THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 286, NO. 29, pp. 26042–26049, July 22, 2011
Printed in the U.S.A.

# Structure of Surface Layer Homology (SLH) Domains from *Bacillus anthracis* Surface Array Protein*[S]

Received for publication, April 5, 2011, and in revised form, May 12, 2011 Published, JBC Papers in Press, May 13, 2011, DOI 10.1074/jbc.M111.248070

**Justin Kern[‡1], Rosemarie Wilton[§], Rongguang Zhang[§], T. Andrew Binkowski[§], Andrzej Joachimiak[§¶], and Olaf Schneewind[‡2]**

*From the [‡]Department of Microbiology, University of Chicago, Chicago, Illinois 60637, the [§]Biosciences Division, Midwest Center for Structural Genomics and Structural Biology Center, Argonne National Laboratory, Argonne, Illinois 60439, and the [¶]Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, Illinois 60637*

**Surface (S)-layers, para-crystalline arrays of protein, are deposited in the envelope of most bacterial species. These surface organelles are retained in the bacterial envelope through the non-covalent association of proteins with cell wall carbohydrates. *Bacillus anthracis*, a Gram-positive pathogen, produces S-layers of the protein Sap, which uses three consecutive repeats of the surface-layer homology (SLH) domain to engage secondary cell wall polysaccharides (SCWP). Using x-ray crystallography, we reveal here the structure of these SLH domains, which assume the shape of a three-prong spindle. Each SLH domain contributes to a three-helical bundle at the spindle base, whereas another α-helix and its connecting loops generate the three prongs. The inter-prong grooves contain conserved cationic and anionic residues, which are necessary for SLH domains to bind the *B. anthracis* SCWP. Modeling experiments suggest that the SLH domains of other S-layer proteins also fold into three-prong spindles and capture bacterial envelope carbohydrates by a similar mechanism.**

Surface layers (S-layers)[3] are para-crystalline sheets of protein, which self-assemble on the surface of microbial cells to form contiguous layers (1, 2). Most organisms that elaborate S-layers do so by abundantly producing and secreting a single protein species (3). Whether an organism produces an S-layer as a component of its envelope structure is assessed by electron microscopy of the cell surface (4). In this manner, species from nearly every branch of the Bacteria and Archaea have been discovered to produce S-layers (2). Proteins within S-layers fulfill variable functions in that they act either as a scaffold or enzyme in the bacterial envelope (5), promote nutrient diffusion or transport (6), or contribute to virulence by enabling microbial adhesion to infected host tissues (7).

Most, but not all, S-layer proteins of bacteria share three tandem ~55 amino acid repeats of the Surface Layer Homology (SLH) domain (8–10). Secreted proteins encoding three tandem SLH domains are tethered to the bacterial envelope by non-covalent interactions between the SLH domains and a secondary cell wall carbohydrate (11). SLH domains are remarkable for being both necessary and sufficient for the incorporation of chimeric proteins into S-layers (12, 13). The SbsC protein of *Geobacillus stearothermophilus* is an example for a class of protein that forms S-layers without SLH domains (14). SbsC binds to the secondary cell wall polysaccharide (SCWP) of *G. stearothermophilus* via its N-terminal domain, which consists of three triple-helical bundles that are connected by two contiguous helices (14). The N-terminal domain of SbsC has high similarity with S-layer proteins from *G. stearothermophilus*, *Geobacillus kaustophilus*, and *Geobacillus tepidamans* (14) and is not similar to proteins with SLH domains.

The Gram-positive bacterium *Bacillus anthracis* is a rod-shaped, spore-forming pathogen of mammalian hosts (15). The envelope of its vegetative forms is composed of a plasma membrane and peptidoglycan layer with attached secondary cell wall polysaccharide (SCWP) (16) and poly-D-γ-glutamic acid capsule (PDGA) (17, 18). The genome of *B. anthracis* encompasses 24 open reading frames whose predicted translation products each contain a secretion signal and three tandem SLH domains (19). An operon of two such genes, *sap* and *eag*, encodes the main S-layer proteins, Sap and EA1, and is adjacent to *csaB*, a gene required for decorating SCWP with ketal-pyruvate (11, 20). *B. anthracis* SCWP is a polymer with the repeating structure $[\rightarrow 6)$-α-GlcNAc-$(1\rightarrow 4)$-β-ManNAc-$(1\rightarrow 4)$-β-GlcNAc-$(1\rightarrow]_n$, where α-GlcNAc is substituted with α-Gal and β-Gal at O3 and O4, respectively, and the β-GlcNAc is substituted with α-Gal at O3 (21). The position of the ketal-pyruvate modification on the SCWP is not known. *B. anthracis* can form S-layers from both extractable antigen 1 (EA1) and surface array protein (Sap) by tethering the SLH domains of these polypeptides to pyruvylated SCWP (20, 22, 23). C-terminal to the SLH domains, S-layer proteins encode crystallization domains, sequences predicted to enable subunit-subunit interactions within the S-layer (22, 24, 25). A simple model for S-layer assembly is that secreted

* This work was supported, in whole or in part, by Grants GM074942 (to A. J.) and AI69227 (to O. S.) from the National Institutes of Health.

*The atomic coordinates and structure factors (code 3PYW) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).*

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Fig. S1.

[1] Supported by National Institutes of Health Training Grant GM007183.

[2] Membership within and support from the Region V "Great Lakes" Regional Center of Excellence in Biodefense and Emerging Infectious Diseases Consortium (GLRCE, National Institute of Allergy and Infectious Diseases Award 1-U54-AI-057153). To whom correspondence should be addressed: 920 E. 58th St., Chicago, IL 6063. Fax: 773-834-8150; E-mail: oschnee@bsd.uchicago.edu.

[3] The abbreviations used are: S-layer, surface layer; SLH, surface layer homology; RMSD, root mean square deviation; SCWP, secondary cell wall polysaccharide; EA, extractable antigen; Sap, surface array protein; IMAC, immobilized metal affinity chromatography; SAD, single-wavelength anomalous diffraction; BSL, *B. anthracis* S-layer-associated protein; PDB, Protein Data Bank.

subunits are recruited to the edge of an extant S-layer network via enthalpy-driven interactions between crystallization domains and are then tethered to the SCWP via the SLH domains (11). This model matches growth of the S-layer(s) with increases in the avidity of these networks for the cell wall. In this manner, bacilli assemble an S-layer on top of their peptidoglycan and thread PDGA capsule between S-layer protein subunits (26, 27). To gain insight into the molecular mechanisms of S-layer assembly, we determined here the three-dimensional structure of the SLH domains of Sap by x-ray crystallography and further explored the mechanism of binding to the SCWP.

## EXPERIMENTAL PROCEDURES

*Bacterial Strains and Plasmids—B. anthracis* Sterne 34F2 or its variants were grown in brain heart infusion broth (BHI) or Luria broth (LB) at 37 °C. *Escherichia coli* strains grown in LB and supplemented with ampicillin (100 $\mu$g ml$^{-1}$) where appropriate. p*sap*$_{SLH}$ was generated by PCR amplification from the *B. anthracis sap* open reading frame with the primers P281 (5′-GGAATTCCATATGGGTAAAACATTCCCAGACGTTC-3′) and P282 (5′-TTT CTCGAGTTCTGTACCGAACTGCTTGTCAG-3′). PCR product and pET16b vector were restricted with NdeI and XhoI, ligated and transformed into *E. coli*. p*gst-SLH*$_{1-3}$ was generated by PCR amplification from the *B. anthracis sap* open reading frame with the primers P277 (5′-TTTGGATCCGGTAAAACATTCCCAGACGTTC-3′) and P278 (5′-TTTGAATTCTTCTGTACCGAACTGCTTGTCAG-3′). The PCR product and pGEX-2TK vector were restricted with EcoRI and BamHI, ligated and transformed into *E. coli*. Truncation variants GST-SLH$_1$ and GST-SLH$_{1-2}$ were generated by PCR amplification with primers P277 and P603 (5′-TTTGAATTCGTAGCTGCTTCTGCACGAGTTAA-3′) or P604 (5′-TTTGAATTCTTCTACAAGAAGAGATGCCATAGAAAC-3′), respectively.

*Protein Crystallization—E. coli* BL21(DE3) cells harboring p*sap*$_{SLH}$ were grown under conditions that promote SeMet incorporation (28). Labeled protein was subjected to immobilized metal affinity chromatography (IMAC) (29), concentrated and purified further by size exclusion chromatography on a Superdex 75 16/60 column equilibrated with 20 mM HEPES, 250 mM NaCl, 2 mM DTT, pH 8.0. A Mosquito liquid dispenser (TTP LabTech) was used to dispense 0.4 $\mu$l of protein (20 mg/ml) and 0.4 $\mu$l of reservoir solution into 96-well CrystalQuick plates (Greiner); droplets were equilibrated against 140 $\mu$l of reservoir solution. Crystals formed in 0.1 M citrate pH 5.5, 2.0 M ammonium sulfate at 4 °C were treated with cryoprotectant (crystallization buffer containing 3.5 M ammonium sulfate), mounted on CryoLoops (Hampton Research) and frozen in liquid nitrogen.

*X-ray Diffraction and Structure Determination*—Single-wavelength anomalous diffraction (SAD) data were collected near the selenium absorption peak at 100 °K from a single SeMet-labeled crystal at the 19ID beamline of the Structural Biology Center at the Advanced Photon Source (Argonne National Laboratory) using the program SBCcollect. Intensities were integrated and scaled with the HKL3000 suite (30). Heavy atom sites were located using the program SHELXD (31) and phased with the program MLPHARE (32). The density modified map output was submitted for model building with programs ARP/wARP. The structure was completed with manual model building using the program COOT (33) and refined with REFMAC (34). The final R was 16.6% and the free R was 18.9% with zero $\Sigma$ cutoff. The stereochemistry of the structure was checked with PROCHECK (35). Comparison of Sap$_{SLH}$ structure with other proteins in the Protein Data Bank (PDB) via the DALI server (Holm) identified only very limited structural homology. The closest homologue is a domain of X-prolyl dipeptidyl aminopeptidase (PDB ID 1LNS, Z-score 5.8, RMSD 3.1 Å) (36) that matches about half of Sap$_{SLH}$.

*GST-SLH Purification and S-layer Assembly—E. coli* BL21 (DE3) carrying p*gst-SLH* or its variants was grown in 1-liter cultures to $A_{600}$ 1 and induced with 1 mM IPTG for 3 h at 37 °C. Cells were harvested by centrifugation, suspended in PBS and disrupted by French press. Cleared lysate was subjected to affinity chromatography on glutathione-Sepharose and eluted with 20 mM reduced glutathione. For circular dichroism studies, GST-SLH variants were bound to glutathione-Sepharose and treated with 50 units of thrombin (Sigma) for 3 h at room temperature. Eluate was treated with 100 $\mu$l of benzamidine-Sepharose to remove thrombin and dialyzed against 8 mM $NaH_2PO_4$, 1.5 mM $Na_2HPO_4$. Purified protein (100 $\mu$g ml$^{-1}$) was subjected to circular dichroism using an AVIV 202 CD Spectrometer at room temperature.

For S-layer assembly assays, *B. anthracis* Sterne was grown overnight in BHI. Vegetative forms were sedimented by centrifugation at 10,000 × *g*, suspended in 3 M urea and heated in a boiling water bath for 30 min to strip the S-layer from murein sacculi with attached SCWP. Cell wall preparations were washed with water and then with PBS and stored in aliquots at −20 °C. GST or GST-SLH were mixed with cell wall suspensions at the indicated $A_{600}$ and incubated at room temperature for 10 min followed by centrifugation at 10,000 × *g*. Supernatant and sediment were mixed with sample buffer, heated to 95 °C and analyzed by SDS-PAGE. Protein was quantified by measuring pixel intensity of acquired images with Adobe Photoshop CS3.

## RESULTS

*Crystal Structure of the SLH Domains of Sap*—The 814 amino acid Sap precursor harbors a N-terminal signal peptide (residues 1–30), three SLH domains (residues residues 34–197) as well as a large C-terminal domain (residues 210–814) that promotes the crystallization of the S-layer protein (25). The nucleotide sequence of the *B. anthracis sap* gene, *i.e.* codons 31 through 210, was cloned into the expression vector pET16b to generate p*sap*$_{SLH}$ (Fig. 1*A*). This plasmid was transformed into *E. coli* BL21(DE3) and T7 RNA polymerase-mediated expression of Sap$_{SLH}$ in bacterial cultures was induced with IPTG. Sap$_{SLH}$ encompasses an N-terminal ten histidyl tag and amino acids Gly$^{31}$-Glu$^{210}$ of Sap (Fig. 1*C*), and was purified from cleared bacterial lysates by affinity chromatography on Ni-NTA-Sepharose and analyzed for purity by Coomassie-stained SDS-PAGE (Fig. 1*B*). Sap$_{SLH}$ crystallized in the tetragonal space group $P4_12_12_1$ with one monomer in the asymmetric unit. Its structure was determined using the single-wavelength anomalous diffraction (SAD) approach with Se-Met labeled

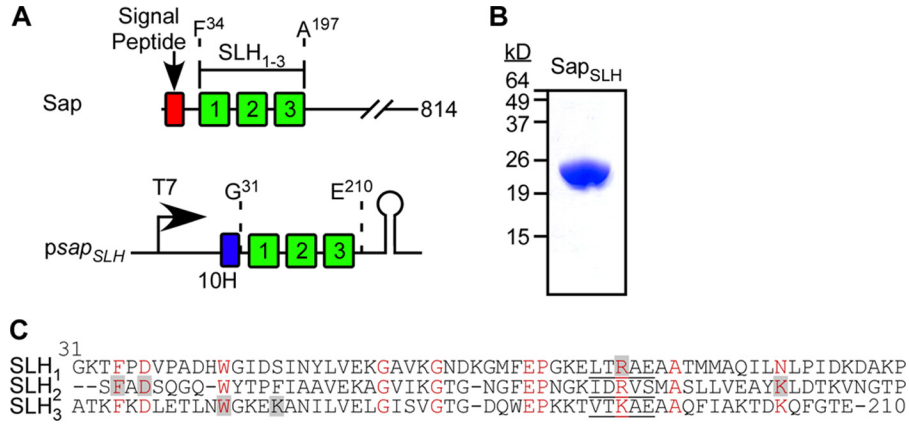


FIGURE 1. **Primary structure, expression and purification of $Sap_{SLH}$.** *A*, genetic organization of the *sap* gene (BAS0841) and its encoded signal peptide and SLH domains on the chromosome of *B. anthracis*. Nucleic acid sequence encoding the SLH domains was cloned into expression vector pET16b to generate p$sap_{SLH}$. *B*, recombinant protein $Sap_{SLH}$ with an N-terminal ten histidyl tag and thrombin cleavage site was purified by affinity chromatography and analyzed on Coomassie-stained SDS-PAGE. *C*, primary structure of the three SLH domains in Sap, aligned via their conserved residues (highlighted in *red*). *Shaded residues* are located in the inter-prong groove 2, positioned between the second and third SLH prong of Sap (see Fig. 5). Residues of the ITRAE motif in each of the three SLH domains are *underlined*.

**TABLE 1**
**Crystallographic studies on $Sap_{SLH}$**

| | |
|---|---|
| **Data Collection** | |
| Space group | $P4_12_12$ |
| Unit cell (Å) | $a = b = 77.77$, $c = 100.41$ |
| MW Da (residue) | 24353.4 (223)[1] |
| Mol (AU) | 1 |
| SeMet (AU) | 6 |
| Wavelength(Å) | 0.9794 (peak) |
| Resolution range (Å) | 54.96–1.80 (1.80–1.85)[2] |
| Number of unique reflections | 27750 (2003) |
| Redundancy | 11.1 (6.3)[2] |
| Completeness (%) | 99.86 (98.6) |
| $R_{merge}$ (%) | 0.099 (0.748)[2] |
| I/ $\sigma$(I) | 26.83 (1.61)[2] |
| **Phasing** | |
| $R_{Cullis}$ (anomalous) (%) | 0.70 |
| Figure of merit (%) | 0.2498 |
| **Refinement** | |
| Resolution | 54.96–1.80 |
| Reflections (work/test) | 27750/1486 |
| $R_{crystal}/R_{free}$ (%) | 16.6/18.9 |
| Rms deviation from ideal geometry | 0.017/1.485 |
| Bond length (Å)/angle (°) | |
| Number of protein atoms | 1741 |
| Number of sulfate/water molecules | 1/241 |
| Mean B-value (Å$^2$) (mainchain/sidechain) | 1.399/3.177 |
| **Ramachandran plot statistic (%)** | |
| Residues in most favored regions | 92.0 |
| Residues in additional allowed regions | 8.0 |

[1] Includes His tag.
[2] Numbers in parenthesis include highest resolution shell.

crystals and refined to 1.80 Å resolution. All residues could be assigned either to the most favored or the allowed regions of Ramachandran plot statistics (Table 1). The structural model that could be derived from these data accounts for the positions of amino acid residues Lys$^{32}$-Thr$^{209}$ of Sap (Fig. 1*C*).

The overall structure of the three SLH domains resembles a three-prong spindle, where each prong is derived from a single SLH domain (Fig. 2). The base of the spindle is assembled from all three domains, each of which contributes a single helix that associates into a three-helical bundle (Fig. 2). The three SLH domains of $Sap_{SLH}$, $SLH_1$ (residues 31–90), $SLH_2$ (91–151) and $SLH_3$ (152–209), share limited sequence identity: $SLH_1$ *versus* $SLH_2$ 26%, $SLH_1$ *versus* $SLH_3$ 39%, and $SLH_2$ *versus* $SLH_3$ 26%. Nevertheless, the structures formed by each of the three SLH domains are nearly identical (Fig. 3). Thus, $Sap_{SLH}$ can be considered as a pseudo-trimer that is assembled from its three SLH domains. When inspected from the C terminus (cargo view), the SLH domain prongs of $Sap_{SLH}$ proceed clockwise from the N terminus and surround the three helical bundle base of the spindle (*a*, *b*, and *c* in Fig. 2*A*). Each SLH domain includes a helix on the lateral side of the molecule, two loops (*A* and *B* in Fig. 2*B*) and the beginning of one linker helix (Fig. 2*B*).

*A Model for SCWP Binding to $Sap_{SLH}$*—A group of five residues, named the ITRAE motif for its consensus sequence (Fig. 1*C*), is partially conserved among the SLH domains of bacterial S-layer proteins (10) and occupies the last four residues of loop B and the first residue of the central helix bundle (Fig. 2*C*). Within the SLH domains of Sap, these motifs have the sequences LTRAE, IDRVS, and VTKAE and contain the cationic residues Arg$^{72}$, Arg$^{131}$, and Lys$^{193}$, respectively (Fig. 1*C*). The corresponding positively charged residues of the ITRAE motif are considered crucial for the incorporation of protein into the S-layer of *Thermoanaerobacterium thermosulfurigenes* (13). Analysis of the solvent accessible surface of $Sap_{SLH}$ revealed three small tunnels at the spindle base (Fig. 4). The tunnels and ITRAE motifs are arranged such that Arg$^{72}$ from $SLH_1$ penetrates the tunnel in $SLH_2$, appearing on the surface between $SLH_2$ and $SLH_3$ (Fig. 2*D*). Similarly, Arg$^{131}$ from $SLH_2$ penetrates the tunnel in $SLH_3$ and appears on the surface between $SLH_1$ and $SLH_3$, whereas Lys$^{193}$ is inserted into the $SLH_1$ tunnel and displayed on the surface between $SLH_2$ and $SLH_3$ (Fig. 2*D*). Thus, all three SLH domains contribute residues to the surface structure of each of the inter-prong grooves (IPG) that are formed by the adjacent prongs of S-layer proteins (Fig. 5, *A* and *B*).

We wondered whether the inter-prong grooves formed by adjacent SLH domains promote association of the S-layer protein with its carbohydrate ligand. In agreement with this conjecture, six amino acids corresponding to the most conserved residues among *B. anthracis* SLH domains contribute to the surface of the inter-prong grooves, for example Arg$^{72}$, Phe$^{95}$, Asp$^{97}$, Lys$^{143}$, Trp$^{164}$, and Lys$^{168}$ of $IPG_2$ (Figs. 5*C* and 1*C*). The SCWPs of different bacterial species represent a complex set of

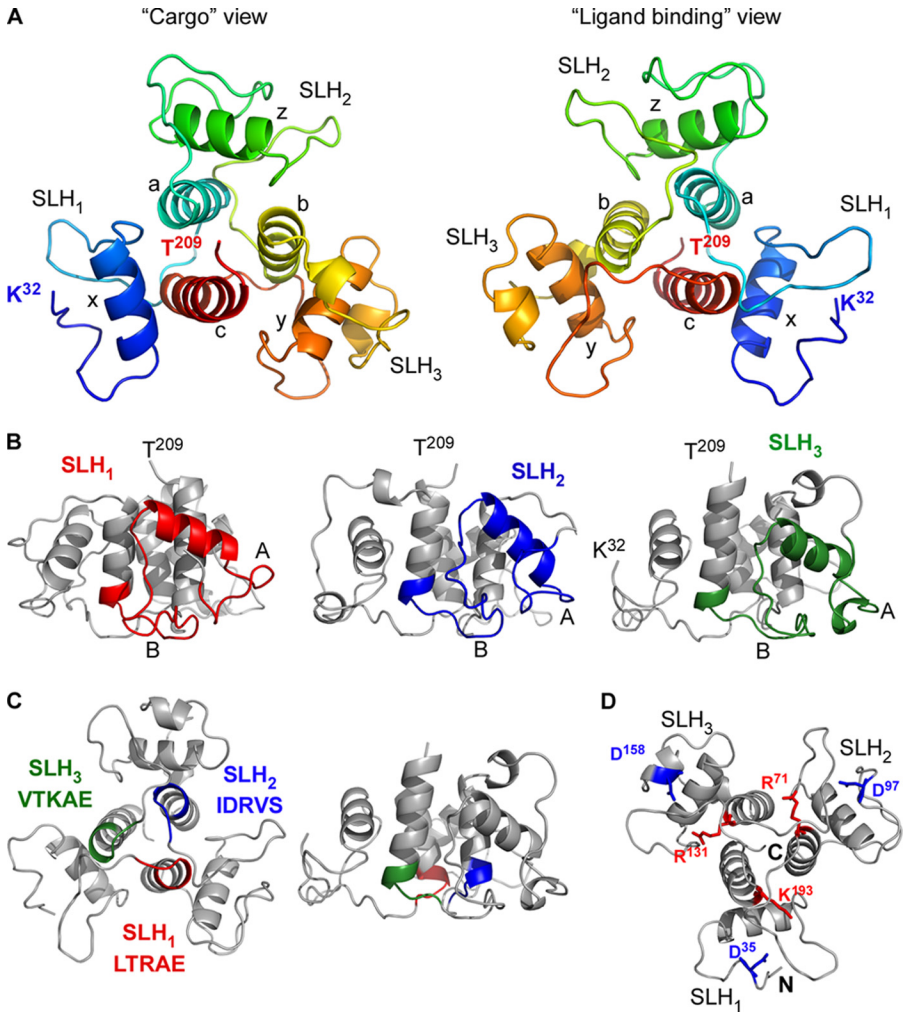


FIGURE 2. **The crystal structure of $Sap_{SLH}$.** *A*, two different views of the $Sap_{SLH}$ pseudo-trimer. Each SLH domain is noted, as is the N and C terminus and the six major helices (a, b, c, x, y, and z). The *left panel* shows the structure from the C terminus, which projects out of the page and away from the SLH domains, and is termed the "cargo" view. The *right panel* shows the ligand binding surface of the SLH-domains. *B*, lateral views of the $Sap_{SLH}$ crystal structure, each SLH domain (as annotated by NCBI) is colored accordingly. The central core of helices a, b, c is not part of the annotated SLH domain. *C*, conserved ITRAE motifs of each SLH domain are colored to show their location and are positioned between each pseudo-trimer at the bottom of helices a, b, and c. *D*, side chains and position of $Asp^{35}$, $Asp^{97}$, $Asp^{158}$, $Arg^{71}$, $Arg^{131}$, and $Lys^{193}$ are highlighted. Charged groups within these residues are in close proximity.

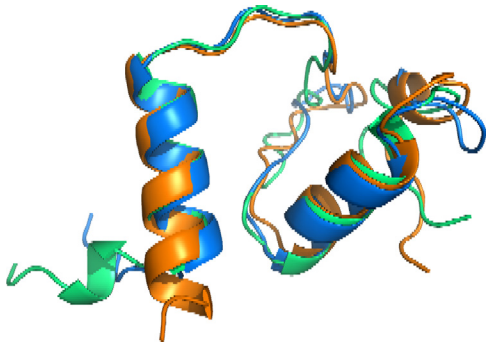

FIGURE 3. **Structural alignment of the three SLH domains of the Sap protein.** The alignment of $SLH_1$ (residues 32–90), $SLH_2$ (91–151), and $SLH_3$ (152–209) was performed with the Swiss Pro Viewer.

carbohydrates with hexose units organized in linear and branched fashion (37). We modeled *B. anthracis* SCWP into the structure of SLH domains to analyze the anchoring of carbohydrates within their three-prong spindle (Fig. 5).

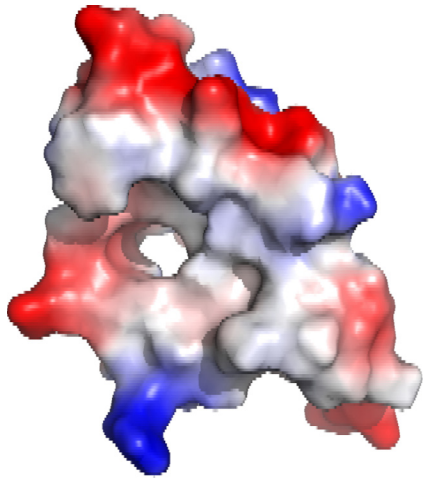

FIGURE 4. **Solvent accessible surface and electrostatic potential of the $SLH_1$ domain of Sap.** Analysis of solvent accessible surfaces identified a tunnel at the center of each of the three SLH domains of Sap.

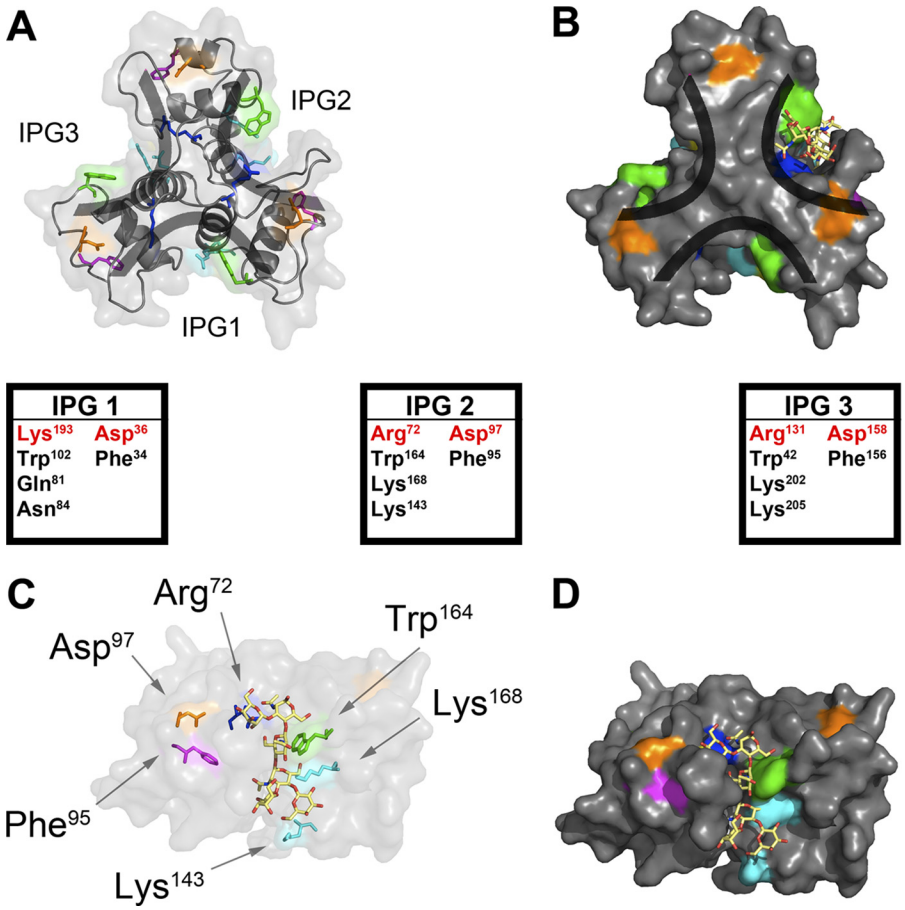


FIGURE 5. **Modeling the SCWP into the structure of the SLH domain of Sap.** The three-prong spindle arrangement of the SLH domain creates three nearly identical binding clefts designated inter-prong grooves ($IPG_{1-3}$). *A*, ribbon model of $Sap_{SLH}$. *B*, space-filling model of $Sap_{SLH}$. *C*, each of the predicted SCWP binding clefts contains a replica of the highly conserved residues, for example in $IPG_2$ $Asp^{97}$ (*orange*), $Arg^{72}$ (*magenta*), $Phe^{95}$ (*cyan*), $Lys^{143}$ (*blue*), $Trp^{164}$ (*green*), and $Lys^{168}$ (*blue*). SCWP was modeled into one of the binding domains, revealing hydrogen bonding and stacking interaction potential with the conserved residues. *D*, *space-filled structure* of the modeling experiment in *C* with the SCWP displayed as a *stick model*.

The SCWP molecule was constructed as described in Choudhury *et al.* (21). A selection of low energy conformers of the molecule were manually modeled into the presumed binding sites of $IPG_1$, $IPG_2$, and $IPG_3$ by performing rigid body rotations and translations of the SCWP to minimize steric clashing with SLH molecule. Once a suitable pose was identified, a short conjugate gradient minimization procedure was performed (38).

The conserved residues in $IPG_2$ and the modeled SCWP are shown in Fig. 5CD. Similar SCWP poses and conserved residue arrangements were observed for the equivalent residues on the surface of $IPG_1$ and $IPG_3$, but are not shown for visual clarity. The conserved phenylalanine residues ($Phe^{34}$, $Phe^{95}$, and $Phe^{156}$) are located at the domain interfaces, most likely playing an important role in domain packing (Fig. 5AB). However, the conserved aspartic acid residues ($Asp^{36}$, $Asp^{97}$, and $Asp^{156}$) are located outside of the SCWP binding domains. Their planar arrangement on the bottom of the molecule suggests that they may be involved in interacting with peptidoglycan or the cell wall linkage units, which provide a tether for SCWP and the envelope of bacilli (20). Of note, our modeling experiments with SLH domains cannot consider the essential contribution of the SCWP ketal-pyruvyl, as the position of this carbohydrate modification is not yet known. We therefore sought to test whether the conserved Asp and Arg/Lys residues in the inter-prong grooves of $Sap_{SLH}$ can indeed contribute to binding SCWPs in the envelope of bacilli.

*Functional Studies of $Sap_{SLH}$ Variants*—Glutathione *S*-transferase (GST) hybrids were used to examine the contribution of individual Sap SLH domains and their key residues toward *B. anthracis* SCWP binding. Purified GST-$SLH_{1-3}$, a hybrid encompassing amino acids 31–210 of Sap fused to the C terminus of GST, was incubated with *B. anthracis* vegetative forms that had been stripped of their S-layer. As a measure of SCWP binding, we monitored co-sedimentation of GST hybrids with bacilli following centrifugation (Fig. 6). As a control, GST alone did not sediment with vegetative forms, whereas all GST-$SLH_{1-3}$ molecules co-sedimented with 10 $A_{600}$ units of bacilli (Fig. 6, *A* and *B*). A reduction in the number of bacilli led to reduced co-sedimentation of GST-$SLH_{1-3}$ (Fig. 6, *A* and *B*). These results are in agreement with the general concept that the SLH domains of S-layer proteins and the SCWP of bacilli represent a receptor-ligand interaction. The removal of individual SLH domains in GST-$SLH_{1-2}$ and GST-$SLH_1$ abolished the association of GST hybrids with the SCWP of bacilli (Fig. 6, *A* and *B*). When compared with $SLH_{1-3}$, the CD spectra of isolated $SLH_{1-2}$ and $SLH_1$ domains displayed stepwise decreases in helix content but retained the overall pattern

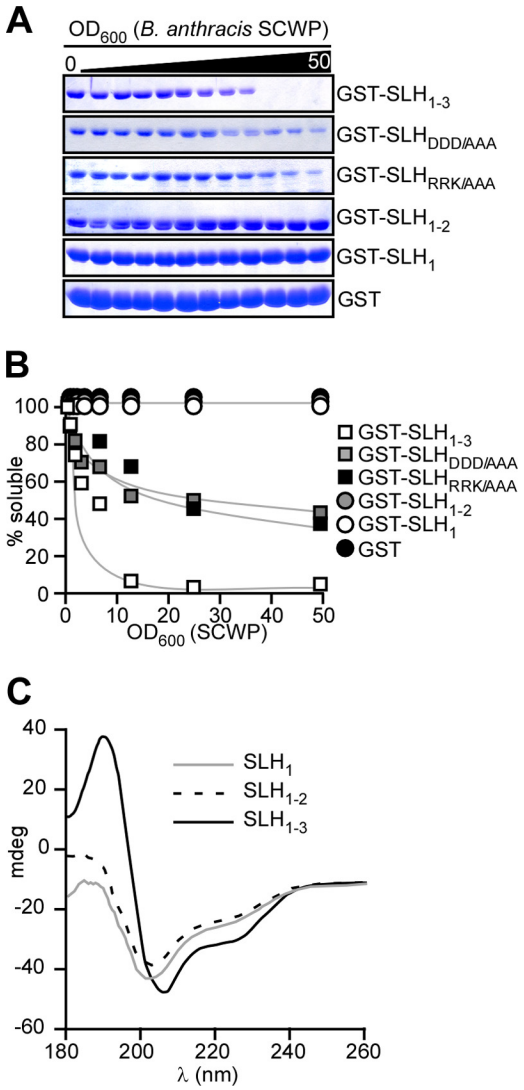


FIGURE 6. **Structural requirements of Sap SLH domains to associate with secondary cell wall polysaccharide of *B. anthracis*.** *A*, purified GST-$SLH_{1-3}$ and variants lacking the third (GST-$SLH_{1-2}$) or the second and third SLH domain (GST-$SLH_1$) or carrying substitutions at conserved acidic (GST-$SLH_{DDD/AAA}$) or basic residues (GST-$SLH_{RRK/AAA}$) were analyzed for their ability to co-sediment with variable numbers of *B. anthracis* vegetative forms that had been stripped of their S-layer proteins. GST (glutathione *S*-transferase) was used as a control for co-sedimentation and depletion of soluble purified protein with the *B. anthracis* cell wall envelope containing secondary cell wall polysaccharide detected by Coomassie-stained SDS-PAGE. *B*, quantification of the data in *panel A*. *C*, circular dichroism (CD) spectra for isolated $SLH_{1-3}$, $SLH_{1-2}$, and $SLH_1$ after their thrombin cleavage from GST hybrids.

expected of helical proteins (Fig. 6*C*). We presume that the lack of GST-$SLH_{1-2}$ and GST-$SLH_1$ association with SCWP is due to the inability of these variants to form the three-pronged spindle structure rather than defects in the overall folding of individual SLH domains. These data demonstrate that a functional binding interface requires the presence of all three tandem SLH domains.

A striking feature of our $Sap_{SLH}$ structure is the juxtaposition of conserved, basic residues extending into the inter-prong grooves of adjacent SLH domains (*vide supra*), to conserved aspartic acid residues positioned in the loop region of each spindle prong (Fig. 2*D*). To test whether these residues are required for SCWP association we generated two variants.

**TABLE 2**
**Alignment of SLH domains from *B. anthracis* S-layer proteins**

| Polypeptide | | Aligment of SLH domains | | | |
|---|---|---|---|---|---|
| NCBI locus | Name | All atoms | RMSD | Cα | RMSD |
| BAS0841 | Sap | – | – | – | – |
| BAS0842 | EA1 | 884 | 0.172 | 152 | 0.119 |
| pXO1–90 | BslA | 749 | 0.362 | 123 | 0.269 |
| pXO1–54 | BslB | 758 | 0.3 | 129 | 0.215 |
| BAS0916 | BslC | 697 | 0.461 | 116 | 0.3 |
| BAS1048 | BslD | 735 | 350 | 116 | 0.214 |
| BAS1049 | BslE | 759 | 0.287 | 130 | 0.205 |
| BAS1050 | BslF | 694 | 0.34 | 118 | 0.254 |
| BAS1787 | BslG | 762 | 0.292 | 132 | 0.21 |
| BAS2160 | BslH | 749 | 0.323 | 125 | 0.232 |
| BAS3093 | BslI | 758 | 0.312 | 125 | 0.203 |
| BAS3425 | BslJ | 752 | 0.317 | 127 | 0.225 |
| BAS1021 | BslK | 831 | 0.216 | 135 | 0.128 |
| BAS1246 | BslL | 755 | 0.227 | 139 | 0.168 |
| BAS2613 | BslM | 739 | 0.274 | 120 | 0.175 |
| BAS4693 | BslN | 748 | 0.281 | 127 | 0.222 |
| BAS1683 | BslO | 714 | 0.257 | 121 | 0.179 |
| BAS0829 | BslP | 639 | 0.327 | 106 | 0.2 |
| BAS3089 | BslQ | 660 | 0.292 | 114 | 0.201 |
| BAS3463 | BslR | 616 | 0.283 | 110 | 0.194 |
| BAS0851 | BslS | 832 | 0.2 | 134 | 0.118 |
| BAS1682 | BslT | 934 | 0.224 | 146 | 0.122 |
| BAS2351 | BslU | 797 | 0.19 | 141 | 0.133 |
| pXO2–42 | AmiA | 808 | 0.18 | 129 | 0.1 |

GST-$SLH_{RRK/AAA}$ harbors alanine substitutions in all three basic residues (R72A, R131A, and K193A), whereas GST-$SLH_{DDD/AAA}$ carries substitutions in the three aspartic acid residues (D36A, D97A, and D158A). GST-$SLH_{RRK/AAA}$ displayed a reduction in its SCWP binding capacity as approximately half of the protein failed to associate with vegetative forms even at 10 $A_{600}$ units (Fig. 6, *A* and *B*). A similar phenotype was observed for GST-$SLH_{DDD/AAA}$ variants. Thus, charged residues conserved in the inter-prong grooves of SLH domains contribute to the binding of S-layer proteins to the SCWP of *B. anthracis*.

*In Silico Prediction of the Three-prong Spindle Structure of SLH Domains from Other S-layer Proteins*—Excepting the amino acids $Phe^{34}$, $Asp^{36}$, $Trp^{42}$, $Gly^{54}$, $Gly^{57}$, $Glu^{64}$, $Pro^{65}$, $Arg^{72}$, $Ala^{76}$, and $Asn^{84}$ which are conserved in many *B. anthracis* SLH domains, the SLH domain peptide sequence is variable (Fig. 1*C*). To ascertain whether the sequence of these domains can satisfy the spatial constraints imposed by the empirically derived $Sap_{SLH}$ structure, we modeled aligned sequences of tandem SLH domain triplets from the *B. anthracis* EA1 S-layer protein and 22 S-layer associated proteins. Each SLH sequence was first aligned with Sap residues 31–209 using ClustalW.

These alignments and the $Sap_{SLH}$ PDB file were used to produce *in silico* predicted structures with MODELLER 9v8 using a standard Python script and models refined with PyMOL (MacPyMOL 1.1r1). *In silico* modeled SLH structures were compared with the empirically derived $Sap_{SLH}$ structure in two ways. First, we used the PyMol align function to measure root mean squared deviation (RMSD) in angstroms (Å) between a set of all atoms in the modeled SLH domains and the $Sap_{SLH}$ structure (Table 2). Using these criteria, every set of three SLH domains produced average RMSD values from $Sap_{SLH}$ equal to or less than 0.461 Å (Table 2). Performing the same calculation with only the α-carbon backbone, we derived alignments where the largest average RMSD distance was calculated to be 0.269 Å for 123 carbons (Table 2). Second, we subjected all 24 SLH

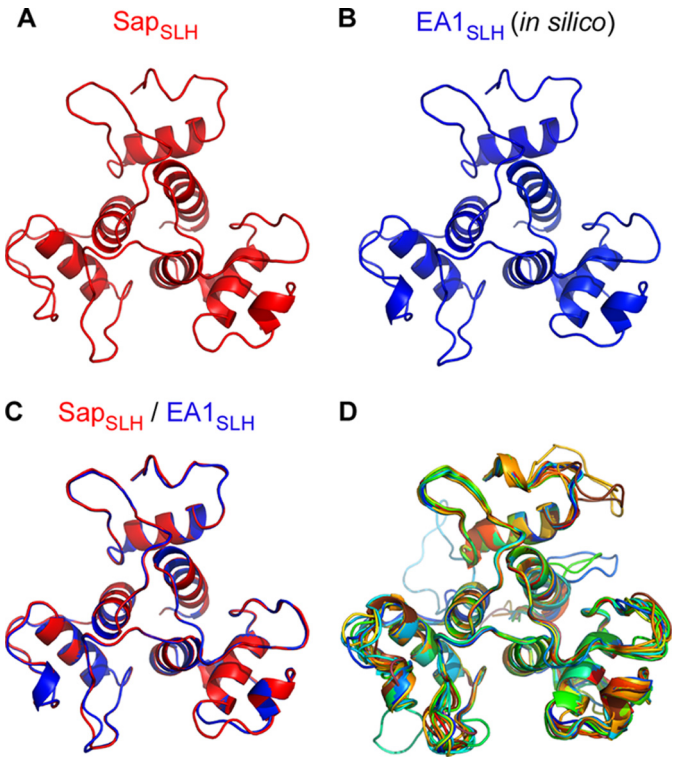


FIGURE 7. **Modeling SLH domains from all *B. anthracis* S-layer proteins to Sap$_{SLH}$.** The crystal structure of Sap$_{SLH}$, shown from the ligand-binding surface (*top left*), was used to model the aligned sequences of all the remaining 23 *B. anthracis* SLH domain sequences into structures. *A*, modeled structure of the SLH domains of EA1 (EA1$_{SLH}$) aligns very closely with the empirically determined structure of SLH domains from Sap (*B*), producing an RMSD of 0.172 Å. The *bottom right panel* shows all 24 SLH domain structures aligned. The RMSD values for each alignment are provided, along with the aligned sequence in Table 2.

domain PDB files to analysis using PROCHECK (35) to judge the stereochemical quality of each modeled SLH domain (and Sap$_{SLH}$). PROCHECK measures whether or not all residues lay within the allowed portions of the Ramachandran plot and scores the likeliness of all modeled bond lengths and angles occurring. Supplemental Table S1 records the PROCHECK outputs for Sap$_{SLH}$ and all 23 modeled SLH domains. Many of our in silico determined structures satisfy all criteria while some fail by containing a single or as many as two amino acids with disallowed F/Ψ angles. As the observed RMSD values, derived from greater than 100 atoms, are low and each of these modeled SLH domains satisfy most, if not all, PROCHECK criteria, we propose that, though highly divergent in primary sequence (Supplemental Table S1), all SLH domains within the *B. anthracis* genome likely adopt a similar fold (Fig. 7).

The high degree of similarity in shape and position can be appreciated by viewing the Sap$_{SLH}$ structure and the modeled structure for the SLH domains of EA1 (EA1$_{SLH}$) when displayed side-by-side or superimposed (Fig. 7, *A* and *B*). The modeled EA1$_{SLH}$ structure aligns closely with the Sap$_{SLH}$ structure (Fig. 7, *A--C*), producing an RMSD of 0.119 Å for 152 $\alpha$-carbons (Table 2) with all amino acids within the allowed regions of the Ramachandran plot (Supplemental Table S1). These results explain earlier observations that the SLH domains of Sap and EA1 bind to the envelope of *B. anthracis* with similar affinity (39). A superimposition of all 23 modeled SLH domains aligned to Sap$_{SLH}$ is provided in Fig. 7*D*. Perfect alignment is not achieved in all cases, as some SLH domain sequences, when aligned to Sap, exhibit higher disparity in their primary sequences (Fig. 7*D*). Nevertheless, these structures still contain remarkably similar secondary structures with all models containing helices a, b, c, x, y, and z in the same position and of the same length. We therefore hypothesize that all SLH domains of S-layer proteins in *B. anthracis* form a three-prong spindle structure and bind to the SCWP in a manner similar to the SLH domain of Sap. Analyzing the predicted secondary structure of SLH domains from S-layer proteins of diverse microbes, Engelhardt and Peters predicted that SLH domains may assume a similar structure (40). Their prediction is corroborated by the results presented in Fig. 7. Structural coordinates for Sap$_{SLH}$ have been deposited in the Protein Database, PDB identification number 3PYW.

## DISCUSSION

Binding of SLH domains to pyruvylated secondary cell wall carbohydrates is thought to be an ancestral mechanism for the anchoring of S-layer proteins to the envelope of bacteria (41). *B. anthracis* elaborates S-layers from two secreted polypeptides with N-terminal SLH domains, Sap and EA1 (22, 42). The SLH domains of Sap and EA1 bind to pyruvylated SCWP, a carbohydrate that is tethered via peptidoglycan linkage units to the murein sacculus of this microbe (20). Sap and EA1 are synthesized by bacilli in great abundance and form two-dimensional para-crystalline arrays on solid surfaces, a feature that is encrypted in their large C-terminal crystallization domains (25). Twenty-two *B. anthracis* S-layer-associated proteins (BSLs) also harbor N-terminal SLH domains (19). Unlike Sap and EA1, BSLs are minor components of the envelope that use their SLH domains to engage SCWPs but do not form para-crystalline arrays (19, 20). One of the S-layer associated proteins, BslA, promotes the adhesion of bacilli to host tissues and is required for the pathogenesis of anthrax disease (7, 19). The S-layer protein BslK binds to hemin, thereby contributing to an iron-uptake pathway that retrieves heme from hemoglobin for subsequent transport across the many layers of the *B. anthracis* envelope (6, 43). The AmiA protein contains three tandem SLH domains and is a peptidoglycan hydrolase (44). The molecular function of other S-layer proteins is not yet known, however, a set of eight, including AmiA, are putative hydrolases, which may act to shape the peptidoglycan layer of bacilli.

Here we used x-ray crystallography to reveal the three-dimensional structure of the SLH domain of Sap, whose overall shape resembles that of a three-prong spindle. Molecular modeling experiments suggest that the SLH domains of other proteins assume a similar structure. Further, the inter-prong grooves of this domain can accommodate both linear and branched carbohydrates, *i.e.* the secondary cell wall polysaccharides that are known to function as the docking platform for the assembly of S-layers. Our experiments identify conserved residues in the SLH domains of Sap and other S-layer proteins of bacilli (Phe$^{34}$, Asp$^{36}$, Trp$^{42}$, and Arg$^{72}$) and we show that two of these, Asp$^{36}$ and Arg$^{72}$, contribute to their interaction with SCWP.

The SCWP of *B. anthracis* is a polymer with the repeating structure $[\rightarrow 6)$-$\alpha$-GlcNAc-$(1\rightarrow 4)$-$\beta$-ManNAc-$(1\rightarrow 4)$-$\beta$-GlcNAc-$(1\rightarrow]_n$, where $\alpha$-GlcNAc is substituted $\alpha$-Gal and $\beta$-Gal at O3 and O4, respectively, and the $\beta$-GlcNAc is substituted with $\alpha$-Gal at O3 (21). CsaB is thought to generate pyruvyl-ketal modifications of SCWP, which are a pre-requisite for S-layer assembly (11). The pyruvyl-ketal introduces a strong negative charge modification into the polysaccharide that promotes binding of the SLH domains in S-layer proteins to the bacterial envelope (20). We surmise that some of the conserved positively charged residues in the inter-prong grooves enable S-layer protein binding to the pyruvyl-ketal of SCWP (20).

Cholera toxin represents another secreted bacterial polypeptide that binds to carbohydrates, specifically the GM1 glycosphingolipid of host cell membranes (45). Cholera toxin B assembles into a pentameric ring structure (46) and docks onto the host cell receptor in a manner whereby subunit interfaces capture the glycosphingolipid ligand (47). Choleragenoid, the toxin B subunit alone, is used as a vaccine antigen to elicit protective antibodies that prevent association of toxin with GM1 receptor. A similar strategy may be plausible to prevent S-layer assembly, and this may provide protection against *B. anthracis* or other infectious agents that also use S-layer proteins for the pathogenesis of their associated diseases (16).

*Acknowledgments—The submitted manuscript has been created by University of Chicago Argonne, LLC, Operator of Argonne National Laboratory ("Argonne"). Argonne, a U.S. Dept. of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government.*

## REFERENCES


1. Houwink, A. L., and Le Poole, J. B. (1952) *Physikalische Verhandlungen* **3,** 98
2. Engelhardt, H. (2007) *J. Struct. Biol.* **160,** 115–124
3. Sleytr, U. B. (1997) *FEMS Microbiol. Rev.* **20,** 5–12
4. Houwink, A. L. (1953) *Biochim. Biophys. Acta* **10,** 360–366
5. Houwink, A. L. (1956) *J. Gen. Microbiol.* **15,** 146–150
6. Tarlovsky, Y., Fabian, M., Solomaha, E., Honsa, E., Olson, J. S., and Maresso, A. W. (2010) *J. Bacteriol.* **192,** 3503–3511
7. Kern, J. W., and Schneewind, O. (2010) *Mol. Microbiol.* **75,** 324–332
8. Fujino, T., Béguin, P., and Aubert, J. P. (1993) *J. Bacteriol.* **175,** 1891–1899
9. Lupas, A., Engelhardt, H., Peters, J., Santarius, U., Volker, S., and Baumeister, W. (1994) *J. Bacteriol.* **176,** 1224–1233
10. Lupas, A. (1996) *Mol. Microbiol.* **20,** 897–898
11. Mesnage, S., Fontaine, T., Mignot, T., Delepierre, M., Mock, M., and Fouet, A. (2000) *EMBO J.* **19,** 4473–4484
12. Mesnage, S., Tosi-Couture, E., and Fouet, A. (1999) *Mol. Microbiol.* **31,** 927–936
13. May, A., Pusztahelyi, T., Hoffmann, N., Fischer, R. J., and Bahl, H. (2006) *Arch. Microbiol.* **185,** 263–269
14. Pavkov, T., Egelseer, E. M., Tesarz, M., Svergun, D. I., Sleytr, U. B., and Keller, W. (2008) *Structure* **16,** 1226–1237
15. Koch, R. (1876) *Beiträge zur Biologie der Pflanzen* **2,** 277–310
16. Leoff, C., Saile, E., Rauvolfova, J., Quinn, C. P., Hoffmaster, A. R., Zhong, W., Mehta, A. S., Boons, G. J., Carlson, R. W., and Kannenberg, E. L. (2009) *Glycobiology* **19,** 665–673
17. Bruckner, V., Kovacs, J., and Denes, G. (1953) *Nature* **172,** 508
18. Fouet, A. (2009) *Mol. Aspects Med.* **30,** 374–385
19. Kern, J. W., and Schneewind, O. (2008) *Mol. Microbiol.* **68,** 504–515
20. Kern, J., Ryan, C., Faull, K., and Schneewind, O. (2010) *J. Mol. Biol.* **401,** 757–775
21. Choudhury, B., Leoff, C., Saile, E., Wilkins, P., Quinn, C. P., Kannenberg, E. L., and Carlson, R. W. (2006) *J. Biol. Chem.* **281,** 27932–27941
22. Mesnage, S., Tosi-Couture, E., Mock, M., Gounon, P., and Fouet, A. (1997) *Mol. Microbiol.* **23,** 1147–1155
23. Mignot, T., Mesnage, S., Couture-Tosi, E., Mock, M., and Fouet, A. (2002) *Mol. Microbiol.* **43,** 1615–1627
24. Couture-Tosi, E., Delacroix, H., Mignot, T., Mesnage, S., Chami, M., Fouet, A., and Mosser, G. (2002) *J. Bacteriol.* **184,** 6448–6456
25. Candela, T., Mignot, T., Hagnerelle, X., Haustant, M., and Fouet, A. (2005) *Microbiology* **151,** 1485–1490
26. Mesnage, S., Tosi-Couture, E., Gounon, P., Mock, M., and Fouet, A. (1998) *J. Bacteriol.* **180,** 52–58
27. Richter, S., Anderson, V. J., Garufi, G., Lu, L., Budzik, J. M., Joachimiak, A., He, C., Schneewind, O., and Missiakas, D. (2009) *Mol. Microbiol.* **71,** 404–420
28. van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L., and Clardy, J. (1993) *J. Mol. Biol.* **229,** 105–124
29. Kim, Y., Dementieva, I., Zhou, M., Wu, R. Y., Lezondra, L., Quartey, P., Joachimiak, G., Korolev, O., Li, H., and Joachimiak, A. (2004) *J. Struct. Funct. Genomics* **5,** 111–118
30. Minor, W., Cymborowski, M., Otwinowski, Z., and Chruszcz, M. (2006) *Acta Crystallogr. D Biol. Crystallogr.* **62,** 859–866
31. Schneider, T. R., and Sheldrick, G. M. (2002) *Acta Crystallogr. D Biol. Crystallogr.* **58,** 1772–1779
32. CCP4 (COLLABORATIVE COMPUTATIONAL PROJECT, N.) (1994) *Acta Crystallogr. D Biol. Crystallogr.* **50,** 760–763
33. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2126–2132
34. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. D Biol. Crystallogr.* **53,** 240–255
35. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26,** 283–291
36. Chich, J. F., Rigolet, P., Nardi, M., Gripon, J. C., Ribadeau-Dumas, B., and Brunie, S. (1995) *Proteins* **23,** 278–281
37. Schäffer, C., and Messner, P. (2005) *Microbiology* **151,** 643–651
38. Case, D. A., Cheatham, T. E., 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K. M., Jr., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) *J. Comput. Chem.* **26,** 1668–1688
39. Chauvaux, S., Matuschek, M., and Beguin, P. (1999) *J. Bacteriol.* **181,** 2455–2458
40. Engelhardt, H., and Peters, J. (1998) *J. Struct. Biol.* **124,** 276–302
41. Cava, F., De Pedro, M. A., Schwarz, H., Henne, A., and Berenguer, J. (2004) *Mol. Microbiol.* **52,** 677–690
42. Etienne-Toumelin, I., Sirard, J. C., Duflot, E., Mock, M., and Fouet, A. (1995) *J. Bacteriol.* **177,** 614–620
43. Maresso, A. W., Garufi, G., and Schneewind, O. (2008) *PLoS pathogens* **4,** e1000132
44. Mesnage, S., and Fouet, A. (2002) *J. Bacteriol.* **184,** 331–334
45. van Heyningen, W. E., Carpenter, C. C., Pierce, N. F., and Greenough, W. B., 3rd (1971) *J. Infect. Dis.* **124,** 415–418
46. Zhang, R. G., Westbrook, M. L., Westbrook, E. M., Scott, D. L., Otwinowski, Z., Maulik, P. R., Reed, R. A., and Shipley, G. G. (1995) *J. Mol. Biol.* **251,** 550–562
47. Merritt, E. A., Sarfaty, S., van den Akker, F., L'Hoir, C., Martial, J. A., and Hol, W. G. (1994) *Protein Sci.* **3,** 166–175