PROTEINS:
Structure, Function, and Bioinformatics

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

Journal:	*PROTEINS: Structure, Function, and Bioinformatics*
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

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

Short title: *Bacillus megaterium* YpeB-C crystal structure

Fatma Işık Üstok[1, 3], 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]Current Address: 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 peptidoglycan, cortex lytic enzyme; SleB; CwlJ; SleL; inhibitory protein

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[1]. 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[2].

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[3].

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[4], and vice versa;[5,6] (ii) YpeB has been shown to be proteolytically cleaved during germination, perhaps breaking the interaction with SleB and permitting activity of the latter[4,7]; (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

M36 metallopeptidases[8]; and (iv) *in vitro* assays conducted with recombinant versions of SleB co-incubated with either full-length YpeB or defined N-terminal and C-terminal domain fragments demonstrate inhibition of SleB activity, to various degrees, in the presence of the variant YpeB proteins[5]. Despite the above, efforts to ascertain whether SleB and YpeB physically interact have as yet delivered negative results[5,6].

In order to gain insight to the molecular mechanisms that regulate cortical depolymerisation during spore germination, high-resolution structures for a number of spore CLEs have recently been solved using X-ray crystallographic methods[9,10]. In a similar vein, we present here the crystal structure of the C-terminal domain of the YpeB protein from *B. megaterium*.

Materials and Methods

Expression and purification of YpeB-C

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

5'-ATGGTTGTTGGATTTGCTGCTTTTCATCAATTAAAAGGTAGAGAA-3' and 5'-TTGGAAGTATAAATTTTCCACTTCATTATATAAAGGTTCTGAGTT-3'. The PCR amplicon was purified and cloned into plasmid pBADcLIC *E. coli* expression plasmid, which is designed to create C-terminal His_{10} fusion proteins. Protein expression was conducted using *E. coli* Top10 cells (Life Technologies), which were cultured in LB medium containing 50 µg/ml carbenicillin at 37°C, 225 rpm, until the optical density at 660 nm (OD_{660}) reached 0.6, when the temperature was reduced to 30°C and protein expression induced by the

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

Purification of YpeB-C entailed defrosting and resuspension of the *E. coli* cellular pellet in 16 ml of ice-cold binding buffer (20 mM sodium phosphate [pH 7.4] plus 500 mM NaCl, 20 mM imidazole and 1 mM phenylmethylsulfonyl fluoride [PMSF]). The cells were lysed by passing the suspension twice through a One Shot Cell Disrupter (Constant Systems Ltd., Northampton, UK) operating at 20 x 10^3 lb/in^2. The cell lysate was then centrifuged (15,000 g, for 20 min at 4°C), before passing the supernatant through a 0.46 µM syringe filter. The clarified lysate was then loaded on to a 1 ml Ni-Sepharose HisTrap HP column (GE Healthcare) fitted to an AKTA Pure protein purification system (GE Healthcare), which had been pre-equilibrated with the same ice-cold buffer. The protein was eluted in the same buffer containing 500 mM imidazole, and then buffer-exchanged and concentrated into 50 mM Tris-HCl (pH 8.0), containing 0.5 mM EDTA and 1 mM DTT, using an Amicon centrifugal filter unit with a 10 kDa MWCO (Merck Millipore, Watford, UK). The C-terminal His_{10} affinity tag was removed from YpeB-C by digesting overnight at 4°C with TEV (S219V) protease (1 µg TEV protease to every 100 µg YpeB-C). The reaction mix was applied to a 1 ml Ni-Sepharose HisTrap HP column, equilibrated with ice-cold 20 mM sodium phosphate (pH 7.4) containing 500 mM NaCl. The affinity tag-free YpeB-C protein, present in the column flow-through fraction, was buffer exchanged into 20 mM Tris-HCl (pH 7.5) 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) plus 1 M NaCl, flow rate 4 ml min^{-1}, with the YpeB-C protein eluting in fractions containing approximately 250 mM NaCl. The fractions were then combined,

concentrated by ultra-filtration, and further purified by gel filtration, using a Superdex 75 column (GE Healthcare) equilibrated with phosphate buffered saline (pH 7.4) at room temperature. Finally, the eluted protein was de-salted using a HiTrap desalting column (GE Healthcare) and concentrated to 12 mg/mL by ultrafiltration (Amicon Centrifugal Filter Units, MWCO 10kDa; Millipore) in 5 mM sodium phosphate (pH 7.0) plus 25 mM NaCl. The purified protein was aliquoted and stored at -80^{o}C.

Crystallisation of YpeB-C

Crystallisation trials were performed using the vapour diffusion sitting-drop technique in 96-well MRC 2-drop crystallisation plates (SWISSCI, Wokingham, UK). □□□□nL of the crystallisation screen conditions were mixed with 200 nL of protein solution (12 mg/ml) and set against 70 µL of reservoir using a crystallisation robot (Crystal Phoenix, Art Robbins Instruments, Inc.). A number of crystallisation trials using various crystallisation screening kits were performed, incubated at 19°C and monitored in a Rock Imager 1000 (Formulatrix, Inc.) automated imaging system. After identifying an initial crystallisation hit in condition #1 of the JCSG+ screen (Qiagen) containing 0.2 M lithium sulphate, 0.1 M sodium acetate pH 4.5 and 50 % (v/v) PEG 400, optimisation trials were conducted using the hanging drop vapour diffusion method at 19^{o}C in 24-well hanging-drop crystallisation plates (Hampton Research) containing varying concentrations of precipitant agents. Diffraction quality crystals for YpeB-C were obtained from 3 µl drops containing a 1:1 mixture of 12 mg/ml protein and a crystallisation solution composed of 0.27 M lithium sulphate and 44 % (v/v) PEG 400 in 0.1 M sodium acetate (pH 4.5). Growth of relatively thin plate-type crystals was evident within 24 – 48 h, with the crystals attaining their maximum size (approx. 0.2 x 0.1 x 0.05 mm^3) after 1-2 weeks. To obtain the heavy metal derivative for the phasing experiments the crystals of YpeB-C were soaked for 4 h in a drop containing the crystallisation condition

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

Diffraction data Collection

The X-ray diffraction datasets for both the native and platinum derivative crystals of YpeB-C were collected at the Diamond Light Source (Oxford, UK), beamline I04. The platinum derivative crystal was subjected to an initial fluorescence scan under an attenuated beam at the platinum K-edge, prior to SAD diffraction data collection from the same crystal at a wavelength of 1.0721 Å. The native crystal dataset was collected at a wavelength of 0.9795 Å. The crystals of YpeB-C, which belonged to the C-centred orthorhombic system and $C222_1$ space group, diffracted to a maximum resolution of 1.80 Å in the case of native dataset and 2.20 Å for the crystal of platinum derivative. The diffraction data were indexed, scaled, and merged using XDS software. Analysis of the crystal solvent content using Matthews' Coefficient indicated the presence of a single molecule in the crystallographic asymmetric unit, resulting in a Matthews coefficient of 2.34 and solvent content of about 47.5%. Crystallographic data collection, phasing and refinement statistics are detailed in Table 1.

Crystal Structure Determination, Model Building and Refinement

The *B. megaterium* YpeB-C crystal structure was solved using the single-wavelength anomalous diffraction (SAD) technique. Experimental phases were obtained from the platinum derivative SAD dataset. The PHENIX[11] software suite was used to perform all crystallographic calculations for structure solution and refinement. The analysis of the anomalous measurability values in the SAD dataset, as defined by the Xtriage module of PHENIX, demonstrated the presence of statistically significant anomalous signal to about 3.3

Å resolution. The search for anomalous atoms, conducted using the HySS (Hybrid Substructure Search) module of the PHENIX software suite, identified the position of 13 possible platinum ion sites in the asymmetric unit. Phases were calculated using Phaser (Figure of Merit 0.32) and further improved by electron density modification using RESOLVE (Figure of Merit 0.62). The resulting experimental electron density map was readily interpretable and an automated model building procedure in PHENIX against the platinum derivative dataset, including all available resolution to 2.20 Å, produced an initial model containing 93 residues out of a total of 234 residues (R_{cryst} = 45.9%, R_{free} = 46.9%). This model was then manually rebuilt using the COOT molecular graphics software suite and refined using PHENIX against the native 1.80 Å resolution dataset. A total of 9 rounds of manual rebuilding and refinement were performed, during which the lesser-defined PepSY domain (PepSY1) was successfully traced. Solvent molecules and sulphate ions were added manually and through an automated procedure as implemented in the PHENIX refinement protocols. The R_{cryst} and R_{free} converged to the values of 19.5% and 22.7%, respectively. The crystallographic statistics and structural validation aspects are shown in Table 1. Atomic coordinates and structure factors for the YpeB-C crystal structure have been deposited with the Protein Data Bank (PDB) under accession code 5BOI.

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_{cryst} and R_{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_β 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[12], 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 – *pep*tidase of M4 and *S*ubtilis *Y*peB protein). Each PepSY domain is comprised of four antiparallel beta strands with an α-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 β-strand 4 to α-helix 2

of PepSY2, and oriented such that interaction between the domains is limited to the antiparallel alignment of β-strand 4 from PepSY1 with β-strand 10 of PepSY2. The major α-helices of both PepSY1 and PepSY2 are on the same face of the molecule, each traversing their respective beta sheets at a similar angle. In contrast to the other PepSY domains, PepSY2 has three short tandem β-strands, namely β6, and the hairpin forming β7 and β8. PepSY3 (S378 – L439) is connected to PepSY2 via an extended short-helix (α3)-loop-long helix (α4) motif, which together with the 4-stranded antiparallel beta sheet, is characteristic of PepSY-like folds adopted in other proteins[13]. The third PepSY domain is aligned broadly in the same plane as PepSY1 and 2, forming a laterally compressed molecule, however, the domain is oriented almost perpendicularly to PepSY1 and PepSY2, such that α-helix 4 forms the base of the molecule.

Structural alignment

Structural comparison between YpeB-C and other protein structures, and indeed comparison of YpeB-C's individual PepSY domains with each other, was examined using both rigid and flexible superposition algorithms, with the *FATCAT* flexible pairwise alignment algorithm delivering the most satisfactory results. Using this method to compare YpeB-C's individual PepSY domains, one at a time against each other, revealed that all three domains are structurally significantly similar (P values < $5x10^{-4}$). Root mean square deviation (r.m.s.d) values range from 2.11 Å (superposition of PepSY2 with PepSY3) to 2.49 Å (superposition of PepSY1 with PepSY2), despite low sequence identity values (<13% in all cases).

Searches for proteins of similar structure to YpeB-C were made using the *DALI*[14] and *FATCAT*[15] servers. Both servers returned a similar range of hits, comprising PepSY, DUF2874 and β-lactamase inhibitor (BLIP)-like proteins. Proteins of the latter two families contain similar 4-strand antiparallel beta sheet-loop-alpha helical structural domains, and,

like PepSY-containing proteins, have been identified as having inhibitory functions[13]. Notably, structural similarity to a YpeB-like protein from the anaerobic spore-former *Clostridium difficile* 630 was identified by these searches. This 205 amino acid protein, encoded at locus CD630_16220 (UniProt Q186H8; PDB 4EXR), has not been characterised functionally in *C. difficile* spores, and does not share an operon with a SleB gene, a candidate for which is encoded at locus CD630_35630. The *C. difficile* YpeB-like protein comprises two tandem repeat PepSY domains, which are oriented to form a C-shaped molecule in which the β-sheets form the inner concave surface and the α-helices form the outer convex surface. In contrast to *Bacillus* species YpeB proteins, the *C. difficile* protein lacks an N-terminal domain, other than the predicted membrane anchor sequence. Superposition of YpeB-C with 4EXR using the *FATCAT* flexible pairwise alignment algorithm revealed that the two structures are significantly similar ($P < 7x10^{-6}$) with an optimised r.m.s.d of 2.49 Å when a single twist is introduced (Figure 2[a]). The aligned structures have 131 equivalent positions while sharing only 15% sequence identity. The requirement for the introduction of a twist to re-orient 4EXR's PepSY domains, enabling significant structural alignment with YpeB-C, presumably explains the failure of early attempts at solving the YpeB-C structure by molecular replacement (MR), using the 4EXR structure as the MR probe.

Other notable structural alignment hits identified by both *DALI* and *FATCAT* servers include the uncharacterised *Bacillus subtilis* YpmB protein (UniProt P54396; PDB 2GU3), which like 4EXR contains two PepSY domains oriented to give a relatively compressed concave shaped molecule, although the placement of the α-helices differs to that observed in 4EXR. Again, the *FATCAT* flexible pairwise alignment algorithm introduces a single twist that permits significant ($P < 3x10^{-4}$) structural alignment of both YpmB PepSY domains with YpeB-C. Here, YpmB superposes with YpeB-C's PepSY1 and PepSY2 domains, with an optimised r.m.s.d. value of 3.33 Å across 122 equivalent positions (sequence identity <11 %).

The YpeB-C structure also aligns significantly ($P < 2x10^{-3}$) with the *B. subtilis* sensory histidine kinase regulatory protein YycI (UniProt Q45612; PDB 2o3o), where the introduction of two twists permits an optimised r.m.s.d. of 2.97 Å, with 174 residues placed in equivalent positions.

In all cases that we have examined then, including several not reported here, optimal structural alignment between YpeB-C and other PepSY (or DUF2874 and BLIP) family proteins requires re-positioning of one or more PepSY domains, indicating that the overall architecture adopted by the triple tandem PepSY repeat in YpeB-C may be unique to this protein. Presumably this will apply also to other *Bacillus* spore YpeB proteins, although this will have to be determined in due course.

Structural insights to YpeB function

Recent work conducted with truncated YpeB proteins in *B. anthracis* spores demonstrated that all three PepSY domains are required for the concomitant stability of YpeB and SleB in dormant spores[6]. In the same study, a variant YpeB protein containing the N-terminal domain and PepSY1 from the C-terminal domain was shown to be relatively stable in the spore, however, SleB abundance was diminished in this strain, indicating that a region beyond PepSY1 is essential to confer stability to SleB in the spore. With this in mind, examination of an electrostatic potential surface representation of YpeB-C, calculated using the Adaptive Poisson-Boltzmann Solver (APBS) Pymol plug-in[16], reveals a large shallow channel or groove that traverses the middle of the molecule. This region is marked by positive surface charge, formed by K345 and K347 from β9 (PepSY2) and K361 and K366 from the short α-helix 3, and may be a candidate site for protein interactions (Figure 2[b]). 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[6]. Analysis of the

corresponding residues in the *B. megaterium* YpeB-C structure reveals that G431 is located in the PepSY3 β13 - β14 connective loop, and it is not immediately obvious why substitution of glycine for alanine would destabilise the protein in the spore. The hydroxyl group from Y411, on the other hand, is sufficiently close (2.7 Å) in the crystal structure to form a hydrogen bond with the backbone carbonyl oxygen atom of I377, which is part of the loop connecting PepSY2 and PepSy3. Accordingly, this intra-molecular interaction may contribute towards the conformational stability of the PepSY3 domain. Alternatively, Y411 is solvent exposed and may be important in interacting with SleB or spore proteases, with disruption to this interaction resulting in proteolytic degradation of YpeB and SleB in the developing spore. Similarly, Y254 can hydrogen bond with the backbone carbonyl oxygen atom of V219, which is located in the loop that precedes α-helix 1. Again, this intra-molecular interaction may contribute towards the structural stability of the domain, or solvent exposed Y254 may be involved in YpeB/SleB-stabilising protein interactions.

To conclude, the crystal structure of the C-terminal domain of the *B. megaterium* YpeB protein has been solved to 1.80 Å. The objective moving forward will be to determine the structure of the full-length protein, with a view to revealing further insights to the precise function of this protein and the molecular mechanisms that underpin its role in regulating SleB activity in the spore.

Acknowledgements

Crystallographic experiments were performed in the Crystallographic X-ray Facility (CXF) at the Department of Biochemistry, University of Cambridge. DYC was supported by CXF.

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.

10. 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. J Bacteriol 2012;194(17):4537-4545.

11. 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 Crystallogr D Biol Crystallogr 2010;66(Pt 2):213-221.

12. Davis IW, Murray LW, Richardson JS, Richardson DC. MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. Nucleic Acids Res 2004;32(Web Server issue):W615-619.

13. Das D, Finn RD, Carlton D, Miller MD, Abdubek P, Astakhova T, Axelrod HL, Bakolitsa C, Chen C, Chiu HJ, Chiu M, Clayton T, Deller MC, Duan L, Ellrott K, Ernst D, Farr CL, Feuerhelm J, Grant JC, Grzechnik A, Han GW, Jaroszewski L, Jin KK, Klock HE, Knuth MW, Kozbial P, Krishna SS, Kumar A, Marciano D, McMullan D, Morse AT, Nigoghossian E, Nopakun A, Okach L, Puckett C, Reyes R, Rife CL, Sefcovic N, Tien HJ, Trame CB, van den Bedem H, Weekes D, Wooten T, Xu Q, Hodgson KO, Wooley J, Elsliger MA, Deacon AM, Godzik A, Lesley SA, Wilson IA. The structure of BVU2987 from *Bacteroides vulgatus* reveals a superfamily of bacterial periplasmic proteins with possible inhibitory function. Acta Crystallogr Sect F Struct Biol Cryst Commun 2010;66(Pt 10):1265-1273.

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

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

16. 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 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[6], 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[16].

Table 1 Crystallographic data collection, phasing and refinement statistics

	SAD dataset K_2PtCl_4 derivative	Native Dataset
Data collection		
Radiation Source	Diamond (UK), I04	Diamond (UK), I04
Wavelength (Å)	1.0721	0.9795
Space group	$C222_1$	$C222_1$
Cell dimensions:		
a, *b*, *c* (Å)	52.77 108.85 85.73	53.21 110.10 85.81
α, β, γ (°)	90.0 90.0 90.0	90.0 90.0 90.0
Resolution (Å)	28.58 – 2.20 (2.32 – 2.20)[1]	43.63 – 1.80 (1.90 – 1.80)[1]
R_{merge}[2] (%)	8.7 (89.8)	6.0 (84.0)
$<I / \sigma(I)>$	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)	
Phasing[3]		
Number of sites found	13	
Overall score (100 * BAYES-CC)	51.7	
Figure of Merit (SOLVE)	0.32	
Figure of Merit (RESOLVE)	0.62	
Density Modification[3]		
R-factor	0.32	
Map skew	0.18	
Corr. of local RMS density	0.88	
Refinement		
Resolution (Å)		46.34 – 1.80
Number of reflections used:		
Total		23,700
R_{free} set		1,997
R_{cryst}[4] / R_{free}[5] (%)		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^2):		
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

[1] The statistics shown in parentheses are for the highest-resolution shell.
[2] $R_{merge} = (\Sigma_{hkl} \Sigma_i | I_i(hkl) - I_{mean}(hkl)|) / \Sigma_{hkl} \Sigma_i I_i(hkl)$.
[3] As calculated by PHENIX software suite
[4] $R_{cryst} = \Sigma_{hkl} ||F_{obs}(hkl)| - |F_{calc}(hkl)|| / \Sigma_{hkl} |F_{obs}(hkl)|$
[5] R_{free} is the same as R_{cryst} but for a random subset of reflections not included in the refinement, about 10% of total reflections.

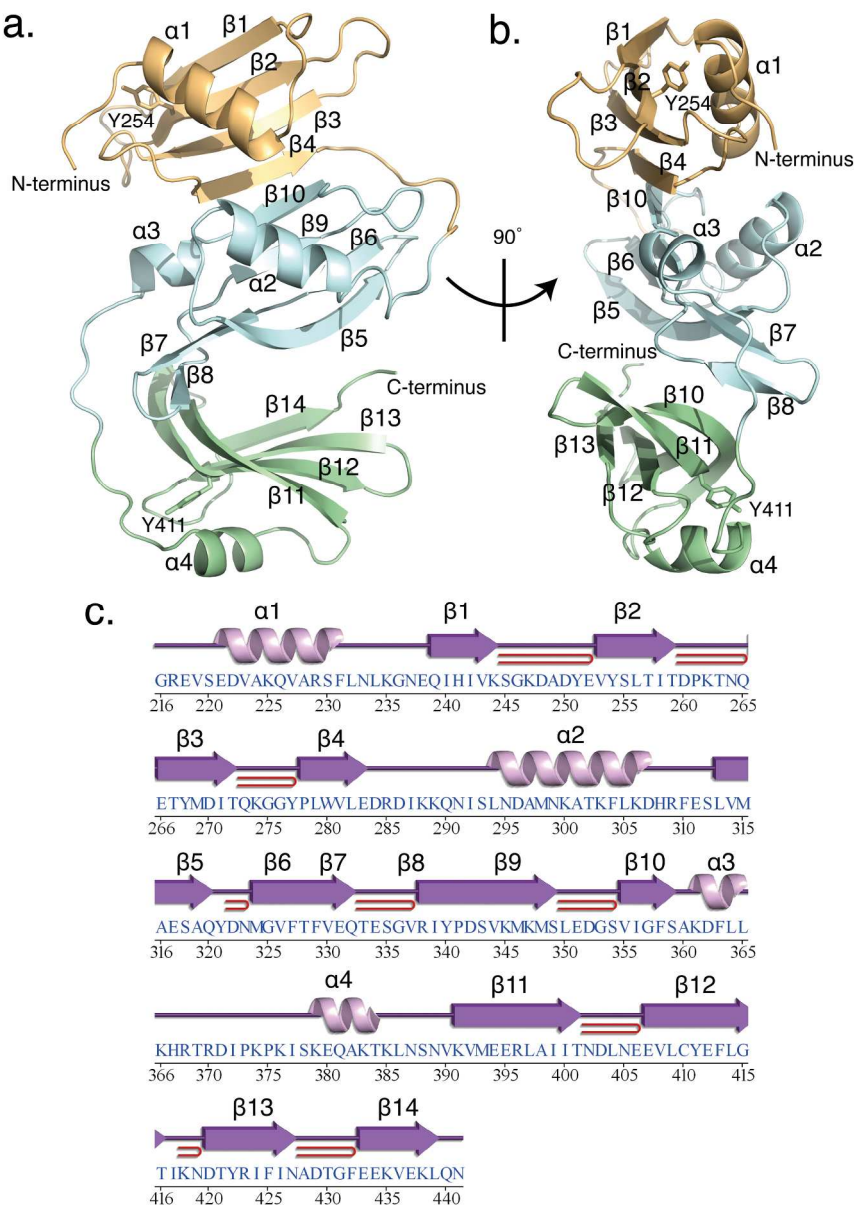

Figure 1. Crystal structure of the C-terminal domain of B. megaterium YpeB. (a) Ribbon representation of the YpeB-C structure. The molecule is rotated 90o counter-clockwise in (b). The molecule is composed of three tandemly arranged PepSY domains (PepSY1, in orange [S220 – D284], PepSY2 in pale blue [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 protein6, 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/).
204x277mm (300 x 300 DPI)

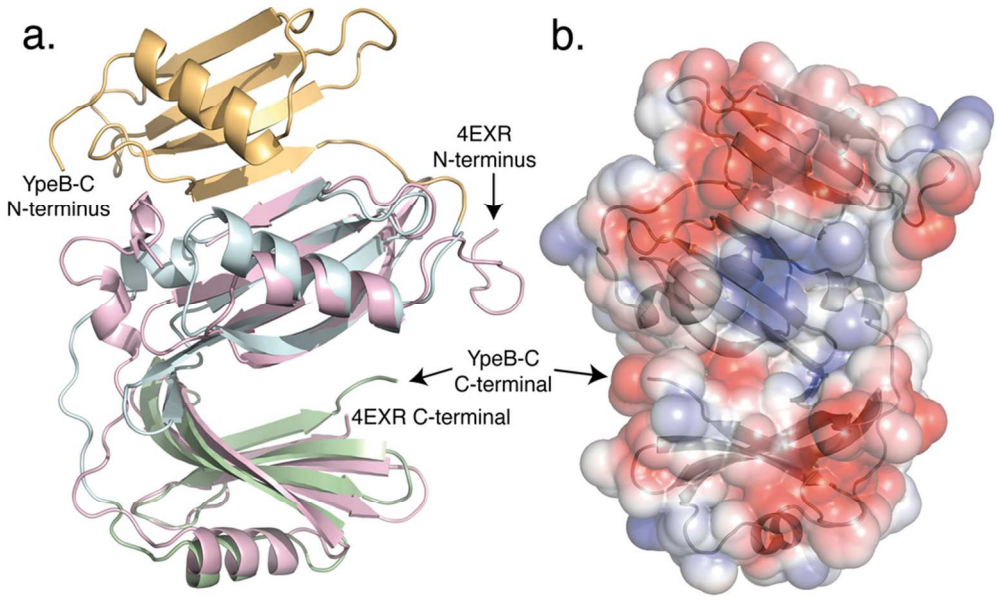

Figure 2. (a) FATCAT flexible pairwise alignment of B. megaterium YpeB-C with Clostridium difficile YpeB-like protein (PDB: 4EXR). The alignment has an optimised r.m.s.d of 2.49 Å when a single twist is introduced in the 4EXR structure, facilitating alignment with the YpeB-C PepSY2 and PepSY3 domains. 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)