MICROBIOLOGY

RESEARCH ARTICLE

Hughes *et al.*, *Microbiology* 2022;168:001274

DOI 10.1099/mic.0.001274

MICROBIOLOGY SOCIETY

OPEN ACCESS

Peptide transport in *Bacillus subtilis* – structure and specificity in the extracellular solute binding proteins OppA and DppE

Adam M. Hughes[1], John F. Darby[1], Eleanor J. Dodson[1], Samuel J. Wilson[1], Johan P. Turkenburg[1], Gavin H. Thomas[2,*] and Anthony J. Wilkinson[1,*]

Abstract

Peptide transporters play important nutritional and cell signalling roles in *Bacillus subtilis*, which are pronounced during stationary phase adaptations and development. Three high-affinity ATP-binding cassette (ABC) family transporters are involved in peptide uptake – the oligopeptide permease (Opp), another peptide permease (App) and a less well-characterized dipeptide permease (Dpp). Here we report crystal structures of the extracellular substrate binding proteins, OppA and DppE, which serve the Opp and Dpp systems, respectively. The structure of OppA was determined in complex with endogenous peptides, modelled as Ser-Asn-Ser-Ser, and with the sporulation-promoting peptide Ser-Arg-Asn-Val-Thr, which bind with K_d values of 0.4 and 2 µM, respectively, as measured by isothermal titration calorimetry. Differential scanning fluorescence experiments with a wider panel of ligands showed that OppA has highest affinity for tetra- and penta-peptides. The structure of DppE revealed the unexpected presence of a murein tripeptide (MTP) ligand, L-Ala-D-Glu-*meso*-DAP, in the peptide binding groove. The mode of MTP binding in DppE is different to that observed in the murein peptide binding protein, MppA, from *Escherichia coli*, suggesting independent evolution of these proteins from an OppA-like precursor. The presence of MTP in DppE points to a role for Dpp in the uptake and recycling of cell wall peptides, a conclusion that is supported by analysis of the genomic context of *dpp*, which revealed adjacent genes encoding enzymes involved in muropeptide catabolism in a gene organization that is widely conserved in *Firmicutes*.

INTRODUCTION

Sporulation is a complex adaptive process enabling long-term survival in adverse conditions and in the bacterium *Bacillus subtilis* is initiated when a threshold concentration of the phosphorylated form of the transcription factor Spo0A (Spo0A~P) is achieved [1]. Phosphorylation promotes dimerization of Spo0A [2] and binding to numerous promoters, ultimately switching on scores of genes required for sporulation [3]. Spo0A is the most downstream element of an expanded two-component signalling system termed the sporulation phosphorelay [4] (Fig. 1). The phosphorelay is fed by up to five sensor kinases (KinA–KinE) which detect and respond to environmental and intracellular signals by using ATP to autophosphorylate on a conserved histidine [5]. The phosphoryl group is then relayed, via Spo0F and Spo0B, onto an aspartate residue in Spo0A. In opposition, the phosphorelay is drained by protein phosphatases belonging to the Rap and Spo0E families that dephosphorylate Spo0F~P and Spo0A~P respectively [6, 7]. This signal integration system ensures that the commitment to the sporulation pathway is made only when other survival strategies have been exhausted.

The identities of the molecular stimuli which lead to activation of the sensor kinases remain unknown, though pyruvate was discovered bound to the distal of two PAS domains in a crystal structure of KinD [8]. In contrast, peptide regulation of Rap phosphatase activity was elucidated early on following the discovery that the *spo0K* lesion [9], which blocks the first step in spore

Received 10 October 2022; Accepted 10 November 2022; Published 01 December 2022

Author affiliations: [1]Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5DD, UK; [2]Department of Biology, University of York, York YO10 5DD, UK.

***Correspondence:** Gavin H. Thomas, gavin.thomas@york.ac.uk; Anthony J. Wilkinson, tony.wilkinson@york.ac.uk

Keywords: sporulation; peptide transport; *Bacillus subtilis*; murein peptide; DppE; OppA.

Abbreviations: ABC, ATP binding cassette; App, another peptide permease; CD, circular dichroism; Dpp, dipeptidepermease; ITC, isothermal titration calorimetry; MMT, malic acid, 2-(N-morpholino)ethanesulfonic acid, Tris; MTP, murein tripeptide; Opp, oligopeptide permease; SBP, substrate binding protein; SPG, succinic acid, sodium dihydrogen phosphate, glycine.

Seven supplementary figures are available with the online version of this article.

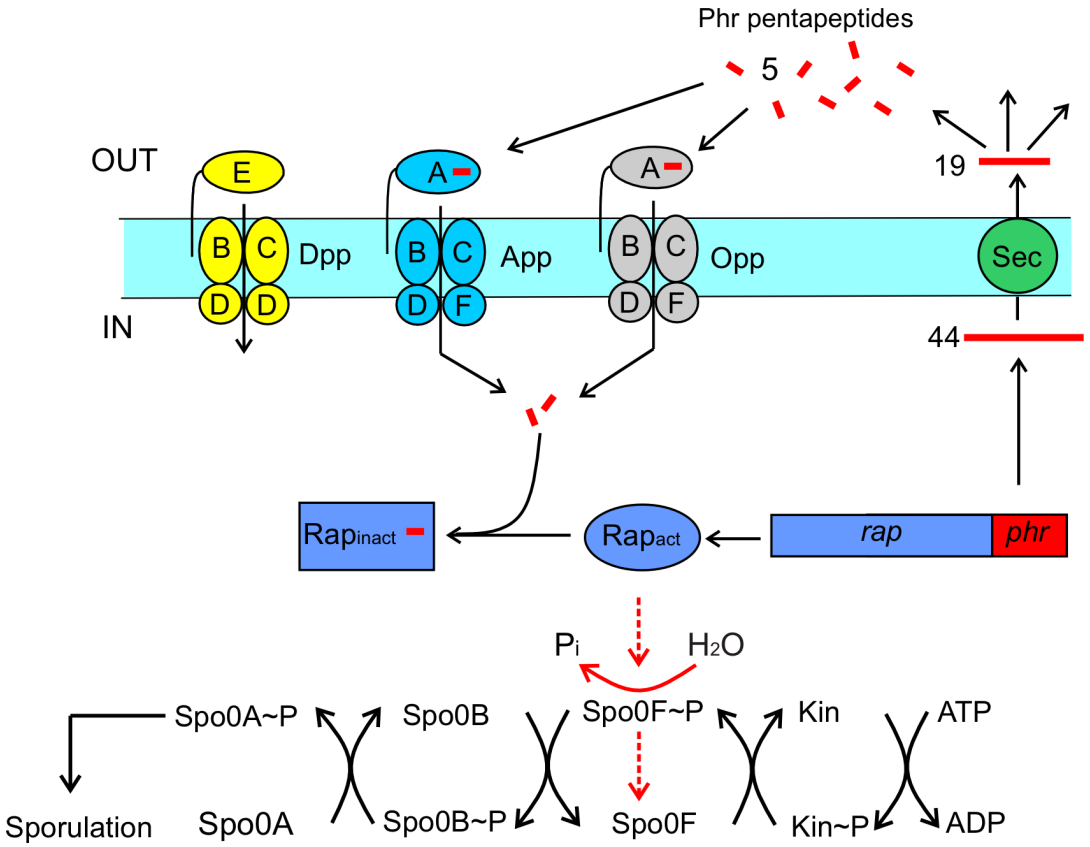


Fig. 1. Peptide transport and sporulation signalling in *B. subtilis*. *B. subtilis* is served by three ATP-binding cassette (ABC)-type peptide uptake systems termed dipeptide permease (Dpp), oligopeptide permease (Opp) and another peptide permease (App). Each consists of lipid-anchored solute binding proteins (DppE, OppA and AppA respectively), membrane spanning channels (DppBC, OppBC and AppBC) and ATPases ($DppD_2$, OppDF and AppDF) associated with the cytoplasmic face of the membrane. In the sporulation phosphorelay shown at the bottom, up to five sensor kinases (Kin) autophosphorylate and relay a phosphoryl group via Spo0F and Spo0B to Spo0A. Flux through the phosphorelay is reduced by the action of Rap phosphatases (blue) which dephosphorylate Spo0F~P (red arrows). The *phr* gene (*phrA* or *phrE*) situated downstream of a cognate *rap* phosphatase gene encodes a 44 residue polypeptide (red) which is exported from the cell by the Sec system with removal of a secretion signal peptide to yield a 19-mer. Under circumstances that favour sporulation, environmental proteases cleave the peptide producing short oligopeptides such as the pentapeptides shown, which may be imported into the cell by either App or Opp. The imported peptides bind to the cognate Rap phosphatase and prevent it from dephosphorylating Spo0F phosphate (Spo0F~P). This allows increased flux through the sporulation phosphorelay and the accumulation of Spo0A~P so that sporulation may commence.

formation, maps to the oligopeptide permease operon, *opp* [10, 11]. Opp is an ATP-binding cassette (ABC) transporter which allows the uptake of peptides of variable length and diverse sequence. The sporulation defects of *opp* mutants can be suppressed by deletions in the *rapA* and *rapB* genes [12]. Genes encoding Rap phosphatases are often adjacent to and transcriptionally coupled to *phr* genes (Fig. 1). Phr proteins are synthesized as precursors comprising approximately 45 residues which are secreted to produce ~20 residue extracellular peptides. Subsequent processing generates pentapeptides, which, following re-uptake by the Opp system, inhibit their cognate Rap phosphatases [13, 14]. Thus, the PhrA (ARNQT) and PhrE (SRNVT) pentapeptides specifically inhibit the phosphatases RapA and RapE respectively. Rap phosphatase inhibition leads to increased flux through the sporulation phosphorelay, accumulation of Spo0A~P and transcription of genes needed for sporulation.

ABC uptake systems in bacteria, such as Opp, comprise five components (Fig. 1). Two form a channel that spans the cell membrane and two are ATPases situated on the cytoplasmic face of the membrane. The fifth component is a substrate binding protein (SBP) which is lipidated and anchored on the extracellular face of the membrane in Gram-positive bacteria. This receptor protein – OppA in the Opp system – defines the specificity of the transporter, capturing extracellular solutes and delivering them to the membrane components for uptake into the cell. Since peptides play an important nutritional role in *B. subtilis*, Opp is expected to handle peptides of variable length and sequence, so that in collaboration with the two other high-affinity peptide transporters, the dipeptide permease, Dpp, and another peptide permease, App, a broad repertoire of peptides can be taken up. The SBPs of the three systems, OppA, AppA and DppE, are sequence- and structurally related and belong to the cluster C structural grouping of ABC SBPs [15, 16].

Table 1. Primers for amplification of sequences encoding extracellular components of *B. subtilis* DppE and OppA

Primer identifier	Sequence*
F_*Bs*OppA*	5′ tccagggaccagcaAAAGGAAAGACGACACTTAACATTAATATTAAAACTGA
R_*Bs*OppA*	5′ tgaggagaaggcgcgTTATTTAAAATATGCGTTTCTGAAATAAACCTCACC
F_*Bs*DppE*	5′ ccagggaccagcaAGCGGAGAAAAGGTGCTGTATGTAAATAAT
R_*Bs*DppE*	5′ tgaggagaaggcgcgTCAGTTTTTATCTGCCCATTTTAAATCGATATAGC
F_YSBLLIC3C	5′ CGCGCCTTCTCCTCACATATGGCTAGC
R_YSBLLIC3C	5′ TGCTGGTCCCTGGAACAGAACTTCC
T7promoter	5′ TAATACGACTCACTATAGGG
T7terminator	5′ GCTAGTTATTGCTCAGCGG

*Sequence complemetary to the pETYSBLLIC vector is in lowercase in contrast to gene-specific sequence in uppercase.

App was discovered as another peptide permease in strains defective in *opp* in which sporulation is restored [17]. Fluorescence titration experiments revealed that its receptor protein, AppA, could bind to a nonapeptide [18], an observation supported by a crystal structure of AppA which revealed the presence of an endogenous nonapeptide ligand [19]. Dpp was first identified following analysis of a transcript that accumulated under conditions which favour sporulation [20]. Sequencing of the associated operon (earlier called *dciA*) revealed a putative ABC-type peptide permease. As null mutations in *dciAE* abolished the capacity of a proline auxotroph to grow on Pro-Gly as the sole proline source, the permease was classified as a dipeptide transporter and renamed Dpp. DppE is the SBP component, with DppB and DppC serving as the transmembrane channel elements and DppD serving as the ATPase component (Fig. 1). This is exceptional for cluster C type ABC transporters, which normally comprise five separate genes, one for each of the components. Analysis of *dpp* expression led to the discovery of its regulation by the transcription factor CodY (Control of dpp Y), a branched chain amino acid and GTP sensor which is now recognized as a global regulator of transcription [21–24]. During vegetative growth, CodY represses genes that are required under conditions of nutrient limitation, suggesting a role for the Dpp system in nutrient scavenging.

Collectively, these studies suggest that the three high-affinity ABC peptide transporters of *B. subtilis* have complementary specificities with respect to the length of the peptides they transport: Dpp handling short substrates (principally dipeptides), Opp covering an intermediate range including pentapeptides and App handling longer peptides. To address the specificity of these peptide uptake systems, we determined the structures of OppA and DppE from *B. subtilis* and measured ligand binding to OppA. Structures of OppA were determined in complex with an endogenous tetrapeptide and with the PhrE peptide, SRNVT, revealing a similar ligand binding site to OppA from *Salmonella typhimurium*, featuring strong interactions with the peptide main chain and large, hydrated side chain pockets [25]. Remarkably, the DppE structure revealed unambiguous electron density for a murein tripeptide and further analysis of its genome context suggests that it plays a dedicated role in the recycling of cell wall peptides.

METHODS

Plasmid construction

DNA fragments encompassing the coding sequences of *dppE* and *oppA* were amplified from *B. subtilis* IG20 (168*trpC2*) genomic DNA by PCR using the oligonucleotide pairs F_BsDppE/R_BsDppE and F_BsOppA-F/R_BsOppA (Table 1). The PCR primers were designed to incorporate specific flanking sequences to facilitate cloning into pETYSBLLIC3C [26], a derivative of pET-28a. In total 50 pmol of each primer pair was mixed with 25 ng of template DNA in a 50 µl reaction containing 0.2 mM dNTPs and 1 Unit of KOD Hot Start Polymerase (Novagen). The expected 1540 and 1559 bp PCR products amplified from *dppE* and *oppA* respectively were purified by PCR clean up (QIAquick PCR Clean up) and each was mixed with linearized pETYSBLLIC3C vector obtained by PCR using the primers F_YSBLLIC3C and R_YSBLLIC3C (Table 1). HiFi reactions (New England Biolabs) were carried out following the supplier's instructions and the ligation products were introduced into *Escherichia coli* NEB 5α competent cells. Plasmids recovered from kanamycin-resistant transformants were sequenced (Eurofins) using the T7promoter and T7terminator primers (Table 1). The verified plasmids pAH101 and pAH102 encode proteins consisting of residues 19–521 of *B. subtilis* DppE and 17–525 of *B. subtilis* OppA respectively, each N-terminally fused to a human rhinovirus 3C cleavable hexahistidine tag.

Protein production and purification

E. coli BL21 (DE3) cells harbouring pAH101 or pAH102 were grown as 1 litre LB cultures supplemented with kanamycin (30 $\mu g ml^{-1}$) at 37 °C to an OD_{600} of 0.6–0.8 before induction of recombinant protein production by addition of IPTG to 1 mM. Four hours later, cells were harvested by centrifugation at 5000 r.p.m. for 50 min using an F10S rotor.

Recombinant DppE and OppA were purified by similar procedures as follows. Cells were resuspended in 35 ml buffer A (50 mM Tris/HCl, 500 mM NaCl, 20 mM imidazole pH 8.0) and a complete EDTA-free protease inhibitor cocktail was added before sonication (five cycles at 16 kHz for 30 s, at 1 min intervals at 4 °C). Cell debris was removed by centrifugation (15 000 r.p.m. for 50 min in an SS34 rotor) and the supernatant was applied to a nickel-charged HisTrap FF column (GE Healthcare) pre-equilibrated with buffer A. This column was developed using a 20–500 mM imidazole gradient in buffer A. Fractions were analysed by denaturing polyacrylamide gel electrophoresis (SDS-PAGE), and those enriched in a recombinant protein of apparent M_r=55–60 kDa were combined and dialysed overnight against buffer A in the presence of HRV 3C protease (1:100 ratio of protease : substrate protein) to remove the N-terminal hexa-histidine tag. The digestion products were passed through a second nickel affinity column and the flow-through fraction was concentrated by pressure membrane filtration (30 000 Da cut off) and subjected to size exclusion chromatography using an S200 16/600 column. Protein-containing fractions were analysed by SDS-PAGE, combined and concentrated. The yields of the purified proteins were typically 20 mg per litre of cell culture.

To obtain unliganded DppE and unliganded OppA, the Ni-NTA column chromatography purification procedure was modified so that after the binding and washing steps, the column was washed with buffer B (buffer A containing 2 M guanidine hydrochloride), to partially unfold the immobilized protein and allow dissociation of endogenous ligands [27]. After washing with 10 column volumes of buffer B, the concentration of guanidinium hydrochloride was decreased to zero in a series of five steps. The protein was then eluted from the column by applying an increasing imidazole concentration gradient as before.

Circular dichroism (CD) spectroscopy

CD spectra for purified proteins were recorded on a J-810 spectropolarimeter (Jasco). Protein samples at 0.2 mg ml^{-1} were analysed in 20 mM Tris pH 8.0 and 50 mM NaCl buffer. Spectra were recorded at 20 °C in a 1 mm path-length quartz cuvette (Starna) over the wavelength range 190–260 nm at 200 nm min^{-1} with 0.5 nm pitch. The spectra were analysed using the SpectraManager version 1.53.00 (Jasco) software.

Crystallization and structure determination

Crystallization experiments were performed in MRC-Wilden 96-well plates using the commercial screens PACT (Molecular Dimensions) and Index (Hampton Research) set up using Hydra 96 and Mosquito liquid handling systems to dispense well and drop solutions respectively. Optimization of initial crystal hits was carried out in 24-well hanging drops set up manually.

OppA

Crystals of OppA were obtained from hanging drops composed of 2 µl of reservoir solution consisting of 0.1 M MMT (malic acid, 3-(N-morpholino)ethanesulfonic acid, Tris), 22.5% PEG 1500 and 2.5% DMSO pH 8.0 and 2 µl protein at 18 $mg ml^{-1}$ with crystal optimization following a seeding protocol. A single crystal was captured in a nylon loop and cryo-protected with mother liquor containing 15% glycerol. X-ray diffraction data were collected at Diamond Light Source on beamline i03 to a resolution of 1.5 Å. The crystal structure of *Bs*OppA was solved by molecular replacement in the program PHASER with the search model 1RKM, the structure of the unliganded form of *S. typhimurium* OppA which shows a 35% sequence identity to *Bs*OppA. There are two molecules of *Bs*OppA in the asymmetric unit with residues 17–525 well defined in the electron density map of chain A and residues 18–525 well defined in chain B. Iterative rounds of model building using COOT [28] and refinement in REFMAC5 [29] followed initial model building. Data collection and refinement statistics are shown in Tables 2 and 3.

Crystals of BsOppA in complex with Ser-Arg-Asn-Val-Thr where grown from the products of an isothermal titration calorimetry experiment described below. This recovered material was fed into a crystallization screening experiment as described above. A single crystal of BsOppA-SRNVT grown from 0.1 M SPG (succinic acid, sodium dihydrogen phosphate, glycine), 25% (w/v) PEG 1500 pH 4.0, was harvested and cryocooled in liquid nitrogen. X-ray diffraction data extending to 1.9 Å spacing were collected at Diamond Light Source on beamline i04-1. The crystal belongs to space group C2 with one protein molecule in the asymmetric unit. The structure was resolved by molecular replacement using the coordinates of BsOppA-SNSS as the search model. Electron density defining the Ser-Arg-Asn-Val-Thr peptide was clearly observed in the peptide binding cavity.

DppE

Crystals of DppE suitable for X-ray analysis were obtained from hanging drops formed by mixing 1 µl of reservoir solution containing 0.1 M Bistris-propane pH 8.5, 0.4 M $MgCl_2$, 22.5% PEG 3350 and 2.5% DMSO with 1 µl of protein at 13 $mg ml^{-1}$. Crystals were transferred to solutions of mother liquor containing 30% ethylene glycol and cryocooled in liquid nitrogen. High-resolution data extending to 1.5 Å resolution were collected from a single crystal at the Diamond Light Source on beamline i03. The

Table 2. Data collection statistics

	OppA-SNSS (8ARN)	OppA-PhrE (8ARE)	DppE (8AZB)	DppE-MTP (8AY0)
Diffraction source	i03	i04	i02	i04
Wavelength (Å)	0.8200	0.9159	0.9795	0.9795
Temperature (K)	100	100	100	100
Space group	C2	I2	$P2_12_12_1$	P1
a, b, c (Å)	101.5, 65.9, 153.3	83.9, 63.7, 100.1	54.5, 91.1, 106.7	60.9, 61.3, 124.1
a, b, g (°)	90, 101.0, 90	90, 114.4, 90	90, 90, 90	78.1, 82.7, 61.6
Resolution range (Å)*	54.96–1.50 (1.53–1.50)	52.21–1.90 (1.94–1.90)	46.09–1.40 (1.42–1.40)	53.17–1.51 (1.54–1.51)
Total reflections*	661094 (32583)	155874 (9854)	777527 (22361)	512020 (25224)
Unique reflections*	158694 (7829)	37997 (2442)	104821 (5040)	233391 (11306)
Completeness (%)*	99.9 (99.9)	100.0 (100.0)	99.6 (98.2)	96.3 (93.8)
Multiplicity*	4.2 (4.2)	4.1 (4.0)	7.4 (4.4)	2.2 (2.2)
$\langle I/\Sigma(I)\rangle$*	12.6 (1.6)	3.2 (1.2)	20.9 (3.4)	8.4 (1.1)
R_{pim}*,†	0.029 (0.479)	0.149 (0.688)	0.026 (0.197)	0.053 (0.759)
$CC_{1/2}$*	0.998 (0.609)	0.955 (0.758)	0.999 (0.886)	0.998 (0.572)
Overall B Factor Wilson plot (Å^2)	20.79	16.91	13.17	17.45

*Values in parentheses correspond to the outer resolution shell.
†R_{pim}=precision-indicating merging R-factor; $\sum_{hkl}[1/(n-1)]^{1/2}\sum_i|I_i - \langle I\rangle|/\sum_{hkl}\sum_i I_i$ where I_i is the intensity of the *i*th measurement of a reflection with indices *hkl* and <*I*> is the statistically weighted average reflection intensity.

data were processed using the program Xia2 [30]. The crystal belonged to space group P1 with three molecules in the asymmetric unit. PHASER [31] and BUCCANEER [32] were used to resolve the structure by molecular replacement using the *Enterococcus faecalis* PrgZ pheromone receptor (coordinate set 4FAJ [33]) as the search model. PrgZ and DppE have 33% sequence identity. Iterative rounds of manual refinement using COOT [34] and REFMAC5 [29, 35] were undertaken to complete model building.

Ligand-free DppE was crystallized in a sitting drop formed by mixing 150 nl of the unliganded protein at 9 mgml^{-1} with 150 nl of 0.1 M MIB buffer pH 4.0, 25% (w/v), PEG 1500. A crystal was captured from the nanodrop in a nylon loop and cooled in liquid nitrogen. Diffraction data extending to 1.4 Å were collected on Diamond Light Source beamline i02. Structure determination was carried out using auto-processed data obtained from Xia2. The space group of the crystal is $P2_12_12_1$ with one molecule in the asymmetric unit. Molecular replacement was performed with PHASER using the DppE coordinate set truncated of ligand and water molecules, as the search model. Anticipating a probable domain opening, two searches were performed with models related by splitting the protein chain at residue 243. Data collection and refinement statistics are shown in Tables 2 and 3.

Ligand binding – OppA

Thermal shift assays were performed with *Bs*OppA and a variety of peptide ligands. Initial experiments were performed in 50 mM Tris, 150 mM NaCl pH 8.0 to establish the optimum protein and dye concentrations for these assays (Fig. S1 available with the online version of this article). Peptide binding experiments were subsequently performed in triplicate with protein at 20 µM and SYPRO Orange dye present in five-fold excess. A 100-fold molar excess of ligand was used for all experiments with ligands dissolved in the same buffer as the protein. A Stratagene Mx3005P real-time quantitative PCR instrument (Qiagen) was used with fluorescence excitation at 517 nm and emission measured at 585 nm following 1 °C/30 s temperature increments over the range 25–96 °C. Data analysis was performed using JSTA.

Isothermal titration calorimetry (ITC) was carried out with *Bs*OppA in 50 mM Tris, 150 mM NaCl, pH 8.0 buffer. The protein was present at a concentration of 90 µM. Titrations were performed with a stock concentration of ligand at 1.6 mM over the course of 19×0.5 µl injections with stirring at 750 r.p.m. at 25 °C. Experiments were performed using a Micro-Cal Auto-iTC200 instrument (Malvern Instruments) with data analysed using MicroCal PEAQ-ITC software. All experiments were performed in triplicate.

Table 3. Structure refinement

	OppA-SNSS (8ARN)	OppA-PhrE (8ARE)	DppE (8AZB)	DppE-MTP (8AY0)
Resolution range (Å)*	55.02–1.50 (1.54–1.50)	52.26–1.90 (1.95–1.90)	46.09–1.40 (1.44–1.40)	60.74–1.51 (1.55–1.51)
Completeness (%)*	99.9 (99.9)	99.4 (99.6)	99.5 (98.5)	96.3 (94.0)
No. of reflections, working set*	150844 (11102)	35978 (2633)	99479 (7225)	221900 (16044)
No. of reflections, test set*	7849 (574)	1811 (152)	5232 (354)	11490 (789)
Final R_{cryst} *,†	0.177 (0.296)	0.252 (0.560)	0.182 (0.226)	0.185, 0.322
Final R_{free} *,‡	0.207 (0.297)	0.308 (0.561)	0.208 (0.267)	0.221, 0.332
No. of non-H atoms				
Protein	8131	4032	4092	11932
Ligand	54	40	–	81
H_2O	884	220	513	1206
Ions/additives	–	5	–	96
Total	9069	4297	4605	13315
R.m.s. deviations				
Bonds (Å)	0.010	0.007	0.012	0.008
Angles (°)	1.614	1.391	1.810	1.500
Average B factors (Å^2)				
Protein	24.88	17.14	18.14	27.26
Ligand	34.53	15.51		22.37
H_2O	35.53	17.78	26.91	34.88
Ions/additives	–	39.70	–	43.18
Ramachandran plot				
Favoured regions (%)	98.3	97.1	98.0	97.8
Additionally allowed (%)	1.7	2.8	2.0	2.4
Outliers (%)	0	0.2	0	0

*Values in parentheses correspond to the outer resolution shell.
†$R_{cryst} = \sum||Fo| - |Fo||/\sum|Fo|$ where *Fo* and *Fc* are the observed and calculated structure factor amplitudes, respectively.
‡R_{free} is the R_{cryst} calculated with 5% of the reflections chosen at random and omitted from the refinement.

Genome analysis

We used the SEED viewer to identify gene clusters homologous to the *B. subtilis dpp* region [36].

RESULTS

OppA and DppE are expected to be lipoproteins anchored to the outer leaflet of the cytoplasmic membrane of *B. subtilis* following secretion and processing of the precursor proteins and their fatty acid acylation on Cys1 of the mature proteins. For soluble recombinant protein production in *E. coli* and subsequent crystallization, we amplified truncated coding sequences to remove the signal peptides, the amino terminal cysteines and further residues at the amino terminus which we judged to be intrinsically unstructured based on sequence and structure alignments. Thus, purified OppA and DppE consist of residues 17–525 and 19–521 respectively, each with a vestigial GPA sequence at the N terminus produced by HRV 3C protease cleavage.

To establish the integrity of the purified proteins, their masses were determined by electrospray ionization MS to be 57902 and 57795 Da for OppA and DppE respectively, in agreement with the expected masses calculated from the sequences of 57934 and 57796 Da. CD spectra recorded using a Jasco J810 spectrophotometer suggested that both proteins possess a complement of

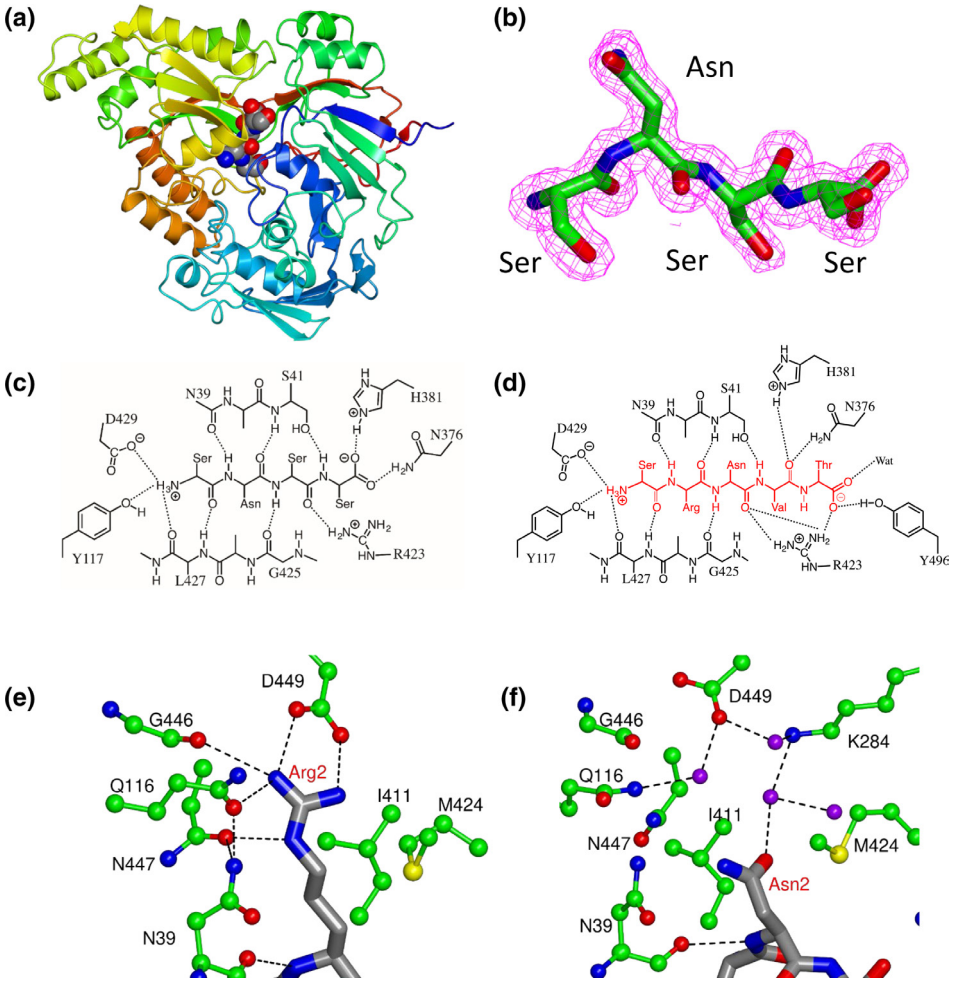


Fig. 2. The structure of OppA from *B. subtilis*. (a) Ribbon tracing of chain A from the crystal structure of OppA. The chain is colour ramped from the N-terminus (blue) to the C-terminus (red). The bound ligand is displayed as spheres coloured by atom type. (b) Electron density maps in the interdomain region of the OppA structure displayed on the ligand from the refined structure. (c and d) Schematic representation of the polar interactions (dashed lines) of OppA with the main chain of (c) the tetrapeptide ligand Ser-Asn-Ser-Ser and (d) the pentapeptide Ser-Arg-Gln-Val-Thr. (e and f) The position 2 side chain binding pocket showing the interactions of the Arg2 (e) and Asn2 (f) side chains of the peptide ligand (cylinder format with grey carbons) of SRNVT and SNSS respectively with the surrounding protein (ball-and-stick format with green carbons). The Arg2 side chain makes abundant direct polar interactions with the protein and displaces four water molecules (purple spheres) which are present when Asn2 is bound.

secondary structure elements consistent with a natively folded protein (Fig. S2). The preservation of this secondary structure following guanidinium hydrochloride treatment to produce the ligand-free protein was also confirmed by CD spectroscopy (Fig. S2).

Structure determination and overall structure of OppA

To elucidate the molecular basis of peptide recognition by *B. subtilis* OppA (BsOppA) we grew crystals of the recombinant protein and resolved and refined its structure with diffraction data extending to 1.5 Å spacing (Tables 2 and 3). The crystals have two molecules in the asymmetric unit and the main chain electron density is contiguous from residue K17 to residue K525 in chain A (Fig. 2a) and K19 to K525 in chain B. The A and B molecules of the asymmetric unit are very closely superimposable with pairwise positional rmsΔ values of 0.2 Å for 507 equivalent Cα atoms. Each chain consists of two lobes with lobe 1 comprising residues 17–276 and 497–525 and lobe 2 made up of residues 277–496 (Figs 2a and S3). A structural similarity search of the protein database in the program PDBeFOLD revealed a number of high scoring hits with Z-values of ~20, and rmsΔ values of ~1.5 Å for ~480 aligned residues. The most closely similar protein structures to BsOppA include PrgZ from *Enterococcus faecalis* (PDB Entry: 4faj), MppA from *Escherichia coli* (3o9p) and the OppA proteins of *Listeria monocytogenes* (5kzt), *Borrelia burgdoferi* (4gl8), *Burkholdderia pseudomallei* (3zs6), *S. typhimurium* (1b58) and *Escherichia coli* (3tcf).

The peptide binding pocket of OppA contains a tetrapeptide

Upon completion of the main chain tracing, we observed unoccupied electron density at the core of both protein chains in the putative peptide binding site (Fig. S4A). We would expect BsOppA to purify with a heterogeneous mixture of peptides bound in the pocket, reflecting the anticipated function of the protein in the uptake of general peptides. In the crystal, this would give rise to electron density that is averaged over these different peptides. Guided by the shape of this electron density, we modelled a tetrapeptide $Ser^{1}Asn^{2}Ser^{3}Ser^{4}$ as the most satisfactory fit to the experimental data (Fig. 2b). There is positive-difference density extending beyond the modelled side chains at positions 1 and 2, indicating that larger side chains are present at these positions. As previously observed for OppA of *S. typhimurium,* the hydrogen bonding and electrostatic potential of the peptide main chain is fulfilled by complementary interactions with the protein (Fig. 2c). Thus, the α-amino group of the ligand forms an ion-pair with the carboxylate of Asp429 and additional charge-dipole interactions with the phenolic hydroxyl of Tyr117 and the >C=O of Leu427. At the other end of the ligand, the α-carboxylate group forms what is expected to be an ion pairing interaction with the side chain of His381 and a further polar interaction to the side chain of Asn376 (Fig. 2c). Meanwhile, the polar groups of the first and second peptide bonds form β-sheet-like interactions with residues 427–425 (anti-parallel) and 39–41 (parallel). The carbonyl and amide groups of the third peptide bond form polar interactions with the side chains of Arg423 and Ser41 respectively. Prominent in the binding site too are the indole side chains of Trp426 and Trp407 whose rings lie parallel to the plane of the first and third peptide bonds respectively. The side chains of Arg423 and Lys27 may contribute favourably to binding of the tetrapeptide's α-carboxylate since their side chain guanidinium/amino nitrogens respectively are situated 3.5 Å from the carboxylate oxygens. With respect to the binding of peptides of different lengths, these side chains are also well placed to form salt bridges with the terminal carboxylate of tripeptide and pentapeptide ligands respectively.

Consistent with the notion that Opp in *B. subtilis* serves a nutritional role as a general peptide transporter and that OppA is a sequence-independent peptide binding protein, there are no strong polar interactions between the protein and the side chains of the tetrapeptide ligand which would confer specificity. Instead, buried water molecules mediate interactions between the peptide side chains and the protein. These water molecules act as molecular cushions, filling volume not occupied by the ligand and possessing hydrogen bond donor and acceptor character which can be adaptably deployed according to the ligand's character.

Specificity and affinity of peptide binding in BsOppA

To measure binding to BsOppA of the Ser-Asn-Ser-Ser peptide that was modelled in the crystal structure, we used a thermal shift (differential scanning fluorimetry) assay in which *Bs*OppA was mixed with the dye, SYPRO-Orange, and the fluorescence emission was monitored as a function of temperature in the presence and absence of the peptide. The dye binds non-specifically to hydrophobic surfaces exposed as the protein unfolds and this is monitored as an enhancement of fluorescence (Fig. S1). The unfolding process can be quantified by a melting temperature T_m derived from plots of relative fluorescence versus temperature [37]. A shift in T_m in the presence of the peptide signals binding. As shown in Table 4 and Fig. S5, we observed a large positive ΔT_m=16 °C in the presence of synthetic Ser-Asn-Ser-Ser. In a follow-up study using ITC, this peptide was shown to bind to *Bs*OppA with high affinity, K_d=0.4 µM, in an enthalpically driven process (Table 5).

The binding of three other tetrapeptides was tested in the thermal shift assay. Lys-Lys-Lys-Lys and Val-Ala-Pro-Gly binding gave ΔT_m values of 4 and 8°C respectively (Table 4) while no shift in T_m was observed for Asp-Asp-Asp-Asp. ITC measurements revealed a K_d value of 10 µM for the Val-Ala-Pro-Gly ligand (Table 5). Next we assayed the binding to BsOppA of Ala-Arg-Asn-Gln-Thr and Ser-Arg-Asn-Val-Thr, which are known substrates of Opp in *B. subtilis* [13, 14]. These peptides are derived from the phosphatase regulators PhrA and PhrE respectively. As shown in Table 4, in the presence of the PhrA and PhrE peptides, unfolding of *Bs*OppA exhibited 12 and 20 °C increases in T_m respectively. ITC measurements revealed K_d values of 23 µM for ARNQT and 2 µM for SRVNT in strikingly entropically driven binding processes (Table 5; Fig. S6).

Table 4. Peptide binding* to *Bs*OppA measured by thermal shift analysis

Peptide	T_m (°C)	SD†	ΔT_m (°C)
None	50.4	0.4	–
KKKK	54.5	0.1	4.2
VAPG	59.1	0.2	8.7
ARNQT (PhrA)	62.4	0.3	12.0
SNSS	66.8	0.2	16.4
SRVNT (PhrE)	71.0	0.2	20.6

*Experiments carried out at 298 K.
†Based on three replicates.

Table 5. Peptide binding* to *Bs*OppA measured by ITC

Peptide	K_d (μM)	*N*	Δ*H* (kJ mol^{-1})	−*T*Δ*S* (kJ mol^{-1})	Δ*G* (kJ mol^{-1})
VAPG	9.97±3.00	0.78±0.02	13.13±1.10	−41.77±0.47	−28.63±0.84
ARNQT (PhrA)	22.53±3.01	1.36±0.03	17.30±0.87	−43.83±0.57	−26.57±0.38
SNSS	0.43±0.03	0.91±0.08	−29.93±0.28	−6.44±0.44	−36.35±0.21
SRVNT (PhrE)	2.16±02.8	0.80±0.01	19.87±0.67	−52.23±0.29	−32.37±0.38

*Experiments carried out in triplicate at 298 K.

The binding of three longer peptides was investigated, the heptapeptide Gly-Arg-Gly-Asp-Ser-Pro-Lys, the octapeptide Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys and the decapeptide Arg-Gly-Asp-Ser-Pro-Ala-Ser-Ser-Lys-Leu. We found no evidence of peptide binding in the thermal shift assays at peptide concentrations up to 2 mM.

The findings described here are in accord with previous peptide binding studies performed with BsOppA in which changes in intrinsic protein fluorescence were monitored [18]. This study showed binding of the tripeptides Arg-Gly-Gly, the tetrapeptide Gly-Leu-Gly-Leu and the pentapeptide Ser-Leu-Ser-Gln-Ser with K_d values of 50, 5 and 1 μM respectively. No binding of the longer peptides tested was observed. Together, with the limited dataset presented here, it appears that BsOppA has highest affinity for tetrapeptides and pentapeptides.

Structure of BsOppA bound to the PhrE-derived peptide

The products of the ITC experiment measuring binding of the PhrE-derived peptide Ser-Arg-Asn-Val-Thr to BsOppA were recovered and fed into a crystallization screening experiment. This yielded a different crystal form whose structure was resolved by molecular replacement using the coordinates of BsOppA bound to SNSS as the search model (Tables 2 and 3). Electron density defining the Ser-Arg-Asn-Val-Thr peptide was clearly observed in the peptide binding cavity.

The SNSS and SRNVT ligand backbones superpose closely and residues 1–3 of the two peptides make the same set of interactions with the protein (Fig. 2d). The >C=O of Val4 forms hydrogen bonds to the imidazole of His381 and the side chain amide of Asn376. The >N of Thr5 in SRNVT, which superposes with a carboxylate oxygen of SNSS, does not form an obvious polar interaction with the protein, while the carboxylate of Thr5 forms an ion pair with Arg423, a charge dipole interaction with the phenolic –OH of Tyr496, a further polar interaction with a water molecule and finally an intramolecular interaction with the side chain amide of Asn3. The Arg423 side chain plays an important role in ligand binding: through direct interactions with residues 3 and 5 of the peptide and by forming bridging interactions to Asn376 and Tyr496, which also interact with the peptide main chain.

As discussed above, the side chains of peptide ligands in the structures of OppA we have resolved form few or no direct interactions with the protein since such interactions would introduce specificity, which is not desirable in a protein whose function is sequence-independent peptide binding and transport [38]. In this regard, Arg2 of the PhrE peptide is of note as its guanidinium group forms five polar interactions with BsOppA including a striking two-pronged salt bridge to the side chain carboxylate of Asp449 (Fig. 2e). Relative to the position 2 binding pocket in the SNSS complex (Fig. 2f), the side chain of Asp449 has an altered conformer and four water molecules have been displaced. Interestingly, Arg2 is the most conserved residue in the Phr peptides of *B. subtilis* [14].

Structure determination and overall structure of DppE

While OppA in *B. subtilis* has been studied previously, the function and specificity of DppE (BsDppE) has not been investigated. We therefore purified and crystallized the protein and solved its structure (Tables 2 and 3). The electron density is contiguous over residues 21–521 of chain B with residues 261–265 and 521 in chain A and residues 256–271 and 521 in chain C missing and presumed to be disordered. The three molecules of the asymmetric unit are very closely superimposable with pairwise positional rmsΔ values of 0.3 Å for 480 equivalent Cα atoms. Each chain consists of two lobes with lobe 1 comprising residues 21–273 and 494–521 and lobe 2 made up of residues 274–493 (Fig. 3a). A structure similarity search revealed that the most similar entry to DppE in the protein data bank is OppA from *B. subtilis* whose structure is described above. Pairwise superposition of the chains from the DppE and OppA structures gives rmsΔ values of 1.2 Å for 480 or so equivalent Cα atoms.

The DppE crystal structure reveals a murein tripeptide binding site

As a result of their moderate to high affinities, the substrate binding proteins of ABC transporters frequently co-purify with their ligands and this has proved a useful tool for assigning specificity and inferring function [39]. After the DppE sequence had been built into the maps, we observed prominent electron density at the heart of all three protein molecules in the site

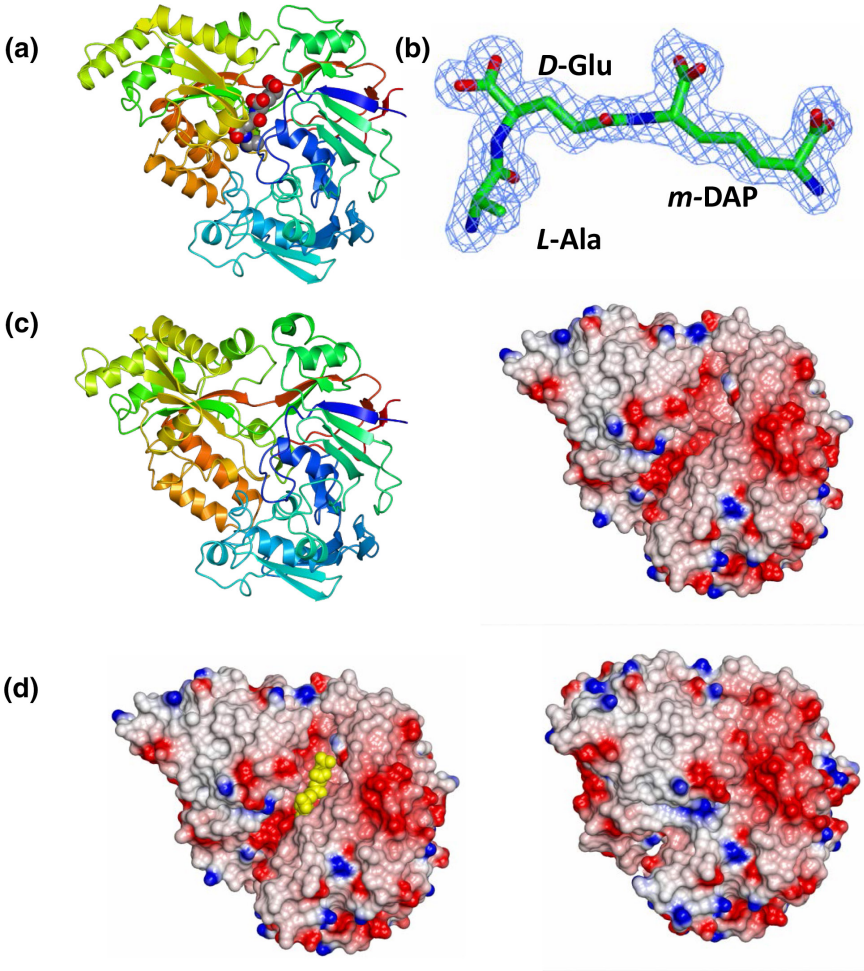


Fig. 3. The structure of DppE from *B. subtilis*. (a) Ribbon tracing of chain B from the crystal structure of DppE. The chain is colour ramped from the N-terminus (blue) to the C-terminus (red). The bound ligand is displayed as spheres coloured by atom type. (b) Electron density map (contoured at 1σ) in the interdomain region of the DppE structure displayed on the murein tripeptide ligand from the refined structure. (c) Ribbon tracing coloured as in (a) (left) and electrostatic surface rendering (right) of unliganded DppE. (d) Surface renderings of the structures illustrating the Venus flytrap binding mechanism. The murein tripeptide ligand, which is shown as green spheres, has been superimposed onto the open structure where it is visible in the binding cleft. Upon binding and domain closure, the ligand is completely enclosed and invisible from the exterior.

normally occupied by the ligand (Figs 3b and S4B). We were unable satisfactorily to model a conventional peptide into this electron density. Experience gained from determination of the structure of MppA in complex with murein tripeptide [40] led us to the realization that the same ligand was present in DppE and L-Ala-D-Glu-*meso*-DAP was very straightforwardly fitted into the maps (Fig. 3b). The murein tripeptide (MTP) ligand is enclosed in the protein interior such that it has no solvent accessible surface area.

Crystal structure of unliganded DppE reveals a 30° domain opening

Unliganded DppE was also prepared and crystallized as described in the Methods. The structure of these crystals was determined using diffraction data extending to 1.4 Å resolution allowing complete tracing of residues 21–521 of the single chain in the asymmetric unit. As shown in Fig. 3c, the unliganded protein is superficially similar in structure to the MTP-bound DppE, with closer inspection revealing that there has been a movement of Lobe 1 relative to Lobe 2. Analysis in the program DynDom suggests the structural change can be described as an inter-lobe rotation of 33° with the most significant changes in main chain conformation occurring at residues 273–277 and 492–494 which form the segments connecting the lobes. This observation is consistent with the Venus fly-trap mechanism of ligand binding reported for other solute binding proteins [41]. According to this mechanism the ligand binds to the open conformation of the protein as modelled in Fig. 3d (left), and is subsequently completely buried in the protein interior upon domain closure (Fig. 3d, right).

Mode of MTP binding

The murein tripeptide (L-Ala-γ-D-Glu-*meso*-DAP) ligand adopts an identical conformation in each of the three molecules of the asymmetric unit and essentially identical interactions are formed with the protein as shown schematically in Fig. 4(a, c). The α-amino group of Ala1 forms an ion-pairing interaction with the carboxylate side chain of Asp426 and charge-dipole interactions with the main chain >C=O of Leu424 and with a water molecule. The carbonyl oxygen of Ala1 forms a hydrogen bond with the >NH of Leu424. The >NH of D-Glu2 in turn forms a hydrogen bond with the main chain carbonyl oxygen of Phe41. The α-carboxylate of D-Glu2 forms a two-pronged salt bridging interaction with the side chain guanidinium group of Arg445. Moving along the peptide the >C=O and the >NH of the γ-peptide linkage form hydrogen bonds with the main chain amide of Asn43 and the side chain hydroxyl of Ser422 (Fig. 4a). The α-carboxylate of *meso*-DAP does not form any direct interactions with DppE, but instead participates in interactions with a number of ordered water molecules (Fig. 4c). At the extremity of the peptide, the free amino group of the DAP forms charge-dipole interactions with the side chain amide oxygen of Gln273 and the main chain >C=O of Asn493 (Fig. 4b). Meanwhile the carboxylate forms a salt bridge to Arg237 in molecules A and C, though interestingly not in molecule B where the Arg237 side chain is poorly defined in the electron density maps. The hydrogen bonding potential of the ligand that is not satisfied by protein–ligand interactions is fulfilled by supplementary interactions with well-defined buried water molecules (Fig. 4c).

Murein tripeptide bound to DppE has a closely similar conformation to Mtp bound to the murein tripeptide binding protein, MppA, from *E. coli* [40] and all of the ligand atoms of the respective structures are closely superimposable (Fig. 4b). There are nevertheless interesting differences in their mode of binding (Fig. 4c, d). Asp426 in DppE aligns with Asp417 in MppA and these residues are functionally equivalent in tethering the α-amino groups of Ala1 (Figs 4c, d and S7). In both cases, the α-carboxylate group of Glu2 forms a two-pronged salt-bridge to an arginine side chain. In this case, Arg445 of DppE does not correspond to Arg402 in MppA on the basis of sequence alignment (Fig. S7), suggesting convergent evolution of a specific interaction to the D-Glu α-carboxylate group in MppA and BsDppE. For the two carboxylate groups of the *meso*-DAP in DppE, the first interacts exclusively with water molecules while the second forms a charge–charge interaction with the side chain of Arg237. For MppA, these interactions are reversed and the first carboxylate forms an ion pair to an arginine (Arg 411) while the second is fully hydrated. Finally, the amino group of *meso*-DAP in both systems forms a charge–dipole interaction with an equivalent residue Gln273 in DppE, and Gln267 in MppA.

Analysis of the Dpp region in *B. subtilis* supports a proposed role for Dpp in peptidoglycan recycling

In a study of the murein tripeptide binding protein MppA from *E. coli*, we noticed that it was encoded on the genome of various gammaproteobacteria adjacent to, and probably co-regulated with, genes with roles in peptidoglycan recycling [40]. We went on to characterize one of these enzymes, MpaA, confirming the existence of the new scavenging pathway [42]. The pathway consists of an ABC transporter that handles murein tri-/tetra- or penta-peptides, a series of enzymes that cleave the peptide from the glycan and progressively shorten the peptide, and an isomerase that converts the D-Glu to L-Glu (Fig. 5a).

Closer analysis of the *dpp* locus in the genomes of *B. subtilis* and related bacteria suggests that a similar pathway exists in *Firmicutes*. The first clue is the presence of the gene, *ykfD*, downstream of the other *dpp* genes. The *ykfD* gene encodes a second ATPase subunit, the first being DppD, which is likely to be the 'missing' element in a five-component transporter, DppBCDE-YkfD (Fig. 5b). Although homodimeric arrangements of ATPase subunits are found in bacterial ABC transporters, heterodimers are much more common especially in the cluster C grouping. Secondly, inserted between the genes encoding the transporter components, DppE and YkfD, are three additional genes, *ykfABC* (Fig. 5b), which are homologous to genes known to be involved in murein recycling pathways. These encode an L-alanine-D/L-glutamate epimerase (YkfB), a γ-D-glutamyl-L-lysine endopeptidase (YkfC) and a putative murein peptide carboxypeptidase (YkfA) [43]. Thirdly, the gene immediately upstream of the *dpp* transporter genes, which is known as *dppA*, encodes a well-characterized D-alanyl-endopeptidase that is able to initiate the catabolism of the murein tetra- and penta-peptides [44, 45] (Fig. 5a). This gene organization is conserved in related *Bacillus* species and components are conserved in more diverse *Firmicutes*, including examples from species of *Listeria, Enterococcus* and *Clostridium* (Fig. 5b).

DISCUSSION

In bacteria, peptides play important signalling roles in adaptations to changes in the cellular environment [46]. Peptide signalling in bacteria was discovered in *B. subtilis* through studies of the regulation of sporulation, following the recognition that the sequence of the *spo0K* locus determined by Perego and Hoch bore striking similarity to the *opp* operon of *S. typhimurium* characterized by Higgins *et al.* [47]. Parallel studies of competence implicated Opp in the control of this adaptation in the same species [10]. The more widespread importance of peptide permeases as conduits for signalling peptides in Gram-positive bacteria became apparent in subsequent years with examples including (i) control of excision and transfer of mobile genetic elements [48], (ii) lysis to lysogeny decision-making in bacteriophages [49], (iii) pheromone-mediated plasmid conjugation of *Enterococcus faecalis* [50], and (iv) virulence in *B. thuringiensis* [51] and *Listeria monocytogenes* [52].

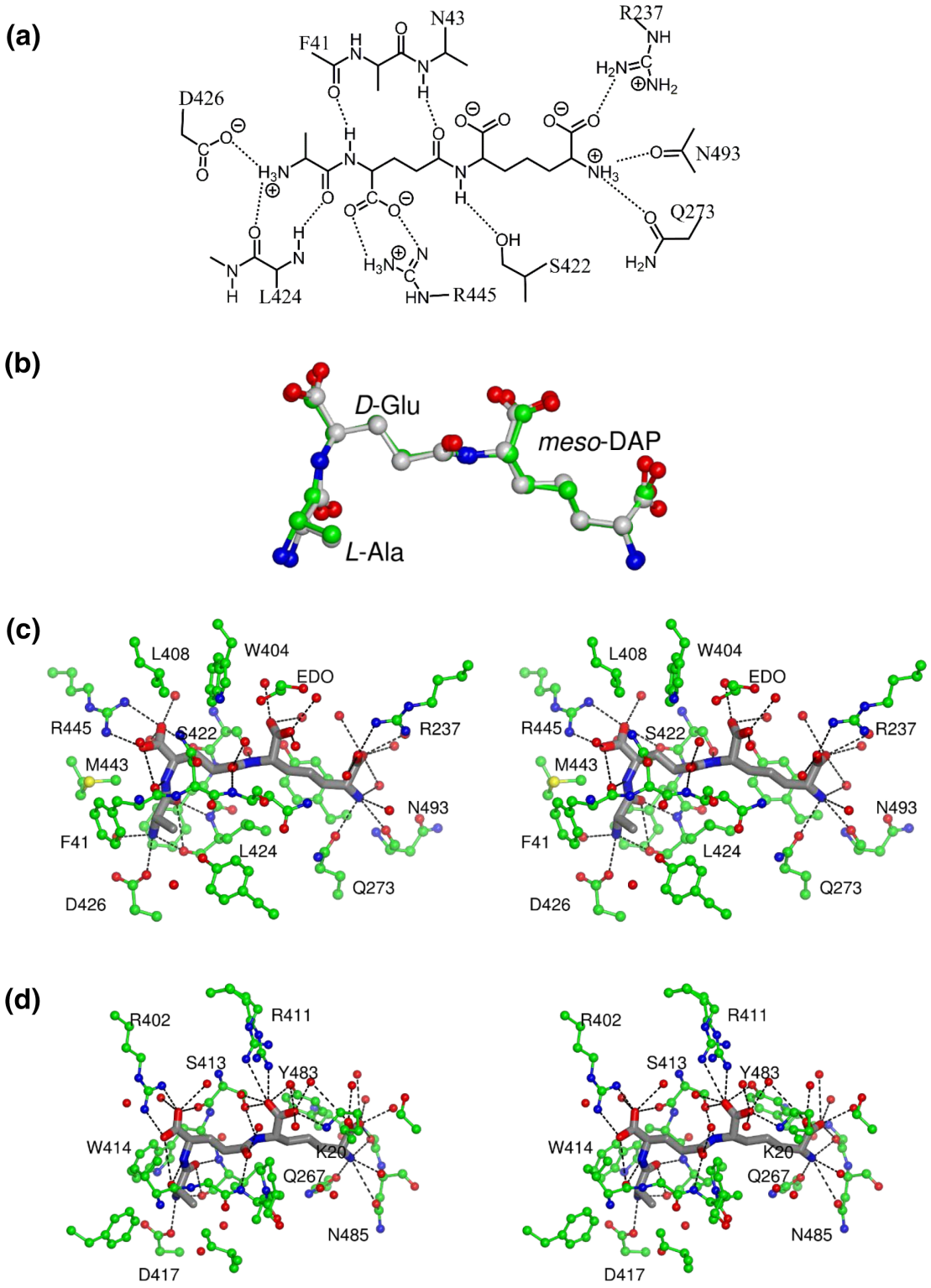


Fig. 4. Mode of murein tripeptide binding in DppE. (a) Schematic representation of the polar interactions (dashed lines) between DppE and the murein tripeptide ligand. (b) Superposition of the murein tripeptide ligands from the crystal structures of DppE from *B. subtilis* and MppA from *E. coli*. The colouring is by atom type: oxygen, red; nitrogen, blue; carbon, green in DppE and grey in MppA. (c and d) Stereo images illustrating the interactions of murein tripeptide shown in cylinder format (grey carbons) and the residues of the binding pocket shown in ball-and-stick format (green carbons). Polar interactions of the Mtp ligand and the surrounding protein and water molecules (red spheres) are indicated by dashed lines. In (c), EDO refers to ethane diol introduced during crystal cryoprotection.

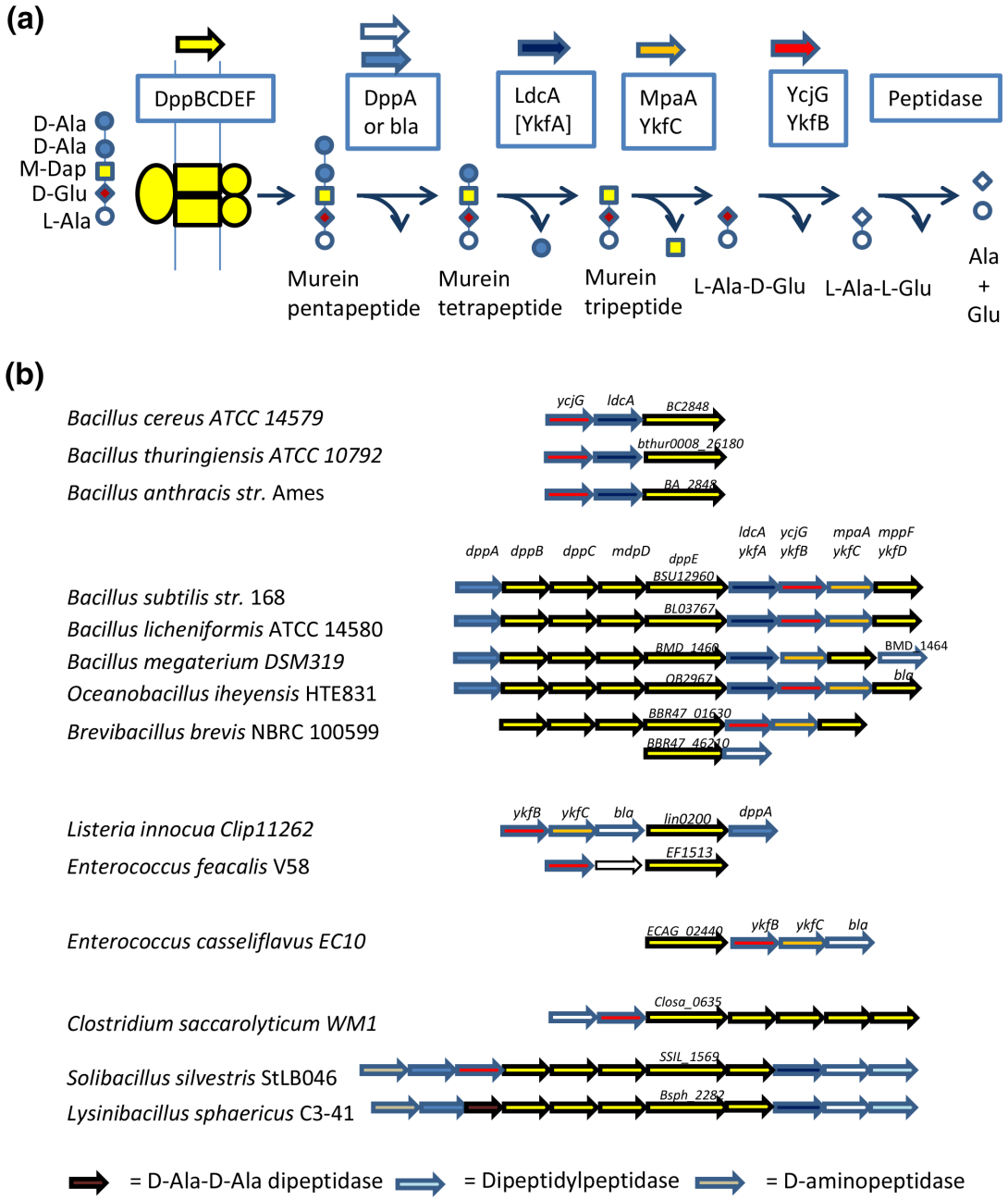


Fig. 5. The muropeptide catabolic pathway in *Firmicutes*. (a) Schematic representation of the steps in the uptake and catabolism of the muropeptide, beginning from the largest substrate, the murein pentapeptide. We propose uptake of the tetra- and tri-peptides is also mediated by Dpp, in which case only a subset of the enzymes would be required for catabolism. (b) Schematic of the layout of the *dpp* transporter genes and the associated murein catabolic enzymes. The colours specifying the gene functions match those for the protein components as defined at the top of part (a) and the bottom of part (b). Note that some of the more divergent clusters have additional/alternative genes that could encode proteins catalysing similar reactions; for example, *bla* is equivalent to *dppA*.

In most but not all cases, signalling peptides will be imported on oligopeptide transporters that play important nutritional roles. Many bacteria have evolved a set of high-affinity peptide uptake systems with distinct but overlapping specificities to enable the transport of peptides of variable length and diverse sequence. The range of ligands handled and the determinants of specificity have been studied most extensively for the oligopeptide permeases of *S. typhimurium* [38] and *Lactococcus lactis* [53]. An outcome of the present work is a complete set of structures of the ABC peptide transporter receptors in *B. subtilis*, namely DppE, OppA and AppA. Electron density in the cavity between the protein lobes clearly reveals the presence of endogenous peptides. In DppE, the ligand is murein tripeptide, while in OppA and AppA, the ligands of best fit are the tetrapeptide Ser-Asn-Ser-Ser and the

nonapeptide Val-Asp-Ser-Lys-Asn-Thr-Ser-Ser-Trp respectively [19]. There is ambiguity in these interpretations consistent with the expectation that Opp and App are, to a large extent, sequence-independent peptide transporters and that OppA and AppA are therefore sequence-independent peptide-binding proteins. The somewhat limited available evidence suggests that Opp transports shorter peptides with highest affinity for tetra- and pentapeptides with App handling longer peptides of up to nine residues. Meanwhile, Dpp may transport dipeptides, but its major role appears to be in the transport of cell wall peptides as discussed below.

In Fig. 6(a), peptide ligands from DppE, OppA and AppA from *B. subtilis,* DppA and MppA from *E. coli*, OppA from *S. typhimurium* and OppA from *Lactococcus lactis* are displayed following superposition of the main chains of the respective proteins. All of the peptides occupy an overlapping volume and there is close superposition of the peptide main chains. With the exception of the nonapeptide ligands in AppA and LlOppA, the α-amino groups of the peptides form a salt-bridge to a conserved aspartate that effectively anchors the ligand (Fig. 6b, c). As a result, peptides of different length differ in the position of their carboxyl termini in this set of SBPs.

It is interesting to consider the length dependence of oligopeptide binding in relation to the conservation of a signature sequence that appears as R^{423}MGWLGD^{429} in BsOppA and which forms an important and extensive surface of the binding site. The side chain of the residue corresponding to D^{429} in OppA binds the ligand's α-amino group while that corresponding to Trp^{426} stacks onto the first of the peptide bonds. The residues equivalent to R^{423} mediate direct or indirect interactions with the α-carboxylate of the ligand. AppA binds its nonapeptide such that residues 4–8 overlap with the pentapeptide in BsOppA, giving a three-residue extension at the N-terminus and an additional residue at the C-terminus. AppA thus lacks the Asp and Arg residues of the signature sequence [19]. LlOppA, which also lacks the signature Asp and Arg residues, binds its nonapeptide so that residues 5–9 correspond to the pentapeptide in BsOppA. In fact, LlOppA does not anchor either the N- or the C-termini of its ligand and instead binds peptides in variable registers according to the sequence [53]. This is a manifestation of its binding to peptides of widely varying length (4–35 residues). Meanwhile, the dipeptide binding protein, DppA, lacks the signature Arg residue as it uses a different residue, Arg^{355}, to bind its shorter substrates [54].

With its D-Glu residue and γ-peptide linkage, murein peptide is very different to a conventional peptide. However, the D-stereochemistry of residue 2 and its isopeptide linkage to *meso*-DAP represent compensating structural changes in three dimensions [40] so that murein tripeptide ligands in DppE and MppA overlap quite well with the conventional peptide ligands in the other proteins (Fig. 6a). Both exploit the conserved Asp residue (Asp^{426} and Asp^{417} respectively) in binding the α-amino group of L-Ala (Fig. 6c). In MppA, the signature Arg^{411} forms a salt-bridge to the terminal carboxylate of the murein tripeptide. As discussed earlier, the murein tripeptide binding pocket in DppE is differently organized – the signature Arg is missing (Fig. 6c) and its role is taken by Arg^{237} (Fig. S7).

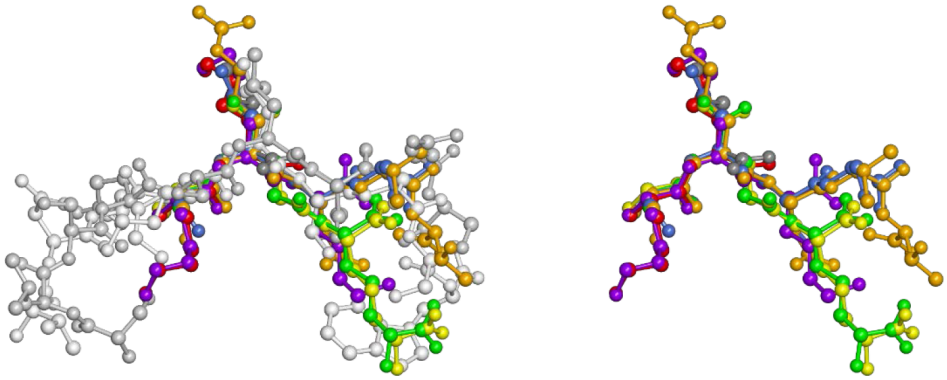
The results presented here suggest that DppE is functionally orthologous to MppA of *E. coli*. However, whereas MppA is an orphan solute binding protein which serves the Opp membrane components, DppE is encoded and cotranscribed with a cognate set of membrane components. In addition to this gene organizational divergence, our analysis shows that BsDppE and EcMppA have evolved distinct modes of murein tripeptide binding, probably from an OppA-like ancestor. This is evident in the differences in the arginine residues deployed in binding the three carboxylate groups of the MTP ligand.

A striking result from this work is the unexpected discovery of murein tripeptide in the binding pocket of DppE, suggesting strongly that Dpp is a murein tripeptide transporter. Insertional mutagenesis of the *dppA* and *dppE* genes had no effect on growth or sporulation in complex media [20]. The clearest phenotype associated with *dppE* mutation was observed following its introduction into a proline auxotrophic strain which lost the ability to grow on minimal media supplemented with Pro-Gly [20]. The capacity of the *dppE* strain to grow on defined media containing murein tripeptide has yet to be tested and represents an interesting and important future experiment.

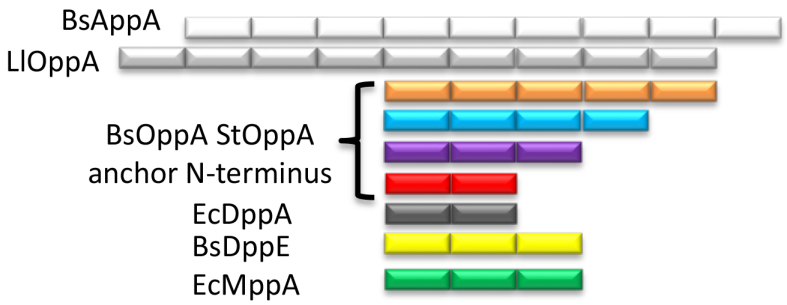
Dpp is under the transcriptional control of CodY in *B. subtilis*. Also contained in the CodY regulon and transcribed from the *dpp* promoter are a group of genes *ykfABCD* and *dppA* located downstream of *dppE*, and upstream of *dppB* respectively [55]. As discussed above and illustrated in Fig. 5, these genes encode a putative murein peptide carboxypeptidase (YkfA), an L-alanine-D/L-glutamate epimerase (YkfB), a γ-D-glutamyl-L-lysine endopeptidase (YkfC) and a D-alanyl-endopeptidase (DppA) that are either known or highly likely to function in cell wall catabolism [44, 45]. These observations reinforce the interpretation that Dpp is a murein tripeptide transporter that functions in the uptake and recycling of cell wall peptides. Moreover, *dpp* regulation by CodY, and the persistence of the *dpp* transcript under nutrient-limiting conditions and during sporulation [55] are consistent with (i) the need to conserve resources in stationary phase and (ii) the extensive cell wall restructuring that accompanies asymmetric septation and mother cell engulfment of the forespore during spore formation [56].

Scavenging of murein peptides is well known in Gram-negative bacteria [57] and there is evidence that it also takes place in Gram-positive bacteria where the murein layer in the cell wall is much thicker. In *B. subtilis* turnover of up to 50% of the murein takes place per generation, accompanied by release of murein tripeptide into the medium [58]. This accumulates to high concentrations in the medium of growing cultures of *B. subtilis* and would constitute an excellent source of carbon, nitrogen and energy at a later stage in growth. The presence of DppA, a well-characterized intracellular D-alanyl-aminopeptidase, suggests that the Dpp

(a)



(b)



(c)

Protein	Ligand	α-NH_3^+	α-CO_2^-	Signature Sequence
BsDppE	Mtp	D^{426}	H_2O/R^{237}	$Q^{420}SSFLPD^{426}$
BsOppA	4-5	D^{429}	R^{423}	$R^{423}MGWLGD^{429}$
BsAppA	9	D^{154}	R^{373}	$V^{416}MGWSLS^{422}$
EcMppA	Mtp	D^{417}	R^{411}/H_2O	$R^{411}ASWVGD^{417}$
EcDppA	2	D^{408}	R^{355}	$M^{402}MGWTGD^{408}$
StOppA	2-4	D^{419}	R^{413}	$R^{413}AGWCAD^{419}$
LlOppA	9	Unfixed	Unfixed	$D^{470}GSWSLA^{476}$

Fig. 6. Comparison of peptide binding across SBP structures. (a) Overlay of the peptide ligands from BsAppA-VDSKNTSSW (1XOC; white), LlOppA-SLSQSLSQS (3RYB; light grey), BsOppA-SRNVT (orange), BsOppA-SNSS (light blue), BsDppE-Mtp (yellow), EcMppA-Mtp (3O9P; green), EcDppA (1DPP; grey), StOppA-KKK (2OLB; purple) and StOppA-KK (2RKM; red) – left with all of the above structures and right with BsAppA-VDSKNTSSW and LlOppA-SLSQSLSQS omitted. The structures were superposed using the SSM superpose routines in CCP4mg [61] applied to protein atoms (matching rmsΔ = 0.3–2.0 Å for 453–508 equivalent atom pairs). Ligands are coloured according to structure. (b) Schematic summary of the overlay in (a) emphasizing the relative positions of the peptides in the respective binding pockets. Each residue is represented by a block and the colouring is as in (a). (c) Variations on the sequence of the peptide binding motif which appears as R^{413}AGWCAD^{419} in StOppA. Residues in red font form interactions with the ligands' charged termini with underlined aromatic residues packing against peptide bonds in the ligands.

system may also mediate the uptake of tetra- and or penta-peptides derived from peptidoglycan. These peptides are likely to be produced by the action of the secreted amidase, AmiE (YbbE), which cleaves the amide linkage in *N*-acetylmuramyl-L-alanine to produce MurNAc and the cell wall-derived stem peptides [59]. Peptidoglycan-derived MurNAc uptake and recycling in *B. subtilis* has subsequently been shown to occur during the transition to stationary phase [60]. We propose that the DppE-containing ABC transporter combined with other gene products in this region, which are similarly produced under nutrient-limiting conditions,

take up and catabolize the released stem peptides (Fig. 5a), providing complete re-assimilation of all the components of the peptidoglycan.

Funding information
This project was funded by the Biotechnology and Biological Sciences Research Council, UK, through a White Rose DTP Studentship to A.M.H.

Acknowledgements
We thank the Diamond Light Source for access to beamlines i02, i03 and i04 (proposal numbers mx-9948 and mx-13587) that contributed to the results presented here.

Author contributions
The project was conceived by A.J.W. and G.H.T. The experiments were carried out by A.M.H. with wet laboratory experimental assistance from S.J.W. Specific expert contributions were provided by J.F.D. (ligand binding), and J.P.T. and E.J.D. (X-ray data collection and structure refinement). The manuscript was drafted by A.M.H., G.H.T. and A.J.W. and edited and revised by all authors.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Hoch JA. Regulation of the phosphorelay and the initiation of sporulation in *Bacillus subtilis*. *Annu Rev Microbiol* 1993;47:441–465.
2. Muchová K, Lewis RJ, Perecko D, Brannigan JA, Ladds JC, *et al*. Dimer-induced signal propagation in Spo0A. *Mol Microbiol* 2004;53:829–842.
3. Molle V, Fujita M, Jensen ST, Eichenberger P, González-Pastor JE, *et al*. The Spo0A regulon of *Bacillus subtilis*. *Mol Microbiol* 2003;50:1683–1701.
4. Burbulys D, Trach KA, Hoch JA. Initiation of sporulation in *B. subtilis* is controlled by a multicomponent phosphorelay. *Cell* 1991;64:545–552.
5. Jiang M, Shao W, Perego M, Hoch JA. Multiple histidine kinases regulate entry into stationary phase and sporulation in *Bacillus subtilis*. *Mol Microbiol* 2000;38:535–542.
6. Perego M, Hanstein C, Welsh KM, Djavakhishvili T, Glaser P, *et al*. Multiple protein-aspartate phosphatases provide a mechanism for the integration of diverse signals in the control of development in *B. subtilis*. *Cell* 1994;79:1047–1055.
7. Ohlsen KL, Grimsley JK, Hoch JA. Deactivation of the sporulation transcription factor Spo0A by the Spo0E protein phosphatase. *Proc Natl Acad Sci U S A* 1994;91:1756–1760.
8. Wu R, Gu M, Wilton R, Babnigg G, Kim Y, *et al*. Insight into the sporulation phosphorelay: crystal structure of the sensor domain of *Bacillus subtilis* histidine kinase, KinD. *Protein Sci* 2013;22:564–576.
9. Piggot PJ, Coote JG. Genetic aspects of bacterial endospore formation. *Bacteriol Rev* 1976;40:908–962.
10. Rudner DZ, LeDeaux JR, Ireton K, Grossman AD. The spo0K locus of *Bacillus subtilis* is homologous to the oligopeptide permease locus and is required for sporulation and competence. *J Bacteriol* 1991;173:1388–1398.
11. Perego M, Higgins CF, Pearce SR, Gallagher MP, Hoch JA. The oligopeptide transport system of *Bacillus subtilis* plays a role in the initiation of sporulation. *Mol Microbiol* 1991;5:173–185.
12. Perego M, Hoch JA. Cell-cell communication regulates the effects of protein aspartate phosphatases on the phosphorelay controlling development in *Bacillus subtilis*. *Proc Natl Acad Sci U S A* 1996;93:1549–1553.
13. Perego M. A peptide export-import control circuit modulating bacterial development regulates protein phosphatases of the phosphorelay. *Proc Natl Acad Sci U S A* 1997;94:8612–8617.
14. Perego M, Brannigan JA. Pentapeptide regulation of aspartyl-phosphate phosphatases. *Peptides* 2001;22:1541–1547.
15. Maqbool A, Horler RSP, Muller A, Wilkinson AJ, Wilson KS, *et al*. The substrate-binding protein in bacterial ABC transporters: dissecting roles in the evolution of substrate specificity. *Biochem Soc Trans* 2015;43:1011–1017.
16. Berntsson RP-A, Smits SHJ, Schmitt L, Slotboom D-J, Poolman B. A structural classification of substrate-binding proteins. *FEBS Lett* 2010;584:2606–2617.
17. Koide A, Hoch JA. Identification of a second oligopeptide transport system in *Bacillus subtilis* and determination of its role in sporulation. *Mol Microbiol* 1994;13:417–426.
18. Picon A, van Wely KHM. Peptide binding to the *Bacillus subtilis* oligopeptide-binding proteins OppA and AppA. *Molecular Biology Today* 2001;2:21–25.
19. Levdikov VM, Blagova EV, Brannigan JA, Wright L, Vagin AA, *et al*. The structure of the oligopeptide-binding protein, AppA, from *Bacillus subtilis* in complex with a nonapeptide. *J Mol Biol* 2005;345:879–892.
20. Mathiopoulos C, Mueller JP, Slack FJ, Murphy CG, Patankar S, *et al*. A *Bacillus subtilis* dipeptide transport system expressed early during sporulation. *Mol Microbiol* 1991;5:1903–1913.
21. Levdikov VM, Blagova E, Young VL, Belitsky BR, Lebedev A, *et al*. Structure of the branched-chain amino acid and GTP-sensing global regulator, CodY, from *Bacillus subtilis J Biol Chem* 2017;292:2714–2728.
22. Molle V, Nakaura Y, Shivers RP, Yamaguchi H, Losick R, *et al*. Additional targets of the *Bacillus subtilis* global regulator CodY identified by chromatin immunoprecipitation and genome-wide transcript analysis. *J Bacteriol* 2003;185:1911–1922.
23. Brinsmade SR, Alexander EL, Livny J, Stettner AI, Segrè D, *et al*. Hierarchical expression of genes controlled by the *Bacillus subtilis* global regulatory protein CodY. *Proc Natl Acad Sci U S A* 2014;111:8227–8232.
24. Slack FJ, Serror P, Joyce E, Sonenshein AL. A gene required for nutritional repression of the *Bacillus subtilis* dipeptide permease operon. *Mol Microbiol* 1995;15:689–702.
25. Tame JR, Murshudov GN, Dodson EJ, Neil TK, Dodson GG, *et al*. The structural basis of sequence-independent peptide binding by OppA protein. *Science* 1994;264:1578–1581.
26. Fogg MJ, Wilkinson AJ. Higher-throughput approaches to crystallization and crystal structure determination. *Biochem Soc Trans* 2008;36:771–775.
27. Hughes A, Wilson S, Dodson EJ, Turkenburg JP, Wilkinson AJ. Crystal structure of the putative peptide-binding protein AppA from *Clostridium difficile*. *Acta Crystallogr F Struct Biol Commun* 2019;75:246–253.
28. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 2010;66:486–501.
29. Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, *et al*. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* 2011;67:355–367.
30. Winter G, Lobley CMC, Prince SM. Decision making in xia2. *Acta Crystallogr D Biol Crystallogr* 2013;69:1260–1273.
31. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, *et al*. Phaser crystallographic software. *J Appl Crystallogr* 2007;40:658–674.

32. Cowtan K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* 2006;62:1002–1011.
33. Berntsson RP-A, Schuurman-Wolters GK, Dunny G, Slotboom D-J, Poolman B. Structure and mode of peptide binding of pheromone receptor PrgZ. *J Biol Chem* 2012;287:37165–37170.
34. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 2010;66:486–501.
35. Evans P. Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* 2006;62:72–82.
36. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, *et al*. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res* 2014;42:D206–14.
37. Niesen FH, Berglund H, Vedadi M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* 2007;2:2212–2221.
38. Sleigh SH, Seavers PR, Wilkinson AJ, Ladbury JE, Tame JR. Crystallographic and calorimetric analysis of peptide binding to OppA protein. *J Mol Biol* 1999;291:393–415.
39. Müller A, Thomas GH, Horler R, Brannigan JA, Blagova E, *et al*. An ATP-binding cassette-type cysteine transporter in *Campylobacter jejuni* inferred from the structure of an extracytoplasmic solute receptor protein. *Mol Microbiol* 2005;57:143–155.
40. Maqbool A, Levdikov VM, Blagova EV, Hervé M, Horler RSP, *et al*. Compensating stereochemical changes allow murein tripeptide to be accommodated in a conventional peptide-binding protein. *J Biol Chem* 2011;286:31512–31521.
41. Mao B, Pear MR, McCammon JA, Quiocho FA. Hinge-bending in L-arabinose-binding protein. The "Venus's-flytrap" model. *J Biol Chem* 1982;257:1131–1133.
42. Maqbool A, Hervé M, Mengin-Lecreulx D, Wilkinson AJ, Thomas GH. MpaA is a murein-tripeptide-specific zinc carboxypeptidase that functions as part of a catabolic pathway for peptidoglycan-derived peptides in γ-proteobacteria. *Biochem J* 2012;448:329–341.
43. Schmidt DM, Hubbard BK, Gerlt JA. Evolution of enzymatic activities in the enolase superfamily: functional assignment of unknown proteins in *Bacillus subtilis* and *Escherichia coli* as L-Ala-D/L-Glu epimerases. *Biochemistry* 2001;40:15707–15715.
44. Cheggour A, Fanuel L, Duez C, Joris B, Bouillenne F, *et al*. The dppA gene of *Bacillus subtilis* encodes a new D-aminopeptidase. *Mol Microbiol* 2000;38:504–513.
45. Remaut H, Bompard-Gilles C, Goffin C, Frère JM, Van Beeumen J. Structure of the *Bacillus subtilis* D-aminopeptidase DppA reveals a novel self-compartmentalizing protease. *Nat Struct Biol* 2001;8:674–678.
46. Slamti L, Lereclus D. The oligopeptide ABC-importers are essential communication channels in Gram-positive bacteria. *Res Microbiol* 2019;170:338–344.
47. Perego M, Higgins CF, Pearce SR, Gallagher MP, Hoch JA. The oligopeptide transport system of *Bacillus subtilis* plays a role in the initiation of sporulation. *Mol Microbiol* 1991;5:173–185.
48. Auchtung JM, Lee CA, Monson RE, Lehman AP, Grossman AD. Regulation of a *Bacillus subtilis* mobile genetic element by intercellular signaling and the global DNA damage response. *Proc Natl Acad Sci U S A* 2005;102:12554–12559.
49. Erez Z, Steinberger-Levy I, Shamir M, Doron S, Stokar-Avihail A, *et al*. Communication between viruses guides lysis-lysogeny decisions. *Nature* 2017;541:488–493.
50. Leonard BA, Podbielski A, Hedberg PJ, Dunny GM. *Enterococcus faecalis* pheromone binding protein, PrgZ, recruits a chromosomal oligopeptide permease system to import sex pheromone cCF10 for induction of conjugation. *Proc Natl Acad Sci U S A* 1996;93:260–264.
51. Gominet M, Slamti L, Gilois N, Rose M, Lereclus D. Oligopeptide permease is required for expression of the *Bacillus thuringiensis* plcR regulon and for virulence. *Mol Microbiol* 2001;40:963–975.
52. Krypotou E, Scortti M, Grundström C, Oelker M, Luisi BF. Control of bacterial virulence through the peptide signature of the habitat. *Cell Rep* 2019;26:1815–1827.
53. Berntsson RP-A, Thunnissen A-MWH, Poolman B, Slotboom D-J. Importance of a hydrophobic pocket for peptide binding in lactococcal OppA. *J Bacteriol* 2011;193:4254–4256.
54. Dunten P, Mowbray SL. Crystal structure of the dipeptide binding protein from *Escherichia coli* involved in active transport and chemotaxis. *Protein Sci* 1995;4:2327–2334.
55. Nicolas P, Mäder U, Dervyn E, Rochat T, Leduc A, *et al*. Condition-dependent transcriptome reveals high-level regulatory architecture in *Bacillus subtilis*. *Science* 2012;335:1103–1106.
56. Chan H, Taib N, Gilmore MC, Mohamed AMT, Hanna K, *et al*. Genetic screens identify additional genes implicated in envelope remodeling during the engulfment stage of *Bacillus subtilis* sporulation. *mBio* 2022;13:e0173222.
57. Park JT. Why does *Escherichia coli* recycle its cell wall peptides? *Mol Microbiol* 1995;17:421–426.
58. Mauck J, Chan L, Glaser L. Turnover of the cell wall of Gram-positive bacteria. *J Biol Chem* 1971;246:1820–1827.
59. Litzinger S, Duckworth A, Nitzsche K, Risinger C, Wittmann V, *et al*. Muropeptide rescue in *Bacillus subtilis* involves sequential hydrolysis by beta-N-acetylglucosaminidase and N-acetylmuramyl-L-alanine amidase. *J Bacteriol* 2010;192:3132–3143.
60. Borisova M, Gaupp R, Duckworth A, Schneider A, Dalügge D, *et al*. Peptidoglycan recycling in Gram-positive bacteria is crucial for survival in stationary phase. *mBio* 2016;7:e00923-16.
61. McNicholas S, Potterton E, Wilson KS, Noble MEM. Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr D Biol Crystallogr* 2011;67:386–394.

Edited by: W. van Schaik and F. M. Commichau