Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

69451 Weinheim, Germany

Conformational Transitions at an S-Layer Growing Boundary Resolved by Cryo-TEM**

Luis R. Comolli, Cristina E. Siegerist, Seong-Ho Shin, Carolyn Bertozzi, William Regan, Alex Zettl, and Jim De Yoreo*

anie_201300543_sm_miscellaneous_information.pdf

Supporting Information

Introduction: a broad application range for these techniques

The use of a single graphene layer as the support for S-layer self-assembly provides an ideal, robust flat surface on which to capture the expanding boundary by flash freezing and cryo-TEM imaging. This approach can be applied to a number of biological polymers. With S-layers, image analysis of the intermediate states spanning the transition from an open boundary to a lattice enables the classification of the main conformational states driving the process. With other self-assembling polymeric proteins, such as fibrils, the focus would be the growing end and the branching points. Although cryo-TEM is not a real time technique, frozen samples contain the ensemble of all conformational states throughout a process and powerful image analysis techniques allow us to sort them out in discrete classes. If we know the structure and stoichiometry of the oligomeric ordered final state, the conformational intermediates can be sorted as function of "reaction time".

Three-dimensional information from rSbpA

Using this polymerization protocol,[S1] the truncated sequence rSbpA (1-1068) self-assembles into ordered, well behaved lattices in bulk suspensions.[S2] However SbpA forms complex 3D patterns in bulk, and requires a supporting surface to grow into extended lattices. Cryogenic electron tomography (cryo-ET) reconstructions of bulk suspensions of rSbpA lattices and sub-volumetric averaging of ~ 3,000 S-layer patches have provided 3D information at ~ 25Å-30Å resolution. The lattice is spanned by homotetrameric subunits of approximately 10 nm in diameter, 13 nm in height, and with a central pore of 2 nm.[S2] We use this information to assign S-layer orientations on our graphene support. Low dose cryo-TEM images optimized for two-dimensional analysis were acquired using a pixel size at the specimen of 0.35 nm, adequate to describe the assembly of a lattice with a period of approximately 13.3 nm.

Amyloid self-assembly is driven by the nucleation of disordered intermediate oligomers which propagate along fibrils trapping misfolded, non-functional proteins. Quite the opposite, the self-assembly of cytoskeletal polymers such as tubulin are driven by the reversible addition of folded, globular polymeric subunits, and require external energy. We describe here a polymerization process which leads to the entropically irreversible incorporation of the fully folded oligomeric unit to the ordered array, utilizing aspects of both processes above. The two-dimensional view of the steps involved in monomer recruitment and conformational rearrangement at the S-layer growing boundary has also provided new information on the intra- and inter-monomer connectivity underlying the lattice (Figures 3, S2-S4). The same sample preparation and image processing methodology should be used for higher resolution work on this and other protein self-assembly problems.

While providing a mechanically strong and elastic support, the one-atom thickness of a single graphene layer, ~ 0.34 nm,[10] is very thin relative to the mean free

path of electrons in biological TEM. Because the probabilities of inelastic and multiple scattering are negligibly small relative to biological samples and relative to alternative supports such as amorphous carbon, graphene contributes minimally to the background noise and to the phase contrast. Furthermore, graphene is periodic with a lattice constant of 2.46 Å, a resolution higher than the goal for most biological cryo-TEM imaging. A single graphene layer is thus transparent in images acquired at lower resolution, as in most current biological imaging applications. Finally, because graphene is a good conductor, typical image degradation artifacts originating from charging can potentially be effectively removed or greatly diminished.

As with all studies using support surfaces for self-assembly and growth of S-layers such as mica, supported lipid bilayers, or graphene in this case, a shortcoming of this study is that it does not directly address the growth mechanism on the surface of microorganisms. Their surface is never covered "*de-novo*", and growth happens by addition of monomers extruded from the cytoplasm after translation to an existing S-layer without boundaries.

Supporting Experimental Section

Cryo-TEM specimen preparation

Wild type SbpA protein was purified as previously described[6] and stored at 4 ºC for later use. The buffer used to induce polymerization and self-assembly consisted of 10 mM Tris pH 7.1, 50 mM $CaCl_2$. To induce self-assembly of 2D lattices, monomeric wt SbpA protein was incubated with the buffer directly on TEM grids. The TEM grids used for this work were made as described before,[10] using 200 and 400 mesh gold TEM grids covered with a thin, lacey carbon film ideal for biological cryo-TEM as support. The graphene sheets rested on the thin lacey carbon film. TEM grids were discharged in an ion plasma cleaner at half power for 20 sec before use. This exposure proved enough for making the grids hydrophilic while preserving their integrity.

The grids were placed on the inside of the covers or caps of 0.6 ml tubes resting upside down. First the grids were "glow-discharged", and then placed on the caps. A drop of approximately 15 µl to 20 µl of polymerization buffer was placed on the grids, and finally an aliquot of the protein suspension was applied (~ 1.5 µl to 2.0 µl to reach a final protein concentration of ~ 100 to to 200 µgr/ml). Equivalently, the buffer and protein were applied to grids held by anti-capillary tweezers and then placed on the caps. The Eppindorf tube was then inserted into the upside down cap to seal the drop and avoid evaporation. The polymerization reaction was incubated for times of between 5 hours (200 µgr/ml) and 12 hours (100 µgr/ml). The grids were then directly taken by a cryo-plunging tweezers, mounted on a cryo-plunger,[15] blotted and flash-frozen. In samples incubated overnight (12 hours) many sheets of S-layers overlapping on the support were found by imaging, while areas with single S-layer sheets and free graphene surface were found only after incubation for shorter times. Samples used for image analysis were made with the latter.

The protein concentration used for these polymerization reactions was approximately ten times lower than for S-layer self-assembly in bulk solution.[11] This is an indirect verification of growth on a supporting surface, driven by adsorption and 2D diffusion rather than by 3D diffusion and is consistent with the findings of previous work using lipid bilayers[6] and bare mica[16] as substrates. The time between the assembly trigger and sample vitrification step was at most 5 hours for all S-layers with analyzed boundaries.

Cryo-TEM imaging

Samples were flash frozen in liquid ethane and stored in liquid nitrogen until they were loaded into the holder for data acquisition. Images were acquired on a JEOL–3100 FFC electron microscope equipped with a FEG electron source operating at 300 kV, an Omega energy filter, a cryo-transfer stage, and a Gatan 795 4Kx4K CCD camera mounted at the exit of an Electron Decelerator.[17] The stage was cooled using liquid nitrogen to 80 K. The decelerator was operated at 250 kV, resulting in images formed by a 50 kV electron beam at the CCD. The images used for analysis were acquired using a nominal magnification of 15 xK resulting in a pixel size at the specimen of 0.35 nm (post-magnification factor of 2.8 x), a dose of 3350 e^-/nm^2 (33.5 $e^-/Å^2$), and a defocus of ~ -6,000 nm.
S-layer growth on this support required approximately ten times lower protein concentration than in bulk solution (Experimental Section), consistent with a surface reaction on a wet support.[6, 15]

We found that a small exposure to ion plasma renders "pristine" graphene hydrophilic, without introducing damage relevant at the scale of this biological problem. Graphene's mechanical properties can potentially help prevent local drift in higher resolution studies than reported here. S-layer growth on this support required approximately ten times lower protein concentration than in bulk solution (Experimental Section), consistent with a surface reaction on a wet support.[6, 15]
Having the S-layer sheets polymerize on a flat transparent surface (graphene) enabled the very clear visualization of the boundaries under the chosen illumination parameters. We used techniques and algorithms designed for the detection of 2D heterogeneities in TEM projections of biological macromolecules to investigate boundaries in actively self-assembling (i.e., growing) S-layers.

Data Processing

The program Imod (http://bio3d.colorado.edu/) was used to select S-layer subunits by placing a model sphere at each subunit, as shown in Figure 1. Image filtering and rotations were done using the suite of programs Priism (http://msg.ucsf.edu/IVE/). Approximately 800 boxes from boundaries and 1,500 boxes from central regions were cropped (using a home-made script).

For boundary subunit-resolved analysis of conformers, the images were cropped using boxes 96 pixels (~ 33.6 nm) on a side initially centered on tetrameric subunits closest to

the open boundaries of the S-layer sheets. For cropping the images the boxes were moved towards the edge by a distance half the periodic repeat of the lattice (Figure S1). For control and reference analysis, central regions of the S-layers further from the boundaries were also boxed. The program Imod (http://bio3d.colorado.edu/) was used to select S-layer subunits by placing a model sphere at each subunit, as shown in Figure 1. Image processing such as X-ray erasing, cropping, and binning were computed with the suit of programs within the package Imod. Image filtering and rotations were done using the suite of programs Priism (http://msg.ucsf.edu/IVE/). Approximately 800 boxes from boundaries and 1,500 boxes from central regions were cropped (using a home-made script).

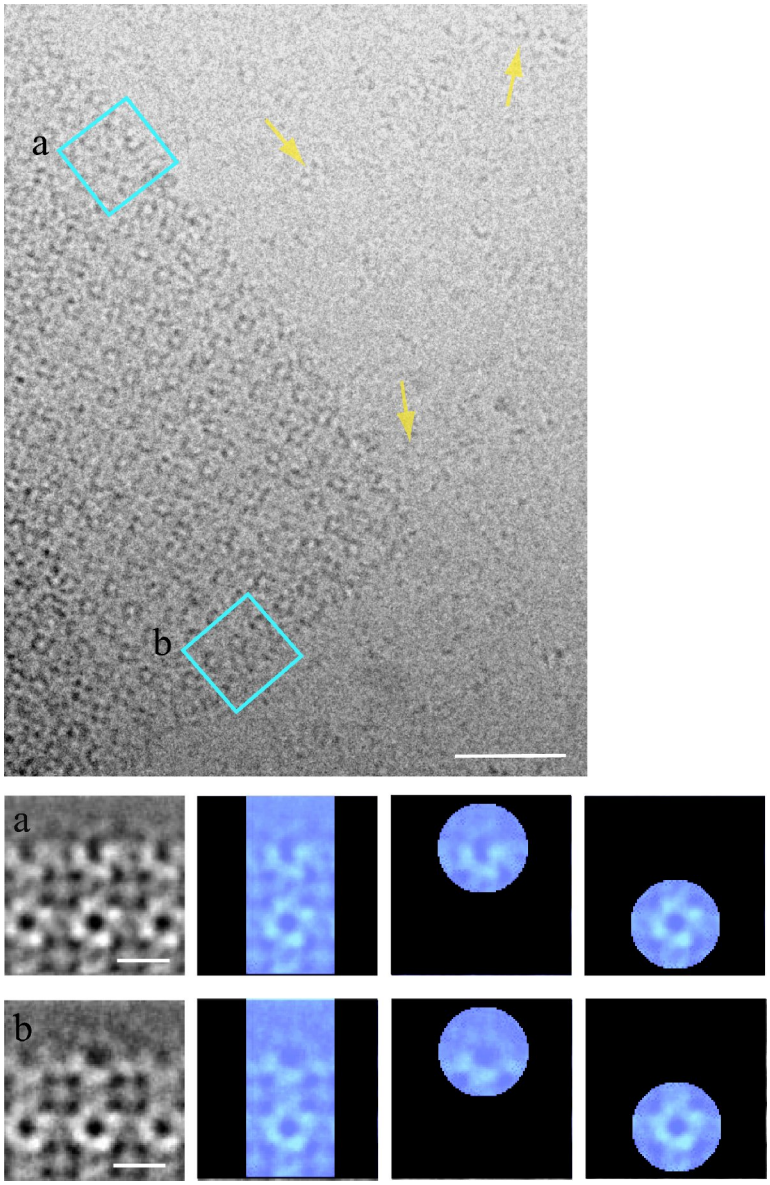
Three different methodological approaches were used and thoroughly compared: Kernel Probability Density Estimator Self-Organizing Map (KerDenSom), Principal Component analysis (PCA), and iterative 2D Maximum Likelihood (ML) classification and analysis. Binary masks were applied in order to process only images of the assembling subunits, or alternatively the assembled subunits inside the boundary (Experimental Section and Figure S1). We also observed protein in extended conformations and small clusters on the graphene surface (Figures 1 & S1). However, low-dose cryo-TEM averaging techniques require the systematic analysis of statistically relevant numbers of projections, and these are ideally provided by the growing boundaries of large S-layer sheets.

Alignment and Classification

Alignment and classification of the boxed images were computed with the various utilities within the "X-Window-based Microscopy Image Processing Package", or Xmipp package, (http://xmipp.cnb.csic.es/twiki/bin/view/Xmipp/WebHome). Several clustering or classification strategies using different algorithms were used in order to validate the results across conceptually different methodologies.
In order to isolate and investigate one assembling subunit at a time, first a rectangular mask 46 pixels wide and 96 pixel long was applied to each box for alignment (the points inside the binary mask are selected). This box includes only a complete homotetrameric subunit and the incomplete subunit at the boundary. The center of the rectangular box was displaced half a lattice period from the full homotetrameric subunit towards the open boundary, and no boxes contained two complete subunits. With this operational design we do not select complete homotetrameric subunits as outermost boundary components, only intermediate steps. The complete homotetrameric subunit is useful for the alignment of boundary boxes as well as for quality control of the results. Wrong alignments were discarded by direct inspection.
In the next step a circular mask 46 pixels in diameter centered at the assembling boundary unit was applied for classification based on Kernel Probability Density Estimator Self-Organizing Map (KerDenSom), Principal Component analysis (PCA), and iterative 2D Maximum Likelihood (ML) classification and analysis (http://xmipp.cnb.csic.es/twiki/bin/view/Xmipp/ListOfReferences & http://xmipp.cnb.csic.es/twiki/bin/view/Xmipp/ListOfPrograms).
KerDenSom clustering results in a minimum of four distinct classes. Runs imposing a larger number of clusters resulted in redundant classes. PCA analysis results were

constrained to the eigenvectors explaining 60% of the variance. As with KerDenSom clustering, four KCMeans clusters of the images projected onto these eingenvectors result in the minimum number of non-redundant classes. When the requested number of "expected representatives" is greater than 4, we obtain redundant classes. Iterative reference-free alignment and classification were computed using the Xmipp algorithm, which relies on a maximum-likelihood (ML) target function. Any number greater than 4 of "expected representatives" or refined references requested results in redundant classes. Four target structures appear to account for the data. Next, the circular mask was centered at the full homotetrameric subunits adjacent to the boundary. Both KerDenSom classification and iterative ML alignment and classification result in three distinct classes. ML was also applied to the full 96 pixels by side boxes without any mask, constraining the process to several numbers of final reference structures. Essentially the same results are obtained in this case. When a smaller and quadrilateral mask is applied to the assembling subunits, the same results are obtained.

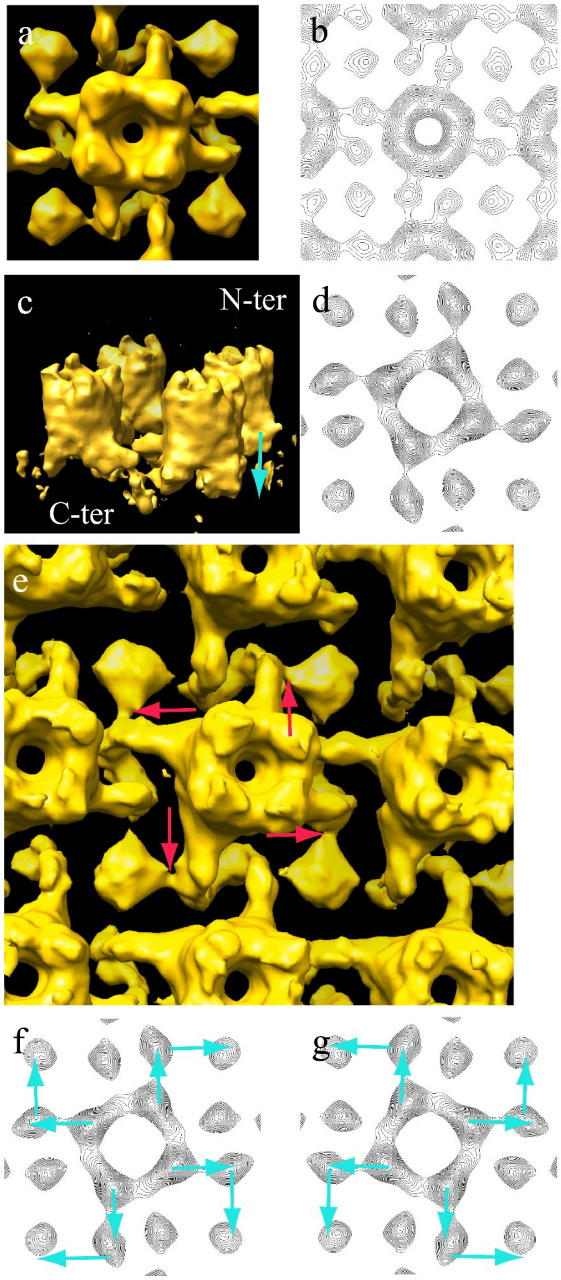
Figure S1



Cryo-TEM image (2D projection) of an S-layer sheet laying on graphene and image cropping for 2D structural analysis. Yellow arrows point to protein adsorbed on the graphene surface, mostly oligomers in extended or partially folded conformations and probably monomers at or below detection level. The blue boxes (**a**) and (**b**) represent the image cropped sections, 96 pixels (36 nm) by side and are directly shown (no averaging) in magnified views below the main panel. The contrast is inverted, a necessary step prior to the computation of class averages (with the collection of all equivalent boxes). The three types of masks shown to the right of (**a**) and (**b**) were used, with the information let through shown in false blue. The rectangular mask, left row, was used in all cases (KerDenSOM, PCA, and ML) for alignment. The results shown in Figures 1 & S3a were obtained using the circular mask letting only the boundary information, middle row.

Essentially the same results are obtained using the circular mask letting only information from the tetrameric subunit adjacent to the boundary, right row, as shown for KerDenSOM in Figure S3c. The result of computing ML alignment and classification on the raw square image sections is shown in Figure S3b; essentially the same results are obtained in terms of discrete set of conformers and coupling between boundary and adjacent tetrameric units conformations. Scale bars: (**a**) 50 nm; insets 10 nm.

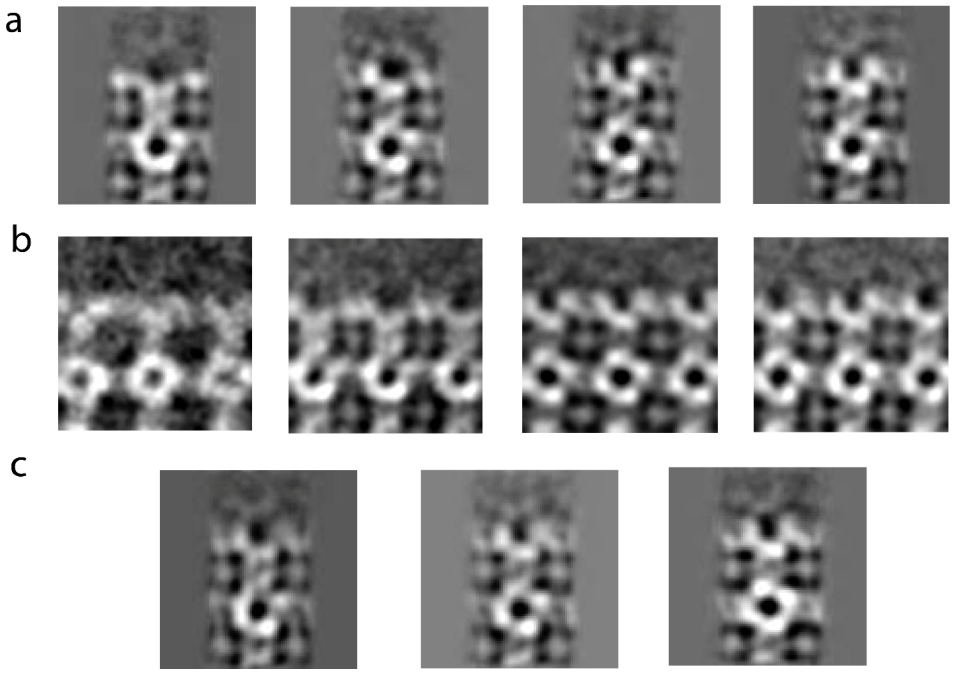
Figure S2



Handedness and connectivity of lattice subunits. (**a**) Three-dimensional view (isosurface rendering) of the rSbpA homotetrameric subunit viewed from the N-terminus. (**b**)

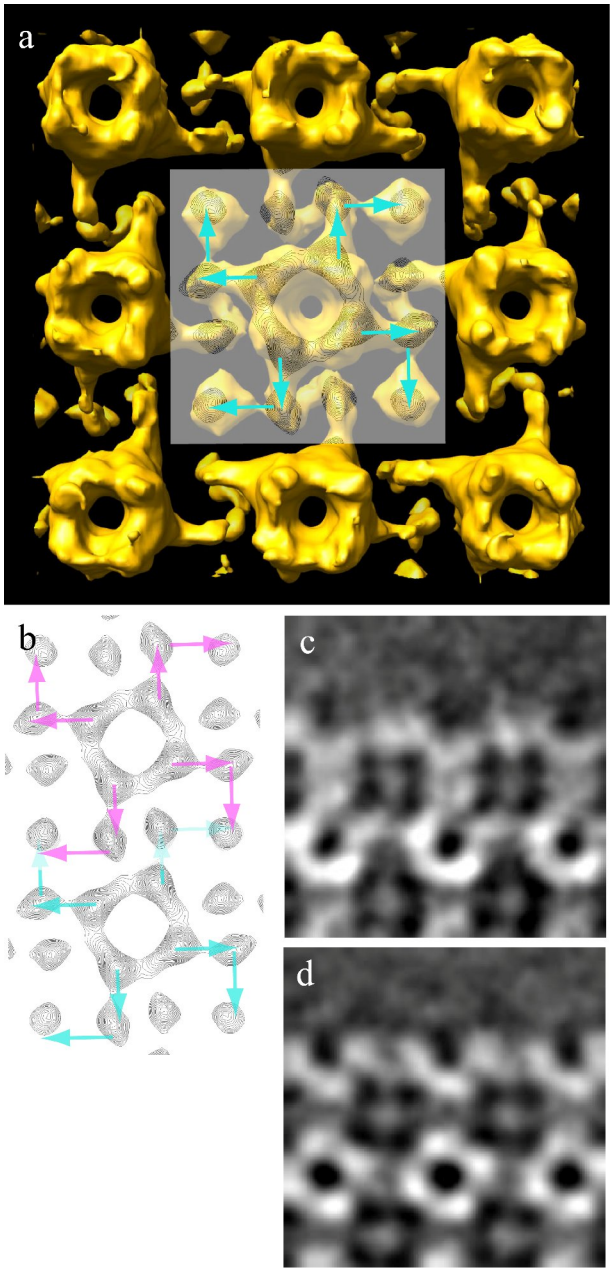
Projection map as a contour plot. (**c**) Side view with assigned N-terminus and C-terminus; blue arrow spans the bottom 30 % of the homotetrameric subunit projected as a contour plot in (**d**). (**e**) N-terminus view of the center region of the lattice averaged patch; red arrows point the connectivity of each subunit making the homotetrameric repeating unit. These 3D views enable us to assign the connectivity of the domains in the projection maps and in 2D cryo-TEM projections. (**f**) The projection map in (**c**) with assigned connectivity and (**g**) the mirror image projecting a lattice viewed from the C-terminus. The handedness of the 2D projections together with an analysis of the 3D reconstructions unequivocally determine which side is in contact with the graphene surface.

Figure S3



Additional classification of conformers at the boundary and adjacent to the boundary. (**a**) PCA analysis of the boundaries (adjacent tetramer was masked). KCMeans clusters of the images projected onto the PCA-based eingenvectors explaining 60% of the variance. The main, non-redundant classes are very similar to those obtained by KerDenSOM and ML analysis. (**b**) Structural classes spanned by the boundaries and subunits adjacent to the boundary from ML iterative classification and averaging (no mask). (**c**) KerDenSOM structural classes from the row adjacent to the boundary (the boundary was masked). Classification was done masking out the boundaries. The homotetrameric subunit adjacent to the boundary structurally changes in a concerted step-wise manner with the evolution of the boundary. The pore at the center of the "barrel-like" subunits shrink in size from approximately 3.85 nm to 3.15 nm in diameter as the assembly becomes more compact and defined. Scale: boxes are ~ 36 nm wide and scale bars 10 nm.

Figure S4



The mature S-layer lattice is connected through small domains ("m" domains in Norville et al.[S3]). Boundary growth coupled to maturation of the lattice requires the assembly of the inter-subunit connecting domains. The projection maps shown in Fig. S2 overlaid on a "top" view of the 3D averaged reconstruction of the rSbpA patch are used to highlight the connectivity of the homotetramers through the small domain at the center of the inter-subunit space. The arrows indicate the connected mass density or linkeage from each monomer of the repeating unit (**a**, **b**). **c**) The assembling extended monomers must couple the formation of these intersubunit connections with their conformational arrangement into the repeating unit. **d**) When compact, folded trimers are assembled at the boundary,

the connections of the mature lattice are already established. The last monomer coming into place must fit into the established scaffold to form the final tetrameric unit. As monomers in extended conformations are attaching to the open boundary and establishing intersubunits connections the last monomer incorporated to form the tetrameric unit is locked in place. This is the more entropically costly step and therefore the rate-limiting step.

Supporting References

[S1] S. Chung, S-H. Shin, C. R. Bertozzi, J. De Yoreo, *Proc. Natl. Acad. Sci. USA*, **2010**, 107 (38), 16536-16451.

[S2] L. R. Comolli, S-H. Shin, C.E. Siegerist, A. Hexemer, K. T. Nam, C. Wang, J.J. De Yoreo, C. Bertozzi, *ACS Nano,* **2013,** Manuscript ID: nn-2013-00263j.
We optimized a method for the self assembly of a truncated sequence of *Lysinibacillus sphaericus* (BS) S layer, SbpA. The truncation of the C-terminus 170 residues results in a sequence, rSbpA, capable of self assembling in a *p4* lattice in bulk. Cryo-ET 3D image analysis of monolayers was used to compute average structures of a patch made of 9 subunits, at a resolution of approximately 25 Å (Shin and Comolli, work in progress). The lattice with a period of approximately 13.3 nm is spanned by homotetrameric subunits of approximately 10 nm in diameter, 13 nm in height, and with a central pore of 2 nm. Lower resolution structures of S-layer on intact BS cells were obtained in order to assign N-terminus and C-terminus in our structure.

[S3] J. E. Norville, D. F. Kelly, T. F. Knight, A. M. Belcher, T. Walz, *J. Struct. Biol.* **2007**, 160, 313-323.