Atomistic Structure of Monomolecular Surface Layer Self-Assemblies: Toward Functionalized Nanostructures

Christine Horejs,[†] Harald Gollner,[†] Dietmar Pum,[†] Uwe B. Sleytr,[†] Herwig Peterlik,[‡] Alois Jungbauer,[§] and Rupert Tscheliessnig[§,*]

[†]Department for Nanobiotechnology, University of Natural Resources and Life Sciences, Vienna, Austria, [‡]Faculty of Physics, University of Vienna, Vienna, Austria, and [§]Austrian Centre of Industrial Biotechnology c/o Institute for Biotechnology, University of Natural Resources and Life Sciences, Vienna, Austria

ABSTRACT The concept of self-assembly is one of the most promising strategies for the creation of defined nanostructures and therefore became an essential part of nanotechnology for the controlled bottom-up design of nanoscale structures. Surface layers (S-layers), which represent the cell envelope of a great variety of prokaryotic cells, show outstanding self-assembly features *in vitro* and have been successfully used as the basic matrix for molecular construction kits. Here we present the three-dimensional structure of an S-layer lattice based on tetrameric unit cells, which will help to facilitate the directed binding of various molecules on the S-layer lattice, thereby creating functional nanoarrays for applications in nanobiotechnology. Our work demonstrates the successful combination of computer simulations, electron microscopy (TEM), and small-angle X-ray scattering (SAXS) as a tool for the investigation of the structure of self-assembling or aggregating proteins, which cannot be determined by X-ray crystallography. To the best of our knowledge, this is the first structural model at an amino acid level of an S-layer unit cell that exhibits p4 lattice symmetry.

KEYWORDS: self-assembly · SAXS · TEM · protein structure · computer simulations · nanoarrays

Surface layers (S-layers) represent an outstanding self-assembly system based on protein subunits, which form the crystalline outermost cell envelope of a great variety of prokaryotic cells.[1] These S-layer proteins self-assemble into monomolecular lattices with symmetries ranging from p1 to p6, exhibiting defined pores and a highly reproducible large-scale order.[2−4] They have been successfully used as the basic unit for molecular construction kits because of their remarkable potential to self-assemble not only in their natural environment but also in solution, on various solid substrates and on lipids.[3,5] A great number of S-layer-carrying organisms has been identified so far, whereas no atomistic structure of a single unmodified S-layer protein could be experimentally determined up to now because they are too large for NMR and they do not crystallize into isotropic three-dimensional crystals as required for X-ray crystallography. Recently, more information on structural details and self-assembly pathways could be gathered.[6−15] However, the defined binding of molecules and nanoparticles on S-layers requires detailed structural information on the underlying S-layer proteins at an amino acid level. Here we present a combination of molecular modeling, small-angle X-ray scattering (SAXS), and transmission electron microscopy (TEM) to reveal the three-dimensional structure of one unit cell of the S-layer protein SbpA from *Bacillus sphaericus* CCM2177.[16] SbpA consists of 1268 amino acids and self-assembles into a square lattice structure (p4). This S-layer protein has been extensively investigated in order to be used for nanobiotechnological applications at the nanoscale.[17−24] In order to model the structure at an amino acid level of one SbpA unit cell consisting of four monomers, we combined three different approaches. In a first step, we defined possible domains of the protein based on information on the self-assembly behavior of truncated and modified recombinant forms of the protein. We already successfully applied this approach to model the tertiary structure of another S-layer protein.[14] We defined seven individual domains of the protein, and on the basis of secondary structure predictions and structural homologies, we premodeled each domain implementing possible secondary structure elements. Consequently, we equilibrated each domain in a water sphere, joined them together, and minimized the whole structure while keeping every domain restraint using molecular dynamic simulations. In a second step, a three-dimensional density model has been calculated by performing tilting experiments with a transmission electron microscope. In a third step, small-angle

* Address correspondence to rupert.tscheliessnig@boku.ac.at.

Received for review December 23, 2010 and accepted February 14, 2011.

Published online March 04, 2011
10.1021/nn1035729

X-ray scattering (SAXS) experiments were performed together with a theory of a fractal mean force potential, which can be adequately used to describe the behavior of S-layers in solution, as we showed in our earlier work.[15] These studies resulted in a distribution of electron densities within an S-layer unit cell. We finally merged these data, the model of one monomer obtained by molecular dynamic simulations, and the density profile as well as the scattering distribution of one unit cell and modeled the three-dimensional structure of an SbpA unit cell at an amino acid level. The information of the location of individual amino acids on the inner and outer surface of the lattice as well as within the pores will facilitate the directed modification of this S-layer protein and the directed binding of molecules and nanoparticles on the lattice for the application of this self-assembly system in nanotechnology.

RESULTS AND DISCUSSION

The S-layer protein SbpA has aroused interest due to the fact that this protein self-assembles into lattices exhibiting a p4 lattice symmetry. The four monomers, which form one unit cell of the lattice, may perfectly serve as a matrix for the generation of functional nanoarrays.[25] The N-terminal part possesses three so-called S-layer homologous (SLH) domains, which is a common structural motive in this class of proteins that is mainly made up of α-helices.[26] The SLH domains are responsible for anchoring the proteins in the underlying cell wall. The successful self-assembly of SbpA demands the presence of bivalent cations,[26] which explains the structural homology of this protein to calcium binding proteins, which could be found by performing homology searches using various different bioinformatic tools. Interestingly, the lattice symmetry can be changed from p4 to p1 if 237 C-terminal amino acids are truncated. If additionally another 113 amino acid residues are removed, the proteins lose the ability to self-assemble.[27] Apparently a structural change of the monomers due to a shortening of the C-terminal domain leads to a loss of the intrinsic function. We also performed secondary structure predictions revealing S-layer proteins' common distribution of secondary structure elements, a mainly α-helical N-terminal region and mainly β-sheets in the central and C-terminal part. On the basis of all this information, we defined the following seven domains: three SLH domains (aa1−aa210, aa211−329, aa330−aa458), one central domain showing structural homology to calcium binding proteins (aa459−aa639), and three C-terminal domains showing homology to Ig-like proteins, which are mainly made up of β-sheets (aa640−aa888, aa889−aa1001, aa1002−aa1238). As a next step, we used the online algorithm PHYRE to model each domain based on fold recognition.[28] Domains one, two, and three (SLH domains) are mainly made up of

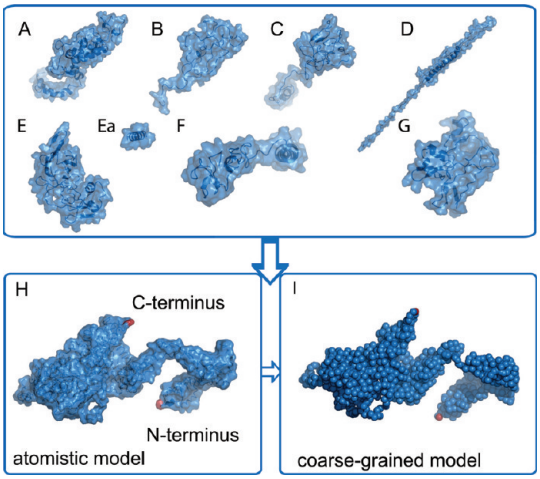

Figure 1. Results of the molecular dynamics simulations of the S-layer protein SbpA. The protein was first split into structurally meaningful domains, which were premodeled using fold recognition to obtain secondary structure elements. These domains were equilibrated in water. The results are given in A−G. (A−C) Three SLH (S-layer homologous) domains located at the N-terminal region of the protein, (D) central domain showing structural homology to Ca^{2+}-binding proteins, (E−G) domains showing structural homology to Ig-like domains. Without domains F and G, the protein loses its ability to self-assemble, and without domain G, the self-assembled products show p1 rather than p4 lattice symmetry.[27] (H) Intermediate structure of the whole protein after the single domains were joined and simulated in vacuum, where the secondary structures were kept rigid, ions were added, and external forces enabled. The termini are marked in red. (I) Coarse-grained protein structure based on the atomistic structure shown in H. Every amino acid is represented by a single bead. Consecutive beads are linked by a harmonic potential.

α-helices and coils as expected. Domains three, five, and seven exhibit parallel and antiparallel β-sheets, and domains four and six could not be modeled by fold recognition and show a rather elongated structure. Every premodeled domain was now equilibrated in a water sphere using molecular dynamics simulations. Figure 1A−G shows the results for every single domain. The three potential SLH domains (Figure 1A−C) are mainly made up of α-helices (especially the first SLH domain, Figure 1A) and random coils after equilibration in water. In clear contrast to our previous work,[14] we could not equilibrate all of these domains in water. Even after a considerable large amount of production runs, two domains lack a sensible secondary structure (Figure 1D,E). The central domain (Figure 1D), which was premodeled using homologies to Ca^{2+}-binding proteins, is supposed to consist of mainly β-sheet structures. β-Sheets are rather complicated secondary structure elements, which may fold due to interactions between distant amino acid residues in the primary structure, which is why they hardly form in the framework of short molecular dynamics simulations. In contrast, those domains, which show homologies to Ig-like domains, show β-sheets, α-helices, and random coil

structures (Figure 1E−G) as expected by the secondary structure predictions.

All domains were now linked together, and an inverse steered molecular dynamics simulation was performed in vacuum. As the protein is charged, Na^+ and Cl^- ions were added to compensate the net charge. Moreover, the anticipated secondary structures, the α-helices, β-sheets, and coils, were kept rigid. In addition, all domains were linked to one another by an external force. This external force is physically motivated by the force that the absent water molecules would impose on the protein. We realize these potentials by harmonic springs, which are necessary to initiate the refolding process. However, the resulting structure, as shown in Figure 1H, represents an intermediate conformation. The resulting model was consequently used to reconstruct the three-dimensional unit cell by rational design based on the density distribution as obtained by electron microscopy studies.

The transmission electron microscope offers a great possibility to determine the density distribution and thereby a three-dimensional reconstruction of an S-layer unit cell by performing tilting studies with a negatively stained self-assembly product due to the symmetry of the S-layer lattice. This method is based on the projection theorem, which states that the two-dimensional Fourier transform of a plane projection of a three-dimensional density distribution is identical to the corresponding central section of the three-dimensional transform normal to the direction of view. Therefore, the three-dimensional transform can be built up section by section and consequently reconstructed by inverse Fourier transform.[29] The phases of the diffraction pattern are directly accessible from an image. Figure 2 shows the results of a tilting series with an SbpA self-assembly product, where the single sections used for the density reconstruction are shown. On the basis of these sections, a density model of the unit cell can be rebuilt, as shown in Figure 2B. This three-dimensional model was consequently used to fit the structural model of one SbpA monomer into the unit cell. Therefore, four monomers are arranged in a way to cover the entire unit cell, where overlaps are avoided. As this arrangement is based on an intermediate structural model, the reconstructed unit cell has to be equilibrated, which demands a coarse-graining of the whole structure (Figure 1I). Each amino acid is represented by a single bead. We introduce three different potentials. First, all consecutive beads are linked by a harmonic potential. As all seven domains but the outer five amino acids of each are kept rigid, the harmonic potential is of minor relevance. The interactions of the amino acids are controlled by two types of pair−pair potentials: an attractive screened Coulomb potential and an associative Gaussian potential. While the first potential type is a consequence of electrostatic forces,

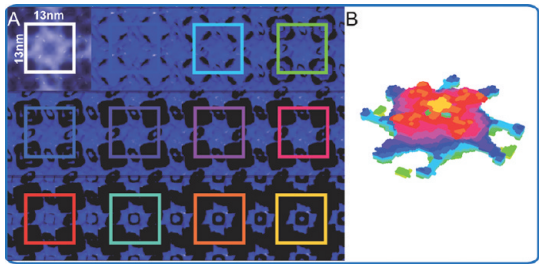

Figure 2. Reconstruction of the three-dimensional density distribution of one SbpA unit cell calculated by inverse Fourier transform of single sections of SbpA self-assembly products. A transmission electron microscope was used to obtain sections of the unit cell by performing tilting measurements. (A) Single sections of an SbpA unit cell. In the top left corner, an electron micrograph of a whole unit cell is shown. The density boundaries are systematically increased (starting in the top left corner). The sections provide the basis for the unit cell density distribution, which is shown by the colored squares. Every color reflects one single section of the unit cell in two dimensions. A superposition of all sections results in the three-dimensional unit cell as shown in B. (B) Three-dimensional density distribution as used for the reconstruction of the unit cell based on the coarse-grained model obtained from molecular dynamics simulations.

the second, in principle, should enable the formation of particular secondary structures. The coarse-grained model is consequently used to model a coarse-grained unit cell. The unit cell is periodic in plane with a period of 13.0 nm, while the periodicity out of plane is 6.0 nm. The amino acids are given appropriate masses, and the entire system is initially kept at 300 K and then cooled to 30 K in order to solidify the lattice.

This final structural model is used to reconstruct the entire small-angle X-ray scattering signal. The whole reconstruction procedure is based on a method presented in our earlier work, where we describe the S-layer system by a fractal mean potential.[15]

In Figure 3B, we give the background-corrected scattering contrasts for the monomeric and the self-assembled samples, where the scattering intensity is shown as a function of the scattering vector *Q*. Black open circles give the scattering contrast of a predominantly monomeric solution, while blue open circles represent the scattering contrast of self-assembled structures. Obviously the monomeric solution lacks any corrugation—all characteristic Bragg peaks are missing. We take this as a clear indication that, although secondary structure elements are formed, the tertiary structure of the protein remains non-native at this stage, suggesting that in their monomeric state the proteins adapt a different conformation than as part of a tetramer. This hypothesis that the proteins first condense into an amorphous cluster in an extended conformation before they restructure to form a crystal of folded tetramers has been recently proposed by Chung *et al.*[12] However, the slope of the scattering intensity is linear for small *Q* values and thus indicates a self-similar system. The fractal dimension of the

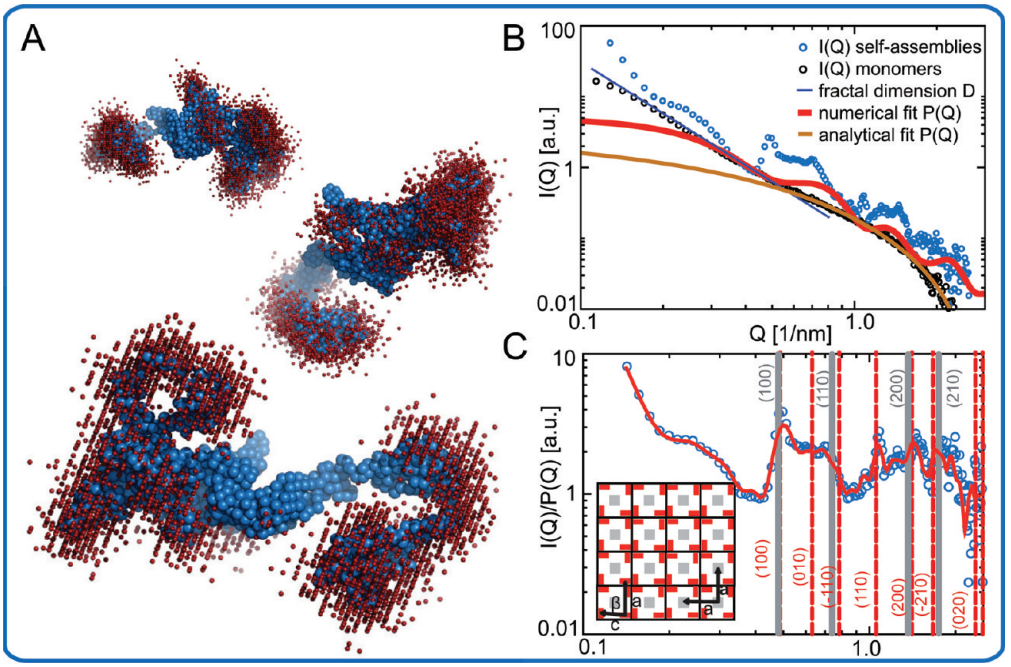

Figure 3. SAXS. (A) Distribution of scattering clusters (red beads) of one SbpA monomer (blue). The scattering sites represent high electron density contrast as determined by small-angle X-ray scattering (SAXS) and a Monte Carlo algorithm. (B) Intensities $I(Q)$ as a function of the scattering vector Q of the monomeric solution and the sample containing self-assemblies are given by black and blue open circles. The slope for the fractal dimension $D = 2.4$ is inserted as a dark blue line. The red line gives the numerical reconstruction using a Monte Carlo algorithm generalized to fractal dimensions and based on the coarse-grained monomer model. The yellow line gives the analytical fit of the monomer data. The reconstruction procedure is based on our previous work.[15] (C) Open circles give scattering intensities divided by the form factor $I(Q)/P(Q)$. The red line is the fit curve from eq 1, which includes the contribution to the scattering signal of the S-layer self-assemblies in solution. On the basis of this fit, the height $H = 5.6$ nm, the radius $R = 7.2$ nm, and the radius of the self-assemblies $R = 72$ nm can be calculated. Vertical lines indicate Bragg reflections, where strong peaks are indicated by gray lines and weak peaks by red lines. Strong peaks indicate an arrangement of the tetramers in a cubic lattice with $a = 13.0 \pm 0.1$ nm, where weak peaks arise due to the arrangement of the proteins in the tetramer. These weak reflections indicate a 2D monoclinic sublattice with unit cell dimensions of $a = 13.0 \pm 0.1$ nm, $c = 10.5 \pm 0.1$ nm, and $\gamma = 80 \pm 1$ (a detailed view of the Bragg reflections is given in Figure S1 in the Supporting Information). A corresponding schematic drawing illustrates the arrangement of the lattice and sublattice.

monomers and self-assemblies is $D = 2.4$ (Figure 3B dark blue), which would be consistent with the formation of layers with a dimension between a two-dimensional plate and a three-dimensional solid. The entire signal with the exception of peaks originating from the structure factor is reconstructed from the baseline, that is, the minima of the scattering curve, by the analytical function $I(Q) \propto K_{D/2-1}(Q\kappa)/(Q\kappa)^{(D/2-1)}$, where K is a Bessel-K function, Q the scattering vector, D the fractal dimension, and κ a characteristic magnitude of the unfolded monomer (Figure 3B yellow line). The details of the reconstruction procedure of a SAXS signal produced by S-layers in solution have been given previously.[15] The scattering contrast of the self-assembled S-layers is fitted numerically on the basis of the coarse-grained model (Figure 1l) and is shown as a red line in Figure 3B. The details of the numerical fitting procedure were given previously.[14] On the basis of the coarse-grained model, it is possible to numerically fit the minima of the Bragg peaks (the red line in Figure 3B follows the pattern of the Bragg peaks in a defined way), which indicates that the proteins only exhibit their tertiary structure when assembled into the lattice structure. However, the fact that we can perfectly fit the scattering contrast of the self-assembled layers using the coarse-grained monomeric model substantiates the presented calculated tertiary structure of the protein, at least on a coarse-grained level.

In a second step, we refine the simulated coarse-grained monomer model (Figure 1l) on the basis of the SAXS data. A reverse Monte Carlo algorithm was applied[15] to reconstruct scattering sites, visible as red beads in Figure 3A. The set of identified clusters corresponds to secondary structure elements at these locations. The scattering intensity arises in particular from those regions with a high electron density contrast: it is visible in Figure 3A that this electron density contrast is concentrated at the terminal domains of the monomer, whereas those parts of the structure, where overlapping takes place to result in the final tetramer, are exhibiting less contrast. This decrease of electron density contrast as a consequence of the formation of bonds supports the presented structural model of the S-layer tetramer.

For the reconstruction of the whole scattering signal of S-layer self-assemblies, we divide the signal by the form factor and thus obtain contributions of the self-assembled monomolecular sheets. The reconstruction

as shown in Figure 3C is calculated using eq 1 (we give the detailed deduction of eq 1 in the Methods section). If the self-assembled monomolecular layers in solution adapt a particular orientation with respect to the scattering vector, the contribution to the scattering contrast is supposed to be dominant over all other orientations. This is in line with the results we observed in our previous work,[15] where another S-layer protein was analyzed in solution by means of SAXS. In that case, we described the self-assembled layers as nanodisks, which were oriented parallel to each other and thereby formed some kind of nematic liquid. Here we refine this theory, as we anticipate that parallel orientated nanodisks contribute most to the SAXS signal.

$$\mathcal{F}(\rho(\zeta))[Q]/\hat{\rho}^{(1)}(Q) = -\int_0^1 \mathrm{d}(\cos\theta) f(\cos\theta)\hat{\delta}_{\parallel}(QR\cos\theta)\,\hat{\delta}_{\perp}(H^2Q\cos\theta)\hat{\delta}_{\parallel}(QR_{\gg}\cos\theta) \quad (1)$$

Therein H represents the height of the layer, R is the in-plane radius of the protein, and $R_{\gg}$ a multiple of R (*i.e.*, the layer size).

On the basis of eq 1, we can reconstruct the entire structure factor and deduce a possible characteristic height $H = 5.6$ nm, radius $R = 7.2$ nm, and nanodisk size $R_{\gg} = 72$ nm. This size, obtained from the model reconstruction, which takes into account the whole scattering curve, is completely consistent with the size obtained from the identified Bragg reflections (Figure 3C vertical gray lines). They indicate an arrangement of the tetramers in a cubic lattice with $a = 13.0 \pm 0.1$ nm. An additional substructure is visible by a number of weaker reflections (Figure 3C vertical red lines). They are attributed to the slightly oblique arrangement of the monomers along the main axes of the cubic lattice and are thus described by a 2D monoclinic sublattice with unit cell dimensions of $a = 13.0 \pm 0.1$ nm, $c = 10.5 \pm 0.1$ nm, and $\gamma = 80 \pm 1°$ (a detailed view of the Bragg reflections is given in Figure S1 in the Supporting Information).

The resulting structural model of one SbpA unit cell is shown in Figure 4. In the tetramer, the monomers are interlocked into one another, where both termini are accessible at either surface. The N-termini are located at the inner surface of the tetramer, which is anchored on the cell surface *via* SLH domains (Figure 4B), whereas the C-termini are accessible at the outer surface of the S-layer unit cell (Figure 4C). Figure 4D,E shows the scattering clusters reconstructed from the SAXS curves as red beads of one monomer in the unit cell. Apparently, the electron density contrast is higher in those domains that are not part of direct overlaps in the lattice. These tetramers build up the S-layer lattice (Figure 4A), leading to the formation of defined pores between the single unit cells. These pores exhibit a

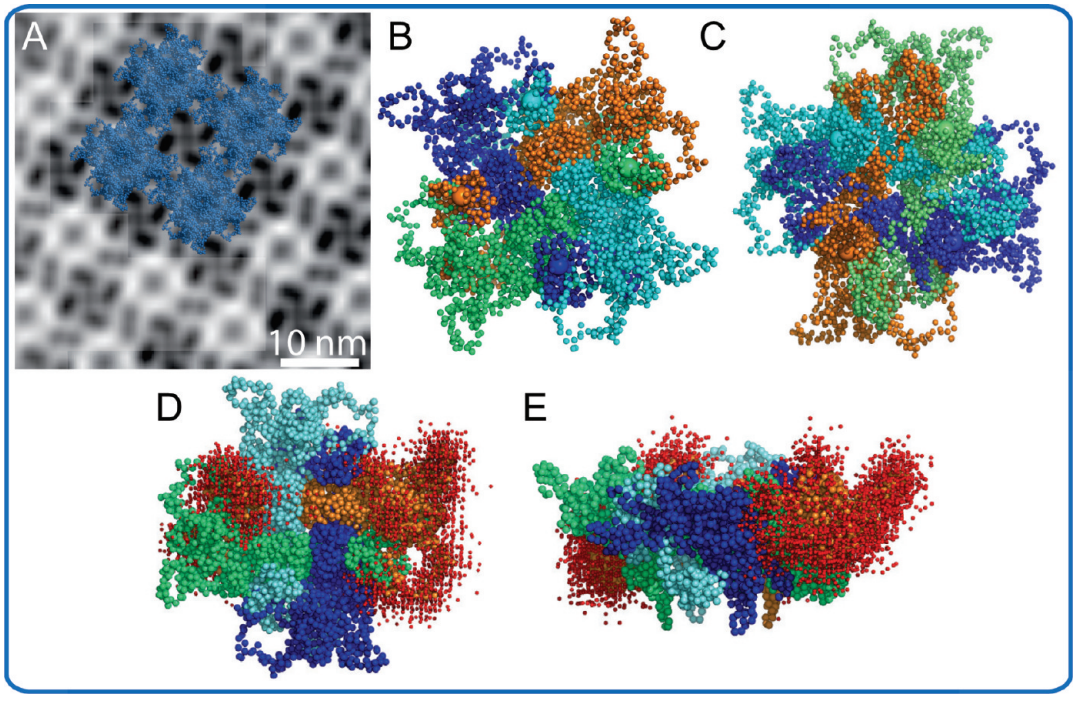

Figure 4. Structure of the resulting SbpA unit cell. (A) Comparison of the calculated structure with an electron micrograph showing an SbpA lattice. (B–E) Every monomer in the tetramer is illustrated in a different color. The proteins are interlocked into each other. (B) Inner surface of the tetramer, which anchors the protein on the cell surface. The N-termini are represented by magnified beads and are accessible on the surface. (C) Outer surface of the tetramer, which is exposed to the surroundings of the cell. The C-termini are also accessible and marked as magnified beads. (D,E) Red beads represent scattering clusters of one monomer of the tetramer as shown in Figure 3A. The overlapping or interacting parts do not show high electron density contrast.

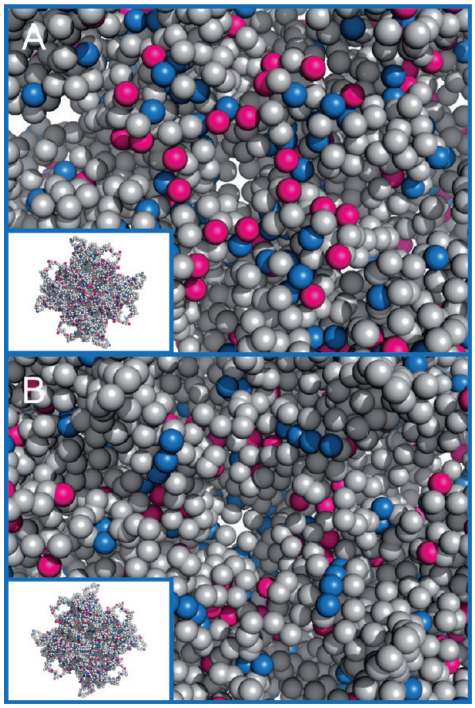

Figure 5. Enlarged view of the central region of one tetramer: (A) inner surface, which is mainly positively charged in the central region; (B) outer surface exhibiting mainly negatively charged amino acids. Pink beads, positively charged amino acids; blue beads, negatively charged amino acids.

very regular cubic arrangement (with a pore-to-pore distance of 13.0 nm) and lead to the strong Bragg peaks in the scattering curve in the high *Q* regime (see Figure 3C and Figure S1 in the Supporting Information). In the center of the unit cells, there is a clear anistropic charge distribution at the outer and inner surface, where at the inner surface the center is surrounded by positively charged residues and at the outer surface by negatively charged residues, as shown in detail in Figure 5A,B. The accessibility of the respective termini and the different charge distribution of the outer and the inner central region of the unit cell have already been demonstrated experimentally.[16,21,25,30] Various S-layer fusion proteins have been experimentally investigated, where truncated forms of the protein were used.[22,26,31,32] Figure 6 shows the correspondent truncated forms using the structural model of the unit cell. The deletion of 237 C-terminal amino acids leads to the change of the lattice symmetry from p4 to p1.[27] Figure 6A,B shows the unit cell, where the deleted amino acids are colored blue. If the C-terminus is shortened by another consecutive 113 amino acids, the proteins are not capable of self-assembling anymore. These residues are shown in Figure 6C,D. Figure 6E,F shows the corresponding monomers, where it becomes clear that the main part of the C-terminal domain, which can be deleted without any loss of functionality, is located outside of the lattice plane and is not part of any distinct overlaps. However, the deletion of this

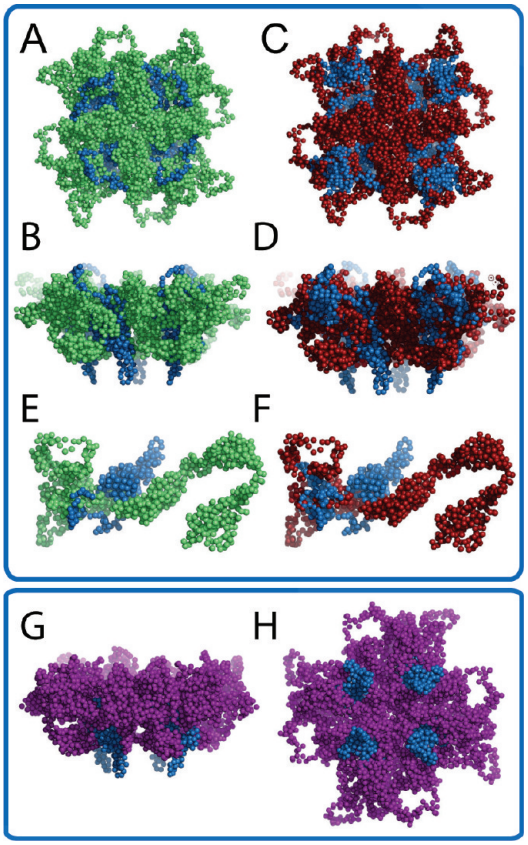

Figure 6. Analysis of the resulting structural model of the SbpA unit cell. The tetramer is compared to experimentally investigated truncated recombinant forms of the protein. (A,B) Blue beads represent the last 237 C-terminal amino acids. The deletion of the blue part leads to a change of the lattice symmetry from p4 to p1.[27] (E) Corresponding monomer. (C,D) Further deletion of 113 amino acids (blue part) leads to a loss of functionality. The proteins are not able to self-assemble anymore. (F) Corresponding monomer. The tetrameric structure shows clearly that the C-terminal part that is not located in the plane of the lattice can be deleted without any loss of function. If additional parts in the plane of the unit cell are deleted, the monomers are not able to form the tetramer. (G,H) Unit cell made up of a recombinant form of the monomer that lacks the last 170 amino acids (rSbpA1068). This form of the protein has been extensively studied for applications as fusion partner for various molecules. It was reported that this recombinant form is better accessible than other truncated forms.[17,20−22,31] The cartoons clearly show that the C-terminal part, which is located out of the plane of the unit cell, is deleted, which smoothes the surface and makes the tetramer better accessible.

separate domain leads to a clear change in the monomer structure, which might very well explain the change in lattice symmetry, where a p4 symmetry demands a more complicated and flexible monomeric structure than a p1 symmetry. Apparently, the further truncation of parts of the protein, which are located inside the plane, leads to a loss of functionality. Interestingly, the electron density contrast, as determined by SAXS (Figure 4D,E), gives very low scattering sites at exactly these locations, which leads to the hypothesis that this part is involved in essential interactions in the tetramer. The most investigated and successfully used

S-layer fusion protein contains a C-terminally truncated form, which maintained the residues 1−1068, where it was reported that this recombinant protein provides a highly accessible fusion partner.[17,20−22,31] Figure 6G,H shows this recombinant protein as part of a unit cell, where the blue part represents the truncated domain. The better accessibility of the fused molecule might be explained by the smoothening of the outer surface through deletion of the C-terminal part, which is located outside of the lattice plane and therefore sticking out of the unit cell.

CONCLUSIONS

In summary, we have demonstrated the successful combination of *in silico* and experimental methods to calculate the three-dimensional structure of an S-layer lattice, which will pave the way for the directed bottom-up design of nanoscale structures based on S-layer proteins. On the basis of an atomistic model calculated using molecular dynamics simulations and molecular modeling, the unit cell was modeled by combining the monomer structure and three-dimensional density distribution data as obtained from tilting studies using electron microscopy. The resulting unit cell was equilibrated again by molecular dynamics simulations and subsequently used to fit small-angle X-ray scattering intensity data of self-assemblies and monomers in solution. Interestingly, the calculated scattering clusters, which represent high electron density contrasts, are located at domains of the monomer that are not part of significant interactions in the tetrameric model, which shows that electron density contrasts are diminished—at least at a resolution achieved by SAXS—if interactions take place. This consistency between the SAXS results and the simulated unit cell structure regarding overlapping parts strongly substantiates the structural model of the unit cell. Furthermore, the SAXS results of the monomeric sample indicate that the proteins are not fully structured in their monomeric state. The hypothesis of the importance of conformational transformations guiding the S-layer self-assembly has been also recently proposed by Chung *et al.*[12] and legitimizes our approach to base the calculation of the tetrameric unit cell on an intermediate monomeric structure. The resulting structural model shows an anisotropy regarding the charge distribution in the center of the tetramer, which has been seen experimentally before. We explained the structure of a tetramer based on truncated recombinant forms of the S-layer protein that have been investigated experimentally and used for nanobiotechnological applications to better understand the behavior of these forms when assembled into S-layer lattices. The presented structure of an S-layer unit cell can serve as a basis for the specific and directed binding of various molecules. So far, the usage of S-layers as building blocks was somehow based on a trial and error approach due to the lack of structural details. The exact location of N- and C-terminal domains in the unit cell, the charge distribution on the inner and outer central region of the S-layer lattice, as well as the altered surface architecture due to the truncation of monomers will help to better understand the behavior of nanostructures based on S-layers and to specifically alter certain parts of the S-layer lattice for the production of nanostructures with different behavior and various architectures. Together with already available experimental data of S-layer proteins, this approach opens the path to determining locations, type, and distribution of amino acids in the S-layer lattice.

METHODS

Protein Preparation. The S-layer protein SbpA from *Bacillus sphaericus* CCM2177 was isolated by default as previously described.[33] After the cell wall preparation, the proteins were extracted with 5 M guanidine hydrochloride (pH 7.2) at room temperature for 30 min. Afterward, the cell wall fragments were removed by centrifugation (30 000 rpm, 60 min, 8°C). The supernatant containing the proteins was dialyzed against 10 mM $CaCl_2$ for 2 h at pH 5.9. The dialyzed sample, which contains self-assemblies of SbpA, was used for the SAXS measurements. For the TEM measurements, one drop of the protein sample was adsorbed on a grid and negatively stained as previously described.[34] A monomer solution was obtained by centrifuging for 30 min at 8 °C and 30 000 rpm.

Molecular Modeling. The S-layer protein SbpA was modeled using different bioinformatics tools and molecular dynamics simulations. Sequence homology searches were performed using BLAST[35] at www.ncbi.nlm.nih.gov/BLAST, where homologies could be found to other S-layer proteins, especially regarding the S-layer homology domains (SLH domains at the N-terminus). Additionally, we found homologies to RTX toxins and related Ca^{2+}-binding proteins. The S-layer protein SbpA needs bivalent cations to self-assemble in solution into the typical two-dimensional sheets, which might explain structural homologies to Ca^{2+}-binding proteins. We also found homologies to Ig-like proteins, which characteristically contain β-sheet structures and to fibronectin type III domains. A comparison of the SbpA sequence with these structural homologues resulted in 26% identity to the S-layer protein SbsB, which has been modeled in our previous work,[14] 34% identity to an Ig-like protein, 25% to RTX toxins and the related Ca^{2+}-binding protein, 32% to a fibronectin type III domain, and 42% identity to the SLH domain of the S-layer protein of *Bacillus cereus*. Secondary structure predictions were made using PSIPRED[36] available at www.psipred.net/psiform.html, DOMPRED[37] at http://bioinfadmin.cs.ucl.ac.uk/dompred, and GOR4[38] at http://pbil.ibcp.fr/htm/index.php. All three algorithms predict mainly α-helices for the N-terminal region and β-sheets for the rest of the protein, which is a common structural motive for S-layer proteins. Finally, we performed domain predictions using CDART[39] available at www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi and Pfam[40] at http://pfam.sanger.ac.uk.

Both algorithms predicted three SLH domains at the N-terminus of the protein. Pfam additionally predicted two Ig-like domains and CDART two fibronectin type III domains. On the basis of all these predictions and the information of the self-assembly behavior of truncated forms of the protein, we defined seven structurally meaningful domains: three SLH domains, one Ca^{2+}-binding domain, three Ig-like or fibronectin type III domains. The molecular dynamic simulations were performed using LAMMPS,[41] which is distributed by the Sandia National Laboratories and free to download at http://lammps.sandia.gov. All simulations were computed at the SUN cluster Phoenix at phoenix.zserv.tuwien.ac.at. Visualizations were done using the PyMOL Molecular Graphics System, Version 1.2r3pre, Schrodinger, LLC.

Transmission Electron Microscopy. Tilt series were performed using a Philips transmission electron microscope CM12 operated at 80 kV. The tilt range was ±60°. For the three-dimensional image reconstruction, we used a software package from Caldiris: CRISP 2.1.a (to edit tilted electron micrographs and perform FFT (fast Fourier transformation)), TriMerge 1.6.a (to reconstruct the three-dimensional density based on Fourier transforms of electron micrographs and to perform the inverse Fourier transform[29]), and TriView 1.3 (to visualize three-dimensional density data). All in all, we performed six tilt series in 5 and 10° steps at a magnification of 37 000.

Small-Angle X-ray Scattering. Small-angle X-ray scattering (SAXS) was performed with Cu Kα radiation from a rotating anode generator (Nanostar, BRUKER AXS, Karlsruhe, Germany) equipped with a pinhole camera and an area detector (VANTEC 2000 from BRUKER AXS). The sample containing self-assembled SbpA proteins was put into capillaries with 1 mm diameter and 10 μm wall thickness (from Hilgenberg) and then sealed with polymeric caps, as the whole equipment operates in vacuum. The SAXS intensity patterns were taken at a sample to detector distance of 109 cm for 6 h. They were corrected for background scattering and then radially averaged to obtain the function $I(q)$, where $q = (4\pi/\lambda)\sin\theta$ is the scattering vector, 2θ the angle between incident and diffracted beam, and $\lambda = 0.1542$ nm the X-ray wavelength. The scattering intensities were normalized to each other in the q range between 2 and 2.5 nm^{-1}. In this region, the scattering intensity is flat and dominated mainly by fluid scattering and an additional smaller contribution from the glass capillary. The scattering intensity from the solution without protein was then subtracted from the solution with protein. We cross-checked that this is consistent with the measured transmission of each sample.

For the analysis of the scattering data, $I(Q)$, we adapt our theory that we have recently formulated to access protein self-assemblies.[15] We follow the Green's approach and give the local density distribution of all proteins in solution by

$$\mathcal{F}(\rho(\zeta))[Q] = I(Q) = \mathcal{F}\left(\int_R^\infty \left(\int_R^\infty G(u'-u'')\rho^{(1)}(u'')\mathrm{d}u''\right)(1-\beta w(\zeta-u'))\mathrm{d}u'\right)[Q] \quad (2)$$

For the Green's function, we assume $G(u'-u'') = \delta(u'-u'')$, where this gives a possible solution of the wave equation. Therein the local density distribution of a single protein is given by $\rho^{(1)}(\zeta)$, where protein−protein interactions are approximated by $(1-\beta w(\zeta - u'))$. This approximation is exact for hard-core potentials. In reciprocal space, the complex convolution given in eq 2 simplifies to $I(Q) = \hat{G}(Q)\hat{w}(Q)^{(1)}(Q)$. Consequently, we calculate the structure factor of the entire system based on the form factor of the protein $\rho^{(1)}(Q)$, which represents the intramean potential of one monomer.[15]

We anticipate that the proteins self-assemble in a nanodisk-like manner.[15] Furthermore, we assume that the nanodisks interact only by hard-core potentials. Then $w(\zeta) = \delta(\zeta^2 - H^2)\Theta(\zeta - R) = \int\pi(R_i)\delta(\zeta^2 - H^2)\delta(\zeta - R_i)\mathrm{d}R_i$, with H being the disk height and R the disk radius. In contrast to our previous work,[15] we introduce a radial distribution function $\pi(R_i)$. If there is no disk radius favored over the other, this probability function may be set to 1. However, we now assume that large disks, that is, disks of a radius $R_\gg$ contribute most. Thus we set $\pi(R_i) = \delta(\zeta - R_\gg)$ and give the protein structure factor by

$$\mathcal{F}(\rho(\zeta))[Q]/\hat{\rho}^{(1)}(Q) = -\int_0^1 \mathrm{d}(\cos\theta) f(\cos\theta)\hat{\delta}_\parallel(QR\cos\theta)\hat{\delta}_\perp(HQ\sin\theta)\hat{\delta}_\parallel(QR_\gg\cos\theta) \quad (3)$$

Therein $\hat{\delta}_{\perp,\parallel}(...) = J_{D\perp,\parallel/2-1}(...)\,(...)^{D\perp,\parallel/2-1}$ indicates the fractal Fourier transform of the respective $\delta_{\perp,\parallel}$ functions.

We emphasize that $f(\cos\theta)$ may be argued as a summation of all possible orientations of the nanodisk with respect to the scattering vector. We now introduce a partition function in terms of the system's mean potential $w(\zeta,\zeta')$, with the system coordinates ζ and an arbitrary linear reaction coordinate ζ'.[42−44] The partition function may be given by $\Pi(\zeta') = \langle\exp(-\beta w(\zeta,\zeta'))\rangle_\zeta$. An equivalent formulation is

$$\Delta\Pi = -\beta^{-1}\left\langle\int_{-\infty}^0 \mathrm{d}\zeta'\partial_{\zeta'}\left(w(\zeta,\zeta')\exp(-\beta w(\zeta,\zeta'))\right)\right\rangle_\zeta \quad (4)$$

with $\beta = 1/k_BT$. In the literature, this approach is described as λ-integration[42,45] or mean force method.[46] To change the integration boundaries, we introduce a coordinate transformation $\xi' = \exp(-\lambda)$ and rewrite eq 4 by

$$\Delta\Pi = -\beta^{-1}\left\langle\int_0^1 \mathrm{d}\lambda\partial_\lambda\exp(-\beta(w(\zeta,\lambda) - \beta^{-1}\ln|J|_\lambda|))\right\rangle_\zeta \quad (5)$$

The Jacobi determinant is essential for the calculation of the mean force,[43] and here it is $\exp(\lambda)$. If we compare eq 2 and eq 5, it is straightforward that $f(\cos\theta)$, which equals $-\beta\partial_{\cos\theta}w(\cos\theta)$, is rather a mean force than a distribution of the orientation of the nanodisks, given by $\cos\theta$. If we aim for the particular distribution function, we have to calculate the Boltzmann weighted mean potential: $\Delta\pi(\cos\theta) = \exp(-\beta w(\cos\theta))$.[47]

We now simplify eq 5 taking into account $\tan\theta = H/Q$:

$$\mathcal{F}(\rho(\zeta))[Q]/\hat{\rho}^{(1)}(Q) = \int_0^1 \mathrm{d}(\cos\theta)(\partial_{\cos\theta}w(\cos\theta) - 1)\hat{\delta}(QR\cos\theta)\hat{\delta}_\perp(H^2\cos\theta)\hat{\delta}_\parallel(QR_\gg\cos\theta) \quad (6)$$

Finally do note that the term $R_\gg\cos\theta$ may as well be interpreted as particular distances ζ that the proteins comprise when forming self-assemblies.

Acknowledgment. The authors thank Jacqueline Friedmann for the preparation of the SAXS samples. C.H. holds a DOC-fFORTE fellowship of the Austrian Academy of Sciences. This project has been supported by the Air Force Project FA9550-09-1-0342.

Supporting Information Available: Detailed view of Figure 3 and the mean potential as determined by SAXS. This material is available free of charge *via* the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

1. Sleytr, U. B. Regular Arrays of Macromolecules on Bacterial Cell Walls: Structure, Chemistry, Assembly, and Function. *Int. Rev. Cytol.* **1978,** *53*, 1–64.
2. Sleytr, U. B. Heterologous Reattachment of Regular Arrays of Glycoproteins on Bacterial Surfaces. *Nature* **1975,** *257*, 400–402.
3. Sleytr, U. B.; Messner, P.; Pum, D.; Sara, M. Crystalline Bacterial Cell Surface Layers (S-Layers): From Supramolecular Cell Structure to Biomimetics and Nanotechnology. *Angew. Chem., Int. Ed.* **1999,** *38*, 1034–1054.
4. Sara, M.; Sleytr, U. B. S-Layer Proteins. *J. Bacteriol.* **2000,** *182*, 859–868.
5. Egelseer, E. M.; Ilk, N.; Pum, D.; Messner, P.; Schäffer, C.; Schuster, B.; Sleytr, U. B. Nanobiotechnological Applications of S-Layers. In *Encyclopedia of Industrial Biotechnology: Bioprocess, Bioseparation, and Cell Technology*; Flicking, M. C., Ed.; Wiley and Sons: Weinheim, Germany, 2009; pp 4424−4448.
6. Pavkov, T.; Oberer, M.; Egelseer, E. M.; Sara, M.; Sleytr, U. B.; Keller, W. Crystallization and Preliminary Structure Determination of the C-Terminal Truncated Domain of the

S-Layer Protein SbsC. *Acta Crystallogr. Sect. D: Biol. Crystallogr. D* **2003,** *59*, 1466–1468.
7. Pavkov, T.; Egelseer, E. M.; Tesarz, M.; Svergun, D. I.; Sleytr, U. B.; Keller, W. The Structure and Binding Behavior of the Bacterial Cell Surface Layer Protein SbsC. *Structure* **2008,** *16*, 1226–1237.
8. Howorka, S.; Sara, M.; Wang, Y.; Kuen, B.; Sleytr, U. B.; Lubitz, W.; Bayley, H. Surface-Accessible Residues in the Monomeric and Assembled Forms of a Bacterial Surface Layer Protein. *J. Biol. Chem.* **2000,** *275*, 37876–37886.
9. Kinns, H.; Badelt-Lichtblau, H.; Egelseer, E. M.; Sleytr, U. B.; Howorka, S. Identifying Assembly-Inhibiting and Assembly-Tolerant Sites in the SbsB S-Layer Protein from *Geobacillus stearothermophilus*. *J. Mol. Biol.* **2010,** *395*, 742–753.
10. Fagan, R. P.; Albesa-Jove, D.; Qazi, O.; Svergun, D. I.; Brown, K. A.; Fairweather, N. F. Structural Insights into the Molecular Organization of the S-Layer from *Clostridium difficile*. *Mol. Microbiol.* **2009,** *71*, 1308–1322.
11. Norville, J. E.; Kelly, D. F.; Knight, T. F., Jr.; Belcher, A. M.; Walz, T. 7 Å Projection Map of the S-Layer Protein SbpA Obtained with Trehalose-Embedded Monolayer Crystals. *J. Struct. Biol.* **2007,** *160*, 313–323.
12. Chung, S.; Shin, S.-H.; Bertozzi, C.; De Yoreo, J. Self-Catalyzed Growth of S Layers *via* an Amorphous-to-Crystalline Transition Limited by Folding Kinetics. *Proc. Natl. Acad. Sci. U.S.A.* **2010,** *107*, 16536–16541.
13. Whitelam, S. Control of Pathways and Yields of Protein Crystallization through the Interplay of Nonspecific and Specific Attractions. *Phys. Rev. Lett.* **2010,** *105*, 088102-1–088102-4.
14. Horejs, C.; Pum, D.; Sleytr, U. B.; Tscheliessnig, R. Structure Prediction of an S-Layer Protein by the Mean Force Method. *J. Chem. Phys.* **2008,** *128*, 065106-1–065106-10.
15. Horejs, C.; Pum, D.; Sleytr, U. B.; Peterlik, H.; Jungbauer, A.; Tscheliessnig, R. Surface Layer Protein Characterization by Small Angle X-ray Scattering and a Fractal Mean Force Concept: From Protein Structure to Nanodisk Assemblies. *J. Chem. Phys.* **2010,** *133*, 175102-1–175102-8.
16. Ilk, N.; Völlenkle, C.; Egelsser, E. M.; Breitwieser, A.; Sleytr, U. B.; Sara, M. Molecular Characterization of the S-Layer Gene, sbpA, of *Bacillus sphaericus* CCM 2177 and Production of a Functional S-Layer Fusion Protein with the Ability To Recrystallize in a Defined Orientation while Presenting the Fused Allergen. *Appl. Environ. Microbiol.* **2002,** *68*, 3251–3260.
17. Badelt-Lichtblau, H.; Kainz, B.; Völlenkle, C.; Egelseer, E. M.; Sleytr, U. B.; Pum, D.; Ilk, N. Genetic Engineering of the S-Layer Protein SbpA of *Lysinibacillus sphaericus* CCM 2177 for the Generation of Functionalized Nanoarrays. *Bioconjugate Chem.* **2009,** *20*, 895–903.
18. Delcea, M.; Krastev, R.; Gutberlet, T.; Pum, D.; Sleytr, U. B.; Toca-Herrera, J. L. Thermal Stability, Mechanical Properties and Water Content of Bacterial Protein Layers Recrystallized on Polyelectrolyte Multilayers. *Soft Matter* **2008,** *4*, 1414–1421.
19. Moreno-Flores, S.; Kasry, A.; Butt, H.-J.; Vavilala, C.; Schmittel, M.; Pum, D.; Sleytr, U. B.; Toca-Herrera, J. L. From Native to Non-Native Two-Dimensional Protein Lattices through Underlying Hydrophilic/Hydrophobic Nanoprotrusions. *Angew. Chem., Int. Ed.* **2008,** *120*, 4785–4788.
20. Ilk, N.; Küpcü, S.; Moncayo, G.; Klimt, S.; Ecker, R. C.; Hofer-Warbinek, R.; Egelseer, E. M.; Sleytr, U. B.; Sara, M. A Functional Chimaeric S-Layer-Enhanced Green Fluorescent Protein To Follow the Uptake of S-Layer-Coated Liposomes into Eukaryotic Cells. *Biochem. J.* **2004,** *379*, 441–448.
21. Huber, C.; Egelseer, E. M.; Ilk, N.; Sleytr, U. B.; Sara, M. S-Layer-Streptavidin Fusion Proteins and S-Layer-Specific Heteropolysaccharides as Part of a Biomolecular Construction Kit for Application in Nanobiotechnology. *Microelectron. Eng.* **2006,** *83*, 1589–1593.
22. Völlenkle, C.; Weigert, S.; Ilk, N.; Egelseer, E. M.; Weber, V.; Loth, F.; Falkenhagen, D.; Sleytr, U. B.; Sara, M. Construction of a Functional S-Layer Fusion Protein Comprising an Immunoglobulin G-Binding Domain for Development of Specific Adsorbents for Extracorporeal Blood Purification. *Appl. Environ. Microbiol.* **2004,** *70*, 1514–1521.
23. Schuster, B.; Gufler, P. C.; Pum, D.; Sleytr, U. B. Interplay of Phospholipase A_2 with S-Layer-Supported Lipid Monolayers. *Langmuir* **2003,** *19*, 3393–3397.
24. Tang, J.; Ebner, A.; Badelt-Lichtblau, H.; Völlenkle, C.; Rankl, C.; Kraxberger, B.; Leitner, M.; Wildling, L.; Gruber, H. J.; Sleytr, U. B.; *et al.* Recognition Imaging and Highly Ordered Molecular Templating of Bacterial S-Layer Nanoarrays Containing Affinity-Tags. *Nano Lett.* **2008,** *8*, 4312–4319.
25. Györvary, E.; Schroedter, A.; Talapin, D. V.; Weller, H.; Pum, D.; Sleytr, U. B. Formation of Nanoparticle Arrays on S-Layer Protein Lattices. *J. Nanosci. Nanotechnol.* **2004,** *4*, 115–120.
26. Ilk, N.; Kosma, P.; Puchberger, M.; Egelseer, E. M.; Mayer, H. F.; Sleytr, U. B.; Sara, M. Structural and Functional Analyses of the Secondary Cell Wall Polymer of *Bacillus sphaericus* CCM 2177 That Serves as an S-Layer-Specific Anchor. *J. Bacteriol.* **1999,** *181*, 7643–7646.
27. Huber, C.; Ilk, N.; Rünzler, D.; Egelseer, E. M.; Weigert, S.; Sleytr, U. B.; Sara, M. The Three S-Layer-like Homology Motifs of the S-Layer Protein SbpA of *Bacillus sphaericus* CCM2177 Are Not Sufficient for Binding to the Pyruvylated Secondary Cell Wall Polymer. *Mol. Microbiol.* **2005,** *55*, 197–205.
28. Kelley, L. A.; Sternberg, M. J. E. Protein Structure Prediction on the Web: A Case Study Using the Phyre Server. *Nat. Protoc.* **2009,** *4*, 363–371.
29. Amos, L. A.; Henderson, R.; Unwin, P. N. Three-Dimensional Structure Determination by Electron Microscopy of Two-Dimensional Crystals. *Prog. Biophys. Mol. Biol.* **1982,** *39*, 183–231.
30. Tang, J.; Badelt-Lichtblau, H.; Ebner, A.; Preiner, J.; Kraxberger, B.; Gruber, H. J.; Sleytr, U. B.; Ilk, N.; Hinterdorfer, P. Fabrication of Highly Ordered Gold Nanoparticle Arrays Templated by Crystalline Lattices of Bacterial S-Layer Protein. *ChemPhysChem* **2008,** *9*, 2317–2320.
31. Pleschberger, M.; Neubauer, A.; Egelseer, E. M.; Weigert, S.; Lindner, B.; Sleytr, U. B.; Muyldermans, S.; Sara, M. Generation of a Functional Monomolecular Protein Lattice Consisting of an S-Layer Fusion Protein Comprising the Variable Domain of a Camel Heavy Chain Antibody. *Bioconjugate Chem.* **2003,** *14*, 440–448.
32. Huber, C.; Liu, J.; Egelseer, E. M.; Moll, D.; Knoll, W.; Sleytr, U. B.; Sara, M. Heterotetramers Formed by an S-Layer-Streptavidin Fusion Protein and Core-Streptavidin as a Nanoarrayed Template for Biochip Development. *Small* **2006,** *2*, 142–150.
33. Sleytr, U. B.; Sara, M.; Küpcü, Z..; Messner, P. Structural and Chemical Characterization of S-Layers of Selected Strains of *Bacillus stearothermophilus* and *Desulfotomaculum nigrificans*. *Arch. Microbiol.* **1986,** *146*, 19–24.
34. Messner, P.; Hollaus, F.; Sleytr, U. B. Paracrystalline Cell Wall Surface Layers of Different *Bacillus stearothermophilus* Strains. *Int. J. Syst. Bacteriol.* **1984,** *34*, 202–210.
35. Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acids Res.* **1997,** *25*, 3389–3402.
36. McGuffin, L. J.; Bryson, K.; Jones, D. T. The PSIPRED Protein Structure Prediction Server. *Bioinformatics* **2000,** *16*, 404–405.
37. Marsden, R. L.; McGuffin, L. J.; Jones, D. T. Rapid Protein Domain Assignment from Amino Acid Sequence Using Predicted Secondary Structure. *Protein Sci.* **2002,** *11*, 2814–2824.
38. Garnier, J.; Gibrat, J.-F.; Robson, B. GOR Method for Predicting Protein Secondary Structure from Amino Acid Sequence. *Methods Enzymol.* **1996,** *266*, 540–553.
39. Geer, L. Y.; Domrachev, M.; Lipman, D. J.; Bryant, S. H. CDART: Protein Homology by Domain Architecture. *Genome Res.* **2002,** *12*, 1619–1623.
40. Finn, R. D.; Mistry, J.; Tate, J.; Coggill, P.; Heger, A.; Pollington, J. E.; Gavin, O. L.; Gunesekaran, P.; Ceric, G.;

Forslund, K.; *et al.* The Pfam Protein Families Database. *Nucleic Acids Res.* **2010,** *38*, 211–222.
41. Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995,** *117*, 1–19.
42. Chandler, D. *Introduction to Modern Statistical Mechanics*; Oxford University Press: New York, 1987.
43. Otter, W. K.; Briels, W. J. The Calculation of Free-Energy Differences by Constrained Molecular-Dynamics Simulations. *J. Chem. Phys.* **1998,** *109*, 4139–4146.
44. Jarzynski, C. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.* **1997,** *78*, 2690–2693.
45. McQuarrie, D. A. *Statistical Mechanics*; University Science Books: Mill Valley, CA, 1976.
46. Tscheliessnig, R.; Geyrhofer, L.; Wendland, M.; Fischer, J. Adsorption from Oversaturated Aqueous Solution: Mean Force Molecular Simulations. *AIChE J.* **2008,** *54*, 2479–2486.
47. Glatter, O.; Kratky, O. *Small Angle X-Ray Scattering*; Academic Press: London, 1982.