitz, G. Direct optical detection in bioanalysis: an update. *Anal. Bioanal. Chem.* **398**, 2363-2372, (2010).

9 Zhang, L., Bailey, J. B., Subramanian, R. H., Groisman, A. & Tezcan, F. A. Hyperexpandable, self-healing macromolecular crystals with integrated polymer networks. *Nature* **557**, 86-91, (2018).

10 Suzuki, Y., Cardone, G., Restrepo, D., Zavattieri, P. D., Baker, T. S. & Tezcan, F. A. Self-assembly of coherently dynamic, auxetic, two-dimensional protein crystals. *Nature* **533**, 369–373, (2016).

11 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).

12 Murray, L. J., Dinca, M. & Long, J. R. Hydrogen storage in metal-organic frameworks. *Chem. Soc. Rev.* **38**, 1294-1314, (2009).

13 Rosi, N. L., Eckert, J., Eddaoudi, M., Vodak, D. T., Kim, J., O'Keeffe, M. & Yaghi, O. M. Hydrogen storage in microporous metal-organic frameworks. *Science* **300**, 1127-1129, (2003).

14 Mason, J. A., Oktawiec, J., Taylor, M. K., Hudson, M. R., Rodriguez, J., Bachman, J. E., Gonzalez, M. I., Cervellino, A., Guagliardi, A., Brown, C. M., Llewellyn, P. L., Masciocchi, N. & Long, J. R. Methane storage in flexible metal–organic frameworks with intrinsic thermal management. *Nature* **527**, 357-361, (2015).

15 Tanaka, T., Fillmore, D., Sun, S.-T., Nishio, I., Swislow, G. & Shah, A. Phase Transitions in Ionic Gels. *Phys. Rev. Lett.* **45**, 1636-1639, (1980).

16 Phadke, A., Zhang, C., Arman, B., Hsu, C.-C., Mashelkar, R. A., Lele, A. K., Tauber, M. J., Arya, G. & Varghese, S. Rapid self-healing hydrogels. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 4383-4388, (2012).

17 Qiu, Y. & Park, K. Environment-sensitive hydrogels for drug delivery. *Adv. Drug Deliv. Rev* **53**, 321-339, (2001).

18 Ruoslahti, E., Bhatia, S. N. & Sailor, M. J. Targeting of drugs and nanoparticles to tumors. *J. Cell Biol.* **188**, 759-768, (2010).

19 Torchilin, V. P. Multifunctional, stimuli-sensitive nanoparticulate systems for drug delivery. *Nat. Rev. Drug Discov.* **13**, 813-827, (2014).

20 Hamidi, M., Azadi, A. & Rafiei, P. Hydrogel nanoparticles in drug delivery. *Adv. Drug Deliv. Rev* **60**, 1638-1649, (2008).

21 Latorre, R. & Miller, C. Conduction and selectivity in potassium channels. *J. Membrain Biol.* **71**, 11-30, (1983).

22 Eisenman, G. & Horn, R. Ionic selectivity revisited: The role of kinetic and equilibrium processes in ion permeation through channels. *J. Membrain Biol.* **76**, 197-225, (1983).

23 Noskov, S. Y., Bernèche, S. & Roux, B. Control of ion selectivity in potassium channels by electrostatic and dynamic properties of carbonyl ligands. *Nature* **431**, 830-834, (2004).

24 Thompson, A. N., Kim, I., Panosian, T. D., Iverson, T. M., Allen, T. W. & Nimigean, C. M. Mechanism of potassium-channel selectivity revealed by Na+ and Li+ binding sites within the KcsA pore. *Nat. Struct. Mol. Biol.* **16**, 1317-1324, (2009).

25 Tsien, R. W., Hess, P., McCleskey, E. W. & Rosenberg, R. L. Calcium Channels: Mechanisms of Selectivity, Permeation, and Block. *Ann. Rev. Biophys. Chem.* **16**, 265-290, (1987).

26 Doyle, D. A., Cabral, J. M., Pfuetzner, R. A., Kuo, A., Gulbis, J. M., Cohen, S. L., Chait, B. T. & MacKinnon, R. The Structure of the Potassium Channel: Molecular Basis of K⁺ Conduction and Selectivity. *Science* **280**, 69-77, (1998).

27 Zhou, Y., Morais-Cabral, J. H., Kaufman, A. & MacKinnon, R. Chemistry of ion coordination and hydration revealed by a K+ channel–Fab complex at 2.0 Å resolution. *Nature* **414**, 43-48, (2001).

28 Thomas, M., Jayatilaka, D. & Corry, B. The Predominant Role of Coordination Number in Potassium Channel Selectivity. *Biophys. J.* **93**, 2635-2643, (2007).

29 Peng, X., Jin, J., Nakamura, Y., Ohno, T. & Ichinose, I. Ultrafast permeation of water through protein-based membranes. *Nat. Nanotechol.* **4**, 353-357, (2009).

30 Zhang, S., Zhang, J., Fang, W., Zhang, Y., Wang, Q. & Jin, J. Ultralarge Single-Layer Porous Protein Nanosheet for Precise Nanosize Separation. *Nano Lett.* **18**, 6563-6569, (2018).

31 Tu, Y.-M., Song, W., Ren, T., Shen, Y.-x., Chowdhury, R., Rajapaksha, P., Culp, T. E., Samineni, L., Lang, C., Thokkadam, A., Carson, D., Dai, Y., Mukthar, A., Zhang, M., Parshin, A., Sloand, J. N., Medina, S. H., Grzelakowski, M., Bhattacharya, D., Phillip, W. A., Gomez, E. D., Hickey, R. J., Wei, Y. & Kumar, M. Rapid fabrication of precise high-throughput filters from membrane protein nanosheets. *Nat. Mater.* **19**, 347-354, (2020).

32 Li, J. R., Kuppler, R. J. & Zhou, H. C. Selective gas adsorption and separation in metal-organic frameworks. *Chem. Soc. Rev.* **38**, 1477-1504, (2009).

33 Ruminski, A. M., Moore, M. M. & Sailor, M. J. Humidity-Compensating Sensor for Volatile Organic Compounds Using Stacked Porous Silicon Photonic Crystals. *Adv. Funct. Mater.* **18**, 3418-3426, (2008).

34 Ruminski, A. M., King, B. H., Salonen, J., Snyder, J. L. & Sailor, M. J. Porous Silicon-Based Optical Microsensors for Volatile Organic Analytes: Effect of Surface Chemistry on Stability and Specificity. *Adv. Funct. Mater.* **20**, 2874-2883, (2010).

35 Ruminski, A. M., Barillaro, G., Chaffin, C. & Sailor, M. J. Internally Referenced Remote Sensors for HF and Cl2 Using Reactive Porous Silicon Photonic Crystals. *Adv. Funct. Mater.* **21**, 1511-1525, (2011).

36 Sohn, H., Létant, S., Sailor, M. J. & Trogler, W. C. Detection of Fluorophosphonate Chemical Warfare Agents by Catalytic Hydrolysis with a Porous Silicon Interferometer. *J. Am. Chem. Soc.* **122**, 5399-5400, (2000).

37 Sailor, M. J. *Porous Silicon in Practice: Preparation, Characterization and Applications*. (Wiley, 2012).

38 Lin, V. S.-Y., Motesharei, K., Dancil, K.-P. S., Sailor, M. J. & Ghadiri, M. R. A Porous Silicon-Based Optical Interferometric Biosensor. *Science* **278**, 840-843, (1997).

39 Dorvee, J. & Sailor, M. J. A low-power sensor for volatile organic compounds based on porous silicon photonic crystals. *Phys. Status Solidi* **202**, 1619-1623, (2005).

40 Kim, D., Zuidema, J. M., Kang, J., Pan, Y., Wu, L., Warther, D., Arkles, B. & Sailor, M. J. Facile Surface Modification of Hydroxylated Silicon Nanostructures Using Heterocyclic Silanes. *J. Am. Chem. Soc.* **138**, 15106-15109, (2016).

41 Park, J.-H., Gu, L., von Maltzahn, G., Ruoslahti, E., Bhatia, S. N. & Sailor, M. J. Biodegradable luminescent porous silicon nanoparticles for in vivo applications. *Nat. Mater.* **8**, 331-336, (2009).

42 Cunin, F., Schmedake, T. A., Link, J. R., Li, Y. Y., Koh, J., Bhatia, S. N. & Sailor, M. J. Biomolecular screening with encoded porous-silicon photonic crystals. *Nat. Mater.* **1**, 39-41, (2002).

43 Joannopoulos, J. D., Villeneuve, P. R. & Fan, S. Photonic crystals: putting a new twist on light. *Nature* **386**, 143-149, (1997).

44 King, B. H., Ruminski, A. M., Snyder, J. L. & Sailor, M. J. Optical-Fiber-Mounted Porous Silicon Photonic Crystals for Sensing Organic Vapor Breakthrough in Activated Carbon. *Adv. Mater.* **19**, 4530-4534, (2007).

45 King, B. H., Gramada, A., Link, J. R. & Sailor, M. J. Internally Referenced Ammonia Sensor Based on an Electrochemically Prepared Porous SiO2 Photonic Crystal. *Adv. Mater.* **19**, 4044-4048, (2007).

46 Lee, J., Mahon, S. B., Mukai, D., Burney, T., Katebian, B. S., Chan, A., Bebarta, V. S., Yoon, D., Boss, G. R. & Brenner, M. The Vitamin B12 Analog Cobinamide Is an Effective Antidote for Oral Cyanide Poisoning. *J. Med. Toxicol.* **12**, 370-379, (2016).

47 Greenawald, L. A., Snyder, J. L., Fry, N. L., Sailor, M. J., Boss, G. R., Finklea, H. O. & Bell, S. Development of a cobinamide-based end-of-service-life indicator for detection of hydrogen cyanide gas. *Sens. Actuators B Chem.* **221**, 379-385, (2015).

48 Greenawald, L. A., Boss, G. R., Snyder, J. L., Reeder, A. & Bell, S. Development of an Inexpensive RGB Color Sensor for the Detection of Hydrogen Cyanide Gas. *ACS Sens.* **2**, 1458-1466, (2017).

49 Ma, J., Dasgupta, P. K., Zelder, F. H. & Boss, G. R. Cobinamide chemistries for photometric cyanide determination. A merging zone liquid core waveguide cyanide analyzer using cyanoaquacobinamide. *Anal. Chim. Acta* **736**, 78-84, (2012).

50 Bala, T., Prasad, B. L. V., Sastry, M., Kahaly, M. U. & Waghmare, U. V. Interaction of Different Metal Ions with Carboxylic Acid Group: A Quantitative Study. *J. Phys. Chem. A* **111**, 6183-6190, (2007).

51 Gail, E., Gos, S., Kulzer, R., Lorösch, J., Rubo, A., Sauer, M., Kellens, R., Reddy, J., Steier, N. & Hasenpusch, W. Cyano Compounds, Inorganic. In Ullmann's Encyclopedia of Industrial Chemistry, (Ed.), (2011).

52 Zhang, S., Alberstein, R., De Yoreo, J. J. & Tezcan, F. A. Assembly of a Patchy Protein into Variable 2D Lattices via Tunable, Multiscale Interactions. *ChemRxiv*, (2020).

53 Ueda, T., Tamura, T. & Hamachi, I. In Situ Construction of Protein-Based Semisynthetic Biosensors. *ACS Sens.* **3**, 527-539, (2018).

54 Rong, G., Corrie, S. R. & Clark, H. A. In Vivo Biosensing: Progress and Perspectives. *ACS Sens.* **2**, 327-338, (2017).

55 Grattieri, M. & Minteer, S. D. Self-Powered Biosensors. *ACS Sens.* **3**, 44-53, (2018).

Chapter 4: Assembly of a patchy protein into variable 2D lattices via tunable, multiscale interactions

4.1 Abstract

The self-assembly of molecular building blocks into higher-order structures is exploited in living systems to create functional complexity and represents a powerful synthetic 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 that of inorganic or polymeric nanoparticles, which lack such attributes. We report the modular self-assembly of an engineered protein into four physicochemically distinct, precisely patterned 2D crystals via control of four classes of interactions acting locally, regionally and globally. 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 new bulk-scale properties for protein-based materials that ensue from such control.

4.2 Introduction

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[1]. 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[2,3]. Yet, as component complexity increases through the introduction of specific interactions and anisotropy—or "patchiness"—so does the resulting structural and functional diversity[3-5].

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)[6,7], regionally (*e.g.*, through associative non-covalent patches or charged surfaces)[8,9] or globally (*e.g.*, through shape complementarity and overall charge)[10,11]. 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[12,13]. This characteristic is universally exploited by living systems to create dynamic and reconfigurable architectures whose structures can be altered by environmental conditions[14] or physical templates (*e.g.*, chaperones[15] and 2D membranes[16]), 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[6,8,17,18] 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[19,20] unless the designed interactions themselves are inherently flexible or tunable[6,7,21,22]. 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.

4.3 Results and Discussion

4.3.1 Structural characterization of solution-grown ^{C98}RhuA lattices

We utilized the highly patchy protein building block L-rhamnulose-1-phosphate aldolase, a C_4-symmetric protein modified at its corners with Cys residues (^{C98}RhuA) (**Figure 4.1a**), 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[7]. 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 ^{C98}RhuA molecules (**Figure 4.1b**)[7,23]. Interestingly, transmission electron microscopy and electron diffraction analyses suggested that the lattices exclusively possessed $p42_12$ symmetry[7], whereby neighbouring ^{C98}RhuA molecules adopted an out-of-plane antiparallel (up-down) arrangement. Atomic force microscopy (AFM) imaging of solution-assembled ^{C98}RhuA crystals directly confirms this periodic, alternating arrangement of ^{C98}RhuA molecules throughout individual lattices (**Figure 4.1d**). The lower protein layer is most clearly resolved at step edges, where tip-protein convolution artefacts on the terrace are minimized (**Figure 4.2**).

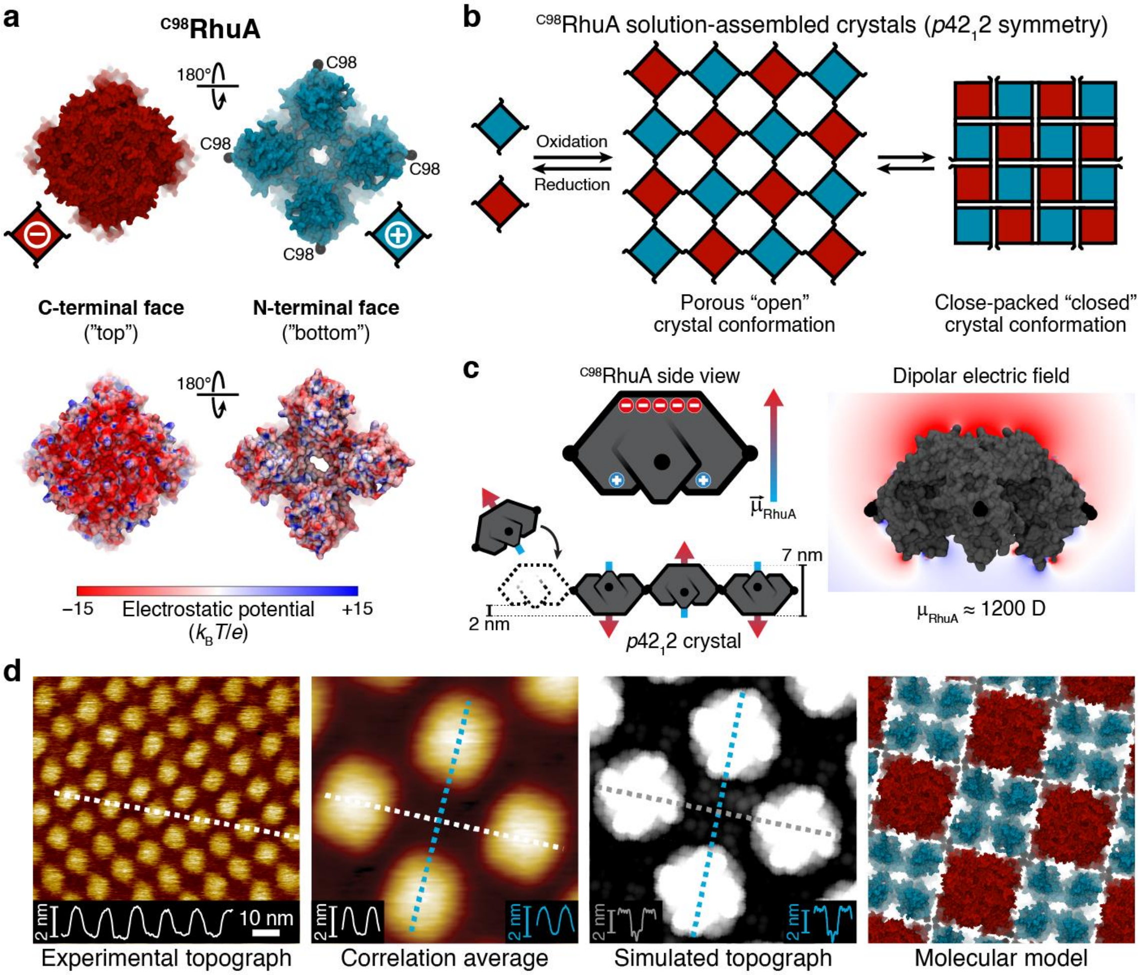


Figure 4.1 | Solution self-assembly of ^{C98}RhuA. a, Solvent-excluded surfaces and electrostatic potential maps of ^{C98}RhuA termini faces, colored according to charge. C98 residues are shown as black spheres. **b,** Schematic depicting ^{C98}RhuA solution self-assembly into 2D $p42_12$ crystals and their emergent auxetic behavior. **c,** Origin of the ^{C98}RhuA macrodipole moment and its "global"-scale electric field (color range: −7.5 to +2.5 k_BT/e). **d,** High-resolution AFM topographs and corresponding structural model of solution-assembled "closed" crystals. The height profiles correspond to the dashed lines.

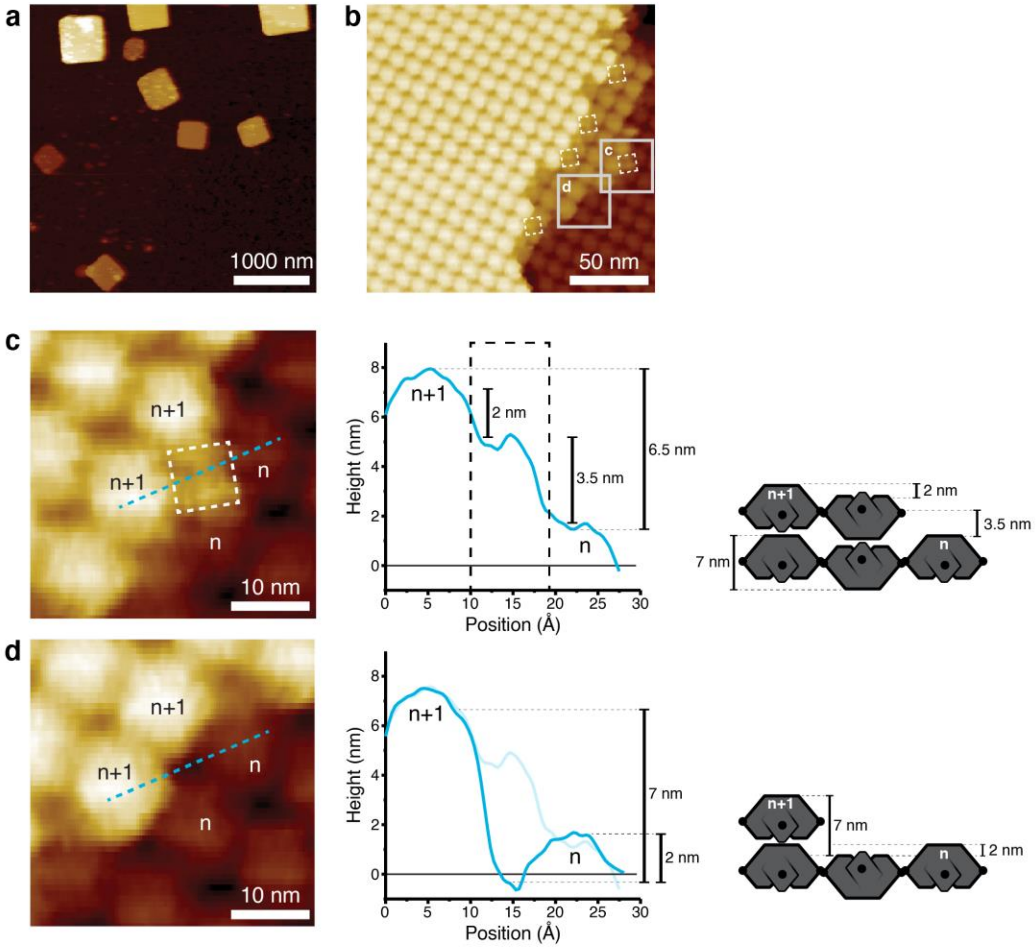


Figure 4.2 | Step-edges of *p*42$_1$2 crystals. a,** Low-magnification image of stacked solution-grown *p*42$_1$2 crystals deposited onto poly-lysine covered *m*-mica. **b, High-magnification image of the crystal edges, showing the stacking of layers. Dangling Nterm-up proteins are highlighted with white dashed boxes in **b** and **c**. The solid light gray boxes in **b** indicate regions selected for more detailed analysis (**c** and **d**). **c,** Close-up AFM image of a terminal Nterm-up protein with corresponding height trace and cartoon to highlight the stepped geometry of the crystal edge. **d,** Same as c but without a terminal Nterm-up protein, showing the expected *ca.* 7 nm drop in height for two stacked *p*42$_1$2 crystals. The height trace from c, shifted by *ca.* −0.374 nm to align its starting height relative to **d** at 0 Å position, is reproduced faintly to facilitate comparison.

4.3.2 Dipolar interactions determine molecular patterning of ^{C98}RhuA crystals in solution

Given that the only inter-protein connections in the ^{C98}RhuA lattices are disulfide bonds[23] (all protein-protein contacts are >5 Å, which is too distant even for weak hydrogen bonds), we originally attributed this distinct up-down arrangement to an energetic bias imposed by different disulfide bond configurations[7]. However, the flexibility of the disulfide hinges suggests a lack of 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 ^{C98}RhuA lattice (**Figure 4.3**). This finding implies there must be indirect inter-protein interactions acting over several nm's (>5 nm's) to yield the alternating ^{C98}RhuA registry.

We next considered the possible impact of dipole-dipole interactions, which have been shown to promote antiparallel packing in covalent organic frameworks[24], small-molecule crystals[25], and nanoparticle lattices[26-28], and have been invoked in stabilization of biomolecular complexes[10], but never implicated as a determinant of protein self-assembly. Examination of the ^{C98}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 (**Figure 4.1a, c**), yielding a sizeable macrodipole moment (*ca*. 1200 D). This polarized structure allows ^{C98}RhuA to be effectively modelled as a physical dipole (**Figure 4.1c**; elaborated in **Section 4.5**), as done previously for inorganic nanoparticles[26,28,29]. Calculation of the pairwise ^{C98}RhuA dipole-dipole interaction over 7–10 nm (dimensions accessible by lattice conformations) reveals a smooth, funnel-shaped energy landscape invariably favouring the antiparallel arrangement of $p42_12$ crystals—relative to parallel—by 0.13–1.11 kcal/mol (0.22–1.87 k_BT), even at such long distances (**Figure 4.4**). Despite this energy being on the order of weak non-covalent bonds, its magnitude is rapidly amplified four-fold by nearest-neighbour interactions (**Figure 4.5**), propagating $p42_12$ symmetry to newly incorporated monomers during solution-phase crystallization. This indicates the collective effect of a multitude of weak, long-range, orientation-dependent interactions controls the precise 2D patterning of ^{C98}RhuA assemblies.

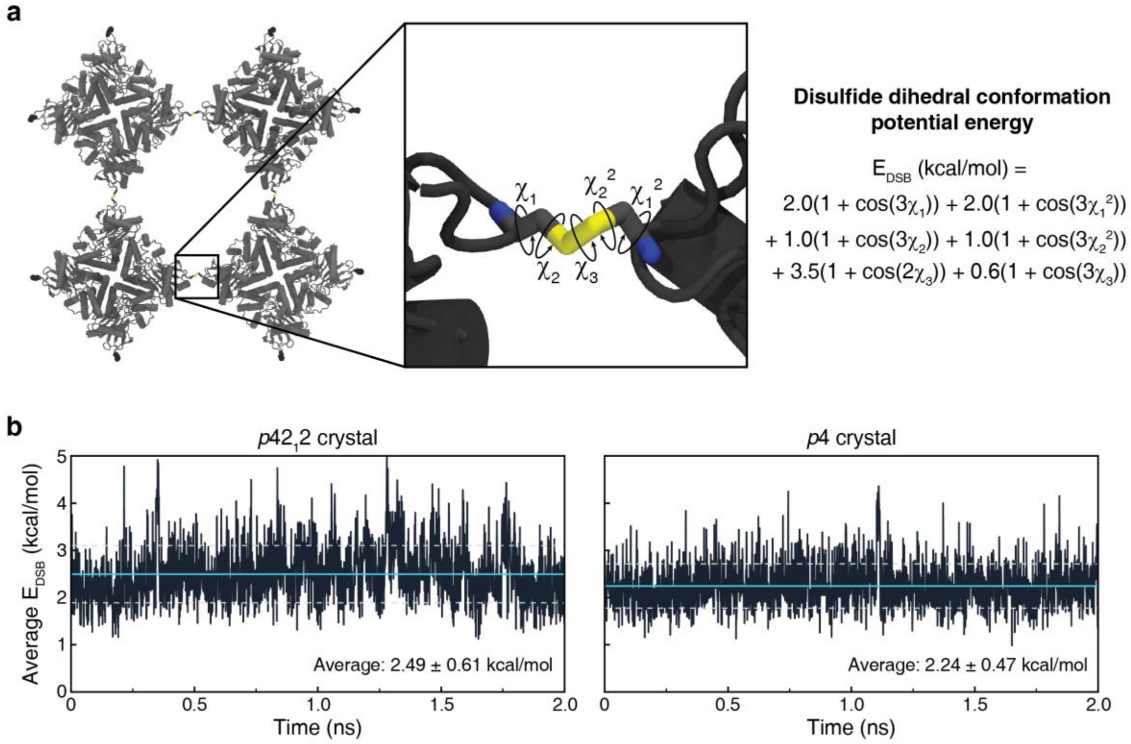


Figure 4.3 | Lack of conformational preference of disulfide bonds. a, Close-up rendering of one of the four disulfide bonds not connected through the periodic boundaries in the 2x2 structure. The five dihedral angles which fully define a disulfide conformation are labeled in the inset, with the equation which relates them to their dihedral potential energy on the right. **b,** Plots of the average dihedral potential energy for the four non-periodic disulfides over the final 2 ns of sampling from the 3D 2x2 simulations. The time average is marked with a cyan line and boundaries of ±1 standard deviation are delineated with dashed light grey lines.

a

Screened pairwise RhuA potentials in 20 mM buffer salts

$$U_{total}(r_{ij},\Delta\varphi_{ij}) = U_{qq}(r_{ij}) + U_{\mu\mu}(r_{ij},\Delta\varphi_{ij})$$

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

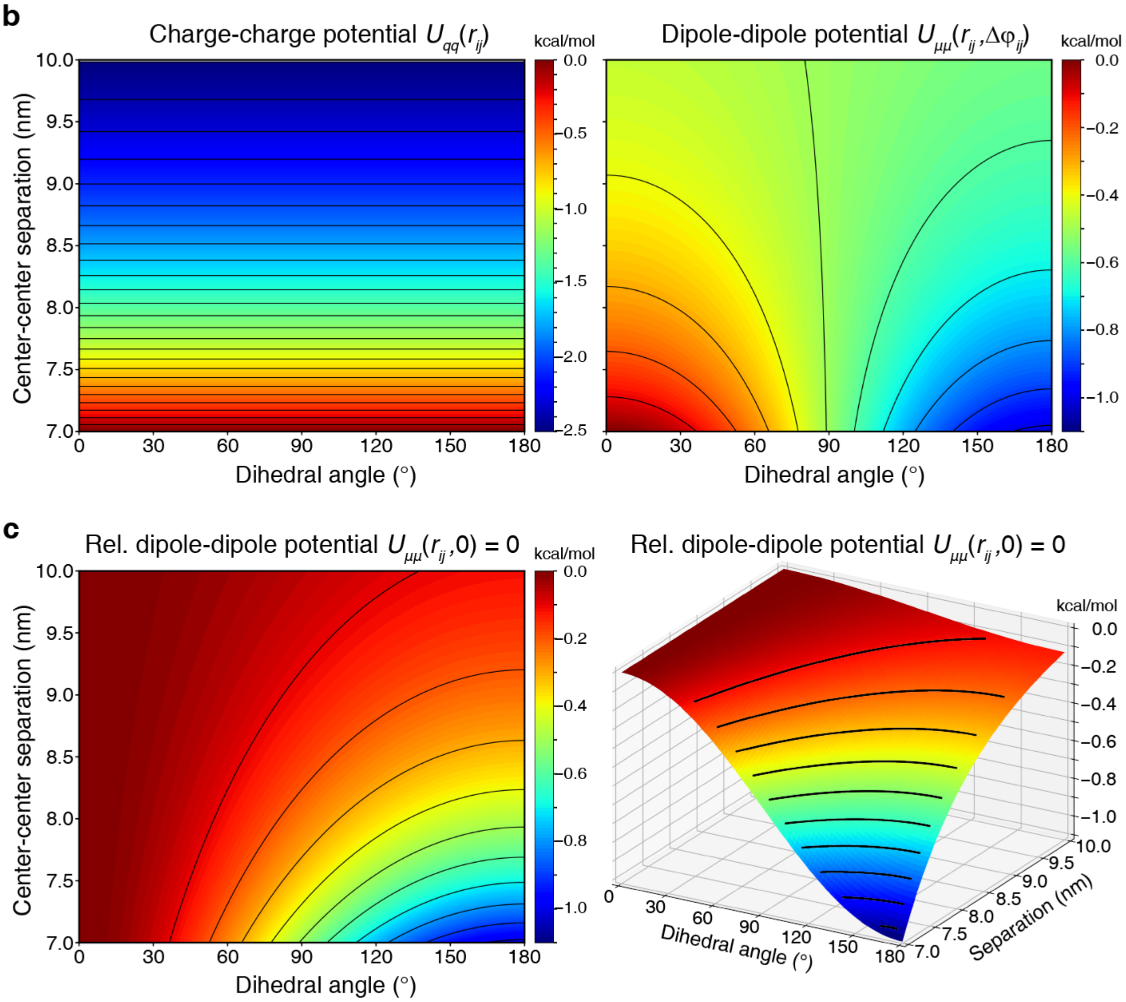
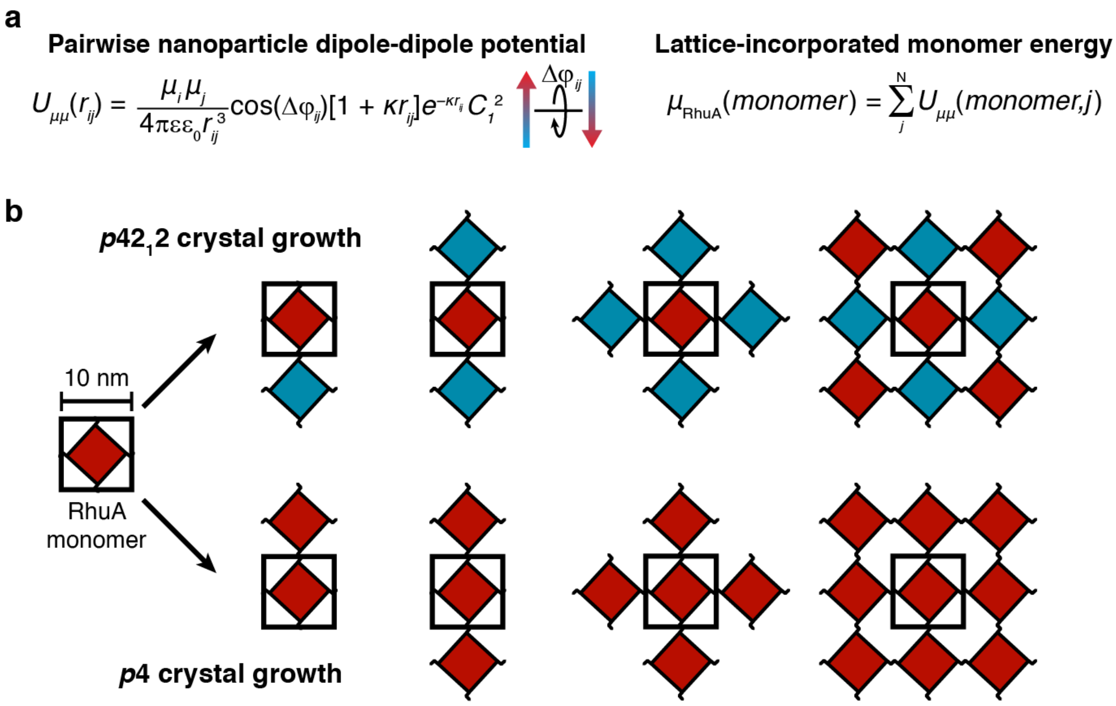


Figure 4.4 | ^{C98}RhuA nanoparticle pairwise electrostatic potential energy surfaces. a, Reduced form of the electrostatic nanoparticle-nanoparticle interaction potential (**Equation 4.4** (in Section 4.6)) and its individual components (charge and dipolar). **b,** Absolute potential energy landscapes for each component of the total potential, shifted such that the maximum energy is at 0.0 kcal/mol, highlighting the dihedral dependence unique to the dipole-dipole potential. The global maximum and minimum energies correspond to $\Delta\varphi_{ij} = 0$ (parallel dipoles) and $\Delta\varphi_{ij} = 180$ (antiparallel dipoles), respectively, at 7 nm separation. **c,** 2D and 3D representations of a relative dipole-dipole potential, shifted such that the energy at $\Delta\varphi_{ij} = 0$ (parallel dipoles) is set to 0.0 kcal/mol for all distances, revealing the potential energy funnel created by dipolar interactions over all r_{ij}. Black contour lines in all plots represent decrements of 0.1 kcal/mol relative to 0.0.



# neighbors	N = 1	N = 2	N = 4	N = 8	
$\mu_{RhuA}(p42_12)$ =	−0.063	−0.126	−0.251	−0.234	(kcal/mol)
$\mu_{RhuA}(p4)$ =	+0.063	+0.126	+0.251	+0.268	(kcal/mol)
$\Delta\mu_{RhuA}$ (open) =	−0.126	−0.251	−0.502	−0.502	(kcal/mol)
$\Delta\mu_{RhuA}$ (closed) =	−1.052	−2.104	−4.208	−4.208	(kcal/mol)

Figure 4.5 | Nearest-neighbor amplification of dipole-dipole interactions. a, Reduced form of the dipole-dipole interaction potential (**Equation 4.4** (in Section 4.6)) and its extension to calculate the total dipolar energy felt by a single monomer as it is incorporated into a lattice under experimental self-assembly conditions (aqueous 20 mM buffer solution). **b,** Cartoon depicting the nucleation of both *p*4 and *p*42$_1$2 crystals around a central ^{C98}RhuA monomer (**top**) and its cumulative dipole potential energy (μ_{RhuA}) as a function of crystal symmetry and number of nearest neighbors (**bottom**). Individual energies for all sizes are included for the open-state crystals (10 nm protein separations), while the potential energy difference ($\Delta\mu_{RhuA}$) is also reported for the closed state (7.071 nm protein separations) to depict the range of magnitudes which promotes antiparallel packing during solution self-assembly. Attenuated energies at the growing edge facilitate reorientation of newly attached monomers before being "locked in" by nearest-neighbor interactions upon incorporation into the bulk lattice. The total potential converges within one shell of neighbors, as shown by the negligible change in $\Delta\mu_{RhuA}$ for N ≥ 4.

4.3.3 Surface-templated self-assembly of ^{C98}RhuA crystals

The above findings suggest the charge anisotropy of ^{C98}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[30]. We found that ^{C98}RhuA readily formed large domains of 2D square lattices on *m*-mica at ≥750-fold lower concentrations than required for solution assembly (≤0.2 μM vs. ≥150 μM). These lattices exhibited both "open" and "closed" conformations resembling those of solution-grown crystals, indicating the designed intermolecular disulfide bonding was preserved and that ^{C98}RhuA-*m*-mica interactions were mediated by the charged surfaces of ^{C98}RhuA (**Figure 4.6**). Indeed, a variant in which Cys98 residues were replaced with Ser (^{S98}RhuA) adsorbed to *m*-mica but failed to form 2D lattices even at concentrations of 50 μM (**Figure 4.7**). ^{C98}RhuA assembly on *m*-mica could also be prevented (reversed) in the presence of 1 mM reductant, which prevents (reverses) disulfide bonding (**Figure 4.8**).

In contrast to the antiparallel registry of solution-assembled lattices, crystals formed on *m*-mica possessed *p*4 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 (**Figure 4.6**). In 5 mM KCl, two classes of *p*4-symmetric ^{C98}RhuA crystals formed concurrently: close-packed crystals with C-terminus up (Cterm-up crystals) and porous crystals with N-terminus up (Nterm-up crystals) (**Figure 4.6a,** 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 ^{C98}RhuA molecules (**Figure 4.9**), the relative

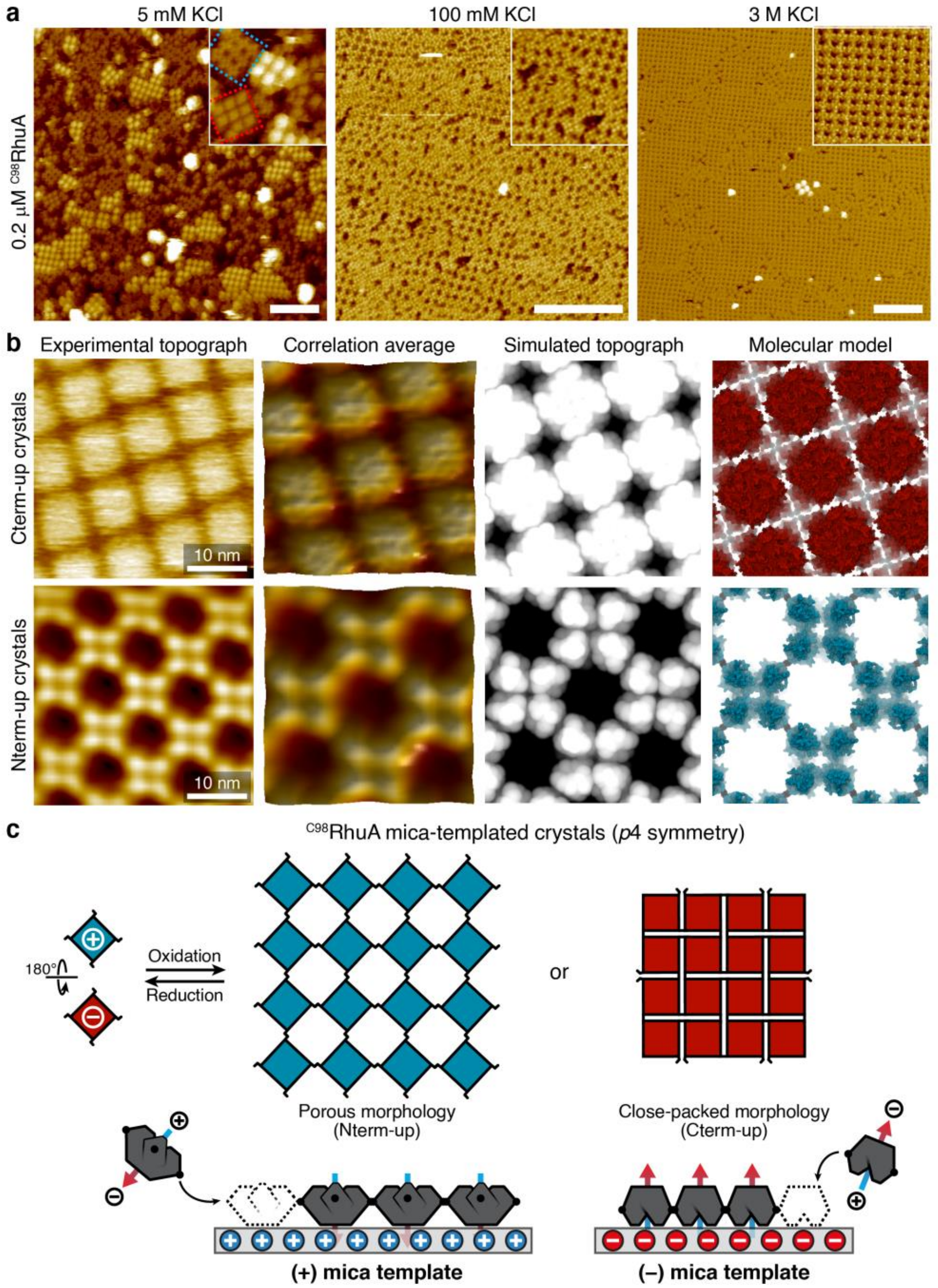


Figure 4.6 | Surface-templated self-assembly of *p*4 ^{C98}RhuA crystals. a, Growth of monolayer ^{C98}RhuA crystals on mica under increasing [K$^+$]. Scale bars are 100 nm. **b,** High-resolution AFM topographs and structural models of mica-templated crystals. **c,** Cartoon overview of mica-templated self-assembly.

orientations of the adsorbed proteins were readily determined from their surface topologies (**Figure 4.1a**, **Figure 4.6b**). Interpretations of the images were corroborated by simulated surface profiles calculated from structural models of *p*4-symmetric ^{C98}RhuA crystals (**Figure 4.6b**). 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 (**Figure 4.6a**, left). At 100 mM K$^+$, poorly ordered Nterm-up crystals dominated (**Figure 4.6a**, middle) and at 3 M K$^+$, self-assembly exclusively yielded highly ordered 2D Nterm-up lattices containing hundreds to thousands of monomers (**Figure 4.6a**, right).

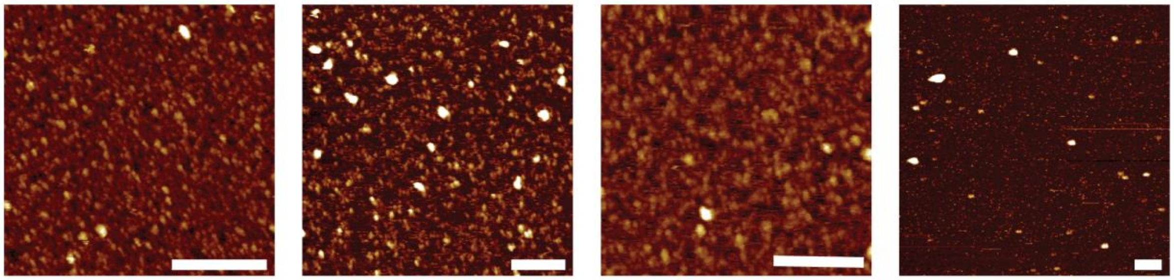
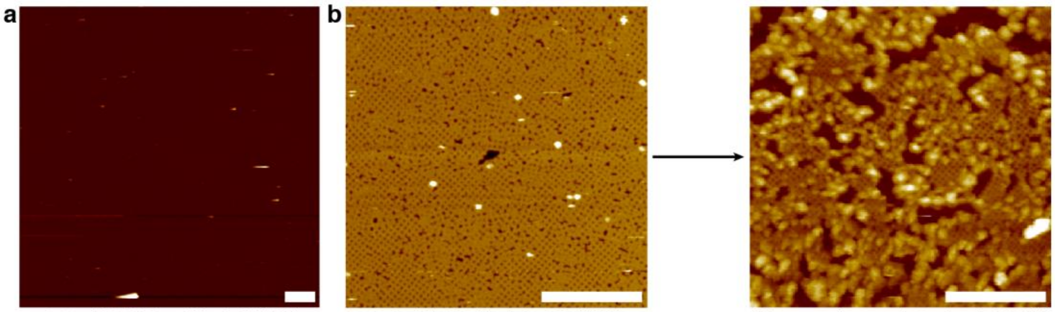

50 μM ^{S98}RhuA, 5 mM KCl
1 μM ^{S98}RhuA, 10 mM KCl
10 μM ^{S98}RhuA, 10 mM KCl
20 μM ^{S98}RhuA, 3 M KCl

Figure 4.7 | Attempted self-assembly of ^{S98}RhuA on *m*-mica. AFM images of ^{S98}RhuA incubated on *m*-mica at various concentrations of protein and KCl. Crystallization did not occur under any conditions explored here. Scale bars: 200 nm.



0.2 μM ^{C98}RhuA in 3 M KCl
+ 1 mM GSH (no assembly)
0.2 μM ^{C98}RhuA in 3 M KCl
+ 0.2 mM GSH
After exchange into 1 mM GSH

Figure 4.8 | Effect of reductant on surface-grown lattices. a, No self-assembly was observed after a 48h incubation on *m*-mica in the presence of 1 mM reduced glutathione (GSH). **b,** Surface-grown crystalline networks (formed after 48h incubation on *m*-mica) undergo disassembly upon the introduction of additional GSH, demonstrating that disulfide bond crosslinks are essential to maintain the protein lattice structure. Scale bars: 200 nm.

This variability in ^{C98}RhuA assembly can be understood from the charge state of *m*-mica (**Figure 4.6c**), which follows a well-documented trend towards positive values with increasing KCl[30]. Freshly cleaved *m*-mica surfaces are negatively charged in neutral aqueous solution. 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[31]. At $[K^+]$ >100 mM, the surface sites are more than 50% occupied, fully compensating the surface charge and resulting in positive zeta potentials[30,32]. Accordingly, while 0.2 μM solutions of ^{C98}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^+ (**Figure 4.6a**). At much lower [^{C98}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 (**Figure 4.10**). Moreover, selectivity for monomer/crystal binding orientation followed well-known trends in specific ion adsorption to *m*-mica[30,33]: low-affinity counterions (*e.g.*, Zn^{2+}) only produced Cterm-up crystals, while Rb^+—a known substitute for K^+[34]—and Mg^{2+} (widely used to induce surface-charge reversal on *m*-mica for DNA adsorption[35]) recapitulated the behaviour with K^+ (**Figure 4.11**).

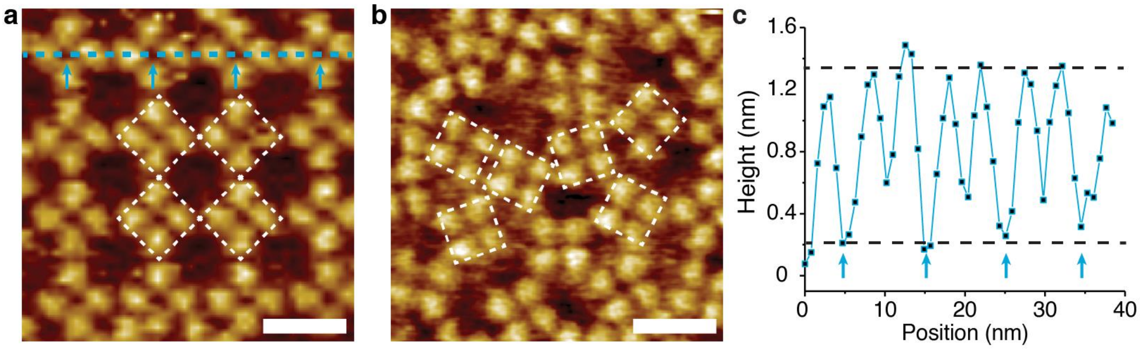


Figure 4.9 | High-resolution images of Nterm-up ^{C98}RhuA proteins. AFM images of Nterm-up ^{C98}RhuA crystals (**a**) and disordered monomers (**b**). White boxes indicate representative individual proteins. **c**, Height profile along the line trace in **a**. The blue arrows in **a** and **c** denote the central dip of the protein N-terminus. Scale bars: 10 nm.

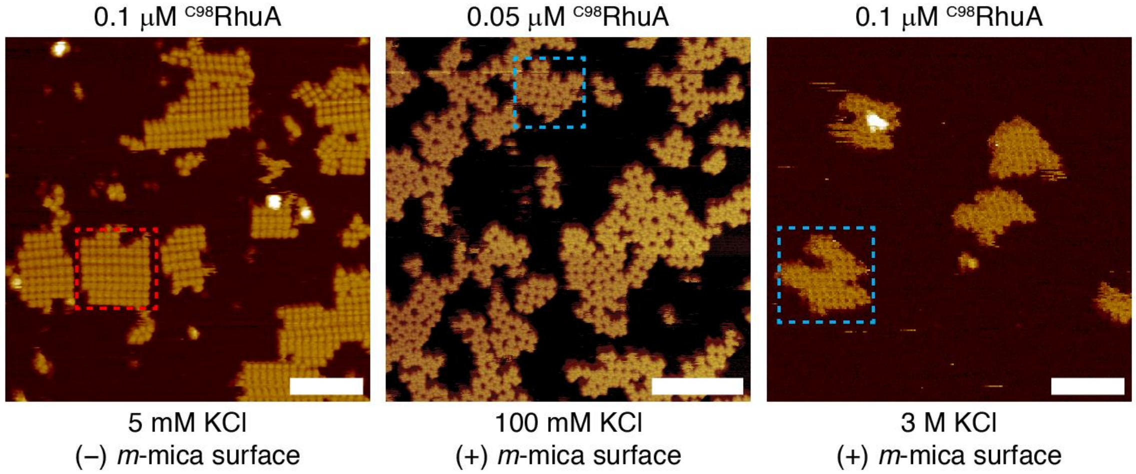
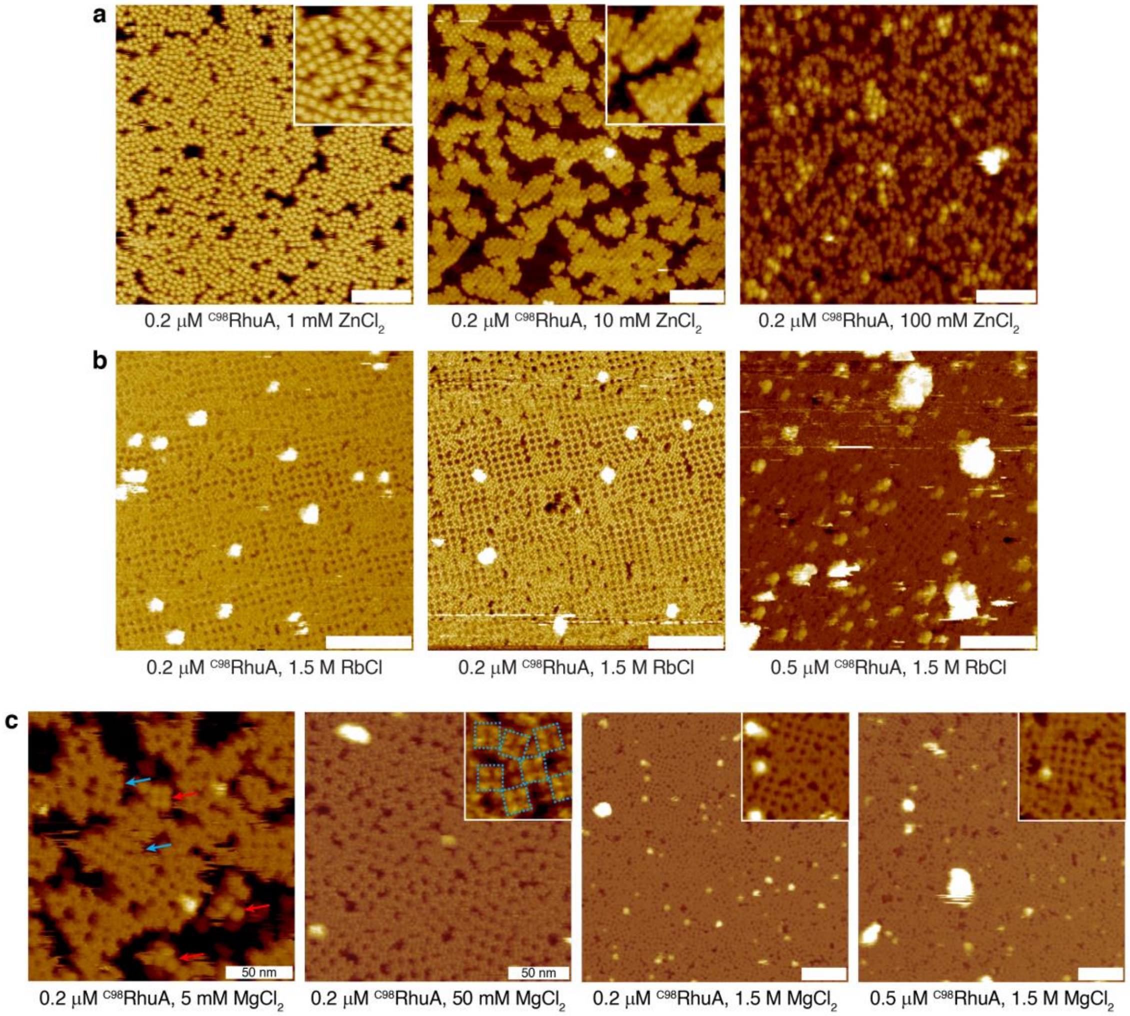


Figure 4.10 | Absolute crystal orientation selectivity reflects mica surface charge. Self-assembly of very low [^{C98}RhuA] onto *m*-mica yields small isolated crystalline domains of identical orientation, likely corresponding to regions of (higher) homogenous local surface charge. At 5 mM KCl, the heterogeneous distribution of K$^+$ ions is insufficient to prevent preferential binding of the cationic ^{C98}RhuA N-terminus to the negative surface. This is in contrast to the observed heterogeneity when the surface is fully covered by protein (**Fig. 2a**). Assembly in ≥100 mM KCl exclusively yields open-state Nterm-up crystals, as observed for higher protein concentrations. Red and blue boxes highlight representative Cterm-up and Nterm-up crystals, respectively. Scale bars: 100 nm.



0.2 μM ^{C98}RhuA, 1 mM $ZnCl_2$ 0.2 μM ^{C98}RhuA, 10 mM $ZnCl_2$ 0.2 μM ^{C98}RhuA, 100 mM $ZnCl_2$

0.2 μM ^{C98}RhuA, 1.5 M RbCl 0.2 μM ^{C98}RhuA, 1.5 M RbCl 0.5 μM ^{C98}RhuA, 1.5 M RbCl

0.2 μM ^{C98}RhuA, 5 mM $MgCl_2$ 0.2 μM ^{C98}RhuA, 50 mM $MgCl_2$ 0.2 μM ^{C98}RhuA, 1.5 M $MgCl_2$ 0.5 μM ^{C98}RhuA, 1.5 M $MgCl_2$

Figure 4.11 | Specific ion effects on ^{C98}RhuA self-assembly on *m*-mica. Mica-templated self-assembly of ^{C98}RhuA at low concentrations in the presence of **a,** Zn^{2+}, **b,** Rb^{+}, and **c,** Mg^{2+}. Zn^{2+} ions do not effectively bind the surface, so it remains negative and enforces exclusively Cterm-up binding, regardless of [Zn^{2+}]. Rb^{+} ions are known substitutes for K^{+} ions in the surface vacancies, and thus yield analogous Nterm-up crystals at molar concentrations. **c**, Increasing concentrations of Mg^{2+} ions recapitulates the trend in bound protein/crystal orientation observed for K^{+} (and Rb^{+}), but reversal occurs at half of the monovalent concentration, suggesting that all three ions modulate the mica surface charge upon adsorption. Red and blue arrows/boxes identify Cterm-up and Nterm-up oriented proteins, respectively. Scale bars are 100 nm except where noted.

4.3.4 Hierarchical self-assembly of bilayer ^{C98}RhuA crystals

When [^{C98}RhuA] was increased to >0.2 μM and [K^+] was held at 3 M, a third 2D lattice arrangement with a height of 9 nm emerged (**Figure 4.12**). The surface topography resembled an array of Cterm-up monomers (**Figure 4.12a**), but this array formed exclusively atop domains of 2D Nterm-up crystals (**Figure 4.12b**, left) and was shorter than twice the height (*ca*. 10 nm) of the underlying Nterm-up monolayer (**Figure 4.12b**, right), implying the ^{C98}RhuA monomers in each layer must interact via interpenetration of the corrugated N-term faces (**Figure 4.12c**). We hypothesized that this tail-to-tail packing is isostructural with the obligate, 9.5-nm tall dimers of the $^{F88/C98}$RhuA variant, whose formation during growth of crystals in bulk solution is stabilized by hydrophobic, Phe88-mediated interactions between the N-term faces[7]. Indeed, the simulated topography for a bilayer based on the $^{F88/C98}$RhuA lattice structure (**Figure 4.12d**, **Figure 4.13**) closely resembles the high-resolution AFM images (**Figure 4.12e**). Importantly, the unique circular shape of the protein units in bilayer crystals seen by AFM is reproduced upon averaging of both substructures of the $^{F88/C98}$RhuA lattice (**Figure 4.13b**). Furthermore, free-energy (*i.e.*, potential of mean force) calculations show that ^{C98}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 (**Figure 4.12f**). N-term intercalated ^{C98}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 *p*4-symmetric monolayer lattices (likely through exfoliation) (**Figure 4.14**) that cannot form without the template.

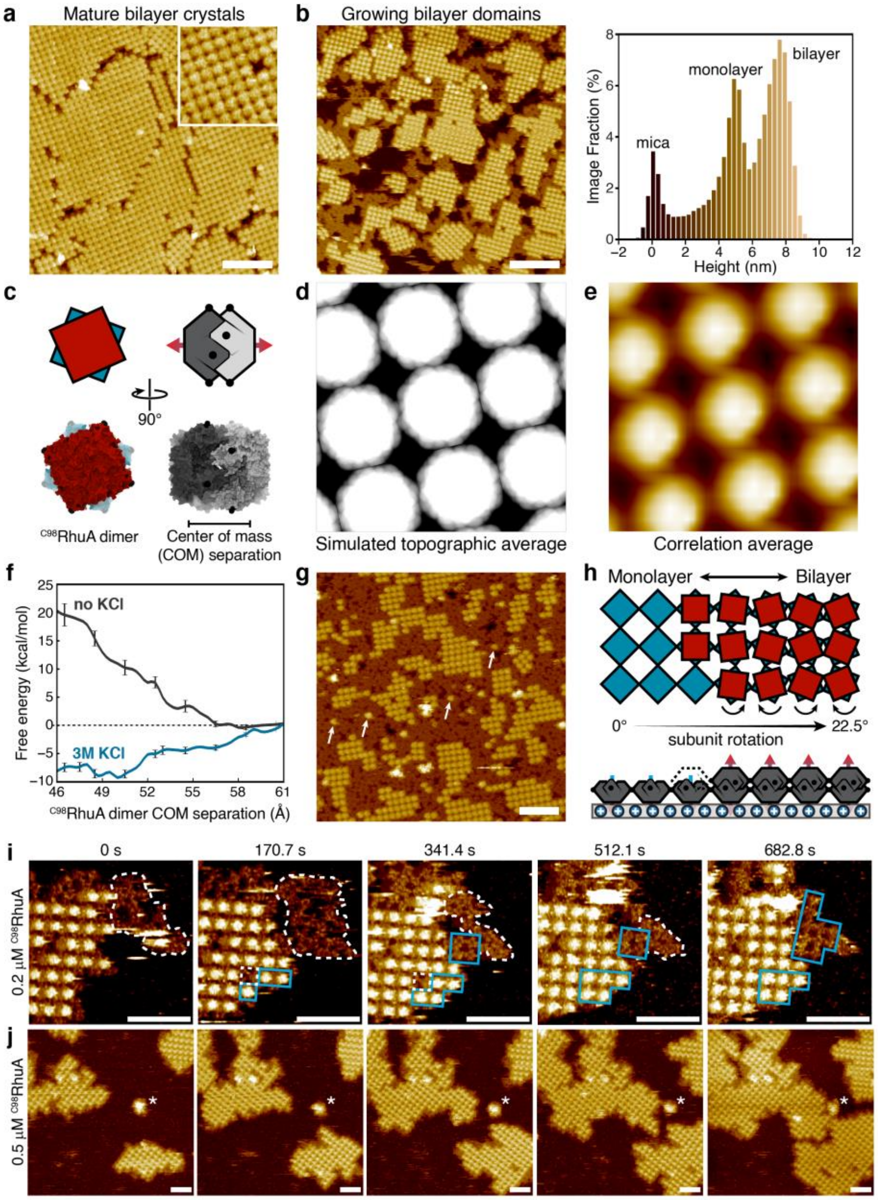


Figure 4.12 | Hierarchical self-assembly of bilayer ^{C98}RhuA crystals. High-resolution AFM images of **a,** mature and **b,** growing bilayer crystals (0.4-0.5 μM ^{C98}RhuA), with image height distribution shown on right. **c,** ^{C98}RhuA dimer structure. **d,** Simulated and **e,** experimental average topographs for the bilayer structure. **f,** Potential of mean force for ^{C98}RhuA dimerization in the presence/absence of 3 M KCl. **g,** Self-assembly of 5 μM ^{C98}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 ^{C98}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: 100 nm (**a**/**b**/**g**), 50 nm (**i**/**j**).

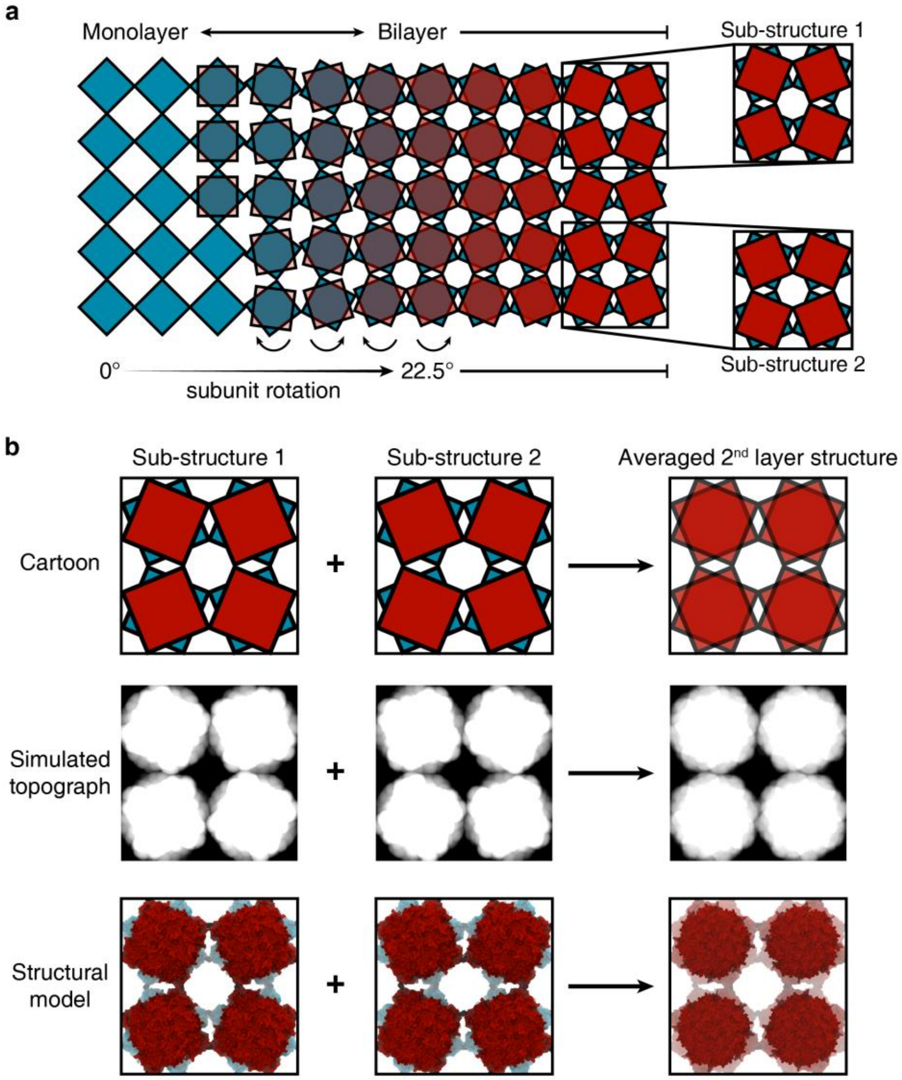
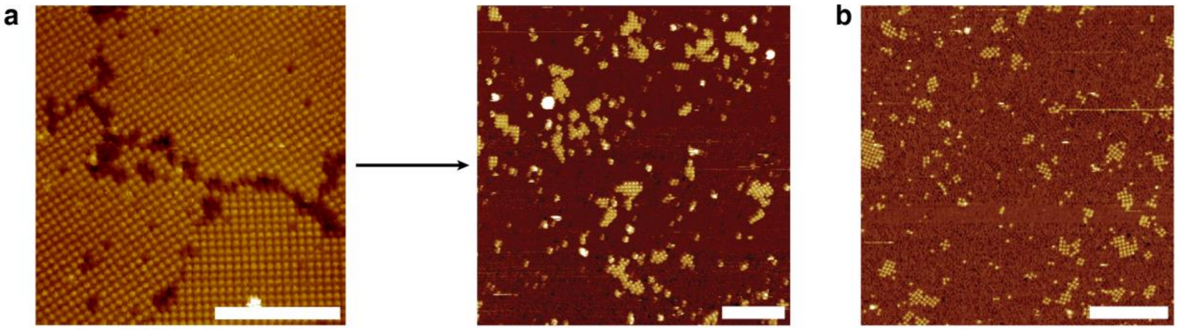


Figure 4.13 | Validation of the ^{C98}RhuA bilayer crystal morphology via tapping simulations. a, Expanded cartoon from **Fig. 3h**, highlighting the requisite 22.5° rotation of the first layer to bring 2nd layer proteins into disulphide-bonding distance (visible underneath low-opacity 2nd layer proteins on left) and resultant “sub-structures” arising from this geometric constraint. **b,** Depiction of the contribution of each sub-structure to the final average topograph, which is consistent with the experimental correlation average for the bilayer crystals.



0.5 μM ^{C98}RhuA assembled in 3 M KCl

After exchange into 100 mM KCl

0.5 μM ^{C98}RhuA assembled in 100 mM KCl

Figure 4.14 | Exfoliation of bilayer crystals. a, AFM images of ^{C98}RhuA bilayer crystals grown in the presence of 3 M KCl before (left) and after (right) exchanging buffer with 100 mM KCl solution. Most of the second layer falls away from the underlying monolayer due to loss of the salting-out effect from the high [KCl]. In this state these crystals resemble ^{C98}RhuA crystals grown directly in the presence of 100 mM KCl (**b**). Scale bars are 200 nm.

The importance of disulfide bonding is reflected by the rarity of isolated 2nd layer monomers (*i.e.*, N-term intercalated dimers) (**Figure 4.12g**). 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%[36], which is 20 times the observed 0.35% coverage of isolated monomers (**Figure 4.15**). Thus, the bilayer morphology is realized via a self-templating effect by the underlying *p*4-symmetric monolayer, whose subunits necessarily undergo a 22.5° rotation to bring the monomers in the second layer into the proximity required for disulfide bonding (**Figure 4.12h**, **Figure 4.13**). All-atom MD simulations carried out in 3 M KCl show that the C-termini of *m*-mica-adsorbed ^{C98}RhuA molecules preferentially reside 8-9 Å away from the mica surface (above the ionic slipping plane) (**Figure 4.16**), suggesting a negligible energetic barrier to in-plane rotation.

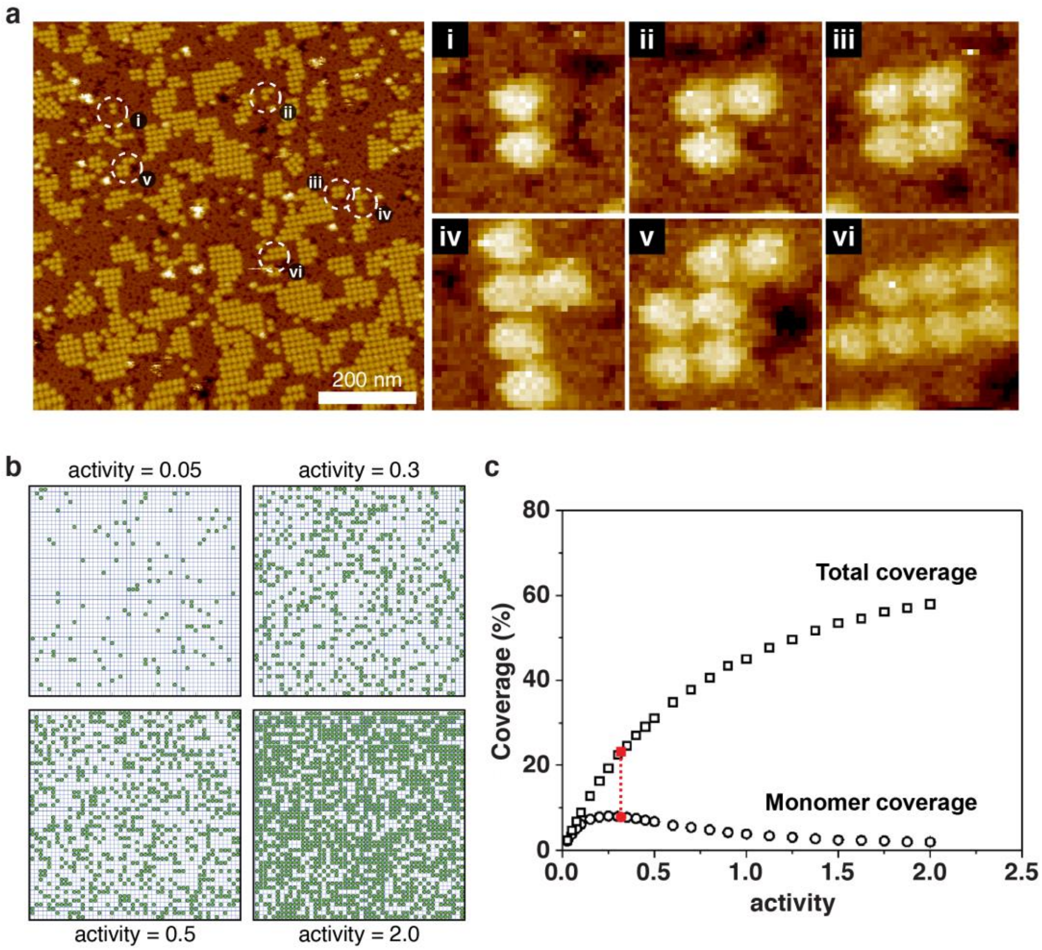


Figure 4.15 | Rarity of isolated 2nd-layer monomers in kinetically-trapped assemblies. a, AFM image of 5 μM ^{C98}RhuA assembled in the presence of 3 M KCl. Large regions of the first protein layer are non-crystalline due to the high concentration of protein, which precludes the formation of large bilayer domains. Only 0.35% of all 2nd layer proteins exist as monomers isolated from any nearest neighbors. Instead, nearly all 2nd layer domains consist of at least two neighboring monomers (**i-vi**; labeled regions and zoom-in images), suggesting that disulphide bonds are essential for the stability of bilayer crystals. The image on the left is the uncropped version of **Fig. 3g**. **b, c**, Relationship between total adsorbate coverage and corresponding isolated monomer coverage determined by grand canonical Monte Carlo simulation for random adsorption, desorption and displacement of non-interacting species with differing solution activities onto a 2D square lattice (adapted from ref. 36). **b** shows the spatial distribution of adsorbates and **c** shows the equilibrium adsorbate coverage, as well as the coverage of isolated monomers. The dashed line marks the total coverage of 23% obtained at 5 μM ^{C98}RhuA and 3 M KCl and the corresponding expected monomer coverage of 7% for non-interacting particles. Were these particles (2nd layer monomers) subject to repulsive interactions, this preference would manifest as a higher percentage of isolated monomers across all total coverage values. The low number of isolated 2nd layer monomers observed in our work (0.35%) thus indicates strongly favorable inter-monomer interactions (disulfide bonding) which are essential for stability of the 2nd layer.

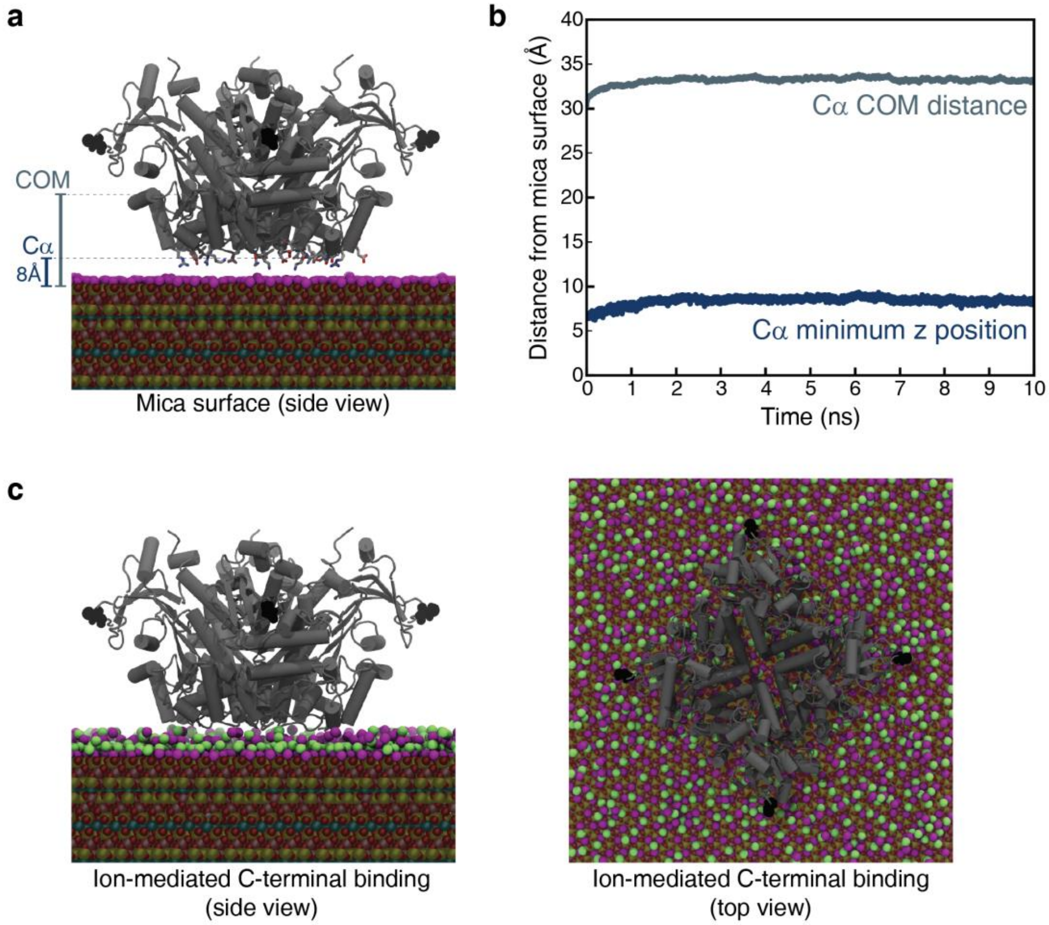


Figure 4.16 | Preferred binding geometry of RhuA C-terminus to *m*-mica in 3 M KCl. a, Rendering of ^{C98}RhuA associated with the *m*-mica surface at its equilibrium position (8-9 Å away from the surface). All residues with any atom ≤7 Å away are shown as sticks. The COM and Cα distances are depicted on the left. **b,** Plot of the change in protein-mica distance over 10 ns of equilibration; the protein reaches its equilibrium distance within 2-3 ns. **c,** Side and top views of **a** with all K^+ and Cl^- ions within 7 Å of the surface depicted as purple and green spheres, respectively. The observed double-layer ion behavior is typical of a charged surface, and mediates all protein-surface interactions, enabling the facile movement of ^{C98}RhuA on the surface by avoiding direct contacts. All renders are from the same snapshot at 5 ns of equilibration.

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 μM [^{C98}RhuA], the first (Nterm-up) layer crystallized via a nonclassical two-step nucleation process[37,38], whereby the growing edges (**Figure 4.12i**) advanced by initial formation of amorphous regions (white dashed contour), which then spontaneously crystallized (blue dashed square). (Interaction with the adjacent "core" that has already crystallized may catalyse the transition, a behaviour that has been observed previously during surface crystallization of S-layer proteins[13].) 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 (**Figure 4.17**), which must be imposed at the time of the transition. This capability for reorganization and ordering implies relatively weak ^{C98}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 (**Figure 4.12i, j**). At much higher [^{C98}RhuA] of 5 μM, the *m*-mica surface became fully covered by the amorphous layer before crystallization could occur, leaving the film kinetically trapped in a disordered state (**Figure 4.12g**). Consequently, only small bilayer domains formed and were surrounded by an amorphous monolayer (**Figure 4.15**).

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[39]. Such tunability underscores the principal advantage of using proteins as patchy building blocks for self-assembled materials.

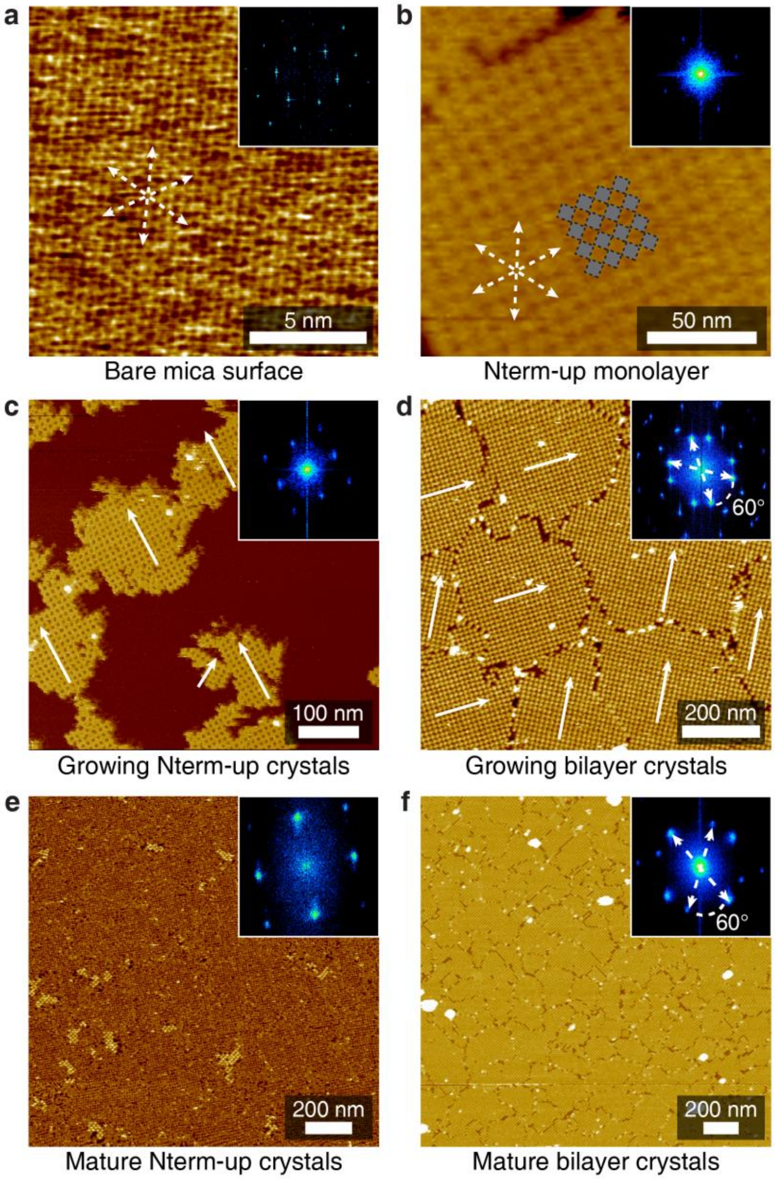


Figure 4.17 | Domain alignment along mica vectors. a, AFM image of the bare mica surface. The hexagonal arrangement of the mica lattice can be directly visualized, and the corresponding lattice vectors are drawn as dashed arrows. **b,** Nterm-up monolayer crystals imaged with the same underlying mica orientation, showing the alignment of ^{C98}RhuA proteins relative to the mica lattice vectors. **c, d,** Examples of crystal domain growth along preferred directions (indicated by arrows). **e, f,** Low-magnification images of mature crystals, demonstrating that crystallite domain orientations are preserved over thousands of unit cells. Growth directions are separated by 60°, consistent with the influence of the underlying substrate. The insets are FFTs.

4.3.5 Diverse morphologies from a singular "patchy" protein

In contrast to our previous work[7], which described the solution self-assembly of different RhuA variants into distinct 2D crystals (*i.e.*, three building blocks – three assemblies), our findings here (**Figure 4.18**) demonstrate that ^{C98}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 ^{C98}RhuA, which is a consequence of weak interactions between neighbouring macrodipoles (operating at a "global" level). 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 charged ^{C98}RhuA. These interactions (at a "regional" level) are strong enough to overcome the dipole-dipole interactions between ^{C98}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 ^{C98}RhuA, leading to formation of bilayer crystals. Such rich phase behaviour from a single building block is rare in nanoparticle self-assembly[4] and unprecedented among self-assembled protein lattices—natural or designed—because intermonomer interaction surfaces in such lattices are too extensive to allow reconfigurability.

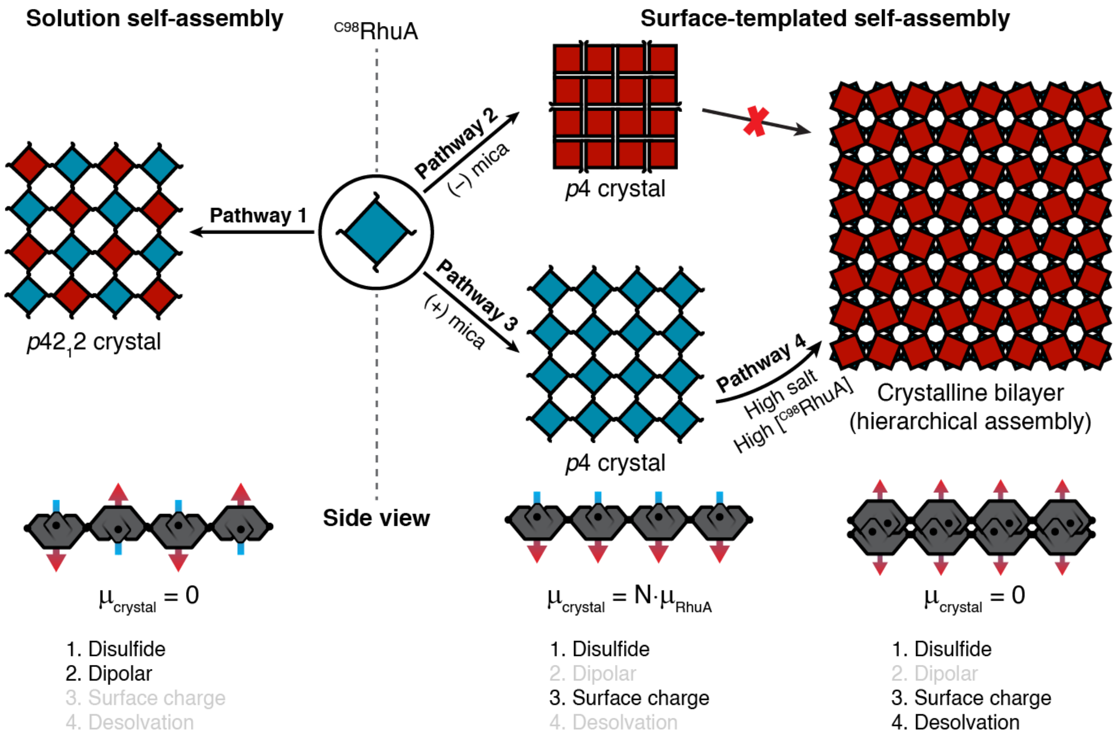


Figure 4.18 | Self-assembly pathways afforded by ^{C98}RhuA "patchiness" and template effects. Overview of all ^{C98}RhuA growth pathways and the self-assembly interactions responsible for their formation. The configuration-dependent dipole moment of each crystal morphology is also indicated.

4.3.6 New functional properties from template-grown *p*4 crystals

Varying the arrangement of molecules or nanoparticles within a material is known to give rise to different emergent bulk-scale properties[40,41]. Consequently, we considered the effect of the dipolar interactions within the new surface-templated 2D phases on their properties. ^{C98}RhuA crystals with *p*4-symmetry (one of only 10 polar crystal classes) contain an extended array of coaligned dipoles (**Figure 4.18**) and are formally "electrets": polarized dielectric materials possessing a permanent electric field[42]. 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 new properties: all-atom MD simulations of infinitely-periodic open-state *p*4-symmetric ^{C98}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 $p42_12$-symmetric crystals (**Figure 4.19a**), consistent with their non-polar crystal classification (**Figure 4.19b**). At equilibrium, a charge differential of 0.08–0.1 e/nm^2 was attained across the membrane, corresponding to a reversal potential of 50–60 mV (**Figure 4.19a**). Both the charge differential (40 *e*) and membrane potential are in excellent agreement with analytical calculations (**Section 4.5**) 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 (0.008–0.016 C/m^2 depending on crystal conformation, **Figure 4.19c**) and resulting electric field strength is intrinsically coupled to in-plane compression and expansion, *p*4-symmetric ^{C98}RhuA crystals are thus expected to exhibit piezoelectric properties[42], while $p42_12$-symmetric crystals are not. Subsequent all-atom simulations directly confirmed this prediction, achieving density-dependent reversal potentials of ≥100 mV at sufficiently compressed conformations (**Figure 4.19c**, **Figure 4.20**). This example illustrates the prospect of obtaining new bulk-scale physical properties through control of specific nanoscale patterning of protein building blocks, which in turn can be designed at the molecular level. Polarized membrane materials such as *p*4-symmetric ^{C98}RhuA crystals should also be promising candidates for new applications in biological nanoelectromechanical systems, including as nanogenerators and passive charge-storage devices, with tunable electrical properties coupled to mechanical in-plane lattice compression.

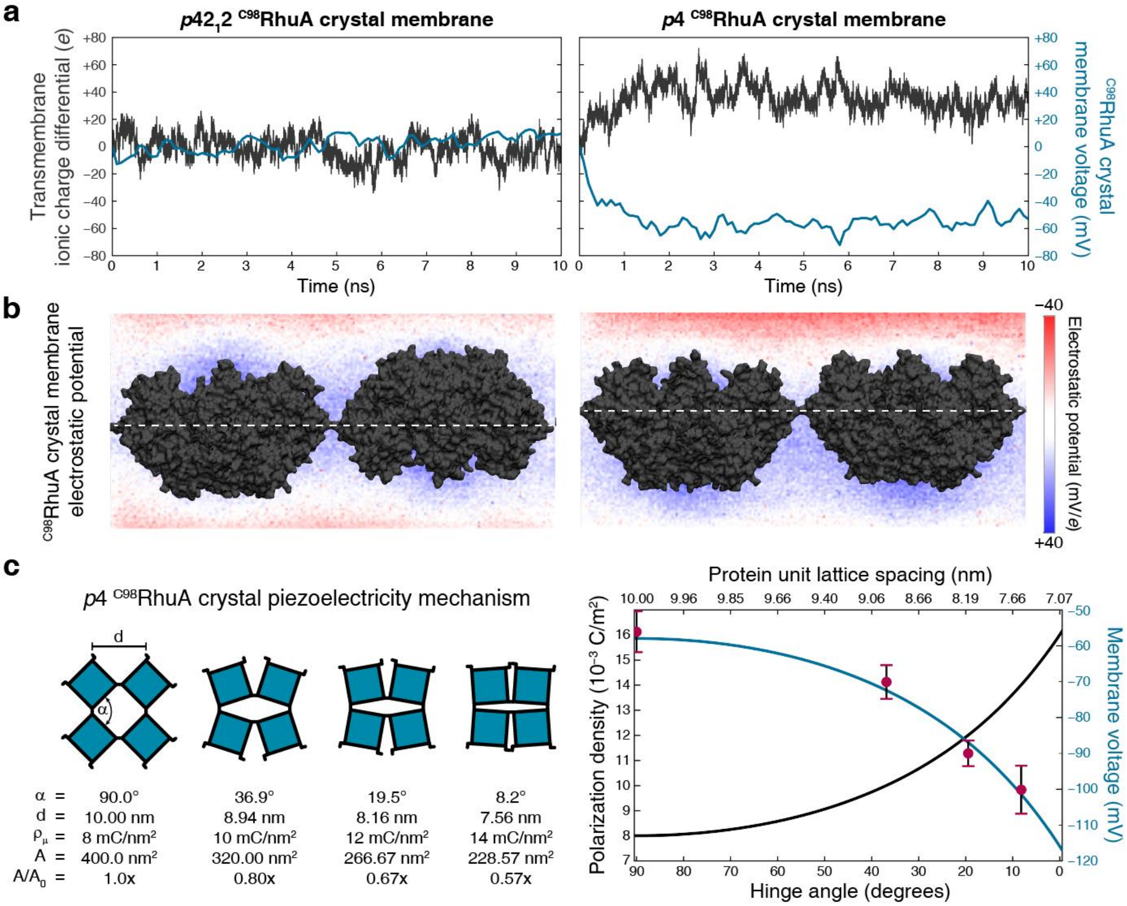


Figure 4.19 | *p*4 ^{C98}RhuA crystals are electrets and piezoelectric materials. a, Net charge imbalance due to ion conduction across the disulfide bonds (half of the simulation box) and resulting electric potential difference across open-state crystals. Dipole-less $p42_12$ crystals induce negligible charge separation, while *p*4 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 *p*4-symmetry ^{C98}RhuA crystals, arising from mechanical coupling between the polarization density (ρ_μ) and lattice conformation (left). The magnitude of this effect was predicted analytically (right) from the ρ_μ over all lattice conformations (as defined by the rotating squares model[7]), while the points reflect numerically determined voltages from all-atom simulations at discrete conformations (**Figure 4.20**). The associated hinge angle (α), lattice spacing (d), ρ_μ, and absolute (A) and relative (A/A_0) 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.

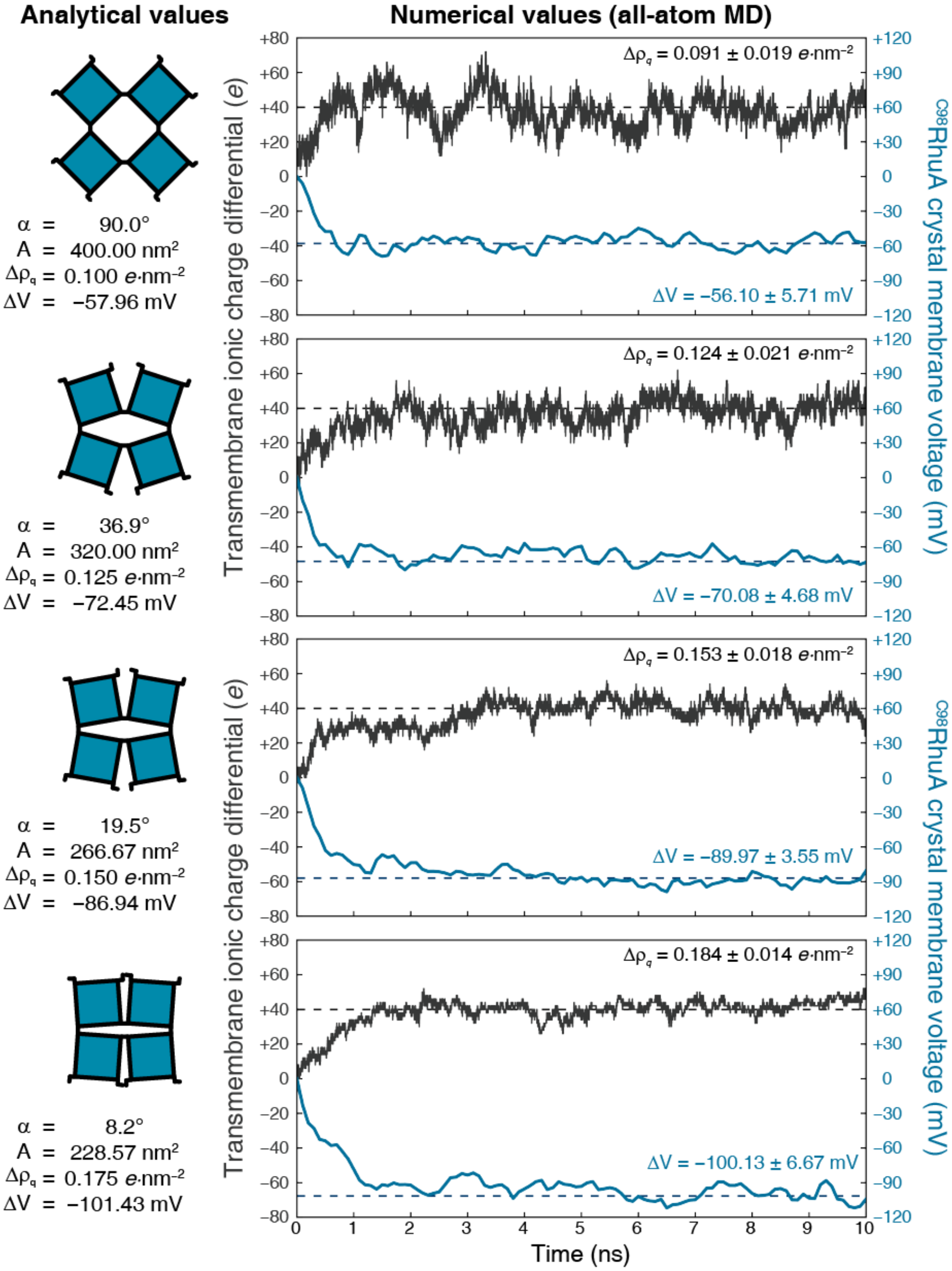


Figure 4.20 | Piezoelectric behavior of *p*4-symmetry ^{C98}RhuA lattices. Results of all-atom MD simulations of *p*4-symmetric ^{C98}RhuA crystals at different conformational states. Cartoons and theoretical values for the hinge angle (α), unit-cell area (A), polarization density (ρ_μ), and electric potential difference (ΔV) are shown at left. The trans-lattice ionic charge differentials (Δq) and resultant voltages (gray and blue traces, respectively) over 10 ns of sampling are shown at right. $\Delta\rho_\mu$ and ΔV values within each plot are the average and s.d. of each quantity over the last 5 ns of simulation. The dashed lines correspond to the idealized analytical value of ΔV for each conformation, and Δq = +40*e* (all conformations).

4.4 Conclusions

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[43-45]. In these contexts, reconfigurable systems like ^{C98}RhuA lattices offer distinct advantages in that the densities and patterns can be facilely altered without the need to engineer new building blocks. Unlike previous work, such as our own that reported the synthesis of singular structures ($p42_12$ crystals) from a single protein[7], the findings presented above demonstrate that ^{C98}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[39,46] and highlight the importance of the relative strengths of interactions over their specific nature in determining the outcome of self-assembly. The results also highlight the capacity of macrodipole moments to influence protein self-assembly; this global scale interaction represents a new addition to the toolkit for designing protein-based materials. Abstracting from the details of the interactions in our specific system, 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) can 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.

4.5 On the approximation of ^{C98}RhuA as a physical dipole

Here we analyze and critically assess our approximation of ^{C98}RhuA proteins as biological nanoparticles possessing internal electric dipoles, which enables the use of rigorous analytical expressions for estimating the magnitude (and thus relative importance) of anisotropic electrostatic interactions in dictating the crystal packing of ^{C98}RhuA monomers within 2D lattices. Though inherently approximate, this approach has been broadly applied with great success to the self-assembly of inorganic nanoparticles[26-28], enabling a quantitative dissection of the energetic drivers which impart morphological selectivity (and diversity), particularly for "patchy" nanoparticles possessing multiple distinct interaction modalities that collectively define the self-assembly free-energy landscape. Indeed, dipole-dipole interactions have been repeatedly identified as the decisive mechanism underpinning the selective formation of 1D chains[28] and antiparallel 2D lattices[26,27] over nonspecific globular aggregates more typical of isotropic particles, and have been exploited to construct crystalline materials with specified physicochemical[40] and electronic properties[41] (*e.g.*, COFs[24], piezoelectrics[45]). Given this extensive precedent for the importance and versatility of dipolar interactions in controlling self-assembly (and emergent materials properties), the present discussion is motivated by the possibility that such effects could be applied to "patchy protein" nanoparticles.

While the influence of dipole moments in proteins has been studied for decades[47-50], these interactions are frequently considered in the context of individual α-helices and have been shown to be relatively localized in space. Consequently, it is essential to establish whether or not the global-scale electric field arising from a protein's folded structure and particular arrangement of charged residues across its surface can be approximated with sufficient accuracy using simple dipoles. It is possible that for many proteins this is not the case, owing to a diffuse distribution of charges on a typical protein's surface. However, the localization of charged residues primarily to the RhuA protein termini surfaces gives rise to a highly polarized atomic structure (and corresponding electric field; **Figure 4.1c**), which is considerably more reminiscent of an electric dipole. Therefore, to validate the use of such an approximation for this system, we compare the predicted electrochemical properties of RhuA crystals—calculated from analytical expressions—directly to results obtained numerically from all-atom MD simulations. As outlined below, we find near-quantitative agreement across all methods. Finally, we note that while the validity of analytical dipole approximations likely holds only for proteins with polarized surfaces, such structures are in principle simple to rationally engineer or design *de novo* into other protein building blocks, facilitating the ability to estimate the magnitude of inter-protein dipolar potentials with reasonable accuracy.

Oriented-dipole electret materials possess a permanent electric field owing to the alignment of polarized units within their structure, and the magnitude of this field will be determined by the magnitude and density of these dipoles. We have reported here trans-lattice ion distributions and calculated electrostatic potential differences across open-state *p*4 and $p42_12$ ^{C98}RhuA crystals (**Figure 4.19a**), as well as for multiple conformations of *p*4 crystals (**Figure 4.19c**, **Figure 4.20**). These results clearly show that the field is present exclusively for the polar *p*4 crystal symmetry, so we next ask whether or not the results are in agreement with expectations based purely on simple dipoles. By doing so, we can critically evaluate whether the magnitude of such dipole-dipole interactions (as calculated from analytical dipoles) are sufficiently accurate to explain the observed antiparallel packing of solution-grown ^{C98}RhuA crystals.

Comparison of expected ion distribution across p4 crystals

To perhaps the most trivial approximation, we can estimate the trans-lattice ion differential expected to arise from dipolar particles simply by converting the 1200 D ^{C98}RhuA macrodipole moment to an equivalent charge × distance "discrete" dipole (0.02082 *e*·nm/D) of 24.983 *e*·nm. Taken to run the 5 nm height of the protein, we obtain our "discrete dipole" approximation of a +5.0*e* and −5.0*e* charge pair separated by a distance of 5 nm. Now applying this to the set of 4 proteins in a 2×2 unit cell which comprise our *in silico* model (**Figure 4.19**), we expect an accumulation of four +5*e* charges on one side and four −5.0*e* charges on the other, for a net ionic charge differential (Δq) of 40*e* across the lattice.

Indeed, we find that for all conformations of *p*4 crystals subjected to all-atom MD simulations, Δq asymptotically approaches 40*e* (**Figure 4.20**), indicating that the predicted distribution of counterions simply equalizes the net charge of these "discrete dipoles", which remains the same regardless of crystal conformation (open vs. closed). We observe in our simulations that as the lattices become more compact, this value becomes more exactly correct (**Figure 4.20**), likely correlating with changes in total pore area through which the ions can exchange freely (suppressing fluctuations when more compact). This analysis suggests that the treatment of ^{C98}RhuA proteins as analytical dipoles accurately approximates the bulk-scale materials properties similarly predicted by expensive all-atom numerical simulations, indicating that RhuA's electric field may be well-represented quite generally even by this extremely simple model.

Comparison to numerically determined potential across p4 crystals

The consequence of maintaining a constant trans-lattice ion differential (*i.e.*, independent of the xy unit cell dimensions) is that the charge density on each side of the crystal must change as the lattice undergoes changes in packing density as a result of in-plane dynamics (**Figure 4.20**). This modulation of the ionic charge density (while the crystal thickness remains constant) necessarily generates corresponding changes to the electrostatic potential difference, and is the fundamental mechanism through which *p*4-symmetry ^{C98}RhuA crystals are predicted to be piezoelectric. With this property directly confirmed by numerical calculations, we can estimate the electrostatic potential expected analytically for a polar crystal through three distinct (though conceptually similar) methods, using the open-state conformation as an example.

1) Analytical potential predicted by trans-lattice ion distribution: Gauss's law enables calculation of the electric field arising from a given distribution of electric charge. In the presence of a static electric field (generated by the polarized crystal), the distribution of mobile solution ions ("free charges") should generate an equivalent field of opposite magnitude at equilibrium. From the "discrete dipoles" (and all-atom MD) we can anticipate $\Delta q = 40e$ for all conformations of a 2×2 unit cell. In the case of an open-state crystal, the cell dimensions are 20×20 nm^2, resulting in a surface charge density of "free charges" $(\sigma_f) = \pm 0.05 \frac{e}{nm^2} = \pm 8.01 \times 10^{-21} \frac{C}{nm^2}$ on each face of the crystal. Such a configuration allows the system to be modeled as an idealized parallel-plate capacitor of surface charge density $\pm\sigma_f$, separated by a distance d = 5 nm (the thickness of a *p*4 RhuA crystal), and mean-field dielectric with permittivity ε. Under these assumptions, the electric field between the plates is constant, and the voltage (from "free charges"; ΔV_f) between the plates can be calculated as $\Delta V_f = \mathbf{E_f} \times d = \frac{\sigma_f}{\varepsilon\varepsilon_0} \times d$. Conservatively assuming a permittivity corresponding to pure water (ε = 78), we obtain the electric field $\mathbf{E_f} = \frac{\sigma_f}{78\varepsilon_0} = 11.60 \frac{mV}{nm}$ and corresponding voltage: $\Delta V_f = \mathbf{E_f} \times 5\ nm = 57.99\ mV$.

2) Analytical potential predicted from bound surface charge density: The generalized formulation of Gauss's law in Maxwell's equations relates the electric displacement field **D** (arising from both "free" and "bound" charges) to the electric field **E** and density of dipole moments **P** (polarization density) within a material as: $\mathbf{D} = \varepsilon\varepsilon_0\mathbf{E} + \mathbf{P}$. Above, we neglected **P** (no "bound charges") to calculate $\mathbf{E}_f$ from the "free charge" displacement field $\mathbf{D}_f$ (= σ_f). Here, we instead consider only the ^{C98}RhuA macrodipole polarization density **P** for a 2×2 open-state crystal (four proteins in a 20×20×5 nm^3 volume) as $\mathbf{P} = \frac{N\mu}{vol.} = \frac{4 \times 1200\, D}{20\times20\times5\, nm^3} \times \frac{3.33564^{-30}\, \frac{C\cdot m}{D}}{10^{-27} \frac{m^3}{nm^3}} = 0.008 \frac{C}{m^2}$. We can then write an expression for the "bound" surface charge density (σ_b) of on each side of the crystal as the dot product of the polarization density and z axis. As all ^{C98}RhuA dipoles within the lattice are oriented perpendicular to the surface normal we obtain the "bound" surface charge density $\sigma_b = \mathbf{P} \cdot \hat{\boldsymbol{z}} = \pm 0.008 \frac{C}{m^2}$ on each face of the crystal, identical to value obtained from the trans-lattice ion distribution (σ_f). We then apply the same parallel-plate capacitor model to calculate the "bound" surface charge potential: $\Delta V_b = \mathbf{E_b} \times d = \frac{\sigma_b}{\varepsilon\varepsilon_0} \times 5\, nm = -57.99\, mV$. The "bound" potential V_b is of equal magnitude to V_f, but opposite sign, due to the opposing orientations of the dipolar bound charge density relative to the free charges (see below).

3) Analytical potential predicted from volumetric polarization density: We can alternatively obtain V_b directly from the volumetric polarization density **P**. Now neglecting free charges (**D** = 0) to calculate the field arising from the "bound" protein dipoles ($\mathbf{E}_b$), we rearrange Gauss's law as $\varepsilon\varepsilon_0\mathbf{E_b} = -\mathbf{P}$. Under the same assumptions as above, the "bound charge" voltage is: $\Delta V_b = \mathbf{E_b} \times d = -\frac{\mathbf{P}}{\varepsilon\varepsilon_0} = -57.99\, mV$.

The very close agreement between the values for Δq and ΔV determined analytically (40*e*, −57.99 mV) and numerically from all-atom MD (37.63 ± 9.22*e*, −56.10 ± 5.71 mV) for open-state *p*4 crystals (**Figure 4.20**) strongly suggests that the predicted macroscale electrochemical properties of ^{C98}RhuA crystals can be effectively estimated from the polarization induced by the oriented macrodipoles of individual RhuA proteins. Importantly, this close agreement is observed for all conformations simulated, demonstrating clear predictive ability from analytical expressions for electric dipoles. We also note that the values provided here conservatively assume a uniform dielectric constant corresponding to pure water. These values could be larger were an effective dielectric constant (*e.g.*, a statistical average of protein and solvent) utilized instead.

4.6 Methods

4.6.1 Protein purification and mutagenesis

^{C98}RhuA was purified as previously described[7]. The plasmid for ^{S98}RhuA was generated via site-directed mutagenesis. Overexpression and purification of ^{S98}RhuA was carried out analogously to ^{C98}RhuA except for the omission of β-mercaptoethanol in the purification buffers. Purified proteins were dialyzed into 20 mM Tris-HCl pH 7.5 and 10 mM reduced L-glutathione (GSH), concentrated to 100-150 μM, flash-frozen in liquid nitrogen, and stored at ≤−60 °C.

4.6.2 Solution self-assembly of ^{C98}RhuA

Crystallization of ^{C98}RhuA was carried out as previously described[7,23], 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 $p42_12$ crystals onto the mica substrate, but also increased the population of stacked 2D crystals. The suspension of $p42_12$ crystals was diluted to 25 μM using

10 mM Tris-HCl (Sigma-Aldrich) pH 7.0 prior to imaging by AFM. 10 μl 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 μl polylysine solution (0.01%; Sigma-Aldrich) on freshly cleaved mica for 1 minute, rinsing with water, and drying under nitrogen prior to use.

4.6.3 Mica-templated self-assembly of ^{C98}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 μM GSH, and desired concentration of KCl, RbCl, $MgCl_2$, or $ZnCl_2$). 100 μl 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, $MgCl_2$, and $ZnCl_2$ were purchased from Sigma-Aldrich, nuclease-free water from Ambion, and muscovite mica from Ted Pella.

4.6.4 Atomic force microscopy

All AFM images of mica-templated and solution-grown ^{C98}RhuA crystals were captured in PeakForce Tapping^{TM} mode on a MultiMode^{TM} 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 100-200 pN, within the typical force range for AFM imaging of biomolecules.[51,52] All offline data processing was done using the SPIP^{TM} software package (Image Metrology, Denmark).

4.6.5 Simulated AFM topographs

AFM topographs were calculated from atomic models with different monomer configurations (extracted from equilibrated 2x2 RhuA crystals; see below) using an *ad hoc* Tcl script executed within VMD[53], 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 Cα atoms was at 0. The probe tip was modeled as a sphere with radius 10.0 Å and center z_{tip} at the end of a cone with half-angle 20° and truncated at z_{cutoff} (**Figure 4.21**). 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_{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, radius) and increasing z_{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 ($z_{min} = z_{tip}$ – radius). This calculation was repeated row-by-row, with the resolution determined by $\Delta x = \Delta y = 0.25$ Å, to generate a 2D array of height values analogous to a true AFM topograph. Output was directly visualized and processed using Gwyddion.

4.6.6 Calculation of the ^{C98}RhuA dimerization potential of mean force

The ^{C98}RhuA dimer structure was prepared from the ^{A88F}RhuA crystal structure (PDB ID: 2UYU) by performing the corresponding D98C and F88A mutations using PyMOL[54]. PSFGEN[53] 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^+ ions (no KCl starting structure) or 2302 K^+ ions and 2270 Cl^- ions (3 M KCl starting structure). The CHARMM36 force field[55] was used for protein atoms and Joung-Cheatham parameters[56] 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 Cα atoms were constrained to their starting positions by a 10 kcal/mol/Å^2 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/Å^2 moving restraint, with the xy coordinates of the Cα 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 **Table 4.1** 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[57]. 100 kcal/mol/deg^2 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 ^{F88}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[58]. PMF simulations were carried out using NAMD 2.12[59].

4.6.7 Molecular modeling and simulation of protein binding to *m*-mica

Equilibration of m-mica: Parameters and initial coordinates for muscovite mica were taken from the INTERFACE force field package[60]. A 5×3 *m*-mica supercell with Al substitutions in agreement with ^{29}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[61] 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^+ ions. These ions were free to exchange with the solution, and at equilibrium 80-85% of sites were occupied by K^+ ions. This state was also attained if the system was initiated with a bare (K^+ occupancy = 0) *m*-mica surface and the same numbers of ions in solution, confirming that it reflects thermodynamic equilibrium.

Equilibration of ^{C98}RhuA: A pre-equilibrated ^{C98}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^+ ions. An additional 3463 K^+ ions and 3463 Cl^- 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.

Simulated binding of ^{C98}RhuA to the m-mica surface: The equilibrated protein system was merged with the equilibrated mica layers into a single structure using TopoTools such that the lowest position of protein Cα 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^- ions were added to the solution to neutralize the system, bringing the solution ion concentrations to approximately 2.67 M K^+ and 3.33 M Cl^-, close to experimental conditions. Solution-exposed K^+ 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^+ 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/Å^2) to their initial positions to prevent drift of the slab. The protein was restrained (by its Cα 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 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 ^{C98}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/Å^2 moving restraint, such that the minimum Cα position was at $z \approx 5$ Å.

Equilibration of ^{C98}RhuA in proximity to an m-mica surface: 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 Cα COM z positions of diagonal protein chains (A-C and B-D) at 0 Å using two 100 kcal/mol 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.

4.6.8 APBS calculations

Continuum electrostatics calculations of the ^{C98}RhuA protein (appropriately protonated using PROPKA 3.1[62]) were carried out using APBS 1.5[63] to reflect experimental solution self-assembly conditions (pH 7.5, [ions] = 20 mM, $\varepsilon_{\text{solvent}} = 78$, $\varepsilon_{\text{protein}} = 2$).

4.6.9 Equilibration of 3D periodic crystals

Open-state 2×2 ^{C98}RhuA crystals were constructed as previously described[23] for both *p*4 and $p42_12$ 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^+ 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/protein restraints were employed to maintain the difference in Cα 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 *p*4 crystal (extracted after 1 ns of equilibration) were manipulated to generate all other *p*4 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 $^{F88/C98}$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.

4.6.10 Calculation of the ^{C98}RhuA macrodipole moment

The dipole moment of ^{C98}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[10] using PDB ID: 1OJR, which yielded a value of 310 D per protein subunit (1240 D total). We used 1200 D for all dipole potential energy calculations as a slightly more conservative (and reflective of the equilibrated protein structure) estimate.

4.6.11 Electret/piezoelectric simulations

The last frame of the equilibrated $p4$ and $p42_12$ 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 Cα atoms were constrained to their starting z position using a

100 kcal/mol 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 = 200.00 Å for *p*42_12 lattices, and spanning 151.19–200.00 Å for different *p*4 lattice conformations; see **Table 4.2** for details), with full-system electrostatics calculated using the Multiscale Summation Method (MSM)[64] 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 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.

4.6.12 Disulfide dihedral energy calculations

Disulfide dihedral angles (as depicted in **Figure 4.3**) for the non-periodic disulfide bonds of *p*4 and *p*42_12 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.

4.6.13 Numerical ^{C98}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[66] plugin in VMD for all 10 ns of sampling (10,000 configurations), split into 100-frame blocks. The maps were calculated over a×b×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 appx. 100 Å air gap between periodic images. An Ewald factor of 0.25 $Å^{-1}$ 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 ($\varepsilon = 78$), and multiplied by the factor 25.87 mV/k_BT for a 300 K simulation. The 5–10 ns time-averaged membrane potentials for *p*4 and $p42_12$ crystals were −56.09 ± 5.71 mV and −3.09 ± 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 *p*4 crystals based purely on the density of dipoles within the lattice (**Section 4.5**), strongly suggesting that the macrodipole moment of RhuA crystals can indeed account for their predicted bulk properties.

4.6.14 Pairwise ^{C98}RhuA nanoparticle electrostatic potentials

The electrostatic potential between individual ^{C98}RhuA proteins was computed using the pairwise potential between two charged nanoparticles (U_{ij}) proposed by Phillies[29] for dilute solutions of proteins and charged colloids (**Equation 4.1**), as employed previously for modelling nanoparticle and protein interactions[26,28,67].

$$U_{ij}\left(r_{ij},\theta_{i,j},\varphi_{i,j}\right) = k_e\frac{q_iq_j}{r_{ij}}e^{-\kappa r_{ij}}C_0^2 + k_e\frac{q_i\mu_j\cos\theta_j + q_j\mu_i\cos\theta_i}{r_{ij}^2}e^{-\kappa r_{ij}}C_0C_1 +$$

$$k_e\frac{\mu_i\mu_j}{r_{ij}^3}\{\cos\theta_j\cos\theta_j\left[2+\kappa r_{ij}+(\kappa r_{ij})^2\right]+\sin\theta_i\sin\theta_j\cos(\varphi_i-\varphi_j)\left[1+\kappa r_{ij}\right]\}e^{-\kappa r_{ij}}C_1^2 \quad (4.1)$$

Where C_0 and C are given by:

$$C_0 = \frac{e^{\kappa a}}{1+\kappa a} \quad (4.2) \qquad\qquad C_1 = \frac{3e^{\kappa a}}{2+2\kappa a+(\kappa a)^2+\frac{(1+\kappa a)}{\varepsilon}} \quad (4.3)$$

Here, q_i and q_j are the net charge of each ^{C98}RhuA protein (−16*e*), μ_i and μ_j are the dipole moments of each ^{C98}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 $(4\pi\varepsilon_0\varepsilon)^{-1}$; ε_0 the permittivity of vacuum; ε the relative solvent permittivity (78); and κ^{-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 ^{C98}RhuA crystals enables Equation 1 to be simplified considerably. It geometrically restricts the dipole moment of each particle pair to be oriented normal to the lattice (so $\theta = 90°$), and $\Delta\varphi_{ij} = 0$ ($p4$ symmetry) or 180 ($p42_12$ symmetry). We can then reduce all sin(θ) terms to unity, and nullify all cos(θ) terms (comprising the entire charge-dipole term and first half of the dipole-dipole term). This gives us the reduced form of U_{ij} (**Equation 4.4**):

$$U_{ij}(r_{ij}, \Delta\varphi_{ij}) = k_e \frac{q_i q_j}{r_{ij}} e^{-\kappa r_{ij}} C_0^2 + k_e \frac{\mu_i \mu_j}{r_{ij}^3} \cos(\Delta\varphi_{ij}) \left[1 + \kappa r_{ij}\right] e^{-\kappa r_{ij}} C_1^2 \qquad (4.4)$$

We can see that the first term is purely repulsive due to the identical net charge on both ^{C98}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 $p42_12$ 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 (**Figure 4.4**). This illustrates the origin of the thermodynamic selectivity for $p42_12$ lattices in solution (**Figure 4.5**).

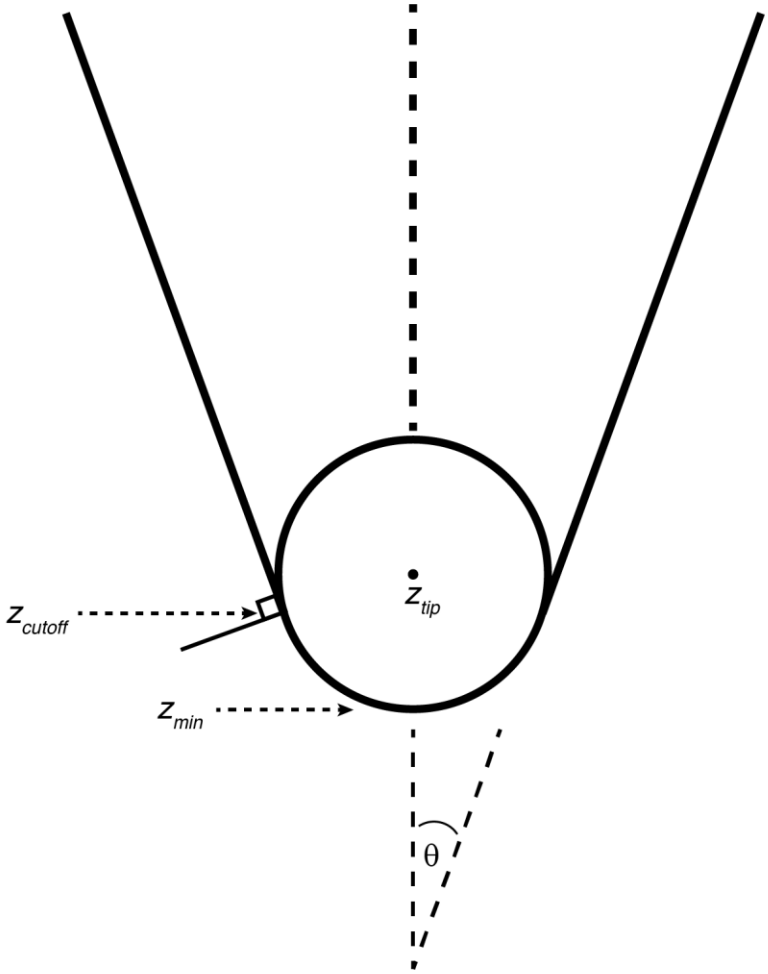


Figure 4.21 | Description of tip convolution simulations. The effect of tip convolution was modeled by looking for overlaps between protein atoms and a tip with the above geometry. Checks were performed by first evaluating the volume encompassed by a sphere centered at z_{tip}, then testing for any overlaps within a cone of half-angle θ which lie tangent to the spherical tip. Upon finding a position at which the number of overlaps met the cutoff criterion, the final height was reported as z_{min}. This effectively captures the coarsening of atomic structure observed during AFM experiments.

Table 4.1 | Umbrella sampling details. Window centers and force constants used for the calculation of the ^{C98}RhuA dimerization PMFs.

no KCl	
Window (Å)	k_f (kcal/mol/Å^2)
46.5	2.0
48.5	2.0
50.5	2.0
52.5	2.0
54.5	2.0
56.5	2.0
58.5	2.0
60.5	2.0

3 M KCl	
Window (Å)	k_f (kcal/mol/Å^2)
46.5	2.0
47.5	4.0
48.5	2.0
50.5	2.0
52.5	2.0
53.0	4.0
54.5	2.0
56.5	2.0
58.5	2.0
60.5	2.0

Table 4.2 | Piezoelectric membrane simulation details. Information regarding the symmetry, crystal conformation (as hinge angle), periodic box dimensions (x = y and are fixed values, z fluctuates), and total number of ions for all simulations used for the piezoelectricity response analysis. Protein center-center distances between disulfide-bonded dimers are equal to half of the periodic box xy dimensions (*e.g.*, 90° angle disulfide-bonded dimers are separated by 100.00 Å).

Symmetry	**Hinge angle (α)**	**PBC xy dims. (Å)**	**Appx. z dim. (Å)**	**# Na$^+$ ions**	**# Cl$^-$ ions**
$p4$	8.2	151.186	101.0	212	212
$p4$	19.5	163.299	102.0	264	264
$p4$	36.9	178.885	103.3	336	336
$p4$	90.0	200.00	104.5	448	448
$p42_12$	90.0	200.00	104.5	448	448

4.7 Acknowledgments

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.

Chapter 4, in part, has been submitted for publication: Zhang, S.*, Alberstein, R.G.*, De Yoreo, J.J., Tezcan, F.A. “Assembly of a patchy protein into variable 2D lattices via tunable, multiscale interactions” (2020). The dissertation author was the co-primary investigator and author of this manuscript.

4.8 References

1. Marsh, J. A. & Teichmann, S. A. Structure, Dynamics, Assembly, and Evolution of Protein Complexes. *Annu. Rev. Biochem.* **84**, 551-575, (2015).

2. 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).

3. Thorkelsson, K., Bai, P. & Xu, T. Self-assembly and applications of anisotropic nanomaterials: A review. *Nano Today* **10**, 48-66, (2015).

4. Glotzer, S. C. & Solomon, M. J. Anisotropy of building blocks and their assembly into complex structures. *Nature* **6**, 557-562, (2007).

5. Gröschel, A. H., Walther, A., Löbling, T. I., Schacher, F. H., Schmalz, H. & Müller, A. H. E. Guided hierarchical co-assembly of soft patchy nanoparticles. *Nature* **503**, 247-251, (2013).

6. Brodin, J. D., Ambroggio, X. I., Tang, C. Y., Parent, K. N., Baker, T. S. & Tezcan, F. A. Metal-directed, chemically tunable assembly of one-, two- and three-dimensional crystalline protein arrays. *Nat. Chem.* **4**, 375–382, (2012).

7. Suzuki, Y., Cardone, G., Restrepo, D., Zavattieri, P. D., Baker, T. S. & Tezcan, F. A. Self-assembly of coherently dynamic, auxetic, two-dimensional protein crystals. *Nature* **533**, 369–373, (2016).

8. Yeates, T. O. Geometric Principles for Designing Highly Symmetric Self-Assembling Protein Nanomaterials. *Annu. Rev. Biophys.* **46**, 23-42, (2017).

9. Kostiainen, M. A., Hiekkataipale, P., Laiho, A., Lemieux, V., Seitsonen, J., Ruokolainen, J. & Ceci, P. Electrostatic assembly of binary nanoparticle superlattices using protein cages. *Nat. Nanotech.* **8**, 52-56, (2013).

10. 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).

11. Simon, A. J., Zhou, Y., Ramasubramani, V., Glaser, J., Pothukuchy, A., Gollihar, J., Gerberich, J. C., Leggere, J. C., Morrow, B. R., Jung, C., Glotzer, S. C., Taylor, D. W. & Ellington, A. D. Supercharging enables organized assembly of synthetic biomolecules. *Nat. Chem.* **11**, 204-212, (2019).

12. Yuan, Y., Tam, M. F., Simplaceanu, V. & Ho, C. New Look at Hemoglobin Allostery. *Chem. Rev.* **115**, 1702-1724, (2015).

13. Shin, S.-H., Chung, S., Sanii, B., Comolli, L. R., Bertozzi, C. R. & De Yoreo, J. J. 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).

14. Engel, A. & Müller, D. J. Observing single biomolecules at work with the atomic force microscope. *Nat. Struct. Mol. Biol.* **7**, 715-718, (2000).

15. Jaswal, S. S., Sohl, J. L., Davis, J. H. & Agard, D. A. Energetic landscape of α-lytic protease optimizes longevity through kinetic stability. *Nature* **415**, 343-346, (2002).

16. Sleytr, U. B., Schuster, B., Egelseer, E. M. & Pum, D. S-layers: principles and applications. *FEMS Microbiol. Rev.*, 823–864, (2014).

17. Koga, N., Tatsumi-Koga, R., Liu, G., Xiao, R., Acton, T. B., Montelione, G. T. & Baker, D. Principles for designing ideal protein structures. *Nature* **491**, 222-227, (2012).

18. Churchfield, L. A. & Tezcan, F. A. Design and Construction of Functional Supramolecular Metalloprotein Assemblies. *Acc. Chem. Res.* **52**, 345-355, (2019).

19. Ringler, P. & Schulz, G. E. Self-assembly of proteins into designed networks. *Science* **302**, 106–109, (2003).

20. Bale, J. B., Gonen, S., Liu, Y., Sheffler, W., Ellis, D., Thomas, C., Cascio, D., Yeates, T. O., Gonen, T., King, N. P. & Baker, D. Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science* **353**, 389-394, (2016).

21. 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).

22. 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).

23. 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., Chen, C., Li, H., Diercks, C. S., Wang, G., Waller, P. J., Li, H., Bredas, J.-L., Yaghi, O. M. & Crommie, M. F. Local Electronic Structure of Molecular Heterojunctions in a Single-Layer 2D Covalent Organic Framework. **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. Bourg, 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).

31. Ricci, M., Spijker, P. & Voïtchovsky, K. Water-induced correlation between single ions imaged at the solid–liquid interface. *Nat. Comm.* **5**, 4400, (2014).

32. 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).

33. Ido, S., Kimiya, H., Kobayashi, K., Kominami, H., Matsushige, K. & Yamada, H. Immunoactive two-dimensional self-assembly of monoclonal antibodies in aqueous solution revealed by atomic force microscopy. *Nat. Mater.* **13**, 264-270, (2014).

34. Pintea, S., de Poel, W., de Jong, A. E. F., Vonk, V., van der Asdonk, P., Drnec, J., Balmes, O., Isern, H., Dufrane, T., Felici, R. & Vlieg, E. Solid–Liquid Interface Structure of Muscovite Mica in CsCl and RbBr Solutions. *Langmuir* **32**, 12955-12965, (2016).

35. Woo, S. & Rothemund, P. W. K. Self-assembly of two-dimensional DNA origami lattices using cation-controlled surface diffusion. *Nat. Commun.* **5**, 4889, (2014).

36. Tao, J., Fijneman, A., Wan, J., Prajapati, S., Mukherjee, K., Fernandez-Martinez, A., Moradian-Oldak, J. & De Yoreo, J. J. Control of Calcium Phosphate Nucleation and Transformation through Interactions of Enamelin and Amelogenin Exhibits the "Goldilocks Effect". *Cryst. Growth. Des.* **18**, 7391-7400, (2018).

37. De Yoreo, J. J., Gilbert, P. U. P. A., Sommerdijk, N. A. J. M., Penn, R. L., Whitelam, S., Joester, D., Zhang, H., Rimer, J. D., Navrotsky, A., Banfield, J. F., Wallace, A. F., Michel, F. M., Meldrum, F. C., Cölfen, H. & Dove, P. M. Crystallization by particle attachment in synthetic, biogenic, and geologic environments. *Science* **349**, aaa6760, (2015).

38. Vekilov, P. G. Two-step mechanism for the nucleation of crystals from solution. *J. Cryst. Growth.* **275**, 65-76, (2005).

39. Whitelam, S. Control of Pathways and Yields of Protein Crystallization through the Interplay of Nonspecific and Specific Attractions. *Phys. Rev. Lett.* **105**, 088102, (2010).

40. Dandekar, P., Kuvadia, Z. B. & Doherty, M. F. Engineering Crystal Morphology. *Annu. Rev. Mater. Res.* **43**, 359-386, (2013).

41. Chung, H. & Diao, Y. Polymorphism as an emerging design strategy for high performance organic electronics. *J. Mater. Chem. C* **4**, 3915-3933, (2016).

42. Hillenbrand, J. & Sessler, G. M. Piezoelectricity in cellular electret films. *IEEE Trans. Dielectr. Electr. Insul.* **7**, 537-542, (2000).

43. Moll, D., Huber, C., Schlegel, B., Pum, D., Sleytr, U. B. & Sára, M. S-layer-streptavidin fusion proteins as template for nanopatterned molecular arrays. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14646-14651, (2002).

44. Mann, S., Archibald, D. D., Didymus, J. M., Douglas, T., Heywood, B. R., Meldrum, F. C. & Reeves, N. J. Crystallization at Inorganic-organic Interfaces: Biominerals and Biomimetic Synthesis. *Science* **261**, 1286-1292, (1993).

45. Hinchet, R., Khan, U., Falconi, C. & Kim, S.-W. Piezoelectric properties in two-dimensional materials: Simulations and experiments. *Mater. Today* **21**, 611-630, (2018).

46. Haxton, T. K. & Whitelam, S. Do hierarchical structures assemble best via hierarchical pathways? *Soft Matter* **9**, 6851-6861, (2013).

47. Brant, D. A. & Flory, P. J. The Role of Dipole Interactions in Determining Polypeptide Configurations. *J. Am. Chem. Sci.* **87**, 663-664, (1965).

48. Gilson, M. K., Rashin, A., Fine, R. & Honig, B. On the calculation of electrostatic interactions in proteins. *J. Mol. Biol.* **184**, 503-516, (1985).

49. Gilson, M. K. & Honig, B. Destabilization of an alpha-helix-bundle protein by helix dipoles. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 1524, (1989).

50. Lockhart, D. & Kim, P. Electrostatic screening of charge and dipole interactions with the helix backbone. *Science* **260**, 198-202, (1993).

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. Schrodinger LLC. *The PyMOL Molecular Graphics System, Version 1.3,* (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., Rosenberg, 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., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L. & Schulten, K. 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. Kohlmeyer, A. & Vermaas, J. TopoTools. (2017).

62. Olsson, M. H. M., Søndergaard, 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., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L. E., Brookes, D. H., Wilson, L., Chen, J., Liles, K., Chun, M., Li, P., Gohara, D. W., Dolinsky, T., Konecny, R., Koes, D. R., Nielsen, J. E., Head-Gordon, T., Geng, W., Krasny, R., Wei, G.-W., Holst, M. J., McCammon, J. A. & Baker, N. A. Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* **27**, 112-128, (2018).

64. Hardy, D. J., Wu, Z., Phillips, J. C., Stone, J. E., Skeel, R. D. & Schulten, K. 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 α-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).

Chapter 5: Investigating the molecular basis of surface patterning of biomolecules at aqueous mineral interfaces

5.1 Abstract

Proteins are necessarily sensitive to physical or chemical changes within their local environment, which enables the regulation of their behavior in a statistical fashion through the imposition of thermodynamic flows (*e.g.*, within the constant flux of nonequilibrium environments of cells). By extension, this sensitivity can be exploited through the use of chemical or physical templates to generate anisotropic environments that, in turn, direct the formation of specified structures (*i.e.*, biomineralization processes). For synthetic chemists and materials scientists, this exquisite sensitivity represents a tremendous opportunity for the construction of highly tunable molecular components and architectures, yet our comprehension of how the subtleties of such molecular scale potentials enable these ends, let alone how to produce them ourselves, is often lacking. Here we report theoretical interrogations of the solid-liquid interfaces of muscovite and fluorophlogopite micas which, despite possessing nearly identical structural and chemical compositions, impose characteristic orientational biases to surface-adsorbed proteins with almost perfect universality via unknown mechanism. Employed synthetically, such a mechanism could afford unprecedented control over protein self-assembly, given that the energetic drivers which comprise its function were fully dissected. Toward this end, we derived classical force-field parameters describing fluorophlogopite for characterization by all-atom MD which, upon combination with a novel mean-field approximation for these aluminosilicate substrates, enabled us to identify and quantify mineral-dependent local structuring of the interfacial solvent that, we propose, is the mechanism by which they encode chemical specificity to adsorbed proteins.

5.2 Introduction

The multiscale organization of matter to achieve complex functions is a hallmark of living systems, whereby (non)equilibrium thermodynamic driving forces are exploited to initiate and direct the "bottom-up" construction of biomolecular materials from smaller, individually tailorable building blocks, namely proteins[1]. Such modularity provides tremendous advantages, as it affords detailed control over the physical, chemical, and mechanical properties of the resultant supramolecular structure through (often subtle) modifications to the subcomponents which comprise it[2,3]. Many biological materials capitalize upon this structural hierarchy by using "lower-level" assemblies as scaffolds to template the nucleation and growth of higher-order architectures through modulation of the free-energy landscapes underpinning their self-assembly. Moreover, these templates need not be biological in nature, but can be physical boundaries such as air-water interfaces (*e.g.*, hydrophobins[4]), polymeric matrices (*e.g.*, biofilms[5]), and inorganic surfaces (*e.g.*, the anchoring of proteins to the mineral matrix of bones[6]), where the roles of the bio/inorganic components may also be reversed (*e.g.*, biomineralization processes[7,8]). Clearly, the ability of proteins (and other biomolecules) to self-assemble at interfaces greatly diversifies the range of structural outcomes, however the physical mechanisms which effect/modulate assembly processes at protein-inorganic interfaces remains nebulous, obscured by the chemical complexity of bioinorganic heterostructures[9].

There are, however, some patterning effects which appear agnostic to complexity of the substance being crystallized, such as epitaxy. This phenomenon, whereby a crystalline substrate promotes the nucleation and growth of surface-adsorbed species into new crystalline layers, is a direct consequence of the periodicity of the underlying substrate, which provides a well-defined pattern that can direct atomic/molecular organization even through weak nonspecific interactions (*e.g.*, physisorption)[10]. For this reason, epitaxial growth is widely utilized in the manufacture of semiconductors[11], where epitaxial films are grown atop semiconductor substrates in order to produce layered structures all sharing a common crystallographic orientation, which is essential to achieve the desired electrical properties[11,12]. Of course, the atomic/molecular constituents of these films and their substrates are often far simpler than proteins (and thus much easier to rationalize), however this phenomenon is no less common for macromolecules and/or complex interfaces. Indeed, several examples of proteins organizing into distinctly patterned arrangements on solid substrates have been reported[13-17], demonstrating a dependence (at a minimum) on the specific protein/surface pairing, as well as the presence of various ions[17,18].

Consequently, the predictive design and construction of new protein-based bioinorganic materials with unique, adaptive, and/or specialized properties (*e.g.*, collagen, which forms bone, tendon, and cartilage as a function of mineral content) will necessitate a similarly comprehensive understanding of the various factors that modulate heterogeneous self-assembly (*e.g.*, solvent, ionic strength/composition, pH, protein-protein, and protein-surface interactions). Here we describe initial efforts to identify and characterize these forces as they apply to the heterogeneous crystallization of a “simple” protein building block on hydrated clay surfaces, and propose general principles for controlling the epitaxial growth of 2D protein assemblies at crystalline interfaces.

5.3 Results and Discussion

5.3.1 Experimentally observed epitaxial alignment of proteins on mica substrates

As our surface template, we selected the layered phyllosilicate minerals muscovite mica (*m*-mica) and fluorophlogopite (F-phlogopite) for computational investigation. These substrates are commonly utilized for the imaging of biomolecules by atomic force microscopy (AFM), as their weak interlayer interactions endows them with nearly perfect basal cleavage, enabling the facile separation of layers to yield an "atomically flat" surface suitable for topographical measurements. Both faces of each layer are defined by a pseudo-hexagonal aluminosilicate structure (bridged by oxygen atoms) with a statistical 1:3 Al:Si composition, creating a −1 charge per unit cell. In the bulk phase, the vacancies of this honeycomb surface structure are occupied by a hexagonal arrangement K^+ ions, which simultaneously neutralizes the net charge and mediates stacking of the layers. What distinguishes these two minerals is their internal structure: muscovite (a dioctahedral sheet silicate) contains a network of trivalent Al^{3+} cations bridged by OH^- anions, while fluorophlogopite (a trioctahedral sheet silicate) instead contains divalent Mg^{2+} cations that are triply coordinated to F^- anions (or OH^- in the non-fluorine-substituted phlogopite). The OH groups of each *m*-mica layer are oriented along one of hexagonal lattice vectors, and angled almost co-planar to the surface. The *m*-mica unit cell contains two such layers, so only 2/3 possible OH orientations exist within the bulk mineral. In contrast, the symmetric coordination by Mg^{2+} in phlogopite orients the bridging (F,OH) groups along the surface normal (**Figure 5.1**). As a result, while both are monoclinic crystal systems (*m*-mica: *C*2/c, phlogopite: *C*2/m), the anisotropic OH orientations and octahedral cation vacancies in *m*-mica break its superficially symmetric structure, while (F,OH)-phlogopite exhibits near-perfect hexagonal periodicity.

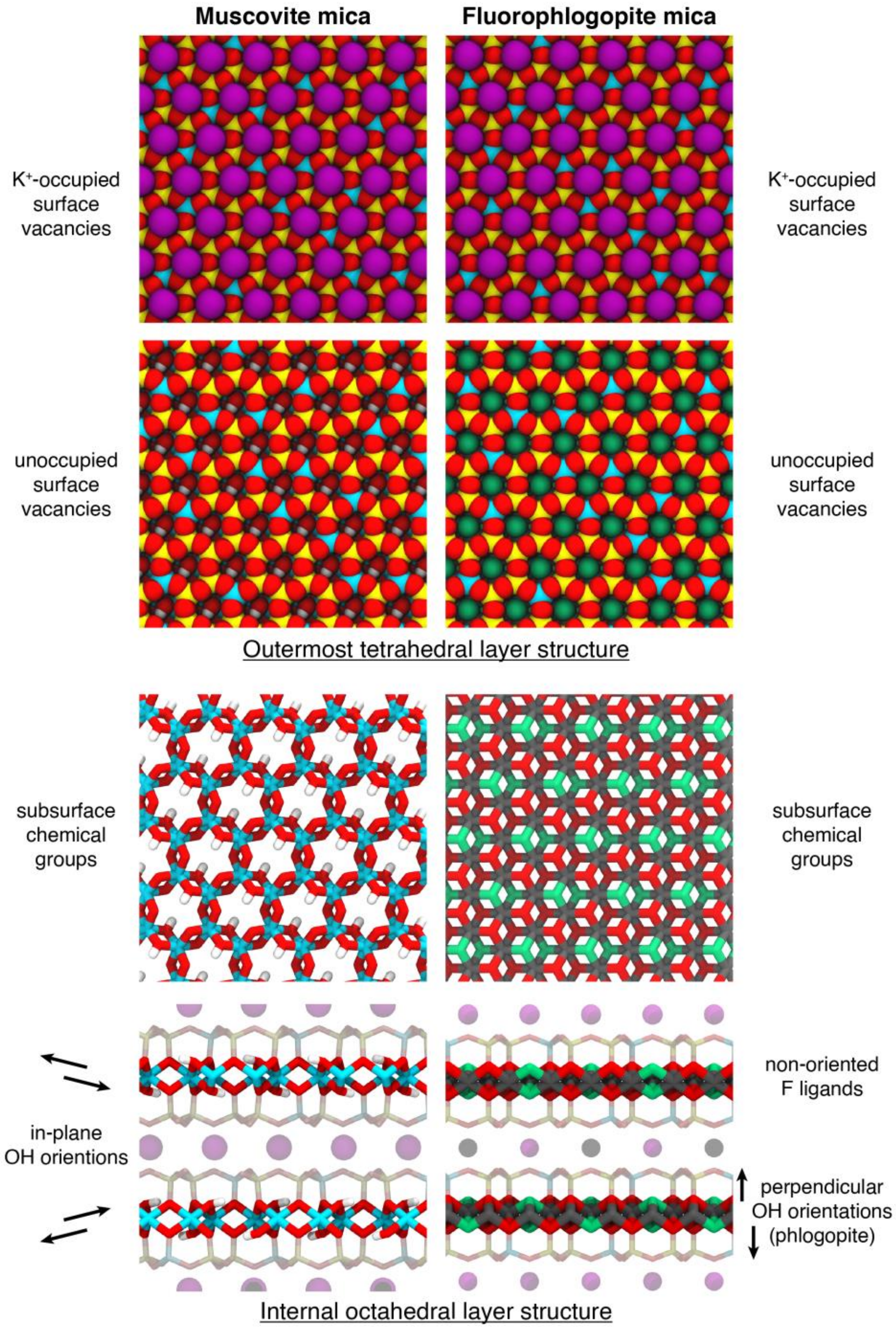


Figure 5.1 | Structural similarities between muscovite and (fluoro)phlogopite micas. The outermost tetrahedral layers of both minerals are structurally and chemically identical (**top**). Within the octahedral layers, they are distinguished primarily by their arrangements of metal ions, with *m*-mica "missing" an Al atom, producing the void into which the OH groups are oriented. Phlogopite instead has a continuous octahedral coordination network, where the isotropic coordination of the bridging ligand position (F shown for the purposes of this study, but is the O_{OH} atom in the unsubstituted mineral), causes it to orient normal to the lattice plane.

Due to the routine use of these substrates (*m*-mica in particular) for biomolecular imaging, the preferential alignment of adsorbed proteins (both natural and designed) has been frequently noted. This is most apparent under conditions which induce surface crystallization of the protein, which typically leads to the formation of small crystalline domains aligned along particular directions (typically the lattice vectors defined by the surface vacancies/K^+ sites), reflecting the orientational anisotropy of the underlying lattice (*i.e.*, heteroepitaxial growth). Furthermore, the distribution of protein/crystal orientations has been shown to depend primarily on the identity of the mineral (*e.g.*, muscovite vs phlogopite)[13,16,17], though it can be further modulated by solution conditions such as electrolyte and/or protein concentration. Surprisingly, this indicates that, despite their extremely similar surface structures and chemistry, protein adsorption and self-assembly processes discriminate between *m*-mica and (F,OH)-phlogopite, with the former only directing growth along one or two lattice vectors, in contrast to all three for the latter. In fact, a recently reported synthetic protein, computationally designed to form geometrically matched interactions between its carboxylate sidechains and the hexagonal surface layer of K^+ ions (*i.e.*, without consideration for the subsurface structure), still exhibited this exact mineral-dependence orientational distribution[16]. This dependence can span all the way to the other end of the design spectrum: the protein ^{C98}RhuA (which self-assembles into 2D lattices via disulfide-bonding, but is otherwise entirely non-self-interacting) was recently shown to crystallize into a variety of distinct morphologies at *m*-mica interfaces via modulation of its self-assembly by the surface template[17]. A clear preference for the growing domains to orient along 1 or 2 of the m-mica lattice vectors was observed for all morphologies, and was sufficiently powerful as to play an important role in enable non-classical nucleation and growth of these domains. We now report previously unpublished data of ^{C98}RhuA assembly on F-phlogopite, which exhibits the exact same differentiability as seen for other proteins (**Figure 5.2**). Thus, despite the seemingly subtle differences between the two silicates, the effect on protein-mica epitaxial alignment appears to be quite general (*i.e.*, independent of protein sequence and structure), and must somehow reflect an effect of remote long-range interactions from deep within the template. We thus set out to perform a quantitative dissection of the structural and energetic principles which give rise to this phenomenon, such that they can be predictably exploited or sidestepped as needed to achieve specific structures and functions.

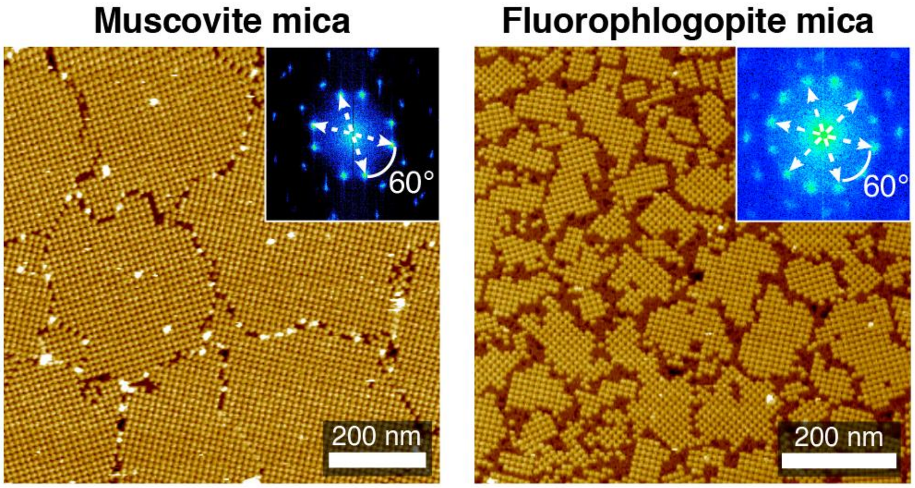


Figure 5.2 | Differential crystal growth behavior of ^{C98}RhuA between *m*-mica and F-phlogopite templates. ^{C98}RhuA was consistently found to preferentially form domains along 2 (occasionally one) of the *m*-mica lattice vectors over the course of extensive characterization. Yet, when incubated with F-phlogopite under the same assembly conditions, equal growth along all three directions was observed instead, unambiguously demonstrating the explicit influence of the underlying substrate on self-assembly.

1.3.2 The simple part: interfacial electrostatics influence protein-surface interactions

Numerous hypotheses have been put forth to rationalize these unambiguous differences, spanning from the highly specific (*e.g.*, the precisely designed lattice matched interactions discussed above[16]), to the more chemically intuitive (*e.g.*, general protein-surface charge complementarity[17]), and less (*e.g.*, the attribution of supramolecular-scale effects to specific individual sidechain interactions[14]). Consequently, proteins such as ^{C98}RhuA are advantageous as model systems to interrogate these predictions, as the certainty afforded by the use of “simple”, strong, flexible, and well-defined interaction motifs (*e.g.*, disulfide bonding) allows their effects to be facilely accounted for and integrated out, laying bare any subtle template-induced effects on assembly that could otherwise be obscured by nonspecific (and/or energetically overpowering) protein-protein interactions. To briefly preface to the remainder of this work, we previously determined that the most consistent method of control over ^{C98}RhuA protein-surface interactions was through varying the solution concentrations of electrolytes (*e.g.*, KCl and RbCl), and thus the surface coverage of ions, which in turn modulates the strength and specific nature of surface-based electrostatic interactions (the fundamental force common to all of the described hypotheses). To semi-quantitatively depict how ion adsorption might manifest distinct interactions, we performed short (<10 ns) simulations of ^{C98}RhuA adsorbed to *m*-mica at 100 mM and 3 M KCl (**Figure 5.3**). The protein orientation in each condition reflects experiment, which we supplemented with computational quantification of the local surface charge and atomic structure of the interface.

High concentrations of K^+ and Cl^- were found to form an ionic double layer at the surface, reversing its charge to positive values and preventing the formation of direct protein-surface contacts (**Figure 5.3**). Conversely, the surface retained its negative charge at low [KCl] concentrations, and the incomplete occupancy of the surface vacancies were indeed explored by amine sidechains (previously singled out as a possible mediator of epitaxial alignment[14]). In short, elements from many different proposed mechanisms may all be correct to some extent. Addressing this complexity motivated the simultaneous investigation of *m*-mica and F-phlogopite reported here, whose close chemical relationship, yet clearly distinguishable behavior, promised to provide essential new context to this longstanding puzzle.

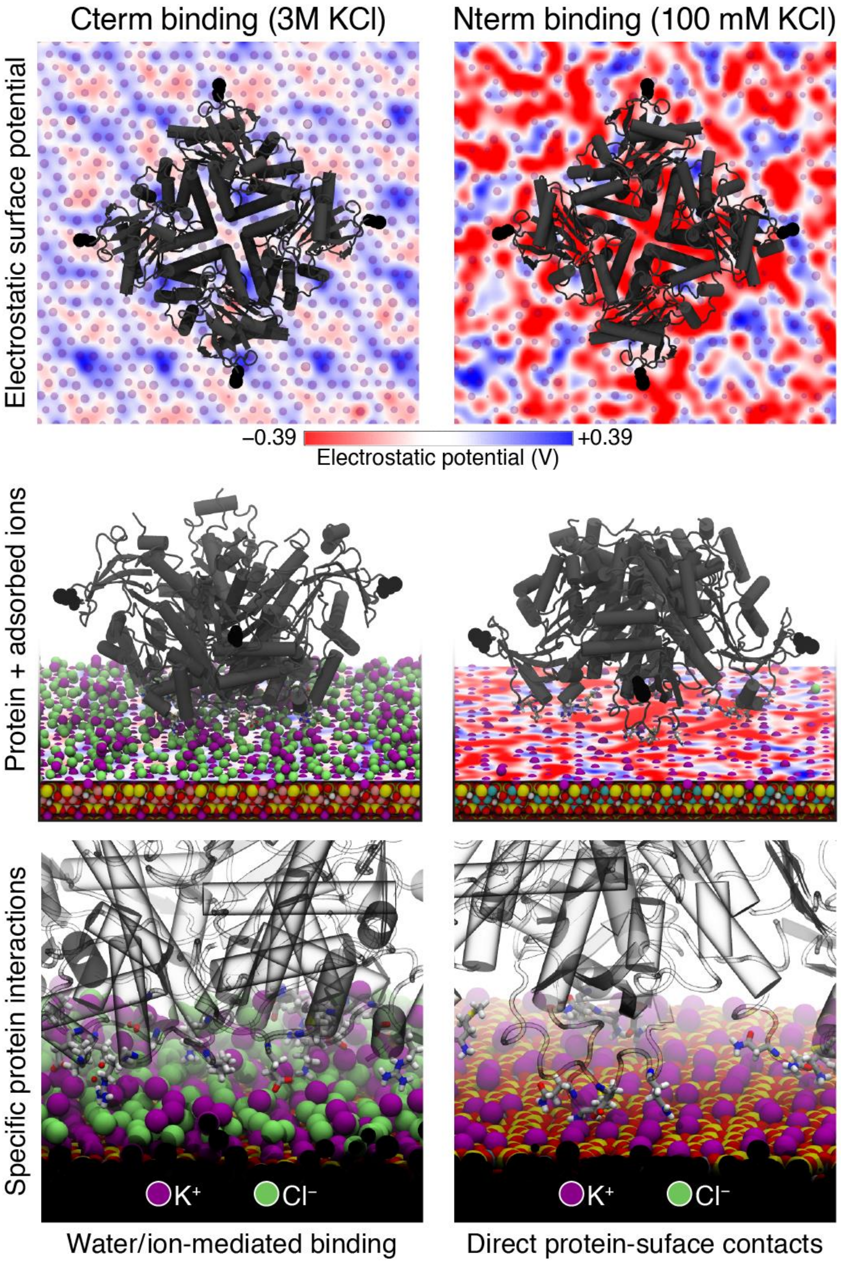


Figure 5.3 | Interfacial ion adsorption dictates both surface charge and accessibility. All-atom MD simulations of ^{C98}RhuA and KCl at the full-size *m*-mica interface revealed completely independent binding mechanisms. The ionic double layer leads to charge reversal of the substrate, and permitted only indirect protein-surface interactions mediated by ions and water molecules, leaving the protein "surfing" several Å above the surface. Low [KCl] preserves the surface charge and does not obstruct physical access to the interface, instead permitting the formation of direct electrostatic interactions between positively charged protein sidechains and the tetrahedral aluminosilicate structure.

1.3.3 Parameter development for the simulation of trioctahedral clays

The experimental interrogation of layered materials such as clays is hindered by many of their natural characteristics, including low symmetry, heterogeneous compositions, defects, and varying degrees of disorder. For example, while the crystal structures of *m*-mica and F-phlogopite have been known for decades, these data provide atomic coordinates (excepting hydrogens), but only statistical information regarding their chemical compositions, which is highly variable due to their propensity for substitutions. Moreover, the presumed complexity of the problem we seek to study (the nano-to-microscale physical chemistry of protein-inorganic interfaces) is exacerbated by the unique physics of surfaces, further constraining the range of applicable experimental techniques (*e.g.*, scanning probe microscopy and surface-sensitive spectroscopic methods). Consequently, computational methods play an essential role in filling knowledge gaps associated with this "invisible region", affording atomic-resolution information regarding the molecular organization at the interface (as well as the energetic factors which produce it) provided that the force fields used to describe these interactions emulate reality with sufficient accuracy.

The aqueous-mica interface has been theoretically modeled using a variety of molecular descriptors with varying degrees of accuracy. DFT calculations have been used to represent the system at the electronic structure level[19], though its computational expense greatly limits the accessible system size and timescales. At the other end of the spectrum are fixed-position atomic coordinates with assumed electrostatic and van der Waals parameters[20], which have since been shown to grossly misrepresent the system energies, and cannot capture dynamic fluctuations of the lattice structure. Thus, a variety of classical force fields have been developed over the years[21-23], which include traditional bonded (bond, angle, dihedral) and nonbonded (Coulombic and van der Waals) terms (and in one case explicit electronic polarization[24]), seeking to strike a balance between accuracy and computational expense. These force fields (even some shown to match predictions from DFT[19]) reproduce experimental observables such as geometries, surface energies, and dynamical properties with varying (sometimes quite poor) levels of accuracy, largely owing to the heuristic origins of their parameters. To address this outstanding issue, the INTERFACE force field (IFF) was developed by Heinz and coworkers[21,25], which is founded on a holistic philosophy of physically justified atomic charges[26] (rationalized from experimental data and multiple lines of reasoning based on fundamental chemical bonding principles), harmonic terms fitted to experimental vibrational spectra, and Lennard-Jones (LJ) terms optimized to reproduce cleavage energies. The result is a highly transferrable force field that reproduces both bulk experimental energies and interfacial properties with great accuracy, while remaining compatible with several other common force fields (*e.g.*, CHARMM[21]). For these reasons, as well as the extensive documentation of parameter derivation, we chose to develop parameters for F-phlogopite within the IFF formalism, which currently has no official descriptors for trioctahedral clays.

As described above, the many similarities between layered silicate materials allows us to adopt many of the required parameters from dioctahedral m-mica. Comparison of the chemical formulas for muscovite ($KAl_2(AlSi_3)O_{10}(OH)_2$) and phlogopite ($KMg_3(AlSi_3)O_{10}(F,OH)_2$) reveals that chemical transmutation of the former into OH-phlogopite is simply the standard dioctahedral-to-trioctahedral substitution of 2 Al^{3+} for 3 Mg^{2+}, with F-phlogopite being the F-substituted variant. With Mg^{2+} parameters already defined within IFF (to represent substitution defects for Al^{3+} in dioctahedral silicates such as montmorillonite), our task was two-fold: balancing the distribution and magnitude of atomic partial charges, and determining reasonable LJ parameters for the F^- ligand.

Following careful consideration of the target structure, we first concluded that this completely restructured octahedral interior meant that Mg^{2+} atoms should not be treated as Al^{3+} substitution defects in this context, but rather an alternative arrangement of the same total charge. Thus, we attempted direct substitution of the 3 Mg^{2+} for 2 Al^{3+} cations (using the partial charges defined in IFF), which resulted in a net charge of +0.4e per unit cell. Following IFF precedent, we determined that the simplest solution was to spread this excess charge evenly to the surrounding oxygen atoms, which are all octahedrally coordinated and thus have near-identical chemical environments. All remaining nonbonded parameters were taken directly from IFF.

Next, we considered the consequences of OH substitution by fluorine. It is known that bulk F-phlogopite has a slightly shorter crystallographic *c* dimension, as the OH group is positioned directly underneath (and oriented towards) the interlayer K^+ ions, producing a basal level of steric- and charge-repulsion. However, F-substituted *m*-mica (whose OH groups are oriented within the layer) does not exhibit this effect, suggesting that the charge and van der Waals parameters of the F^- ligand are likely to be similar to the hydroxyl oxygen. Furthermore, the Mg-F bond length is ~2.02 Å, suggesting a primarily ionic bond character requiring larger partial charges. Direct substitution of the OH-phlogopite structure produces a deficit of 1.1*e*, suggesting a F^- partial charge of −0.55e, 0.233e lower than the O_{OH} charge. Instead, we reverted the oxygen charges from the previous step and placed the excess negative charge on the F ligand, following its higher electronegativity relative to O, resulting in an O_{OH}-like partial charge of −0.683e. Finally, we adapted LJ parameters derived for solvated F^- ions (Heinz, personal communication) by reducing the ε parameter to 0.024 kcal/mol, making it both slightly larger and more weakly dispersive relative to O^{2-}, in line with elemental trends. All model parameters are summarized in **Table 5.1**.

Table 5.1 | Nonbonded force field parameters derived for di- and trioctahedral clays.

Atom type	Charge (*e*) *m*-mica	Charge (*e*) F(OH)-phlogopite	r_0 (pm)	ε (kcal/mol)
$K^{+\ \text{(interlayer)}}$	+1	+1	380	0.20
Si^{surface}	+1.1	+1.1	400	0.05
Al^{surface}	+0.8	+0.8	420	0.05
$Al^{\text{octahedral}}$	+1.45		420	0.05
$Mg^{\text{octahedral}}$		+1.1	420	0.05
O^{surface}	−0.55	−0.55	350	0.025
O^{apical}	−0.758	−0.758 (−0.825)	350	0.025
O^{hydroxyl}	−0.783	(−0.850)	350	0.025
H^{hydroxyl}	+0.3	(+0.3)	109.8	0.013
F		−0.683	366.5	0.024

1.3.4 A new mean-field representation for aluminosilicate surfaces

As described above, there is a representability problem for 1:3 aluminosilicate surfaces such as *m*-mica and phlogopite, namely the specific positions of the Al substitutions. The stoichiometry is purely statistical, but the distribution is also not totally random, obeying (at least) Lowenstein's rule, which states that Al sites are always separated by at least one Si atom (*i.e.*, there are no Al-O-Al bonds). The oldest and most common method of dealing with this issue is to (essentially) ignore it. This is done by constructing rows of Al atoms, which alternate between substituting 1 and 2 Si atoms within adjacent cells (**Figure 5.4**), yielding the correct 1:3 stoichiometry but obliterating a key characteristic of these materials. This model has no basis in reality. It's not just inaccurate, it's *artificially* periodic, and thus antithetical to the present study, which aims to tease out subtle energetic factors controlling epitaxial growth. A logical alternative, though rarely reported, is to computationally generate random sites for substitution and reject any which do not meet the criterion of Lowenstein's rule. Setting aside the scientific concern of hindering others from replicate the study (without directly receiving the exact structure used), the distribution is not *truly* random, but rather the sum of independent probabilities of individual sites having 1–3 substitutions. These individual distributions have been determined by ^{29}Si NMR, and a modeled 2.5×2.7 nm^2 supercell of *m*-mica containing an Al^{3+} substitution pattern which matches these experiments is publicly available from IFF[21] (**Figure 5.4**). While this is treatment is by far the most appealing, it cannot quite solve a final issue associated with the study described here. Our protein of interest (^{C98}RhuA) is a roughly 7×7 nm^2 along its edges and 10 nm along the diagonal, so to simulate even one protein adsorbed to the surface, the requisite surface area necessitates that the provided supercell be tiled 25 times over, once again artificially introducing periodicity into our surface structure (visible in **Figure 5.2**, left), though in a manner significantly closer to reality. While in principle the same fitting procedure to produce the IFF model could be extended to arbitrary sizes of mica, this does not completely avoid the issue of reproducibility.

To address this in a general manner, we instead propose the use of a mean-field approach, whereby the partial charges and vdW radii of the surface Si and Al atoms (which already have identical ε values) are set to their statistical 1:3 average (**Figure 5.4**). This reproduces the average macroscopic surface charge density (and therefore the resulting electric field in solution), and eliminates the caveats associated with explicit substitutions. Our initial findings strongly indicate that this novel treatment of aluminosilicate surfaces is not only of comparable quality to the explicit Al substitution models, but actually enables the detection of subtle molecular patterning that is otherwise obscured. With all the necessary parameters for our simulations now established, we can finally address the intended research topic.

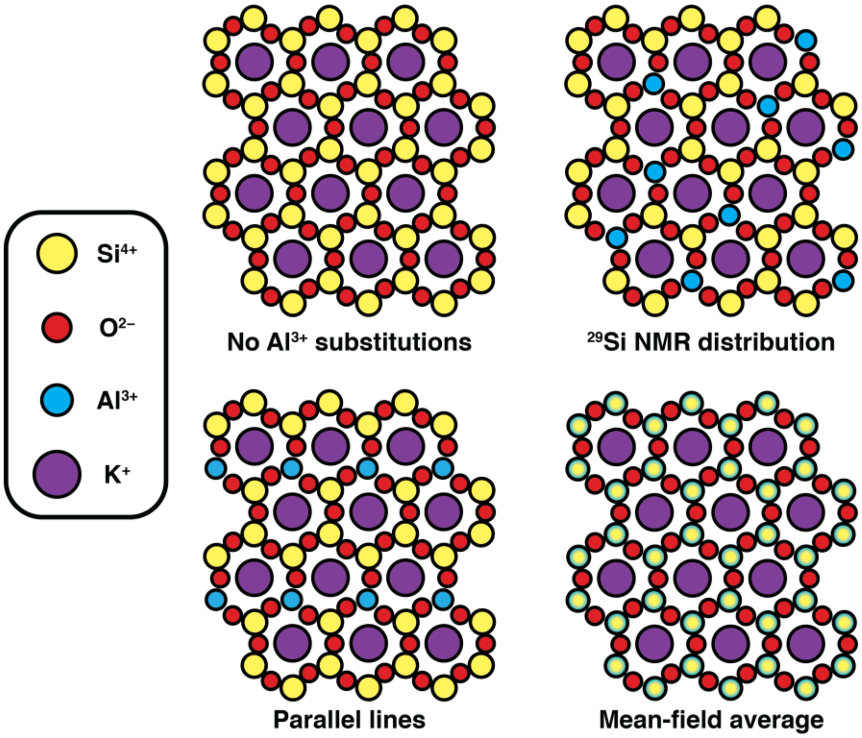


Figure 5.4 | Different representations the same 1:3 Al:Si surface stoichiometry. Al atoms substitute for Si atoms within the outer tetrahedral structure of silicate minerals in a statistical fashion, though computational modeling of these materials has typically included these substitutions explicitly with varying degrees of realism. Though fitting to NMR data greatly improved the realism of these models relative to the older method of parallel lines, tiling of the supercell to study protein adsorption creates artificial periodicity. Here we propose and characterize a mean-field description of the system where all Al/Si sites are represented as one atom type defined as the 1:3 linear combinations of their nonbonded terms, providing a "truly statistical" representation of the surface at all length scales.

1.3.5 Both micas yield indistinguishable molecular distributions along the z-axis

We first characterized the aqueous response of a 3 M KCl electrolyte solution to both minerals (as well as their explicit vs mean-field variants) by standard methods. As is commonly performed, we calculated the slab-averaged 1D positional distributions of all solution species along the surface normal (the z-axis), producing "Z distribution functions" (ZDFs). This quantitatively describes the local structuring of solvent and ions induced by the presence of the highly charged surface, though any patterning within the xy plane is completely averaged out. Consequently, the behavior of both minerals should be very similar if the derived parameters for F-phlogopite are appropriate. A cursory inspection immediately confirmed this to be the case (**Figure 5.5**).

While all ZDFs were found to depict generally similar profiles, several distinctions could be identified that will prove germane to our analysis. As our baseline reference point, we first examined the model with greatest precedent: the explicit-Al^{3+} *m*-mica surface (**Figure 5.5a**, left). Closest to the interface are surface-adsorbed K^+ ions, which occupy a split density peak, reflecting their population at two different sites: K^+ ions at z = 1.4 Å are situated in the hexagonal surface vacancies and comprise the Stern layer, while ions centered at z = 2.0 Å are adsorbed directly onto the explicit Al substitutions of the surface. This behavior has been widely reported in the literature and is therefore unsurprising, however one can already begin envision the consequences of this distribution on patterning in the xy plane. Even at this extremely high electrolyte concentration, our simulations predict that fewer than half of the surface vacancies are occupied by K^+ ions (using the density minimum at z = 1.6 Å as a cutoff), which is consistent with computational literature precedent, but at odds with experimental measurements, where estimates for the KCl concentration at which 50% of sites are occupied range from 0.1–1.0 M. While this highlights a potentially serious limitation of these models, the essential takeaway is the close similarity between the F-phlogopite surface ZDFs (albeit exhibiting slightly different density maxima) and the *m*-mica reference system (**Figure 5.5a**, right). However, the comparable distribution of adsorbed K^+ ions between the two surfaces induces similar local structuring of the solution, producing a dense layer of water and diffuse layer of Cl^- counterions, which together comprise the electric double layer.

In contrast, the mean-field models paint a vastly different picture. The Stern layer is significantly more well-defined, producing a single peak centered at a height of 1.5 Å comprising >90% occupancy of all surface sites (**Figure 5.5b**). This result proved to be the first clear demonstration of the utility of this mean-field approximation, bringing gold-standard parameters even more in line with experiment through a simple modification. The observed increase in ion structuring of the Stern layer produces a knock-on effect, sharpening and intensifying the solvent and ion density peaks even at distances nearly 1 nm away from the interface. This long-range ordering, which arises from our more experimentally consistent model, provides tantalizing evidence for a mechanism by which the (sub)surface chemistry of the mineral is structurally communicated to adsorbed macromolecules via the induced organization of water and ions.

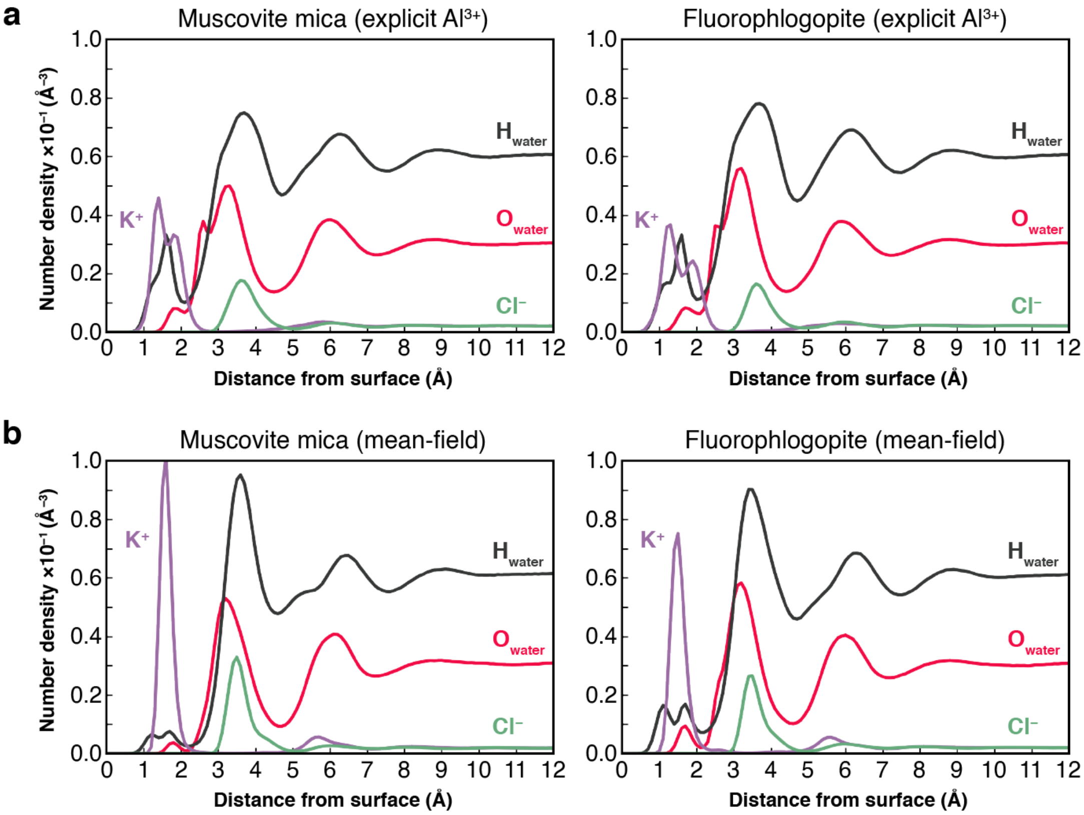


Figure 5.5 | 1D density distributions of aqueous 3 M KCl solutions upon hydration of each model surface. a, ZDFs for the newly derived F-phlogopite parameters (right) are comparable to the previously characterized *m*-mica structure (both containing the ^{29}Si NMR-based Al substitution pattern), indicating that they are physically reasonable. **b,** ZDFs for the corresponding mean-field representations produce a significant increase in K^+ occupancy (~90%) of the surface vacancies, much closer to experimental measurements than the ≤50% occupancy predicted in **a**. Consequently, the greater degree of solvent structuring produced as a results should also better reproduce experimental observables.

As a preliminary exploration of this possibility, we next calculated the angular distribution of water molecules (relative to the surface normal) as a function of distance from the surface (**Figure 5.6**). As seen for the ZDFs, the slab-averaged aqueous response is broadly similar between the two minerals, though here the distributions nearly identical for the explicit Al-substituted models, while the average surfaces exhibiting a modest, yet easily distinguished, difference between them (elaborated below). As before, the transition from an explicit to implicit surface representation results in markedly different solvent responses, reflecting new long-range ordering now captured by our model.

All systems exhibit up to 3 distinct peaks in the angle/distance distributions, corresponding to (from closest to greatest distance): water-occupied surface cavities, aluminosilicate-bonded waters, and K^+-coordinated waters above the Stern layer at $z \approx 3.2$ Å, with their dipoles oriented upwards away from the interface (**Figure 5.6**). We note, however, that the first two peaks nearly vanish for the averaged mineral models, reflecting their occupancy by K^+ ions instead. Thus, consistent with their chemical and structural similarities, both *m*-mica and F-phlogopite induce nearly identical structuring of interfacial solvent and ions (as averaged within the plane), though the modest discrepancies identified between the two mean-field models betrays their subtle distinguishability. As these are *not* captured with explicit atom typing, our methodology promises to finally shed new light onto the enigmatic physics of solid-liquid interfaces.

1.3.6 Observations of substrate-specific patterns encoded within the interfacial water density

At of the time of writing, our quantification of solvent structure above these aluminosilicate surfaces remains extremely limited, however preliminary observations appear to support this narrative as described. To investigate the possibility of in-plane patterning above the surface, we calculated volumetric density maps for the solvent and ions within 1 nm of a 13.0×13.5 nm^2 *m*-

mica surface. 2D slices parallel to the interface indeed revealed a variety of distinctive in-plane patterns that rapidly interchanged as a function of distance from the surface (**Figure 5.7**). The apparent sensitivity of xy solvent patterning to z position indeed suggests that the mineral substrate exerts a powerful form of molecular "imprinting", which encodes its chemical makeup into the local molecular solvent structure over a large contiguous volume.

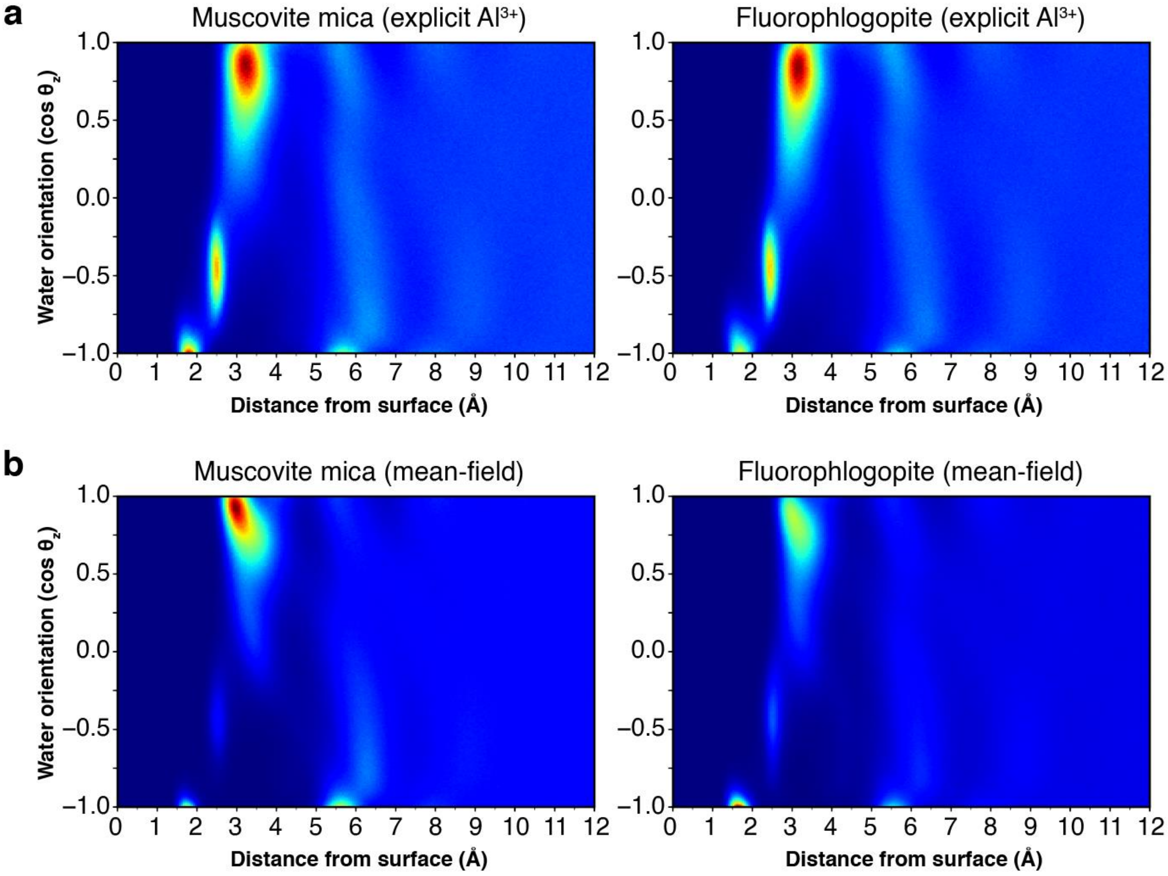


Figure 5.6 | Orientational distributions of solvent molecules above the mineral interface. **a,** 2D distributions depicting the solvent dipolar response of water are identical for the Al-substituted models. **b,** Upon charge-averaging, the two closest distance peaks disappear (due to their occupation by K+ ions), and we observe a difference between the two minerals for the first time. Note the small peak at 5.5–6 Å/−1.0 (both hydrogens oriented down) which is more pronounced for *m*-mica. The continuous vertical band from 6–6.5 Å for the explicit-Al systems shows that water adopts all orientations (randomly) at that distance, while they become more ordered for the mean-field model. This is the same distance range where striping is observed (**Figure 5.7**)

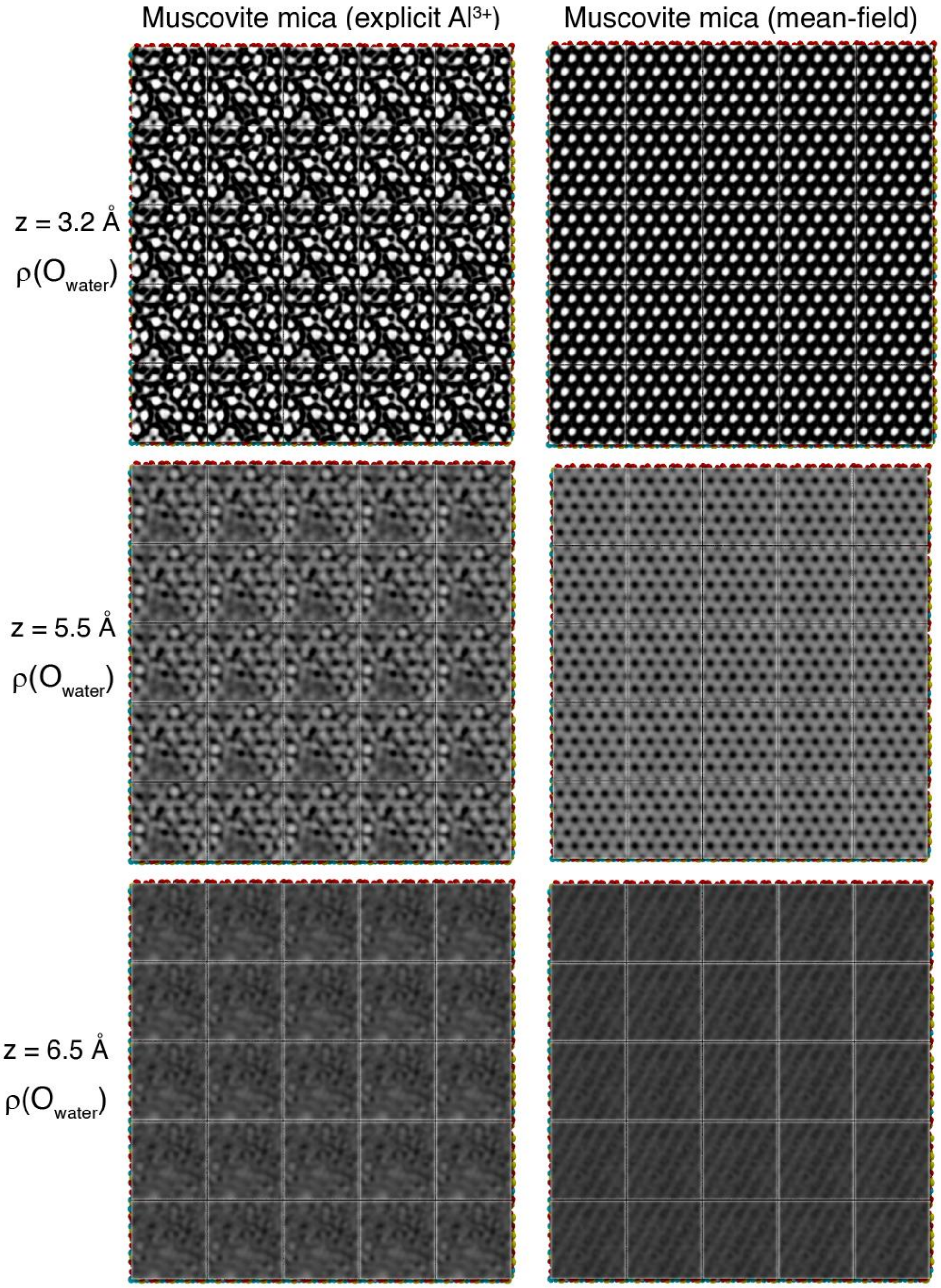


Figure 5.7 | Surface-induced patterns are encoded as spatial variations of the solvent density. The influence of the underlying substrate is imprinted onto the solvent density, as seen here for the 5x3 supercell tiled to an area capable of binding ^{C98}RhuA. On the left, this pattern is dominated by the heterogeneity of explicit Al substitutions. However, when this charge is averaged over the whole surface, we obtain hexagonal arrangements which spontaneously reorganize into 1D striping far from the interface. Darker colors indicate regions of higher density.

Directly at the surface, water and ions snake between the Stern layer cations, producing honeycomb-like outlines, though this quickly dissipates into noise over the next few angstroms. For the explicit Al-substituted structure, no clear patterns ever reemerge beyond this point, though the mild periodicity associated with tiling of the substitution pattern is visible (**Figure 5.7**). In contrast, z ≈ 5.5 Å above the mean-field *m*-mica model, high density peaks for the water oxygens appear, producing a hexagonal array of solvent molecules. This matches a small peak in the dipolar response at that height, where this model exhibits the narrowest spread of solvent orientations (**Figure 5.6b**). Shockingly, further away still, the hexagonal pattern suddenly collapses into close-packed 1D stripes, first manifesting in the H_{water} density at z ≈ 6.1 Å and quickly followed by analogous striping in the O_{water} density at z ≈ 6.4 Å (**Figure 5.7, Figure 5.8**). The collinear striping

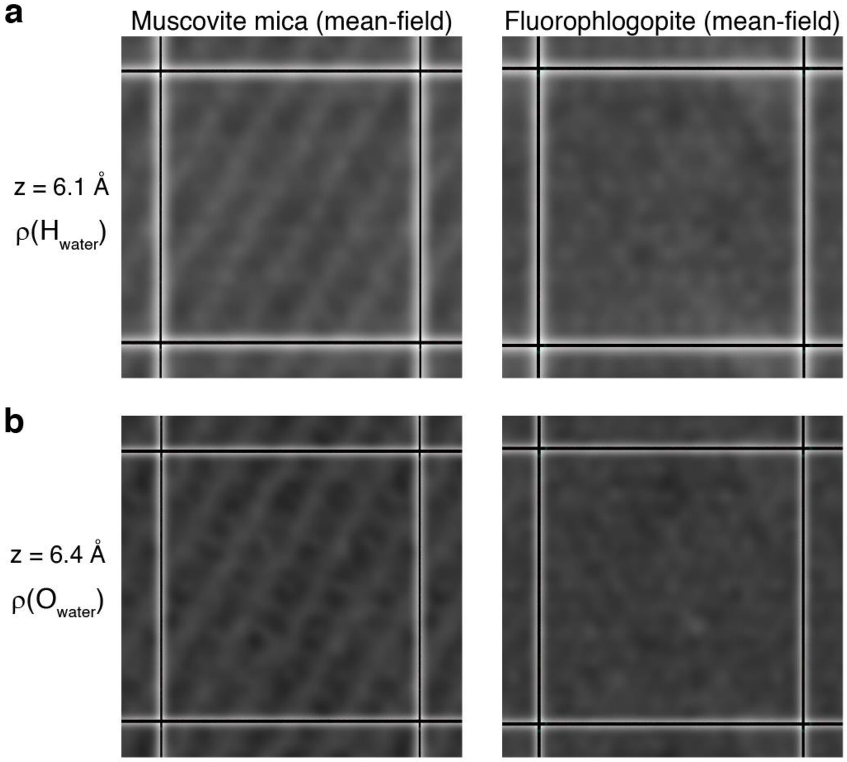


Figure 5.8 | 1D striping is exclusive to muscovite mica. Zoomed-in slices through the solvent density reveal striping of both water hydrogen (**a**) and oxygen atoms (**b**) which is observed for m-mica, but not F-phlogopite, indicating it is not an artifact of the mean-field approximation, but truly reflects the surface chemistry.

of both elements implies that these solvent molecules are in fact oriented within the plane, reflected in their narrower distribution of dipole angles (**Figure 5.6b**). Moreover, the specific direction of the stripes perfectly aligns with the orientation of the subsurface OH groups, potentially validating a long-hypothesized origin of preferential alignment on *m*-mica. Importantly, this difference in patterning is also clearly resolved for full-size (5×5 tiling of the supercell) *m*-mica models, and is thus not a finite-size artifact of the periodic boundaries for the standard supercell size (**Figure 5.9**). Cross-referencing these results with the ZDF profiles, we note that this well-defined ordering spontaneously emerges only at the most distant (significant) peak in the solvent density, consistent with the perceived ability to orient proteins residing ~8 Å above the surface, as observed during simulations of ^{C98}RhuA absorption at the *m*-mica surface (**Figure 4.16**).

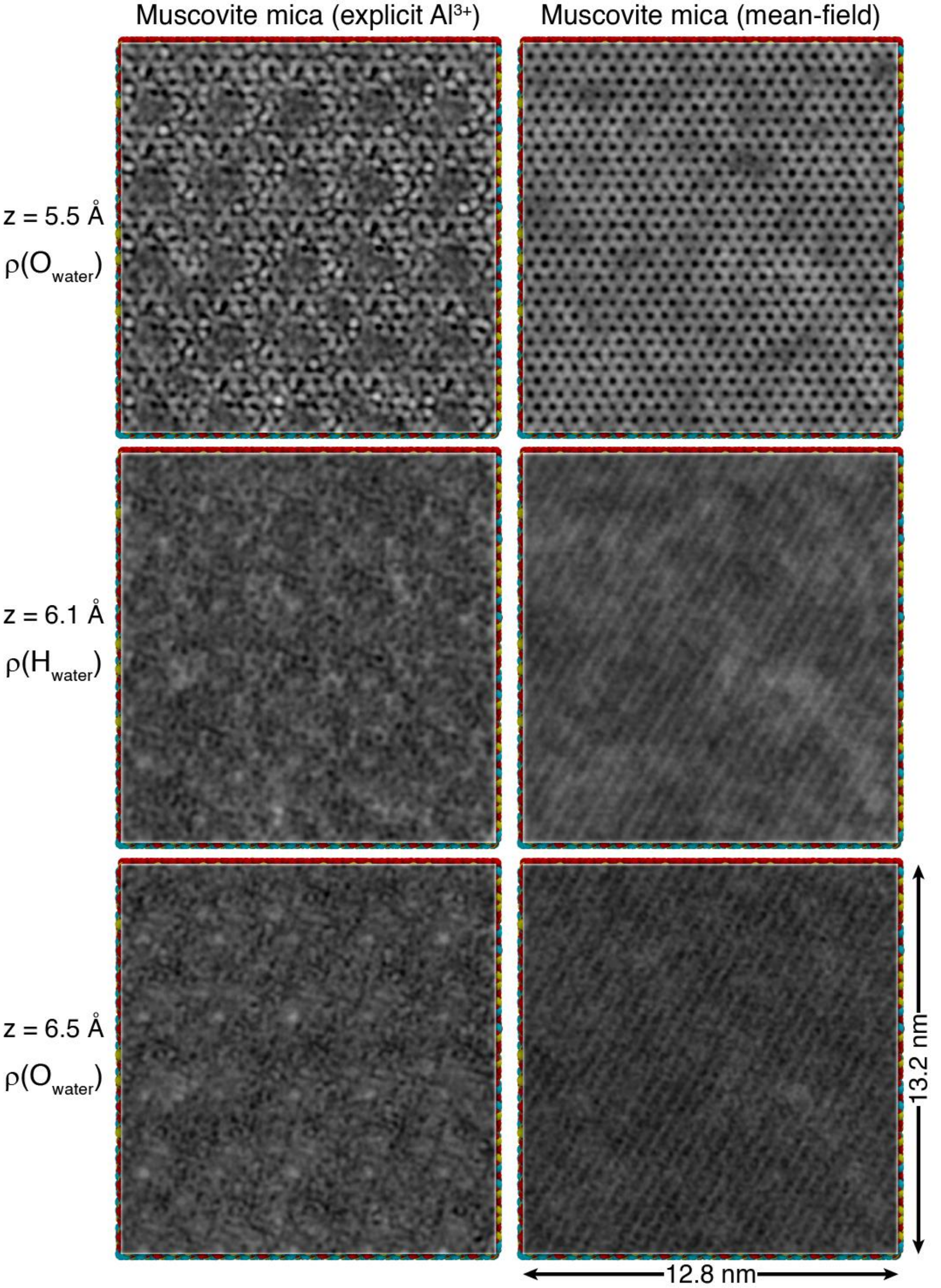


Figure 5.9 | Solvent density patterns are scale invariant. Volume slices of the solvent density for a full-size singular *m*-mica surface equivalent in area to the 5×5 tiling shown in **Figure 5.7** reveals the same patterning. The artificial (roughly) four-fold periodicity associated with tiling the ^{29}Si-NMR-based Al substitutions is particularly noticeable for the O_{water} density at 5.5 Å. Dimensions are provided in the lower right for context. Thus, the mean-field model truly captures the average behavior at all scales, and patterning is not a finite-size artifact from the periodic boundaries.

5.4 Conclusions

Proteins are unmatched in their ability to carry out arbitrary functions or exhibit particular properties, making them highly valuable as nanoscale building blocks. However, this exceptional tunability, which makes them so desirable, is a natural consequence of their remarkable sensitivity to subtle energetic factors, which also renders the energy landscapes underpinning their self-assembly fraught with complexity. Often, such subtleties can be overcome through the use of templates, which reshape these landscapes through the imposition of physical or chemical forces of variable magnitude, though nature is rarely so heavy-handed. Consequently, our ability to utilize templates to design new functional biological materials (*e.g.*, the tunable mechanical properties of collagen) is ultimately limited by our understanding of the delicate interplay of interactions which enable their construction from heterogeneous components.

The experimentally observed alignment of many proteins (including ^{C98}RhuA) along specific lattice vectors fits the general classification of epitaxial growth, however their ability to discriminate between superficially identical patterned substrates (despite residing several angstroms away from the surface) exemplifies the unique promises, and challenges, of their utility as sensing/responsive elements or components of self-assembled materials. Seeking to provide new clarity into the molecular-scale mechanism responsible for this behavior, we developed new parameters for the phyllosilicate mineral (F,OH)-phlogopite (the trioctahedral analog of *m*-mica), which, despite differing only in subsurface composition, produces a distinct epitaxial patterning of ^{C98}RhuA (among many others). We further reported the first application of a mean-field approach to the simulation of aqueous aluminosilicate interfaces, which (in addition to producing a physical picture significantly more in line with experimental expectations) enabled fresh insights into the molecular organization of solvent and ions at the mineral surface via MD simulations, further elevating its utility as a tool to probe complex interfacial phenomena.

In summary, the initial results afforded by our simulations have suggested a new mechanism for substrate patterning. We now posit that the physicochemical properties of the substrate are first mapped directly into the local solvent structure, directly producing strong, collective density fluctuations which in turn produce induced local potentials to which macromolecules are sensitive. In this way, the solvent serves to both translate physical patterns into chemical potentials and communicate this information over long distances. If true, such a mechanism may represent a general physical framework which (at least partially) explains the universal ability of proteins to distinguish such subtle chemical differences from remote distances.

5.5 Methods

5.5.1 Development and testing of classical force field parameters for fluorophlogopite

Initial atomic coordinates to describe the structural model of fluorophlogopite were taken from the 24 °C single-crystal x-ray structure of synthetic F-phlogopite as reported by Takeda and Morosin[27]. From these coordinates, a *P*1 unit cell was constructed, and a CHARMM-type topology file was manually constructed to describe it. This cell was then tiled to yield a 5×3 supercell (resembling the *m*-mica model structure from IFF) and all missing bonds were retroactively added via PSFGEN and the topology file defined for the unit cell. All bond distances and angles were extracted, verified against reported values, and transcribed into a CHARMM-type parameter file, multiplying the bond lengths by 1.05 as prescribed by IFF[21]. Si positions corresponding to the surface Al substitutions on the IFF 5×3 *m*-mica cell were manually identified and converted to Al, creating a final 5×3 F-phlogopite supercell perfectly analogous to the model for *m*-mica, with the intention of minimizing the possibility of artifacts which could be mistaken for true differences between the minerals. The construction of large 5×5 tilings of these supercells was performed as previously described[17]. Briefly, a custom Tcl build script automates the removal of the periodic bonds included in the original supercell, generates tiled copies of the cell, and reconstructs all missing bonds (periodic and nearest-neighbor).

As described in Section 5.3.3, nearly all parameters for F-phlogopite could be directly largely adopted from the IFF silicate parameters used to represent muscovite. Only the selection of F parameters required significant deliberation, though extensive personal communication with H. Heinz has convinced us both that the charge and LJ parameters are quite suitable.

5.5.2 Molecular modeling and simulation of hydrated mineral interfaces

Triple-stacks of each both mineral slabs were constructed and subjected to simulation (with minor modification) as described previously[17]. Briefly, simulations were carried out at constant pressure while maintaining a fixed xy area corresponding to the crystallographic dimensions of the mineral surface for each 5×3 supercell, at a temperature of 300 K, hydrated by a solution of 3M KCl ions described using the parameters of Joung and Cheatham[28]. All simulations were performed using NAMD 2.13[29]. After merging the equilibrated water/KCl system with the *m*-mica triple-stack, the coordinates were minimized for 500 steps. All *m*-mica atoms were constrained to their initial positions by a 1 kcal/mol restraint which was relaxed over the first 5 ns, followed by 5 ns equilibration and then production runs with the central layer (non-H atom) COM held in place. RDFs from 3M KCl solutions of J/C ions using CHARMM TIP3P were found to overestimate the strength of K^+–Cl^- pairing relative to standard TIP3P when referenced against neutron diffraction data of Soper and coworkers[30,31] (data not shown), so NBFIX terms were applied to the CHARMM TIP3P hydrogen-J/C ion parameters to reproduce the correct ion behavior while retaining CHARMM TIP3P behavior for protein simulations with the CHARMM36 parameter set.

5.5.3 Volumetric mapping of the solvent and ion structuring near mineral interfaces

Trajectories were first aligned to place the average position of the bridging oxygens of the silicate surface at $z = 0$, and all atoms wrapped into the central box. The solvent density was then calculated from $z = 0$–10 Å using the volmap tool in VMD with all atom radii scaled to 1.0, as averaged over 100 ns (5×3 supercell) or 20 ns (full-size 5×5 supercell tiling) of simulation. The spatial variation of the density was then visualized as slices through the volumes within VMD.

5.5.4 Calculation of 1D density distributions

Coordinates for the z-positions of all atoms within 20 Å of the surface were extracted for all 100 (20) ns of simulations, then binned into a 1D histogram via NumPy to produce the "ZDFs". For the orientational distributions, cos θ was calculated by calculation of the dot product between the water dipole moment (calculated using the measure dipole in VMD) and the (0,0,1) vector for all waters with their O_{water} within 20 Å of the surface, then binned into a 2D histogram via NumPy.

5.6 Acknowledgments

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. 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.

Chapter 5, in part, is currently being prepared for submission for publication: Alberstein, R.G.*, Prelesnik, J.L.*, Mundy, C.J., Tezcan, F.A. "Molecular basis for surface patterning of biomolecules at aqueous mineral interfaces". The dissertation author was the co-primary investigator and author of this manuscript.

5.7 References

1. Marsh, J. A. & Teichmann, S. A. Structure, Dynamics, Assembly, and Evolution of Protein Complexes. *Annu. Rev. Biochem.* **84**, 551-575, (2015).

2. Pereira-Leal, J. B., Levy, E. D. & Teichmann, S. A. The origins and evolution of functional modules: lessons from protein complexes. **361**, 507-517, (2006).

3. Levy, E. D., Erba, E. B., Robinson, C. V. & Teichmann, S. A. Assembly reflects evolution of protein complexes. *Nature* **453**, 1262-1265, (2008).

4. Wösten, H. A. B. Hydrophobins: Multipurpose Proteins. *Annu. Rev. Microbiol.* **55**, 625-646, (2001).
5. Branda, S. S., Vik, Å., Friedman, L. & Kolter, R. Biofilms: the matrix revisited. *Trends Microbiol.* **13**, 20-26, (2005).

6. Sodek, J., Ganss, B. & McKee, M. D. Osteopontin. *Crit. Rev. Oral Biol. Med.* **11**, 279-303, (2000).

7. Tao, J., Fijneman, A., Wan, J., Prajapati, S., Mukherjee, K., Fernandez-Martinez, A., Moradian-Oldak, J. & De Yoreo, J. J. Control of Calcium Phosphate Nucleation and Transformation through Interactions of Enamelin and Amelogenin Exhibits the "Goldilocks Effect". *Cryst. Growth. Des.* **18**, 7391-7400, (2018).

8. Burazerovic, S., Gradinaru, J., Pierron, J. & Ward, T. R. Hierarchical self-assembly of one-dimensional streptavidin bundles as a collagen mimetic for the biomineralization of calcite. *Angew. Chem. Int. Ed. Eng.* **46**, 5510-5514, (2007).

9. Mann, S., Archibald, D. D., Didymus, J. M., Douglas, T., Heywood, B. R., Meldrum, F. C. & Reeves, N. J. Crystallization at Inorganic-organic Interfaces: Biominerals and Biomimetic Synthesis. *Science* **261**, 1286-1292, (1993).

10. Yang, W., Chen, G., Shi, Z., Liu, C.-C., Zhang, L., Xie, G., Cheng, M., Wang, D., Yang, R., Shi, D., Watanabe, K., Taniguchi, T., Yao, Y., Zhang, Y. & Zhang, G. Epitaxial growth of single-domain graphene on hexagonal boron nitride. *Nature Mater.* **12**, 792-797, (2013).

11. Suntola, T. Atomic layer epitaxy. *Thin Solid Films* **216**, 84-89, (1992).

12. Chambers, S. A. Epitaxial growth and properties of thin film oxides. *Surf. Scie. Rep.* **39**, 105-180, (2000).

13. Leow, W. W. & Hwang, W. Epitaxially Guided Assembly of Collagen Layers on Mica Surfaces. *Langmuir* **27**, 10907-10913, (2011).

14. Czajkowsky, D. M., Li, L., Sun, J., Hu, J. & Shao, Z. Heteroepitaxial Streptavidin Nanocrystals Reveal Critical Role of Proton "Fingers" and Subsurface Atoms in Determining Adsorbed Protein Orientation. *ACS Nano* **6**, 190-198, (2012).

15. Loo, R. W. & Goh, M. C. Potassium Ion Mediated Collagen Microfibril Assembly on Mica. *Langmuir* **24**, 13276-13278, (2008).

16. 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).

17. Zhang, S., Alberstein, R., De Yoreo, J. J. & Tezcan, F. A. Assembly of a Patchy Protein into Variable 2D Lattices via Tunable, Multiscale Interactions. *ChemRxiv*, (2020).

18. Wang, L., Guo, Y., Li, P. & Song, Y. Anion-Specific Effects on the Assembly of Collagen Layers Mediated by Magnesium Ion on Mica Surface. *J. Phys. Chem. B* **118**, 511-518, (2014).

19. Prakash, A., Pfaendtner, J., Chun, J. & Mundy, C. J. Quantifying the Molecular-Scale Aqueous Response to the Mica Surface. *J. Phys. Chem. C* **121**, 18496-18504, (2017).

20. Skipper, N. T., Refson, K. & McConnell, J. D. C. Computer simulation of interlayer water in 2:1 clays. *J. Chem. Phys.* **94**, 7434-7445, (1991).

21. 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).

22. Cygan, R. T., Liang, J.-J. & Kalinichev, A. G. Molecular Models of Hydroxide, Oxyhydroxide, and Clay Phases and the Development of a General Force Field. *J. Phys. Chem. B* **108**, 1255-1266, (2004).

23. Cygan, R. T., Greathouse, J. A., Heinz, H. & Kalinichev, A. G. Molecular models and simulations of layered materials. *J. Mater. Chem.* **19**, 2470-2481, (2009).

24. Tesson, S., Louisfrema, W., Salanne, M., Boutin, A., Rotenberg, B. & Marry, V. Classical Polarizable Force Field To Study Dry Charged Clays and Zeolites. *J. Phys. Chem. C* **121**, 9833-9846, (2017).

25. Heinz, H., Koerner, H., Anderson, K. L., Vaia, R. A. & Farmer, B. L. Force Field for Mica-Type Silicates and Dynamics of Octadecylammonium Chains Grafted to Montmorillonite. *Chem. Mater.* **17**, 5658-5669, (2005).

26. Heinz, H. & Suter, U. W. Atomic Charges for Classical Simulations of Polar Systems. *J. Phys. Chem. B* **108**, 18341-18352, (2004).

27. Takedia, H. & Morosin, B. Comparison of Observed and Predicted Structural Parameters of Mica at High Temperature. *Acta Cryst.* **B31**, 2444-2452, (1975).

28. 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).

29. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L. & Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781-1802, (2005).

30. Mancinelli, R., Botti, A., Bruni, F., Ricci, M. A., Soper A. K. Hydration of Sodium, Potassium, and Chloride Ions in Solution and the Concept of Structure Maker/Breaker. *J. Phys. Chem. B* **111**, 13570-13577, (2007).

31. Mancinelli, R., Botti, A., Bruni, F., Ricci, M. A., Soper A. K. Perturbation of water structure due to monovalent ions in solution. *Phys. Chem. Chem. Phys.* **9**, 2959-2967, (2007).

Chapter 6: Reflection and Outlook

Our appreciation of, and respect for, the beauty and complexity of life stretches back millennia, likely, I would wager, predating when humans became humans. The appearance of self-sustaining living systems on this “mote of dust” floating through the cosmos inspires a sense of awe, and curiosity, for the physical laws which give rise to them. This sensation has only grown, as our ability to investigate biological processes in finer detail has revealed an endless number of elegant solutions to the challenges which must be surmounted for life to go on. Of these, there are none more elegant than the biological machinery of proteins, which fold, wiggle, construct, deconstruct, switch, signal, read and write biological information, regulate and respond, just to name a few of their myriad functions. This represents my personal inspiration.

To enable this broad functional employ, proteins must be able to recognize and interface with specific partners, as well as respond to environmental stimuli, in prescribed fashions. Their ability to do so in “uncontrolled” crowded cellular environment underscores the specificity by which they follow thermodynamic driving forces to generate forward motion from a scale dominated by random fluctuations. Consequently, our desire to harness these capabilities for our own technological ends necessitates an understanding of how nanoscale organization manifests at the macroscale, and how we might control this organization in a logical manner. To paraphrase Feynman: it is not that molecules are too small, it is that we are too big! This dissertation has attempted to elucidate some of these thermodynamic drivers as they pertain to the emergence of unusual and/or subtle processes.

In Chapter 2, we carried out extensive computational and experimental characterization of a synthetic protein lattice capable of undergoing large-scale dynamic breathing modes. All-atom calculations predicted that the reorganization of solvent enabled the stabilization of the nonporous conformation over all others, a preference which we able to overcome through the installation of repulsive metal-binding residues, which created a chemical switch that prevented closure of the lattice until specific metal ions were introduced. This solvent-exclusion effect is appreciated as a general biological phenomenon, particularly in crowded environments, and these lattices may provide a convenient method for systematic studies on the consequences of perturbation to the local hydration environment of proteins. Eventually, this may provide another mechanism by which we may exert control over the assembly/dynamics of proteins in an indirect fashion.

Chapter 3 capitalized on the switching behavior of these new lattices in order to provide a specialized protective coating for an inorganic photonic sensor. In the nonporous state, this coating protected the sensor from arbitrary chemical species, only becoming permeable in the presence of HCN, which displaced the bound metal to “pop open” the lattices. What remains to be seen is to what extent this methodology can be extended to other analytes. Many species bind transition metal ions, which is both a strength and weakness of this method. However, the ability of proteins to bind molecular structures with very high specificity suggests that more complex designs may enable arbitrary species to be targeted instead.

In Chapter 4, we described the self-assembly of our protein of interest into a diversity of 2D crystalline lattices. This control was afforded through use of a mineral surface as a template, whose charge and chemical composition could be modulated by solution conditions. By characterizing the energetic drivers responsible for each morphology, we delineated a complete picture of all self-assembly pathway and how various factors, both intrinsic and external, enables their formation. Clearly this is a powerful method for constructing new protein-based materials with specified properties, as exemplified by the predicted piezoelectric behavior of the surface-templated lattices. Of immediate interest would be the electrochemical characterization of these properties, which may render these lattices effective as electroosmotic molecular membranes or for passive power generation through the segregation of ions. It also paves the way for dipole-dipole interactions to be intentionally exploited by protein engineers as a way to provide subtle energetic nudges to the assembly landscape to obtain specific products.

Subtlety remained a theme in Chapter 5, where we reported initial characterizations a complex aqueous inorganic interfaces. Our results suggest a possible mechanism by which these surfaces imprint their physical and chemical structures into the local structure of the solvent in a manner which is recognizable by biomolecules. Experimental demonstrations of this prediction (should it continue to hold under further theoretical investigation) will likely require careful quantitative mapping of the interfacial solvent via high-resolution AFM. Furthermore, how such a mechanism may be exploited to control protein self-assembly remains to be determined. At the very least, we may take inspiration from the lattice-matching strategy employed by natural (*e.g.*, ice-binding) proteins to design new structures that recognize specific substrates.

Our ability to master these molecular forces (and how biological structures respond to them) will certainly come to define our capabilities as molecular engineers over the foreseeable future, particularly as we seek to recreate the intricate and powerful capabilities of some of nature’s most advanced machines for our own purposes. After all, making tools is what we humans do best!