DOI: 10.1002/ange.201300543

Self-Assembly

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*

S-layers are two-dimensional protein or glycoprotein lattices that cover the surfaces of many bacteria and archaea. Because they constitute the first interface of interaction between microorganisms and their environment, hosts, and predators, they are of great biological interest. Moreover, owing to their nanoscale, periodic, porous structure and relative ease of manipulation, they have the potential to be useful for both nano-biotechnological and materials applications. However, details of the assembly process are not yet known for any S-layer and high resolution structural information is very limited. Herein, we report a two-dimensional (2D) structural analysis of the expanding boundary of an isolated *Lysinibacillus sphaericus* S-layer (SbpA) growing on a graphene support. The results reveal previously unknown steps in the conformational transformation that drives the well-documented non-classical pathway of S-layer assembly and show how the fully-folded oligomeric repeating unit is entropically locked into the ordered array. In addition, our results provide the first demonstration that the unique physical properties of graphene offer superior image quality for cryogenic transmission electron microscopy (cryo-TEM) of biological macromolecules.

[*] L. R. Comolli
Life Sciences Division, Lawrence Berkeley National Laboratory
Berkeley, CA 94720 (USA)
E-mail: lrcomolli@lbl.gov
C. E. Siegerist
http://www.cristinasiegerist.com/ComputingVisualization/index.html
Berkeley, CA 94708 (USA)
S.-H. Shin, J. De Yoreo
Molecular Foundry, Lawrence Berkeley National Laboratory
Berkeley, CA 94720 (USA)
W. Regan, Prof. A. Zettl
Materials Sciences Division, Lawrence Berkeley National Laboratory
Berkeley, CA 94720 (USA)
and
Department of Physics, University of California, Berkeley
Berkeley, CA 94720 (USA)
Prof. C. Bertozzi
Department of Chemistry, University of California, Berkeley
Berkeley, CA 94720 (USA)

[**] This work was performed at Lawrence Berkeley National Laboratory, with support from the Office of Science, Office of Basic Energy Sciences, Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Supporting information for this article (including experimental details) is available on the WWW under http://dx.doi.org/10.1002/anie.201300543.

S-layers assemble from a single protein or glycoprotein sequence to form a 2D lattice that covers the entire cell surface of microorganisms, including the cell poles and division sites. They are non-covalently attached to peptidoglycans and related polymers of gram-positive cell walls, linked to the outer membrane of gram-negative bacterial cell walls, and integrated into the cytoplasmic membrane through trans-membrane domains in gram-negative archaea.[1,2] The primary sequence of the single protein or glycoprotein species contains all the information needed for assembly. Although they are often called "crystalline" cell surface layers, they are better described as "quasi-crystalline" or "paracrystalline".[3]

S-layers also self-assemble in vitro in the presence of Ca^{2+} ions, either on support films or in bulk solution, into ordered arrays with long-range order, substantially larger than a single cell.[4–8] Previous studies found that self-assembly of SbpA (1268 residues from *Lysinibacillus sphaericus*) on lipid bilayers follows a multi-step pathway.[6] It starts with the aggregation of monomers that adsorb onto the lipids in an extended conformation to form amorphous or liquid-like clusters. These clusters subsequently crystallize into the characteristic lattice of homotetrameric units, which grows by addition of new tetramers to the lattice edge sites. The rate of tetramer addition increases linearly with protein concentration, implying that monomers are added one at a time. However, the pathway through which the individual monomeric units become integrated with the correct conformation into the homotetrameric units—arguably the single most important step in S-layer assembly—remains unknown.

To gain insight into S-layer assembly at the level of the tetramer subunits in the intact solution state, we obtained cryo-TEM images of single sheets, plunge-frozen while growing on graphene. Because the active self-assembling S-layers are instantly frozen, all the conformational states present at the expanding boundary on the graphene flat support are captured. Image alignment and averaging provide a view of the steps leading to subunit recruitment and maturation in S-layer self-assembly.

For this study, we chose the surface layer protein SbpA (1268 residues), from the gram-positive bacterium *Lysinibacillus sphaericus*, which naturally forms a 2D quasi-crystalline cell envelope.[7] We carried out the reconstitution of SbpA on single graphene sheets supported by TEM grids designed for cryo-TEM. The great potential of graphene for use as a support for biological cryo-TEM samples has recently been discussed.[9] While mechanically strong and elastic, the 0.246 nm lattice constant and one-atom thickness of a single graphene layer, approximately 0.34 nm,[10] make them trans-

parent for this application (see the Supporting Information for more information).

In this study, having a flat and isotropic support was critical to ensuring that we were studying the growth of the S-layer boundary and not variable interactions with a heterogeneous support. We thus decided to use pristine graphene monolayers instead of oxidized graphene, thus avoiding the typical heterogeneities reported previously.[9a] The graphene support also enabled our image analysis, because the growing S-layer on graphene was always parallel to the image plane. Thus, the 2D analysis of the projected S-layer boundary could be carried out with systematic self-consistency.

S-layer growth was stopped by flash-freezing at early points during the polymerization reactions, with excess free protein still in the solution phase. The flash-frozen samples contained an ensemble of all states, enabling us to isolate the main steps leading to formation of the compact and fully assembled homotetrameric subunit forming the lattice. The absence of any empty—i.e., protein-free—regions near the open boundaries of the S-layers verified that there was available free protein near those boundaries at the instant of flash-freezing (Figure 1; Supporting Information, Figure S1), implying that we indeed obtained views of a frozen dynamic process that had not gone to completion. Selection of the well-defined homotetrameric repeating units immediately adjacent to open boundaries was used as a means to capture the incomplete, unknown structures at the growing edge within a box for image processing (Figure S1).

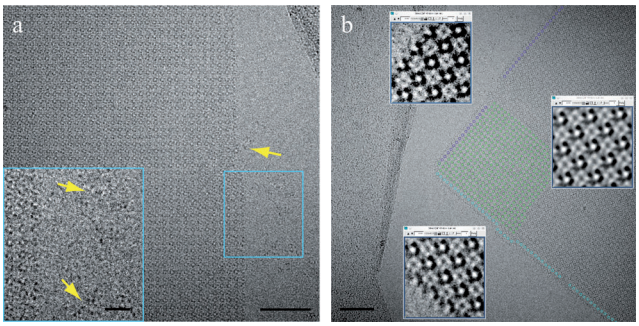

Figure 1. Cryo-TEM image (2D projection) of an S-layer sheet lying on graphene and image cropping for 2D structural analysis. a) An S-layer sheet covers the underlying flat support, with loose un-polymerized protein to the right of the boundary. Inset: magnified view of the blue square; yellow arrows = unfolded wtSbpA protein not yet incorporated into the lattice at the time of flash-freezing (see also Figure S1). The S-layer and the loose un-polymerized protein are on the same plane, the surface of a graphene sheet embedded within a thin film of vitreous ice. A small boundary of the lacey carbon film supporting the graphene sheet is visible in the top, right corner. The highly uniform square lattice spanned by the S-layer, and the flatness or in-plane distribution of the mass density is apparent. b) Method of boundary-box selection and preliminary averages with enhanced signal-to-noise. Full homotetrameric subunits adjacent to the open, growing boundaries (marked red and blue) were selected in a first step for structural analysis of the unknown boundary. Preliminary averages along the top (red, left top inset) and bottom (blue, left bottom inset) boundaries show a slightly different mass distribution for open, non-tetrameric oligomers (incomplete units with a large dominance of trimers). Subunits from the interior of the sheets (marked green) were selected for control purposes; average shown in the right inset. A smaller section was used for iterative classification and averaging, details in the Supporting Information. Scale bars: a) 100 nm, inset: 25 nm; b) 100 nm, insets: 50 nm per side.

To understand which side of SbpA is facing the graphene support in our 2D projections, the C-terminus or N-terminus, we used a 3D sub-volumetric average structure of the nine-subunit lattice patch of bulk rSbpA (fully assembled regions) reconstructed by cryogenic electron tomography (cryo-ET).[11] Both SbpA and rSbpA form square lattices, but the connectors between units break the symmetry and give a handedness to the lattice. In our SbpA 2D projections, the handedness of the lattice is the same in 85% of the data (17 out of 20 images; Figure 2a). Only these were used for analysis and the remaining 15% (Figure 2b) were discarded. Figure 2c shows a projection of a reconstructed rSbpA 3D patch looking down from the C-terminus towards the N-terminus of the protein, with oblique connectors between the units (as in Figure 2a). Views as in Figure 2b can only be obtained through an image inversion (a mirror).

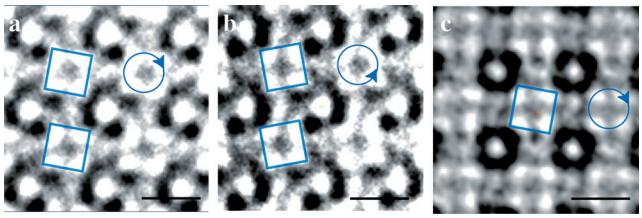

Figure 2. Assignment of the orientation and S-layer face on the graphene sheet: the handedness of the lattice allows identification of top and bottom views. a) and b) are mirror projections obtained from different cryo-grid areas (different sheets). Blue boxes highlight the oblique connections between units spanning a square lattice, enabling the distinction between top and bottom views. Blue circles with arrows show the path direction in (a) is a mirror image of (b). c) Projection of the reconstructed 3D structure of a small sub-volume of the lattice, viewed from the C-terminus towards the N-terminus (see also Supporting Information). Although (c) is the lattice formed by the truncated rSbpA sequence and the connectors are thinner, the architecture is the same as for the untruncated sequence and projection (c) helps to interpret (a) and (b). Scale bars: 10 nm.

This means that the S-layer is generally oriented with the N-terminus of the subunits making contact with the graphene surface and the C-terminus exposed to the bulk solution (Figure 2; Figure S2). In a bacterial cell, the N-terminus is associated with the peptidoglycan through strong ionic interactions. The ability to distinguish "up" and "down" in 2D projections of the S-layer allows for a self-consistent structural analysis of the lattice and the growing boundary. In the first few hours after polymerization, the lattice was not yet mature. We often found a slight departure from a full fourfold symmetry, with the directionality biased towards the largest open boundary, as seen in Figure 1 (insets) and Figure 2a,b.

We were able to clearly identify structures that correspond to discrete steps in folding events and conformational transformations to oligomeric forms, which are inherent parts of the assembly process. These discrete classes are nearly identical across the three mathematically independent classification methods used for image enhancement of our low-

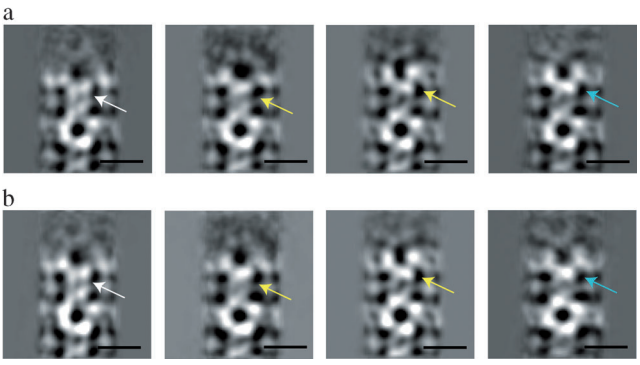

Figure 3. Main conformers at the growing boundary according to different methods. KerDenSOM maps (a), and maximum likelihood-based iterative image classification and averaging (b) result in four distinct structural classes. These methods are mathematically independent. The homotetrameric subunit adjacent to the boundary structurally changes in a concerted step-wise manner with the evolution of the boundary. The discrete steps identified by 2D classification methods are: incorporation of monomers in extended conformations, while the repeating homotetramer is not fully assembled (white arrows); density, probably representing three monomers, is arranged at the boundary while the adjacent homotetramer forms a compact near symmetric oligomer (yellow arrows); a trimer is formed at the boundary and a fully symmetric homotetramer at the adjacent position (blue arrows). The number of individual members per class is 168–210. Images were cropped using square boxes, 96 pixels on a side, as shown. A rectangular mask 46 pixels wide and 96 pixel long was applied to each box for alignment, gray-scale areas on display. Scale bars: 10 nm.

dose data (Experimental Section; Figure 3 a,b). As previously described,[6] the assembly pathway started with monomers of extended conformation. Lattice growth was not driven by the incorporation of new fully assembled and folded donut-shaped tetramers. Instead, tetramers matured into the final fourfold oligomeric conformation (Figure 1 and Figure 2), as monomers in the extended conformation were concomitantly added to the boundary (Figure 3). These extended conformations at the open boundary progressively rearranged into a distinct trimer with the fourfold symmetry of the mature subunit, as the adjacent homotetrameric subunit folded into the mature donut-shaped repeating unit spanning the lattice. These results are also robust: masking the tetramer adjacent to the boundary for classification of the boundary mass density (Figure 3; Figures S1,S3), masking the open boundary and classifying the adjacent tetramer (Figure S3), or classifying unmasked images containing both the boundary and adjacent tetramer resulted in the same classification and relationship between conformations (Figure S3). The discrete steps in the maturation of the homotetrameric subunit were coupled to the same discrete steps in the organization of the boundary across methods and image processing strategies.

In summary, the lattice does not grow through the incorporation of fully assembled and folded homotetrameric subunits. Moreover, tetrameric subunits adjacent to a growing boundary are not fourfold symmetric during the adsorption of new, extended monomers, until a series of discrete rearrangements take place. Thus, the repeating subunit only forms within the lattice, and loses the freedom to leave it.

According to our data, the most common growth unit structure at the edge sites is a trimer open to the edge. Thus the rate-limiting step is the closure of the trimer by addition of the last monomer to form the tetramer at the growing boundary. This step carries the highest entropic cost because the last monomer must dock into a scaffold in the step, binding to two existing sub-units and thus completing and locking one unit of the lattice. This is equivalent to protein folding in strong confinement, where entropy plays the dominant role (over enthalpy) decreasing the ability of the protein to experience unfolded states.[12]

These observations are consistent with the findings of Chung et al.,[6] who observed S-layer formation using in situ AFM. First, they reported that the initial clusters of SbpA did not grow by adsorption of folded tetramers. Instead they found that tetramers are formed from adsorbed monomers and that lattice units tended to complete entire rows successively before beginning the next row. Second, they found that the kinetics of growth scaled linearly with the S-layer concentration. If the growth required formation of a complete tetrameric unit, then the growth rate would be dependent on the concentration to the fourth power. Linear kinetics imply that attachment of a single monomer is rate limiting. Our data show that this is the case and, moreover, that it is attachment of the last monomer that determines the growth rate. Third, once formed, homotetrameric units were never observed to leave the lattice; growth was irreversible. Our data suggests the reason for this irreversibility is that tetramer addition does not drive boundary growth, rather tetramers form within the lattice through a monomer addition process and once the last monomer has been locked into place, it has a low probability of leaving. The last monomer coming into place establishes the inter-subunit interactions within an already ordered lattice scaffold, and provides interactions for free monomers in extended conformations. This final conformational arrangement in "confinement" with a loss of degrees of freedom constitutes the most entropically costly—and therefore the rate-limiting—step. Finally, Chung et al. concluded that the emergence of the ordered lattice catalyzes the growth process by lowering the barrier for tetramer formation. We found that this is indeed the case, because monomers in extended conformations must interact with the lattice to form new tetramers.

The nucleation of partially disordered oligomers and the subsequent transformations drive the growth of amyloid fibrils, trapping misfolded non-functional proteins.[13] In drastically different polymerization processes, cytoskeletal proteins self-assemble from fully functional and folded oligomeric subunits. In tubulin for example, fully folded catalytic heterodimers are added reversibly to the growing ordered array.[14] Furthermore, in microtubules assembling through an intermediate polymeric sheet/ribbon, these are made of identical repeating subunits and a collective transition closes the sheets into cylinders or tubes. Polymerization requires external energy, GTP, for the presence of functional units.

The SbpA S-layer self-assembly process combines elements of these two classes of processes. It grows through addition of extended subunits at the boundaries of the

ordered array, and involves a transition of the adsorbed monomers to the repeating unit spanning the stable ordered array. As a result the repeating oligomeric units are added non-reversibly. And while the process requires oligomeric unfolded intermediates, these make a transition to the repeating unit, spanning the stable ordered array without spending external energy (ATP or GTP). Thus, SbpA offers a window for the study of this hitherto less well-investigated class of protein assembly.

Understanding the rules for SbpA assembly offers the potential for controlling nanoscale 3D patterns of materials. Moreover, the same techniques applied here to S-layers can be used to study other complex protein self-assembly problems, and can be extended to three dimensions. The use of graphene supports allows for the solution of a 2D problem, while optimizing high quality cryo-TEM data and opens the way for higher resolution, time-resolved structural analysis of the critical domains and interactions in wild-type and engineered proteins that undergo self-assembly. This approach can also be used to investigate assembly of modified polypeptide chains in which domains critical for self-assembly have been engineered (truncated or modified) and surface interactions introduced, based on quantitative structural and dynamic information.

Received: January 21, 2013
Revised: March 6, 2013
Published online: April 8, 2013

Keywords: conformational transitions · electron microscopy · nanostructures · oligomeric intermediates · self-assembly

[1] U. B. Sleyter, B. Schuster, E. M. Egelseer, D. Pum, C. M. Horejs, R. Tschelienssnig, N. Ilk, *PMBTS* **2011**, *103*, 277 – 353.
[2] T. Pavkov-Keller, S. Howorka, W. Keller, *PMBTS* **2011**, *103*, 73 – 130.
[3] F. Amat, L. R. Comolli, J. Smit, F. Moussavi, K. H. Downing, M. Horowitz, *J. Bacteriol.* **2010**, *192*, 4847 – 4858.
[4] J. F. Nomellini, S. Kupcu, U. B. Sleytr, J. Smit, *J. Bacteriol.* **1997**, *179*, 6349 – 6354.
[5] J. E. Norville, D. F. Kelly, T. F. Knight, A. M. Belcher, T. Walz, *J. Struct. Biol.* **2007**, *160*, 313 – 323.
[6] S. Chung, S.-H. Shin, C. R. Bertozzi, J. De Yoreo, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 16536 – 16541
[7] U. B. Sleytr, P. Messner, D. Pum, M. Sara, *Angew. Chem.* **1999**, *111*, 1098 – 1120; *Angew. Chem. Int. Ed.* **1999**, *38*, 1034 – 1054.
[8] a) Y. Sierra-Sastre, S. Choi, S. T. Picraux, C. A. Batt, *J. Am. Chem. Soc.* **2008**, *130*, 10488 – 10489; b) W. Shenton, D. Pum, U. B. Sleytr, S. Mann, *Nature* **1997**, *389*, 585 – 587.
[9] a) R. S. Pantelic, J. C. Meyer, U. Kaiser, W. Baumeister, J. M. Plitzko, *J. Struct. Biol.* **2010**, *170*, 152 – 158; b) R. S. Pantelic, J. W. Suk, C. W. Magnuson, J. C. Meyer, P. Wachsmuth, U. Kaiser, R. S. Ruoff, H. Stahlberg, *J. Struct. Biol.* **2011**, *174*, 234 – 238; c) R. S. Pantelic, J. W. Suk, Y. Hao, R. S. Ruoff, H. Stahlberg, *Nano Lett.* **2011**, *11*, 4319 – 4323.
[10] W. Regan, N. Alem, B. Aleman, B. Geng, C. Girit, L. Maserati, F. Wang, M. Crommie, A. Zettl, *Appl. Phys. Lett.* **2010**, *96*, 113102.
[11] L. R. Comolli, S.-H. Shin, C. E. Siegerist, A. Hexemer, K. T. Nam, C. Wang, J. J. De Yoreo, C. Bertozzi, unpublished results. See Supporting Information for more details.
[12] T. A. Knotts, N. Rathore, J. J. de Pablo, *Biophys. J.* **2008**, *94*, 4473 – 4483.
[13] a) W.-F. Xue, S. W. Homans, S. E. Radford, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 8926 – 8931; b) Y. Liang, D. G. Lynn, K. M. Berland, *J. Am. Chem. Soc.* **2010**, *132*, 6306 – 6308.
[14] a) E. Nogales, H.-W. Wang, *Curr. Opin. Cell Biol.* **2006**, *18*, 179 – 184; b) H.-W. Wang, E. Nogales, *Nature* **2005**, *435*, 911 – 915; c) E. Nogales, H.-W. Wang, *Curr. Opin. Struct. Biol.* **2006**, *16*, 221 – 229.
[15] L. R. Comolli, R. Duarte, D. Baum, B. Luef, K. H. D. Downing, D. Larson, *R. Csencsits, J. Banfield, Microsc. Res. Tech.* **2012**, *75*, 829 – 836.
[16] S.-H. Shin, S. Chung, B. Sanii, L. R. Comolli, C. R. Bertozzi, J. J. De Yoreo, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 12968 – 12973.
[17] K. H. Downing, P. E. Mooney, *Rev. Sci. Instrum.* **2008**, *79*, 043702.

