PROTEINS:
Structure, Function, and Bioinformatics

The crystal structure of *Clostridium perfringens* SleM, a muramidase involved in cortical hydrolysis during spore germination

Journal:	*PROTEINS: Structure, Function, and Bioinformatics*
Manuscript ID	Prot-00120-2016.R1
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Al-Riyami, Bahja; University of Cambridge, Department Chemical Engineering and Biotechnology Ustok, Fatma; University of Cambridge, Department of Haematology Stott, Katherine; University of Cambridge, Department Biochemistry Chirgadze, Dimitri; University of Cambridge, Crystallography and Biocomputing Unit Christie, Graham; University of Cambridge, Department Chemical Engineering and Biotechnology
Key Words:	cortex lytic enzyme, peptidoglycan lysin, spore, GH25 family

The crystal structure of *Clostridium perfringens* SleM, a muramidase involved in cortical hydrolysis during spore germination

Short title: SleM crystal structure

Bahja Al-Riyami[1], Fatma Işık Üstok[3], Katherine Stott[2], Dimitri Y. Chirgadze[2] and Graham Christie[1*]

[1]Institute of Biotechnology, Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge, United Kingdom

[2]Department of Biochemistry, University of Cambridge, Cambridge, United Kingdom

[3]Department of Haematology, Division of Structural Medicine and Thrombosis Research Unit, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, United Kingdom

*Correspondence to: Graham Christie, Institute of Biotechnology, Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge, United Kingdom, Tel (+44) 1223 334-166; E-mail: gc301@cam.ac.uk

Keywords: cortex lytic enzyme, peptidoglycan lysin, spore, GH25 family

Abstract

Clostridium perfringens spores employ two peptidoglycan lysins to degrade the spore cortex during germination. SleC initiates cortex hydrolysis to generate cortical fragments that are degraded further by the muramidase SleM. Here we present the crystal structure of the *C. perfringens* S40 SleM protein at 1.8 angstroms. SleM comprises an N-terminal catalytic domain that adopts an irregular α/β-barrel fold that is common to GH25 family lysozymes, plus a C-terminal fibronectin type III domain. The latter is involved in forming the SleM dimer that is evident in both the crystal structure and in solution. A truncated form of SleM that lacks the FnIII domain shows reduced activity against spore sacculi indicating that this domain may have a role in facilitating the position of substrate with respect to the enzyme's active site.

Introduction

Bacterial species belonging to the orders *Bacillales* and *Clostridiales* initiate sporulation in response to nutrient starvation [1]. The resultant endospores (spores) are equipped with several morphological and structural features that enable them to persist in the environment in a metabolically dormant state for many years. One such feature is the thick layer of cortical peptidoglycan (PG) that is deposited between the proteinaceous spore coat and the membrane-bound spore protoplast (core). The cortex is essential for the maintenance of the relatively dehydrated state of the spore core, which contributes to metabolic dormancy and spore heat resistance [2,3]. An essential step during spore germination concerns the depolymerisation of the spore cortex by germination specific PG lysins, since this is required to permit hydration of the protoplast to levels that are commensurate with the resumption of metabolism [4].

Spores of *Bacillus* species employ the semi-redundant SleB and CwlJ lytic transglycosylases to initiate cortex hydrolysis during spore germination [5,6], generating cortical fragments that are then degraded further by the N-acetyl glucosaminidase SleL [7-9]. In contrast, most members of the *Clostridiales* appear to rely upon the SleC lytic transglycosylase to initiate cortex hydrolysis [10-12], with the SleM muramidase acting upon the cortical fragments generated in a role that appears to be analogous to that of SleL in *Bacillus* spores [12,13].

Intriguingly, cortex lytic enzymes (CLEs) involved in the germination of both *Bacillus* and *Clostridium* spores are typically present in the dormant spore in mature forms. SleC is the exception, being present in the spore as an inactive zymogen that is subsequently cleaved and activated by Csp proteases during germination [11,14,15]. Regardless, the molecular mechanisms involved in the regulation of CLE catalytic activity in dormant and germinating spores, and the basis of CLE cortical substrate specificity, are poorly understood. Recent

efforts to address these questions have employed X-ray crystallography to reveal the three-dimensional structures of several proteins involved in cortex hydrolysis [16-19]. In addition to placing mechanistic studies of CLE activity and specificity on a solid structural footing, this information should enable a structure-led approach to the design of inhibitors or stimulants of spore germination, which may in turn facilitate the development of novel therapeutic and decontamination strategies.

In this study, X-ray crystallography was used to solve the crystal structure of the *Clostridium perfringens* S40 SleM CLE to 1.8 Å, which represents the first high resolution *Clostridiales* CLE structure to be determined. In common with *Bacillus* CLEs for which structures are available, the protein is modular, comprising a modified α/β-barrel catalytic domain, plus a C-terminal fibronectin type III (FnIII) domain not previously observed in spore CLEs.

Materials and Methods

Expression and purification of SleM

A DNA fragment encoding the full length *C. perfringens* S40 SleM protein (UniProt identifier O06496), codon optimised for expression in *E. coli*, was obtained from GeneArt Gene Synthesis (Paisley, UK). The entire *sleM* open reading frame (ORF), minus the stop codon, was subsequently amplified by PCR using primers (sequences available upon request) that included additional 5' nucleotides to facilitate ligation independent cloning. The *sleM* amplicon was then purified and cloned into the pBADcLIC *E. coli* expression vector, which is designed to add a cleavable His_{10} tag at the C-terminal of the protein to facilitate purification [20].

Protein expression was conducted using *E. coli* Top10 cells (Life Technologies Ltd., Paisley, UK), which were cultured in 2 L baffled flasks containing 500 ml LB medium

supplemented with 50 µg/ml carbenicillin at 37°C and 225 rpm. The temperature was reduced to 30°C when the optical density of the culture at 600 nm (OD_{600}) reached 0.6, upon which protein expression was induced by the addition of arabinose to a final concentration of 0.2 % (w/v). Protein expression continued for 5 h, and then the cells were harvested by centrifugation (8,000 g, for 10 min at 4°C), before washing the cellular pellets with buffer (50 mM Tris-HCl [pH 8.0], 100 mM NaCl) and storing at –80°C.

SleM purification comprised thawing and resuspension of the cellular pellet in 16 ml of ice-cold binding buffer (20 mM sodium phosphate, pH 7.4, 500 mM NaCl, 20 mM imidazole, 1 mM phenylmethylsulfonyl fluoride [PMSF]). Cell lysis was achieved by three passes through a One Shot Cell Disrupter (Constant Systems Ltd., Northampton, UK) operating at 20 x 10^3 lb/in^2. The cell lysate was centrifuged (15,000 g, for 20 min at 4°C), the supernatant passed through a 0.45 µM syringe filter, and then loaded on to a 1 ml Ni-Sepharose HisTrap HP column (GE Healthcare, Little Chalfont, UK) fitted to an AKTA Pure protein purification system (GE Healthcare), which had been pre-equilibrated with the same ice-cold buffer (minus PMSF). The protein was eluted in the same buffer containing 500 mM imidazole, and then buffer-exchanged and concentrated into 25 mM sodium phosphate, pH 7.0, 150 mM NaCl, 2 mM EDTA, 10% (v/v) glycerol, 1 mM DTT, using a 10 kDa MWCO Amicon centrifugal filter unit (Merck Millipore, Watford, UK). The C-terminal His_{10} affinity tag was subsequently removed by incubating overnight at 4°C with His_6-tagged TEV (S219V) protease [21] (1 µg TEV protease to every 100 µg SleM). The reaction mix was subject to a second round of Ni^{2+}-NTA affinity chromatography, using the same 1 ml HisTrap column equilibrated with ice-cold 20 mM sodium phosphate, pH 7.4, 500 mM NaCl. The SleM protein, now minus the affinity tag, was collected in the flow through fraction and then buffer exchanged into 20 mM Tris-HCl, pH 7.5 (buffer A), before loading onto a 1 ml Resource Q anion-exchange column (GE Healthcare) equilibrated with the same buffer at

room temperature. A salt gradient was applied using 20 mM Tris-HCl, pH 7.5, 1 M NaCl (buffer B), flow rate 4 ml min^{-1}, with SleM eluting in fractions that contained approximately 250 mM NaCl. SleM-containing fractions were pooled, concentrated by ultra-filtration, and subjected to gel filtration using a Superdex 75 column (GE Healthcare) equilibrated with 20 mM Tris-HCl, pH 7.0, 150 mM NaCl at room temperature. The purified protein, which contained vector-derived MGGGFA and ENLYFQ residues at the respective N- and C-termini, was desalted using a HiTrap desalting column (GE Healthcare) and concentrated to 9.5 mg/ml by ultra-filtration in 20 mM Tris-HCl, pH 7.0, 150 mM NaCl. The purified protein was aliquoted and stored at –80°C.

Crystallisation of SleM

Crystallisation trials were performed using the vapour diffusion sitting-drop technique in 96-well MRC 2-drop crystallisation plates (Swissci UK, Wokingham, UK), mixing 200 nL of the crystallisation screen conditions with 200 nL of protein solution (9.5 mg/ml) and setting this against 70 µL of reservoir using a crystallisation robot (Crystal Phoenix, Art Robbins Instruments, Inc.). Crystallisation trials were conducted at 19°C, using several commercial screens, with automated monitoring via a Rock Imager 1000 (Formulatrix, Inc.) imaging system. Initial hits were identified in PACT (condition B10) and PEGS I (condition G4) crystallisation screens (Qiagen, Manchester, UK) before optimising conditions in 24-well hanging-drop crystallisation plates (Hampton Research). Diffraction quality crystals for SleM were obtained from 3 µl drops containing a 1:1 mixture of 9.5 mg/ml protein and a crystallisation solution comprising 0.1 M MES buffer, pH 6.0, 0.25 M $MgCl_2$, 16% (w/v) PEG6000. Crystals appeared after a few days when incubated at 19°C, attaining maximum dimensions of approximately 0.2 mm x 0.5 mm x 0.5 mm after about 2 weeks.

Diffraction Data Collection and Processing

Fully grown crystals were cryo-protected by mounting in loops prior to immersion in a drop containing the crystallisation condition plus 26% (v/v) ethylene glycol for a few seconds, and then flash-frozen in liquid nitrogen. The X-ray diffraction dataset was collected using a copper rotating anode X-ray diffraction system (wavelength of 1.5418 Å) equipped with a confocal mirror monochromator, a kappa geometry goniometer, and Platinum 135 CCD detector (X8 PROTEUM, Bruker AXS, Ltd.) at a temperature of 100K provided by the COBRA Cryostream cryogenic cooling device (Oxford Cryosystems, Ltd.). The exposure time was set to 20 sec for a single phi-oscillation image of 1 degree, and the total of 480 oscillation images were collected in 3 different kappa geometry orientations. The dataset was indexed, scaled and merged using PROTEUM2 data processing software (Bruker AXS, Ltd). The crystal belongs to the monoclinic $P2_1$ space group with cell parameters a= 50.60 Å, b=85.85 Å, c=87.21 Å, $\alpha = \gamma = 90°$, and $\beta = 105.51°$, and diffracted to a maximum resolution of 1.8 Å. Analysis of the crystal solvent content using Matthews Coefficient indicated that two molecules of SleM are present in the crystallographic asymmetric unit. This composition results in about 50% of solvent content and Matthews Coefficient of 2.44. The crystallographic data collection statistics are shown in Table 1.

Crystal Structure Determination, Model Building and Refinement

The SleM crystal structure was solved by the Molecular Replacement (MR) method. The crystal structure of bacterial lysozyme (cellosyl) from *Streptomyces Coelicolor* (PDB-ID: 1JFX) was used as the MR search probe. The sequence identity between the search probe and SleM is 29% over 210 residues, whereas the crystallized protein contains 332 residues (including vector derived residues). All MR calculations were performed in PHASER, part of the PHENIX crystallographic software suite [22]. The positions of the two SleM molecules

within the asymmetric part of the unit cell were successfully identified. The translation factor Z-score for this solution was 44.8, indicating that an unambiguous solution has been found. The model obtained was subjected to several rounds of alternating manual rebuilding performed in the molecular graphics software suite *Coot* [23] and crystallographic refinement calculations in the PHENIX crystallographic software suite. Hydrogen atoms were added in their riding positions to the protein atoms but not to the water molecules. The R_{cryst} and R_{free} factors converged to the values of 15.1% and 18.2%, respectively. The final model has no amino acid residues in disallowed and more than 99% in favoured regions of the Ramachandran phi-psi plot. The crystallographic statistics and structural validation aspects are shown in Table 1. The final model of the SleM crystal structure contains residues 4-320 of molecule A, and residues 7-320 of molecule B in the asymmetric unit of the unit cell. The model also contains 1,140 water molecules, four MES molecules, and two Mg^{2+} ions. The metal ions were identified and placed into the electron density using PHENIX ion identification algorithms as implemented in the refinement protocols.

Construction of variant SleM proteins

The pBADcLIC-SleM plasmid described above served as a template to construct truncated (Q2-A229; M225-G321) and variant (D13N, D108N and E110Q) forms of SleM via PCR or employment of a QuikChange Lightning Site-Directed Mutagenesis kit (Agilent Technologies, Wokingham, UK) where appropriate. Similarly, the pBADc-LIC-GFP plasmid was employed to create full-length, catalytic (Q2-A229) and FnIII-domain (M225-G321) C-terminal GFP-fusion proteins. Plasmids with the correct substitutions were identified by DNA sequencing and then introduced by transformation to *E. coli* Top 10 cells. Native and variant SleM proteins were expressed and purified essentially as described above.

Enzyme assays

Enzymatic activity of the various forms of SleM was assessed by incubating purified enzymes with *B. subtilis* spore PG sacculi suspended at an optical density (A600 nm) of 0.5 in 50 mM Tris-HCl, pH 7.5, containing 150 mM NaCl. Sacculi were incubated with recombinant *B. megaterium* SleB [24] (0.1 μM) plus designated SleM variant proteins (1 μM) at 37°C for 60 min. Reactions were monitored by recording changes in absorbance (600 nm) using a PerkinElmer Envision-Xcite multilabel plate reader.

Analytical Ultracentrifugation

Sedimentation velocity experiments were conducted with an Optima XL-I (Beckman Coulter) centrifuge using an An60 Ti four-hole rotor. Standard double-sector Epon centrepieces equipped with sapphire windows contained 400 μL of SleM or SleM catalytic domain (Q2 – G217) at 2.0 mg/mL. Interference data were acquired at time intervals of 330 s and rotor speeds of 45 krpm (SleM) or 40 krpm (SleM catalytic domain), at a temperature of 20°C with systematic noise subtracted. The density and viscosity of the buffer and the partial specific volume of the protein were calculated using Sednterp [25]. Multi-component sedimentation coefficient distributions were obtained from 75 scans (SleM) or 200 scans (SleM catalytic domain; even-numbered scans only were used for the fit) by direct boundary modelling of the Lamm equation using Sedfit v.14.1 [26].

Accession codes

Atomic coordinates and structure factors for the SleM crystal structure have been deposited with the Protein Data Bank (PDB) under accession code 5JIP.

Results

SleM crystal structure

Crystals of *C. perfringens* SleM were obtained after heterologous expression of the protein in *E. coli* and subsequent purification to homogeneity. The orthologous *C. botulinum* protein also proved amenable to expression and purification from *E. coli* but failed to yield crystals. The SleM structure was solved by molecular replacement and refined to 1.8 Å, with an R_{cryst} of 15.1% and R_{free} of 18.2% (Table 1). The asymmetric unit contained two SleM molecules arranged as a dimer, plus 1,140 water molecules, four MES molecules, three Mg^{2+} ions and one Na^{+} ion. The final electron density map shows clear density for both protein molecules in their entirety except for disordered sections comprising the final six C-terminal residues (NGEFLG) and N- and C-terminal residues introduced as cloning artefacts. Similarly, electron density for the side chain of K289 in molecule A was not observed, hence only the C_β of the side chain has been modeled. All residues were in allowed regions of the Ramachandran plot. Superposition of both SleM protomers revealed that they are essentially identical (root mean square deviation [r.m.s.d] of 0.36 Å over 314 residues).

SleM comprises two structurally distinct domains (Figure 1). The N-terminal catalytic domain (Q2 – S220) is formed from a modified α/β-barrel fold (strictly $[\alpha/\beta]_5[\beta]_3$) that is characteristic of lysozymes belonging to the GH-25 family, whereas the C-terminal domain (M225 to N315) adopts a fibronectin III (FnIII)-type fold, comprising a seven-stranded β-sandwich. Searches conducted with the *Dali* [27] and *Fatcat* [28] servers failed to identify any proteins that share similarly placed α/β-barrel and FnIII domains. The α/β-barrel and FnIII domains are connected by a long loop (P218 – N234) that extends to β-1 of the FnIII domain, which is interrupted by a short 3_{10}-helix (I221 – L224). A magnesium ion is bound close to the 3_{10}-helix, in a solvent inaccessible location coordinated by backbone carbonyl groups from residues D222 and M225. Both residues are positioned at distances (2.0 Å and 2.2 Å

respectively) that are consistent with ligand-Mg^{2+} bonds [29]. Temperature factor values in this region of the protein are indicative of low intrinsic mobility, suggesting that the bound Mg^{2+} ion may have a role in reducing inter-domain flexibility. The latter may be of biological significance since SleM has been shown previously to have a strict requirement for divalent metal ions for activity [13].

FnIII domain (M225 to N315) and dimerisation

Fibronectin type III domains are extremely common in animal modular proteins, where they are often linked in tandem to form extracellular matrix proteins or as part of the ectodomains of receptors. They have also been identified in several bacterial chitinolytic enzymes, including *Serratia marcescens* ChiA, ChiC and *Bacillus circulans* A1, where they appear to facilitate substrate binding to the catalytic domain [30,31]. Comparative structural searches conducted with the *Dali* server identified the C-terminal located FnIII domain from the *C. perfringens* GH84C multi-modular N-acetylglucosaminidase as being the closest structural neighbour to the SleM FnIII domain (1.9 Å r.m.s.d. over 86 residues, Z score 12.6). The analogous domain from *B. circulans* chitinase A1 was also identified as a close structural match (2.1 Å r.m.s.d. over 84 residues, Z score 11.5).

In the case of SleM, the FnIII domain appears to provide an interface for protein dimerisation, which results in the burying of 2210 $Å^2$ of the two subunits. The dimer is formed by salt bridges between charged residues located on α4 of the α/β-barrel and charged residues located on the FnIII domain of the second molecule (D127:K275, D131:K275, E135:K241), and hydrogen bonds between predominantly FnIII-located residues (N255:N255, Y265:R138, T268:N258). Analytical ultracentrifugation was conducted to ascertain whether the SleM dimer present in the crystalline state is representative also of the protein's quaternary structure in solution, or whether the observed dimer is a consequence of

crystal packing in the presence of high protein concentration. The results of these analyses indicate that SleM is principally dimeric in solution, with sedimentation velocity [$c(s)$] distributions indicating that the protein exists predominantly as a single species in solution with a molecular weight of 73.4 kDa (the calculated protomer MW is 37.5 kDa) (Figure S1). However, a truncated version of the protein, comprising only the catalytic domain (Q2-A229) was shown to exist as a monomer in solution with a mass of 28.6 kDa (exactly matching the MW calculated from the sequence).

The truncated monomeric form of SleM retained the ability to digest cortical PG fragments generated by limiting concentrations of SleB in reaction mixtures, as evident by the reduction in absorbance (A600 nm) of suspensions of spore sacculi co-incubated with both proteins (Figure 2). However, the reduction in absorbance was significantly reduced compared to reactions containing full-length SleM (P<0.001 between 15 and 60 min), perhaps stemming from a reduced ability to recruit or bind substrate. Pull down assays conducted with GFP fusion proteins and purified spore sacculi, which indicate that only the full length SleM protein has discernible carbohydrate binding capacity (Figure S2), provide some evidence to support this hypothesis.

SleM catalytic domain (Q2 – S220)

The catalytic domain of SleM comprises an irregular α/β-barrel, in which only the first five β-strands are flanked by α-helices and β8, which is antiparallel with respect to the other strands, closes the barrel. The topology of the SleM catalytic domain is therefore consistent with other GH25 family lysozymes for which structures are available [32-37]. The most closely related structures, identified from the *Dali* server, included several bacteriophage-associated endolysins. These included the Psm lysin from phage phiSM101, which targets *Clostridium perfringens* (1.7 Å r.m.s.d. over 196 residues, Z score 26.5), PlyB from the *Bacillus anthracis*

BcpI phage (1.9 Å r.m.s.d. over 172 residues, Z score 20.4), Ctp1L from ϕCTP1, which targets *Clostridium tyrobutyricum* (2.0 Å r.m.s.d. over 183 residues, Z score 23.1) and Cpl-1 from *Streptococcus pneumonia* phage CP-1 (2.4 Å r.m.s.d. over 176 residues, Z score 18.7). Non-phage associated structural hits include a *Bacillus anthracis* GH25 lysozyme (2.1 Å r.m.s.d. over 189 residues, Z score 21.7) and the *Streptomyces coelicolor* cellosyl lysozyme that was used as the MR probe (1.8 Å r.m.s.d. over 187 residues, Z score 23.8). Comparable r.m.s.d values and high Z scores are indicative of significant structural similarity between SleM and the aforementioned proteins.

A structure-based multiple sequence alignment of the catalytic domains of SleM and related molecules is shown in Figure S3. From this, the carboxylate pair (D108 and E110), which share the D*X*E sequence motif common to GH25 enzymes, and which are proposed to catalyse cleavage of the substrate glycan chain via a "neighbouring group" mechanism [34], were readily identifiable. A third residue, D13, which is also conserved in the SleM sequence alignment, would most likely also be directly involved in catalysis if it proceeded via the classical "inverting" mechanism that has also been proposed for GH25 enzymes [38], although the actual mode has yet to be established. As well as being positionally conserved, *Coot*-facilitated overlays of the catalytic domains of SleM and related proteins revealed that all three carboxylate residues are spatially conserved within the active centres of the respective enzymes. Functional assays conducted with variant SleM proteins bearing individual substitutions at these locations (D13N, D108N and E110Q) revealed a loss of detectable enzyme activity against purified spore PG sacculi co-incubated with the SleB CLE (Figure 2).

In terms of substrate binding, SleM's α/β-barrel has a notable cleft that traverses the C-terminal face (~30 Å) of the catalytic domain, measuring approximately 7 Å wide by 9 Å deep. Two MES molecules were observed within this groove in the SleM crystal structure,

presumably as a result of the high concentration (0.1 M) of this non-substrate molecule present in the crystallisation buffer (Figure 1d). Superposition of SleM's α/β-barrel with the analogous domain from other GH25 enzymes reveals that residues that form the negatively charged pit towards the centre of the putative substrate binding cleft – around which the putative catalytic residues and one of the bound MES molecules are localised - are highly conserved both spatially and in the primary sequence alignment (Y68, F70, V106, W171, Y144, D213) (Figure S3). Similarly, as expected for a carbohydrate binding protein, aromatic residues feature heavily in the predicted substrate-binding groove (F17, Y68, F70, Y144, F148, F149), although sequence and spatial conservation with other GH25 enzymes is less evident here, which presumably reflects substrate specificity. By overlaying the coordinates of Cpl-1 with bound ligand (tetrasaccharide pentapeptide) (PDB-ID: 2J8G) with SleM's α/β-barrel we can infer that the entrance to the SleM substrate binding site is formed by L147 and F148 on one side of the cleft and S180 and N181 on the other. While the pentapeptide moiety of the overlaid ligand cannot be accommodated within the SleM binding site - which isn't surprising since this moiety is not present in spore PG - we can use the location of the NAG residues at the +1 and +3 positions to identify the probable binding pocket for the muramic-acid lactam (MAL) moiety (Figure 3). The latter is unique to spore PG and serves to differentiate between cortical and germ cell wall PG for CLE activity during spore germination [39,40]. Hence, if we consider tetrasaccharide tetrapeptide (or alanine) as a candidate SleM substrate, which has a glycan backbone comprising NAM-NAG-MAL-NAG, then catalysis would occur between NAM and NAG located at the -1 and +1 locations respectively. Accordingly, MAL would be bound at the +2 location in this model, accommodated in the pocket formed by T113, Y144, G146, F148, F149 and catalytic E110.

Discussion

The SleM crystal structure represents the first *Clostridium* CLE structure to be solved and the third CLE – after *B. cereus* and *B. megaterium* SleL [19] - for which complete structures have been determined. Crystal structures for the catalytic domains of *B. cereus* [17] and *B. anthracis* SleB [16], and *B. subtilis* YdhD (PDB-ID: 3CZ8), have also been solved by X-ray crystallography, meaning that of the major CLEs, three dimensional structural information is lacking only for CwlJ and SleC. With the exception of the aforementioned CwlJ, which is capable of initiating cortex depolymerisation in *Bacillus* spores, and may also have a role in germination of some *Clostridial* species, all CLEs adopt a modular structure. Typically, this entails a single catalytic domain plus one (e.g. SleB) or two (e.g. SleL) substrate binding domain(s). The SleM structure revealed the presence of a C-terminal located FnIII domain, which wasn't identified from primary sequence analyses. The precise role of this domain hasn't been clarified, although evidently FnIII-located residues participate in formation of the SleM dimer that is observed in the crystal structure and in solution. A role in dimer formation has been demonstrated recently for FnIII-type domains, namely in the irisin myokine associated with the human FNDC5 receptor [41]. However, a truncated variant of SleM, comprising only the α/β-barrel catalytic domain, was shown to be monomeric in solution while displaying reduced activity against pre-digested cortical PG. Hence, neither the FnIII domain nor dimer formation is essential to SleM catalytic function, although these data perhaps indicate a role in recruitment or correct orientation of PG substrate relative to the catalytic domain. Similar roles for FnIII-type domains have been characterised or alluded to previously for chitinolytic and PG lysins [30,31]. Indeed, the SleM FnIII domain has several surface-exposed aromatic residues (Y265, F283 and Y284 on one side of the domain, with Y235 and F243 on the other), which have been shown to be of functional importance in FnIII-carbohydrate interactions in *Serratia marcescens* ChiB [30]. If this applies to the SleM

FnIII domain then it is intriguing that at least three different folds are employed by CLE substrate-binding domains to interact with the structurally unique spore cortical PG.

Unfortunately, none of the CLE crystal structures solved to date contain PG ligands bound to the catalytic or substrate-binding domains. As such, precise information on residues involved in substrate specificity – in particular, selective binding of PG containing the MAL moiety – has yet to be ascertained. It seems also that appropriately liganded structures will be required to yield insight to the structural basis for differentiation between CLEs that can cleave intact spore PG (e.g. SleB, SleC) and those that appear only to be active against pre-digested or fragmented cortical PG. All three cortical fragment lytic enzymes for which three-dimensional structural information is available (SleM, SleL and YdhD) share α/β-barrel-type folds, whereas SleB, which can lyse intact PG, adopts a fold reminiscent of family-1 transglycosylases, albeit with unique topology. The latter results in a catalytic domain with a wide accessible substrate binding cleft compared to the convoluted α/β-barrel enzyme clefts, and perhaps this is essential in accommodating cortical PG as arranged in the dormant spore sacculus. Future studies of enzyme-substrate complexes should help to clarify this situation, and will contribute to on-going work aimed at developing novel therapeutics and agents for improved control of spores.

Acknowledgements

Bahja Al-Riyami is the recipient of a scholarship from the Sultanate of Oman government. Crystallographic experiments were performed in the Crystallographic X-ray Facility, and analytical centrifugation experiments within the Biophysical Facility, both at the Department of Biochemistry, University of Cambridge. The authors have no conflicts of interest to declare.

References

1. Paredes-Sabja D, Setlow P, Sarker MR. Germination of spores of *Bacillales* and *Clostridiales* species: mechanisms and proteins involved. Trends Microbiol 2011;19(2):85-94.

2. Gerhardt P, Marquis RE. Spore thermoresistance mechanisms. In: Smith I, Slepecky RA, Setlow P, editors. Regulation of prokaryotic development - structural and functional analysis of bacterial sporulation and germination. Washington, DC: American Society for Microbiology; 1989. p 43-63.

3. Imae Y, Strominger JL. Relationship between cortex content and properties of *Bacillus sphaericus* spores. Journal of bacteriology 1976;126(2):907-913.

4. Setlow P. Germination of spores of *Bacillus* species: what we know and do not know. Journal of bacteriology 2014;196(7):1297-1305.

5. Chirakkal H, O'Rourke M, Atrih A, Foster SJ, Moir A. Analysis of spore cortex lytic enzymes and related proteins in *Bacillus subtilis* endospore germination. Microbiology (Reading, England) 2002;148(Pt 8):2383-2392.

6. Ishikawa S, Yamane K, Sekiguchi J. Regulation and characterization of a newly deduced cell wall hydrolase gene (*cwlJ*) which affects germination of *Bacillus subtilis* spores. Journal of bacteriology 1998;180(6):1375-1380.

7. Chen Y, Fukuoka S, Makino S. A novel spore peptidoglycan hydrolase of *Bacillus cereus*: biochemical characterization and nucleotide sequence of the corresponding gene, *sleL*. Journal of bacteriology 2000;182(6):1499-1506.

8. Lambert EA, Popham DL. The *Bacillus anthracis* SleL (YaaH) protein is an N-acetylglucosaminidase involved in spore cortex depolymerization. Journal of bacteriology 2008;190(23):7601-7607.

9. Ustok FI, Packman LC, Lowe CR, Christie G. Spore germination mediated by *Bacillus megaterium* QM B1551 SleL and YpeB. Journal of bacteriology 2014;196(5):1045-1054.

10. Burns DA, Heap JT, Minton NP. SleC is essential for germination of *Clostridium difficile* spores in nutrient-rich medium supplemented with the bile salt taurocholate. Journal of bacteriology 2010;192(3):657-664.

11. Miyata S, Moriyama R, Miyahara N, Makino S. A gene (sleC) encoding a spore-cortex-lytic enzyme from *Clostridium perfringens* S40 spores; cloning, sequence analysis and molecular characterization. Microbiology (Reading, England) 1995;141 (Pt 10):2643-2650.

12. Paredes-Sabja D, Setlow P, Sarker MR. SleC is essential for cortex peptidoglycan hydrolysis during germination of spores of the pathogenic bacterium *Clostridium perfringens*. Journal of bacteriology 2009;191(8):2711-2720.

13. Chen Y, Miyata S, Makino S, Moriyama R. Molecular characterization of a germination-specific muramidase from *Clostridium perfringens* S40 spores and nucleotide sequence of the corresponding gene. Journal of bacteriology 1997;179(10):3181-3187.

14. Shimamoto S, Moriyama R, Sugimoto K, Miyata S, Makino S. Partial characterization of an enzyme fraction with protease activity which converts the spore peptidoglycan hydrolase (SleC) precursor to an active enzyme during germination of *Clostridium perfringens* S40 spores and analysis of a gene cluster involved in the activity. Journal of bacteriology 2001;183(12):3742-3751.

15. Adams CM, Eckenroth BE, Putnam EE, Doublie S, Shen A. Structural and functional analysis of the CspB protease required for *Clostridium* spore germination. PLoS pathogens 2013;9(2):e1003165.

16. Jing X, Robinson HR, Heffron JD, Popham DL, Schubot FD. The catalytic domain of the germination-specific lytic transglycosylase SleB from *Bacillus anthracis* displays a unique active site topology. Proteins 2012;80(10):2469-2475.

17. Li Y, Jin K, Setlow B, Setlow P, Hao B. Crystal structure of the catalytic domain of the *Bacillus cereus* SleB protein, important in cortex peptidoglycan degradation during spore germination. Journal of bacteriology 2012;194(17):4537-4545.

18. Ustok FI, Chirgadze DY, Christie G. Crystal structure of the PepSY-containing domain of the YpeB protein involved in germination of *bacillus* spores. Proteins 2015;83(10):1914-1921.

19. Ustok FI, Chirgadze DY, Christie G. Structural and functional analysis of SleL, a peptidoglycan lysin involved in germination of *Bacillus* spores. Proteins 2015;83(10):1787-1799.

20. Geertsma ER, Poolman B. High-throughput cloning and expression in recalcitrant bacteria. Nature methods 2007;4(9):705-707.

21. Kapust RB, Tözsér J, Fox JD, Anderson DE, Cherry S, Copeland TD, Waugh DS. Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. Protein Engineering 2001;14(12):993-1000.

22. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta crystallographica Section D, Biological crystallography 2010;66(Pt 2):213-221.

23. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta crystallographica Section D, Biological crystallography 2010;66(Pt 4):486-501.

24. Li Y, Butzin XY, Davis A, Setlow B, Korza G, Ustok FI, Christie G, Setlow P, Hao B. Activity and regulation of various forms of CwlJ, SleB, and YpeB proteins in degrading cortex peptidoglycan of spores of *Bacillus* species in vitro and during spore germination. Journal of bacteriology 2013;195(11):2530-2540.

25. Laue TM, Shah BD, Ridgeway TM, Pelletier SL. Computer-aided interpretation of analytical sedimentation data for proteins. In: Harding SE, Rowe AJ, Horton J, editors. Analytical ultracentrifugation in biochemistry and polymer science. Cambridge, UK: The Royal Society of Chemistry; 1992. p 90-125.

26. Schuck P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. Biophysical Journal 2000;78(3):1606-1619.

27. Holm L, Kaariainen S, Rosenstrom P, Schenkel A. Searching protein structure databases with DaliLite v.3. Bioinformatics (Oxford, England) 2008;24(23):2780-2781.

28. Ye Y, Godzik A. FATCAT: a web server for flexible structure comparison and structure similarity searching. Nucleic acids research 2004;32(Web Server issue):W582-585.

29. Harding MM. Small revisions to predicted distances around metal sites in proteins. Acta crystallographica Section D, Biological crystallography 2006;62(Pt 6):678-682.

30. Katouno F, Taguchi M, Sakurai K, Uchiyama T, Nikaidou N, Nonaka T, Sugiyama J, Watanabe T. Importance of exposed aromatic residues in chitinase B from *Serratia marcescens* 2170 for crystalline chitin hydrolysis. Journal of biochemistry 2004;136(2):163-168.

31. Watanabe T, Ito Y, Yamada T, Hashimoto M, Sekine S, Tanaka H. The roles of the C-terminal domain and type III domains of chitinase A1 from *Bacillus circulans* WL-12 in chitin degradation. Journal of bacteriology 1994;176(15):4465-4472.

32. Dunne M, Leicht S, Krichel B, Mertens HD. Crystal Structure of the CTP1L Endolysin Reveals How Its Activity Is Regulated by a Secondary Translation Product. 2016;291(10):4882-4893.

33. Korczynska JE, Danielsen S, Schagerlof U, Turkenburg JP, Davies GJ, Wilson KS, Taylor EJ. The structure of a family GH25 lysozyme from *Aspergillus fumigatus*. Acta crystallographica Section F, Structural biology and crystallization communications 2010;66(Pt 9):973-977.

34. Martinez-Fleites C, Korczynska JE, Davies GJ, Cope MJ, Turkenburg JP, Taylor EJ. The crystal structure of a family GH25 lysozyme from *Bacillus anthracis* implies a neighboring-group catalytic mechanism with retention of anomeric configuration. Carbohydrate research 2009;344(13):1753-1757.

35. Porter CJ, Schuch R, Pelzek AJ, Buckle AM, McGowan S, Wilce MC, Rossjohn J, Russell R, Nelson D, Fischetti VA, Whisstock JC. The 1.6 A crystal structure of the catalytic domain of PlyB, a bacteriophage lysin active against *Bacillus anthracis*. Journal of molecular biology 2007;366(2):540-550.

36. Rau A, Hogg T, Marquardt R, Hilgenfeld R. A new lysozyme fold. Crystal structure of the muramidase from *Streptomyces coelicolor* at 1.65 A resolution. The Journal of biological chemistry 2001;276(34):31994-31999.

37. Tamai E, Yoshida H, Sekiya H, Nariya H, Miyata S, Okabe A, Kuwahara T, Maki J, Kamitori S. X-ray structure of a novel endolysin encoded by episomal phage phiSM101 of *Clostridium perfringens*. Molecular microbiology 2014;92(2):326-337.

38. Hermoso JA, Monterroso B, Albert A, Galan B, Ahrazem O, Garcia P, Martinez-Ripoll M, Garcia JL, Menendez M. Structural basis for selective recognition of

pneumococcal cell wall by modular endolysin from phage Cp-1. Structure (London, England : 1993) 2003;11(10):1239-1249.

39. Popham DL, Helin J, Costello CE, Setlow P. Muramic lactam in peptidoglycan of *Bacillus subtilis* spores is required for spore outgrowth but not for spore dehydration or heat resistance. Proceedings of the National Academy of Sciences of the United States of America 1996;93(26):15405-15410.

40. Warth AD, Strominger JL. Structure of the peptidoglycan of bacterial spores: occurrence of the lactam of muramic acid. Proceedings of the National Academy of Sciences of the United States of America 1969;64(2):528-535.

41. Schumacher MA, Chinnam N, Ohashi T, Shah RS, Erickson HP. The structure of irisin reveals a novel intersubunit beta-sheet fibronectin type III (FNIII) dimer: implications for receptor activation. The Journal of biological chemistry 2013;288(47):33738-33744.

42. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. Proceedings of the National Academy of Sciences of the United States of America 2001;98(18):10037-10041.

Figure Legends

Figure 1 Crystal structure of *Clostridium perfringens* SleM. (a) Cartoon representation of the SleM dimer, rotated by 90° with respect to (a) to show (b) top, and (c) side views of the molecules. The C-terminal face of the catalytic α/β-barrel as depicted in (c), is shown in surface representation in (d), which also shows the location of two MES molecules that were bound in the enzyme's substrate binding groove.

Figure 2 Purified *Bacillus subtilis* spore sacculi incubated with variant *C. perfringens* SleM proteins. Spore sacculi were resuspended at an OD_{600} ~0.5 in 50 mM Tris-HCl (pH 7.5) containing 150 mM NaCl, and then co-incubated at 37°C with 0.1 µM *B. megaterium* SleB plus 1 µM of full length SleM (●), $SleM^{cat}$ Q2-A229 (▲), SleM D13N (◊), SleM D108N (×) or SleM E110Q (□). Control reactions individually containing SleM (■) or SleB (o) are also shown. Hydrolysis of sacculi was measured by following changes to the optical density (A600) of the suspension as described in the Materials and Methods. Data shown are the mean ± standard deviation of three independent assays. For clarity, error bars are not shown for reactions containing SleB alone, and SleM D13N, D108N and E110Q variant proteins. Standard deviations in these cases were <15% of mean values.

Figure 3 Electrostatic surface of the SleM catalytic barrel as calculated by APBS [42], ranging from –3 kT/e in red (most negative) to +3 kT/e in blue (most positive), superposed with a peptidoglycan analogue from the Cpl-1-ligand complex (PDB-ID: 2J8G) [the catalytic barrels from both enzymes overlay with an r.m.s.d. of 2.4 Å over 176 residues, Z score 18.7]. The spore-specific muramic-acid lactam (MAL) moiety is predicted to be located in the +2 position instead of NAM as shown in the figure, with cleavage of the substrate taking place between NAM and NAG located at the respective -1 (not shown) and +1 subsites. Catalytic

D13, D108 and E110, which surround the electronegative centre of the enzyme, are labelled and shown in stick representation.

Figure S1 Analytical ultracentrifugation sedimentation velocity data. The residuals are from the fit with the continuous $c(s)$ distribution model. (**A**) Component sedimentation coefficient distribution for SleM at 2 mg/mL showing populations of monomeric and dimeric species fitting to a uniform frictional ratio of $F_{k,w} = 1.3$. The r.m.s.d. was 0.02. (**B**) Distribution for SleM (2-217) also at 2 mg/mL contains monomeric species only ($s^{o}_{20,w} = 2.68$). The fitted frictional ratio and mass were $F_{k,w} = 1.3$ and 28.6 kDa, respectively. The r.m.s.d. was 0.01.

Figure S2 Pulldown assay conducted with full length and truncated SleM-GFP proteins against spore sacculi. Essentially, 100 µM of each purified protein was incubated with 30 µL of purified *B. subtilis* spore sacculi (OD~50 at 600 nm) in 20 mM sodium phosphate buffer (pH 7.0), and the samples (200 µL) incubated with gentle agitation at 4°C for 2 hours. Samples were then subjected to centrifugation at 15,000 g for 10 minutes before being analysed on a FastGene blue light LED illuminator (Geneflow Ltd., Lichfield, UK). Strong fluorescence was associated with the pelleted sacculi in only the full length SleM-GFP sample, indicating that both catalytic and FnIII domains are required for efficient binding to spore peptidoglycan.

Figure S3 Clustal Omega sequence alignment of the catalytic α/β-barrel domains from a selection of family GH25 lysozymes for which three-dimensional structural information is known. Asterisks denote conserved residues. Residues that form the negatively charged pit that represents the active site of the enzyme, which are highly conserved both spatially and in the primary sequence alignment with other GH25 family enzymes, are boxed in red. These

include residues implicated directly in catalysis (D13, D108 and E110), which are boxed in blue.

Table 1. Crystallographic X-ray data collection, phasing and refinement statistics

	Native Dataset
Data collection	
Radiation Source	In-house, Copper Rotating Anode
Wavelength (Å)	1.5418
Space group	$P2_1$
Cell dimensions:	
a, *b*, *c* (Å)	50.60 85.85 87.21
α, β, γ (°)	90.0 105.51 90.0
Resolution (Å)	42.40 – 1.80 (1.90 – 1.80)[a]
R_{sym} [b] (%)	7.4 (35.7)
$\langle I / \sigma(I) \rangle$	24.5 (5.0)
Completeness (%)	99.6 (98.2)
Redundancy	3.9 (2.9)
Number of unique reflections	66,317 (9,699)
Refinement	
Resolution (Å)	42.40 – 1.80
Number of reflections used:	
Total	66,249
R_{free} set	1,990
R_{cryst}[c] / R_{free}[d] (%)	14.9 / 18.3
Solvent content, %	49.7
Number of protein molecules in asymmetric unit	2
Number of non-hydrogen of atoms in asymmetric unit:	
Protein atoms	5,052
Mg^{2+} ion atoms	3
Na^{+} ion atoms	1
MES atoms	48
Water molecules	1,144
B-factor, (Å^2):	
Average	17.0
Wilson	14.1
Ramachandran plot analysis, number of residues in:	
Favoured regions, %	99.36
Allowed regions, %	0.64
Disallowed regions, %	0
R.m.s. deviations:	
Bond lengths (Å)	0.011
Bond angles (°)	1.300

[a] The statistics shown in parentheses are for the highest-resolution shell.
[b] $R_{sym} = \Sigma_i \, | I_i(hkl) - I_{mean}(hkl) | \, / \, \Sigma_{hkl} \, \Sigma_i \, I_i(hkl)$
[c] $R_{cryst} = \Sigma_{hkl} \, ||F_{obs}(hkl)| - |F_{calc}(hkl)|| / \Sigma_{hkl} \, |F_{obs}(hkl)|$
[d] R_{free} is the same as R_{cryst} for a random subset not included in the refinement of about 10% of total reflection.

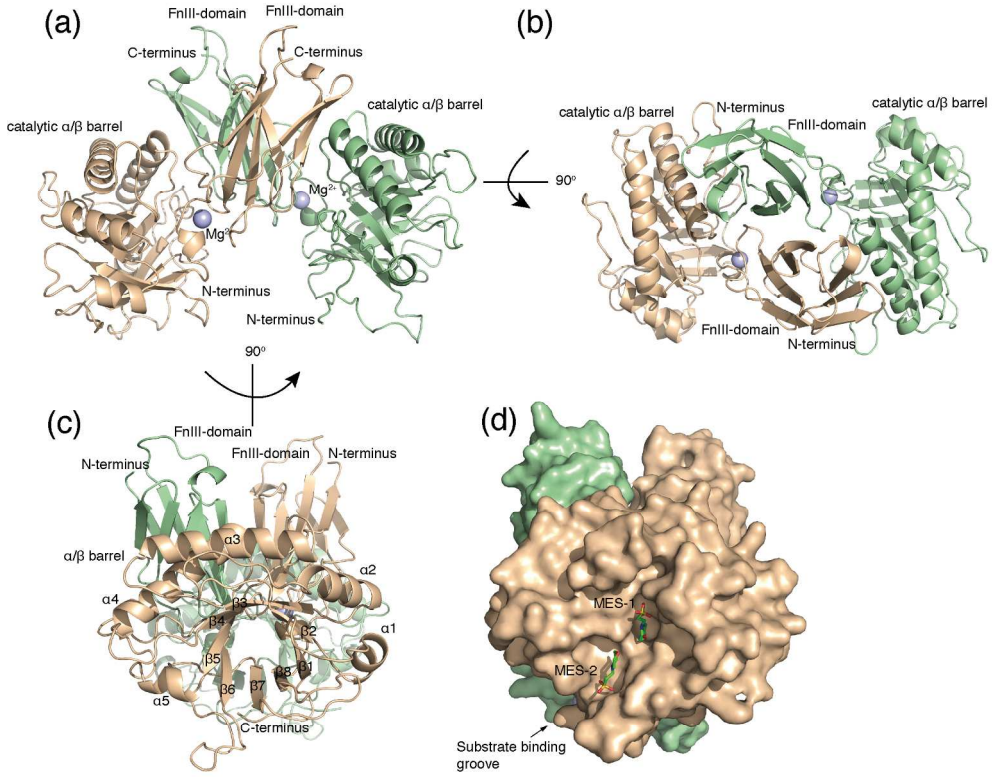

Figure 1 Crystal structure of *Clostridium perfringens* SleM. (a) Cartoon representation of the SleM dimer, rotated by 90° with respect to (a) to show (b) top, and (c) side views of the molecules. The C-terminal face of the catalytic α/β-barrel as depicted in (c), is shown in surface representation in (d), which also shows the location of two MES molecules that were bound in the enzyme's substrate binding groove.

Figure 1
201x172mm (300 x 300 DPI)

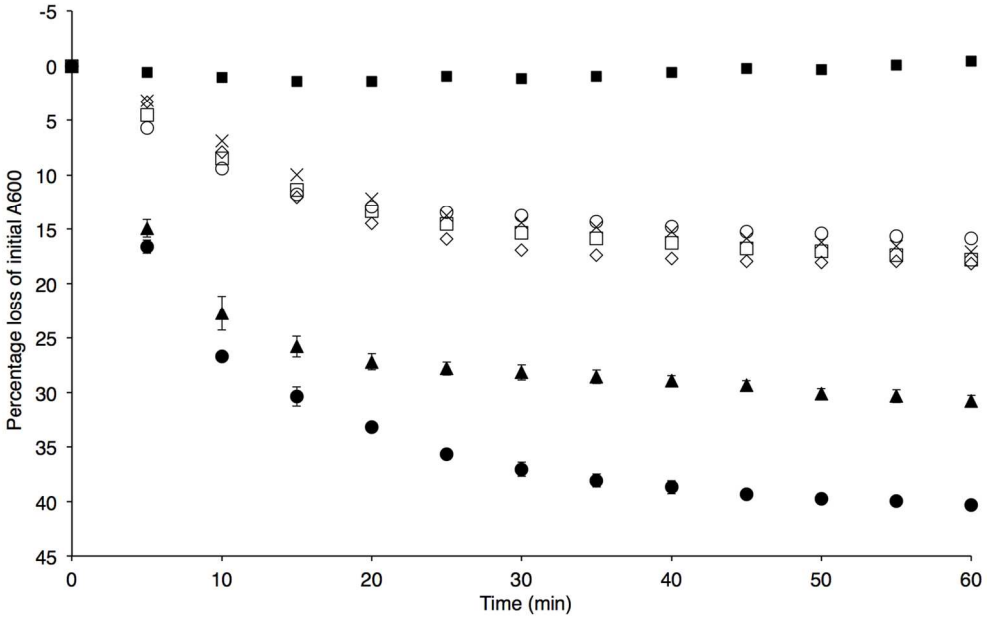

Figure 2 Purified *Bacillus subtilis* spore sacculi incubated with variant *C. perfringens* SleM proteins. Spore sacculi were resuspended at an OD600 ~0.5 in 50 mM Tris-HCl (pH 7.5) containing 150 mM NaCl, and then co-incubated at 37°C with 0.1 μM *B. megaterium* SleB plus 1 μM of full length SleM (●), SleMcat Q2-A229 (▲), SleM D13N (◇), SleM D108N (×) or SleM E110Q (□). Control reactions individually containing SleM (■) or SleB (o) are also shown. Hydrolysis of sacculi was measured by following changes to the optical density (A600) of the suspension as described in the Materials and Methods. Data shown are the mean ± standard deviation of three independent assays. For clarity, error bars are not shown for reactions containing SleB alone, and SleM D13N, D108N and E110Q variant proteins. Standard deviations in these cases were <15% of mean values.

Figure 2
140x86mm (300 x 300 DPI)

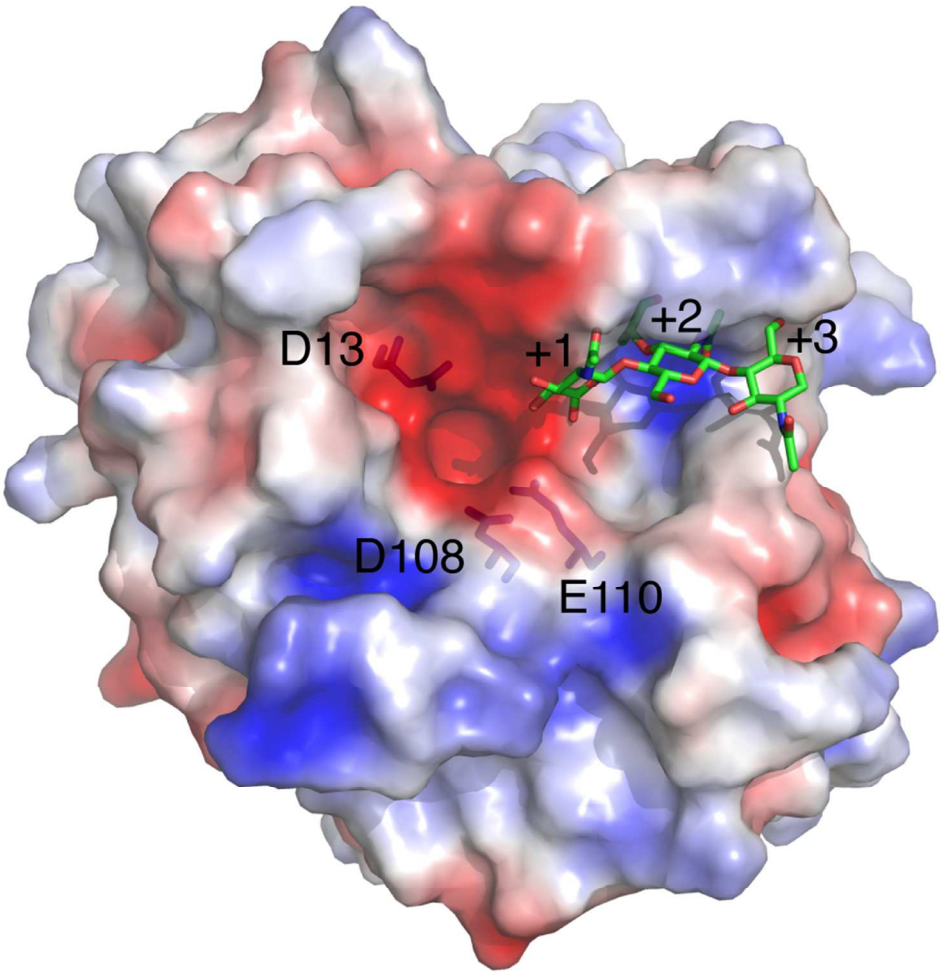

Figure 3 Electrostatic surface of the SleM catalytic barrel as calculated by APBS [42], ranging from –3 kT/e in red (most negative) to +3 kT/e in blue (most positive), superposed with a peptidoglycan analogue from the Cpl-1-ligand complex (PDB-ID: 2J8G) [the catalytic barrels from both enzymes overlay with an r.m.s.d. of 2.4 Å over 176 residues, Z score 18.7]. The spore-specific muramic-acid lactam (MAL) moiety is predicted to be located in the +2 position instead of NAM as shown in the figure, with cleavage of the substrate taking place between NAM and NAG located at the respective -1 (not shown) and +1 subsites. Catalytic D13, D108 and E110, which surround the electronegative centre of the enzyme, are labelled and shown in stick representation.

Figure 3
90x90mm (300 x 300 DPI)