A bacterial surface layer protein exploits multistep crystallization for rapid self-assembly

Jonathan Herrmann[a,b], Po-Nan Li[b,c], Fatemeh Jabbarpour[a,b], Anson C. K. Chan[d], Ivan Rajkovic[b], Tsutomu Matsui[b], Lucy Shapiro[e], John Smit[d], Thomas M. Weiss[b], Michael E. P. Murphy[d] [ID], and Soichi Wakatsuki[a,b,1] [ID]

[a]Department of Structural Biology, Stanford University, Stanford, CA 94305; [b]Biosciences Division, SLAC National Accelerator Laboratory, Menlo Park, CA 94025; [c]Department of Electrical Engineering, Stanford University, Stanford, CA 94305; [d]Department of Microbiology and Immunology, University of British Columbia, Vancouver, BC V6T 1Z3, Canada; and [e]Department of Developmental Biology, Stanford University, Stanford, CA 94305

Surface layers (S-layers) are crystalline protein coats surrounding microbial cells. S-layer proteins (SLPs) regulate their extracellular self-assembly by crystallizing when exposed to an environmental trigger. However, molecular mechanisms governing rapid protein crystallization in vivo or in vitro are largely unknown. Here, we demonstrate that the *Caulobacter crescentus* SLP readily crystallizes into sheets in vitro via a calcium-triggered multistep assembly pathway. This pathway involves 2 domains serving distinct functions in assembly. The C-terminal crystallization domain forms the physiological 2-dimensional (2D) crystal lattice, but full-length protein crystallizes multiple orders of magnitude faster due to the N-terminal nucleation domain. Observing crystallization using a time course of electron cryo-microscopy (Cryo-EM) imaging reveals a crystalline intermediate wherein N-terminal nucleation domains exhibit motional dynamics with respect to rigid lattice-forming crystallization domains. Dynamic flexibility between the 2 domains rationalizes efficient S-layer crystal nucleation on the curved cellular surface. Rate enhancement of protein crystallization by a discrete nucleation domain may enable engineering of kinetically controllable self-assembling 2D macromolecular nanomaterials.

protein self-assembly | Cryo-EM time course | microbiology | biophysics | crystal nucleation

The structural determinants of macromolecular phase transitions are important to govern a wide variety of biological processes and disease states (1, 2). Microbial surface-layer proteins (SLPs) are highly expressed and ubiquitous proteins that self-assemble on the outside of bacteria and archaea to form crystalline protein coats (3–5). SLPs undergo a phase transition from aqueous to solid as part of their biological assembly and function (6, 7). RsaA, the 98-kDa SLP from *Caulobacter crescentus*, undergoes environmentally triggered self-assembly, forming sheets of crystalline protein with hexameric 22-nm repeats when exposed to calcium (7, 8). Recent work from our laboratory demonstrated that fast, efficient protein crystallization drives S-layer assembly in vivo (9). Therefore, we examined RsaA self-assembly in vitro using time-resolved small angle X-ray scattering (SAXS), circular dichroism (CD) spectroscopy, and a time course of electron cryo-microscopy (Cryo-EM) images to determine the structural basis for S-layer assembly.

RsaA consists of 2 domains, both of which are necessary for successful S-layer assembly in vivo (Fig. 1A) (10, 11). The N-terminal domain, here defined as RsaA$_{1-246}$, is responsible for anchoring the protein to the cell surface via a noncovalent interaction with specific sugar moieties of the lipopolysaccharide (LPS) outer membrane (11–13). In contrast, the C-terminal region is responsible for making the intermolecular contacts necessary to form the physiological S-layer lattice (10). A recent 2.7-Å resolution crystal structure revealed that the C terminus, RsaA$_{249-1026}$, consists of a series of β-strands arranged in an L-shape and connected by loops interspersed with calcium-binding Repeat-in-Toxin (RTX) motifs (Fig. 1 A and B) (10, 14). Docking this

crystal structure into a 7.4-Å resolution electron cryo-tomographic reconstruction of the repeating hexameric unit of the native S-layer lattice suggests the identity of the side chains coordinating the physiological S-layer crystal contacts (Fig. 1A and *SI Appendix*, Fig. S1) (10). This S-layer reconstruction also revealed the location and overall topology of the N-terminal domain, which forms an anchoring ring beneath and around the center of each hexameric repeat made by the C-terminal crystallization domain (Fig. 1A) (10, 11).

Using superresolution microscopy and single-molecule tracking, we have recently shown that during native S-layer assembly, RsaA monomers diffuse while noncovalently attached to the cell surface until incorporated into a growing crystal lattice (9). When no previous S-layer lattice exists, RsaA nucleates 2D crystals at remarkably low concentrations to form stable crystalline rafts with as few as ~50 molecules (9). Advances in imaging modalities such as atomic force microscopy (AFM) and Cryo-EM have allowed for the direct observation of macromolecular nucleation and crystallization for a select few systems at the nanometer scale (15–18). These studies have revealed canonical and noncanonical assembly
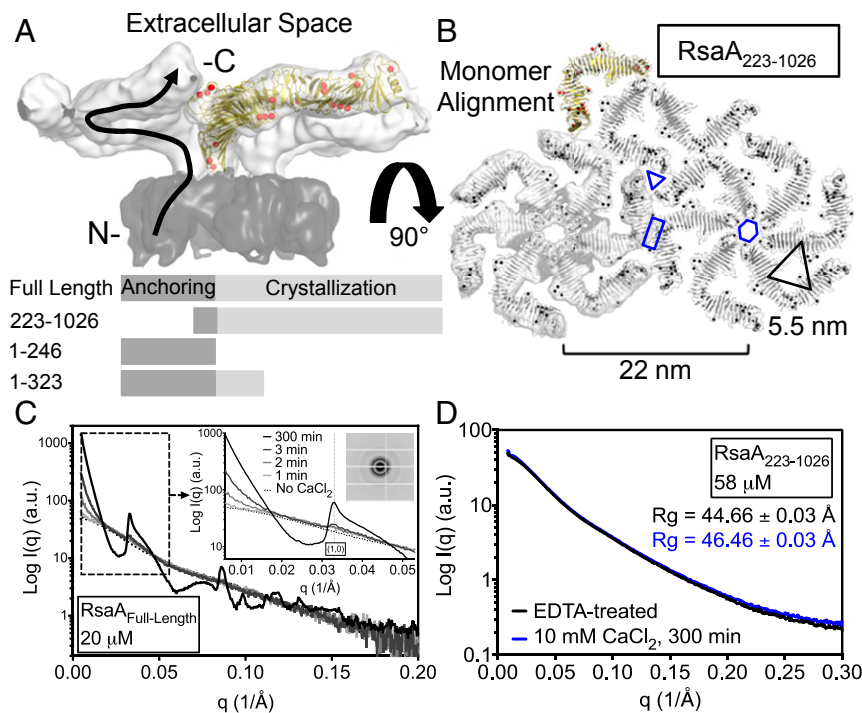
Fig. 1. The RsaA N terminus enhances the rate of crystallization. (*A*) The published cryo-tomographic reconstruction of the RsaA hexameric repeat is shown with the crystallization domain surface in light gray and the anchoring domain surface in dark gray (EMD-3604). The previously determined crystal structure of the crystallization domain is docked into 1 arm of the hexamer with protein ribbon in gold and calcium ions as red spheres (PDB ID code 5N97). Below, a graphical representation of the 4 RsaA truncation constructs used in this study is given. (*B*) The RsaA$_{223-1026}$ crystal structure is shown (gray ribbons with black spheres for strontium ions) with symmetry mates forming the physiological S-layer lattice. A hexamer is shown aligned with the cryo-tomographic reconstruction from *A* (gray surfaces), and an alignment of monomers from both crystal structures is shown at the top. Conserved dimeric, trimeric, and hexameric interfaces are noted by a blue rectangle, triangle, and hexagon, respectively. The 22-nm characteristic spacing between hexameric repeats is noted at the bottom, and a black 5.5-nm equilateral triangle notes a pseudosymmetric center. (*C*) Time-resolved SAXS of 20 μM RsaA$_{Full-Length}$ shows the appearance of powder diffraction when calcium is added. Bragg signal due to powder diffraction (detector image *Inset*) appears within 2 min after calcium addition (plot inset marked by dashed box). (*D*) Time-resolved SAXS of 58 μM RsaA$_{223-1026}$ shows no evidence of self-association after 5 h of incubation with 10 mM CaCl$_2$.

pathways, including other bacterial S-layer proteins, which can exhibit structurally distinct kinetic intermediates (15, 17, 19–21). Therefore, we asked how the kinetic pathway of RsaA self-assembly might enable efficient S-layer assembly observed in vivo. Using rationally designed RsaA truncations (Fig. 1*A*) and a combination of X-ray crystallography, SAXS, Cryo-EM, stability assays, and CD spectroscopy, we identify in vitro structural dynamics that enable rapid 2D crystallization of the *C. crescentus* SLP, RsaA.

Results

The RsaA N Terminus Enhances the Rate of Crystallization. Recent attempts to determine the high-resolution structure of RsaA$_{Full-Length}$ resulted in a 2.7-Å resolution crystal structure of the C-terminal domain (RsaA$_{249-1026}$) (10). This 3-dimensional (3D) crystal is composed of stacked sheets of hexameric 22-nm repeats characteristic of the 2D physiological S-layer crystal lattice (10). To determine whether the unknown structure of the N terminus contributed to the formation of stacked sheets of the S-layer lattice, RsaA$_{223-1026}$, was cloned, purified, crystallized by the hanging drop method, and diffraction data collected to 2.1-Å resolution (22). RsaA$_{223-1026}$ was crystallized with strontium as a calcium analog, and the previously reported structure (10) was used as a search model for the molecular replacement method (*SI Appendix,* Table S1). The RsaA$_{223-1026}$ crystal structure consists of stacked hexameric sheets with substantially similar calcium-binding sites and crystal contacts within and between subunits as previously determined (Fig. 1*B* and *SI Appendix,* Fig. S1) (10). This crystal packing indicates that the crystallization domain can stably

assemble into stacked sheets of the S-layer lattice without the N-terminal domain.

To compare the kinetics by which RsaA$_{Full-Length}$ and RsaA$_{223-1026}$ self-assemble in the presence of CaCl$_2$, we performed time-resolved SAXS on both protein forms. Previous work showed that purified RsaA$_{Full-Length}$ crystallizes into sheets upon the addition of 10 mM CaCl$_2$ (7, 8). RsaA$_{Full-Length}$ at 20 μM exhibited Bragg peaks consistent with a 2D protein crystal lattice within 2 min of calcium addition (Fig. 1*C*) (7). By 5 h, a clear powder diffraction pattern emerged, consistent with a lattice of hexameric 22-nm repeats (Fig. 1*C*). Time-resolved SAXS of 62 and 100 μM RsaA$_{Full-Length}$ using a stopped-flow mixer to introduce calcium quickly (within 50 ms) showed evidence of rapid assembly. At these concentrations, Bragg peaks appeared within seconds of calcium addition (*SI Appendix,* Fig. S2). In stark contrast, 58 μM RsaA$_{223-1026}$ remained soluble and monomeric 5 h after calcium addition (Fig. 1*D*). Therefore, the RsaA N terminus enhances the rate of in vitro crystallization by at least 3 orders of magnitude.

A Calcium-Triggered Conformational Change Precedes Nucleation. Calcium-triggered self-assembly of RsaA was examined by subjecting multiple concentrations of RsaA$_{Full-Length}$ to SAXS analysis precisely 1 min after calcium addition (Fig. 2 *A–C*). Crystallization in suspension can be monitored by quantifying scattering signal in the scattering angles corresponding to the (1,0) diffraction peak of the RsaA crystal lattice ($0.032 < q < 0.048$ Å^{-1}). RsaA crystallization as measured 1 min after calcium addition is concentration-dependent, with initial signal appearing at 25 μM and steady-state crystal growth observed for concentrations above 35 μM (Fig. 2*A*).
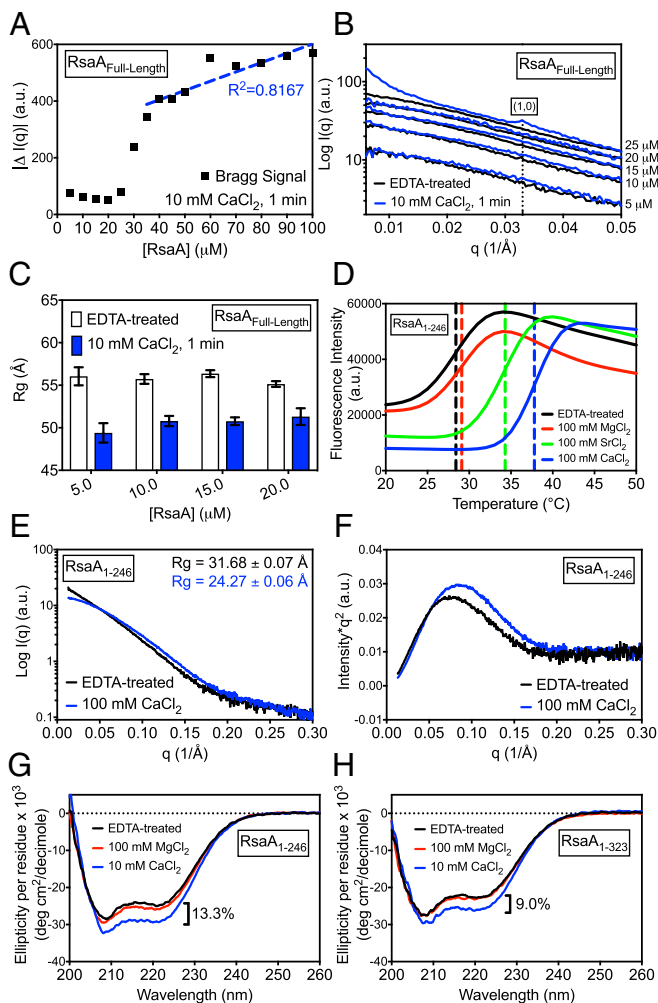
Fig. 2. A conformational change within the N terminus precedes nucleation. (*A*) Bragg signal from SAXS profiles of a variety of concentrations of RsaA$_{Full-Length}$ 1 min after 10 mM CaCl$_2$ addition reveals steady-state crystallization at concentrations above 35 μM. Bragg signal was measured for the (1,0) diffraction peak located at $0.032 < q < 0.048$ Å^{-1}. (*B*) SAXS profiles of EDTA-treated (black) or prenucleation (10 mM CaCl$_2$ for 1 min, blue) RsaA$_{Full-Length}$ samples reveal a calcium-induced change in scattering profile. (*C*) Differences in Rg of EDTA-treated RsaA$_{Full-Length}$ (black) or samples 1 min after adding 10 mM CaCl$_2$ (blue) show a decrease in Rg by ∼4 Å before crystallization. Error bars are SEM. (*D*) Differential scanning fluorimetry assays of RsaA$_{1-246}$ treated with EDTA (black), 100 mM MgCl$_2$ (red), 100 mM SrCl$_2$ (green), or 100 mM CaCl$_2$ (blue) shows calcium-specific stabilization with melting temperatures denoted by vertical dashed lines. (*E* and *F*) SAXS profiles and Kratky plots of EDTA-treated (black) or 100 mM CaCl$_2$-added (blue) RsaA$_{1-246}$ shows a change in scattering profile corresponding to a decrease in Rg by 7.4 Å. (*G* and *H*) Circular dichroism spectra of RsaA$_{1-246}$ and RsaA$_{1-323}$ are shown treated with EDTA (black), 100 mM MgCl$_2$ (red), or 10 mM CaCl$_2$ (blue). A decrease in the relative signal change induced by calcium indicates that the observed folding change occurs entirely within RsaA$_{1-246}$.

To determine whether any conformational changes occur within the protein after calcium addition, but before the appearance of crystals, SAXS data from monomeric RsaA were compared to data collected 1 min after calcium addition for RsaA at and below 25 μM (Fig. 2*B*). A change in scattering intensity was observed at low scattering angles prior to the appearance of Bragg diffraction signal (Fig. 2*B*). This signal corresponds to a decrease in the calculated radius of gyration (Rg) from ∼55 Å for calcium-free monomeric protein to ∼51 Å for precrystallized (calcium added) RsaA$_{Full-Length}$ (Fig. 2*C* and *SI Appendix*, Fig. S3). Therefore, a

calcium-triggered conformational change occurs after calcium addition, but before self-association.

The N Terminus Increases Secondary Structure when Exposed to Calcium. SAXS profiles and CD spectra of RsaA$_{223-1026}$ with and without 10 mM CaCl$_2$ show no significant differences despite many known calcium-binding sites within this region of the protein (Fig. 1 *B* and *D* and *SI Appendix*, Fig. S4). The N-terminal domain may then be responsible for the calcium-dependent conformational change observed in Fig. 2*B*. Although RsaA$_{1-246}$ lacks putative calcium binding motifs, calcium-specific stabilization was observed by differential scanning fluorimetry (Fig. 2*D* and *SI Appendix*, Fig. S5). To probe structural changes due to calcium binding, we collected SAXS profiles of RsaA$_{1-246}$ with and without 100 mM CaCl$_2$ (Fig. 2*E*). RsaA$_{1-246}$ displayed a significant change in scattering profile corresponding to a decrease in Rg of 7.4 Å and an increase in globularity (Fig. 2 *E* and *F* and *SI Appendix*, Fig. S4).

Given the calcium-specific conformational change observed for RsaA$_{1-246}$, we utilized CD spectroscopy to compare the secondary structure of the N terminus in the presence of 5 mM ethylenediaminetetraacetic acid (EDTA), 10 mM CaCl$_2$, or 100 mM MgCl$_2$. A calcium-specific response was evident, corresponding to a 13.3% increase in secondary structure (Fig. 2*G*). To determine whether this effect was confined to the N terminus, we created a hybrid construct including all of the N terminus and the first 7 β-strands (77 amino acids) of the C terminus, RsaA$_{1-323}$ (Fig. 1*A*). CD spectra of this construct revealed a calcium-specific 9.0% increase in secondary structure (Fig. 2*H*), corresponding to a 32% decrease in relative signal with a 27% increase in relative mass between the 2 constructs. Therefore, the C-terminal 77 amino acids in RsaA$_{1-323}$ do not contribute to the observed calcium-induced change in secondary structure, which occurs entirely within the N-terminal domain, RsaA$_{1-246}$.

A Short-Lived Intermediate Crystal Lattice Is Observed during Nucleation. To observe the steps leading to the appearance of the RsaA crystal lattice, we imaged RsaA nucleation using a time course of Cryo-EM images. Samples of 20 μM RsaA$_{Full-Length}$ were frozen on an electron microscopy (EM) grid 30 s, 60 s, 90 s, and 120 s after adding 10 mM CaCl$_2$. Multiple structural states of RsaA were observed (Fig. 3). The first structure, termed the intermediate lattice, appears within 30 s after calcium addition (*SI Appendix*, Fig. S6) and dominates the viewing area at the 60-s time point (Fig. 3*A* and *SI Appendix*, Fig. S6). The intermediate lattice exhibits a characteristic spacing of 5.5 nm, which may reflect pseudosymmetric trigonal spacing formed between neighboring monomers within each repeating hexamer (Figs. 1*B* and 3 *A*, *Insets*). At 90 s, the sample becomes heterogeneous, consisting of the previously observed intermediate lattice as well as the mature lattice, which exhibits the expected 22-nm spacing in Fourier space (*SI Appendix*, Fig. S6). At 120 s, single and stacked layers of mature lattices are predominantly observed (Fig. 3*B* and *SI Appendix*, Figs. S6 and S7). Quantification of the appearance of the intermediate and mature lattices throughout the time course reveals the rapid appearance and disappearance of the intermediate lattice within 120 s (Fig. 3*C*). Therefore, a time course of Cryo-EM imaging revealed a short-lived crystalline intermediate during RsaA nucleation.

The Intermediate Crystal Lattice Is Partially Ordered. To elucidate the structural differences between the intermediate and mature lattices at roughly nanometer resolution, 2D averages of the repeating hexamers composing each crystalline state were calculated. These structures were then compared to the previously determined subtomographic average of the physiological S-layer (EMD-3604) and a truncated version corresponding to the crystallization domain alone (using PDB ID code 5N97
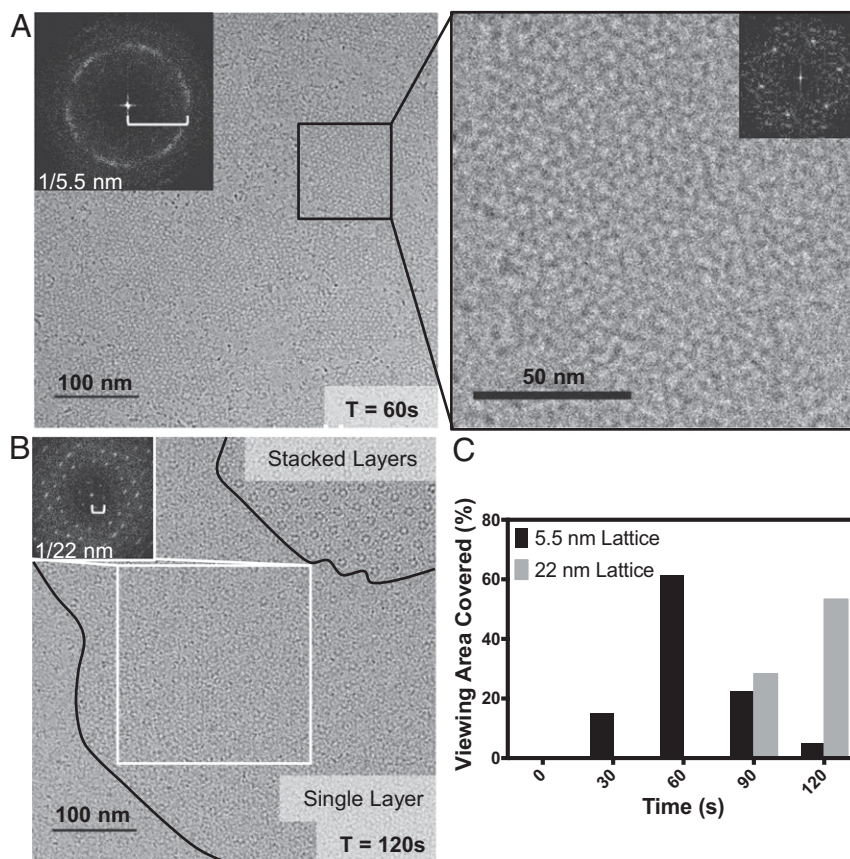
Fig. 3. RsaA exhibits a short-lived intermediate crystal lattice during nucleation. (*A*) A Cryo-TEM image of 20 μM RsaA$_{Full-Length}$ 60 s after 10 mM CaCl$_2$ addition shows crystal lattices with a characteristic spacing of 5.5 nm (*Insets* are FFTs). (*B*) A Cryo-TEM image of 20 μM RsaA$_{Full-Length}$ 120 s after calcium addition shows a crystal lattice with single-layer and stacked-layer sections (boundaries noted by black lines) exhibiting a characteristic spacing of 22 nm (*Inset* is FFT of white box). (*C*) Quantitating the viewing area covered by 5.5 nm and single or stacked layer 22-nm crystal lattices for at least 2.9 μm^2 of total area at each time point shows appearance and disappearance of the 5.5-nm lattice.

as a guide) (Fig. 4 *A–D*). The intermediate state possesses the pinwheel structure characteristic of the crystallization domain alone (Fig. 4 *A* and *B*) and contrasts with the physiological RsaA lattice, which possesses significant density in a ring near the center of each RsaA hexamer (Fig. 1 *A* and *B*). Averaging hexameric repeats from single-layer mature lattices shows electron
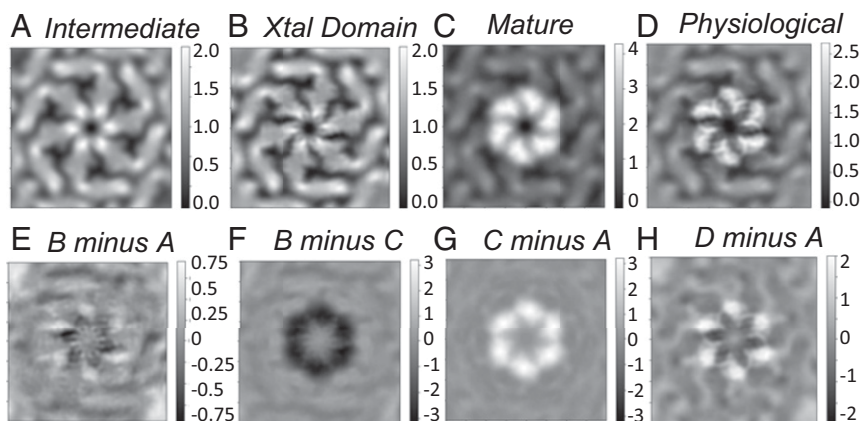

Fig. 4. The N-terminal domain exhibits motional dynamics in the crystalline intermediate. (*A*) A 2D average of 1,044 hexameric particles from intermediate lattices (both 30 s and 60 s time points) as shown in Fig. 3*A* is produced. (*B*) The subtomographic average of the native crystallization domain alone is shown (EMD-3604). (*C*) A 2D average of 1,163 hexameric particles from mature single layer lattices as shown in Fig. 3*B* is produced. (*D*) The subtomographic average of the complete native hexameric repeating unit is shown (EMD-3604). (*E*) The difference map subtracting the intermediate (*A*) from the crystallization domain alone (*B*) shows near complete signal removal. (*F*) The difference map subtracting the time-resolved mature lattice (*C*) from the crystallization domain alone (*B*) shows negative differences around the center of the repeating hexameric unit (black). (*G*) The difference map subtracting the time-resolved intermediate lattice (*A*) from the time-resolved mature lattice (*C*) produces positive differences around the center of the repeating hexameric unit (white). (*H*) Subtracting the time-resolved intermediate (*A*) lattice from the physiological lattice (*D*) produces localized positive differences at the location of the N-terminal anchoring/nucleation domain (white). Scale bars denote pixel intensities.
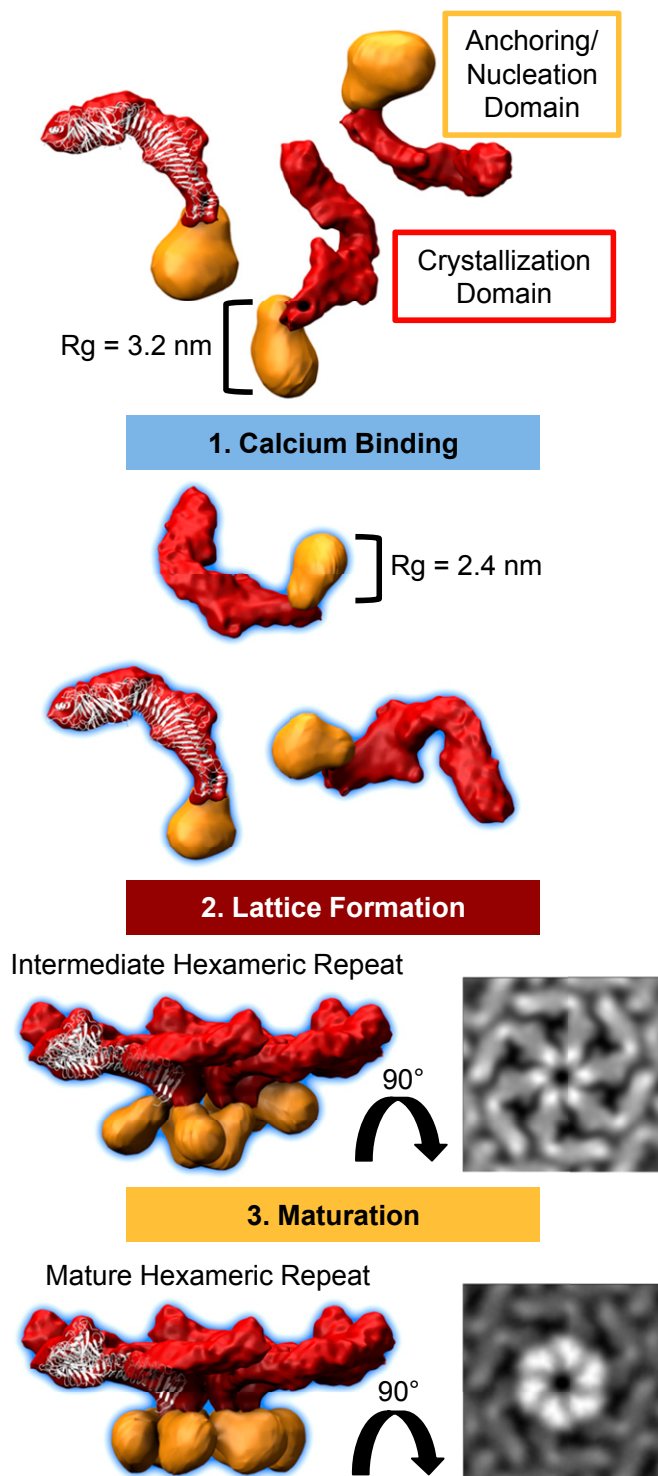
Fig. 5. Rapid in vitro RsaA crystallization involves a multistep pathway. Without calcium, the crystallization domain of RsaA (red surface, white crystal structure) is folded, but the N terminus (orange surface) exists in a partially unfolded state. Upon calcium introduction, the N terminus folds and compacts (Rg decreases by 7.4 Å). Then, monomers begin to assemble into a lattice using the crystallization domain. However, the N terminus retains motion and therefore does not appear in the 2D average of the intermediate state as observed by Cryo-TEM. Later, the N terminus locks into place as confirmed by another 2D average, forming a state resembling the physiological crystal lattice. Thus, a multistep pathway enables rapid protein crystallization for this bacterial SLP.

density that more closely approximates the physiological lattice (Fig. 4 *C* and *D*).

To better define the structural dynamics within and between these states, difference maps were calculated between these averaged images (Fig. 4 *E–H*). Subtracting the intermediate state (Fig. 4*A*) from the crystallization domain (Fig. 4*B*) yields little residual signal (Fig. 4*E*), indicating that this domain is the predominant region of the protein that is ordered during this state. Subtracting the single-layer mature lattice (Fig. 4*C*) from the crystallization domain alone (Fig. 4*B*) creates negative differences in a ring around the center of the repeating hexameric unit (Fig. 4*F*). Strikingly, subtracting the intermediate lattice (Fig. 4*A*) from either the mature or physiological lattices (Fig. 4 *C* and *D*) produces positive electron density in a similar ring around the symmetric center of the repeating hexameric unit (Fig. 4 *G* and *H*). These difference maps indicate that the intermediate lattice lacks electron density at the location corresponding to the N-terminal anchoring domain. Thus, the N-terminal domain retains flexibility in the intermediate state until the subsequent mature conformation is completed.

Discussion

Despite the vast diversity of SLPs (5, 23), the molecular interactions responsible for maintaining a physiological S-layer lattice have been elucidated by crystallography and confirmed by EM for only RsaA; although, a near-atomic model has been proposed for SbsB, the SLP from the Gram-positive bacterium *Geobacillus stearothermophilus*, which is based on nanobody-aided 3D crystallization (6, 10). Calcium binding in SbsB was shown to directly mediate interactions between domains, thereby compacting the structure into the conformation suitable for crystallization (6). In the case of RsaA, although calcium-binding sites are evident throughout the crystallization domain, all crystal contacts between monomers occur through amino acid interactions, not by coordinating divalent calcium ions (Fig. 1*B* and *SI Appendix*, Fig. S1) (10). This observation was supported by our SAXS and CD analyses of RsaA$_{223-1026}$ indicating that calcium binding does not significantly change the structure of the crystallization domain (Fig. 1*D* and *SI Appendix*, Fig. S4). However, calcium stabilizes the secondary structure of RsaA and binds near residues directly involved in crystal contacts such as P693 and T758 at the dimeric interface and T256 at the hexameric interface (*SI Appendix*, Figs. S1 and S5) (7, 10).

The 2.1-Å resolution crystal structure of RsaA$_{223-1026}$ revealed that although calcium binding along the crystallization domain does not induce an observable conformational change, this binding behavior is sufficient to assemble stacked sheets of the physiological S-layer lattice (Fig. 1*B*). However, this macroscopic assembly of RsaA$_{223-1026}$ occurred over multiple days in the presence of precipitants and was aided by a slow concentration process due to vapor equilibration (22). In stark contrast, upon adding 10 mM CaCl$_2$ to purified RsaA$_{Full-Length}$, self-assembly occurs readily as observed with small angle X-ray diffraction within 2 min at 20 µM (Fig. 1*C*). As a direct comparison, RsaA$_{223-1026}$ failed to self-assemble in 5 h under the same buffer conditions and a higher protein concentration (Fig. 1*D*). Therefore, the N-terminal domain of RsaA enhances self-assembly kinetics and acts as a nucleation domain.

Monitoring protein crystal nucleation at the nanoscale by a time course of Cryo-EM images (Fig. 3) revealed that after calcium binding, RsaA crystallizes into sheets via a multistep pathway. At 20 µM, RsaA$_{Full-Length}$ self-assembles into mature crystals within 120 s (Figs. 1*C* and 3*B*) by first forming a structurally distinct short-lived crystalline intermediate at earlier time points (Figs. 3 and 4). Within this intermediate state, the nucleation domain appears flexible with respect to the rigid lattice formed concurrently by the crystallization domain (Figs. 4 and 5). The nucleation domain appears to reach its mature conformation

by later forming a hexagonally symmetric ring around the center of each hexameric repeat made by the crystallization domain (Figs. 4C and 5). This mature conformation agrees with the physiological S-layer structure, in which the nucleation domain resides beneath the crystallization domain (Figs. 1A, 4D, and 5). Crystallization of SbpA, the calcium-triggered SLP from *Lysinibacillus sphaericus*, was recently observed by time-resolved AFM and exhibited a structural intermediate evidenced by discrete changes in crystal height (21, 24). Delayed assembly of specific domains, such as the N-terminal domain of RsaA (Fig. 5), could be responsible for intermediate conformations observed for other SLPs like SbpA.

The crystallization domain possesses all of the residues necessary to form the long-range S-layer lattice, yet we observe a nucleation domain-dependent multistep assembly pathway of RsaA$_{Full-Length}$. How does the nucleation domain mediate assembly kinetics? The precise mechanism by which this nucleation domain enhances the rate of self-assembly without directly participating in most intermolecular contacts remains unclear. By isolating and examining RsaA truncations, we determined that the nucleation domain undergoes a calcium-specific conformational change before crystallization, likely mediated by specific structural changes including increasing helical content of the domain upon binding calcium (Fig. 2). This observed calcium-induced conformational change might affect the formation of the hexameric interface of the crystallization domain, which is just downstream in the RsaA sequence. This change may also expose additional intermolecular interfaces to be utilized during delayed maturation of the nucleation domain. Segregating the biochemical steps in this assembly pathway would likely require high-resolution structural knowledge of the nucleation domain to guide rational mutagenesis. Alternatively, in the context of a short-lived crystalline intermediate, flexibility between domains may provide access to a larger landscape of conformational microstates, which could act as an additional entropic driving force for successful crystal nucleation. A careful study of the energetics of RsaA self-assembly may be required to fully elucidate the mechanism of nucleation rate enhancement observed herein.

S-layer assembly in vivo occurs through continuous protein crystallization as individual secreted RsaA monomers diffuse on the LPS outer membrane until reaching the edge of a growing S-layer crystal (9). Copious environmental calcium ensures that RsaA assumes its crystallizable conformation upon secretion (7). Whether this conformation is required for S-layer anchoring is unknown; however, calcium depletion causes S-layer shedding, which could also be explained by partial unfolding and aggregation of the protein (7, 9). Upon reintroducing calcium to a liquid culture of *Caulobacter* cells that have shed their S-layer, RsaA nucleates quickly and efficiently at concentrations as low as 1–10 nM (9), suggesting an evolutionary basis for noncanonical crystallization kinetics if the proposed assembly pathway occurs on the cellular surface. High-resolution imaging of the *Caulobacter* S-layer in situ either by Cryo-EM or by superresolution fluorescence imaging has only observed S-layer assembly postnucleation, thereby precluding direct visual evidence of this multistep assembly mechanism in vivo (9, 10). However, the topology of the bacterial surface has been shown to affect S-layer crystal growth in vivo (9). Flexibility between domains might allow the crystallization domain to position itself for lattice formation while relieving torque originating from anchoring a 2D lattice on a variably curved surface. Our results further indicate that SLPs may segregate kinetic regulation of assembly using a structurally remote nucleation domain that enables fast and robust crystallization through an intermediate. This modularity raises the possibility of designing biologically inspired, self-assembling macromolecular nanomaterials with controllable nucleation kinetics.

Materials and Methods

Protein Purification. RsaA$_{Full-Length}$ and RsaA$_{223-1026}$ were purified from *C. crescentus* cells while RsaA$_{1-246}$ and RsaA$_{1-323}$ were recombinantly expressed and purified from *Escherichia coli*. All proteins were purified by size exclusion chromatography prior to use, and purity was assessed by SDS/PAGE (*SI Appendix*, Fig. S8). A detailed description of the method used to purify each RsaA construct can be found in *SI Appendix*.

X-ray Crystal Structure Solution and Refinement. Structure determination and refinement of RsaA$_{223-1026}$ was performed by molecular replacement and is described fully in *SI Appendix*. Refinement and modeling statistics can be found in the *SI Appendix*, Table S1. Map and coordinate files are available as ID code 6P5T in the Worldwide Protein Data Bank.

SAXS/Diffraction. SAXS/D measurements were collected at the bio-SAXS beamline 4-2 at Stanford Synchrotron Radiation Lightsource (SSRL) (25) as well as the Structurally Integrated Biology for Life Sciences (SIBYLS) beamline 12.3.1 at the Advanced Light Source (ALS) (26). SAXS/D experiments are fully described in *SI Appendix*.

Differential Scanning Fluorimetry. Stability measurements of purified RsaA protein were performed using the Thermofluor assay, which consisted of 45 µL of protein solution between 0.2 mg/mL and 1 mg/mL mixed with 5 µL of 10× ion solution (or water) as well as 0.5 µL of 10,000× SYPRO Orange Protein Gel Stain (excitation/emission wavelength, λ = 490/590 nm) (Thermo Fisher Scientific). Temperature was increased at a rate of 1 °C per minute from 4 °C to 100 °C, and fluorescence was measured every minute with a qPCR thermocycler in FRET mode (Bio-Rad). Aggregation temperature was determined by locating the global minimum of the second derivative of the raw data. Binding data were fit to a single-site binding model using Prism (GraphPad).

CD Spectroscopy. CD measurements were performed using a J-815 Circular Dichroism Spectrometer (Jasco). Far-UV spectra (200–260 nm) were recorded in a 1-mm path-length quartz cell with an exposure time of 1 s/nm. The sample cell was maintained at 10 °C, and 3 scans were collected and averaged for each sample. RsaA$_{1-246}$, RsaA$_{1-323}$, and RsaA$_{223-1026}$ were brought to a final concentration of 1–3 µM and a uniform volume of CaCl$_2$, MgCl$_2$, or buffer was added to each sample before acquiring spectra. A buffer spectrum with the appropriate added ion was subtracted from each sample spectrum before plotting.

EM. A time course of Cryo-EM imaging was performed using a 200-keV TF20 electron microscope (Thermo Fisher Scientific) equipped with a K2 direct electron detector (Gatan). A complete description of EM data collection and analysis can be found in *SI Appendix*.

Data Availability Statement. Supplementary information figures (9), 1 table, and supplemental methods are provided with this manuscript, which include representative images from all datasets and time points. Full Cryo-EM time course datasets can be found in the Stanford Digital Repository, https://purl.stanford.edu/tz311fw8492, and materials are available from S.W. upon reasonable request. The atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB), http://www.wwpdb.org/ (PDB ID code 6P5T).

BIOPHYSICS AND COMPUTATIONAL BIOLOGY

1. R. Schönherr, J. M. Rudolph, L. Redecke, Protein crystallization in living cells. *Biol. Chem.* **399**, 751–772 (2018).
2. S. Boeynaems *et al*., Protein phase separation: A new phase in cell biology. *Trends Cell Biol.* **28**, 420–435 (2018).
3. J. H. Y. Lau, J. F. Nomellini, J. Smit, Analysis of high-level S-layer protein secretion in *Caulobacter crescentus*. *Can. J. Microbiol.* **56**, 501–514 (2010).
4. D. Pum, J. L. Toca-Herrera, U. B. Sleytr, S-layer protein self-assembly. *Int. J. Mol. Sci.* **14**, 2484–2501 (2013).
5. C. Zhu *et al*., Diversity in S-layers. *Prog. Biophys. Mol. Biol.* **123**, 1–15 (2017).
6. E. Baranova *et al*., SbsB structure and lattice reconstruction unveil Ca2+ triggered S-layer assembly. *Nature* **487**, 119–122 (2012).
7. J. Herrmann *et al*., Environmental calcium controls alternate physical states of the *Caulobacter* surface layer. *Biophys. J.* **112**, 1841–1851 (2017).
8. J. F. Nomellini, S. Kupcu, U. B. Sleytr, J. Smit, Factors controlling *in vitro* re-crystallization of the *Caulobacter crescentus* paracrystalline S-layer. *J. Bacteriol.* **179**, 6349–6354 (1997).
9. C. J. Comerci *et al*., Topologically-guided continuous protein crystallization controls bacterial surface layer self-assembly. *Nat. Commun.* **10**, 2731 (2019).
10. T. A. M. Bharat *et al*., Structure of the hexagonal surface layer on *Caulobacter crescentus* cells. *Nat. Microbiol.* **2**, 17059 (2017).
11. M. J. Ford, J. F. Nomellini, J. Smit, S-layer anchoring and localization of an S-layer-associated protease in *Caulobacter crescentus*. *J. Bacteriol.* **189**, 2226–2237 (2007).
12. M. D. Jones, E. Vinogradov, J. F. Nomellini, J. Smit, The core and O-polysaccharide structure of the *Caulobacter crescentus* lipopolysaccharide. *Carbohydr. Res.* **402**, 111–117 (2015).
13. S. G. Walker, D. N. Karunaratne, N. Ravenscroft, J. Smit, Characterization of mutants of *Caulobacter crescentus* defective in surface attachment of the paracrystalline surface layer. *J. Bacteriol.* **176**, 6312–6323 (1994).
14. A. Chenal, J. I. Guijarro, B. Raynal, M. Delepierre, D. Ladant, RTX calcium binding motifs are intrinsically disordered in the absence of calcium: Implication for protein secretion. *J. Biol. Chem.* **284**, 1781–1789 (2009).
15. M. Sleutel, A. E. S. Van Driessche, Nucleation of protein crystals–A nanoscopic perspective. *Nanoscale* **10**, 12256–12267 (2018).
16. A. E. S. Van Driessche *et al*., Molecular nucleation mechanisms and control strategies for crystal polymorph selection. *Nature* **556**, 89–94 (2018).
17. M. Sleutel, J. Lutsko, A. E. S. Van Driessche, M. A. Durán-Olivencia, D. Maes, Observing classical nucleation theory at work by monitoring phase transitions with molecular precision. *Nat. Commun.* **5**, 5598 (2014).
18. S. Chung, S.-H. Shin, C. R. Bertozzi, J. J. De Yoreo, Self-catalyzed growth of S layers via an amorphous-to-crystalline transition limited by folding kinetics. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 16536–16541 (2010).
19. A. Sauter *et al*., Real-time observation of nonclassical protein crystallization kinetics. *J. Am. Chem. Soc.* **137**, 1485–1491 (2015).
20. A. Breitwieser, J. Iturri, J. L. Toca-Herrera, U. B. Sleytr, D. Pum, In vitro characterization of the two-stage non-classical reassembly pathway of S-layers. *Int. J. Mol. Sci.* **18**, 5–8 (2017).
21. S.-H. Shin *et al*., Direct observation of kinetic traps associated with structural transformations leading to multiple pathways of S-layer assembly. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 12968–12973 (2012).
22. M. D. Jones, A. C. K. Chan, J. F. Nomellini, M. E. P. Murphy, J. Smit, Surface-layer protein from *Caulobacter crescentus*: Expression, purification and X-ray crystallographic analysis. *Acta Crystallogr. F Struct. Biol. Commun.* **72**, 677–680 (2016).
23. P.-N. Li *et al*., Nutrient transport suggests an evolutionary basis for charged archaeal surface layer proteins. *ISME J.* **12**, 2389–2402 (2018).
24. B. Stel, F. Cometto, B. Rad, J. J. De Yoreo, M. Lingenfelder, Dynamically resolved self-assembly of S-layer proteins on solid surfaces. *Chem. Commun. (Camb.)* **54**, 10264–10267 (2018).
25. I. L. Smolsky *et al*., Biological small-angle X-ray scattering facility at the Stanford Synchrotron Radiation Laboratory. *J. Appl. Cryst.* **40** (suppl. 1), s453–s458 (2007).
26. K. N. Dyer *et al*., "High-throughput SAXS for the characterization of biomolecules in solution: A practical approach" in *Structural Genomics*, Y. W. Chen, Ed. (Methods in Molecular Biology, Humana Press, Totowa, NJ), pp. 245–258.