INTRODUCTION

The surfaces of many Archaea and Bacteria are covered by a two-dimensional crystalline array known as the surface layer (S-layer). This macromolecular structure results from the non-covalent, entropy-driven, self-assembly of proteins (Sára & Sleytr, 2000). Most S-layers consist of a single protein species. Electron microscopy and image analysis procedures have shown that all S-layer proteins possess two morphological domains (Baumeister et al., 1989; Engelhardt & Peters, 1998). Biochemical, biophysical and genetic approaches have shown that one of these domains is the cell-wall-analyzing domain (Lupas et al., 1994; Mesnage et al., 1999; Jarosch et al., 2001; Smit et al., 2001; Rünzler et al., 2004). Gram-positive bacteria contain various anchoring domains that interact with different secondary cell wall polymers. The triple repetition of S-layer-homology (SLH) motifs constitutes an N-terminal SLH domain that interacts with a pyruvylated polysaccharide (Ilk et al., 1999; Mesnage et al., 2000). S-layer proteins devoid of SLH domains are anchored to different types of secondary cell wall polymers through their N- or C-terminal domains (Sára et al., 1996; Egelseer et al., 1998; Jarosch et al., 2000; Smit et al., 2001).

The other domain appears to be a crystallization domain (Bingle et al., 1987; Mesnage et al., 1997, 1999; Howorka et al., 2000; Jarosch et al., 2001; Smit et al., 2002; Mader et al., 2004). Recent studies have focused on the structure–function relationships of this domain. Mutagenesis approaches, be they point mutations or deletions, have identified regions within the crystallization domain that are critical for array assembly (Howorka et al., 2000; Sillanpää et al., 2000; Jarosch et al., 2001; Smit et al., 2002). However, these analyses have not provided an overall picture of how an S-layer protein self-assembles to generate an ordered array. A conclusion is that only small portions of the crystallization domain can generally be deleted before its capacity to self-assemble is lost (Sillanpää et al., 2000; Jarosch et al., 2001; Smit et al., 2002). Only a few
research groups have carried out structural analysis of S-layer protein crystallization (Claus et al., 2002; Jing et al., 2002; Pavkov et al., 2003), and the atomic structure remains unknown. The residues lining the pores and required for self-assembly remain to be identified.

*Bacillus anthracis*, the aetiological agent of anthrax, synthesizes two S-layer proteins: Sap and EA1 (Etienne-Toumelin et al., 1995; Mesnage et al., 1997). Both proteins have the same modular organization: an N-terminal anchoring domain consisting of three SLH motifs is fused to a C-terminal domain starting at residue 211, which is thought to be the crystallization domain (Mesnage et al., 1999). In rich medium, *B. anthracis* cells are surrounded by a Sap S-layer during the exponential phase, which is replaced by an EA1 S-layer as the cells enter the stationary phase (Mignot et al., 2002; Couture-Tosi et al., 2002). The Sap S-layer presents a fibril-like structure, and the array unit cell has the following dimensions: a, 81 Å (8-1 nm); b, 184 Å (18-4 nm); and c, 96 Å (Couture-Tosi et al., 2002). There is a possible p2 symmetry, but due to poor resolution on deflated bacteria, the molecular boundaries of each molecule could not be unambiguously defined. The projection map revealed several domains that repeat themselves along the two axes of the crystal. This sub-domain organization is consistent with the analysis of Sap following digestion with proteinase K, since resistant fragments smaller than, and internal to, the 600 aa C-terminal domain were obtained (Mesnage et al., 1999, and unpublished results).

In this work, we combined electron microscopy and genetics to analyse the 604 aa C-terminal (Sapc) domain, which is suggested to be the crystallization domain. For the genetic analysis, we took advantage of a bacterial (Bordetella pertussis) two-hybrid system to study the domains necessary for the interaction (Karimova et al., 1998). By both approaches, we showed that Sapc is the crystallization domain. The peptides responsible for the observed interactions were further sought by screening interactions of a representative library of fragments with Sapc. We describe here the polypeptides obtained.

**METHODS**

**Bacterial strains, plasmids and culture media.** The bacterial strains and plasmids used are listed in Table 1. *Escherichia coli* was grown in Luria–Bertani (LB) broth or on LB agar (Miller, 1972). *B. anthracis* cells were grown in BHI (Difco) broth or on BHI agar. The following antibiotics were added to the culture medium on which *E. coli* was grown: ampicillin, 100 μg ml⁻¹, kanamycin, 40 μg ml⁻¹, and chloramphenicol, 25 μg ml⁻¹. The ability of *E. coli* to ferment sugars was screened on MacConkey agar plates supplemented with 1% maltose (MK1%m).

**The bacterial two-hybrid system.** This method is based on the fact that the *Bor. pertussis* adenylate cyclase is inactive when the T18 and T25 domains are produced separately. Activity is restored when these domains are brought together, such as when fused T18 and T25 polypeptides interact. The interaction-mediated reconstitution of *Bor. pertussis* adenylate cyclase activity is easily monitored in the adenylate-cyclase (cya)-deficient *E. coli* strain DHPl (Karimova et al., 2000). The catalytic domain of the adenylate cyclase was reconstituted through the interaction of Sapc peptides (Sapc) corresponds to the potential crystallization domain, i.e. the C-terminal two-thirds of Sapc fused to the T18 and T25 fragments of the adenylate cyclase cloned into pUT18 or pUT18, and pT25 or pKT25, respectively. Red colonies on MK1%m indicate a positive interaction.

**β-Galactosidase assays.** These assays were performed on toluene-treated bacterial suspensions, as described by Miller (1972). Each assay was carried out at least four times using two independent cultures.

**Plasmid constructions.** A DNA fragment harbouring the 3'-terminal two-thirds of the sap gene, termed sapc, was amplified by PCR using Vent DNA polymerase (New England Biolabs) with oligonucleotides sap1 (TCGCAGAAATGTAGATCGAAAAGTCGTTAC) and sap2 (ATTTTGTCGGTCTTGCTTTTAAATAGAAA), and using *B. anthracis* 9131 chromosomal DNA as a template. The PCR fragments were phosphorylated using T4 polynucleotide kinase. The vectors pT25 and pT18, which encode T25 and T18 of *Bor. pertussis* adenylate cyclase, respectively, were cut with Smal, and ligated to the phosphorylated DNA fragments, giving rise to pT25-sapc and pT18-sapc, respectively.

pKT25-n, pKT25-m and pKT25-c, which were obtained during the library screening (see Results and Discussion), were cut with BamHI and Acc65I, and ligated into pUT18 digested with the same enzymes, giving rise to pUT18-n, pUT18-m and pUT18-c, respectively.

**Library construction.** A library was constructed in pKT25. Five micrograms of sap, PCR product was randomly sheared by sonication with two pulses of 15 s, each at 40 W, from a Branson sonifier cell disruptor B15. DNA fragments were treated with Vent DNA polymerase, and then phosphorylated and ligated into pKT25, which had been previously digested with Smal, and dephosphorylated with calf intestine alkaline phosphatase.

**Transformations.** *E. coli* strains TG1 and DHPl were transformed for library construction and two-hybrid library screening, respectively, by electroporation. Electroporation was carried out with a Bio-Rad Gene Pulser apparatus at a capacitance of 25 μF, a resistance of 200 Ω, and a voltage of 2-5 kV. Other transformations were carried out by the heat-shock method described by Chung & Miller (1988).

**His-Sap, purification.** Recombinant His-Sap, was overproduced and extracted according to Mignot et al. (2002), and was purified according to the procedure recommended by the manufacturer (Pharmacia). The final concentration of the protein was between 1-3 and 1-5 mg ml⁻¹.

**Two-dimensional crystallization on a lipid film.** A 1 μl volume of a lipid mixture was spread on the surface of a drop in a Tetlon well (60 μl) containing a buffer consisting of Tris/HCl 50 mM, pH 8.0, NaCl 250 mM. The lipid mixture was made of the ligand lipid 1,2-dioleoyl-sn-glycero-3-[N(5-amino-1-carboxypentyl)limino-diadicet acid[jucinyl] (nickel salt) (Avanti Polar Lipids) and the diluting lipid dioleyl-phosphatidylcholine (Avanti Polar Lipids) at a molar ratio of 1:4 in chloroform/methanol (9:1, v/v), and at a final concentration of 0-1 mg ml⁻¹. After overnight incubation at room temperature in a humid chamber, 1 μl Sapc protein solution was injected below the lipid layer. Crystallization samples were incubated at room temperature in a humid chamber.

After 2–24 h, plane carbon-coated grids were placed on top of the crystallization wells, and left in contact with the interface. After 15 min, the grids were picked up, and the transferred surface was negatively stained with 2% phosphotungstic acid.
Table 1. Bacterial strains and plasmids

<table>
<thead>
<tr>
<th>Strain or plasmid</th>
<th>Characteristics</th>
<th>Origin or reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E. coli</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BTH101</td>
<td>F⁻ Δaca99 rpsL(SrtS) hsdR2 galE15 galK16 mcrA1 mcrB1</td>
<td>Karimova et al. (1998)</td>
</tr>
<tr>
<td>DHP1</td>
<td>F⁻ gnlV44 (AS) recA1 endA1 glyA96 (Na⁺) thi-1 hsdR17 spoT1 rfbD1 cya-854 ilv-691:::Tn10</td>
<td>Karimova et al. (2001)</td>
</tr>
<tr>
<td>TG1</td>
<td>F² traD36 lacI⁴ ΔlacZΔ1 MIS proA⁺ B⁺ supE Δ(hdsM–mcrB)</td>
<td>Maniatis et al. (1982)</td>
</tr>
<tr>
<td><strong>B. anthracis</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9131</td>
<td>Plasmidless</td>
<td>Etienne-Toumelin et al. (1995)</td>
</tr>
</tbody>
</table>

**Plasmids**

<table>
<thead>
<tr>
<th>Insert and markers</th>
<th>Origin or reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pT25*</td>
<td>T25 adenylate cyclase subdomain; CamR</td>
</tr>
<tr>
<td>pT18*</td>
<td>T18 adenylate cyclase subdomain; AmpR</td>
</tr>
<tr>
<td>pKT25*</td>
<td>T25 adenylate cyclase subdomain; KanR</td>
</tr>
<tr>
<td>pUT18*</td>
<td>T18 adenylate cyclase subdomain; AmpR</td>
</tr>
<tr>
<td>pT25-Sapc</td>
<td>Sapc fused to T25; CamR</td>
</tr>
<tr>
<td>pT18-Sapc</td>
<td>Sapc fused to T18; AmpR</td>
</tr>
<tr>
<td>pKT25-n</td>
<td>N-Sapc subdomain fused to T25; KanR</td>
</tr>
<tr>
<td>pKT25-m</td>
<td>M-Sapc subdomain fused to T25; KanR</td>
</tr>
<tr>
<td>pKT25-c</td>
<td>C-Sapc subdomain fused to T25; KanR</td>
</tr>
<tr>
<td>pUT18-n</td>
<td>N-Sapc subdomain fused to T18; AmpR</td>
</tr>
<tr>
<td>pUT18-m</td>
<td>M-Sapc subdomain fused to T18; AmpR</td>
</tr>
<tr>
<td>pUT18-c</td>
<td>C-Sapc subdomain fused to T18; AmpR</td>
</tr>
</tbody>
</table>

*pT18 or pUT18 are compatible with pKT25 or pT25, and can be used in any combination to test two-hybrid interactions.

**RESULTS AND DISCUSSION**

The 604 aa C-terminal domain of Sap is the crystallization domain

The Sap S-layer protein is a modular protein, composed of an anchoring domain, the SLH domain, followed by a potential anchoring domain. Mesnage et al. (1999) showed that Sap has a relatively protease-resistant core, termed Sapc, which starts immediately after the SLH domain, and corresponds to the 604 C-terminal amino acids. We tested by structural and genetic approaches whether this 604 aa domain is the crystallization domain.

The structural approach was a crystallization assay. A His₆-Sapc two-dimensional crystallization was tested on a His-ligand-containing lipid film (Fig. 1). The position of the histidine residues (N-extremity of the Sapc domain) was identical to that of the SLH anchoring domain in the Sap S-layer protein. Electron micrographs of two-dimensional crystals were selected by optical diffraction for their highly ordered area. The best images showed the (2,10) diffraction order in the power spectrum corresponding to 18 Å resolution. This result indicates that Sapc is sufficient to form a crystalline structure. Furthermore, the unit-cell parameters were calculated as: a, 81 Å; b, 180 Å; γ, 96°. These values are in agreement with those of the Sap S-layer found at the surface of negatively stained deflated bacteria, which are: a, 81 Å; b, 184 Å; and γ, 96° (Couture-Tosi et al., 2002). This result indicates that Sapc constitutes the crystallization domain. This is consistent with other studies that have shown that the C-terminal domains of S-layer proteins, which begin with an anchoring domain, are the crystallization domains (Jarosch et al., 2001; Rünzler et al., 2004; Mader et al., 2004).

For the genetic approach, a bacterial two-hybrid system was used, which was based on a reconstituted signal transduction pathway assaying the interaction of two independently produced subdomains of *Bor. pertussis* adenylate cyclase. The restored activity of the functional *Bor. pertussis* cyclase, obtained when T18 and T25 domains are brought together, can easily be monitored because the synthesis of cAMP triggers the transcription of catabolic operons that are otherwise silent in an *E. coli* cya mutant (see Methods). Sapc was fused to the T18 and T25 domains, and the resulting proteins were co-expressed in an *E. coli* cya-deficient strain. Red colonies were obtained on MK1 %m sucrose plates, whereas white colonies were obtained when each fused Sapc was co-expressed with T18 or T25 alone. This meant that a functional adenylate cyclase was reconstituted in the...
presence of two Sapc domains. Sapc is therefore folded in a manner yielding a stable auto-interaction. This result indicates that Sapc is involved in the auto-assembly process, and that the SLH domain is not required for Sap–Sap interaction. These findings correlate with Sapc being the crystallization domain.

Construction and analysis of a Sapc library

The same two-hybrid approach was used to identify interacting subdomains or peptides of Sapc, but this time with small random fragments of the sapc sequence. A fusion library was constructed in pKT25. To create a representative and random sapc DNA library of small DNA fragments, the PCR product was sonicated, and a large variety of DNA fragments was obtained. Fragments of between about 0·1 and 1·8 kb (i.e. the size of sapc), with the vast majority around 0·2 kb, were obtained (data not shown). To harbour all putative interacting peptides, the library should contain at least 3300 clones. If we consider the average polypeptides (approx. 60 aa), and if each Sapc amino acid is in turn the first of the peptide, then 544 polypeptides (604–60) are required to cover Sapc completely. As only one insert out of six will be in-frame with the ORFs encoding T25, the library should contain at least 3264 (544 × 2 × 3) inserts. We analysed the content of the library (made in pKT25). It contained 10 000 clones, of which 70% were recombinant. The size of the inserted fragments was estimated by amplifying them from randomly chosen plasmids representing 0·5% of the population. Most inserts were approximately 200 bp long, thus similar to the size of the sonicated DNA fragments. Therefore, the library should harbour enough clones to cover the entire Sapc peptide sequence. This is the first time that this two-hybrid system has been used to screen for interactions with a generated random library. Prior to this study, it had only been used to assay for interactions with proteins or chosen peptides (for example, Jobling & Holmes, 2000; Gilmour et al., 2003; Karimova et al., 2001; Gropp et al., 2001).

Screening of the library for Sapc peptides that mediate the interaction with Sapc

We screened for peptides that interact with Sapc by expressing the library in the presence of T18-Sapc. Thirty-five clones yielded a positive signal, and were selected for further investigation. To confirm the specificity of the interactions, we checked that the expression of the peptide from each clone with T18 alone did not result in a detectable interaction. Seventeen of the 35 candidates were shown to be true positives. Plasmids from all positive clones were extracted, and the inserts were sequenced (Fig. 2). As expected, all sequenced fragments were in-frame with the ORFs encoding T25, the library should contain at least

<table>
<thead>
<tr>
<th>pT18</th>
<th>pT18-Sapc</th>
<th>pUT18-n</th>
<th>pUT18-m</th>
<th>pUT18-c</th>
</tr>
</thead>
<tbody>
<tr>
<td>pT25 white 55 white 81 white 50 white 45 white 38</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pT25-Sapc white 80 red 1250 red 1750 red 2083 red 1848</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pKT25-n white 62 red 1580 white 12 white 46 white 38</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pKT25-m white 58 red 875 white 24 red 1147 white 23</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pKT25-c white 65 red 1050 white 13 white 23 white 31</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We screened for peptides that interact with Sapc by expressing the library in the presence of T18-Sapc. Thirty-five clones yielded a positive signal, and were selected for further investigation. To confirm the specificity of the interactions, we checked that the expression of the peptide from each clone with T18 alone did not result in a detectable interaction. Seventeen of the 35 candidates were shown to be true positives. Plasmids from all positive clones were extracted, and the inserts were sequenced (Fig. 2). As expected, all sequenced fragments were in-frame with the ORF encoding T25. This confirms the validity of the screen. It was noteworthy that no small fragments were isolated. Indeed, the sequenced inserts were between 465 bp (155 aa) and 1812 bp [604 residues, i.e. the size of the complete sapc (Sapc) sequence] (Fig. 2). This is significantly higher than the mean fragment size of the library. The screening method is not biased towards fairly large fragments: small fragments have been shown to be efficient in promoting the adenylate cyclase domain interaction (35 aa for the leucine zipper; Karimova et al., 1998). The fact that we did not isolate any small fragments hampered our initial goal, which was, in view of mutagenesis analysis, to determine small interacting peptides. However,
it interestingly suggests that effective interactions require folded large polypeptides rather than specific sequence motifs.

Interestingly, all polypeptides starting after residue 200 ended at residue 604. This indicates that the C-terminal polypeptides yielded an interaction only if the protein was not truncated at its C-terminal end. Furthermore, most polypeptides were more than 300 aa long. This indicated that only large polypeptides fold to form a stable interaction with Sapc that is selected in this system.

Subdomain interactions

Analysis of the Sapc-interacting polypeptides showed that they are not clustered in a particular region of Sapc. The existence of three non-overlapping and quasi-contiguous polypeptides of between 155 and 210 aa (from clones 8, 9 and 13; Fig. 2), each interacting with Sapc, suggested that Sapc is organized in three ‘subdomains’. These were termed N, M and C for the N-terminal, the central and the C-terminal polypeptide, respectively (Fig. 2).

To check these interactions, we constructed the reciprocal two-hybrid system. The inserts from pKT25 encoding the N, M and C polypeptides were thus subcloned into pUT18, generating pUT18-n, pUT18-m and pUT18-c, respectively (Table 1). The polypeptides encoded by all these plasmids interacted with T25-Sapc. Indeed, all transformants were red (Table 2, row two). This confirms that the N, M and C polypeptides can interact with Sapc. It further suggests that their folding is independent of the nearby heterologous polypeptide, which is compatible with these being subdomains of the Sapc domain.

To study the polypeptide–Sapc interaction further, we analysed putative interactions between these polypeptides. We took advantage of having constructs with each polypeptide in both vectors to test polypeptide interactions. Due to the possible orientation bias, all nine combinations were tested by co-transformation of the E. coli cya strain (Table 2, rows three, four and five). Only polypeptide M was able to yield an interaction, which was an auto-interaction. The failure of the other polypeptides to yield any interaction, except with the complete Sapc domain, suggests that the observed interactions, including the M–M interaction, are significant. The unique M–M interaction suggests that a similar interaction could exist in the crystal S-layer.

Concluding remarks

In this study we have shown by two independent approaches that Sapc is the crystallization domain. We took advantage of our genetic system to further dissect the molecular interactions that lead to the assembly of the crystal. Proportionally very few peptides gave rise to a positive signal, indicative of selectivity. These were large polypeptides, showing that this two-hybrid system can accommodate such fragments. The requirement for large polypeptides to drive the Sapc interaction may be characteristic of S-layer protein assembly, as biochemical analyses carried out on other S-layer proteins showed that large polypeptides are required for crystallization and self-assembly (Jarosch et al., 2001; Sillanpää et al., 2000; Smit et al., 2002). Three-dimensional crystallography has never been successfully applied to complete S-layer proteins. To understand the structural role of the polypeptides defined herein, we will carry out three-dimensional crystallography on these Sapc fragments. The projection map of Sap can then be reappraised in view of the atomic-level determination of the subdomain structure (Couture-Tosi et al., 2002).

ACKNOWLEDGEMENTS

The authors thank G. Karimova and D. Ladant for the gift of strains and plasmids, and for the critical reading of the manuscript. We acknowledge E. Bastien for constructing the pUT18 recombinant plasmids. S. Mesnage, P. Goossens and J.-L. Ranck are thanked for the communication of unpublished results, and helpful discussions. Thanks to M. Mock, in whose laboratory this work was conducted, for her constant interest. T.M., T.C. and X.H. were funded by the Ministère de l’Education Nationale, de la Recherche et de la Technologie.

REFERENCES


