Protein-mineral interactions and inorganic nucleation through designed protein interfaces

Fátima Angélica Dávila Hernández

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2023

Reading committee:
David Baker, Chair
James J. De Yoreo
Lilo D. Pozzo

Program Authorized to Offer Degree:
Molecular Engineering

University of Washington

Abstract

Protein-mineral interactions and inorganic nucleation through designed protein interfaces

Fátima Angélica Dávila Hernández

Chair of the Supervisory Committee:
Professor David Baker
Department of Biochemistry

Biomolecules have the ability to regulate the formation of hierarchically structured biominerals through their interactions with inorganic crystals. However, the details of the atomic structure at the organic-inorganic interface that governs this process are not yet known. This work will present a set of design principles for the creation of molecular templates targeted to interact with calcium carbonate and hematite. The starting hypothesis states that a structured flat molecular template could achieve heterogeneous nucleation of calcium carbonate or facet-specific and oriented binding of proteins to hematite by pre-organizing binding moieties for calcium or iron on its surface. To test this, helical repeat proteins displaying regularly spaced carboxylate arrays on their surfaces were designed. It was discovered that these protein templates directly nucleate nano-calcite with non-natural (110) or (202) facets. These proteins also allow for the bypassing of vaterite, which forms in the absence of the proteins. The resulting nanocrystals then come together by oriented attachment to form calcite mesocrystals. By altering the protein length and manipulating their surface chemistry, the nanocrystal size and nucleation rate can be adjusted. As the size of the carboxylate arrays decreased, the nanocrystal diameters increased. Furthermore, the nucleation activity was eliminated by partially replacing the carboxylates with lysines. In the case of hematite, binding studies suggested that proteins with a target spacing of 10.9 Å (but a most likely interhelical spacing of 11.2 Å) are capable of interacting specifically with the target (012) surface. This was observed for incubation conditions under which the surface potential is identical for both test facets. Binding on the off-target (001) surface was heterogeneous and domain-dependent. Nucleation studies of hematite in the presence of proteins suggested that several designed and non-designed proteins had a degree of inhibition in the effect of nucleation of hematite starting with a ferrihydrite precursor. In sum, these templates achieve a degree of tunability only accessible through protein design and represent one of the most programmable systems for broader biomineralization studies. These advances open the possibility to use de novo protein design to program biomineralization, offering a pathway to creating advanced hybrid materials.

ACKNOWLEDGEMENTS

Grad school has been an adventure and I wouldn't have been able to make it through it without an extensive support network within and outside UW including friends and family. First and foremost, I would like to thank my family. My parents Angélica Hernández and Miguel Dávila. My siblings David and Marifer Dávila. They have been my main source of support, inspiration, respite and optimism and I love them deeply.

I'm thankful to Luz Mejía and Alvarado Rodríguez, who without knowing me opened the doors of their house when I first arrived to Seattle without having a place where to stay and allowed me to feel like family.

Rahul Tewari and Manny have been there with me for almost five years now. Having faith in me, cheering me on, allowing me to rest with them and celebrating together. They have been invaluable companions and I'm grateful for their love and because I found them in my life.

I'm thankful for my grad school friends Breena Sperry, Julian Freedland, Reid Schur and Jake Geiger. I'm also happy that I had amazing and supportive flat mates such as Carmen Hom and Victoria Tran. My friends in Seattle and family away from home Melody Wang, Julia Gall and Felipe Lima.

I'm also grateful for the financial support from the Center for the Science of Synthesis Across Scales (CSSAS), a Energy Frontiers Research Center (EFRC) from the Department of Energy (DOE), as well as support from the Molecular Engineering and Sciences Program.

This work is done in close collaboration with members of the De Yoreo Group. Discussions with them and their guidance have been invaluable in shaping this project and formulating new, exciting research questions. Special thanks and acknowledgement to Professor James J. De Yoreo; Professor Shuai Zhang, for collection and analysis of all the AFM data presented here; to Dr. Guomin Zhu, who carried out all the hematite nucleation experiments and the data collection; and to Dr. Biao Jin, my close collaborator in the calcium carbonate project who collected the EM data presented in the following chapters. Special thanks also to Amy Stegmann for her friendship and scientific insight.

I would also like to thank the entire Baker Lab and members of the Institute for Protein Design (IPD) — with special acknowledgement to Dr. Harley Pyles for his guidance, support, and friendship — and of course, Professor David Baker who has given me the opportunity to research in his group. I'm also grateful for the support and friendship from current and former IPD members including Susana Vásquez Torres, Amijai Saragovi, Meerit Said, Franziska Seeger, Stacey Gerben, Christian Richardson, Audrey Olshefsky, Erin Yang, Anindya Roy, and Christine Kang.

Thank you to the members of the Protein Production and Crystallography Cores at the IPD, who carried out protein production for crystallography trials, helped to set and screen crystal trays, as well as running several characterization methods included in this work (LCMS, SECMALS).

1 Chapter 1: BACKGROUND and INTRODUCTION

1.1 Biomineralization

1.1.1 Overview

Nature's ability to create inorganic crystals and shape them into remarkably intricate and regular structures has fascinated scientists for centuries. Biomineralization is the process by which living organisms deposit inorganic material in a controlled manner. This can be as a byproduct of their metabolism, or intentionally as an adaptation mechanism imparting them structural, optical, or environmental sensing advantages. These processes range from membrane bound nanoparticles in unicellular organisms, to complex processes ranging multiple length scales in multicellular organisms.

The story and developments in biomineralization have followed the evolution of microscopy and structural characterization techniques since the late 17th century[1]. From the first optical microscopy studies by Van Leeuwenhoek studying osteons in bones, to following studies through the development of polarized light microscopy, X-ray diffraction, transmission electron microscopy (TEM), scanning electron microscopy (SEM), and the cryo or in situ techniques developed for the latter two microscopy techniques. Advances within the field in the past ten years have been centered around transcriptomic and proteomic studies of mineralizing tissue, or following structural developments and interactions between the organic matrix and the mineralized component. Together, they have painted a picture of a biomineralization field where sequence to structure to function relationships are key. These relationships together with the properties of important biominerals will be explored in the following section.

1.1.2 Biominerals

The resulting biominerals have unique properties that go beyond what is achievable by their synthetic or geological counterparts. For example ice-binding (IBP) and ice-nucleating proteins (INP) are capable of the molecular recognition of ice and modification of its properties such as the non-equilibrium depression of its melting point (thermal hysteresis), ice recrystallization inhibition, and dynamic ice shaping[2,3]. These proteins are used as an adaptation mechanism for organisms living at sub-zero temperatures. The remarkable activity of these molecules has led to the development of small molecule and small peptide mimics whose potency remains variable and not on par with that of native proteins[4]. Hence, numerous structure-functional studies[5–9] have been carried out with the final goal of rationally designing potent mimetics with improved stability and synthesizability. Biomineralized calcium carbonate has superior mechanical properties, e.g. in nacre, the strength of a nacre shell is up to 3000 higher than of pure $CaCO_3$ [10]. This is achieved by layers deposited of a calcium carbonate mineral called aragonite sandwiched between organic matter believed to template and concentrate the precipitated inorganic phase[11].

In general, the way by which these properties are achieved can be related to the ability of organisms to control the shape, position, phase, and orientation of the deposited minerals. This can be in turn obtained by carefully regulating the environment, i.e. the flux of electrons, pH, and ionic strength. An additional control includes the use of templates and constrained spaces for the deposition of minerals. The combined use of solution condition controls and confined environments with templates manipulates the free energy landscape of nucleation and growth and effectively lowers the activation energy barrier to produce minerals with fewer energetic requirements compared to minerals attained by synthetic routes or geological processes[12]. The following section will explain how this is possible from the perspective of classical nucleation theory (CNT).

1.2 Nucleation theory

This section will first frame the broader concepts of biomineralization through CNT in solution and at surfaces. Crystallization is a phase transition from a high energy solvated state to a low energy crystalline lattice. Living organisms control the phase, habit, and growth of a crystal by manipulating the free energy landscape of this phase transformation. This can be achieved by modulating the speciation of ions in solution, controlling the pH, in addition to potentially using templates to guide the nucleation as described by the principles of homogeneous and heterogeneous nucleation outlined below.

1.2.1 Homogeneous vs heterogeneous nucleation

The contents presented in this section are adapted from De Yoreo & Vekilov[13]. For ions and molecules to transition from a solvated state to a solid, crystalline phase, a driving force is required. The formation of a precipitate, in this case the new phase, is only thermodynamically favorable when the free energy of the new phase in the final solution is lower compared to the free

energy of its constituents in solution, i.e. the old solvated phase at equilibrium $\Delta g_{equilibrium}$. This condition is achieved when the activity product of the reactants (AP) exceeds the equilibrium activity product (K_{sp}), which is also known as supersaturation (σ). Therefore, the chemical potential (Δμ) or driving force for nucleation is determined by the degree of supersaturation as described by the equations below. Here k_B is Boltzmann's constant and T is the absolute temperature.

$$\Delta\mu = \Delta g_{supersaturated} - \Delta g_{equilibrium} \quad (1.1)$$

$$= k_B T ln AP - k_B T ln K_{sp} \quad (1.2)$$

$$= k_B T ln\left(\frac{AP}{K_{sp}}\right) \quad (1.3)$$

$$= k_B T \sigma \quad (1.4)$$

$$\sigma \equiv ln\left(\frac{AP}{K_{sp}}\right) \quad (1.5)$$

Ions in solution, pH, redox potential, and ionic strength have all an effect on the degree of supersaturation. The strong effect of solution conditions on supersaturation highlights how regulation of solution conditions by organisms can induce and control nucleation. Local fluctuations of particles then drive the formation and dissolution of clusters at a rate governed by the chemical potential as a driving force, and the interfacial free energy (α) as a barrier for further growth (Fig. 1.1, upper pathway).

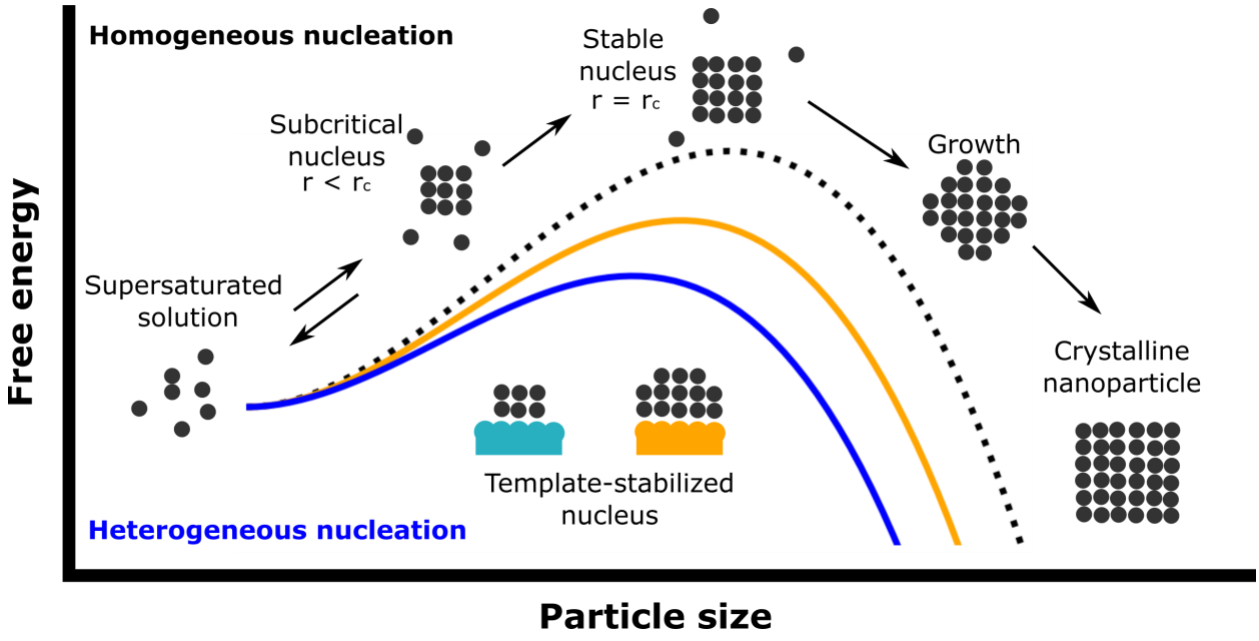

Figure 1.1. Homogeneous and heterogeneous nucleation (reproduced from Davila-Hernandez et al.[14]). The upper pathway and dotted black line show the nucleation process in the absence of a template. Fluctuations of particles need to overcome a free energy barrier dictated by the interfacial free energy of the growing particle in order to reach equilibrium at the critical radius and then grow further. The presence of a template (in orange and blue) lowers both the critical radius and the free energy barrier. Better templates which can further reduce the interfacial free energy may create even smaller nuclei (compare orange with blue).

The interfacial free energy can be defined as the energy required to form and stabilize a surface or boundary between two dissimilar phases. To relate the driving force and barriers to nucleation, another common key assumption is made about the shape of the nuclei. The nucleus is assumed to adopt a spherical shape, as this is most likely to maximize the amount of molecules in the new phase with the lowest surface area, hence adopting the lowest energy conformation. Considering this, the following expression describes the sum of the bulk (Δg_b) and surface (Δg_s) free energy terms, where Ω is the molecular volume and r is the radius of the sphere.

$$\Delta g = \Delta g_b + \Delta g_s \tag{1.6}$$

$$= -\left\{\frac{\left(\frac{4}{3}\right)\pi r^3}{\Omega}\right\}\Delta\mu + 4\pi r^2\alpha \tag{1.7}$$

By plotting this equation as a function of the particle radius, a maximum point is reached at a critical radius (r_c) where the fluctuations building or dissolving the new phase are at

equilibrium (Fig. 1.1). Ultimately, under supersaturation conditions, the new phase will form, but the critical size of the nucleus and the related interfacial free energy and supersaturation will determine the rate and the onset of nucleation (J_n) as described in (Eq. 1.8) where A is a factor describing geometry and material parameters. In brief, this equation is derived by finding first r_c and then obtaining the free energy barrier of nucleation (Δg_n) by substituting the expression for r_c back into (Eq. 1.7).

$$J_n = Aexp(-\Delta g_n/k_BT) \tag{1.8}$$

$$= Aexp(-B\alpha^3/\sigma^2) \tag{1.9}$$

$$B \equiv \frac{16\,\pi\Omega}{3k_BT} \tag{1.10}$$

These concepts summarize the main factors of homogeneous nucleation. The next step will be to discuss heterogeneous nucleation as a way to change the rate of nucleation by changing factors affecting the interfacial free energy.

The introduction of a potential template for nucleation results in the formation of additional surfaces or boundaries and their associated energies. Elements composing the total interfacial free energy in (Eq. 1.11) below are: α_{lc} between the solvent or liquid and the new phase – in this case the crystal nucleus; α_{sc} between the template or substrate and the crystal nucleus; and α_{ls} between the liquid and the substrate.

$$\alpha_{tot} = \alpha_{lc}\{1 - (\alpha_{ls} - \alpha_{sc})/2\alpha_{lc}\} \tag{1.11}$$

The goal of a template is to significantly reduce the total interfacial free energy. Examples by which this can be achieved include presenting chemical moieties compatible with the crystal surface, by structuring its surface to closely match that of the nucleus and minimize lattice strain, and to interact with the solvent to favor ion desolvation at the interface. A combination of these effects can diminish the total interfacial free energy and as a consequence have a deep impact on the barrier for nucleation, shift the size of r_c to smaller values, hence changing the rate of

nucleation(Fig. 1.1). Templates can offer additional features such as determining the plane from which nucleation can start, the number, location, and orientation of the nuclei, and indirectly affect the final size of the particles produced. Different ways by which templates can achieve these features will be described in subsequent sections.

1.2.2 Ostwald law of phases

The contents presented in this section are adapted from De Yoreo & Vekilov[13]. The pathway of nucleation in CNT is commonly assumed to proceed directly from solution to the formation of nuclei with a crystalline structure that is identical to that of the eventual bulk crystal. This is based on the capillary approximation, which assumes that the molecular arrangement in a crystal's embryo is the same as in a large crystal, making the surface free energy of the nuclei equivalent to that of the crystal interface. However, this assumption is unlikely to hold because the energy barrier leading to a more disordered, less stable state is lower than that leading to the most stable state. This revision is encapsulated by the Ostwald-Lussac law of phases, which explains that the pathway to the final crystalline state will pass through all less stable states in order of increasing stability.

The rate of nucleation equation (Eq. 1.9) shows that the only way a less stable phase can nucleate more rapidly is by reducing the interfacial free energy by supersaturation ratio (α^3/σ^2). For a less stable phase, this implies that the K_{sp} is larger, which means that σ is smaller. Therefore, α must be reduced for the ratio to become smaller. The interfacial free energy (α) is proportional to the difference between the change in free energy in forming a crystal with a surface (Δg) and that of forming an infinite crystal (Δg_b) as evidenced by (Eq. 1.12), which is in turn derived from (Eqs. 1.6 & 1.7).

$$\alpha = (\Delta g - \Delta g_b)/4\pi r^2 \tag{1.12}$$

Enthalpic and entropic contributions to the free energy terms can help explain the differences between the less or the more stable phases (i.e., Δg_i=$\Delta h_i - T\Delta s_i$). The less stable phase would have both bulk and surface free energies that are smaller than that of the more stable phase. This is ensured by the poorer bonding corresponding to the smaller enthalpy contribution (Δh_i) and reflecting as well on the higher level of disorder which impacts the entropy term (Δs_i).

1.3 Minerals of interest and their biogenic or bio-inspired synthesis

1.3.1 Calcium carbonate & other calcium biominerals

The known polymorphs of calcium carbonate are six: amorphous calcium carbonate (ACC), crystalline monohydrate, ikaite, vaterite, aragonite and calcite[15]. Solubility of these polymorphs is inversely proportional to their stability. When nucleated in solution in the absence of additives, and when supersaturated for ACC, the formation of the minerals typically transitions from ACC through one or more of the rest of the metastable phases according to the Ostwald-Volmer rule to finally transform into calcite [16]. The case of hydroxyapatite (HAP), a calcium-based biomineral, is similar to calcium carbonate having an amorphous phase as a precursor which transitions to several crystalline polymorphs. There is a large body of literature using synthetic and natural additives to mimic the complex order, phase, shape, size, and orientation of calcium carbonate crystals found in biominerals. These mostly rely on salts, small organic molecules[17–19], polydisperse polymers[20], block copolymers, and natural or engineered proteins or peptides. A commonality between the soft templates reported is the presence of anionic groups, potentially responsible for the binding and concentration of calcium ions and a main driver for crystallization. In this section, we will first review the biological and then the bio-inspired mineralization of calcium carbonate and other calcium biominerals like HAP.

Natural biomineralization proteins are known to nucleate specific $CaCO_3$ polymorphs, but are intrinsically disordered, insoluble, or have no known tertiary structures[21,22]. One example of this can be found in mollusk shells, where highly repetitive, acidic proteins are proposed to be involved in the nucleation and growth of aragonite. X-ray studies have shown a close relationship between aragonite and the β-chitin found in the organic matrix[23,24]. This protein is proposed to self-assemble on the chitin, and create a periodic array interacting with the mineral. A protein with

a known tertiary structure and a hypothesized binding mode to its inorganic partner is osteocalcin. Structural analysis revealed how acidic groups, including **γ**-carboxylated glutamic acid, coordinated a row of calcium ions at the interface of the crystallographic dimer[25]. The arrangement of these calcium ions matched two planes of hydroxyapatite belonging to the family of planes observed in biogenic crystals. Fine details of this proposed structure to function relationship between osteocalcin binding and bone nucleation or formation is not confirmed, but the mechanism suggested can play a big role, together with self-assembly of protein monomers within a larger organic matrix.

With regards to the bio-inspired mineralization of calcium carbonate, the use of rationally designed proteins and peptides has also been explored. At least three different strategies have been employed by the following examples. In the first case, ubiquitin is modified using click chemistry to contain three phosphate groups lined on one side of the protein[26]. When mineralization is induced, the protein creates a macroscopic Polymer-Induced Liquid-Precursor (PILP) film that breaks down into amorphous nanoparticles which then mature into crystals with unique morphology. In the next example, a protein composed of three domains: a chitin-binding, a silk fibroin and a calcium-binding domain, is used to nucleate hollow spheres of vaterite composed by radially distributed nanoparticles and nanoplatelets[27]. In this case, the effect of confining the protein and binding it to the chitin surface, coupled to the structural formation of beta strands of the silk domain and the concomitant formation of an alpha helix of the calcium-binding domain during the mineralization reaction are all parameters that tie the deposition of the mineral with a structural component of the protein. Thirdly, six designed peptides showed a large effect on the shape of growing calcite crystals[12]. Of these low complexity, high flexibility peptide sequences comprising 16 amino acids, only a couple showed clear differences between the effects produced

by the designed peptides compared to the scrambled variants. This meant that potential sequence to structure and function relationships could only be traced to a select few peptides with a net positive charge and positive and negatively-charged residues. Peptides with negative charge seemed to have a stronger effect both in the designed and scrambled variants, leading the researchers to think that charge overrides the specificity that can be conferred in the presence of positive and negative residues. In sum, all of these examples have sought to underline structure to function relationships where chemical compatibility moieties were explored in conjunction with some sort of structuring element: ubiquitin fold for the first example, binding and structuring domains for the second, and helical propensity of a modeled sequence for the final.

1.3.2 Iron oxides with an emphasis on Hematite

Of the known biominerals, 40% are iron based with a diverse set of anions used and polymorphs formed[28]. Iron is both highly abundant in the biosphere and poorly soluble in water[29], which might explain why so many organisms opted to use iron, but had to evolve different ways of making it available and easy to transform. Biogenic iron oxides tend to be highly hydrated and less crystalline, making them more soluble and amenable for transformation by organisms.

At room temperature, at atmospheric oxygen partial pressure, and over a significant range of pH, hematite is the most stable form of the iron oxides. Its crystal structure lacks hydroxylation between the layers of octahedrally-coordinated Fe(III) observed in other iron oxides. These octahedrons are stacked in a corundum structure with two thirds of the octahedral interstices filled with Fe(III)[30]. Nucleation of hematite is considered to occur from an amorphous iron oxide precursor, or it can also be epitaxially-templated by other iron oxides like goethite[31]. To date and to the best of my knowledge, there are no known organisms capable of synthesizing hematite.

Hematite surface chemistry has gained attraction because of its use in photochemical applications such as degradation of organic contaminants and generation of hydrogen gas from water[32,33]. There are three well characterized and naturally occurring iron oxide surfaces (001), (012), and (113). Chatman, Zarzycki, and Rosso[34] modeled and measured the pH dependance of the surface potential for each of the main hematite planes described above. The ratio and distribution of two different kinds of protonation and deprotonation sites is unique to each of these surfaces, and explains their different behavior at a given pH value. Surface chemistry and structure characterization of these surfaces reveal the expected different atomic arrangements and symmetries found in the bulk[35,36], but different chemical terminations and degrees of hydroxylation and protonation when in contact with water[37,38]. Broadly, terminations include vacancies of oxygen atoms and modifications to the number of iron atoms at the top two layers as expected from the bulk structure. These different terminations can have an effect on the reactivity and surface potential.

Regarding natural proteins known to interact with and nucleate iron oxides, we can find Mms6, a protein with an array of negative residues that requires self-assembly into larger structures to create a mineral. This is a membrane-associated protein with evidence linking it with magnetite growth in magnetotactic bacteria[39]. Although there is no molecular structure for the protein, the monomers appear to form micelles in solution where the negatively charged c-terminus is exposed to solvent, and the hydrophobic n-terminus points towards the inside[40]. After following changes in assembly by TEM, DLS, and NMR at different iron oxidation states and concentrations, it was concluded that the protein appears to have preferential binding for Fe(II) and to recruit Fe(III) by simple electrostatic charges[41,42]. The binding and nucleating mechanism proposed involves a disc-shaped protein assembly on the inner walls of a membrane (similar to what would be expected

in the magnetosome), where the spacing of the monomers allows a fine alternating pattern for selective binding of Fe(II)[43]. In the same operon, there are multiple related proteins whose function is closely related, but most of their structures are unknown. The commonality for most of them is the presence of sites for iron binding, iron transport, sites for iron redox, transmembrane domains, and their capability to oligomerize at the interface[44,45].

Regarding biomimetic production of iron oxides, one noteworthy example is that of a protein-induced α-Fe_2O_3 synthesis using a hyperthermostable ferritin cage[46]. Ferritin is a native protein composed of 24 subunits of α-helical proteins forming a hollow nanoparticle with high geometry. Capture and transport of iron through pores in the protein shell, followed by non-specific precipitation in the charged interior, ends with the precipitation of amorphous iron hydroxide as a form of iron storage in natural systems[47]. After the formation of the iron hydroxide, extensive heat treatment of the loaded protein cage lead to the formation of α-Fe_2O_3[46]. This provides some feasibility of nucleation of hematite by proteins, and the opportunity to test the proposed set of interfacial design properties that can mediate protein binding and nucleation of inorganic crystals.

1.4 Protein-inorganic crystal interactions

In this section, different natural and designed interfaces between proteins and inorganic crystals will be introduced. From these descriptions, key design rules that can be incorporated into future idealized systems will be summarized.

1.4.1 Intrinsically disordered proteins

Intrinsically disordered proteins are a class of proteins or protein regions that do not have a defined three-dimensional structure or adopted secondary structure under native conditions and instead exist as an ensemble of structures with no equilibrium atom positions or bond angles[48]. There is a large body of literature studying the effect of intrinsically disordered proteins (IDPs) on the natural and engineered precipitation of crystalline matter[49,50]. This interest arises from the observation that proteins annotated for biomineralization are predicted to have a great degree of disorder (mean of 53%)[51]. Common characteristics for IDPs involved in biomineralization are low complexity, repetitiveness, small molecular weights, highly charged regions, and in some cases the presence of partially structured domains. The general way in which IDPs are thought to participate in nucleation and growth modification of minerals is through structuring after interacting with their binding partners such as ions, other proteins, the inorganic crystal, an organic matrix (e.g. chitosan), or after oligomerization[52].

Disorder in proteins may promote the nucleation of inorganic matter for two reasons. Firstly, by changing between various conformations, it can interact with different binding partners. Secondly, the creation of ensembles that bind and pre-organize ions can encourage precipitation in a way similar to how disordered regions play a role in enzyme catalysis[48,53]. Upon strong binding, these molecules are capable of shaping or nucleating crystals, following the common rule that good binders are good nucleators[21]. Novel bio-inspired synthesis systems have been developed based on this using bio-panning studies where selection using phage display systems yielded peptides or viral capsids capable of nucleating materials such as metal oxides for lithium-oxygen batteries[54] and other architectures by self-assembly and interactions with ions, other organic molecules and nanoparticles[55]. However, these peptides appear to be promiscuous in their binding and nucleating capabilities. Thus, in order to engineer intentional nucleating and binding capabilities of biomaterials, structured domains need to be used in order to produce sequence to structure to function rules capable of imparting specificity.

1.4.2 Ice-binding and ice-nucleating proteins

IBPs are an example of a well-understood relationship between an inorganic crystal and its binding protein partner. X-ray crystallography and ice binding or modulation studies have shown that, independent of the topology, the activity of these proteins seems to be closely related to the binding of water and creation of a polypentagonal semi-clathrate networks in register with one or multiple planes of ice[56–58]. This feat is accomplished by a large, flat, and mostly hydrophobic surface with a sparse array of sidechain or backbone atoms that form hydrogen bonds to water. Several IBPs have β-solenoid repeat protein topologies that space their interacting residues in close registry with the lattice spacing of water molecules on specific planes of ice. This creates a flat, repetitive, semi-rigid structure that allows maximum compatibility with the interacting

inorganic phase and reduces the lattice strain of epitaxially matching the inorganic to the organic component.

INPs are capable of forming ice above the homogeneous ice nucleation temperature. There are no structures of INPs to date, but models draw parallels to IBPs and suggest they share a similar mechanism of action[9]. INPs are thought to mimic within a monomer multiple end-to-end IBP subunits or to be products of IBP oligomerization. This has been explored by modeling studies[8] and limited experimental evidence from ice nucleation experiments suggesting oligomerization as a way of modulating IBP to INP behavior[59].

1.4.3 Peptide self-assembly on carbon nanostructures

Epitaxy is defined as the long range crystalline order of a deposited phase, which is induced through interactions with a crystalline surface[60]. This self-assembly process is guiding reported interactions between peptides and carbon nanomaterials such as nanotubes (Single Walled carbon Nanotubes, or SWNTs) or graphene. One strategy designs interactions between small proteins and SWNTs by first defining interaction elements such as the beta-carbon of alanine on a helix and interacting with the resonant, six-member ring structure of the extended carbon nanotube sheet[61]. This is then followed by placement of an additional, interacting secondary structure element (in this case another alpha helix) in such a way that it samples designable interfaces within the symmetry constraints of a unit cell of the target carbon nanotube chirality. Sequences constrained to repeats with alternating hydrophobic and charged interactions resulted in beta-strands with opposed hydrophobic and charged faces[62]. These formed oligomeric interactions through hydrogen bonding, as well as interactions with graphene by forming nanoribbons along one of the six directions of the graphene's symmetry. Both of these design cases on graphene and SWNTs consider the coherent interactions between the protein and the carbon nanomaterial, as well as the cooperative interactions between the alpha and beta secondary structure elements. To summarize, these interactions are preorganized enough to enable selective binding, but not so inflexible that a perfect fit is necessary for selective recognition to occur.

1.4.4 Interactions between inorganic crystals and designed proteins

Designed proteins have been lattice-matched to the potassium sub-lattice on the surface of muscovite mica (001)[63]. Different assemblies where the protein was aligned with the underlying symmetry of the surface were achieved including nanoribbons formed by end-to-end oligomerization and the formation of honeycomb structures. In another case, a designed beta-propeller pizza protein with an engineered metal site formed a cadmium chloride nanocrystal created at the interface between two trimers of the protein[64]. This is made possible by a coordination motif lying on the symmetry axis of the trimer, which in turn matches the symmetry of the crystal, and by solvent exclusion. Once again, both matching the symmetry of the system and having compatible coordinating moieties precisely positioned are key components to successful design of protein-inorganic crystal interactions.

1.4.5 Lessons learned from designed and natural protein-inorganic crystal interactions

In summary, lessons learned from natural and synthetic protein-inorganic crystal interactions highlight the importance of (i) a binding interface with repetitive domains, (ii) the use of compatible moieties to either create special coordination environments or interaction motifs, (iii) forming favorable intersubunit cooperative interactions, (iv) and most importantly symmetry matching of the target material. This interface is further expanded by oligomerization and self-assembly on membranes or by association with other organic materials or a crystalline matrix. The organic matrix and other elements in the biomineralization system again emphasize the importance of fine tuning solution chemistry by concentrating ions, changing their redox state, and confining them. However, a common limitation is the lack of structural data and/or experimental evidence tying the structure of a protein with specific binding or nucleating of a mineral and showing how this strategy generally applies. Based on these examples we seek to explore if this mechanism could be applicable to direct binding and nucleation of inorganic materials using designed proteins.

Designing proteins that can mimic to some extent the intrinsically disordered repetitive domains described in natural biomineralization systems, and confine them in a regular secondary and tertiary structure, is an alternative to relying less on the multiple conformations these peptides might adopt in solution. Regularity and tunability is the characteristic of designed proteins we would like to explore.

1.5 Protein design

The following section is adapted from Huang et al.[65]. Amino acid sequences, which fold spontaneously from extended polypeptide chains into three-dimensional structures, encode most natural biological functions. Through selective pressure on random variants of primordial proteins over millions of years, the structure to function relationships of these sequences were explored. However, the sequence space for a protein of typical length of 200 amino acids is vast, with 20^{200} possible sequences to explore. Evolutionary processes have only explored a small region of this sequence space through incremental and random means. De novo protein design is the systematic and biophysics-based exploration of this sequence space to generate new proteins, allowing for the discovery of sequences unrelated to those found in nature.

1.5.1 Principles for protein design

Protein design's physical principles are rooted in the hypothesis initially presented by Christian Anfinsen, which posits that native-like protein conformations are unique, thermodynamically stable conformations with low energy accessible through their amino acid sequences[66]. These folded states are located in energy landscape minima and have a net favorable Gibbs free energy change compared to the unfolded state. In order to satisfy this low energy state, some design rules include the burial of hydrophobic residues in the core away from the solvent, minimizing the size of the cavity that the protein occupies in the solvent, and maximizing Van der Waals forces. Regarding hydrogen bonding, polar groups buried in the core must be engaged in them and the formation of secondary structures is guided by interactions between the polar carbonyl and amide groups of the polypeptide backbone. Another key interaction is that of the side chains with backbone atoms, where steric and torsional effects favor certain backbone geometries and hydrogen bonding stabilizes the termini of α-helices.

1.5.2 Navigating the energy landscape using computational protein design

To predict and design proteins, algorithms are needed to sample the space of protein sequences and structures. These are then evaluated using an energy function that approximates the energetics of the system to identify a low-energy sequence and conformation (refer to Section 2.1 for protein design using Rosetta, and Section 2.1.1 for the Rosetta energy function). In contrast to the structure prediction and fixed backbone design problem, de novo design involves an unknown sequence and exact backbone structure. To reduce the search space, protein backbone design tasks are typically limited by using short peptide backbone segments found in the Protein Data Bank (PDB) or by algebraic equations that specify geometry parametrically. These strategies are particularly useful for designing repeat proteins, which have an internal symmetry consisting of a single idealized unit repeated multiple times. As a result, the sampling space is reduced to a small sequence-structure combination that can be reused, and by utilizing parametric sampling, it is possible to exercise remarkable control over the global structure. In summary, the protein shape and energy landscape can be navigated using parametric design of repeat proteins to simplify the sequence and structure search space while biasing towards a target conformation.

1.5.3 Protein-inorganic crystal design principles enabled by protein design

As outlined in section 1.4.1, even though IDPs are commonly involved in the precipitation of inorganic material, structured domains are apparently necessary to establish a sequence-to-structure relationship that enables specific mineral interactions, including nucleation. The desired template characteristics are summarized in preorganized interactions using a flat, repetitive, semi-rigid structure allowing maximum compatibility with the interacting inorganic phase that reduces lattice strain and creates a close symmetry match to the inorganic component. These preorganized interactions need to use compatible moieties in fully or partially coordinated environments. All of these requisites can be fulfilled by designed repeat proteins, as their tertiary structure can be parametrically tuned to be flat, their nature is repetitive, and their high thermostability makes them robust scaffolds for multiple surface mutations that can present intended interactive motifs that can create partially coordinated environments.

1.6 Scope of this work

Our main motivation is to use molecular design to interrogate a set of hypotheses that would be intractable to address with natural biomineralization systems or the engineered proteins to date. We are asking specific questions including: is a structured interface that can pre-organize ions and mimic a crystalline structure required to create a good template?, what set of interactions or surface chemistries are key to binding or nucleation?, and what is the role of interface size and shape? Answering these questions will allow us to better understand biomineralization processes and bring us closer to the rational design of artificial biomineralization systems. Our proposed key innovation in this case is that protein design will allow us to create structured, tunable templates from which we can draw and test design rules. These new systems could enable an alternative method for the bottom-up nanofabrication of materials, and provide opportunities to study how natural biomineralization systems work and find ways to manipulate them. Other long term applications include therapies for biomineralization-related disease and development of new materials for clean energy.

2 Chapter 2: METHODS and their BACKGROUND

This chapter contains background and limited reproduction of content from a manuscript in preparation by Davila-Hernandez, Jin, Pyles et al.[14].

2.1 Protein design with Rosetta

The following section is adapted from Leman et al.[67] Developed in the mid-1990s, Rosetta is a comprehensive framework for computational structural biology that was originally designed to predict protein structure and folding. Other modeling and structure prediction tools and suites that rely on molecular mechanics, molecular dynamics, and quantum mechanics calculations include Schrödinger[68], the Molecular Operating Environment[69], and Discovery Studio[70]. Complementary to Rosetta are molecular dynamics packages like CHARMM[71], AMBER[72], and GROMACS[73], which simulate most atoms explicitly using a physics-based energy function that solves Newton's equations of motion. Additionally, there are various other tools used for specialized tasks, such as de novo modeling. Examples of such tools include AlphaFold[74], QUARK[75], RoseTTAFold[76], and RaptorX[77]. This is by no means an exhaustive list of all the tools and suites used for structural biology modeling.

The Rosetta protocol typically begins by taking a molecular conformation, referred to as a Pose, which is then modified either deterministically or stochastically using a Mover routine. The resulting conformation is then evaluated by a custom energy function described in Section 2.1.1, and its acceptance is determined by the Metropolis criterion based on the computed energy difference between the original and new conformation. Multiple independent trajectories are generated, and the final models are assessed based on the scientific objective. However, there are two main limitations to evaluating structures or groups of structures using Rosetta: sampling and scoring. In terms of scoring, it is assumed that Rosetta's ScoreFunction is a close approximation

of the biophysics of the system. Regarding sampling, it is assumed that the set of conformations includes enough representatives. These concepts are summarized in Fig. 2.1.

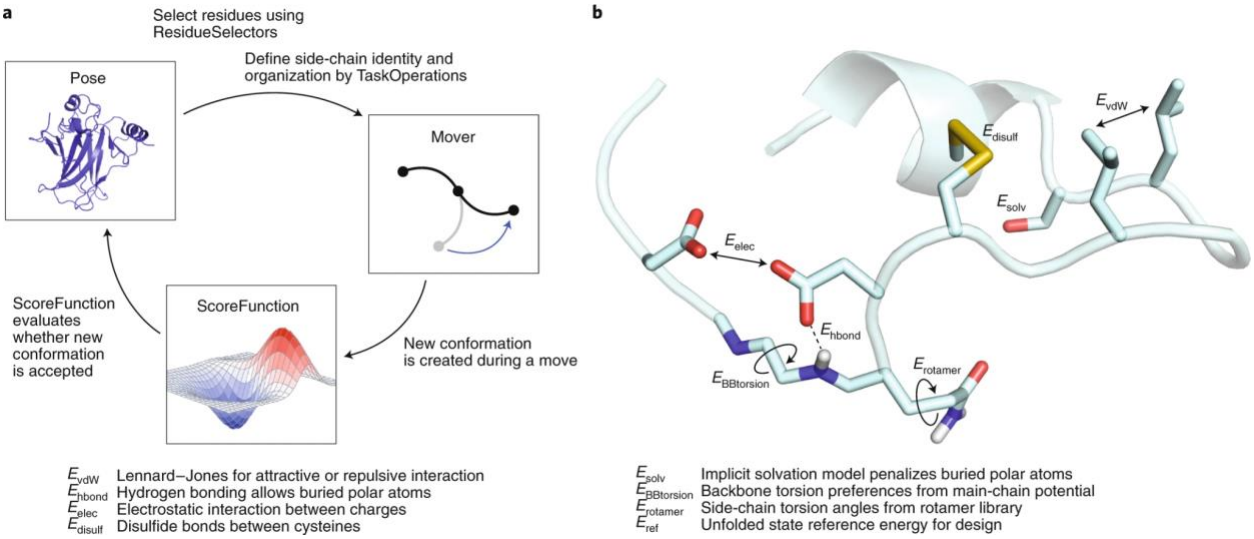

Figure 2.1. The sampling and scoring protocol of the Rosetta Macromolecular Modeling Suite, reproduced from Leman et al.[67]. **(a)** Starting from a conformation, dubbed pose, a protocol called mover is used to modify the conformation which is then evaluated by a custom score function (described below in Section 2.1.1) before further selection or introduction into a new protocol proceeds. **(b and lower panel)** Description of Rosetta's score function energy terms and their illustration using a model protein.

2.1.1 The Rosetta Energy Function

As described in section 1.5.1, proteins achieve their three-dimensional structures based on their amino acid sequence and a balance between enthalpic and entropic factors of non-covalent interactions. Energy functions aim to approximate this process. This section adapts content from Alford et al.[78]. The score function is a linear combination of weighted score terms, which are weighted to balance between statistically derived potentials and physics-based potentials. The total energy (ΔE_{total}) is determined based on the geometric degrees of freedom (Θ_i) and chemical identities (aa_i), and each term is multiplied by a weight (w_i) as indicated in Eq. 2.1.

$$\Delta E_{total} = \sum_i \quad w_i E_i(\Theta_i, aa_i) \tag{2.1}$$

These potentials include van der Waals energies, hydrogen bonds, electrostatics, disulfide bonds, residue solvation, backbone torsion angles, sidechain rotamer energies, and an average reference

energy for the unfolded state (as described in Fig. 2.1). Some energy terms are further broken down, and each component is parametrized separately. Additionally, guidance energy terms are incorporated to either encourage certain features (such as amino acid composition and hydrogen bond networks) or discourage others (such as voids and buried unsatisfied hydrogen bond donors and acceptors) based on the design objective.

The score function has several important limitations. Firstly, it indirectly models entropy by incorporating factors such as rotamer bond angles, solvation, peptide fragments, and pair terms that represent only a half of the entropy at relevant temperatures and solvation densities. Additionally, it is a knowledge-based function that relies on crystal structures, so it cannot capture the full range of biologically relevant dynamic structures, and it is less generalizable and understandable than a physics-based model. Finally, the implicit solvation model means that it cannot explicitly model ions, water molecules, or lipid environments.

2.1.2 Design of helical repeat proteins with Rosetta

The design of flat, helical repetitive proteins which would fulfill the requirements of an ideal template according to what is discussed in Sections 1.5.3, can be achieved through either fragment-based assembly with biased superhelical parameters, or by parametric backbone generation followed by translation of these helical elements on the XY plane. These are described below.

For the fragment-based assembly with biased superhelical parameters, the protocol used is adapted from Brunette et al.[79], and Hicks et al.[80], where initial structures were generated using RosettaRemodel followed by scoring with a coarse-grained function weighted with target helical parameters. We tested a number of combinations of sampled fragment lengths of helices and loops to then determine that the best was to sample helices of equal length or with a difference in length

within three residues. Helix lengths of 16 to 30 residues were sampled. The helical parameters were biased to have a radius of 500 nm, a rise of 0 Å and a curvature of 0 rad, yielding a protein with a flat, repetitive surface. Inter-repeat distance was biased to either 8.7 or 10.9 Å using harmonic constraints and allowing for a deviation of 0.05 Å. However, no satisfactory solutions were found for the 8.7 Å spacing. Sequence design was carried out using FastDesign with an enforced inner repeat symmetry and enforced interhelical repeat spacing preventing the backbone minimization cycles to create scaffolds with spacings deviating from the target. Residues on one side of the surface of the protein were mutated to glutamates either manually or by using a custom PyRosetta protocol.

In the case of parametric generation of helical bundles followed by translation on the XY plane, the design method is reported by Huddy et al.[81] and is briefly described below: A helical secondary structure element is placed along additional secondary structure elements that will be part of the repeating unit using BundleGridSampler mover in RosettaScripts. The rigid body transformation for the repeat propagation is set by translating a copy of the original helix; copies are added as needed. Degrees of freedom are limited to the helix phase, displacement of repeating helices on the XY plane, and change in height between adjacent helices. In the case of this molecule, the inter-repeat distance was set to 8.7 Å. Sequence design was carried out using FastDesign with an enforced inner repeat symmetry. Residues on one side of the surface of the protein were mutated manually or using a custom PyRosetta protocol to glutamates.

The top protein designs were selected based on various factors such as protein energetics, core packing, secondary structure shape complementarity, helix quality, number of secondary structures in contact, and buried unsatisfied hydrogen bonds. These designs were then subjected to forward folding simulations to identify those that exhibited a funnel shape leading towards a set

of low energy relaxed models near the original model, as evidenced by a low root-mean-square deviation (RMSD). Another metric, the area to the left of the folding funnel from the lowest energy point to +8 rosetta energy units, may also be used to select designs with a value less than 25, as previously published[80,82].

The protein named DHR49-Neg was originally named DHR49 in Brunette et al.[79] and was resurfaced manually to contain aspartates and glutamates on one side of its surface. In addition, the inner repeat sequence was conserved at the termini, creating the possibility to create end-to-end oligomers.

2.1.3 "Flattening" of scaffolds

To create idealized versions of the original protein designs that are perfectly flat and have a set inter-repeat spacing the following protocol was implemented in Rosetta. In general terms, this protocol performed monte carlo sampling of torsion angles that are propagated according to the repeat symmetry while enforcing constraints. These constraints ensure that first, there is a given distance between helices (harmonic constraints that are set between carbon-alpha atoms in each repeat). Then, angle constraints between the termini and the center repeats ensure that there is no curvature (to approximate 180º). Finally dihedral constraints on the termini of the protein ensure that there is no twist along the other axis of the protein, this is translated in practice to a dihedral angle close to 0º between two atoms on the first repeat and two atoms on the final repeat. Scoring was then performed by BetaNov16 score function without weights enforcing the constraints.

2.1.4 Docking of scaffolds on different surfaces of calcite

The flattened, idealized models were then taken to perform docking simulations. A simulation starts by initializing a pose with the interacting surface facing the mineral surface, and creating a random rotation in 360º. The protein is then brought into contact with the surface while

minimizing by small rigid body movements the docked configuration. This is then fed to Monte Carlo simulated annealing protocol that iteratively minimizes the protein rigid body movements while sampling the side chain configuration at the interface. The side chains are constrained in such a way that the same configuration is chosen at every equivalent repeat position and thus gives us an idea of likely interactions that are repeated and idealized by the interacting scaffold. Finally, the binding energy was measured as ddG calculated in Rosetta Energy Units.

2.2 Experimental testing of the designed proteins

2.2.1 Protein expression and purification

Genes encoding the designs were then ordered through Genscript. Constructs with a N-terminal His6-tag followed by a TEV Cleavage site were cloned into either pET-28b+ or pet21b between NdeI and XhoI sites. An additional flexible linker with a tryptophan was added to help with protein concentration determination by absorbance at 280 nm. The cloned genes were transformed into either Lemo21(DE3) E. coli from New England Biolabs (NEB) or in BLR(DE3) E. coli cells from Novagen. Expression then proceeded for 24 hours at 37 ºC using 0.5 L cultures in 2L flasks using Studiers M2 autoinduction media with 50 µg/mL kanamycin or 50 µg/mL carbenicillin for pET-28b+ or pet21b, respectively. Cells were pelleted at 4000 g for 30 minutes at 12 ºC, then resuspended in ~40 mL of lysis buffer (20 mM Tris, 500 mM NaCl, 30 mM Imidazole, 0.25% Chaps, 1mM PMSF, 1 mg/mL DNAse, pH 8) and finally lysed after homogenization using a microfluidizer (Microfluidics M110P) at 18K pounds force per square inch. The lysate was clarified at 24000g for 30 minutes at 12 ºC, and the soluble fraction was filtered through 0.7 µm syringe filters and set to do overnight batch binding with 1.5 mL of Ni-NTA resin (Qiagen) equilibrated in wash buffer (20 mM Tris, 500 mM NaCl, 30 mM Imidazole, 0.25% Chaps, 5% glycerol, pH 8). This was then transferred to a gravity column and washed with

25 mL of wash buffer before elution in 3 mL of elution buffer (20 mM Tris, 500 mM NaCl, 500 mM Imidazole, 0.25% Chaps, 5% glycerol, pH 8). Eluate was then dialysed in 3.5 kDa molecular weight cut-off dialysis cassettes (Thermo) into 5 L of TEV cleavage buffer (50 mM Tris, 50 mM NaCl, pH 8) three times before starting overnight cleavage by adding TEV protease in a ratio of 1 mg for each 25 mg of tagged protein. Secondary IMAC was carried out to remove the TEV protease and uncleaved product. The flowthrough was collected for fractionation by size exclusion chromatography. with an AKTA pure chromatography system on a Superdex 200 Increase 10/300 GL column in TBS (20 mM Tris pH 8.0, 100 mM NaCl). The purified proteins were then dialyzed into MOPS buffer (10 mM) adjusted to pH 7 in the case of the proteins used for calcium carbonate nucleation, or into 20 mM Tris at pH either 7 or 8 for the hematite-binding experiments. Dialysis was carried out three times overnight with a dialysis ratio of 1:10000 volume each time and aliquots of 20 µl were snap-frozen for long-term storage.

2.2.2 Measuring protein concentration

Absorbance at 280 nm wavelength of 2 µl of protein samples was measured using a Nanodrop 8000 spectrometer (Thermo Scientific). The concentration was then calculated based on the measured absorbance and the known extinction coefficient following the Beer-Lambert law.

2.2.3 Circular dichroism

Circular dichroism was used to determine the helical content or secondary structure of the designed proteins as well as their thermal stability. This was achieved using a Jasco J-1500 CD spectrometer. Measurements were taken on a sample with a concentration of 0.3 mg/mL in 20 mM Tris pH 8 and 100 mM NaCl, using a 1 mm path length cuvette. The raw CD signal was divided by N x C x L x 10 to convert it to mean residue ellipticity, where N is the number of residues, C is

the protein concentration, and L is the path length of 0.1 cm. Thermal melts were by measuring spectra at 95 ºC.

2.2.4 Small Angle X-ray Scattering (SAXS)

To confirm the tertiary structure and quaternary structure of the designed models, SAXS measurements were performed. The SIBYLS High Throughput SAXS Advanced Light Source in Berkeley, California[83]was used to collect data. Each sample was exposed to the beam for 0.3 s for a total of 10.2 s, which resulted in 33 frames per sample. The data was collected at both low (~1 mg/mL) and high (~5 mg/mL) protein concentrations in SAXS buffer (25mM Tris pH 8.0, 150mM NaCl, 2% glycerol). The SIBYLS website's "SAXS FrameSlice" tool[84] was used to analyze the data for the high and low concentration samples and determine the best dataset. If there was evidence of aggregation in any of the 33 frames, only the data points before aggregation occurred were used in the Gunier region; otherwise, all data was included. All data was used for Porod and Wide regions. The resulting dataset was then compared to the predicted SAXS profile generated from the design model using the FoxS SAXS server[85].

2.2.5 Size Exclusion Chromatography with Multiple Angle Light Scattering (SEC-MALS)

After the initial SEC run, the purified samples were combined and adjusted to a final concentration of 2 mg/mL. A high-performance liquid chromatography system (Agilent) with a Superdex 200 10/300 GL column was used to run 100 uL of each sample. A multi-angle light scattering detector (Wyatt) was integrated with the fractionation runs to obtain the absolute molecular weights for each of the designed proteins, as previously described[86].

2.2.6 Native Mass Spectrometry (nMS)

nMS experiments were carried out as described before[63]. To analyze sample purity and integrity, an UltiMate 3000 RSLC (Thermo Fisher Scientific) coupled with an Exactive Plus EMR

Orbitrap instrument (Thermo Fisher Scientific) was used with a modified quadrupole mass filter and surface-induced dissociation[87]. The procedure involved injecting 5 µl of 8µM protein in TBS and on-line buffer exchanging it to 200 mM ammonium acetate, pH 6.8 using a self-packed buffer exchange column[88] (P6 polyacrylamide gel, BioRad) at a flow rate of 100 µl per minute. Mass spectra were then recorded, and UniDec version 2.6.5[89] was used to deconvolute the spectra with the expected masses calculated using the NIST Mass and Fragment Calculator v1.3242[90]. The injection time was set to 200 ms, and voltages were adjusted to minimize unintentional ion activation while allowing ion transmission. The recorded mass spectra covered a range of 1,000–12,000 m/z with a resolution of 8,750 defined at 200 m/z, with sample mass taken every 1 Da, and a peak full-width at half-maximum of 1 Thompson using a Gaussian peak shape function.

2.2.7 Crystallography

Crystallization experiments were conducted using the sitting drop vapor diffusion method and crystallization trials were set up in 200 nL drops using the 96-well plate format at 20 ˚C. Crystallization plates were set up using a Mosquito LCP from SPT Labtech, then imaged using UVEX microscopes and UVEX PS-256 from JAN Scientific. Diffraction quality crystals formed in 0.1 M TRIS pH 6.5, and 25% w/v Polyethylene glycol 3,350. Crystals were flash frozen in liquid nitrogen before sending them to the synchrotron.

Diffraction data was collected at the Advanced Photon Source beamline on 24-ID-C. X-ray intensities and data reduction were evaluated and integrated using XDS[91] and merged/scaled using Pointless/Aimless in the CCP4 program suite[92]. Structure determination and refinement starting phases were obtained by molecular replacement using Phaser[93] using the designed model for the structures. Following molecular replacement, the models were improved using

phenix.autobuild[94]; efforts were made to reduce model bias by setting rebuild-in-place to false, and using simulated annealing and prime-and-switch phasing. Structures were refined in Phenix[94]. Model building was performed using COOT[95].

2.3 $CaCO_3$ crystallization experiments

Sections 2.5 & 2.7 describe the experimental setup for the characterization of the nucleation of $CaCO_3$ and the interactions with $CaCl_2$ in the presence of proteins. In a typical crystallization experiment, 0.5 ml 10 mM $CaCl_2$ was mixed with 10 mM $NaHCO_3$ as a control group. In the experimental group, we mixed 10 mM $CaCl_2$ with 2 μM of protein first and then added 10 mM NaHCO3 to initiate the nucleation and growth of $CaCO_3$.

2.4 Hematite nucleation experiments

Ferrihydrite was prepared by adjusting a 0.02 M $FeCl_3$ solution to pH 7. This was then incubated at 80 ºC for 15 hours in the presence or absence of protein. Incubation with the protein was done either directly or separated by a membrane. The membrane-separated setup incubated a 2 mL reaction containing ferrihydrite in the outer chamber while the inner chamber retained either hematite nanoparticles, or protein. Ex-situ TEM characterization was performed as described in Section 2.5.1.

2.5 Transmission Electron Microscopy

2.5.1 Ex situ TEM

Ex-situ TEM samples were prepared by dropping time-resolved 0.6 μl reaction solutions on a carbon-coated Cu-grid (300 mesh) purchased from Ted Pella which was treated by a plasma cleaner. TEM tests were performed in a FEI Titan ETEM 80–300 kV operated at 300 kV.

2.5.2 Cryo TEM

Cryo-TEM experiments were performed in a Titan ETEM 80–300 kV. Prior to the vitrification procedure, a pure lacey carbon grid is surface plasma-treated to make it hydrophilic. Using an automated vitrification robot (FEI Vitrobot Mark III, blot time: 3s), a 3.0 μl mixed aqueous solution including 5 mM $CaCl_2$ and proteins as well as 5 mM $NaHCO_3$ was loaded onto a grid and plunged into liquid ethane. The frozen sample was saved and transferred in a cryogenic holder and cryo-TEM (FEI ETEM operated at 300 kV) under liquid nitrogen conditions, and thus the microscope maintained its temperature near −192 °C throughout the experiment.

2.5.3 Liquid-phase TEM

All of the liquid-cell chips (Hummingbird Scientific), consisting of two square 4 mm^2 silicon chips with 50 nm thick silicon nitride (Si3N4) membranes in 50 × 200 μm^2 windows for imaging, were plasma cleaned in a Plasma Cleaner (Harrick Plasma) for 2 min before use. In a typical experiment, 0.3 μl 10 mM $CaCl_2$ containing 2.16 μM protein was dropped onto the bottom chip, followed by adding 0.3 μl 10 mM $NaHCO_3$ solution, and finally sealed the reaction solution using a window chip. The sealed chips were assembled inside the liquid cell holder (Hummingbird Scientific) and did leak-checking. After that, the holder was immediately inserted into the TEM for observation within ~5 min. TEM was carried out in a field emission Titan ETEM 80–300 kV (Thermo Fisher Scientific) operated at 300 kV. TEM images were acquired using an Eagle CCD (1,024 × 1,024 pixels). In situ movies were recorded using free software called Camstudio for Screen and Video Recorder. All images from in situ movies were processed using the open-source software ImageJ. To minimize beam effects, a low electron dose rate (~100 $e/nm^2 \cdot s$) was used to observe the formation process by adjusting the condenser aperture size (50 μm), and spot size (3).

2.6 UV-Vis and DLS

The $CaCl_2$ incubated protein solution was measured by UV-Vis spectrophotometer (Ultrospec 2100 pro) and dynamic light scattering (DLS, using a Malvern Zetasizer) to evaluate the protein-Ca interactions.

2.7 Attenuated Total Reflection Fourier Transform Infrared Spectroscopy

The crystallization process of $CaCO_3$ was monitored in situ using a Bruker LUMOS II Fourier transform infrared (FTIR) spectrometer with Superior μ-attenuated total reflectance (ATR) FT-IR Capabilities. The retractable diamond crystal is controlled by a high-precision piezoelectric motor and it is integrated into the lens, which allows us to control the detector location exactly in the reaction solution. The crystallization experiments were initiated by adding 200 μl 10 mM $NaHCO_3$ into 200 μl 10 mM $CaCl_2$ with or without DHR proteins. For each FTIR spectrum recorded, 8 scans were carried out at a resolution of 2 cm^{-1} using H_2O as background. The first spectrum was recorded ~5 s after mixing the solutions.

2.8 Atomic Force Microscopy

2.8.1 AFM for the characterization of protein binding to different hematite surfaces at different pH values

Hematite substrates were sonicated in water and 2-propanol for 10 min, respectively. Then the substrates were annealed from 1000 °C. The protein stock solution was diluted to a desired concentration with 20 mM Tris buffer. 50 μl diluted protein solution was incubated on annealed hematite substrate for 30 min before AFM imaging. The AFM characterization was done on MultiMode VIII AFM (Bruker, CA) with PeakForce Tapping mode and ScanAsyst-Air (Bruker, CA) probes. The data processing was done using SPIP software (Image Metrology, Denmark).

Nuclease-free water was bought from Ambion. Tris buffer (pH 7 or pH 8) was bought from Sigma-Aldrich.

2.8.2 AFM for the characterization of protein-protein and protein-calcium complexes

For DHR49-Neg, the protein stock solution was diluted to 0.5 μM in a 20 mM Tris buffer with 3M KCl. Then 100 μl diluted protein solution was incubated on freshly cleaved mica for 30 min. For FD31, the protein stock solution was diluted to 1.0 μM in nuclease-free water with 5 mM CaCl2. Then 100 μl diluted protein solution was incubated on freshly cleaved mica for 10 min. The as-assembled proteins on mica were imaged using Cypher-ES AFM (Asylum Research, CA) in a 20 mM Tris buffer with 3M KCl, and 5 mM CaCl2 , respectively. The amplitude modulation mode and SNL-10-A probe (Bruker, CA) were used in the AFM experiments. The data processing was done with SPIP software (Image Metrology, Denmark). Nuclease-free water was bought from Ambion. Tris buffer (pH 7) and KCl were bought from Sigma-Aldrich. Mica was bought from Ted Pella, CA.

3 Chapter 3: DIRECTING POLYMORPH SPECIFIC CALCIUM CARBONATE FORMATION WITH DE NOVO PROTEIN TEMPLATES

The material in this chapter is partially reproduced from a manuscript in preparation by Davila-Hernandez, Jin, Pyles et al.[14].

3.1 Introduction and protein design principles

Natural proteins nucleate specific $CaCO_3$ polymorphs[96–99], but none have known stable tertiary structures and many are intrinsically disordered or insoluble[21]. Additives including small organic molecules[17–19], polymers[20], amino acids[100], and peptides[12,101] affect $CaCO_3$ crystallization, but have not been shown to template nucleation. Stereochemical matched self-assembled monolayers (SAM)[102] modulate the nucleation of $CaCO_3$, suggesting that geometric lattice matching may provide a general route to controlling mineralization[103,104]; [105,106]. In further support of this concept, the structures of ice binding proteins contain surfaces with an epitaxial-like lattice matching to the ice lattice[107] which enables modulation of ice formation by binding ice nuclei through preorganized ice-like waters.

Guided by the ice binding protein example, we hypothesized that designed proteins with flat surfaces displaying functional groups lattice matched to a mineral of interest could be used to modulate mineralization by promoting ion association in a pattern consistent with the mineral lattice, thus directly reducing the free energy of formation of the critical nucleus (As described in detail in Section 1.2). Advances in protein design now enable the creation of new proteins with precisely specified structures, and in previous work we showed that lattice matched interactions between arrays of carboxylate side chains on a designed protein surface and preformed mica crystals direct formation of precise protein-mica assemblies[63].

As a first test of the potential of designed proteins for directly nucleating inorganic mineralization, we chose $CaCO_3$ as a model system due to its numerous polymorphs, well documented nucleation behavior with and without proteins[108], and prevalence in natural biominerals. We then sought to design proteins with structures suitable for nucleating crystalline $CaCO_3$. We aimed to generate flat repetitive surface arrays of calcium-coordinating carboxylate groups (Fig. 3.1a) that could bind to and stabilize $CaCO_3$ nuclei (Fig. 3.1b). To enable exploration of the contribution of the size of the protein-mineral interface keeping the chemical composition and geometry otherwise constant, we employed Designed Helical Repeat (DHR; [79]) proteins that can be shrunk or extended simply by changing the number of repeating units. These proteins have >10 nm^2 flat surfaces with comparable or better structural definition and superior tunability relative to SAMs, but within individual genetically encoded soluble molecules that can be readily reprogrammed.

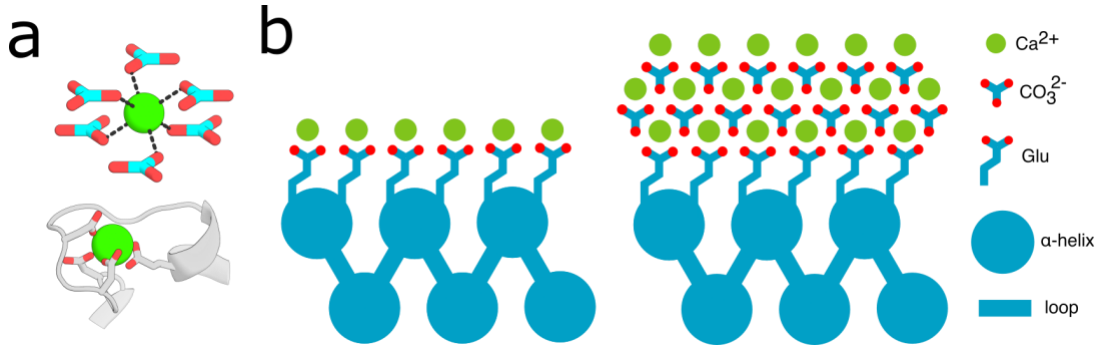

Figure 3.1. Design principles for the heterogeneous nucleation of calcium carbonate guided by a protein template. **(a)** Coordination of calcium (Top) by carbonate ions within a unit cell of calcite, and (Bottom) by carboxylate containing glutamate and aspartate residues in the structure of calmodulin (PDB ID: 1A29, residues 18-33). **(b)** Tessellating binding moieties across repeated α-helices within a designed protein capable of pre-organizing bound calcium ions.

These flat repeat helix-turn-helix-turn protein backbones were designed according to what is described in Section 2.1.2 and were characterized according to the methodology described in Section 2.2. In contrast to typical designed protein surfaces containing a mix of positive, negative, and non-charged hydrophilic residues to promote solubility, and to hydrophobic protein surfaces

designed to bind other proteins, we designed surfaces consisting entirely of carboxylate side chains spaced at regular intervals to mimic the carbonate sublattice on various crystallographic planes of $CaCO_3$ crystals as seen in Fig. 3.1b.

3.2 Results

3.2.1 Testing and biophysical characterization of the designed proteins

We selected 40 designs for experimental characterization and obtained synthetic genes encoding them. Eight of these expressed at high levels in E. coli and were highly soluble. Two designs, FD15 and FD31, were selected for detailed characterization. Size exclusion chromatography (SEC) showed that both proteins were monodisperse in solution with molecular weights expected for the protein monomer (Fig. 3.2a,d,g). Circular dichroism (CD) experiments showed the expected all alpha-helical structure, and the proteins were found to be hyperstable, remaining folded at temperatures up to 95 ºC (Fig. 3.2b,e,h). Small angle X-ray scattering (SAXS) profiles of FD31 were consistent with profiles computed for the design model (Fig. 3.2c). The crystal structure of FD15 was solved at 3 Å resolution and matched the design model with an RMSD of 0.45 Å (Fig. 3.2i). This perfectly straight repeat protein possesses a 8.7 Å spacing between helices in adjacent repeat subunits. We also designed an array of carboxylate side chains onto the surface of DHR49, a previously designed flat DHR [79]. In order to allow protein-protein assembly, as in some native biomineralization proteins[40,109], capping elements on DHR49's terminal repeats were removed to produce head-to-tail protein-protein interfaces. AFM experiments suggest that this protein (DHR49-Neg) forms fibers through designed head-to-tail interfaces at the mica-liquid interface (Fig. 3.2f), but some monomeric form seems to prevail in solution (Fig. 3.2d).

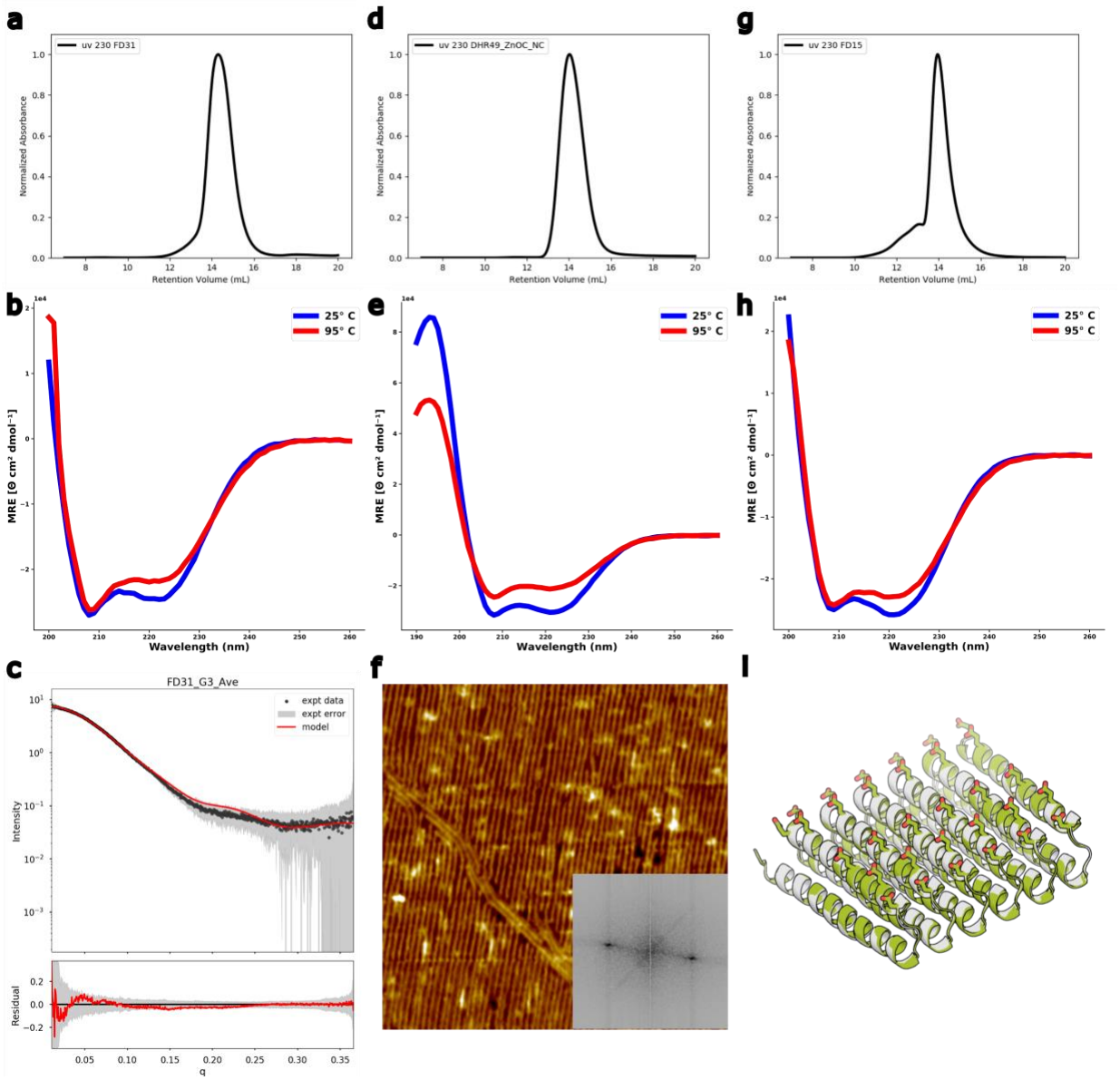

Figure 3.2. Biophysical characterization of designed proteins. **(a, b ,c)** Characterization of FD31, **(d, e, f)** DHR49-Neg, **(g, h, i)** and FD15. First row shows the monomeric, monodisperse behavior of the protein by Size Exclusion Chromatography traces and the second row shows that the protein is hyperthermostable by Circular Dichroism thermal melt traces. **(c)** Small Angle X-ray Scattering confirms the quaternary structure of FD31. **(f)** Atomic Force Microscopy on the surface of mica shows the formation of end-to-end oligomers. **(i)** The crystal structure of FD15 in gray overlaid on the design model in green (solved at 3 Å resolution, RMSD of 0.45 Å).

3.2.2 Mineralization experiments of CaCO3 in the presence of DHR proteins

We first investigated the effect of the designed proteins on calcium carbonate mineralization using in-situ liquid-phase attenuated total reflectance-Fourier transform infrared

spectroscopy (ATR-FTIR) and time-dependent ex-situ transmission electron microscopy (TEM) (Fig. 3.3). A mineralization solution of 5 mM $CaCl_2$ and 5 mM $NaHCO_3$, which is supersaturated for both vaterite and calcite but undersaturated for amorphous calcium carbonate (ACC), was prepared with and without 1 μM protein. Vaterite appeared as the first solid phase in the protein-free mineralization solution or in the presence of FD15 or bovine serum albumin (BSA) (Fig. 3.3c,k,l), which grew to hundreds of nm diameter spherical particles. In contrast, in the presence of DHR49-Neg predominantly calcite nanocrystals ~5 nm in size were initially observed (Fig. 3.3f). TEM examination showed that FD31 even more effectively promoted the direct nucleation of nanocrystalline calcite (Fig. 3.3i), which was supported by selected area electron diffraction (SAED) (Fig. S3a), cryo-TEM (Fig. S3b), and high-resolution TEM (HR-TEM) (Fig. S3c). In contrast, FD15 did not alter the crystallization pathway from that seen with the negative controls. Given that FD15, FD31, and DHR49-Neg all have flat surfaces with regularly arrayed carboxylate groups, these differences likely stem from differences in repeat spacing, net charge, number of carboxylate groups, ratio of aspartate to glutamate side chains, solvent accessible surface area, and surface hydrophobicity (Table 3.1).

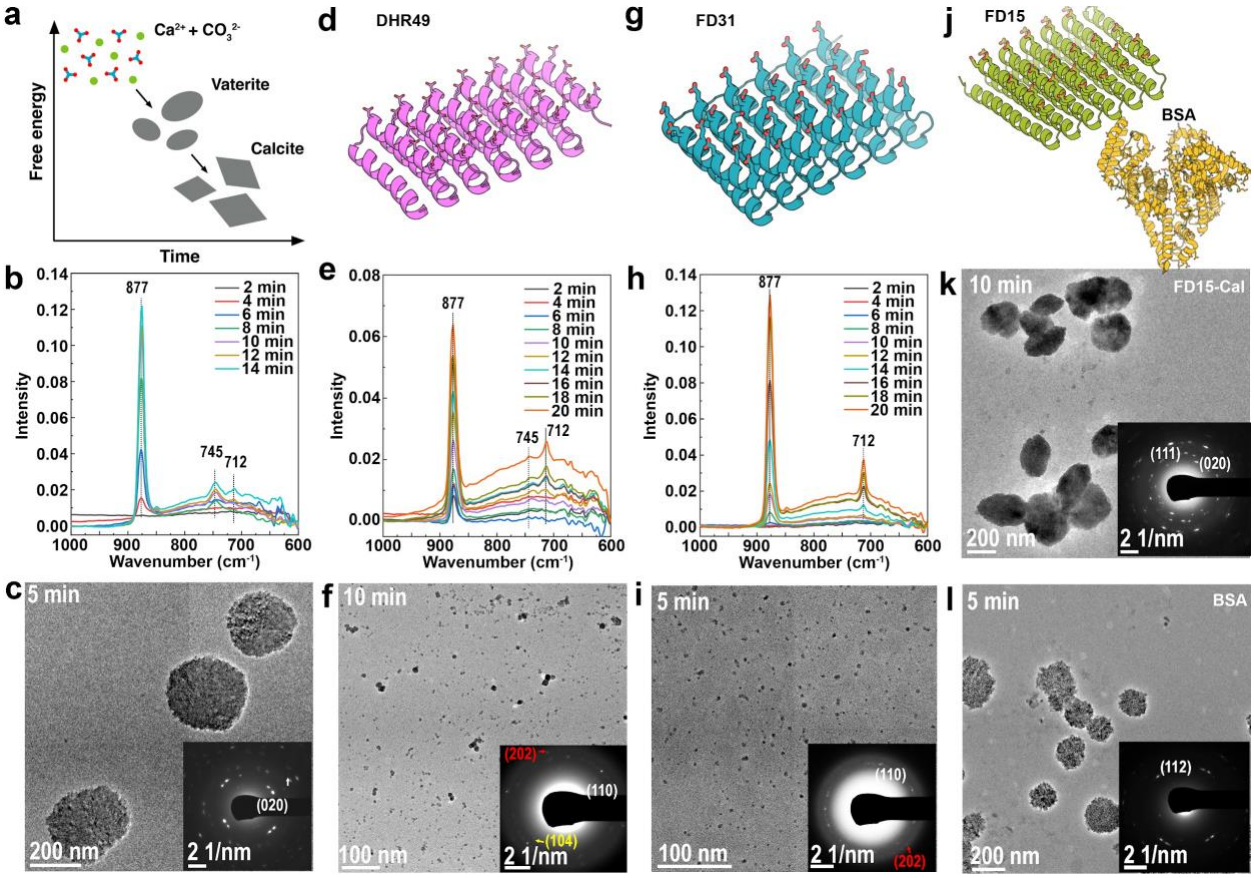

Figure 3.3. Designed proteins modulate CaCO3 crystallization. The CaCO3 crystallization process in supersaturated solutions containing 5 mM $CaCl_2$ and 5 mM $NaHCO_3$ was monitored in the absence and presence of different designed proteins. **(a)** Schematic showing the nucleation and transformation of calcium carbonate in the absence of additives. First, vaterite forms followed by transformation to micron-sized calcite [110]. **(b)** in situ ATR-FTIR and **(c)** TEM of $CaCO_3$ crystallization in the absence of protein showing formation of vaterite. **(d-l)**, Impact of 1μM designed proteins on mineralization. Protein design models **(d,g)** and corresponding in situ ATR-FTIR **(e,h)** showing formation of predominantly calcite in the presence of DHR49-Neg, and exclusively calcite in the presence of FD31. Representative TEM images confirm the formation of **(c)** vaterite in the absence of the proteins, **(f)** primarily calcite crystals with DHR49-Neg, and **(i)** calcite nanocrystals in the presence of FD31. In contrast, in the presence of **(j)** FD15 or a BSA control **(j)**, only vaterite is formed **(k,l)**.

Table 3.1 Biochemical and structural properties of the proteins tested in this study.

Name	Size (nm) (l x w x h)	Repeat Spacing (nm)	Net Charge	Number of carboxylate groups	Number of carboxylates at the interface	Ratio of aspartate to glutamate	Solvent accessible surface area (nm^2)	Percent surface hydrophobic
BSA	8.3 x 7.4 x 6.1	n.a.	-18	98	n.a.	39/59	279.33	57.8
FD15	6.4 x 5.5 x 1.8	0.87	-32	67	30	10/57	154.54	51.5
DHR49-Neg	6.7 x 3.7 x 2.3	0.97	-54	72	42	30/42	115.39	46.4
FD31	7.9 x 4.6 x 2.4	1.12	-47	81	36	5/76	150.93	49.3
FD31-Rep3	4.6 x 4.6 x 2.4	1.12	-26	45	18	2/43	89.93	50.1
FD31-Rep9	11.2 x 4.6 x 2.4	1.12	-68	117	54	8/109	220.13	47.4
FD31-Gln-Chkr	7.9 x 4.6 x 2.4	1.12	-29	63	18	5/58	152.12	48.9
FD31-Asp	7.9 x 4.6 x 2.4	1.12	-47	81	36	29/52	148.77	48.7
FD31-Lys-Chkr	7.9 x 4.6 x 2.4	1.12	-23	69	24	5/64	155.38	50.2

3.2.3 Nucleation pathways in the presence of FD31

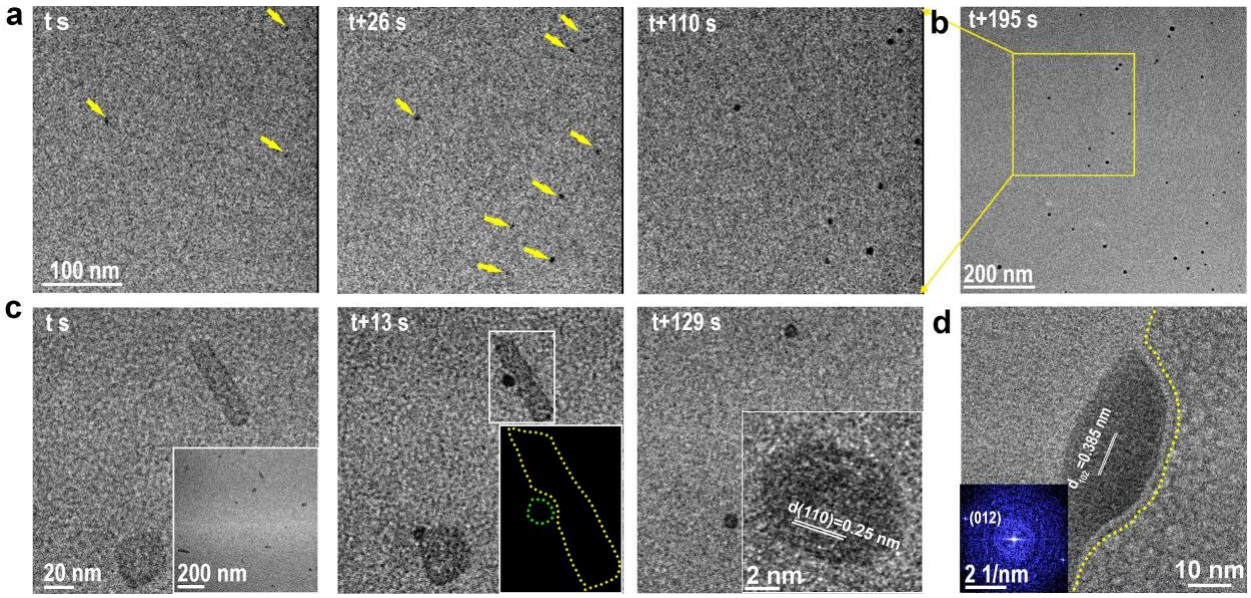

Figure 3.4. FD31 mediated nucleation of calcite nanocrystals. **(a)** Sequential LP-TEM images showing the nucleation of eight $CaCO_3$ nanoparticles in precursor solution containing FD31 protein. **(b)** Low-mag LP-TEM image shows more nucleated nanoparticles. **(c)** Sequential LP-TEM images showing the initial nucleation and growth of two $CaCO_3$ nanoparticles around the sheet-like templates after FD31 was preincubated with $CaCl_2$ for 5 minutes before addition of $NaHCO_3$. Inset at t s displays multiple template formation at an early stage. A schematic (Inset at t+13 s) shows the disconnected interface with a width of 1.2 nm between the as-formed particle and the substrate. At t+129 s, the protein assembly disappears, likely due to depletion of Ca, and the protein disperses into solution. Inset at t+129 s shows LP-HR-TEM image of newly formed particle with the {110} lattice distance of calcite. **(d)** In situ LP-HR-TEM image reveals the interface structure between calcite and the template. The FFT image in the inset shows the nucleated particle is calcite. As in Fig 2, FD31 protein is at 1 μM and $CaCl_2$ and $NaHCO_3$ at 5 mM.

To investigate the mechanism underlying these nucleation effects we focused on FD31, which most effectively promoted direct formation of nano-calcite. Supersaturated solutions containing 1 μM protein were sealed into a liquid cell for LP-TEM observations. Two nucleation pathways were observed. In the first, calcite nanocrystals nucleated throughout the solution (Fig. 3.4a), reaching a diameter of ~7 nm and a number density comparable to that of the protein monomers (~1.6E15 vs ~6.5E14). Due to their low contrast and small size, individual protein monomers are not observable with LP-TEM, however AFM indicates their presence in solution

(Fig. 3.6e). The calcite nanoparticles were also observed to be already present when the electron beam was moved to neighboring regions and imaging was performed at lower magnification (Fig. 3b), indicating that electron beam effects were not driving the process. Given that FD31 is 8 nm by 5 nm by 3 nm in size, which is similar to the 1~5 nm critical nucleus size of $CaCO_3$ [111,112], the size and number density of calcite nanocrystals suggests nucleation is driven by individual proteins. Further support for protein monomer-driven nucleation comes from the observation that as protein concentration is increased from 0.5 μM to 1 μM and 2 μM there is an increase in the number density of nano-calcite particles (Fig. 3.5a-c).

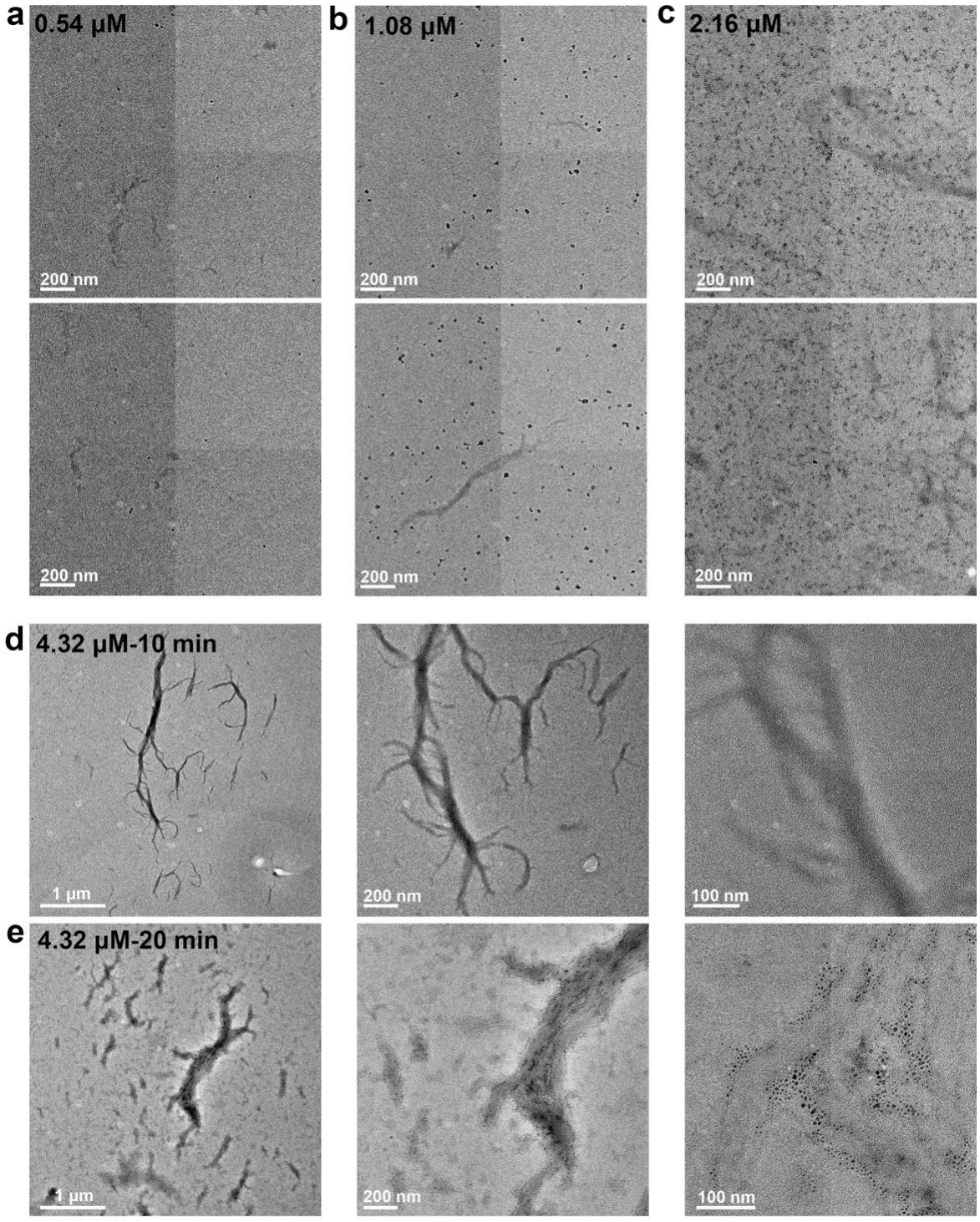

Figure 3.5. FD31 concentration effects on CaCO3 nucleation in supersaturated solutions containing 5 mM $CaCl_2$ and 5 mM $NaHCO_3$. FD31 was incubated with 10 mM $CaCl_2$ for 5 minutes prior to addition of an equal volume of 10 mM NaHCO3. TEM samples were prepared 10-20 minutes after addition of $NaHCO_3$. **(a-c)** Representative TEM images show the resulting $CaCO_3$ particles on copper grids in the presence of 0.54, 1.08, 2.16 μM FD31 proteins. **(d-e)** TEM images show the formation of CaCO3 nanoparticles on pre-assembled calcium-protein templates.

We observed a second nucleation pathway driven by sheet-like assemblies when FD31 was incubated with Ca^{2+} prior to addition of $NaHCO_3$. (Fig. 3.4c; Fig. 3.5d-e). Preincubation of 2 μM FD31 with 10 mM Ca^{2+} in the absence of HCO_3^-, led to the formation of supramolecular assemblies (inset in Fig. 3.4c). Upon addition of an equal volume of 10 mM $NaHCO_3$ at 5 mins, we observed the nucleation of nanoparticles (Fig. 3.4c) identified as calcite by in situ high resolution (HR)-TEM (Inset in Fig. 3c). The initial radial growth rate of the calcite nanocrystals is ~6 nm/s. As the calcite particles continue to grow, the protein-Ca^{2+} supramolecular assemblies dissolve (Fig. 3.4c). Incubation of FD31 with $CaCl_2$ for 30 mins to 24 hours shows coordination of Ca^{+2} by carboxylate groups as evidenced by the UV-Vis peak shift (Fig 3.6a,b) accompanied by formation of assemblies, which were found to be crystalline as per beamline XRD and HR-TEM (Fig. 3.6g-i). This suggests the assemblies may dissolve after $NaHCO_3$ is added and calcite is formed because the available Ca^{2+} in their vicinity is depleted. At concentrations of 4 μM, protein the sheets persist after numerous calcite particles form on their surface (Fig. 3.5d,e).

At favorable viewing angles, a ~1.2 nm interface layer was observed between the protein-Ca^{2+} assemblies and the nascent calcite nanocrystals (Fig. 3.4c,d). This is consistent with the ~1 nm thickness of hydration layering on calcite [113] and suggests nucleation is an interface-driven process, as observed in other mineral systems in which surface ligands present carboxyl rich interfaces[114,115]. The flat periodic array of acidic residues on the surface of the proteins may direct the absorption and organization of Ca^{2+} ions and water molecules in the hydration layer into an arrangement consistent with the structure of calcite, or the interfacial region may alter the activities of the $CaCO_3$ phases to modulate the supersaturation for calcite vs vaterite.

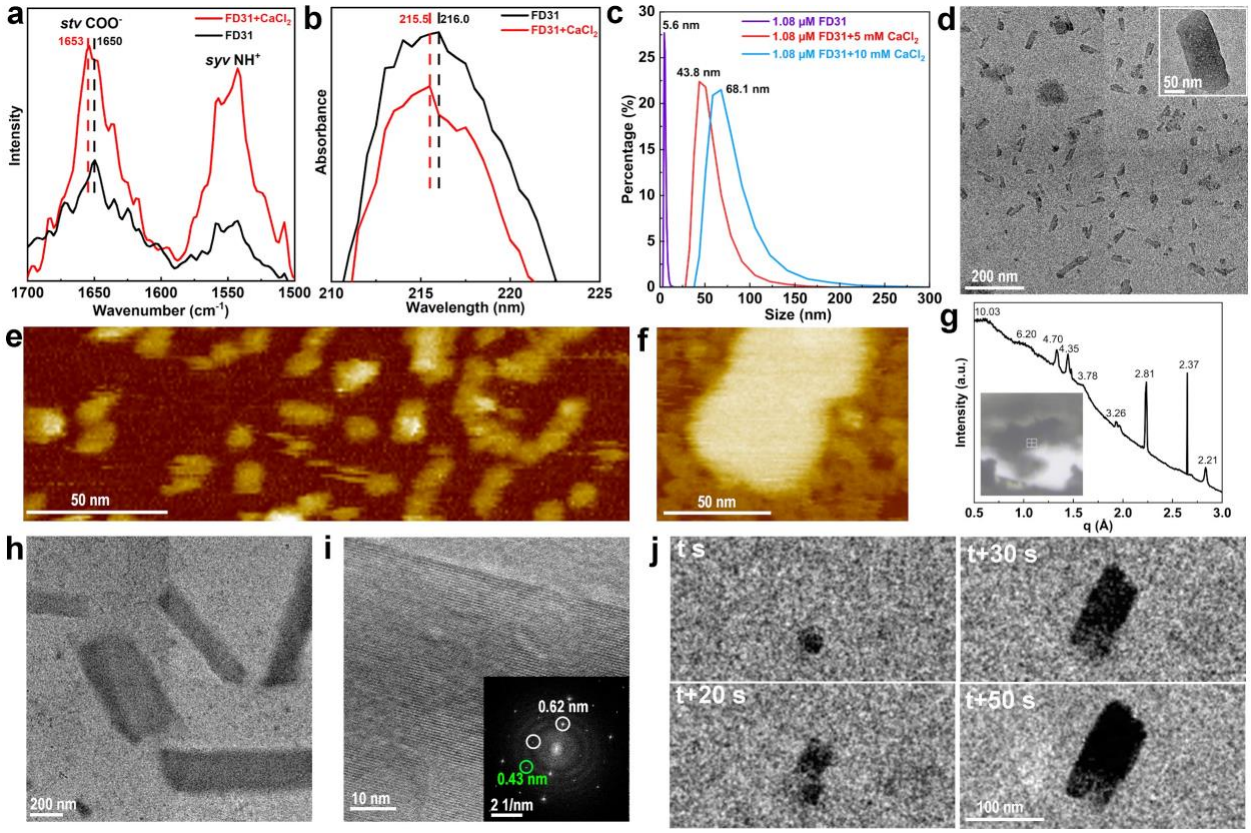

Figure 3.6. The structural characterization of the FD31-Ca complex. **(a)** Liquid-phase ATR-FTIR spectra and **(b)** UV-vis spectra of pure FD31 (black) and FD31 plus $CaCl_2$ (red) solutions. The peak shift shows the coordination of Ca^{2+} with carboxylate groups on protein. **(c)** DLS size measurements of FD31 with different $CaCl_2$ concentrations (0, 5, and 10 mM). **(d)** LP-TEM image showing the sheet-like assemblies. **(e-f)** Liquid-phase AFM images showing the sheet-like individual FD31 molecules and FD31-Ca complex. AFM image reveals a sheet-like shape of the FD31 protein molecules dispersed in MOPS buffer solution with a size of ~8 nm by 5 nm by 2 nm (Length by width by height), which is consistent with their simulated "Flattened" Rosetta structure model (~7.8 nm by 4.5 nm by 2.5 nm). The measured length difference from the model might be attributed to their conformational change in solution, as predicted by AlphaFold2 models. **(g)** Beamline XRD of assembled protein structure. **(h-i)** Representative TEM image showing the sheet-like assemblies **(j)** Time-dependent LP-TEM images show the formation of the FD31-Ca complex.

3.2.4 The growth process of calcite in the presence of FD31

We next followed the growth of FD31 nucleated calcite nanocrystals over time using LP-TEM. During the ~5 min following nucleation, individual nuclei assembled into ~10-20 nm particles through multiple attachment events, rather than growing through a classical pathway (Fig. 3.7a). Ex-situ HR-TEM shows the larger agglomerated particles are surrounded by smaller

particles (Fig. 3.7b-c), and the assemblies have continuous crystal lattices (Fig. 3.7d). Similar agglomerates were also observed after growth with DHR49-Neg at this time point (Fig. 3.8). Based on these observations we propose that oriented particle attachment[116] is the dominant pathway for calcite growth in the presence of FD31 and DHR49-Neg.

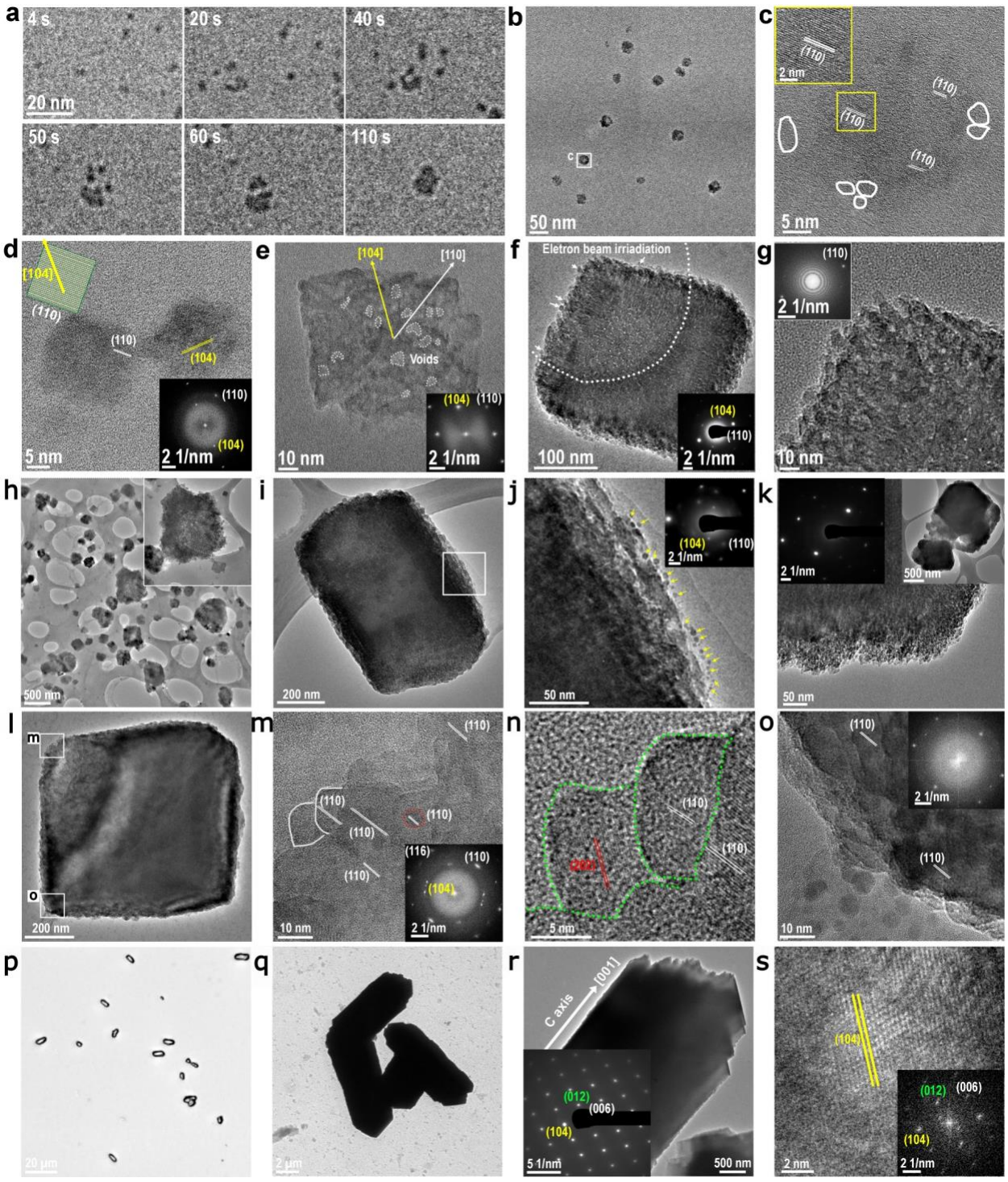

Figure 3.7. Calcite nanocrystal growth and assembly in the presence of FD31. **(a)** In situ LP-TEM image sequence shows the attachment of multiple calcite nanocrystals into a larger one. **(b)** Ex situ TEM image showing calcite grown through particle attachment. **(c)** HR-TEM image showing the aggregated crystal with (110) lattice orientation. White frames mark several individual calcite nanocrystals around the larger particle. **(d)** HR-TEM image shows additional calcite nanocrystals with a size of ~20 nm and (110) side faces. **(e)** TEM image shows a ~100 nm rhombohedron-like

calcite. The inset FTT shows the crystal orientation and single crystallinity. **(f)** TEM image showing rhombohedral calcite with a size of ~300 nm. The embedded SAED image demonstrates the single crystalline feature. Arrows mark several calcite nanocrystals on the surface. **(g)** HR-TEM and corresponding FFT images in the inset confirm the single crystalline feature and that the rough surface is composed of attached nanoparticles.**(h-o)**TEM images of calcite at 30 mins in the presence of 1.08 μM FD31 protein. **(h)** Irregular calcite consisting of nanocrystals. Inset shows an individual calcite aggregate. **(i,j)** One calcite crystal is surrounded by some calcite nanoparticles. Inserted SAED image confirms the single crystalline feature. **(k)** Rhombohedral calcite crystals consist of smaller nanocrystals. **(l)** TEM image shows a rhombohedral calcite with a size of ~800 nm. **(m,n,o)** Higher magnification views of the corners of the particle in **(l)**. **(m)** HR-TEM image of the corner part, showing the same (110) lattice orientation for the crystal body. **(n)** HR-TEM image showing several single-crystal units on the surface. **(o)** HR-TEM image showing numerous small nanocrystals located around the rhombohedral calcite. **(p)** OM **(q-s)**, TEM and SAED images show the formation of rod-like calcite with the addition of 4 μM FD31 protein.

The ~10-20 nm agglomerated calcite crystals formed with FD31 and DHR49-Neg also express different preferred facet orientations, (110) for FD31 (Fig. 3.7b-d) and (202) for DHR49-Neg (Fig. 3.8b-f), both of which are distinct from the natural (104) facets of calcite grown in protein-free solutions. These differences likely reflect differences in the structure and chemistry of the two proteins (Highlighted in Table 3.1). First, while the proteins are both composed of repeating alpha helices, these are spaced at different distances, leading to different extents of epitaxially matching for different calcite surfaces (see Section 3.3.2 for further discussion, Fig 3.13). Second, although the two proteins have similar net-charges and number of carboxylate groups, they contain different ratios of aspartate and glutamate residues, which present carboxylate groups in different orientations, resulting in distinct stereochemical alignments relative to carbonate ions in the calcite surface (see Section 3.3.2 for further discussion, Fig 3.12).

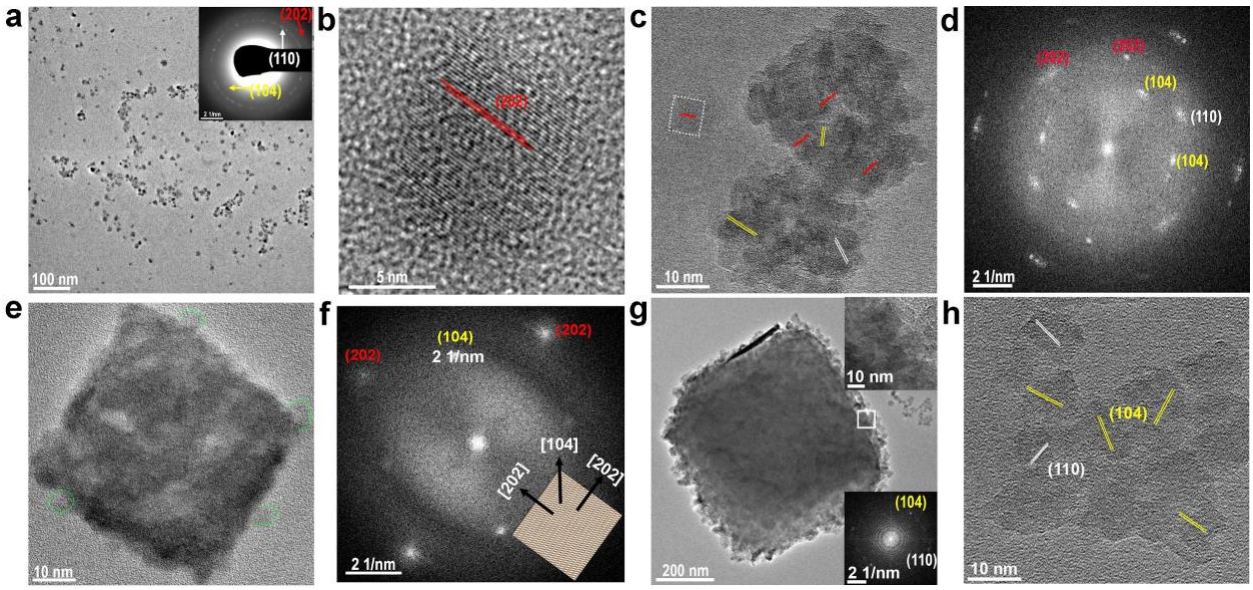

Figure 3.8. The CaCO3 crystallization process in the presence of DHR49-Neg. **(a)** TEM and inserted SAED images show the existence of calcite nanocrystals. **(b)** HR-TEM image shows an individual nanocrystal with (202) lattice. **(c, d)** HR-TEM and corresponding FFT images show aggregated calcite single crystals. **(e, f)** HR-TEM and corresponding FFT images show a cubic-like single crystal with exposed (202) facets. **(g, h)** Individual rhombohedral calcite with some nanocrystals on or around it. Inserted HR-TEM and FFT images in panel "g" to confirm that calcite nanocrystals are incorporated into the pre-existed calcite.

At 20 min we observed ~100 nm calcite crystals with a pseudo-rhombohedral morphology (Fig. 3.7e), their habit resembling the thermodynamically stable {104} calcite rhombohedron, but with terraced corners and low contrast features that suggest a discontinuous lattice separated by voids created by the inclusion of proteins. At 30-40 mins, the calcite crystals grew to ~300 nm to ~800 nm (Fig. 3.7f-i), and retained a similar morphology with rough surfaces, internal voids, and multiple 4-8 nm particles on their surfaces (Fig. 3.7j-o), suggesting growth through repeated attachment of the nanocrystals. This conclusion is supported by LP-TEM data (Fig. 3.9e). DHR49-Neg also formed calcite rhombohedra of similar size and habit (Fig. 3.8g). These results show that, in the presence of the proteins, the 5-7 nm primary calcite nanocrystals assemble through oriented attachment to ultimately form micron-scale mesocrystalline rhombohedral calcite. When the FD31 concentration was increased fourfold, calcite crystals elongated along the

c-axis ([001] direction) formed (Fig. 3.7q-s). Thus, specific designed proteins can direct the formation of nano-calcite, inhibit their further growth directly from solution species, and switch the crystallization pathway from classical single crystal growth to a nonclassical oriented particle attachment and produce calcite-microcrystals with morphologies that vary with protein design and concentration.

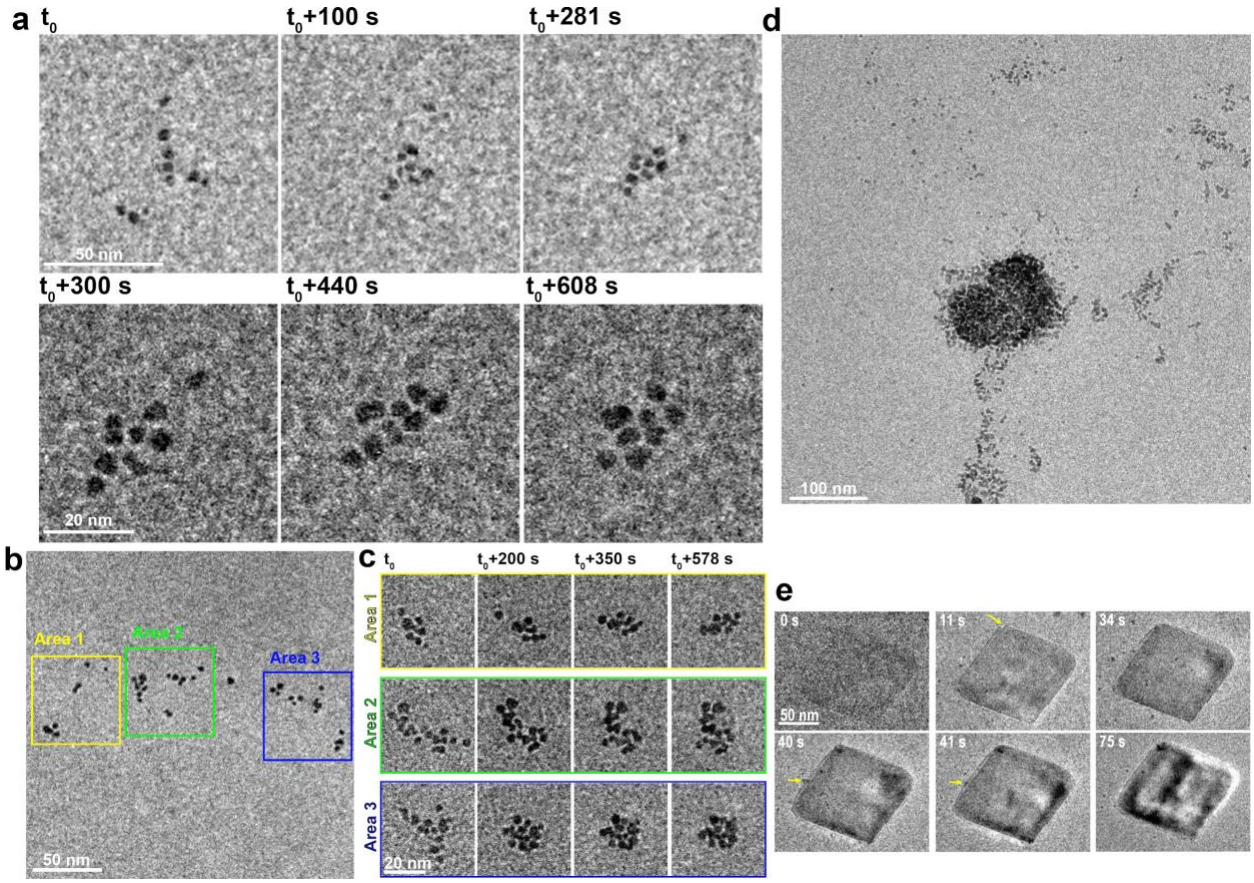

Figure 3.9. LP-TEM observation of dynamical behavior of FD31 1 μM nucleated calcite nanoparticles in liquid-cell. **(a)** Sequential TEM images showing the aggregation process of multiple particles. **(b)** LP-TEM image showing the initial distribution of particles. **(c)** Series of TEM images showing the formation of three aggregates. **(d)** LP-TEM image showing the aggregated particles. **(e)** Sequential TEM images show the attachment of particles into pre-existed rhombohedral calcite.

3.2.5 Protein length and surface chemistry effects on $CaCO_3$ crystallization

We next investigated what specific features of FD31 contribute to its potent calcite nucleation activity by generating versions of the protein with varied structural and chemical

features. The effect of template length (the extent of the carboxylate arrays) was investigated by generating versions of the protein with different numbers of repeat subunits. FD31 has 6 repeat units and a 5 x 8 nm interface containing 36 carboxylate groups, and we produced a 3 repeat version with a 5 x 5 nm interface containing 18 carboxylate groups and a 9 repeat version with a 5 x 11 interface containing 54 carboxylate groups. The 3, 6, and 9 repeat templates all drive the formation of nano-calcite, but the 3 repeat version leads to larger nanoparticles (Fig. 3.10a-c), likely due to reduced stabilization of the critical nucleus or weaker binding to calcite surfaces following nucleation. Nucleation driven by monomers of FD31-Rep9 was observed by liquid-phase TEM (Fig. 3.10d,e).

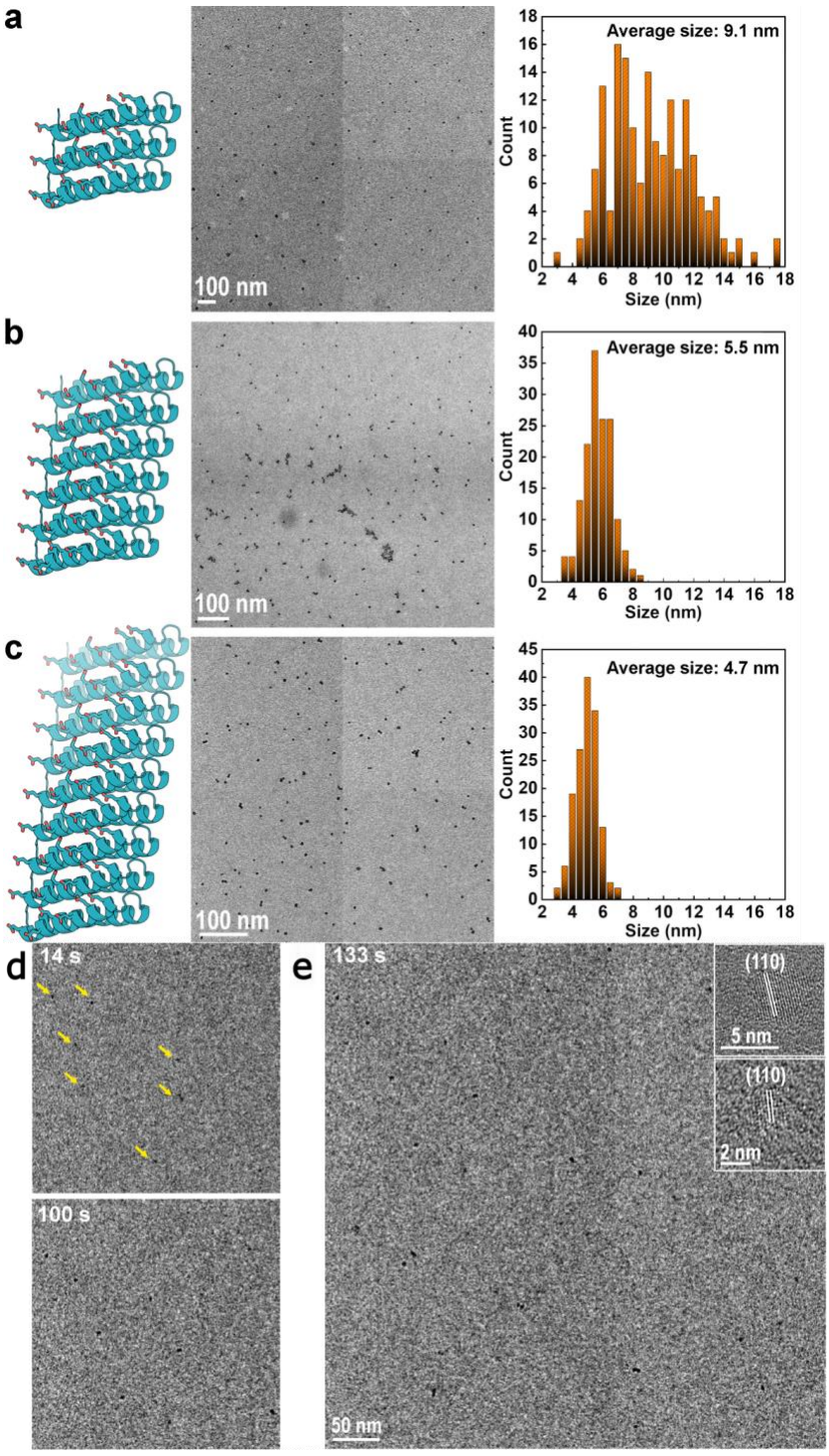

Figure 3.10. Tuning calcite nucleation by varying FD31 length and surface chemistry. **(a-c)** Ex-situ TEM and the size distribution of particles formed in the presence of **(a)** FD31-Rep3 with 3 repeats, **(b)** FD31 with 6 repeats, and **(c)** FD31-Rep9 with 9 repeats. In situ LP-TEM observation of CaCO3 crystallization process in the presence of FD31-Rep9 protein. **(d)** Sequential TEM images showing the nucleation of $CaCO_3$ nanoparticles. **(e)** TEM image showing multiple CaCO3 nanoparticles in liquid-cell. Inserted HR-TEM image confirms the calcite nanocrystals.

To investigate the effect of surface chemistry on calcite nucleation activity, we modified the surface chemistry of FD31 in three ways. We either substituted half of the glutamate residues to nearly isosteric but non-charged glutamine residues (FD31-Gln-Checker), replaced 24 out of 36 glutamate residues with shorter aspartic acid residues (FD31-Asp), or mutated 13 out of 36 of the glutamate to basic lysine residues (FD31-Lys-Checker)(Fig. 3.11). FD31 (containing an all glutamate interface) and FD31-Gln-Checker generated calcite-dominant nanocrystals (Fig. 3.3g,h,i, Fig. 3.11a-d), both vaterite and calcite phases appeared in the presence of FD31-Asp (Fig. 3.11e-h), and FD31-Lys-Checker produced primarily vaterite (Fig. 3.11i-l). These variants provide direct experimental evidence that the structure and chemistry of the pre-organized side chain array in our DHR protein affects its templating of calcite nucleation.

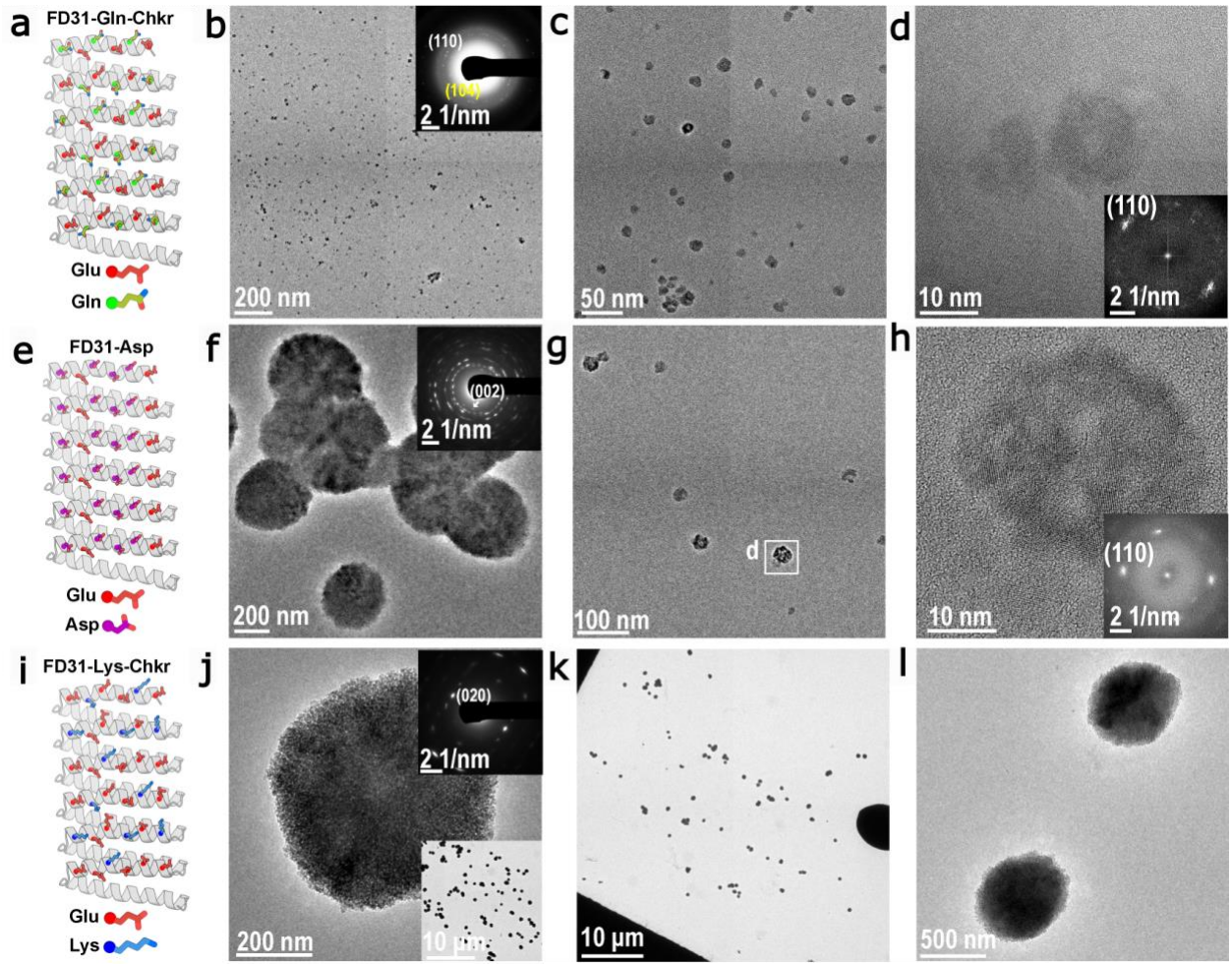

Figure 3.11. The resulting $CaCO_3$ in the presence of different FD31 mutants. **(a-d)** TEM, SAED and FFT images show the formation of calcite-dominant nanocrystals in the presence of FD31-Gln-Checker. **(e-h)** TEM, SAED, and FFT images show the predominant formation of vaterite and some calcite nanocrystals in the presence of FD31-Asp. **(i-l)** TEM and SAED images show the formation of a vaterite dominant phase in the presence of FD31-Lys-Checker.

3.3 Discussion

3.3.1 Nucleation pathways and growth mechanisms in the presence of FD31

Our designed protein templates promote the direct formation of nano calcite, bypassing the typical formation of vaterite microcrystals, by both monomer and protein-Ca^{2+} assembly driven processes. The latter pathway resembles the multistep process in bone and dentin calcification where Ca^{2+} induced assembly of an acidic matrix protein precedes formation of apatite[117].

The Ca-protein template driven nucleation is more effective at driving nucleation as there are no additional nucleation events observed in solution but the ones near the interface of the template. In the case of the nucleation route driven by individual monomers, this route seems to be the most prevalent one since the number of nuclei is comparable to the number of protein monomers in solution. In this case, there is a higher number of calcite nuclei compared to the number of monomers in solution. This implies that the monomers could be playing a role where they catalyze the precipitation reaction and then proceed to either stabilize the nuclei, or to nucleate more particles. There is precedent for ligands acting both as nucleators and stabilizing agents.

Initial studies in nanosciences showed that the synthesis of small nanoparticles was achieved through the presence of ligands that would passivate the surface of a growing nucleus[118]. The particle would keep growing until the spacing of the tail groups, which decreased with decreasing particle curvature, was reduced to the point where their repulsive interactions became self-limiting. This combined effect of surface passivation and steric hindrance would limit the size into which these nanoparticles would grow and their self-assembly pathway

guided through packing by shape. In this work, the interstitial spacing between particles undergoing attachment (Fig. 3.9a,b) is in the order of 1 to 2 nm, leaving enough space for roughly one monomer interacting face-on shared among at least two particles. The fate of the monomer is then to be expelled from the growing aggregate or to be trapped and form the voids and interstices observed in larger particles (Fig. 3.7e,f). This expulsion from the growing aggregate together with the stabilization of multiple particles by a single monomer supports the hypothesis of a monomer being recycled after stabilization of particle aggregates. Considering the surface area of the spherical nucleus and of the interacting face of the protein monomers, 9 monomers should roughly stabilize a single calcite particle. This is unless the incipient nuclei are pinned and stabilized by a single monomer, similar to what is done by individual protein monomers acting as antifreeze proteins[3].

The proteins also drive post-nucleation growth through oriented attachment of 5-7 nm nanocrystals into micron-scale rhombohedral single crystals with terraced edges and abundant low contrast inclusions. Growth by oriented particle attachment can be interpreted or achieved through three separate explanations[119,120]. First, particles come in close proximity and low energy interfaces interact and orient the particle, followed by healing of the interface by atomic rearrangement. Second, particles can be attracted to one another through temporary or permanent dipoles. Third, particles are attached through random orientations, but coalignment occurs through dissolution-recrystallization processes or atomic rearrangement after attachment. HR-LP-TEM would be necessary to exactly determine the mechanism by which growth by attachment is carried out in this case. However, evidence of orientation of particles and then attachment through low-energy interfaces is possibly revealed in LP-TEM videos showing rotation of the particles before they are brought into contact (Fig. 3.9a).

3.3.2 The role of stereochemistry and epitaxial matching in the mineralization outcomes

The most significant differences between the designed proteins that could explain the differences in outcome are the interhelical spacing and the surface chemistry, particularly the presence of groups capable of stereochemical matching of different surfaces.

First, the use of odd (Asp) vs. even (Glu) number of carbons in the side chains of the binding moieties could bias the formation of different interfaces by virtue of the angle at which the carboxyl group is displayed on the surface of the protein, similar to what is found for SAMs [111]. If the preferred angle of the carboxyl group matches the orientation of the carbonate groups on that facet, the interaction will be favored. This angle is parallel to the substrate surface in the case of Glu and in the case of Asp there is a preference for an oxygen atom to point out. Given these orientations, Glu is likely to interact with {110} and Asp can interact most likely with {202}(Fig. 3.12).

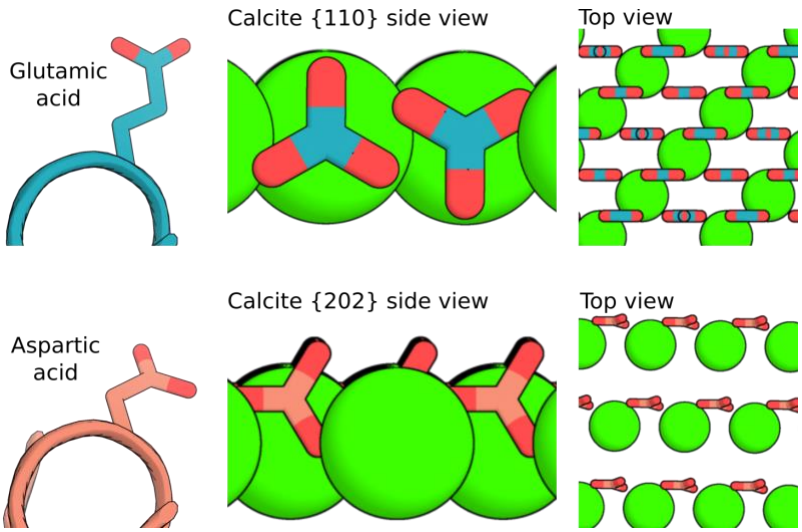

Figure 3.12. Orientations of the carboxylate groups in different binding moieties could explain the bias towards the formation of a particular calcite facet.

In the case of geometric matching, the interhelical repeat distance of a protein can be compared to the distance between atoms on a plane. For this, we modeled the Rosetta energy of flattened protein models with different interhelical spacings (See Section 2.1.3). The shape of the

curve for the energy vs. interhelical spacing can provide information of the flexibility and spacings that a protein can adopt, whereas the lowest energy in the curve can approximate the most likely spacing adopted by the protein. Flat proteins were modeled since they represent the lowest possible interfacial free energy driving the least possible mismatch between the crystal and the template. The flatter curve after ~11.5 Å spacing for FD31 (blue circles, Fig. 3.13), indicates that it could be the most flexible, accommodating different interatomic spacings at 11.5 Å and higher. This could explain why FD31 has a higher nucleation activity compared to DHR49-Neg and FD15 (which have increasingly narrower distributions). The minima for each protein model is close to the crystallographic median interhelical spacing, for FD15 it's 8.7 Å, for DHR49-Neg it is ~10.5 vs 9.7 Å (modeled vs. crystal structure), and for FD31 is ~11.5 vs ~11.2 Å of the AlphaFold2 structure prediction.

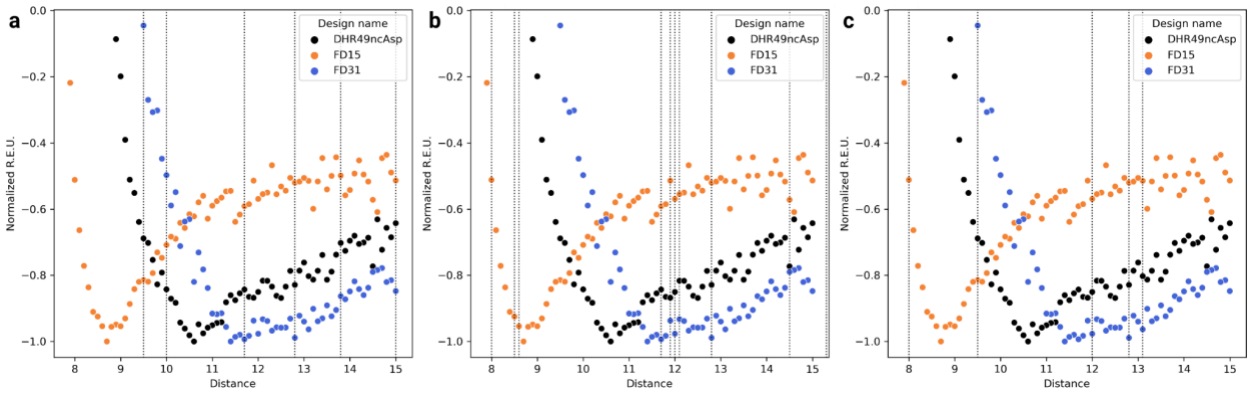

Figure 3.13. Normalized Rosetta Energy as a function of the interhelical spacing distances for different protein models. Units on the x axis are angstroms. Vertical dotted lines represent the unique interatomic spacings in plane on different surfaces of calcite: **(a)** (202) surface **(b)** (110) surface **(c)** (104) surface.

Beyond stereochemical matching arguments, there are additional elements to consider like the effect of the solvent. The guiding hypothesis of structured proteins with carboxylate groups on a regular surface considers that it can bind and pre-organize ions from solution. This leads to further consideration of the ions, which can create some solvent structuring, and how proteins can structure solvent at their surface. Interactions among ions and between ions and the solvent are competing with interactions between the proteins and the ions. So this might point to the water pre-organization capabilities of a protein scaffold to be the key for nucleation as opposed to simple

stereochemical matching. Still, the structuring of the solvent would in the end be dependent on the surface chemistry, the shape of the protein, and the different degrees of structuring of this template. Further support for this structuring of water and ions as a driving force for nucleation is the observation of a ~1 nm separation between the template and the nascent nuclei for the formation of nanocalcite at the interface of the Ca-protein complexes (Fig. 3.4c,d).

3.3.3 Understanding template surface chemistry and size effects

The ways in which the protein can be interacting with calcium carbonate can be explained considering previous MD studies[121]. Deriving from those studies it can be suggested that the greater effect is imparted by the highly charged side chains presenting carboxylate groups which could interact with Ca^{+2} ions directly and with CO_3^{-2} ions through salt bridges or water-mediated interactions. Positively charged arginine and lysine groups can interact with CO_3^{-2} ions. Other interactions include hydrogen-bonding/ionic bonding of both ions with the oxygen atom of serine or the nitrogen backbone atoms; electrostatic interactions of Ca^{+2} with the hydroxyl group of threonine; and coordination of either ion with the polar amide groups in asparagine or glutamine. One possible reason as to why we see nanocalcite in the presence of FD31-Gln-Chkr but not in that of FD31-Lys-Chkr is that the former does not disrupt Ca^{+2} binding (these groups are capable of coordination of the same ion as the negatively charged side chains), as opposed to positively charged side chains in the latter protein being able only to interact with CO_3^{-2} ions. However, it is surprising to see that FD31-Asp nucleates both vaterite and calcite. Possibly, interactions between the calcite nuclei and the negatively-charged residues create a mismatch at the interface and the length and the rigidity of the side chains plays a role (aspartate chains have more restricted side chain conformations compared to glutamate).

A separate set of interactions can be considered for the protein surface hydrophobic residues and the solvent. Alanines, threonines, and other hydrophobic moieties in protein side chains have been computationally modeled[7] and partially found through crystallographic and

functional studies[56,122] to be involved in the formation of clathrate water networks as a mechanism of interaction between proteins and ice. It stands to reason that similar networks can arise in designed proteins, in the case of FD31, there is a total exposed hydrophobic surface area of 7440 $Å^2$, compared to the 5354 $Å^2$ of DHR49-Neg. Ordering water molecules on the surface of proteins has an entropic cost, this can be compensated by the release of water back to the bulk after desolvation of ions at the interface with the protein template. This can in turn facilitate interactions of ions with the protein and the formation of an incipient nucleus according to CNT at heterogeneous interfaces described in Section 1.2.1, Equation 1.11, where a high interfacial free energy between the liquid and the substrate α_{ls} can drive the overall interfacial free energy down.

The smaller template nucleated larger calcite particles, either because it has a higher interfacial free energy and is a less effective nucleator, leaving more Ca^{2+} and CO_3^{2-} available for post-nucleation classical growth, or because it is a weaker surface ligand on the nucleated crystals. Given that template-mineral binding is directly related to the interfacial free energy between the template and the crystal[123], these two effects are expected to go hand in hand [21].

3.3.4 Advantages and disadvantages compared with other systems

The closest, rationally-designed protein or peptide systems that have been used to modulate the nucleation and growth of calcium carbonate, and that are comparable to the molecules described in this chapter, are described in Section 1.3.1. There are a number of advantages regarding the general approach of using DHRs over previous attempts. These include that there is a large interface that can be modulated (as opposed to the described click chemistry approach, which is limited to restricted sites[26]); there is no need for immobilization to surfaces to drive self-assembly and structuring (as in the case of the ChiSifiCa protein[27]); and that structured helical elements can be pre-organized into a tertiary structure and guide nucleation of calcium

carbonate instead of growth only (like in the case of the small, designed helical peptides[12]). Tunability can be enhanced by introducing non-canonical amino acids at the interface enabling the creation of different coordination chemistries with click chemistry (however, this is not a trivial task and would require either re-design and/or optimization of expression conditions). An additional advantage of de novo design in biomineralization is the potential to target specific faces or non-natural faces since directed evolution cannot target a specific face and is limited to the stable crystal faces formed[124]. One key disadvantage is that they cannot achieve the same degree of coverage and number density compared to SAMs.

3.4 Conclusion

Our designed proteins possess greater structural regularity and tunability than previously studied native and engineered biomineralization proteins[26,98,125] and drive polymorph-specific $CaCO_3$ nucleation. This tunability allows rigorous interrogation of the interface driving this templating effect. Our DHR proteins. We compared DHR templates with similar chemistries in arrays spaced at distance of 0.9 nm, 1.0 nm, and 1.1 nm intervals (FD15, DHR49-Neg, and FD31, respectively; Table 3.1) and found the 0.9 nm protein did not template calcite (FD15), the 1.0 nm protein (DHR49-Neg) templated calcite to some extent, and the 1.1 nm protein (FD31) was a strong nucleator of calcite. FD31 nucleated nano-calcite, bypassing the formation of micron-sized vaterite, by two distinct pathways: one driven by individual monomers, the other by interfacial nucleation near a sacrificial template formed by structured ca-protein interactions. Produced nanocrystals assembled by oriented attachment to create mesocrystalline calcite in the micron scale. The 1.0 nm protein and the 1.1 nm protein produced nanocrystals with different preferred orientations, potentially due to the formation of better epitaxial matches with different calcite planes (Section 3.3.2). We tuned the length of FD31 to 5 nm, 8 nm and 11 nm by producing

versions with 3, 6, and 9 repeat units (FD31-Rep3, FD31, FD31-Rep9, respectively). The unprecedented mutability of our scaffold allowed us to interrogate the effect of various surface chemistries on nucleation outcomes. The observation that FD31-Gln-Checker nucleates calcite better than FD31-Asp suggests that stereochemical alignment is more important to the activity of our template than the total number of negative moieties. Steric hindrance and/or the inclusion of positive moieties within the template may explain the inability of FD31-Lys-Checker to nucleate calcite.

Computational protein design allows biomineralization-like processes to be studied with genetically encoded molecular templates that have known stable 3D structures, are chemically and structurally tunable, and are amenable to further engineering[65,79,126]. Compared to previous studies of natural and engineered proteins, using de novo designed proteins allows more rigorous testing of how structurally defined biomolecular surfaces can control inorganic crystallization. This new approach sets the stage for the programmable control of crystal polymorph, nucleation pathway, growth mechanism, and crystal morphology, and the development of novel hybrid materials with functional properties.

4 Chapter 4: HEMATITE BINDING and NUCLEATION

4.1 Introduction

Recent work using atomic force microscopy (AFM) to study de novo helical-repeat (DHR) proteins with interfaces designed to bind the mica (001) K+ ion sublattice showed a lattice matching approach could control the orientation of proteins adsorbed on an inorganic crystal[63]. Additionally, the authors showed the strength of binding and the structures formed by proteins on the surface were affected by the size of the proteins and by the addition of designed protein-protein interfaces. This work also highlighted that the structure of the protein-inorganic interfaces were highly dependent on solution conditions and entropic interactions between the rod-shaped proteins, where aspect ratio is a feature with a large effect.

Building on the work of repeat proteins designed to self-assemble on mica, we seek to further test the lattice-matching approach by designing binding towards surfaces with a different chemistry and spacings. The high abundance and diversity of iron oxides in natural and synthetic systems, together with their broad applications, was the motivation to center our focus on iron oxides, and particularly on hematite. Figure 4.1 explains the general approach to be used, and provides a guide of how to interpret the data presented in further sections.

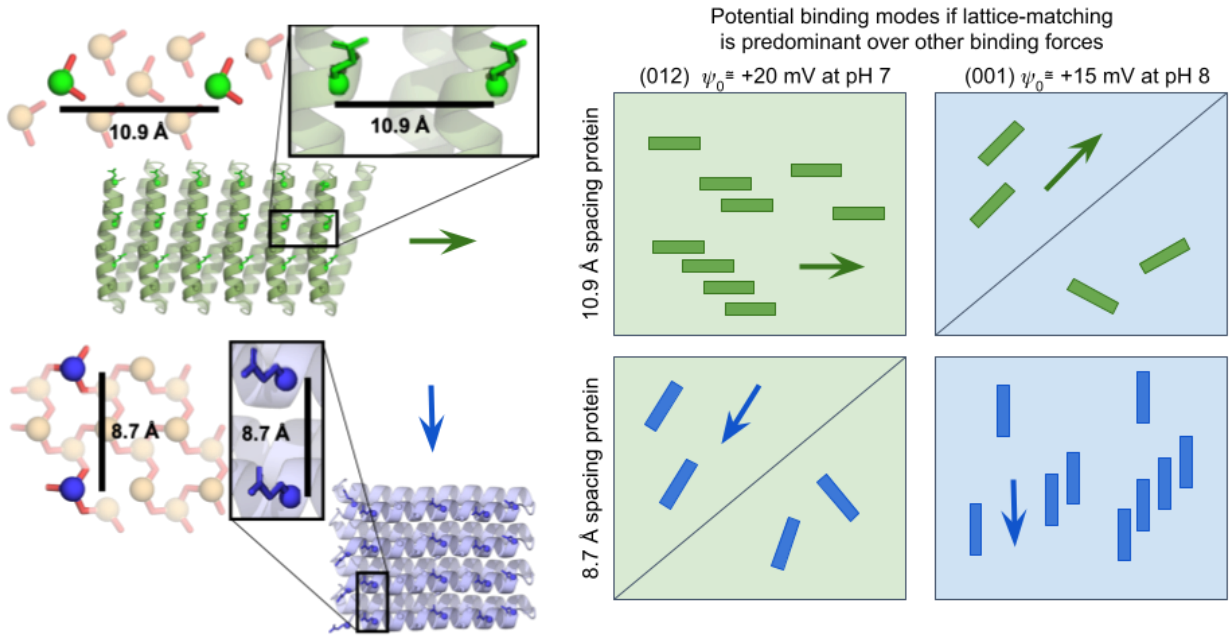

Figure 4.1. Lattice matching design principle for targeted and oriented binding on two hematite surfaces. A green designed helical repeat protein (DHR) with a spacing of 10.9 Å between its subunits matches twice the lattice spacing of iron atoms on the (012) surface of hematite. A blue DHR with a spacing of 8.7 Å between its subunits matches twice the lattice spacing of iron atoms on the (001) surface. On the right, an illustration of the expected binding on the two surfaces. Off-target interactions are expected to have a lower coverage or affinity, and may lack a preferred orientation. Note that the (001) plane has three-fold symmetry, so we expect binding along three equal orientations, but only one is shown for simplicity. Spacing of chemical moieties and further tuning of the interface would lead to better understanding the design rules for targeted oriented binding between proteins and iron oxides and other inorganic crystal surfaces.

Once again, a designed helical repeat topology was chosen due to advancements in the field that make it possible to tune this topology with a relatively high degree of success. Specific DHR scaffolds were generated from scratch according to the methodology described in Section 2.1.2 with a precise repeat-spacing to match the hematite (012) surface, (green proteins in Fig. 4.1 with a 10.9 Å spacing) or biased match to the hematite (001) surface (blue proteins in Fig. 4.1 with an 8.7 Å spacing).

4.2 Results

4.2.1 Biochemical characterization of the first generation of hematite-binding proteins

From an initial set of 12 proteins, three protein variants from a single original template sequence were characterized. These proteins, collectively named the R-Rep4 series, are expected to fold into helical repeat proteins where the repeating elements are spaced at an average distance of 10.9 Å, but according to models generated by AlphaFold2 the median interhelical repeat distance is 11.2 Å. Versions of this scaffold with four repeating elements behaved as monomers in solution as confirmed by SEC-MALS (Fig. 4.2). CD characterization confirmed their thermostability, giving the characteristic signal of alpha-helical secondary structure, which was largely maintained after heating to 95 ºC. The original spectrum was recovered at 25ºC, and a thermal melt screen following changes at 222 nm shows no evidence of cooperative unfolding at higher temperatures (Fig. 4.2). Experimental SAXS traces closely fit the traces predicted for two of the protein models, confirming their shape and oligomeric state in solution.

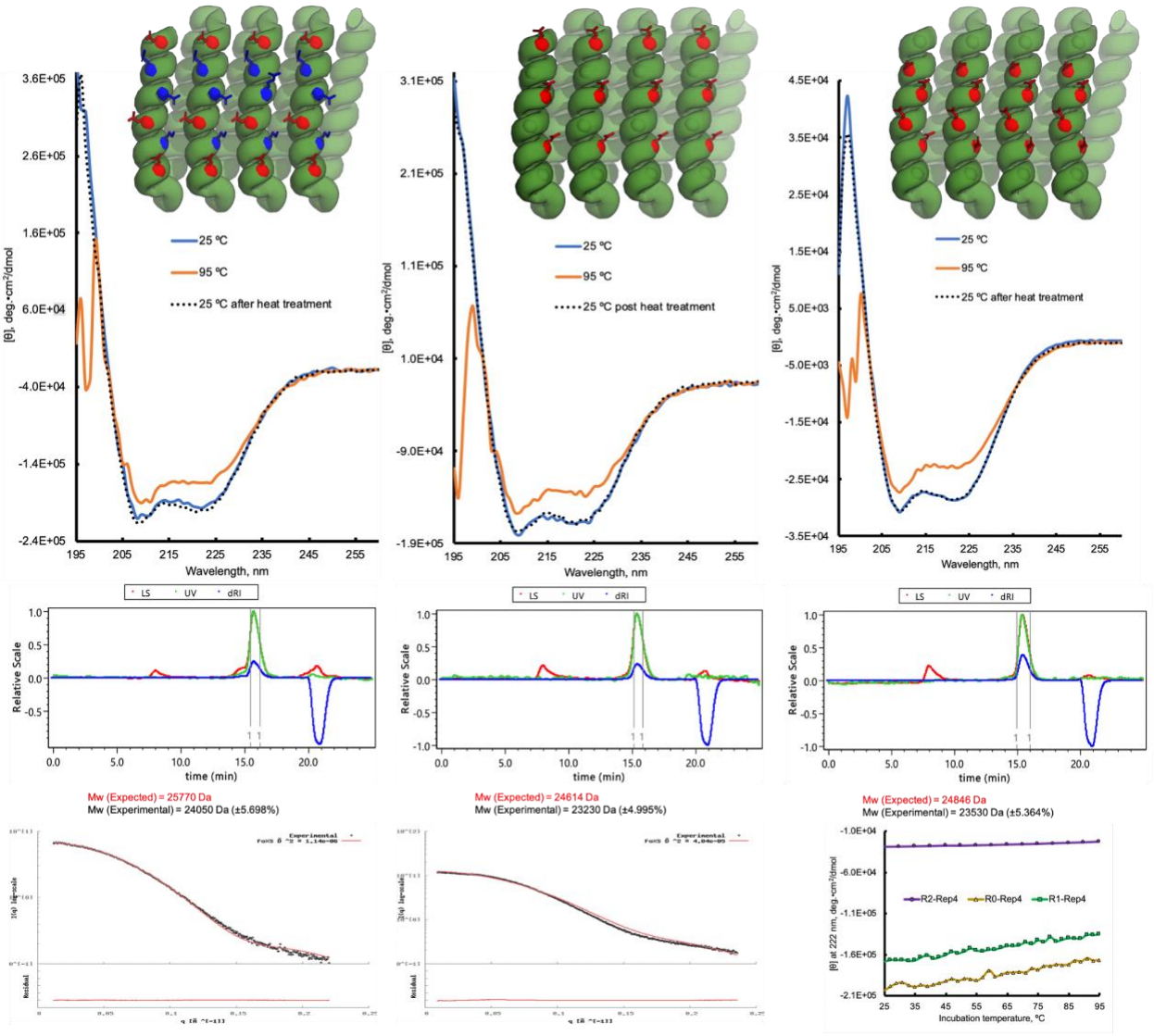

Figure 4.2. Biochemical characterization of three members of the R-Rep4 series. Divided in three columns, each corresponding to a different sequence. The protein models at the top of each column show the alpha-helical secondary structure in cartoon representation. The exposed residues at the binding interface in stick representation are colored red for negatively charged residues and blue for positively charged. Left column R0-Rep4 characterization data from top to bottom: CD, SEC-MALS with the expected and observed molecular weights, model fit to SAXS data. Middle column R1-Rep4 characterization data from top to bottom: CD spectrum, SEC-MALS chromatogram with the expected and observed molecular weights, model fit to SAXS data. Right column R2-Rep4 characterization data from top to bottom: CD spectrum, SEC-MALS chromatogram with the expected and observed molecular weights. Bottom right: Thermal melt curves measured by CD.

4.2.2 In situ AFM binding characterization of R-Rep4 proteins binding on the (012) surface

Preliminary characterization of the four-repeat variant using in-situ AFM aimed to look at differences in coverage between a set of three proteins. The target surface for all of them was the (012) plane of hematite, and the main difference between the variants is the number and arrangement of negative groups on one side of their surface. All of them had a negative net charge, but R0-Rep4 had a patch of positive residues on both sides of the protein (Fig. 4.2). Only one of the other two variants, R1-Rep4, was tested in this preliminary set. The main difference of this protein is the presence of a surface composed entirely of negative and small hydrophobic (alanine) residues on one side of the molecule (Fig. 4.2). The in situ imaging conditions were done at pH 7 with 20 mM Tris. We observed that having positive patches on both sides of the interface on R0-Rep4 were not sufficient to prevent it from interacting with a surface (Fig. 4.3) with a positive electric potential at pH 7 of $\psi_0 \cong +15$ mV[34].

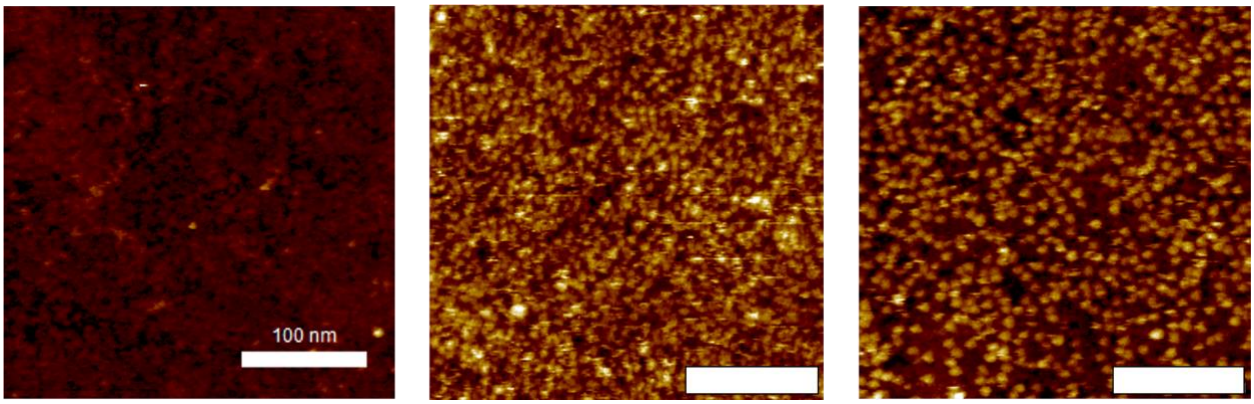

Figure 4.3. R-Rep4 proteins [0.1 μM] imaged on R-Cut hematite at pH 7 ($\psi 0 \cong +15$ mV). From left to right: bare hematite substrate using an ozone cleaning protocol, R0-Rep4, and R1-Rep4. Scale bar is 100 nm. All experiments in 20 mM Tris.

The coverage of the surface was greater for the protein with positive and negative charges and a higher net charge. Determining if lattice-matched proteins bind target surfaces preferentially and with a defined orientation will require further testing to confirm that this interaction goes beyond a simple electrostatic attraction. In addition, the aspect ratio for the four-repeat version of

these proteins is too low (roughly 1:1, 4.6 nm X 5.4 nm) to determine whether a preferred orientation existed on this substrate. A protein with a larger aspect ratio and a comparison test with an off target surface were proposed to ask these questions.

4.2.3 Biochemical and binding characterization of six-repeat variants on the R and C-cuts

Increasing the aspect ratio of the designed proteins is important to resolve the relative orientation of the proteins on the surface and compare it with the crystallographic orientation of the surface to evaluate if the lattice-matching approach is having an effect. Two known and tested ways of increasing the aspect ratio of proteins containing amino acid repeats are to either simply increase the number of inner repeats within a single protein monomer, or to design an interface between monomers in a way that could form end-to-end oligomerization and form fiber-like assemblies. Both approaches were tried with marginal success in producing and purifying six-repeat and nine-repeat versions from the original four-repeat "R" variants. The six-repeat variants showed some degree of oligomerization when characterized by SEC and SEC-MALS (Fig. 4.4). However, measurements of the protein's dimensions when adsorbed on the hematite surface are consistent with the estimated size and shape of monomers contacting the surface face-on (Fig. 4.5).

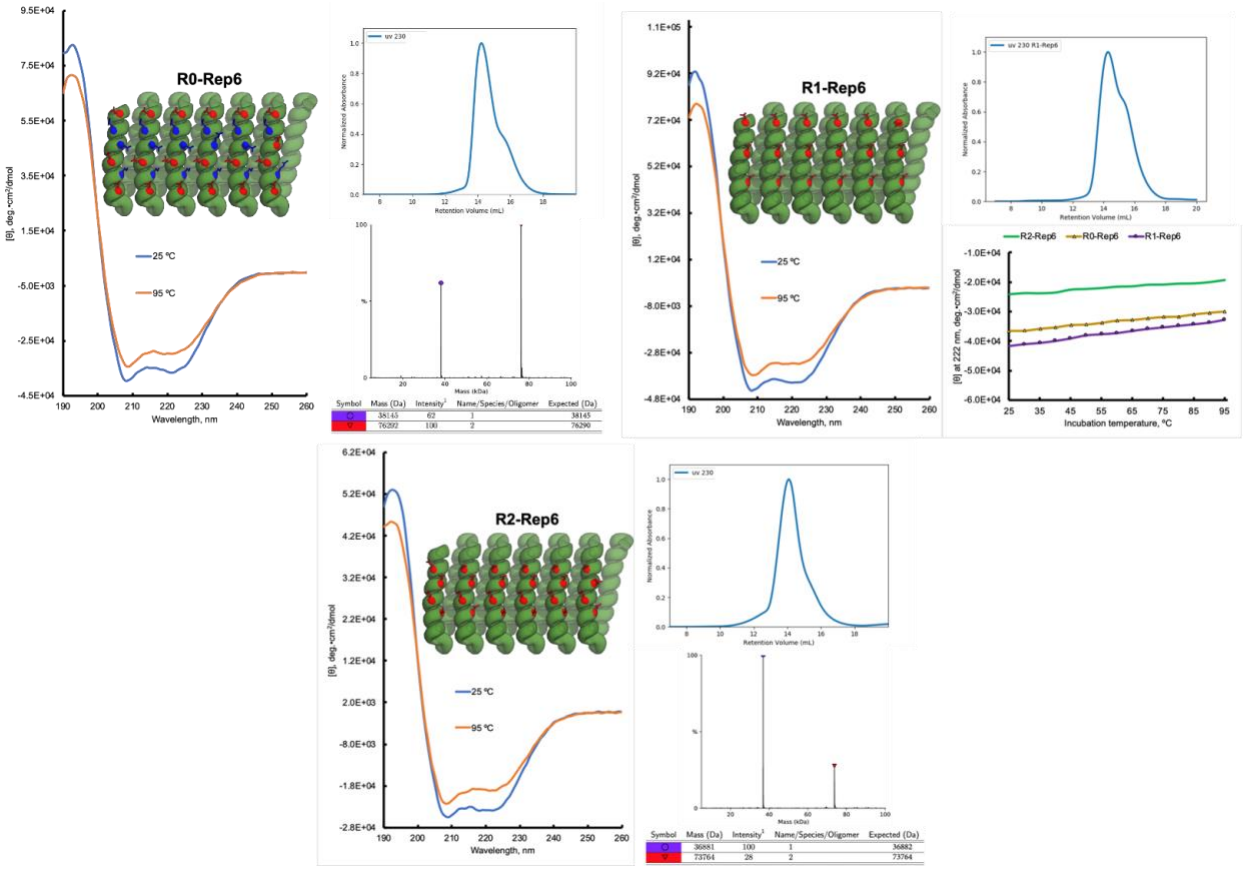

Figure 4.4. Biochemical characterization data for the expanded R-Rep6 proteins. DHR protein models depicted in green with cartoon α-helices spaced 10.9 Å apart. Negative residues highlighted in red, positive in blue. Alongside each protein model is the secondary structure determination by CD confirming the presence α-helices which are mostly retained during incubation at 95 ºC. Oligomeric state characterization results by nMS for the SEC main peak are shown below each chromatogram. SEC peak broadness and retention volume are consistent with the partial formation of dimers in solution at Tris 20mM pH 8 and NaCl 100 mM.

A different trend emerged for the six-repeat variants after incubating them on the R-cut surface. Incubation of 0.1 µM R2-Rep6 with the R-cut surface yielded an apparent array of oriented proteins, which retained its orientation after the surface was cleaned, rotated 90º, and tested again (Fig. 4.5a,b). This is strong evidence of oriented attachment of the protein, but we have not yet confirmed that the target orientation is the same as the observed. We proceeded by determining if the R2-Rep6 design preferentially bound to the target R-cut hematite surface vs. the off-target C-cut hematite surface. When incubated with 0.02 µM of R2-Rep6, binding events were not observed on the C-cut surface (Fig. 4.5c), but a few remained in the R-cut surface (Fig. 4.6c).

Repetition of this experiment at higher concentrations, 0.1 µM, yielded no coverage in 3 out of 5 surveyed areas (Fig. 4.5d). While R2-Rep6 covered extensively the R-cut surface at this high concentration conditions in an apparent oriented array as described above, the behavior of this protein on the C-cut surface appeared to be drastically domain-dependent. Most areas had no binding events, but a small fraction of the surface displayed large domains of oriented protein (Fig. 4.5e,f). This suggested that surface nanostructure and differences in the chemical terminations potentially explain the domain-dependent binding behavior (See Section 4.3.1).

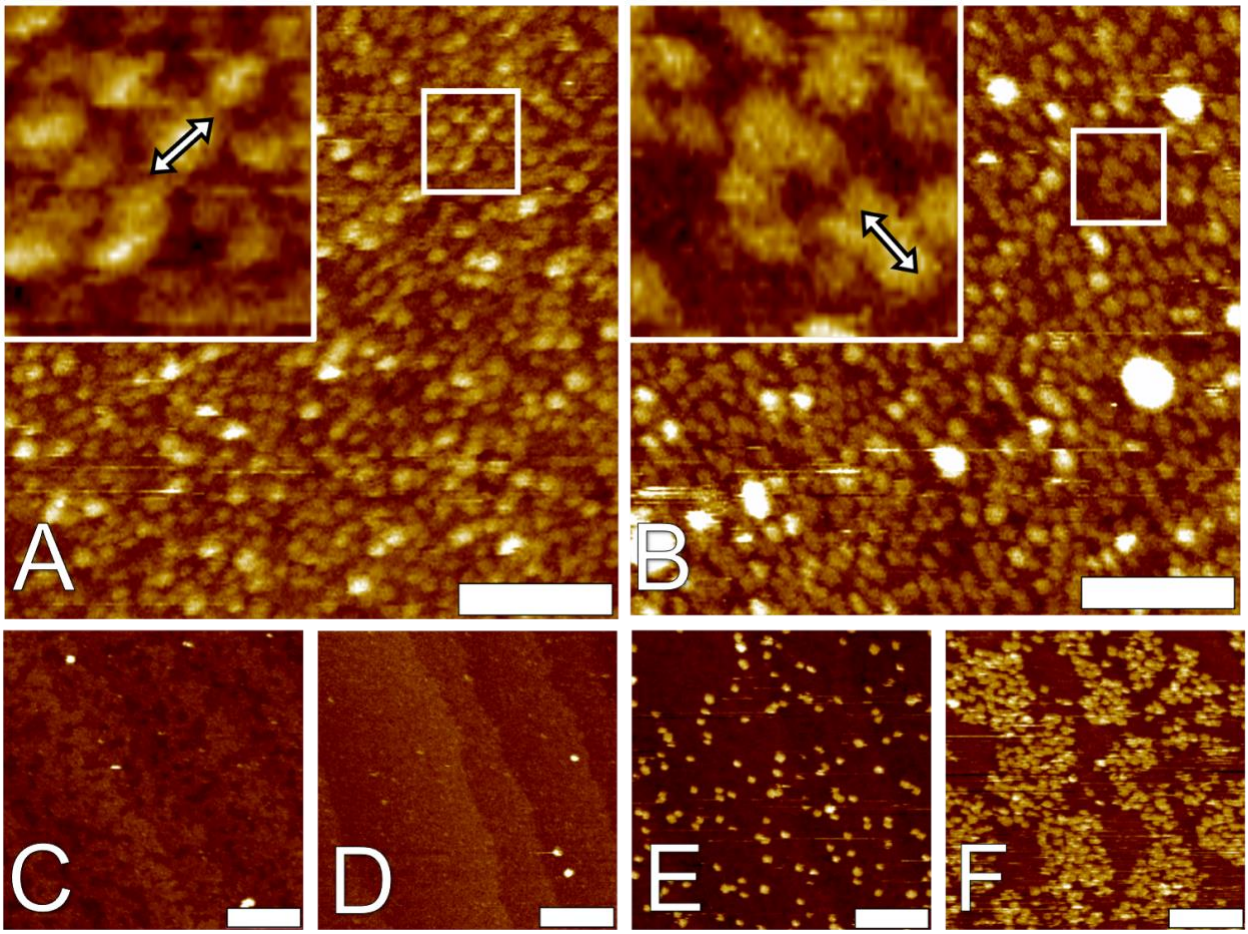

Figure 4.5. Biochemical R2-Rep6 apparent oriented binding on the R-Cut and heterogeneous binding on the C-Cut at pH 7. **(a)** Inset shows the possible orientation of R2-Rep6 (expected charge at pH 7 of -49.61) on the R-cut ($\psi 0 \cong +15$ mV). **(b)** Apparent orientation is retained after cleaning the substrate, rotating it 90º and incubating with the same protein type and amount [0.1 µM]. **(c-f)** R2-Rep6 incubated on the C-Cut at pH 7 ($\psi 0 \cong +30$ mV). **(c)** No binding of R2-Rep6 [0.02 µM] observed on the C-Cut. Protein concentration was increased to [0.1 µM] to match incubation

conditions from **(a)** and **(b)**. **(d-f)** Heterogeneous covering of R2-Rep6 on the C-cut at 0.1 μM. **(d)** Out of five surveyed areas, this was representative of three fields showing no binding. **(e)** In the same experiment, an example of some binding events on the C-Cut. **(f)** Extensive binding on one field. All scale bars are 100 nm.

On the R-Cut target surface at pH 7, we observe binding events for R1-Rep6 and R2-Rep6 (Fig. 4.6b,c). These binding events represent a higher coverage on the R-cut when compared to what could be found on the C-cut if a five-fold concentration difference is considered. Here, the R-cut surface is incubated with 0.02 μM, whereas the C-Cut is incubated with 0.1 or 0.15 μM of protein (Fig. 4.6e,f). This is the first preliminary evidence of a potential designed specificity, independent of electrostatic attraction. Comparing the same proteins on the off-target C-cut surface at pH 7, the R0-Rep6 protein (Fig. 4.6d) had a larger coverage compared to its more negatively-charged counterparts. However, R1-Rep6 had a slightly larger coverage despite having a less negative charge than R2-Rep6 (Fig. 4.6e-f). Simple attraction of opposite charges do not seem to be responsible for these binding events, as the coverage of the surface increases with the net charge of the protein, but again with the caveat of domain-dependent behavior.

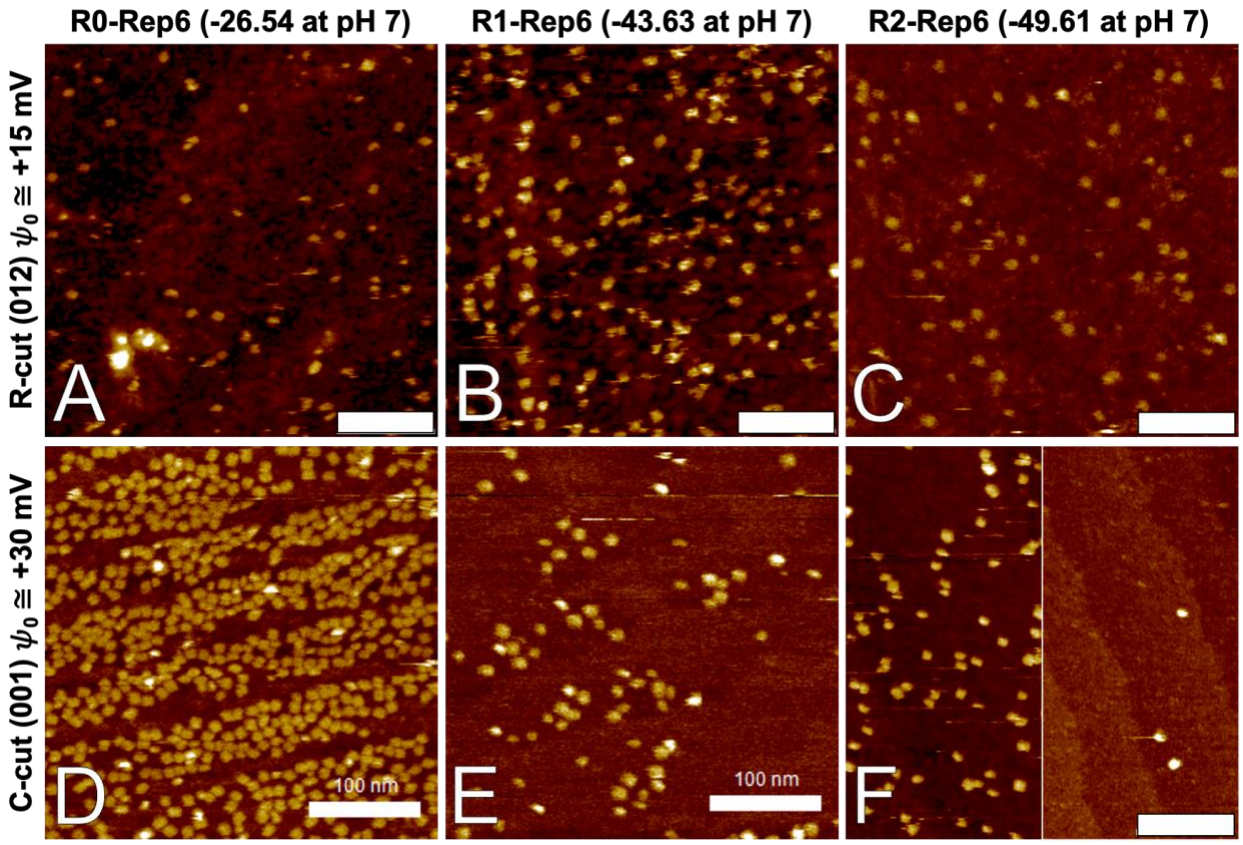

Figure 4.6. R-Rep6 proteins on R-cut and C-cut at pH 7. **(a-c)** Proteins on R-Cut at 0.02 μM. **(d)** R0-Rep6 at 0.15 μM on the C-cut. **(e)** R1 and **(f)** R2-Rep6, respectively at 0.1 μM on the C-Cut. **(f)** Split pane with representative views of heterogeneous coverage for B2-Rep6 as discussed in Figure 4.5d-f. All scale bars are 100 nm.

4.2.4 Repeat 9 proteins on the R and C-cut at pH 7 and 8

This set was expanded to longer R-Repeat9 proteins and was tested on the hematite R-cut and C-cut surfaces at pH 7 and 8. Our initial hypothesis was that if the behavior of these proteins could be described only by the electrostatic binding, at a given concentration we would observe the most binding events on the C-cut surface at pH 7. In contrast, if lattice-matching and other details of the designed interface are important, a larger amount of oriented binding should be seen on the target R-cut surface. In agreement with expectations for a predominant effect of the designed surface over the non-specific electrostatic binding, all of the proteins in the R-rep9 series showed

low or no binding at 0.1 μM and pH 8 on the off-target C-cut surface (Fig. 4.7a-c). Surface charge at pH 8 is expected to be $\psi_0 \cong +15$ mV, roughly half of that expected at pH 7 on the same (001) surface. On the same off-target (001) surface, non-specific binding events are noticeable when incubating at pH 7, this represents an increase in ψ_0 from +15 mV to +30 mV, potentially driving nonspecific interactions higher (Fig. 4.7d). These binding events are noticeable at 0.02 μM, five times less the concentration used in the incubation at pH 8 on the top of the figure (Fig. 4.7a-c). However, when tested at the same pH value and half of the concentration on the target R-cut surface, there are more binding events (Fig. 4,7e). This condition with multiple binding events occurs at the same ψ_0 as in Fig. 4.7a even though the concentration is 10 times lower. R0-Rep9 in this case is showing specific binding to the R-Cut over the C-Cut.

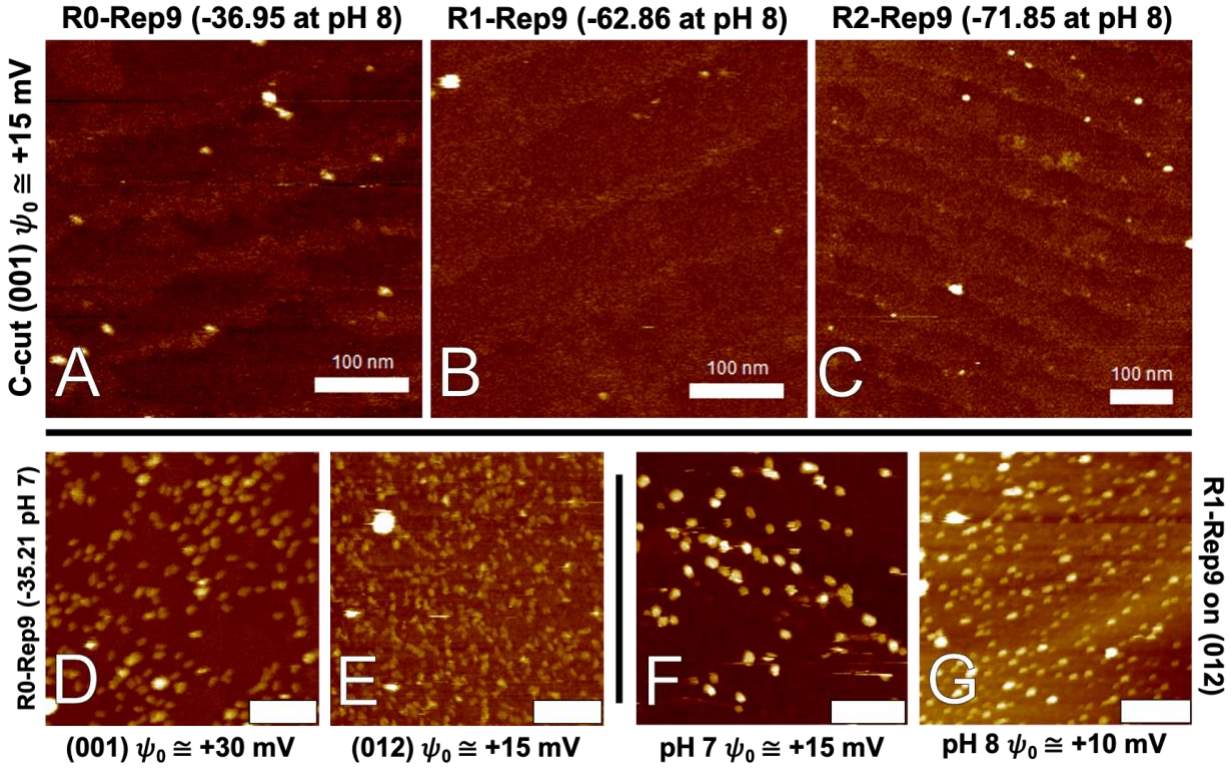

Figure 4.7. R-Rep9 proteins on R-Cut and C-Cut at pH 7 and 8. **(a-c)** Low or no binding observed for any R-Rep9 protein [0.1 μM] on the C-cut at pH 8. **(d)** Lower relative binding affinity R0-Rep9 of the C-Cut [0.02 μM] at pH 7 over **(e)** R0-Rep9 on the R-Cut [0.01 μM]. Binding to R-cut by R1-

Rep9 proteins [0.1 μM] is heterogeneous at **(f)** pH 7 and **(g)** pH 8. Images are representative of fields with binding events. Scale bar is 100 nm for all images.

R1-rep9 proteins, designed with one negatively-charged side, had an apparent higher coverage at pH 8 ($\psi_0 \cong +10$ mV) compared to pH 7 ($\psi_0 \cong +15$ mV) on the target R-cut surface (Fig. 4.7f,g). This behavior seemed to be heterogeneous on the same hematite piece, with large domains with no binding events. This is in contrast to the R0-Rep9 series, which had a large coverage even though the concentration is tenfold less.

4.2.5 Nucleation trials of hematite in the presence of proteins

Encouraged by this initial evidence of specific targeted behavior, we proceeded with testing for the effect of these scaffolds on the nucleation of hematite. A small difference in the interfacial energy would in theory have a large effect on the rate of nucleation of hematite nanoparticles, as described before in Section 1.2. These small changes in interfacial energy can be related to differences in coverage observed by AFM studies. The first set of trials involved mixing directly the protein with a ferrihydrite precursor as described in Section 2.4 (Fig. 4.8). Due to the thermostability of these proteins, supported by the CD experiments described in Sections 4.2.1 and 4.2.3, we expected that the designed proteins would hold the designed secondary and tertiary structures at these temperatures. Incubation with the protein was compared to a positive control with no additives, and Bovine Serum Albumin (BSA) was used to compare the effect of a native, negatively-charged, globular protein. The buffer of the protein was always MOPS 10 mM, as this was found to avoid nucleation inhibition in our control experiments. BSA, R2-Rep6, R0-Rep6, and C1-Rep6 were all found to inhibit the nucleation of hematite by different degrees, but there was no clear trend or drastic effect (Fig. 4.8). An alternative nucleation setup was pursued using a microdialysis membrane as described in Section 2.4. The objective of separating the ferrihydrite precursor from the protein by a membrane was to eliminate the potentially confounding variable

of proteins interacting with and binding to ferrihydrite, thus changing its role as a potential template for nucleation. A key test experiment for this involved a hematite seed in the inner chamber while ferrihydrite remained on the outer chamber (Fig. 4.8). If the membrane had no effect on the dissolution and transport of ions to the hematite seed, we would expect to see the growth of the nanocrystal inside the chamber in the absence of the protein, but the nanocrystal remained unchanged. An additional preliminary test suggested that even when ferrihydrite was incubated at high pH in the outer chamber, that is under dissolution conditions for ferrihydrite, no ions were detected in the inner chamber. Another setup was tested where a protein with a thiol group was immobilized on a gold surface (data not shown). This surface was then incubated in a solution containing ferrihydrite. The setup yielded inconclusive results. C1-Rep6 is the only protein that has been expressed successfully with a designed 8.7 Å spacing between repeating elements, whose target is the C-cut surface of hematite (Nucleation trial shown in Fig. 4.8). This protein appears to be mostly monomeric by SEC and nMS describes it as monomeric, but SEC-MALS shows evidence of a minor dimer peak, which can be interfering in the SAXS structural determination of this monomer (data not shown).

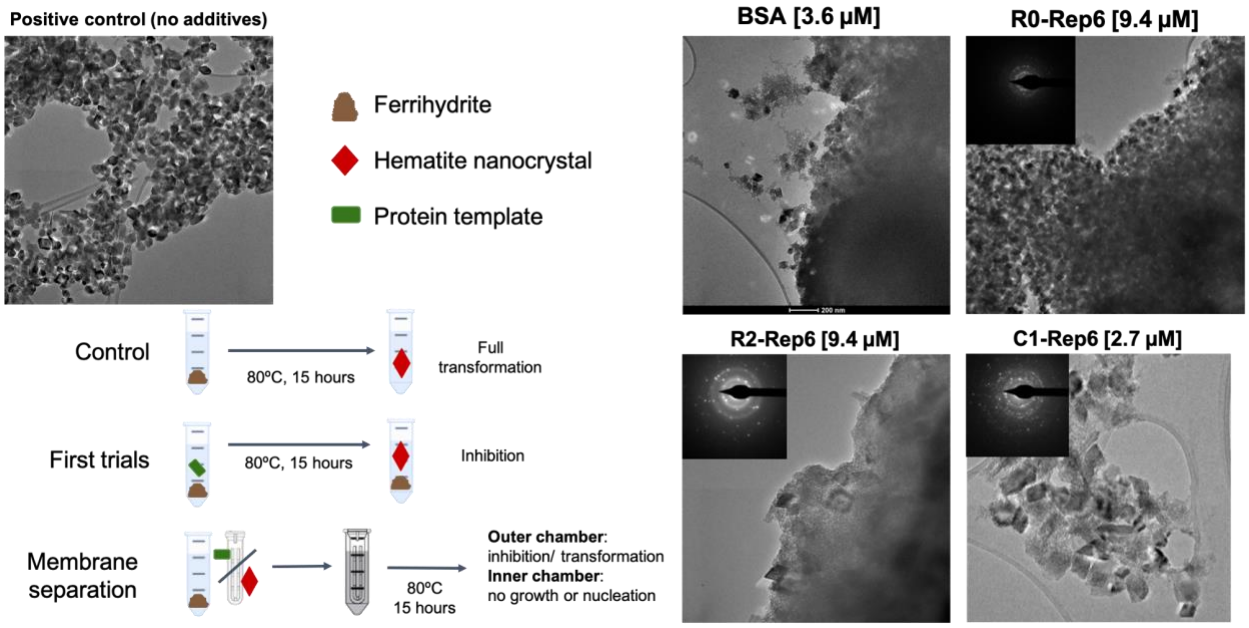

Figure 4.8. Hematite nucleation trial summary of preliminary data. Left: a schematic representing the two main setups tried and an EM showing hematite nanoparticles synthesized under conditions described for the control. Right: Hematite nucleation or inhibition outcomes for different protein identities and concentrations incubated directly with ferrihydrite as described in the 'First trials' scheme to the left.

4.3 Discussion

4.3.1 Explaining differential binding on C-Cut surface

Differences in binding on the C-Cut surface are consistent with reported heterogeneity in the terminations. There are at least four potential terminations on the (001) hematite surface: iron terminated, oxygen terminated, and versions of the previous two where there are some iron vacancies. Crystal Truncation Rod (CTR) studies have shown that if the hematite surface is fresh or aged, there will be different terminations[127]. In the case of the fresh hematite, it will be iron terminated, and oxygen terminated in the case of the aged sample, this is accompanied by a hydroxylation layer with varying degrees of coordination[128]. The aged sample also has more structured water on its surface. Binding through structured water can explain the binding heterogeneity observed in the C-Cut.

In comparison to the C-Cut, the R-Cut of hematite has more structured water as determined by in situ high resolution specular X-ray reflectivity measurements[129]. The same study points out that doubly-coordinated Fe groups are largely uncharged at circumneutral pH in the C-cut, whereas single coordinated groups on the R-Cut are always charged, even when the surface charge is zero. The atomic structure reported for the R-Cut describes two terminations where iron is under coordinated by the absence of one or two oxygens in high vacuum[35]. This is expected to be hydroxylated in water.

4.3.2 Explaining differential binding depending on surface potential

An important consideration for comparing binding events on either surface is the effect of pH on the surface potential. At pH 7, the R-cut surface has a surface potential of $\psi_0 \cong +15$ mV and the C-cut surface a potential of $\psi_0 \cong +30$ mV[34]. With a higher surface potential at pH 7, the C-cut surface could potentially lead to more instances of non-specific binding driven purely by electrostatic interactions between the negatively charged protein and the positively charged surface. Repeating these experiments at a set pH value where the surface potential is the same for both surfaces will allow us to understand the effect of electrostatic interactions between the surface and the protein, their role on specific binding, and explore the effect lattice-matching may have.

4.3.3 The effect of surface chemistry on binding to hematite

The original expectation was for proteins that had a more negative charge and a designed surface with exclusively negatively-charged side chains and small hydrophobics to have the highest affinity for the positively-charged surface. In contrast, proteins with positive and negative charges displayed the highest coverage and therefore binding affinity as seen in Fig. 4.3 and 4.7. This can be explained by the different coordination chemistries that have been discovered through biopanning studies and in native proteins interacting with metal oxides. For example, a hydroxyl-

proline-hydroxyl motif has been proposed as an archetype motif for binding hematite by a biopanning study[130], this motif is also found in the terminal region of OmcA, an extracellular cytochrome involved in mineral reduction[131]. MD simulations proposed hydrogen bonding to the hydroxyl layer of the surface as the binding mode. In a separate biopanning study, sequence analysis found that positively charged residues occurred two times more often than negatively charged ones, suggesting that interactions were driven by Pearson hard acid-hard base interactions and hydrogen bonding[132]. In sum, positive and negative charges have been found to interact with iron oxides and in the case of hematite it seems that hydrogen bonding interactions might be key.

4.4 Conclusion

Designed helical repeat templates offer a tunable surface, in terms of structure and chemistry, that is not found in natural or engineered systems interacting with iron oxides to date. Proteins with nine repeats, positively and negatively charged surfaces, and a target spacing of 10.9 Å (but a most likely interhelical spacing of 11.2 Å) are capable of interacting specifically with the target R-Cut surface. This happens under incubation conditions where the surface potential is comparable between this surface and the C-Cut surface and holds true even when the concentration of the binding protein is reduced ten-fold on the target surface. In the case of six-repeat proteins, molecules with a negative and hydrophobic binding surface bind with a preferred orientation on the R-Cut surface and their binding behavior on the C-Cut surface is heterogeneous and domain-dependent. Testing of other shapes and surface chemistries will be required to assess whether these effects are exclusive of this set of related molecules and to find design rules that will enable arbitrary targeting of oriented binding on a given iron oxide surface.

Nucleation studies showed that when incubating directly in the presence of ferrihydrite all of the tested proteins showed some degree of inhibition. Further attempts to separate ferrihydrite while keeping it as a source of ions were unsuccessful in establishing a proof-of-concept control experiment or their results were inconclusive. Further trials should consider the use of different precursors or synthesis methods and will be elaborated in Section 5.3.4.

The next suggested studies, described in Section 5.3, are required to gain a complete understanding of the degree of specificity of different variants, the termination chemistry and orientation of the substrate and the structuring of water at the interface and on the protein.

5 Chapter 5: FUTURE DIRECTIONS

5.1 Protein design enabled new avenues of research

As established in previous chapters, repeat proteins offer a topology that simplifies the sequence and structure search space and also provide the regularity and repetitiveness required for matching a crystalline lattice. So far, only a DHR topology has been tested, but there are alternatives such as alpha-beta repeat topologies as well as only beta-solenoid topology. Advantages of these topologies include a higher diversity of inter-repeat spacings becoming available, in particular very small spacings compared to what is possible for DHRs (8 Å in DHRs compared to 4.7 Å in beta-solenoids). As seen in Section 3.3.2 inter repeat distance together with the flexibility of the scaffold could be the key for epitaxial matching of different surfaces and for driving the nucleation of a given polymorph. Therefore, a priority for future work could be to test similar negatively-charged chemistries but threaded on scaffolds with different degrees of flexibility and inter-repeat distances. One characteristic that has been kept constant for simplicity and to minimize the degree of mismatch between the protein and the inorganic crystal is the flatness of the scaffold. However, testing scaffolds with different degrees of curvature and twist could reveal the relative importance of a flat template. In sum, varying the scaffold shape while preserving its chemistry would be a desirable next step.

The converse would be to keep the scaffold shape constant but to vary its chemistry. This is particularly important in the case for the hematite binders, since a mutation series for FD31 regarding calcium carbonate has already revealed some polymorph-specific nucleation behavior connected to the choice of surface chemistry (See Section 3.2.5). Native coordination motifs of proteins interacting with iron include carboxylate, imidazole, and oxygen phenolate groups[133]. Besides, MD simulations together with biopanning studies have pointed to serine and threonine

hydrogen-bonding groups as interaction motifs with the hydroxylate layer of hematite surfaces[130]. Further coordination groups that have been observed as canonical interaction motifs for calcium in proteins include the oxygen atoms in asparagines, glutamines, the carbonyl from backbones, and (to a lesser extent) serine and threonine[134]. These are not only responsible for binding directly to calcium, but also to establish a network of hydrogen bonding sites capable of stabilizing calcium-binding sites. Also, as described in Section 3.3.3, desolvation of ions at the interface driven by hydrophobic groups could lead to enhanced nucleation rates. Generally speaking, threading in partial coordination motifs with more diverse chemistries together with ion desolvation groups could lead to differences in binding and nucleation which will complement our vision of the surface chemistry required for good interactions with inorganic crystals.

The same variations in surface chemistry and scaffold shape could be explored for the formation of assemblies like the ones observed during the incubation of FD31 with calcium ions (Section 3.2.3). The shape and surface chemistry of these assemblies could dictate the formation of nuclei with different shapes, like spindles[135]. Another potential avenue for design could be ways of stabilizing the ca-protein assembly (so far they have been acting as sacrificial templates) and observing the resulting nucleation and assembly outcomes. Systematic testing would lead to design rules for the formation of ca-protein assemblies with different degrees of crystallinity that have target shapes and surface chemistries.

Going back to the lessons from natural systems, it is noteworthy that many of the effects of nucleation can be traced back to large interfaces of charged molecules in a regular spacing. Whichever is the organism's choice of a template, it does not act in isolation, but in environments and with other molecular assemblies that help to change the solution chemistry and concentration of ions. Such environments can be emulated using protein design by designing closed oligomers

with defined interior surfaces or by appending charged tails to recruit ions. The Baker Lab has designed a large variety of helical repeat proteins with different shapes, sizes, oligomeric states, and higher order symmetries[79,126]. Confinement effects can only be observed in oligomers. The geometry and shape of these cavities together with the surface chemistry are important aspects to consider.

5.1.1 Design enabled by molecular dynamics

MD simulations are required for the understanding of a diversity of phenomena tied to the nucleation of calcium carbonate by protein templates. Examples of this include: How can different templates capture ions and pre-organize them to different degrees? Based on this is it possible to rank or explain the observed differences in nucleation capabilities of different scaffolds or chemistries? Can MD give us an idea or model the structure of the ca-protein assemblies and their behavior in solution? What sort of interactions (perhaps weak) are required to explain the dissolution of these templates? Previous MD studies have found that a certain rigidity of the binding site is required for the effective behavior of ice-binding proteins[7], where slowing down the kinetics of water is essential. This translates to having a rigid side chain and binding site that can drive the immobilization of ions. In contrast, templates with more rigid side chains in the case of calcium carbonate lead to a lower nucleation efficiency. An alternative explanation could be by the interplay of the rigidity of the binding site with the flexibility of the scaffold so that interactions are preorganized enough to enable selective binding, but not too inflexible that a perfect fit is necessary[61]. Flexibility can also mean that the protein is able to accommodate more and different epitaxial matches, which could be driving up the affinity and nucleation activity of the template. Particularly in the case of proteins interacting with hematite, it is important to model the ordering of water on the surface of the protein and compare it with the ordering of water on the surface of

the hematite with different terminations. Water-mediated interactions seem to be key to understanding the binding behavior of these proteins since there seems to be different degrees of water ordering and different degrees of hydroxylation depending on the termination and face being exposed[127,129].

5.2 Calcium carbonate biomineralization

5.2.1 High-throughput testing alternatives

So far, only a low throughput method for the characterization of the effect of different proteins, which is the ex-situ TEM of reactions, has been tried. This is not only slow, but prone to drying effects. In order to screen more proteins with different shapes and surface chemistries, a high-throughput method is desired. To this end, three different methods were attempted. The first was to display proteins on the surface of yeast and then use optical microscopy to search for variants that promoted the formation of microcrystals, i.e. to look for changes in the number, size, and shape of the crystals being formed. A second effort consisted of the same microscopy readout, but using purified versions of the protein and different crystal growth setups. Nevertheless, these two setups proved to produce unreliable results and noisy data in addition to not providing polymorph-specific information (crystal shape is an unreliable way of determining crystal polymorph). A spectroscopy-based assay was tried next using ATR-FTIR. The formation of different polymorphs over time was monitored. A problem with this setup was the need to use different precursors from what was originally used for the nucleation studies in the presence of proteins. Different precursors meant in this case a significant change in the speciation of ions and supersaturation, making the initial results hard to compare with the ones obtained by this method.

Binding and nucleation have a direct relationship through the interfacial free energy[21]. It is possible, therefore, to search for potential nucleators using a screening assay panning for

binders. This can be done using a pull-down assay or follow a new method for high-throughput screening developed by members of the Baker group. A pull-down assay would consist of incubation of nanoparticles with proteins, followed by centrifugation, washes, and finally resuspension in a denaturing buffer for analysis by gel electrophoresis. The second high-throughput screening method consists of incubation of the nanoparticles with yeast displaying the protein followed by sorting by a flow cytometry assay. For the second method, original trials have been unsuccessful, but particle dispersion and stability can be screened using zeta potential measurements, Dynamic Light Scattering (DLS), and the presence of stabilizing agents such as polymers, e.g. EA3007, $(EO)_{68}$-$(MAA)_8$-$C_{12}H_{25}$[136]. In this case and ideally, the binding mechanism would be through the displacement of stabilizing ligands, although non-specific interactions of the protein with the polymer are also possible.

5.2.2 Alternatives and additional nucleation assays

There are alternative precursors and methodologies to carry out a nucleation-based screening of the effect of different proteins. By means of carbonate ester hydrolysis (dimethyl or diethyl carbonate), calcium carbonate precipitation can be induced from a completely homogeneous solution, and titration of carbonate can be done to a solution containing a constant calcium concentration[16]. This method is capable of establishing first heterogeneous and then homogeneous nucleation on a timescale suitable for kinetic investigations. Preliminary tests have found that DLS characterization at low protein concentrations (1 μM) does not detect the hydrodynamic radius of the protein, making it possible to detect the onset of particle formation (calcium carbonate scatters more light than proteins when in dilute conditions) driven by heterogeneous nucleation on the surface of proteins. The interfacial free energy of the template could be measured by varying the chemical potential and measuring this way the onset of

precipitation. This may be followed by end-point measurements using ATR-FTIR to determine the polymorph distribution of the formed nanocrystals. In essence, a change in the choice of a precursor can lead to studies of how effective a template is by measuring the interfacial free energy and to have additional polymorph information through ATR-FTIR.

An alternative way to determine the relative nucleation efficiency of different designed proteins is through a Cryo-TEM experiment. By measuring the nucleation density of a reaction quenched by a cryo-plunge in a quantitative EM experiment, the relative efficiency of a template can be determined in addition to providing information about the nuclei size and shape. Nonetheless, preliminary efforts pursuing this route have been unsuccessful since it seems that sample heterogeneity is dominating. Future research should focus on optimizing reaction and vitrification conditions to ensure a homogeneous sample from which quantitative data can be derived.

5.3 Hematite binding and nucleation

5.3.1 Alternatives in methodology

Another possibility would be a hybrid approach between protein design and bio-panning that could involve the rational design coupled to directed evolution of protein surfaces. For that, a high-throughput screening method would be needed. Efforts within our biomineralization group are exploring the possibility of using protein display methods and flow cytometry as a way to search the sequence and structure space of repeat proteins for different binding modes spanning a range of net charges, surface hydrophobicity and other charged or hydrogen-bonding interactions.

5.3.2 Completing AFM coverage and orientation experiments

In agreement with the lattice-matching hypothesis, we would like to observe preferential binding of lattice-matched proteins on their target surface, despite incubation conditions where

non-specific binding driven by electrostatics could be a major driving force in the system. This would be measured either by relative coverage under the two established pH conditions, or after establishing a different method to bias the surface potential (perhaps a target pH where a given surface potential is reached). In addition, we would like to observe the same specificity trend for the new set of proteins targeting the C-cut surface. An additional control protein would be desirable, where the lattice parameter is also off-target from both surfaces, and a comparison with native globular and non-repetitive proteins such as BSA and Lysozyme. Looking for quantitative ways of measuring the binding affinity from these AFM experiments or other methods would also be desired to see if small changes in affinity have an effect on nucleation or inhibition.

One aspect that has been a source of additional complexity, making results difficult to interpret, is the presence of different crystallographic domains and chemical terminations on the same surface, as well as the modification of these terminations introduced by cleaning treatments, incubation conditions and other contaminants. It is recommended to characterize the hematite substrate using X-ray Crystal Truncation Rod (CTR) and grazing incidence diffraction total scattering measurements. CTR will serve as a coarse mapping of crystallographic domains and their terminations; while grazing incidence diffraction total scattering will provide information of water structuring at the interface. This data can be compared with other AFM methods, finely mapping charge and surface potential (Kelvin probe AFM), and hydration structures (3D FFM AFM). Mapping interactions with the crystallographic domains and terminations would allow us to have a better model to understand why and how proteins are interacting with the surface.

5.3.3 Design, characterization and structural validation of proteins

It is a priority to generate long aspect ratio and unbounded one-dimensional oligomers that can confirm the oriented binding hypothesis or to help us understand why alternative orientations might be adopted and if there are any conditions that can bias a particular orientation. We are also working to obtain crystal structures for the key parent design. Structural characterization would validate the designed spacing.

In addition or complementary to the protein screen and re-design discussed in the section above, further design efforts should include a large screening effort to design scaffolds that are monodisperse in solution, can tolerate multiple mutations, are flat, are thermostable, and span a range of previously unexplored spacings to test for affinity and interactions between repeating elements and different metal oxide surfaces. Once this set is obtained, thorough surface mutations screening for different ordering of groups, a screen of net charges and hydrophobicity, and a measure of relative flexibility of the scaffold could give a systemic insight of structural and chemical features on the surface of repeating proteins that can tune their behavior with different metal oxide surfaces.

5.3.4 Nucleation studies

Regarding hematite or iron oxide nucleation, it is possible to extract information from inhibition studies. Another option is to look for selective growth inhibition of hematite nanocrystal seeds in the presence of proteins. Additionally, we are discussing an effort to observe heterogeneous nucleation of goethite, since the synthesis of this mineral does not involve the formation of ferrihydrite as an amorphous solid precursor and because hematite and goethite have similar spacings, making it compatible with the designed proteins. One more aspect to be tested is

to append peptides with the capability to recruit ions, or to catalyze non-specific precipitation. This would be in the hope to provide complementary capabilities to the templating effects that the proteins are designed to have.

5.4 Significance

Previous work was done either on natural systems, self-assembled monolayers, interactions of proteins with surfaces with no demonstration of their nucleation capabilities, or bio-panning of peptide libraries against different surfaces and demonstration of their nucleation efficiency. Designing protein building blocks to test hypotheses framed by classical nucleation theory would advance understanding of these interactions that span observations on all of the previous systems above. It is not only a new way of probing this problem, but also a way that is highly designable and that captures more information compared to the natural counterparts and other engineering systems.

The scope of this work is twofold: (i) to provide initial evidence of directed nucleation and oriented attachment of calcium carbonate driven by designed proteins with tunable size and surface chemistry; and (ii) to offer the first studies of oriented and specific attachment of designed proteins on two hematite facets at different pH values by expanding the lattice-matching hypothesis used for the design protein arrays on mica to iron oxide surfaces.

Lessons from these studies can offer understanding on how to rationally design mineral-protein interactions to target binding, nucleation and growth of other inorganic crystals of interest. In addition, lessons learned from this synthetic system can be applied to control biomineralization in natural systems, as a way of understanding disease and developing therapies.

6 References

1. Weiner S. Biomineralization: a structural perspective. J Struct Biol. 2008;163: 229–234.

2. Stevens CA, Bachtiger F, Kong X-D, Abriata LA, Sosso GC, Gibson MI, et al. A minimalistic cyclic ice-binding peptide from phage display. Nat Commun. 2021;12: 2675.

3. Davies PL. Ice-binding proteins: a remarkable diversity of structures for stopping and starting ice growth. Trends Biochem Sci. 2014;39: 548–555.

4. Ampaw A, Charlton TA, Briard JG, Ben RN. Designing the next generation of cryoprotectants - From proteins to small molecules. Pept Sci. 2019;111: e24086.

5. Naullage PM, Metya AK, Molinero V. Computationally efficient approach for the identification of ice-binding surfaces and how they bind ice. J Chem Phys. 2020;153: 174106.

6. Hudait A, Odendahl N, Qiu Y, Paesani F, Molinero V. Ice-Nucleating and Antifreeze Proteins Recognize Ice through a Diversity of Anchored Clathrate and Ice-like Motifs. J Am Chem Soc. 2018;140: 4905–4912.

7. Hudait A, Qiu Y, Odendahl N, Molinero V. Hydrogen-Bonding and Hydrophobic Groups Contribute Equally to the Binding of Hyperactive Antifreeze and Ice-Nucleating Proteins to Ice. J Am Chem Soc. 2019;141: 7887–7898.

8. Qiu Y, Hudait A, Molinero V. How Size and Aggregation of Ice-Binding Proteins Control Their Ice Nucleation Efficiency. J Am Chem Soc. 2019;141: 7439–7452.

9. Hudait A, Moberg DR, Qiu Y, Odendahl N, Paesani F, Molinero V. Preordering of water is not needed for ice recognition by hyperactive antifreeze proteins. Proc Natl Acad Sci U S A. 2018;115: 8266–8271.

10. Jackson AP, Vincent JFV, Turner RM. Comparison of nacre with other ceramic composites. J Mater Sci. 1990;25: 3173–3178.

11. Sun J, Bhushan B. Hierarchical structure and mechanical properties of nacre: a review. RSC Adv. 2012;2: 7617–7632.

12. Masica DL, Schrier SB, Specht EA, Gray JJ. De novo design of peptide-calcite biomineralization systems. J Am Chem Soc. 2010;132: 12252–12262.

13. De Yoreo JJ, Vekilov PG. 3. Principles of crystal nucleation and growth. In: Dove PM, De Yoreo JJ, Weiner S, editors. Biomineralization. Berlin, Boston: De Gruyter; 2003. pp. 57–94.

14. Davila-Hernandez FA, Jin B, Pyles H, Zhang S, Wang Z, Huddy TF, et al. Directing polymorph specific calcium carbonate formation with de novo protein templates. 2023.

15. Fu L-H, Dong Y-Y, Ma M-G, Yue W, Sun S-L, Sun R-C. Why to synthesize vaterite polymorph of calcium carbonate on the cellulose matrix via sonochemistry process? Ultrason Sonochem. 2013;20: 1188–1193.

16. Dietzsch M, Andrusenko I, Branscheid R, Emmerling F, Tremel W. Snapshots of calcium carbonate

formation - A step by step analysis. Zeitschrift für Kristallographie – Crystalline Materials. 2017;232: 255–265.

17. Kababya S, Gal A, Kahil K, Weiner S, Addadi L, Schmidt A. Phosphate-water interplay tunes amorphous calcium carbonate metastability: spontaneous phase separation and crystallization vs stabilization viewed by solid state NMR. J Am Chem Soc. 2015;137: 990–998.

18. Tobler DJ, Blanco JDR, Dideriksen K, Sand KK, Bovet N, Benning LG, et al. The Effect of Aspartic Acid and Glycine on Amorphous Calcium Carbonate (ACC) Structure, Stability and Crystallization. Procedia Earth and Planetary Science. 2014;10: 143–148.

19. Tobler DJ, Rodriguez-Blanco JD, Dideriksen K, Bovet N, Sand KK, Stipp SLS. Citrate Effects on Amorphous Calcium Carbonate (ACC) Structure, Stability, and Crystallization. Adv Funct Mater. 2015;25: 3081–3090.

20. Yu S-H, Cölfen H. Bio-inspired crystal morphogenesis by hydrophilic polymers. J Mater Chem. 2004;14: 2124–2147.

21. Hamm LM, Giuffre AJ, Han N, Tao J, Wang D, De Yoreo JJ, et al. Reconciling disparate views of template-directed nucleation through measurement of calcite nucleation kinetics and binding energies. Proc Natl Acad Sci U S A. 2014;111: 1304–1309.

22. Liu X, Pu J, Zeng S, Jin C, Dong S, Li J. Hyriopsis cumingii Hic52—A novel nacreous layer matrix protein with a collagen-like structure. Int J Biol Macromol. 2017;102: 667–673.

23. Sarashina I, Endo K. Primary structure of a soluble matrix protein of scallop shell; implications for calcium carbonate biomineralization. Am Mineral. 1998;83: 1510–1515.

24. Addadi L, Weiner S. Control and design principles in biological mineralization. Angew Chem Int Ed Engl. 1992;31: 153–169.

25. Hoang QQ, Sicheri F, Howard AJ, Yang DSC. Bone recognition mechanism of porcine osteocalcin from crystal structure. Nature. 2003;425: 977–980.

26. Ruiz-Agudo C, Lutz J, Keckeis P, King M, Marx A, Gebauer D. Ubiquitin Designer Proteins as a New Additive Generation toward Controlling Crystallization. J Am Chem Soc. 2019;141: 12240–12245.

27. Ping H, Wan Y, Xie H, Xie J, Wang W, Wang H, et al. Organized Arrangement of Calcium Carbonate Crystals, Directed by a Rationally Designed Protein. Cryst Growth Des. 2018;18: 3576–3583.

28. Weiner S, Dove PM. An Overview of Biomineralization Processes and the Problem of the Vital Effect. Rev Mineral Geochem. 2003;54: 1–29.

29. Schwertmann U, Cornell RM. Iron oxides in the laboratory: preparation and characterization. 2nd ed. Schwertmann U, Cornell RM, editors. Weinheim; New York: Wiley-VCH Verlag; 2008.

30. Finger LW, Hazen RM. Crystal structure and isothermal compression of Fe2O3, Cr2O3, and V2O3 to 50 kbars. J Appl Phys. 1980;51: 5362–5367.

31. Lin M, Tng L, Lim T, Choo M, Zhang J, Tan HR, et al. Hydrothermal Synthesis of Octadecahedral Hematite (α-Fe2O3) Nanoparticles: An Epitaxial Growth from Goethite (α-FeOOH). J Phys Chem

C. 2014;118: 10903–10910.

32. Ding D, Huang Y, Zhou C, Liu Z, Ren J, Zhang R, et al. Facet-controlling agents free synthesis of hematite crystals with high-index planes: Excellent photodegradation performance and mechanism insight. ACS Appl Mater Interfaces. 2016;8: 142–151.

33. Huang X, Hou X, Jia F, Song F, Zhao J, Zhang L. Ascorbate-promoted surface iron cycle for efficient heterogeneous Fenton alachlor degradation with hematite nanocrystals. ACS Appl Mater Interfaces. 2017;9: 8751–8758.

34. Chatman S, Zarzycki P, Rosso KM. Surface potentials of (001), (012), (113) hematite (α-Fe2O3) crystal faces in aqueous solution. Phys Chem Chem Phys. 2013;15: 13911–13921.

35. Kraushofer F, Jakub Z, Bichler M, Hulva J, Drmota P, Weinold M, et al. Atomic-Scale Structure of the Hematite α-Fe2O3(1$\bar{1}$02) "R-Cut" Surface. The Journal of Physical Chemistry C. 2018. pp. 1657–1669. doi:10.1021/acs.jpcc.7b10515

36. Wang X-G, Weiss W, Shaikhutdinov SK, Ritter M, Petersen M, Wagner F, et al. The Hematite (\alpha- Fe_2O_3) (0001) Surface: Evidence for Domains of Distinct Chemistry. Phys Rev Lett. 1998;81: 1038–1041.

37. von Rudorff GF, Jakobsen R, Rosso KM, Blumberger J. Fast Interconversion of Hydrogen Bonding at the Hematite (001)-Liquid Water Interface. J Phys Chem Lett. 2016;7: 1155–1160.

38. Lützenkirchen J, Heberling F, Supljika F, Preocanin T, Kallay N, Johann F, et al. Structure--charge relationship--the case of hematite (001). Faraday Discuss. 2015;180: 55–79.

39. Uebe R, Schüler D. Magnetosome biogenesis in magnetotactic bacteria. Nat Rev Microbiol. 2016;14: 621–637.

40. Wang L, Prozorov T, Palo PE, Liu X, Vaknin D, Prozorov R, et al. Self-assembly and biphasic iron-binding characteristics of Mms6, a bacterial protein that promotes the formation of superparamagnetic magnetite nanoparticles of uniform size and shape. Biomacromolecules. 2012;13: 98–105.

41. Rawlings AE, Bramble JP, Hounslow AM, Williamson MP, Monnington AE, Cooke DJ, et al. Ferrous iron binding key to Mms6 magnetite biomineralisation: A mechanistic study to understand magnetite formation using pH titration and NMR spectroscopy. Chemistry. 2016;22: 7885–7894.

42. Zhang H, Liu X, Feng S, Wang W, Schmidt-Rohr K, Akinc M, et al. Morphological transformations in the magnetite biomineralizing protein Mms6 in iron solutions: a small-angle X-ray scattering study. Langmuir. 2015;31: 2818–2825.

43. Staniland SS, Rawlings AE. Crystallizing the function of the magnetosome membrane mineralization protein Mms6. Biochem Soc Trans. 2016;44: 883–890.

44. Nudelman H, Zarivach R. Structure prediction of magnetosome-associated proteins. Front Microbiol. 2014;5: 9.

45. Zeytuni N, Ozyamak E, Ben-Harush K, Davidov G, Levin M, Gat Y, et al. Self-recognition mechanism of MamA, a magnetosome-associated TPR-containing protein, promotes complex assembly. Proc Natl Acad Sci U S A. 2011;108: E480–7.

46. Klem MT, Young M, Douglas T. Biomimetic synthesis of photoactive α-Fe2O3templated by the hyperthermophilic ferritin from Pyrococus furiosus. J Mater Chem. 2010;20: 65–67.

47. Bradley JM, Le Brun NE, Moore GR. Ferritins: furnishing proteins with iron. J Biol Inorg Chem. 2016;21: 13–28.

48. Oldfield CJ, Dunker AK. Intrinsically disordered proteins and intrinsically disordered protein regions. Annu Rev Biochem. 2014;83: 553–584.

49. Boskey AL, Villarreal-Ramirez E. Intrinsically disordered proteins and biomineralization. Matrix Biol. 2016;52-54: 43–59.

50. Wojtas M, Dobryszycki P, Ozyhar A. Intrinsically Disordered Proteins in Biomineralization. researchgate.net; 2012.

51. Kalmar L, Homola D, Varga G, Tompa P. Structural disorder in proteins brings order to crystal growth in biomineralization. Bone. 2012;51: 528–534.

52. Sharma V, Srinivasan A, Nikolajeff F, Kumar S. Biomineralization process in hard tissues: The interaction complexity within protein and inorganic counterparts. Acta Biomater. 2021;120: 20–37.

53. Henzler-Wildman K, Kern D. Dynamic personalities of proteins. Nature. 2007;450: 964–972.

54. Oh D, Qi J, Han B, Zhang G, Carney TJ, Ohmura J, et al. M13 virus-directed synthesis of nanostructured metal oxides for lithium-oxygen batteries. Nano Lett. 2014;14: 4837–4845.

55. Levin A, Hakala TA, Schnaider L, Bernardes GJL, Gazit E, Knowles TPJ. Biomimetic peptide self-assembly for functional materials. Nature Reviews Chemistry. 2020;4: 615–634.

56. Mahatabuddin S, Fukami D, Arai T, Nishimiya Y, Shimizu R, Shibazaki C, et al. Polypentagonal ice-like water networks emerge solely in an activity-improved variant of ice-binding protein. Proc Natl Acad Sci U S A. 2018;115: 5456–5461.

57. Sun T, Lin F-H, Campbell RL, Allingham JS, Davies PL. An antifreeze protein folds with an interior network of more than 400 semi-clathrate waters. Science. 2014;343: 795–798.

58. Ye Q, Eves R, Campbell RL, Davies PL. Crystal structure of an insect antifreeze protein reveals ordered waters on the ice-binding surface. Biochem J. 2020;477: 3271–3286.

59. Eickhoff L, Dreischmeier K, Zipori A, Sirotinskaya V, Adar C, Reicher N, et al. Contrasting Behavior of Antifreeze Proteins: Ice Growth Inhibitors and Ice Nucleation Promotors. J Phys Chem Lett. 2019. Available: https://pubs.acs.org/doi/abs/10.1021/acs.jpclett.8b03719

60. Nanda V. Heterogeneous Epitaxy: Designed Peptides Scale Graphene's Surface. Biophysical journal. 2016. pp. 2291–2292.

61. Grigoryan G, Kim YH, Acharya R, Axelrod K, Jain RM, Willis L, et al. Computational design of virus-like protein assemblies on carbon nanotube surfaces. Science. 2011;332: 1071–1076.

62. Mustata G-M, Kim YH, Zhang J, DeGrado WF, Grigoryan G, Wanunu M. Graphene Symmetry Amplified by Designed Peptide Self-Assembly. Biophys J. 2016;110: 2507–2516.

63. Pyles H, Zhang S, De Yoreo JJ, Baker D. Controlling protein assembly on inorganic crystals through

designed protein interfaces. Nature. 2019;571: 251–256.

64. Voet ARD, Noguchi H, Addy C, Zhang KYJ, Tame JRH. Biomineralization of a cadmium chloride nanocrystal by a designed symmetrical protein. Angew Chem Int Ed Engl. 2015;54: 9857–9860.

65. Huang P-S, Boyken SE, Baker D. The coming of age of de novo protein design. Nature. 2016;537: 320–327.

66. Epstein CJ, Goldberger RF. The genetic control of tertiary protein structure: studies with model systems. Cold Spring Harbor. 1963. Available: http://symposium.cshlp.org/content/28/439.short

67. Leman JK, Weitzner BD, Lewis SM, Adolf-Bryfogle J, Alam N, Alford RF, et al. Macromolecular modeling and design in Rosetta: recent methods and frameworks. Nat Methods. 2020;17: 665–680.

68. Schrödinger. Biologics Design. 2020. Available: https://www.schrodinger.com/science-articles/biologics-design

69. Chemical Computing Group. Molecular Operating Environment (MOE). 2020. Available: https://www.chemcomp.com/Products.htm

70. Systèmes D. BIOVIA, Discovery Studio Modeling Environment. 2016. Available: https://www.3ds.com/products-services/biovia/products/molecular-modeling-simulation/biovia-discovery-studio/

71. Brooks BR, Brooks CL 3rd, Mackerell AD Jr, Nilsson L, Petrella RJ, Roux B, et al. CHARMM: the biomolecular simulation program. J Comput Chem. 2009;30: 1545–1614.

72. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. J Comput Chem. 2004;25: 1157–1174.

73. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. GROMACS: fast, flexible, and free. J Comput Chem. 2005;26: 1701–1718.

74. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596: 583–589.

75. Zheng W, Li Y, Zhang C, Pearce R, Mortuza SM, Zhang Y. Deep-learning contact-map guided protein structure prediction in CASP13. Proteins. 2019;87: 1149–1164.

76. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al. Accurate prediction of protein structures and interactions using a three-track neural network. Science. 2021;373: 871–876.

77. Xu J, Wang S. Analysis of distance-based protein structure prediction by deep learning in CASP13. Proteins. 2019;87: 1069–1081.

78. Alford RF, Leaver-Fay A, Jeliazkov JR, O'Meara MJ, DiMaio FP, Park H, et al. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. J Chem Theory Comput. 2017;13: 3031–3048.

79. Brunette TJ, Parmeggiani F, Huang P-S, Bhabha G, Ekiert DC, Tsutakawa SE, et al. Exploring the repeat protein universe through computational protein design. Nature. 2015;528: 580–584.

80. Hicks DR, Kennedy MA, Thompson KA, DeWitt M, Coventry B, Kang A, et al. De novo design of protein homodimers containing tunable symmetric protein pockets. Proc Natl Acad Sci U S A. 2022;119: e2113400119.

81. Huddy TF, Hsia Y, Xu J, Kibler RD, Bethel N, Nagarajan D, et al. Geometrically programmable nanomaterial construction using regularized protein building blocks. 2023.

82. Koepnick B, Flatten J, Husain T, Ford A, Silva D-A, Bick MJ, et al. De novo protein design by citizen scientists. Nature. 2019;570: 390–394.

83. Dyer KN, Hammel M, Rambo RP, Tsutakawa SE, Rodic I, Classen S, et al. High-Throughput SAXS for the Characterization of Biomolecules in Solution: A Practical Approach. In: Chen YW, editor. Structural Genomics: General Applications. Totowa, NJ: Humana Press; 2014. pp. 245–258.

84. SAXS FrameSlice. [cited 22 Apr 2023]. Available: https://sibyls.als.lbl.gov/ran

85. Schneidman-Duhovny D, Hammel M, Tainer JA, Sali A. Accurate SAXS profile computation and its assessment by contrast variation experiments. Biophys J. 2013;105: 962–974.

86. Fallas JA, Ueda G, Sheffler W, Nguyen V, McNamara DE, Sankaran B, et al. Computational design of self-assembling cyclic protein homo-oligomers. Nat Chem. 2017;9: 353–360.

87. VanAernum ZL, Gilbert JD, Belov ME, Makarov AA, Horning SR, Wysocki VH. Surface-Induced Dissociation of Noncovalent Protein Complexes in an Extended Mass Range Orbitrap Mass Spectrometer. Anal Chem. 2019;91: 3611–3618.

88. Waitt GM, Xu R, Wisely GB, Williams JD. Automated in-line gel filtration for native state mass spectrometry. J Am Soc Mass Spectrom. 2011;19: 239–245.

89. Marty MT, Baldwin AJ, Marklund EG, Hochberg GKA, Benesch JLP, Robinson CV. Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. Anal Chem. 2015;87: 4370–4376.

90. Kilpatrick EL, Liao W-L, Camara JE, Turko IV, Bunk DM. Expression and characterization of 15N-labeled human C-reactive protein in Escherichia coli and Pichia pastoris for use in isotope-dilution mass spectrometry. Protein Expr Purif. 2012;85: 94–99.

91. Kabsch W. XDS. Acta Crystallogr D Biol Crystallogr. 2010;66: 125–132.

92. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, et al. Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr. 2011;67: 235–242.

93. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr. 2007;40: 658–674.

94. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr. 2010;66: 213–221.

95. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004;60: 2126–2132.

96. Bahn SY, Jo BH, Choi YS, Cha HJ. Control of nacre biomineralization by Pif80 in pearl oyster. Sci

Adv. 2017;3: e1700765.

97. Belcher AM, Wu XH, Christensen RJ, Hansma PK, Stucky GD, Morse DE. Control of crystal phase switching and orientation by soluble mollusc-shell proteins. Nature. 1996;381: 56–58.

98. Rao A, Roncal-Herrero T, Schmid E, Drechsler M, Scheffner M, Gebauer D, et al. On Biomineralization: Enzymes Switch on Mesocrystal Assembly. ACS Cent Sci. 2019;5: 357–364.

99. Raz S, Hamilton PC, Wilt FH, Weiner S, Addadi L. The transient phase of amorphous calcium carbonate in sea urchin larval spicules: The involvement of proteins and magnesium ions in its formation and stabilization. Adv Funct Mater. 2003;13: 480–486.

100. Jiang W, Pacella MS, Athanasiadou D, Nelea V, Vali H, Hazen RM, et al. Chiral acidic amino acids induce chiral hierarchical structure in calcium carbonate. Nat Commun. 2017;8: 15066.

101. DeOliveira DB, Laursen RA. Control of Calcite Crystal Morphology by a Peptide Designed To Bind to a Specific Surface. J Am Chem Soc. 1997;119: 10627–10631.

102. Aizenberg J, Black AJ, Whitesides GM. Oriented Growth of Calcite Controlled by Self-Assembled Monolayers of Functionalized Alkanethiols Supported on Gold and Silver. J Am Chem Soc. 1999;121: 4500–4509.

103. Addadi L, Weiner S. Interactions between acidic proteins and crystals: stereochemical requirements in biomineralization. Proc Natl Acad Sci U S A. 1985;82: 4110–4114.

104. Mann S. Molecular recognition in biomineralization. Nature. 1988;332: 119–124.

105. Duffy DM, Harding JH. Simulation of organic monolayers as templates for the nucleation of calcite crystals. Langmuir. 2004;20: 7630–7636.

106. Freeman CL, Hu Q, Nielsen MH, Tao J, De Yoreo JJ, Harding JH. Surface Selectivity of Calcite on Self-Assembled Monolayers. J Phys Chem C. 2013;117: 5154–5163.

107. Voets IK. From ice-binding proteins to bio-inspired antifreeze materials. Soft Matter. 2017;13: 4808–4823.

108. Niu Y-Q, Liu J-H, Aymonier C, Fermani S, Kralj D, Falini G, et al. Calcium carbonate: controlled synthesis, surface functionalization, and nanostructured materials. Chem Soc Rev. 2022;51: 7883–7943.

109. Fang P-A, Conway JF, Margolis HC, Simmer JP, Beniash E. Hierarchical self-assembly of amelogenin and the regulation of biomineralization at the nanoscale. Proc Natl Acad Sci U S A. 2011;108: 14097–14102.

110. Nielsen MH, Aloni S, De Yoreo JJ. In situ TEM imaging of $CaCO_3$ nucleation reveals coexistence of direct and indirect pathways. Science. 2014;345: 1158–1162.

111. Hu Q, Nielsen MH, Freeman CL, Hamm LM, Tao J, Lee JRI, et al. The thermodynamics of calcite nucleation at organic interfaces: Classical vs. non-classical pathways. Faraday Discuss. 2012;159: 509–523.

112. Zhu Y, Li Q, Kim D, Min Y, Lee B, Jun Y-S. Sulfate-Controlled Heterogeneous CaCO3 Nucleation and Its Non-linear Interfacial Energy Evolution. Environ Sci Technol. 2021.

doi:10.1021/acs.est.1c02865

113. Söngen H, Schlegel SJ, Morais Jaques Y, Tracey J, Hosseinpour S, Hwang D, et al. Water orientation at the calcite-water interface. J Phys Chem Lett. 2021;12: 7605–7611.

114. Zhu G, Sushko ML, Loring JS, Legg BA, Song M, Soltis JA, et al. Self-similar mesocrystals form via interface-driven nucleation and assembly. Nature. 2021;590: 416–422.

115. Cheng Y, Tao J, Zhu G, Soltis JA, Legg BA, Nakouzi E, et al. Near surface nucleation and particle mediated growth of colloidal Au nanocrystals. Nanoscale. 2018;10: 11907–11912.

116. De Yoreo JJ, Gilbert PUPA, Sommerdijk NAJM, Penn RL, Whitelam S, Joester D, et al. Crystallization by particle attachment in synthetic, biogenic, and geologic environments. Science. 2015;349: aaa6760.

117. He G, Dahl T, Veis A, George A. Nucleation of apatite crystals in vitro by self-assembled dentin matrix protein 1. Nat Mater. 2003;2: 552–558.

118. Boles MA, Ling D, Hyeon T, Talapin DV. Erratum: The surface science of nanocrystals. Nat Mater. 2016;15: 364.

119. Li D, Nielsen MH, Lee JRI, Frandsen C, Banfield JF, De Yoreo JJ. Direction-specific interactions control crystal growth by oriented attachment. Science. 2012;336: 1014–1018.

120. Liu L, Nakouzi E, Sushko ML, Schenter GK, Mundy CJ, Chun J, et al. Connecting energetics to dynamics in particle growth by oriented attachment using real-time observations. Nat Commun. 2020;11: 1045.

121. Kumar A, Mohanram H, Li J, Le Ferrand H, Verma CS, Miserez A. Disorder–Order Interplay of a Barnacle Cement Protein Triggered by Interactions with Calcium and Carbonate Ions: A Molecular Dynamics Study. Chem Mater. 2020;32: 8845–8859.

122. Garnham CP, Campbell RL, Davies PL. Anchored clathrate waters bind antifreeze proteins to ice. Proc Natl Acad Sci U S A. 2011;108: 7363–7367.

123. Giuffre AJ, Hamm LM, Han N, De Yoreo JJ, Dove PM. Polysaccharide chemistry regulates kinetics of calcite nucleation through competition of interfacial energies. Proc Natl Acad Sci U S A. 2013;110: 9261–9266.

124. Schüler T, Renkel J, Hobe S, Susewind M, Jacob DE, Panthöfer M, et al. Designed peptides for biomineral polymorph recognition: a case study for calcium carbonate. J Mater Chem B Mater Biol Med. 2014;2: 3511–3518.

125. Lakshminarayanan R, Kini RM, Valiyaveettil S. Investigation of the role of ansocalcin in the biomineralization in goose eggshell matrix. Proc Natl Acad Sci U S A. 2002;99: 5155–5159.

126. Hsia Y, Mout R, Sheffler W, Edman NI, Vulovic I, Park Y-J, et al. Design of multi-scale protein complexes by hierarchical building block fusion. Nat Commun. 2021;12: 2294.

127. Lützenkirchen J, Heberling F, Supljika F, Preocanin T, Kallay N, Johann F, et al. Structure-charge relationship - the case of hematite (001). Faraday Discuss. 2015;180: 55–79.

128. Trainor TP, Chaka AM, Eng PJ, Newville M, Waychunas GA, Catalano JG, et al. Structure and

reactivity of the hydrated hematite (0001) surface. Surf Sci. 2004;573: 204–224.

129. Catalano JG. Weak interfacial water ordering on isostructural hematite and corundum (001) surfaces. Geochim Cosmochim Acta. 2011;75: 2062–2071.

130. Lower BH, Lins RD, Oestreicher Z, Straatsma TP, Hochella MF Jr, Shi L, et al. In vitro evolution of a peptide with a hematite binding motif that may constitute a natural metal-oxide binding archetype. Environ Sci Technol. 2008;42: 3821–3827.

131. Edwards MJ, Baiden NA, Johs A, Tomanicek SJ, Liang L, Shi L, et al. The X-ray crystal structure of Shewanella oneidensis OmcA reveals new insight at the microbe–mineral interface. FEBS Lett. 2014;588: 1886–1890.

132. You F, Yin G, Pu X, Li Y, Hu Y, Huang Z, et al. Biopanning and characterization of peptides with Fe3O4 nanoparticles-binding capability via phage display random peptide library technique. Colloids Surf B Biointerfaces. 2016;141: 537–545.

133. Baker HM, Anderson BF, Baker EN. Dealing with iron: common structural principles in proteins that transport iron and heme. Proc Natl Acad Sci U S A. 2003;100: 3579–3583.

134. Pidcock E, Moore GR. Structural characteristics of protein binding sites for calcium and lanthanide ions. J Biol Inorg Chem. 2001;6: 479–489.

135. Wang F, Cao B, Mao C. Bacteriophage Bundles with Prealigned Ca2+ Initiate the Oriented Nucleation and Growth of Hydroxylapatite. Chem Mater. 2010;22: 3630–3636.

136. Faatz M, Gröhn F, Wegner G. Amorphous calcium carbonate: Synthesis and potential intermediate in biomineralization. Adv Mater. 2004;16: 996–1000.