



PROTEINS:  
Structure, Function, and Bioinformatics

## Crystal structure of the PepSY-containing domain of the YpeB protein involved in germination of Bacillus spores

Journal:	<i>PROTEINS: Structure, Function, and Bioinformatics</i>
Manuscript ID:	Prot-00202-2015.R1
Wiley - Manuscript type:	Structure Note
Date Submitted by the Author:	n/a
Complete List of Authors:	Christie, Graham; University of Cambridge, Department Chemical Engineering and Biotechnology Ustok, Fatma; University of Cambridge, Department of Haematology Chirgadze, Dimitri; University of Cambridge, Department of Biochemistry
Key Words:	Bacillus, Spore, Germination, Peptidoglycan, Hydrolysis, Inhibition

SCHOLARONE™  
Manuscripts

Review

1  
2  
3 **Crystal structure of the PepSY-containing domain of the YpeB protein involved in**  
4  
5 **germination of *Bacillus* spores**  
6  
7

8  
9  
10 Short title: *Bacillus megaterium* YpeB-C crystal structure  
11

12  
13  
14 Fatma Işık Üstok<sup>1,3</sup>, Dimitri Y. Chirgadze<sup>2</sup>, and Graham Christie<sup>1\*</sup>  
15  
16

17  
18 <sup>1</sup>Institute of Biotechnology, Department of Chemical Engineering and Biotechnology,  
19 University of Cambridge, Cambridge, United Kingdom  
20

21  
22 <sup>2</sup>Department of Biochemistry, University of Cambridge, Cambridge, United Kingdom  
23

24  
25 <sup>3</sup>Current Address: Department of Haematology, Division of Structural Medicine  
26 and Thrombosis Research Unit, Cambridge Institute for Medical Research, University of  
27 Cambridge, Cambridge, United Kingdom  
28  
29  
30

31  
32  
33  
34  
35  
36  
37  
38 \*Correspondence to: Graham Christie, Institute of Biotechnology, Department of Chemical  
39 Engineering and Biotechnology, University of Cambridge, Cambridge, United Kingdom,  
40

41  
42 Tel (+44) 1223 334-166; E-mail: gc301@cam.ac.uk  
43  
44  
45

46  
47 **Keywords:** cortex peptidoglycan, cortex lytic enzyme; SleB; CwlJ; SleL; inhibitory protein  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Abstract**

The crystal structure of the C-terminal domain of the *Bacillus megaterium* YpeB protein has been solved by X-ray crystallography to 1.80 Å resolution. The full-length protein is essential in stabilising the SleB cortex lytic enzyme in *Bacillus* spores, and may have a role in regulating SleB activity during spore germination. The YpeB-C crystal structure comprises three tandemly repeated PepSY domains, which are aligned to form an extended laterally compressed molecule. A predominantly positively charged region located in the second PepSY domain may provide a site for protein interactions that are important in stabilising SleB and YpeB within the spore.

## Introduction

Bacterial cells of the genera *Bacillus* and *Clostridium* initiate the process of sporulation in response to nutrient starvation. The resultant endospores (or spores) have several unique morphological and structural features that result in metabolic dormancy and an ability to persist in the environment for perhaps millennia<sup>1</sup>. In order to re-initiate vegetative growth and metabolism, the spore has to undergo the process of germination, which is triggered typically by the presence of defined nutrient molecules in the spore environment<sup>2</sup>.

A major event in germination concerns the enzymatic degradation of the thick layer of structurally distinct peptidoglycan, or cortex, that surrounds the spore protoplast. A limited number of cortex-lytic enzymes (CLEs) are responsible for conducting cortical depolymerisation during germination. Spores of *Bacillus* species, and a few *Clostridia*, require functional SleB or CwlJ to initiate this process, with additional enzymes, including SleL and perhaps YdhD, having roles in further degrading large cortical fragments generated by SleB and CwlJ activity<sup>3</sup>.

*Bacillus* CLEs are present in the spore in a mature but inactive form, and as yet, little is known of the mechanisms that limit their activity to a defined window within the germination process. One hypothesis is that the SleB protein is somehow held in an inactive state during dormancy by an interacting partner protein, namely YpeB, and that disruption of this interaction during germination permits SleB activity. Several lines of evidence support this hypothesis, namely (i) YpeB and SleB display a reciprocal relationship concerning their presence in the spore i.e. SleB is missing in YpeB null mutant spores<sup>4</sup>, and vice versa;<sup>5,6</sup> (ii) YpeB has been shown to be proteolytically cleaved during germination, perhaps breaking the interaction with SleB and permitting activity of the latter<sup>4,7</sup>; (iii) bioinformatic analyses predict that the C-terminal domain of the protein contains 2 – 3 tandem repeats of the PepSY motif (Pfam accession PF03413) that has been shown to inhibit peptidase activity in M4 and

1  
2  
3 M36 metallopeptidases<sup>8</sup>; and (iv) *in vitro* assays conducted with recombinant versions of  
4  
5 SleB co-incubated with either full-length YpeB or defined N-terminal and C-terminal domain  
6  
7 fragments demonstrate inhibition of SleB activity, to various degrees, in the presence of the  
8  
9 variant YpeB proteins<sup>5</sup>. Despite the above, efforts to ascertain whether SleB and YpeB  
10  
11 physically interact have as yet delivered negative results<sup>5,6</sup>.  
12  
13

14 In order to gain insight to the molecular mechanisms that regulate cortical  
15  
16 depolymerisation during spore germination, high-resolution structures for a number of spore  
17  
18 CLEs have recently been solved using X-ray crystallographic methods<sup>9,10</sup>. In a similar vein,  
19  
20 we present here the crystal structure of the C-terminal domain of the YpeB protein from *B.*  
21  
22 *megaterium*.  
23  
24

## 25 26 27 **Materials and Methods**

### 28 29 *Expression and purification of YpeB-C*

30  
31 A DNA fragment encoding the predicted C-terminal domain of the *B. megaterium* QM  
32  
33 B1551 YpeB protein (UniProt accession D5DRI0), comprising codons 216 – 449, was  
34  
35 amplified by PCR using purified genomic DNA as template. The following primer pair, both  
36  
37 of which include additional nucleotides at the 5' ends to facilitate ligation independent  
38  
39 cloning, were used in the PCR reaction:  
40  
41

42  
43 5'-ATGGTTGTTGGATTTGCTGCTTTTCATCAATTAAGGTAGAGAA-3' and 5'-  
44  
45 TTGGAAGTATAAATTTCCACTTCATTATATAAAGGTTCTGAGTT-3'. The PCR  
46  
47 amplicon was purified and cloned into plasmid pBADcLIC *E. coli* expression plasmid, which  
48  
49 is designed to create C-terminal His<sub>10</sub> fusion proteins. Protein expression was conducted  
50  
51 using *E. coli* Top10 cells (Life Technologies), which were cultured in LB medium containing  
52  
53 50 µg/ml carbenicillin at 37°C, 225 rpm, until the optical density at 660 nm (OD<sub>660</sub>) reached  
54  
55 0.6, when the temperature was reduced to 30°C and protein expression induced by the  
56  
57  
58  
59  
60

1  
2  
3 addition of arabinose to a final concentration of 0.2 % (w/v). Protein expression continued  
4  
5 for 6 h, before the cells were harvested by centrifugation (8,000 g, for 10 min at 4°C). The  
6  
7 resultant cellular pellets were washed with buffer (50 mM Tris-HCl [pH 8.0], 100 mM NaCl)  
8  
9 and then stored at -80°C.

10  
11  
12 Purification of YpeB-C entailed defrosting and resuspension of the *E. coli* cellular  
13  
14 pellet in 16 ml of ice-cold binding buffer (20 mM sodium phosphate [pH 7.4] plus 500 mM  
15  
16 NaCl, 20 mM imidazole and 1 mM phenylmethylsulfonyl fluoride [PMSF]). The cells were  
17  
18 lysed by passing the suspension twice through a One Shot Cell Disrupter (Constant Systems  
19  
20 Ltd., Northampton, UK) operating at 20 x 10<sup>3</sup> lb/in<sup>2</sup>. The cell lysate was then centrifuged  
21  
22 (15,000 g, for 20 min at 4°C), before passing the supernatant through a 0.46 µm syringe  
23  
24 filter. The clarified lysate was then loaded on to a 1 ml Ni-Sepharose HisTrap HP column  
25  
26 (GE Healthcare) fitted to an AKTA Pure protein purification system (GE Healthcare), which  
27  
28 had been pre-equilibrated with the same ice-cold buffer. The protein was eluted in the same  
29  
30 buffer containing 500 mM imidazole, and then buffer-exchanged and concentrated into 50  
31  
32 mM Tris-HCl (pH 8.0), containing 0.5 mM EDTA and 1 mM DTT, using an Amicon  
33  
34 centrifugal filter unit with a 10 kDa MWCO (Merck Millipore, Watford, UK). The C-  
35  
36 terminal His<sub>10</sub> affinity tag was removed from YpeB-C by digesting overnight at 4°C with  
37  
38 TEV (S219V) protease (1 µg TEV protease to every 100 µg YpeB-C). The reaction mix was  
39  
40 applied to a 1 ml Ni-Sepharose HisTrap HP column, equilibrated with ice-cold 20 mM  
41  
42 sodium phosphate (pH 7.4) containing 500 mM NaCl. The affinity tag-free YpeB-C protein,  
43  
44 present in the column flow-through fraction, was buffer exchanged into 20 mM Tris-HCl (pH  
45  
46 7.5) before loading onto a 1 ml Resource Q anion-exchange column (GE Healthcare)  
47  
48 equilibrated with the same buffer at room temperature. A salt gradient was applied using 20  
49  
50 mM Tris-HCl (pH 7.5) plus 1 M NaCl, flow rate 4 ml min<sup>-1</sup>, with the YpeB-C protein eluting  
51  
52 in fractions containing approximately 250 mM NaCl. The fractions were then combined,  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 concentrated by ultra-filtration, and further purified by gel filtration, using a Superdex 75  
4 column (GE Healthcare) equilibrated with phosphate buffered saline (pH 7.4) at room  
5 temperature. Finally, the eluted protein was de-salted using a HiTrap desalting column (GE  
6 Healthcare) and concentrated to 12 mg/mL by ultrafiltration (Amicon Centrifugal Filter  
7 Units, MWCO 10kDa; Millipore) in 5 mM sodium phosphate (pH 7.0) plus 25 mM NaCl.  
8  
9 The purified protein was aliquoted and stored at -80°C.  
10  
11  
12  
13  
14  
15  
16  
17

### 18 *Crystallisation of YpeB-C*

19  
20 Crystallisation trials were performed using the vapour diffusion sitting-drop technique in 96-  
21 well MRC 2-drop crystallisation plates (SWISSCI, Wokingham, UK). □□□□nL of the  
22 crystallisation screen conditions were mixed with 200 nL of protein solution (12 mg/ml) and  
23 set against 70 μL of reservoir using a crystallisation robot (Crystal Phoenix, Art Robbins  
24 Instruments, Inc.). A number of crystallisation trials using various crystallisation screening  
25 kits were performed, incubated at 19°C and monitored in a Rock Imager 1000 (Formulatrix,  
26 Inc.) automated imaging system. After identifying an initial crystallisation hit in condition #1  
27 of the JCSG+ screen (Qiagen) containing 0.2 M lithium sulphate, 0.1 M sodium acetate pH  
28 4.5 and 50 % (v/v) PEG 400, optimisation trials were conducted using the hanging drop  
29 vapour diffusion method at 19°C in 24-well hanging-drop crystallisation plates (Hampton  
30 Research) containing varying concentrations of precipitant agents. Diffraction quality  
31 crystals for YpeB-C were obtained from 3 μl drops containing a 1:1 mixture of 12 mg/ml  
32 protein and a crystallisation solution composed of 0.27 M lithium sulphate and 44 % (v/v)  
33 PEG 400 in 0.1 M sodium acetate (pH 4.5). Growth of relatively thin plate-type crystals was  
34 evident within 24 – 48 h, with the crystals attaining their maximum size (approx. 0.2 x 0.1 x  
35 0.05 mm<sup>3</sup>) after 1-2 weeks. To obtain the heavy metal derivative for the phasing experiments  
36 the crystals of YpeB-C were soaked for 4 h in a drop containing the crystallisation condition  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 and 5 mM potassium tetrachloroplatinate (II). The crystals were then back-soaked in a drop  
4 containing the crystallisation condition and 26 % (v/v) ethylene glycol for 10-20 min and  
5  
6 then flash frozen in liquid nitrogen until the X-ray data collection.  
7  
8  
9

### 10 11 *Diffraction data Collection*

12  
13 The X-ray diffraction datasets for both the native and platinum derivative crystals of YpeB-C  
14 were collected at the Diamond Light Source (Oxford, UK), beamline I04. The platinum  
15 derivative crystal was subjected to an initial fluorescence scan under an attenuated beam at  
16 the platinum K-edge, prior to SAD diffraction data collection from the same crystal at a  
17 wavelength of 1.0721 Å. The native crystal dataset was collected at a wavelength of 0.9795  
18 Å. The crystals of YpeB-C, which belonged to the C-centred orthorhombic system and  
19 C222<sub>1</sub> space group, diffracted to a maximum resolution of 1.80 Å in the case of native dataset  
20 and 2.20 Å for the crystal of platinum derivative. The diffraction data were indexed, scaled,  
21 and merged using XDS software. Analysis of the crystal solvent content using Matthews'  
22 Coefficient indicated the presence of a single molecule in the crystallographic asymmetric  
23 unit, resulting in a Matthews coefficient of 2.34 and solvent content of about 47.5%.  
24  
25 Crystallographic data collection, phasing and refinement statistics are detailed in Table 1.  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42

### 43 *Crystal Structure Determination, Model Building and Refinement*

44  
45 The *B. megaterium* YpeB-C crystal structure was solved using the single-wavelength  
46 anomalous diffraction (SAD) technique. Experimental phases were obtained from the  
47 platinum derivative SAD dataset. The PHENIX<sup>11</sup> software suite was used to perform all  
48 crystallographic calculations for structure solution and refinement. The analysis of the  
49 anomalous measurability values in the SAD dataset, as defined by the Xtriage module of  
50 PHENIX, demonstrated the presence of statistically significant anomalous signal to about 3.3  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3 Å resolution. The search for anomalous atoms, conducted using the HySS (Hybrid  
4 Substructure Search) module of the PHENIX software suite, identified the position of 13  
5 possible platinum ion sites in the asymmetric unit. Phases were calculated using Phaser  
6 (Figure of Merit 0.32) and further improved by electron density modification using  
7 RESOLVE (Figure of Merit 0.62). The resulting experimental electron density map was  
8 readily interpretable and an automated model building procedure in PHENIX against the  
9 platinum derivative dataset, including all available resolution to 2.20 Å, produced an initial  
10 model containing 93 residues out of a total of 234 residues ( $R_{\text{cryst}} = 45.9\%$ ,  $R_{\text{free}} = 46.9\%$ ).  
11 This model was then manually rebuilt using the COOT molecular graphics software suite and  
12 refined using PHENIX against the native 1.80 Å resolution dataset. A total of 9 rounds of  
13 manual rebuilding and refinement were performed, during which the lesser-defined PepSY  
14 domain (PepSY1) was successfully traced. Solvent molecules and sulphate ions were added  
15 manually and through an automated procedure as implemented in the PHENIX refinement  
16 protocols. The  $R_{\text{cryst}}$  and  $R_{\text{free}}$  converged to the values of 19.5% and 22.7%, respectively.  
17 The crystallographic statistics and structural validation aspects are shown in Table 1. Atomic  
18 coordinates and structure factors for the YpeB-C crystal structure have been deposited with  
19 the Protein Data Bank (PDB) under accession code 5BOI.  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Results and Discussion

### *YpeB-C crystal structure*

Crystallisation experiments conducted with full-length YpeB protein and an N-terminal domain construct failed to yield diffraction quality crystals. However, the C-terminal domain of the *B. megaterium* YpeB protein (hereafter YpeB-C) crystallised readily, with SAD phasing permitting the solution of the three dimensional structure at 1.80 Å resolution. The respective  $R_{\text{cryst}}$  and  $R_{\text{free}}$  values were 19.5% and 22.7% (Table 1). The crystallised protein contained residues 216 – 449 of the full-length protein, plus the cloning artefacts MGGGFA and ENLYFQ at the respective N- and C-termini. The analysed crystal contained a single YpeB-C monomer in the asymmetric unit, plus 100 water molecules and two sulphate ions derived from the crystallisation buffer. The calculated electron density map allowed largely unambiguous tracing of most of the crystallised protein residues, although neither the N- or C-terminal cloning artefact residues were visible in the map. Similarly, the final eight YpeB-C residues (S442 through to V449) were not observed in the electron density map, presumably as a result of structural disorder at the C-terminus. In two cases (R217 and K418) electron density was such that only  $C_{\beta}$  of side chain atoms could be placed with certainty (residue numbering refers to the amino acid position in the full length protein sequence). The Ramachandran plot, produced by MolProbity<sup>12</sup>, revealed that 97% and 100% of amino acids were in the favoured and allowed regions, respectively.

The YpeB-C molecule adopts an elongated S-shaped structure with approximate dimensions of 60 x 40 x 20 Å (Figure 1). As predicted from bioinformatic analysis, the molecule is formed of three tandem PepSY domain repeats (**PepSY – peptidase of M4 and Subtilis YpeB protein**). Each PepSY domain is comprised of four antiparallel beta strands with an  $\alpha$ -helix positioned on the convex side of the beta sheets. PepSY1 (S220 – D284) and PepSY2 (I292 – R368) are connected by a long loop extending from  $\beta$ -strand 4 to  $\alpha$ -helix 2

1  
2  
3 of PepSY2, and oriented such that interaction between the domains is limited to the  
4 antiparallel alignment of  $\beta$ -strand 4 from PepSY1 with  $\beta$ -strand 10 of PepSY2. The major  $\alpha$ -  
5 helices of both PepSY1 and PepSY2 are on the same face of the molecule, each traversing  
6 their respective beta sheets at a similar angle. In contrast to the other PepSY domains,  
7 PepSY2 has three short tandem  $\beta$ -strands, namely  $\beta_6$ , and the hairpin forming  $\beta_7$  and  $\beta_8$ .  
8  
9  
10  
11  
12  
13  
14 PepSY3 (S378 – L439) is connected to PepSY2 via an **extended short-helix ( $\alpha_3$ )-loop-long**  
15 **helix ( $\alpha_4$ ) motif, which together with the 4-stranded antiparallel beta sheet**, is characteristic  
16 of PepSY-like folds adopted in other proteins<sup>13</sup>. The third PepSY domain is aligned broadly  
17 in the same plane as PepSY1 and 2, forming a laterally compressed molecule, however, the  
18 domain is oriented almost perpendicularly to PepSY1 and PepSY2, such that  $\alpha$ -helix 4 forms  
19 the base of the molecule.  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29

### 30 *Structural alignment*

31  
32 **Structural comparison between YpeB-C and other protein structures, and indeed comparison**  
33 **of YpeB-C's individual PepSY domains with each other, was examined using both rigid and**  
34 **flexible superposition algorithms, with the *FATCAT* flexible pairwise alignment algorithm**  
35 **delivering the most satisfactory results. Using this method to compare YpeB-C's individual**  
36 **PepSY domains, one at a time against each other, revealed that all three domains are**  
37 **structurally significantly similar (P values  $< 5 \times 10^{-4}$ ). Root mean square deviation (r.m.s.d)**  
38 **values range from 2.11 Å (superposition of PepSY2 with PepSY3) to 2.49 Å (superposition**  
39 **of PepSY1 with PepSY2), despite low sequence identity values ( $< 13\%$  in all cases).**  
40  
41  
42  
43  
44  
45  
46  
47  
48

49  
50 Searches for proteins of similar structure to YpeB-C were made using the *DALI*<sup>14</sup> and  
51 *FATCAT*<sup>15</sup> servers. Both servers returned a similar range of hits, comprising PepSY,  
52 DUF2874 and  $\beta$ -lactamase inhibitor (BLIP)-like proteins. Proteins of the latter two families  
53 contain similar 4-strand antiparallel beta sheet-loop-alpha helical structural domains, and,  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 like PepSY-containing proteins, have been identified as having inhibitory functions<sup>13</sup>.  
4  
5 Notably, structural similarity to a YpeB-like protein from the anaerobic spore-former  
6  
7 *Clostridium difficile* 630 was identified by these searches. This 205 amino acid protein,  
8  
9 encoded at locus CD630\_16220 (UniProt Q186H8; PDB 4EXR), has not been characterised  
10  
11 functionally in *C. difficile* spores, and does not share an operon with a SleB gene, a candidate  
12  
13 for which is encoded at locus CD630\_35630. The *C. difficile* YpeB-like protein comprises  
14  
15 two tandem repeat PepSY domains, which are oriented to form a C-shaped molecule in which  
16  
17 the  $\beta$ -sheets form the inner concave surface and the  $\alpha$ -helices form the outer convex surface.  
18  
19 In contrast to *Bacillus* species YpeB proteins, the *C. difficile* protein lacks an N-terminal  
20  
21 domain, other than the predicted membrane anchor sequence. Superposition of YpeB-C with  
22  
23 4EXR using the *FATCAT* flexible pairwise alignment algorithm revealed that the two  
24  
25 structures are significantly similar ( $P < 7 \times 10^{-6}$ ) with an optimised r.m.s.d of 2.49 Å when a  
26  
27 single twist is introduced (Figure 2[a]). The aligned structures have 131 equivalent positions  
28  
29 while sharing only 15% sequence identity. The requirement for the introduction of a twist to  
30  
31 re-orient 4EXR's PepSY domains, enabling significant structural alignment with YpeB-C,  
32  
33 presumably explains the failure of early attempts at solving the YpeB-C structure by  
34  
35 molecular replacement (MR), using the 4EXR structure as the MR probe.  
36  
37  
38  
39

40  
41 Other notable structural alignment hits identified by both *DALI* and *FATCAT* servers  
42  
43 include the uncharacterised *Bacillus subtilis* YpmB protein (UniProt P54396; PDB 2GU3),  
44  
45 which like 4EXR contains two PepSY domains oriented to give a relatively compressed  
46  
47 concave shaped molecule, although the placement of the  $\alpha$ -helices differs to that observed in  
48  
49 4EXR. Again, the *FATCAT* flexible pairwise alignment algorithm introduces a single twist  
50  
51 that permits significant ( $P < 3 \times 10^{-4}$ ) structural alignment of both YpmB PepSY domains with  
52  
53 YpeB-C. Here, YpmB superposes with YpeB-C's PepSY1 and PepSY2 domains, with an  
54  
55 optimised r.m.s.d. value of 3.33 Å across 122 equivalent positions (sequence identity <11 %).  
56  
57  
58  
59  
60

1  
2  
3 The YpeB-C structure also aligns significantly ( $P < 2 \times 10^{-3}$ ) with the *B. subtilis* sensory  
4 histidine kinase regulatory protein YycI (UniProt Q45612; PDB 2o3o), where the  
5 introduction of two twists permits an optimised r.m.s.d. of 2.97 Å, with 174 residues placed  
6 in equivalent positions.  
7  
8  
9

10  
11 In all cases that we have examined then, including several not reported here, optimal  
12 structural alignment between YpeB-C and other PepSY (or DUF2874 and BLIP) family  
13 proteins requires re-positioning of one or more PepSY domains, indicating that the overall  
14 architecture adopted by the triple tandem PepSY repeat in YpeB-C may be unique to this  
15 protein. Presumably this will apply also to other *Bacillus* spore YpeB proteins, although this  
16 will have to be determined in due course.  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26

### 27 *Structural insights to YpeB function*

28  
29 Recent work conducted with truncated YpeB proteins in *B. anthracis* spores demonstrated  
30 that all three PepSY domains are required for the concomitant stability of YpeB and SleB in  
31 dormant spores<sup>6</sup>. In the same study, a variant YpeB protein containing the N-terminal  
32 domain and PepSY1 from the C-terminal domain was shown to be relatively stable in the  
33 spore, however, SleB abundance was diminished in this strain, indicating that a region  
34 beyond PepSY1 is essential to confer stability to SleB in the spore. With this in mind,  
35 examination of an electrostatic potential surface representation of YpeB-C, calculated using  
36 the Adaptive Poisson-Boltzmann Solver (APBS) Pymol plug-in<sup>16</sup>, reveals a large shallow  
37 channel or groove that traverses the middle of the molecule. This region is marked by  
38 positive surface charge, formed by K345 and K347 from  $\beta$ 9 (PepSY2) and K361 and K366  
39 from the short  $\alpha$ -helix 3, and may be a candidate site for protein interactions (Figure 2[b]).  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
Bernhards *et al.*, also identified a number of defined residues (Y254, Y410 and G430) that  
appear to be important to YpeB stability in *B. anthracis* spores<sup>6</sup>. Analysis of the

1  
2  
3 corresponding residues in the *B. megaterium* YpeB-C structure reveals that G431 is located  
4  
5 in the PepSY3  $\beta$ 13 -  $\beta$ 14 connective loop, and it is not immediately obvious why substitution  
6  
7 of glycine for alanine would destabilise the protein in the spore. The hydroxyl group from  
8  
9 Y411, on the other hand, is sufficiently close (2.7 Å) in the crystal structure to form a  
10  
11 hydrogen bond with the backbone carbonyl oxygen atom of I377, which is part of the loop  
12  
13 connecting PepSY2 and PepSY3. Accordingly, this intra-molecular interaction may  
14  
15 contribute towards the conformational stability of the PepSY3 domain. Alternatively, Y411  
16  
17 is solvent exposed and may be important in interacting with SleB or spore proteases, with  
18  
19 disruption to this interaction resulting in proteolytic degradation of YpeB and SleB in the  
20  
21 developing spore. Similarly, Y254 can hydrogen bond with the backbone carbonyl oxygen  
22  
23 atom of V219, which is located in the loop that precedes  $\alpha$ -helix 1. Again, this intra-  
24  
25 molecular interaction may contribute towards the structural stability of the domain, or solvent  
26  
27 exposed Y254 may be involved in YpeB/SleB-stabilising protein interactions.  
28  
29  
30

31  
32 To conclude, the crystal structure of the C-terminal domain of the *B. megaterium*  
33  
34 YpeB protein has been solved to 1.80 Å. The objective moving forward will be to determine  
35  
36 the structure of the full-length protein, with a view to revealing further insights to the precise  
37  
38 function of this protein and the molecular mechanisms that underpin its role in regulating  
39  
40 SleB activity in the spore.  
41  
42  
43  
44

#### 45 **Acknowledgements**

46  
47 Crystallographic experiments were performed in the Crystallographic X-ray Facility (CXF) at  
48  
49 the Department of Biochemistry, University of Cambridge. DYC was supported by CXF.  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**References**

1. Nicholson WL, Munakata N, Horneck G, Melosh HJ, Setlow P. Resistance of *Bacillus* Endospores to Extreme Terrestrial and Extraterrestrial Environments. *Microbiology and Molecular Biology Reviews* 2000;64(3):548-572.
2. Setlow P. Germination of spores of *Bacillus* species: what we know and do not know. *J Bacteriol* 2014;196(7):1297-1305.
3. 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.
4. 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* 2002;148(Pt 8):2383-2392.
5. 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. *J Bacteriol* 2013;195(11):2530-2540.
6. Bernhards CB, Popham DL. Role of YpeB in cortex hydrolysis during germination of *Bacillus anthracis* spores. *J Bacteriol* 2014;196(19):3399-3409.
7. Bernhards CB, Chen Y, Toutkoushian H, Popham DL. HtrC is involved in proteolysis of YpeB during germination of *Bacillus anthracis* and *Bacillus subtilis* spores. *J Bacteriol* 2015;197(2):326-336.
8. Yeats C, Rawlings ND, Bateman A. The PepSY domain: a regulator of peptidase activity in the microbial environment? *Trends Biochem Sci* 2004;29(4):169-172.
9. 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.



- 1  
2  
3 10. Li Y, Jin K, Setlow B, Setlow P, Hao B. Crystal structure of the catalytic domain of  
4 the *Bacillus cereus* SleB protein, important in cortex peptidoglycan degradation during spore  
5 germination. *J Bacteriol* 2012;194(17):4537-4545.  
6  
7
- 8  
9 11. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ,  
10 Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read  
11 RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. PHENIX: a comprehensive  
12 Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol*  
13 *Crystallogr* 2010;66(Pt 2):213-221.  
14  
15
- 16  
17 12. Davis IW, Murray LW, Richardson JS, Richardson DC. MOLPROBITY: structure  
18 validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids*  
19 *Res* 2004;32(Web Server issue):W615-619.  
20  
21
- 22  
23 13. Das D, Finn RD, Carlton D, Miller MD, Abdubek P, Astakhova T, Axelrod HL,  
24 Bakolitsa C, Chen C, Chiu HJ, Chiu M, Clayton T, Deller MC, Duan L, Ellrott K, Ernst D,  
25 Farr CL, Feuerhelm J, Grant JC, Grzechnik A, Han GW, Jaroszewski L, Jin KK, Klock HE,  
26 Knuth MW, Kozbial P, Krishna SS, Kumar A, Marciano D, McMullan D, Morse AT,  
27 Nigoghossian E, Nopakun A, Okach L, Puckett C, Reyes R, Rife CL, Sefcovic N, Tien HJ,  
28 Trame CB, van den Bedem H, Weekes D, Wooten T, Xu Q, Hodgson KO, Wooley J, Elsliger  
29 MA, Deacon AM, Godzik A, Lesley SA, Wilson IA. The structure of BVU2987 from  
30 *Bacteroides vulgatus* reveals a superfamily of bacterial periplasmic proteins with possible  
31 inhibitory function. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 2010;66(Pt 10):1265-  
32 1273.  
33  
34
- 35  
36 14. Holm L, Kaariainen S, Rosenstrom P, Schenkel A. Searching protein structure  
37 databases with DaliLite v.3. *Bioinformatics* 2008;24(23):2780-2781.  
38  
39
- 40  
41 15. Ye Y, Godzik A. FATCAT: a web server for flexible structure comparison and  
42 structure similarity searching. *Nucleic Acids Res* 2004;32(Web Server issue):W582-585.  
43  
44
- 45  
46 16. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of  
47 nanosystems: application to microtubules and the ribosome. *Proceedings of the National*  
48 *Academy of Sciences of the United States of America* 2001;98(18):10037-10041.  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



## Figure Legends

**Figure 1.** Crystal structure of the C-terminal domain of *B. megaterium* YpeB. (a) Ribbon representation of the YpeB-C structure. The molecule is rotated 90° counter-clockwise in (b). The molecule is composed of three tandemly arranged PepSY domains (PepSY1, in orange [S220 – D284], PepSY2 in pale blue [I292 – R368], and PepSY3 in light green [S378 – L439]). Residues Y254 and Y411, located within PepSY1 and PepSY3 respectively, which have been shown to be of functional importance in the orthologous *B. anthracis* YpeB protein<sup>6</sup>, are shown in stick representation. (c) Secondary structure elements associated with YpeB-C amino acid sequence (generated by PDBsum; <http://www.ebi.ac.uk/pdbsum/>).

**Figure 2.** (a) *FATCAT* flexible pairwise alignment of *B. megaterium* YpeB-C with *Clostridium difficile* YpeB-like protein (PDB: 4EXR). The alignment has an optimised r.m.s.d of 2.49 Å when a single twist is introduced in the 4EXR structure, facilitating alignment with the YpeB-C PepSY2 and PepSY3 domains. YpeB-C is coloured as in Figure 1; 4EXR is coloured pink. (b) Molecular surface representation of YpeB-C coloured according to the local electrostatic potential, ranging from -3 kT/e in red (most negative) to +3 kT/e in blue (most positive). The local electrostatic potential was calculated using the APBS Pymol plug-in<sup>16</sup>.

**Table 1** Crystallographic data collection, phasing and refinement statistics

	SAD dataset K <sub>2</sub> PtCl <sub>4</sub> derivative	Native Dataset
<b>Data collection</b>		
Radiation Source	Diamond (UK), I04	Diamond (UK), I04
Wavelength (Å)	1.0721	0.9795
Space group	C222 <sub>1</sub>	C222 <sub>1</sub>
Cell dimensions:		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	52.77 108.85 85.73	53.21 110.10 85.81
<i>α</i> , <i>β</i> , <i>γ</i> (°)	90.0 90.0 90.0	90.0 90.0 90.0
Resolution (Å)	28.58 – 2.20 (2.32 – 2.20) <sup>1</sup>	43.63 – 1.80 (1.90 – 1.80) <sup>1</sup>
<i>R</i> <sub>merge</sub> <sup>2</sup> (%)	8.7 (89.8)	6.0 (84.0)
< <i>I</i> / <i>σ</i> ( <i>I</i> )>	20.2 (3.5)	15.7 (2.5)
Completeness (%)	99.9 (100.0)	99.9 (99.9)
Redundancy	16.6 (16.3)	7.3 (7.5)
Number of unique reflections	12,907	23,741
Anomalous Completeness (%)	100.0 (100.0)	
Anomalous Redundancy (%)	8.7 (8.3)	
<b>Phasing</b> <sup>3</sup>		
Number of sites found	13	
Overall score (100 * BAYES-CC)	51.7	
Figure of Merit (SOLVE)	0.32	
Figure of Merit (RESOLVE)	0.62	
<b>Density Modification</b> <sup>3</sup>		
R-factor	0.32	
Map skew	0.18	
Corr. of local RMS density	0.88	
<b>Refinement</b>		
Resolution (Å)		46.34 – 1.80
Number of reflections used:		
	Total	23,700
	<i>R</i> <sub>free</sub> set	1,997
<i>R</i> <sub>cryst</sub> <sup>4</sup> / <i>R</i> <sub>free</sub> <sup>5</sup> (%)		19.5 / 22.7
Solvent content, %		47.5
Number of protein molecules in asymmetric unit		1
Number of non-hydrogen of atoms in asymmetric unit:		
	Protein atoms	1833
	ion	10
	Water atoms	100
B-factor, (Å <sup>2</sup> ):		
	Average	53.7
	Wilson	30.0
Ramachandran plot analysis, number of residues in:		
	Favoured regions, %	97.36
	Allowed regions, %	2.64
	Disallowed regions, %	0
R.m.s. deviations:		

Bond lengths (Å)	0.008
Bond angles (°)	1.109

<sup>1</sup> The statistics shown in parentheses are for the highest-resolution shell.

<sup>2</sup>  $R_{\text{merge}} = (\sum_{hkl} \sum_i |I_i(hkl) - I_{\text{mean}}(hkl)|) / \sum_{hkl} \sum_i I_i(hkl)$ .

<sup>3</sup> As calculated by PHENIX software suite

<sup>4</sup>  $R_{\text{cryst}} = \sum_{hkl} (|F_{\text{obs}}(hkl)| - |F_{\text{calc}}(hkl)|) / \sum_{hkl} |F_{\text{obs}}(hkl)|$

<sup>5</sup>  $R_{\text{free}}$  is the same as  $R_{\text{cryst}}$  but for a random subset of reflections not included in the refinement, about 10% of total reflections.

For Peer Review

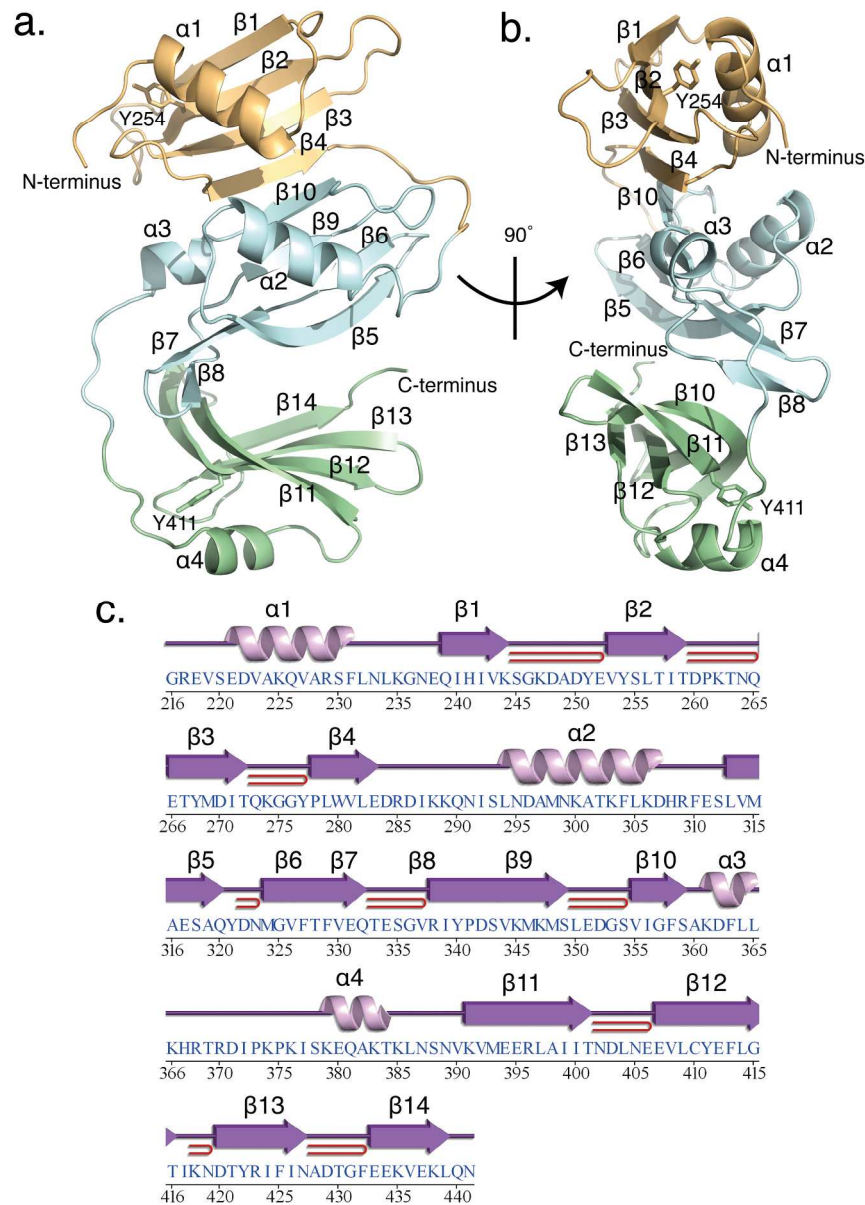


Figure 1. Crystal structure of the C-terminal domain of *B. megaterium* YpeB. (a) Ribbon representation of the YpeB-C structure. The molecule is rotated 90° counter-clockwise in (b). The molecule is composed of three tandemly arranged PepSY domains (PepSY1, in orange [S220 – D284], PepSY2 in pale blue [I292 – R368], and PepSY3 in light green [S378 – L439]). Residues Y254 and Y411, located within PepSY1 and PepSY3 respectively, which have been shown to be of functional importance in the orthologous *B. anthracis* YpeB protein<sup>6</sup>, are shown in stick representation. (c) Secondary structure elements associated with YpeB-C amino acid sequence (generated by PDBsum; <http://www.ebi.ac.uk/pdbsum/>).

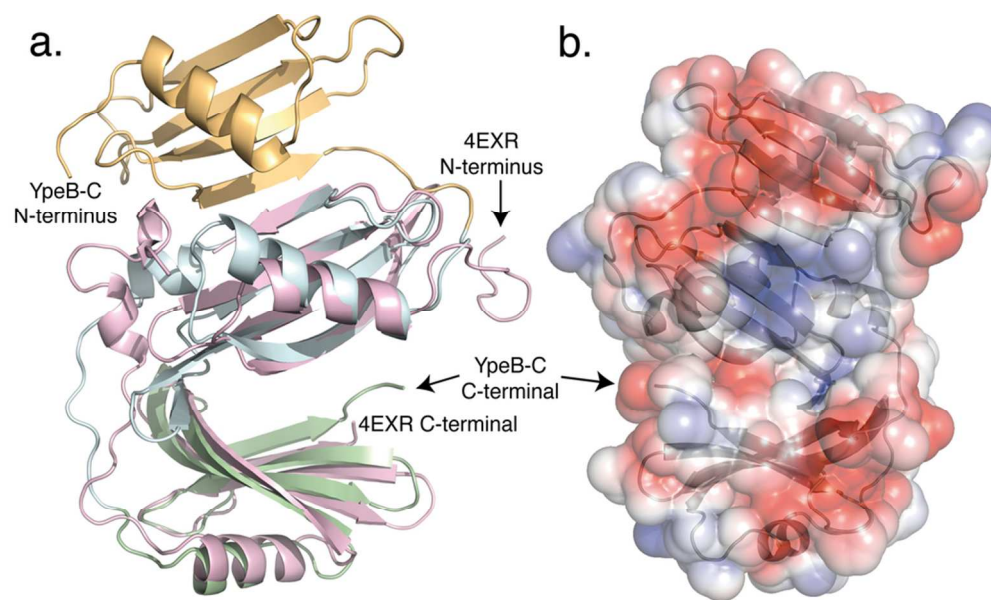


Figure 2. (a) FATCAT flexible pairwise alignment of *B. megaterium* YpeB-C with *Clostridium difficile* YpeB-like protein (PDB: 4EXR). The alignment has an optimised r.m.s.d of 2.49 Å when a single twist is introduced in the 4EXR structure, facilitating alignment with the YpeB-C PepSY2 and PepSY3 domains. YpeB-C is coloured as in Figure 1; 4EXR is coloured pink. (b) Molecular surface representation of YpeB-C coloured according to the local electrostatic potential, ranging from -3 kT/e in red (most negative) to +3 kT/e in blue (most positive). The local electrostatic potential was calculated using the APBS Pymol plug-in16.  
88x52mm (300 x 300 DPI)