

Qingping Xu,^{a,b} Polat Abdubek,^{b,c}
Tamara Astakhova,^{b,d} Herbert L.
Axelrod,^{a,b} Constantina Bakolitsa,^{b,e}
Xiaohui Cai,^{b,d} Dennis Carlton,^{b,f}
Connie Chen,^{b,c} Hsiu-Ju Chiu,^{a,b}
Michelle Chiu,^{b,c} Thomas Clayton,^{b,f}
Debanu Das,^{a,b} Marc C. Deller,^{b,f} Lian
Duan,^{b,d} Kyle Ellrott,^{b,d} Carol L. Farr,^{b,f}
Julie Feuerhelm,^{b,c} Joanna C. Grant,^{b,c}
Anna Grzechnik,^{b,f} Gye Won Han,^{b,f}
Lukasz Jaroszewski,^{b,d,e} Kevin K. Jin,^{a,b}
Heath E. Klock,^{b,c} Mark W. Knuth,^{b,c}
Piotr Kozbial,^{b,e} S. Sri Krishna,^{b,d,e}
Abhinav Kumar,^{a,b} Winnie W. Lam,^{a,b}
David Marciano,^{b,f} Mitchell D.
Miller,^{a,b} Andrew T. Morse,^{b,d} Edward
Nigoghossian,^{b,c} Amanda Nopakun,^{b,f}
Linda Okach,^{b,c} Christina Puckett,^{b,c}
Ron Reyes,^{a,b} Henry J. Tien,^{b,f}
Christine B. Trame,^{a,b} Henry van den
Bedem,^{a,b} Dana Weekes,^{b,e} Tiffany
Wooten,^{b,c} Andrew Yeh,^{a,b} Keith O.
Hodgson,^{b,g} John Wooley,^{b,d}
Marc-André Elsliger,^{b,f} Ashley M.
Deacon,^{a,b} Adam Godzik,^{b,d,e} Scott A.
Lesley,^{b,c,f} and Ian A. Wilson^{b,f*}

^aStanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, CA, USA, ^bJoint Center for Structural Genomics, <http://www.jcsg.org>, USA, ^cProtein Sciences Department, Genomics Institute of the Novartis Research Foundation, San Diego, CA, USA, ^dCenter for Research in Biological Systems, University of California, San Diego, La Jolla, CA, USA, ^eProgram on Bioinformatics and Systems Biology, Sanford–Burnham Medical Research Institute, La Jolla, CA, USA, ^fDepartment of Molecular Biology, The Scripps Research Institute, La Jolla, CA, USA, and ^gPhoton Science, SLAC National Accelerator Laboratory, Menlo Park, CA, USA

Correspondence e-mail: wilson@scripps.edu

Received 13 April 2010
Accepted 3 June 2010

PDB Reference: YkfC–L-Ala- γ -D-Glu complex, 3h41.

Structure of the γ -D-glutamyl-L-diamino acid endopeptidase YkfC from *Bacillus cereus* in complex with L-Ala- γ -D-Glu: insights into substrate recognition by NlpC/P60 cysteine peptidases

Dipeptidyl-peptidase VI from *Bacillus sphaericus* and YkfC from *Bacillus subtilis* have both previously been characterized as highly specific γ -D-glutamyl-L-diamino acid endopeptidases. The crystal structure of a YkfC ortholog from *Bacillus cereus* (BcYkfC) at 1.8 Å resolution revealed that it contains two N-terminal bacterial SH3 (SH3b) domains in addition to the C-terminal catalytic NlpC/P60 domain that is ubiquitous in the very large family of cell-wall-related cysteine peptidases. A bound reaction product (L-Ala- γ -D-Glu) enabled the identification of conserved sequence and structural signatures for recognition of L-Ala and γ -D-Glu and, therefore, provides a clear framework for understanding the substrate specificity observed in dipeptidyl-peptidase VI, YkfC and other NlpC/P60 domains in general. The first SH3b domain plays an important role in defining substrate specificity by contributing to the formation of the active site, such that only murein peptides with a free N-terminal alanine are allowed. A conserved tyrosine in the SH3b domain of the YkfC subfamily is correlated with the presence of a conserved acidic residue in the NlpC/P60 domain and both residues interact with the free amine group of the alanine. This structural feature allows the definition of a subfamily of NlpC/P60 enzymes with the same N-terminal substrate requirements, including a previously characterized cyanobacterial L-alanine- γ -D-glutamate endopeptidase that contains the two key components (an NlpC/P60 domain attached to an SH3b domain) for assembly of a YkfC-like active site.

1. Introduction

Cell-wall turnover, an enzymatic process that results in the loss of peptidoglycan (PG) components, has been reported in many bacteria, including *Escherichia coli* and *Bacillus subtilis* (Doyle *et al.*, 1988). The products of the turnover are generally re-utilized through a process known as PG recycling (Park & Uehara, 2008). The molecular processes involved in cell-wall turnover and recycling are not currently well understood in comparison to cell-wall synthesis, particularly in bacteria other than *E. coli* (Park & Uehara, 2008; Uehara & Park, 2003, 2004, 2007; Uehara *et al.*, 2005). In *E. coli*, the cell wall is degraded by lytic transglycosylases that release anhydromuropeptides (GlcNAc-anhMurNAc-L-Ala- γ -D-Glu-DAP-D-Ala, where DAP is *meso*-diaminopimelic acid), which are imported into the cytoplasm, primarily by AmpG permease, and subsequently processed by *N*-acetyl-anhydromuramyl-L-alanine amidase (which cleaves between GlcNAc-anhMurNAc and L-Ala) and LD-carboxypeptidase LdcA (which cleaves between DAP and D-Ala). Two fates are possible for the generated murein tripeptide L-Ala- γ -D-Glu-DAP. Under normal growth conditions, these tripeptides are recycled by Mpl ligase and returned to the peptidoglycan-biosynthetic pathway. During nutrient-limiting conditions, an additional pathway (Fig. 1) is likely to be involved in the murein tripeptide metabolism, as proposed in *E. coli* (Uehara & Park, 2003). MpaA endopeptidase, a metalloprotease, specifically cleaves L-Ala- γ -D-Glu-DAP to produce L-Ala- γ -D-Glu and DAP. L-Ala- γ -D-Glu is then converted to

L-Ala-L-Glu and subsequently to L-Ala and L-Glu by YcjG epimerase and PepD peptidase, respectively.

Some of the enzymes in the cell-wall recycling of *E. coli*, such as AmpG, AmpD, Mpl and MpaA, have no orthologs in *B. subtilis* (Park & Uehara, 2008), suggesting that the mechanism of cell-wall recycling may differ between the two bacteria. In *B. subtilis*, PG was proposed to be cleaved by a muramidase and an amidase to produce GlcNAc-MurNAc and stem peptides (Park & Uehara, 2008). The free peptides are imported into the cytoplasm by an unidentified permease and subsequently processed by YkfABC enzymes. YkfA, an LD-carboxypeptidase, removes the terminal D-Ala. The generated tripeptide is further metabolized by a pathway that is functionally equivalent to that of *E. coli*, with YkfC as the γ -D-Glu-DAP endopeptidase and YkfB as the L-Ala-D-Glu epimerase (Fig. 1) (Schmidt *et al.*, 2001). Interestingly, while YkfB is homologous to YcjG of *E. coli*, YkfC is unrelated to MpaA in sequence and structure despite having an equivalent function.

YkfC contains a C-terminal NlpC/P60 cysteine peptidase domain. NlpC/P60 is a large family of cell-wall related cysteine peptidases that are broadly distributed in bacteria, viruses, archaea and eukaryotes (Anantharaman & Aravind, 2003; Bateman & Rawlings, 2003; Rigden *et al.*, 2003). Characterized NlpC/P60 enzymes are almost all γ -D-Glu-DAP (or γ -D-Glu-Lys) endopeptidases. While their biochemical function seems to be conserved, the physiological roles of NlpC/P60 proteins are diverse, including involvement in cell separation, expansion, differentiation, cell-wall turnover, cell lysis, protein secretion and virus infection (Smith *et al.*, 2000). Secreted NlpC/P60 proteins also have other roles in pathogenesis. The autolysin P60 of *Listeria monocytogenes* is involved in host-cell invasion (Kuhn & Goebel, 1989), enterotoxin FM of *B. cereus* in food poisoning (Asano *et al.*, 1997), and SagA of *Enterococcus faecium* is a secreted antigen that binds to extracellular matrix proteins (Teng *et al.*, 2003).

NlpC/P60 proteins can be lethal to bacteria owing to their ability to compromise cell-wall integrity or cell-wall biosynthesis. Therefore, their activities are tightly controlled through multiple mechanisms (Smith *et al.*, 2000); their expression is regulated at the transcription level and their cellular localization is dependent on their physiological roles. Furthermore, their atomic structures are highly optimized to precisely define their substrate specificity (Xu *et al.*, 2009). NlpC/P60 proteins are often fused to auxiliary domains, many of which are known cell-wall binding modules (*e.g.* LysM and the choline-binding domain). Thus, it is generally assumed that these auxiliary domains function as targeting domains which localize their proteins to the cell wall. The functional synergy between the NlpC/P60 domains and their auxiliary domains is currently not fully understood. We have previously determined the crystal structure of a γ -D-Glu-DAP endopeptidase from cyanobacteria (AvPCP/NpPCP;

Anabaena variabilis/Nostoc punctiforme PG cysteine peptidase; Xu *et al.*, 2009) and showed that it contained an N-terminal bacterial SH3 (SH3b) domain and a C-terminal NlpC/P60 domain. We proposed that the SH3b domain of this enzyme is important in defining the substrate specificity of the peptidase domain. However, the mechanism of substrate recognition by NlpC/P60 and SH3b was not firmly established. Here, we report the crystal structure of YkfC from *B. cereus* (BcYkfC) in complex with L-Ala- γ -D-Glu. BcYkfC shares 40% sequence identity with YkfC from *B. subtilis*, which has previously been biochemically characterized (Schmidt *et al.*, 2001). Thus, we now have the first detailed view of substrate recognition by an NlpC/P60 protein.

2. Material and methods

2.1. Sequence analysis

Homologs of BcYkfC were identified using *PSI-BLAST* (Altschul *et al.*, 1997; three iterations) against the nonredundant (nr) protein sequence database at the National Center for Biotechnology Information (NCBI) using the sequence of the catalytic domain of BcYkfC as the probe (residues 210–333). An alignment length of ≥ 70 and an *E* value of ≤ 0.02 were used to extract a subset of hits. These proteins (2599 sequences) were aligned using *HMMALIGN* (Eddy, 1998) against the default global alignment profile of NlpC/P60 domains (PF00877) from the PFAM database v.23 (Bateman *et al.*, 2004). YkfC-subfamily candidates were extracted from the above aligned subset based on the presence of an aspartate corresponding to position 256 of BcYkfC. The full-length sequences were then aligned and clustered using *PIPEALIGN* (Plewniak *et al.*, 2003). Only sequences that also contained a conserved tyrosine corresponding to position 118 of BcYkfC were classified into the YkfC subfamily. Plots of sequence conservation in the active site were prepared using *WEBLOGO* (Crooks *et al.*, 2004).

2.2. Protein expression and purification

Clones were generated using the Polymerase Incomplete Primer Extension (PIPE) cloning method (Klock *et al.*, 2008). The gene encoding BcYkfC (GenBank NP_979181; Swiss-Prot Q736M3) was amplified by polymerase chain reaction (PCR) from *B. cereus* NRS248 ATCC 10987 genomic DNA using *PfuTurbo* DNA polymerase (Stratagene) and I-PIPE (Insert) primers (forward primer, 5'-ctgtacttccagggcGAAGAGAAGAAAGATAGTAAGGCGT-3'; reverse primer, 5'-aattaagtcgcttaAGGTAAGTAACGACGCGCAC-CAGCG-3'; target sequence in upper case) that included sequences for the predicted 5' and 3' ends. The expression vector pSpeedET, which encodes an amino-terminal tobacco etch virus (TEV) protease-cleavable expression and purification tag (MGSDKIHSHHHHHEN-LYFQ/G), was PCR-amplified with V-PIPE (Vector) primers (forward primer, 5'-taacgagactaattaactcgtttaaacggtctccagc-3'; reverse primer, 5'-gccctggaagtacaggttttctgtgatgatgatgatg-3'). V-PIPE and I-PIPE PCR products were mixed to anneal the amplified DNA fragments together. *E. coli* GeneHogs (Invitrogen) competent cells were transformed with the I-PIPE/V-PIPE mixture and dispensed onto selective LB-agar plates. The cloning junctions were confirmed by DNA sequencing. Using the PIPE method, the gene segment encoding residues Met1–Ala23 was omitted as these residues were predicted to form a signal peptide. Expression was performed in a selenomethionine-containing medium with suppression of normal methionine synthesis. At the end of fermentation, lysozyme was added to the culture to a final concentration of 250 $\mu\text{g ml}^{-1}$ and the cells were harvested and frozen. After one freeze–thaw cycle, the

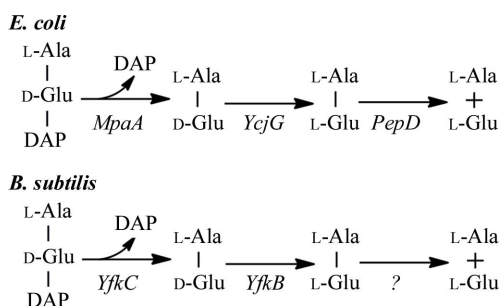


Figure 1
Proposed metabolic pathways for murein peptides in *E. coli* and *B. subtilis*.

cells were sonicated in lysis buffer [50 mM HEPES pH 8.0, 50 mM NaCl, 10 mM imidazole, 1 mM tris(2-carboxyethyl)phosphine-HCl (TCEP)] and the lysate was clarified by centrifugation at 32 500g for 30 min. The soluble fraction was passed over nickel-chelating resin (GE Healthcare) pre-equilibrated with lysis buffer, the resin was washed with wash buffer [50 mM HEPES pH 8.0, 300 mM NaCl, 40 mM imidazole, 10% (v/v) glycerol, 1 mM TCEP] and the protein was eluted with elution buffer [20 mM HEPES pH 8.0, 300 mM imidazole, 10% (v/v) glycerol, 1 mM TCEP]. The eluate was buffer-exchanged with TEV buffer (20 mM HEPES pH 8.0, 200 mM NaCl, 40 mM imidazole, 1 mM TCEP) using a PD-10 column (GE Healthcare) and incubated with 1 mg TEV protease per 15 mg of eluted protein. The protease-treated eluate was run over nickel-chelating resin (GE Healthcare) pre-equilibrated with HEPES crystallization buffer (20 mM HEPES pH 8.0, 200 mM NaCl, 40 mM imidazole, 1 mM TCEP) and the resin was washed with the same buffer. The flowthrough and wash fractions were combined and concentrated to 18.8 mg ml⁻¹ as determined using the Coomassie Plus Protein Assay Reagent (Pierce) by centrifugal ultrafiltration (Millipore) for crystallization trials. The oligomeric state of BcYkfC was determined using a 0.8 × 30 cm Shodex Protein KW-803 column (Thomson Instruments) pre-calibrated with gel-filtration standards (Bio-Rad).

2.3. Crystallization

BcYkfC was crystallized by mixing 200 nl protein solution with 200 nl crystallization solution and equilibrating against a 50 µl reservoir solution using the nanodroplet vapor-diffusion method (Santarsiero *et al.*, 2002) with standard Joint Center for Structural Genomics (JCSG; <http://www.jcsg.org>) crystallization protocols (Lesley *et al.*, 2002). The crystallization solution was composed of 0.2 M sodium chloride, 50% (v/v) PEG 200 and 0.1 M phosphate-citrate pH 4.2. A needle-shaped crystal of approximate dimensions 100 × 15 × 15 µm was harvested after 29 d at 277 K. No additional cryoprotectant was added to the crystal. Initial screening for diffraction was carried out using the Stanford Automated Mounting system (SAM; Cohen *et al.*, 2002) at the Stanford Synchrotron Radiation Lightsource (SSRL, Menlo Park, California, USA).

2.4. Data collection, structure solution and refinement

Multi-wavelength anomalous diffraction (MAD) data were collected at wavelengths corresponding to the peak, high-energy remote and inflection wavelengths of a selenium MAD experiment at 100 K using a MAR CCD 325 detector (Rayonix) on SSRL beamline 11-1. Processing of the diffraction data and initial structure solution were carried out using the automatic structure-solution script *autoXDSp* developed at the JCSG (unpublished work). This script shepherds the structure-determination process, as summarized below, using preset rules through a decision-tree that mimics that used by an experienced crystallographer. The calculations are parallelized on a computer cluster such that initial maps and models can usually be obtained within 1 h of the completion of data collection. In summary, the MAD data were integrated and reduced using *XDS* and then scaled with the program *XSCALE* (Kabsch, 1993, 2010). Selenium sites were located with *SHELXD* (Sheldrick, 2008). Phase refinement and automatic model building were performed using *autoSHARP* (Bricogne *et al.*, 2003) and *ARP/wARP* (Cohen *et al.*, 2004). This automated process produced an initial model that was 92% complete. Further model completion and refinement were performed manually with *Coot* (Emsley & Cowtan, 2004) and *REFMAC* (Murshudov *et al.*, 1997) from the *CCP4* suite (Collaborative Computational Project,

Table 1

Data-collection, phasing and refinement statistics (PDB code 3h41).

Values in parentheses are for the highest resolution shell. The high-resolution cutoff was chosen such that the mean $I/\sigma(I)$ in the highest resolution shell was around 2.

	λ_1 MADSe, peak	λ_2 MADSe, remote	λ_3 MADSe, inflection
Space group	C2		
Unit-cell parameters (Å, °)	$a = 95.2, b = 59.4, c = 61.3, \beta = 103.3$		
Data collection			
Wavelength (Å)	0.9786	0.9184	0.9799
Resolution range (Å)	41.5–1.79 (1.88–1.79)	41.5–1.84 (1.94–1.84)	41.5–1.86 (1.96–1.86)
No. of observations	116822	107729	104295
No. of unique reflections	31085	28629	27652
Completeness (%)	98.0 (95.1)	98.6 (98.1)	98.5 (97.9)
Mean $I/\sigma(I)$	11.3 (2.1)	12.5 (2.8)	13.1 (2.8)
R_{merge} on I^\dagger (%)	11.1 (71)	10.0 (54)	10.0 (56)
MAD phasing			
Resolution	41.5–1.79		
No. of Se sites	5		
Mean figure of merit	0.36		
Model and refinement statistics			
Resolution range (Å)	41.5–1.79		
No. of reflections (total)	31083		
No. of reflections (test)	1564		
Completeness (%)	98.0		
Data set used in refinement	λ_1 MADSe		
Cutoff criterion	$ F > 0$		
$R_{\text{cryst}}^\ddagger$ (%)	16.3		
R_{free}^\S (%)	19.7		
Stereochemical parameters			
Restraints (r.m.s.d. observed)			
Bond lengths (Å)	0.015		
Bond angles (°)	1.50		
Average isotropic B value (Å ²)	19.9		
ESU [¶] based on R_{free} (Å)	0.11		
Protein residues/atoms	305/2463		
Waters/peptide/other ligands	265/1/7		
<i>MolProbity</i> statistics			
All-atom clash score	5.07		
Ramachandran favored (%)	97.7		
No. of Ramachandran outliers	1		
No. of rotamer outliers	1		

[†] $R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$. [‡] $R_{\text{cryst}} = \frac{\sum_{hkl} |F_{\text{obs}}| - |F_{\text{calc}}|}{\sum_{hkl} |F_{\text{obs}}|}$, where F_{calc} and F_{obs} are the calculated and observed structure-factor amplitudes, respectively. [§] R_{free} is the same as R_{cryst} but for 5.0% of the total reflections chosen at random and omitted from refinement. [¶] Estimated standard uncertainty in atomic coordinates.

Number 4, 1994). Data and refinement statistics are summarized in Table 1. Analysis of the stereochemical quality of the model was accomplished using *MolProbity* (Chen *et al.*, 2010). All molecular graphics were prepared with *PyMOL* (DeLano Scientific) unless specifically stated otherwise. Atomic coordinates and experimental structure factors for BcYkfC at 1.8 Å resolution have been deposited in the PDB with accession code 3h41.

2.5. Molecular modeling

Molecular docking was performed using the same protocol as described previously (Xu *et al.*, 2009) using *Glade* v.5.0 (Schrödinger LLC). The positions of the bound ligand were used as restraints such that the L-Ala- γ -D-Glu portion of the docked substrate adopted a similar conformation as seen in the crystal structure. In order to perform the docking studies, the tri-oxidized cysteine (OCS238) in the crystal structure was replaced with the reduced form, which is needed for reaction in the papain family of cysteine peptidases. Furthermore, only the side-chain conformer with higher occupancy in the crystal structure was considered in the docking experiments when multiple conformations were observed for an active-site residue.

Table 2

Structural comparisons of BcYkfC and other bacterial proteins that share at least one common domain.

The alignment was performed by the DALI structural comparison server using full-length BcYkfC and individual domains (SH3b1 and NlpC/P60) as search probes. For proteins with multiple SH3-like domains (PDB codes 1m9s and 1xov) or multiple chains, only the best match is shown.

Aligned substructure(s)	PDB code	Reference	R.m.s.d. (Å)	Aligned length	No. of residues†	Z score	Sequence identity (%)	Comments
SH3b + NlpC/P60	2fg0	Xu <i>et al.</i> (2009)	2.1	193	221	21.3	26	Cyanobacterial NpPCP
	2hbw	Xu <i>et al.</i> (2009)	2.3	193	220	21.2	29	Cyanobacterial AvPCP
SH3b	1m9s	Marino <i>et al.</i> (2002)	1.9	60	523	6.0	12	TD‡ of internalin B
	1r77	Lu <i>et al.</i> (2006)	2.3	57	103	4.7	12	TD‡ of PG hydrolase ALE-1
	1xov	Korndorfer <i>et al.</i> (2006)	3.0	51	317	3.5	10	TD‡ of endolysin PlyPSA
	2akk	Srisailam <i>et al.</i> (2006)	1.8	44	74	2.7	7	PhnA-like protein
	1x27	Nasertorabi <i>et al.</i> (2006)	1.9	54	164	5.7	13	Eukaryotic SH3
NlpC/P60	2jyx	Aramini <i>et al.</i> (2008)	1.9	119	129	17.8	29	<i>E. coli</i> lipoprotein Spr
	2p1g	NYSGXRC§	2.5	110	230	9.6	18	Uncharacterized
	2ioa	Pai <i>et al.</i> (2006)	3.2	106	589	7.3	9	CHAP domain of GspS¶

† Number of residues present in the model used for comparison. ‡ Targeting domain. § New York SGX Research Center for Structural Genomics (unpublished work). ¶ Glutathionylspermidine synthetase/amidase.

3. Results and discussion

3.1. Genomic context

Full-length BcYkfC (strain *B. cereus* ATCC 10987) contains 333 residues (molecular weight 37.3 kDa), the first 23 of which are predicted to be a signal peptide by the Phobius web server (Kall *et al.*, 2007). Despite being homologous to BcYkfC, *B. subtilis* YkfC (296 residues) does not contain a signal peptide, suggesting that the two proteins function at different cellular locations.

As in *B. subtilis*, the *ykfC* gene (BCE_2878) is adjacent to the *ykfB* L-Ala-D-Glu epimerase gene (BCE_2879) in the *B. cereus* genome. The YkfB epimerases (sequence identity 50%) from both bacteria do not contain predicted signal peptides. The genome association of *ykfB* and *ykfC* is also observed in other bacteria (*e.g.* *Listeria innocua*, *Acidobacterium* sp., *Solibacter usitatus*, *Bacteroides thetaiotaomicron* and *Gramella forsetii*). However, the genome context of *ykfBC* is different in *B. subtilis* compared with *B. cereus* (Fig. 2). The *YkfA–D* genes of *B. subtilis* are located next to the *dppB–E* dipeptide ABC transporter operon, whereas the *YkfBC* genes of *B. cereus* are located downstream of *divIC*, which encodes a putative cell-division protein. Downstream of *ykfC* is an *oppA* gene that is homologous to *dppE* (34% sequence identity). Both genes are predicted to encode periplasmic dipeptide-binding proteins.

Based on the presence/absence of the signal peptide on YkfC and YkfB, as well as their genomic contexts, we suggest that the strategy for metabolizing murein peptides is likely to differ between *B. subtilis* and *B. cereus*. In *B. cereus*, the murein peptides are likely to be broken down outside the cell, with the resulting dipeptide L-Ala-γ-D-Glu being imported into cytoplasm for further processing by YkfB, while the reactions catalyzed by both YkfC and YkfB are likely to occur in the cytoplasm in *B. subtilis*.

3.2. Structure determination and quality of the model

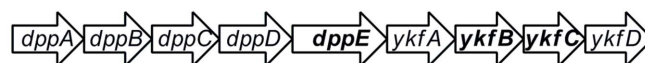
The crystal structure of BcYkfC was determined using the high-throughput structural genomics pipeline implemented at the JCSG (Lesley *et al.*, 2002). The selenomethionine derivative of BcYkfC was expressed in *E. coli* with an N-terminal TEV-cleavable His tag and purified by metal-affinity chromatography. In order to improve the chance of obtaining crystals, the predicted N-terminal signal peptide (residues 1–23) was not included in the cloned construct. The data were indexed in space group C2 and the structure was determined at 1.79 Å resolution with one molecule per asymmetric unit using the MAD method ($R_{\text{cryst}} = 16.3\%$, $R_{\text{free}} = 19.7\%$). The electron density for the main chain was well defined throughout the entire molecule.

The mean residual error of the coordinates was estimated to be 0.11 Å by the diffraction-component precision index (DPI) method (Cruickshank, 1999). The model of BcYkfC displays good geometry, with an all-atom clash score of 5.07, and the Ramachandran plot produced by MolProbity (Chen *et al.*, 2010) shows that all residues, but one, are in allowed regions, with 97.7% in favored regions. Only one residue is flagged as a rotamer outlier. The Ramachandran (Pro305) and rotamer (His303) outliers are supported by well defined electron density. Since these residues are either close to or part of the active site, these structural deviations from ideality are likely to be of functional relevance. Additionally, two *cis*-peptides (Asn57–Pro58 and Asn154–Pro155) are also supported by clear electron density. The final model of BcYkfC contains residues 29–333, one dipeptide L-Ala-γ-D-Glu, one phosphate, six polyethylene glycol (PEG) fragments from the crystallization solution and 265 waters. The residual residue (Gly0) from the cleaved N-terminal purification tag, residues 24–28 and the side chains of Glu201 and Arg270 were disordered and were not included in the final model. Data-collection, refinement and model statistics are summarized in Table 1.

3.3. Overall structure

BcYkfC is likely to be a monomer in solution, as supported by crystal-packing analysis and analytical size-exclusion chromatography. The structure of BcYkfC consists of three domains: two SH3b domains (SH3b1, residues 29–129; SH3b2, residues 130–207) and a C-terminal NlpC/P60 cysteine peptidase domain (residues 208–333) (Figs. 3a and 3b). The two SH3b domains are similar to each other, with an r.m.s.d. of 2.2 Å for 57 aligned C α atoms (sequence identity of 13%). A structural similarity search using DALI (Holm & Sander, 1995) did not find any other structures with the same three-domain architecture. However, a number of significantly similar substructures were identified and are summarized in Table 2. The SH3b domains of BcYkfC are similar to other SH3b domains (Holm & Sander, 1995),

B. subtilis



B. cereus

**Figure 2**Genomic context of the *ykfB* and *ykfC* genes in *B. subtilis* and *B. cereus*.

including the N-terminal domain of cyanobacterial γ -D-Glu-DAP endopeptidases (Xu *et al.*, 2009), the GW domains of internalin B (Marino *et al.*, 2002), the cell-wall-targeting domain of glycyglycine endopeptidase ALE-1 (Lu *et al.*, 2006), PhnA-like protein (Srisailam *et al.*, 2006) and endolysin PlyPSA (Korndorfer *et al.*, 2006), as well as many eukaryotic SH3 domains. A β -hairpin within the so-called 'RT loop' (*i.e.* the loop between β A and β B) region appears to be a common and unique structural feature of SH3b domains compared with their eukaryotic counterparts (Xu *et al.*, 2009). The BcYkfC

SH3b domains both contain this conserved structural motif (β A1– β A2). However, SH3b1 contains a novel helical insertion (α 1– α 3) that is not seen in previous SH3b structures or in SH3b2 (Fig. 3*b*). Prokaryotic SH3-like domains have also been implicated in poly-peptide binding (Wylie *et al.*, 2005) and metal binding (Pohl *et al.*, 1999). Although these SH3-like domains contain a similar five-stranded (β A– β E) core, they display much larger structural differences (and are not among significant DALI hits with $Z > 2.0$) and lack the β -hairpin in the RT-loop region when compared with SH3b

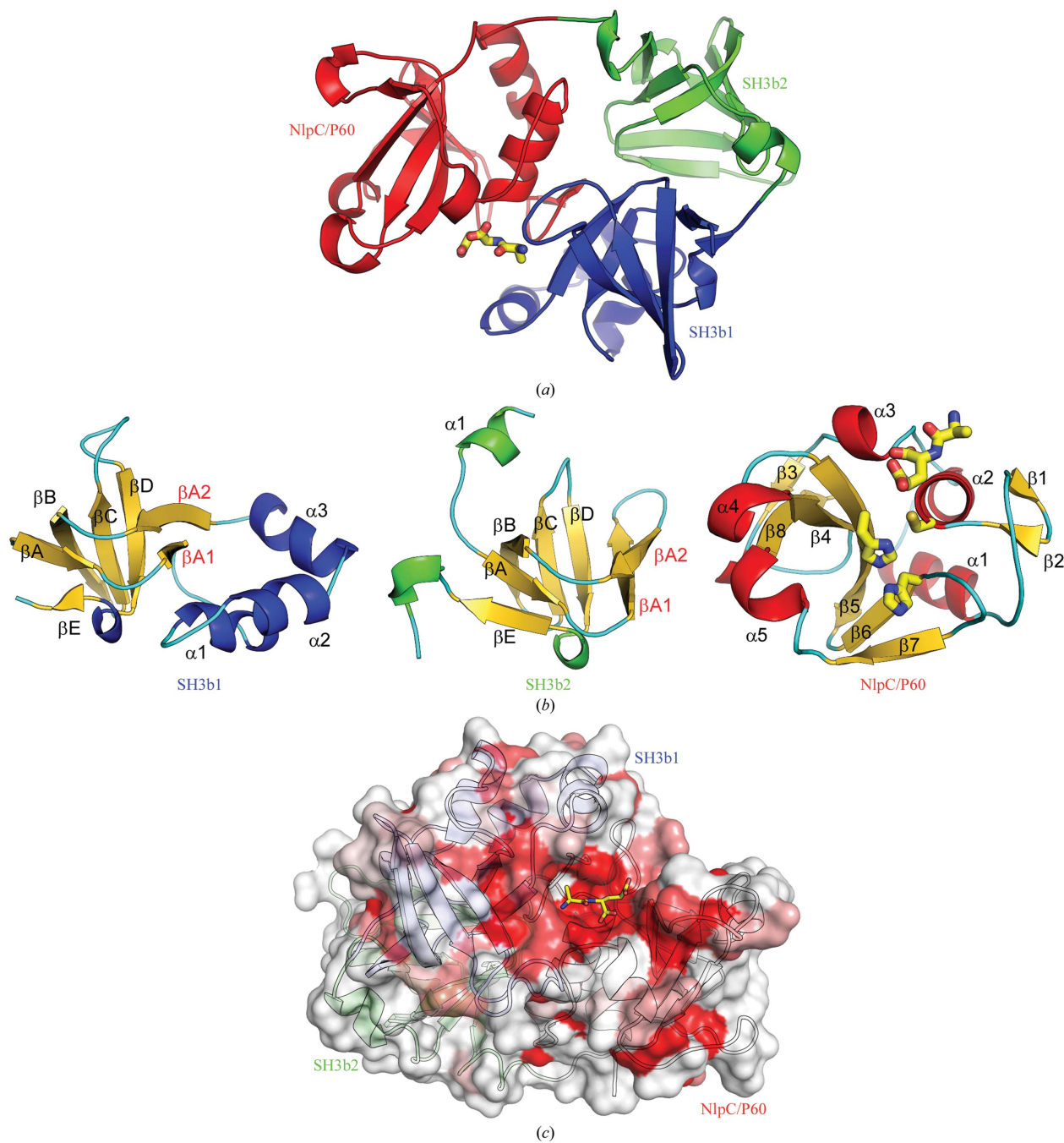


Figure 3 Crystal structure of YkfC from *B. cereus* in complex with L-Ala- γ -D-Glu. (a) Ribbon representation of BcYkfC, highlighting its domain organization. SH3b1 is depicted in blue, SH3b2 in green and NlpC/P60 in red. The bound L-Ala- γ -D-Glu is shown as a stick model. (b) Ribbon representations of individual domains, showing the secondary-structure elements. (c) Molecular surface of YkfC colored by sequence conservation. The surface color gradient indicates the level of sequence conservation from the most conserved residues (deep red) to nonconserved residues (white).

domains. The C-terminal NlpC/P60 catalytic domain of BcYkfc is remotely related to the papain family of cysteine peptidases (Anantharaman & Aravind, 2003), with highest similarity to cyanobacterial γ -D-Glu-DAP endopeptidases (Xu *et al.*, 2009), *E. coli* lipoprotein Spr (Aramini *et al.*, 2008) and two uncharacterized proteins (PDB codes 2p1g and 2im9; NYSGXRC, unpublished work) (Table 2).

The three domains of BcYkfc are arranged in a triangle such that each domain interacts with the two other domains. The interface ($\sim 570 \text{ \AA}^2$ per domain) between the two SH3b domains is mostly hydrophobic and is centered on interactions between the β A– β A1 and β A2– β B loops of SH3b1 and the β A2 and β B strands and β D– β E loop of SH3b2. The SH3b1–NlpC/P60 domain interface (903 \AA^2 buried surface per domain) is mediated through α 2– α 3– β A1 and the β C– β D loop of SH3b1, and α 1– α 2– α 3 and the β 6– β 7 loop of NlpC/P60. The active site is located at this SH3b1–NlpC/P60 interface, whereas the SH3b2 domain is distal to the active site. A multiple sequence alignment of full-length homologs of Ykfc (37 sequences; average sequence identity of 54%) indicates that most of the highly conserved residues are either buried inside the protein or clustered around the active site (Fig. 3c).

3.4. Active site

The catalytic triad of the peptidase domain consists of Cys238, His291 and His303. The catalytic Cys238 is oxidized (OCS) in the

crystal based on the electron density (Fig. 4a). Since an oxidized cysteine can no longer function as a nucleophile in the reaction (Storer & Ménard, 1994), the enzyme in the crystal is inactive. The conformation of the cysteine side chain is not significantly affected by the oxidation as its side chain is in a similar location and conformation as in other NlpC/P60 structures (Xu *et al.*, 2009).

The dipeptide L-Ala- γ -D-Glu was identified from well defined electron density in the active site of BcYkfc (Fig. 4a). As the dipeptide was not added during protein purification or crystallization, it was most likely to have been obtained during protein expression in *E. coli*. Since this dipeptide is one of the reaction products of BcYkfc (Fig. 2), it unequivocally identifies the S2–S1 binding-site cavity, which is formed by residues from both the SH3b1 (Glu83, Thr84 and Tyr118) and NlpC/P60 domains (Tyr226, Trp228, Ala229, Asp237, Arg255, Asp256, Ser257, His290 and His291). Two charged residues, Asp237 and Arg255, which are highly conserved in NlpC/P60 domains, are involved in a hydrogen-bond network that connects many residues in the active site (Fig. 4b).

Two active-site residues, Ser257 and His291, display two discrete side-chain conformations (Fig. 4b). The first conformer of His291 (occupancy modeled as 0.7), which facilitates hydrogen bonding to His303, is identical to the corresponding histidine of the catalytic dyad in papain and other NlpC/P60 enzymes. The second rotamer points His291 towards the solvent. The side-chain isomerism in the active site is likely to be a consequence of the observed oxidized state

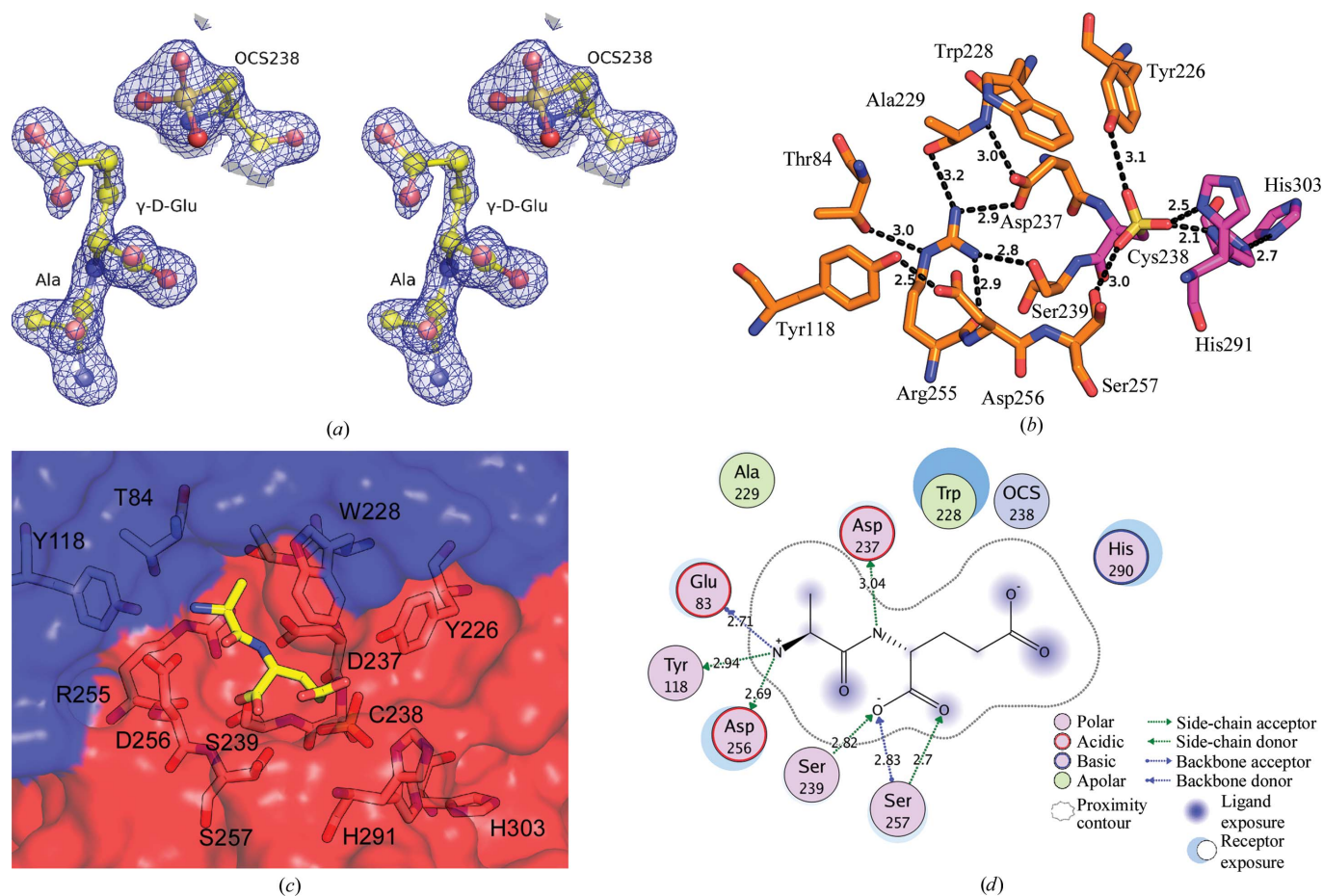


Figure 4 Active site and recognition of L-Ala- γ -D-Glu by BcYkfc. (a) Stereoview of a $2F_o - F_c$ OMIT map, where L-Ala- γ -D-Glu and OCS238 were omitted from phasing/refinement, contoured at 1.5σ . (b) The extensive hydrogen-bond network in the active site of BcYkfc. Hydrogen bonds and distances are shown as dashed lines. (c) L-Ala- γ -D-Glu (stick representation; yellow C atoms) is located in the active site at the interface of the SH3b1 domain (blue) and the NlpC/P60 domain (red). (d) The interaction between L-Ala- γ -D-Glu and the active site of Ykfc. This figure was generated using the program MOE 2008.10 (Chemical Computing Group Inc.).

of the catalytic Cys238, since the oxidized Cys238 makes multiple hydrogen-bond interactions with nearby side chains, including Tyr226, Ser257 and His291 (Fig. 4b).

3.5. Recognition of L-Ala- γ -D-Glu

The active-site residues form a pocket that is highly complementary in shape and chemical properties to the ligand (Figs. 4c and 4d). The average *B* value for the bound ligand is 26 Å² (the overall *B* value of the protein is 18 Å²), indicating that it is well ordered with almost full occupancy in the active site. The interface between the dipeptide and the protein buries a total surface area of 490 Å². The

dipeptide is stabilized by multiple hydrogen bonds. The free amine of L-Ala in the S2 pocket makes hydrogen bonds to Glu83 O and the side chains of Tyr118 (OH) and Asp256 (O^{δ1}), which are highly conserved in the YkfC subfamily of NlpC/P60 enzymes (see below). The α -NH group of γ -D-Glu in the S1 site forms a weak hydrogen bond to Asp237 O^{δ2}. The α -carboxyl of D-Glu is stabilized by hydrogen-bonding interactions with Ser239 and Ser257 (Fig. 4d). On the solvent-exposed side of the substrate, waters are also involved in the hydrogen-bond network with the substrate and the enzyme (not shown). The L-Ala methyl side chain points towards a hydrophobic pocket defined by the side chains of Trp228 and Ala229. Additionally, the aliphatic C atoms of D-Glu are involved in hydrophobic contacts

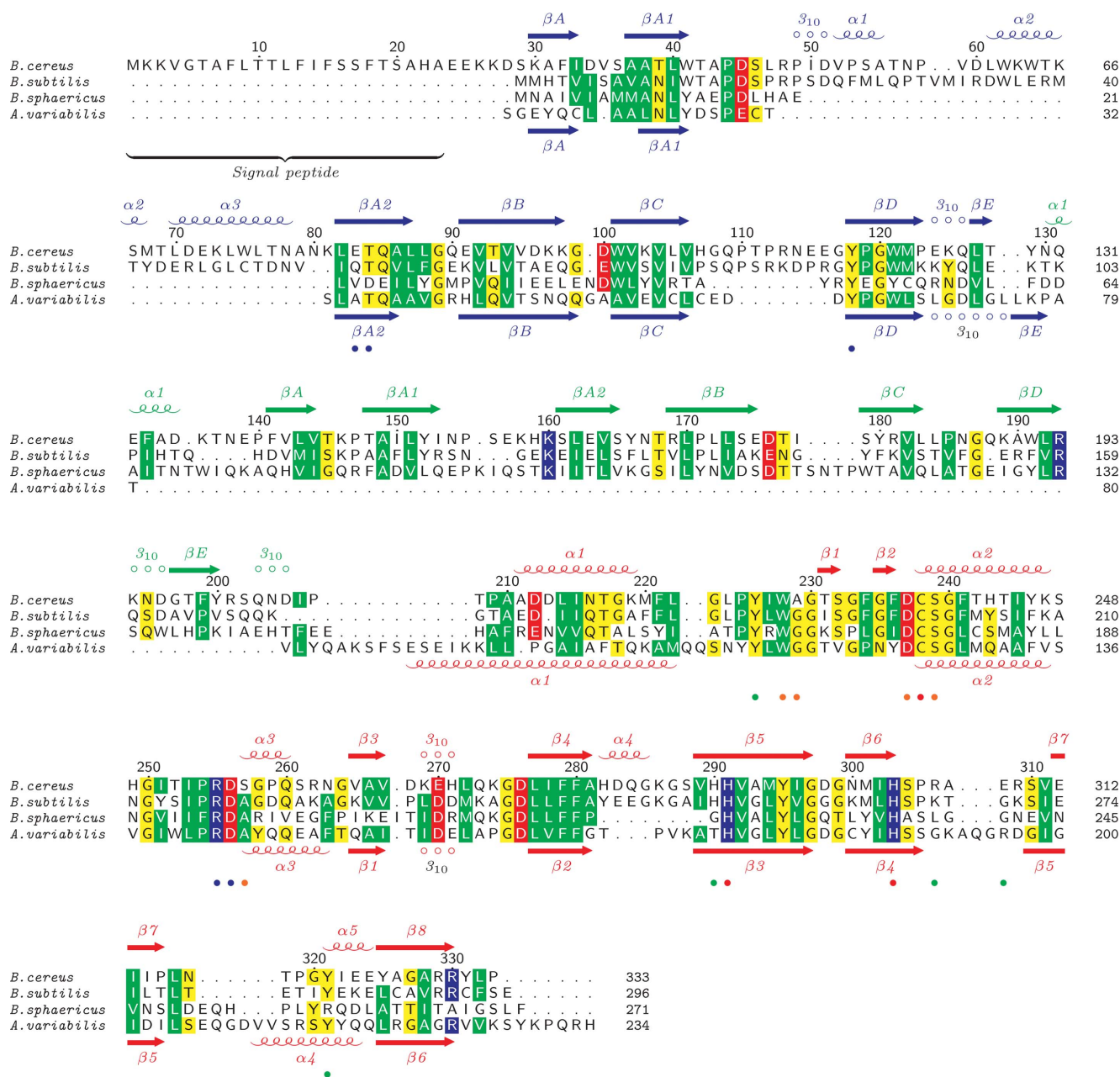


Figure 5 Sequence alignment of YkfC from *B. subtilis* and *B. cereus*, dipeptidyl-peptidase VI (DPP VI) from *B. sphaericus* and a γ -D-glutamyl-L-diamino acid endopeptidase from *A. variabilis* (AvPCP). The sequence numbering and secondary-structure elements of YkfC from *B. cereus* and AvPCP are indicated at the top and bottom, respectively. The alignment was generated by merging and manually editing the structure-based sequence alignment of BcYkfC and AvPCP with the sequence alignment of the top three sequences. The active-site residues are marked with colored dots at the bottom (blue, S2; orange, S1; red, catalytic triad; green, potential S' sites).

with Trp228. As a result, Trp228 contributes to both the S2 and S1 binding sites.

3.6. YkfC is specific for murein peptides with free N-terminal L-Ala

The active site of *B. subtilis* YkfC is highly conserved compared with that of BcYkfC (Fig. 5). Dipeptidyl peptidase VI (DPP VI) from

B. sphaericus is a γ -D-Glu-DAP(Lys) dipeptidase that is found in the cytoplasm during sporulation (Vacheron *et al.*, 1979). DPP VI has strict specificity for murein peptides with an N-terminal L-Ala. The crystal structure of BcYkfC provides a structural basis for this specificity. DPP VI and BcYkfC have only 21% sequence identity, but the same sets of key residues are conserved, as expected given their similar folds and function (Fig. 5). Thus, DPP VI is clearly a homolog

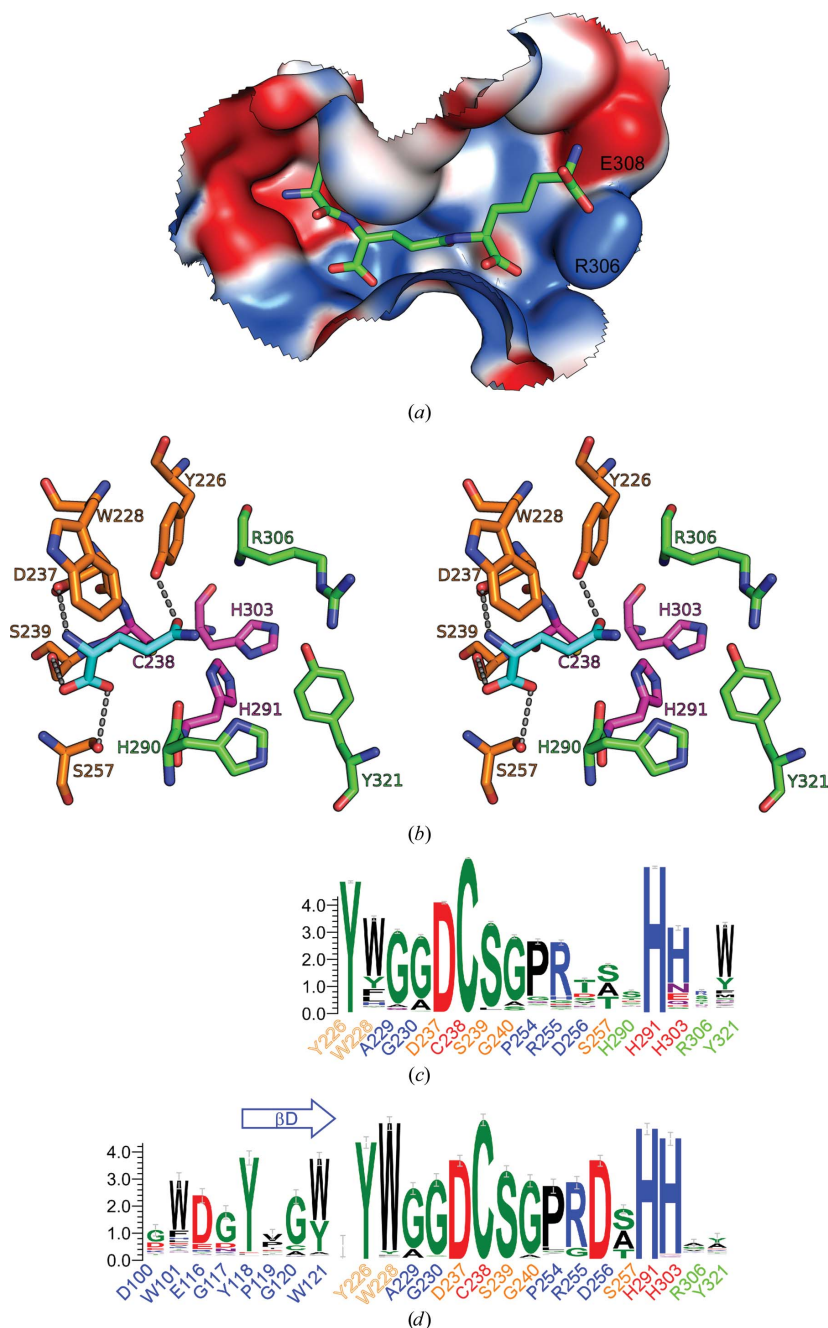


Figure 6

Models of substrate recognition by BcYkfC. (a) L-Ala- γ -D-Glu-DAP was docked into the active site of BcYkfC. The protein surface is colored according to a gradient in electrostatic potential from negative (red) to positive (blue) (MOE 2008.10; Chemical Computing Group Inc.). (b) Stereoview of the specific interactions (four polar, one nonpolar) of γ -D-Glu (cyan) in the context of the tripeptide by five residues of YkfC (Tyr226, Trp228, Asp237, Ser239 and Ser257). The protein residues are colored according to subsite (S1, orange; catalytic triad, magenta; S1', green). (c) Sequence conservation of the active sites in NlpC/P60 domains based on 2277 NlpC/P60 domains with an intact catalytic dyad (Cys238 and His291) and a conserved Tyr226 (blue, S2; orange, S1; red, catalytic triad; green, S' sites). (d) Sequence conservation of the active sites in the YkfC subfamily of NlpC/P60 enzymes based on 282 sequences selected based on the presence of a conserved aspartate residue at position 256 of BcYkfC. The conservation of this residue is highly correlated with a conserved tyrosine in the β D region of the SH3b domain (Tyr118 of BcYkfC); both of these residues interact with the free amine of L-Ala of the substrate.

of BcYkfc, except that its SH3b1 domain has no large helical insertion ($\alpha 1$ – $\alpha 3$) between $\beta A1$ and $\beta A2$ and it has a shorter βC – βD loop (Fig. 5). Furthermore, the S2 and S1 binding sites for L-Ala- γ -D-Glu are highly conserved between DPP VI and BcYkfc, including the conserved tyrosine (Tyr118 of Ykfc), which is the only residue from SH3b1 whose side chain interacts with L-Ala. The residues from the catalytic domain that contribute to the S2 and S1 sites are identical in BcYkfc and DPP VI (except for an Ala/Ser mutation at position 257 of BcYkfc). Therefore, we conclude that both the *B. subtilis* and the *B. cereus* Ykfc are likely to have very similar substrate specificity for murein peptides that contain an N-terminal L-Ala. Enzymes with this specificity cannot cleave PG-biosynthesis precursors such as UDP-MurNAc-pentapeptide and thus are not likely to interfere with the PG-synthesis pathway.

3.7. S' sites of Ykfc and docking studies

B. subtilis Ykfc has previously been shown to process the tripeptide L-Ala- γ -D-Glu-L-Lys, the tetrapeptide L-Ala- γ -D-Glu-L-Lys-D-Ala and the pentapeptide L-Ala- γ -D-Glu-L-Lys-D-Ala-D-Ala (Schmidt *et al.*, 2001). DPP VI also displayed no particular specificity towards the additional residues attached after DAP (or Lys; Vacheron *et al.*, 1979). Residues in BcYkfc that potentially form the S' sites include His290, His291, Arg306, Glu308, Arg309, Tyr321 and Glu324. These residues are generally less conserved among close BcYkfc homologs (Fig. 5), suggesting that the specificity of Ykfc is determined primarily at the S2 and S1 sites as described above. The catalytic mechanism of NlpC/P60 is currently unknown, but it is likely to be similar to that of papain based on the similar arrangement of their catalytic residues (Anantharaman & Aravind, 2003; Aramini *et al.*, 2008; Xu *et al.*, 2009), except for two interesting variations of important catalytic residues. The third polar residue of the catalytic triad in NlpC/P60 is more often a histidine rather than the asparagine in papain (Bateman *et al.*, 2004). A tyrosine (Tyr226 of BcYkfc) is located at the position equivalent to Gln119 in papain that is thought to be important for catalysis by stabilizing the transition state (Xu *et al.*, 2009; Aramini *et al.*, 2008). Therefore, we docked the tripeptide L-Ala- γ -D-Glu-L-DAP into the active site of BcYkfc in order to deduce a likely mode of interaction and to assess the effects of the differences (Fig. 6a). The C ^{δ} carboxyl group of γ -D-Glu is displaced owing to the oxidation of Cys238 in the crystal structure. In the modeled structure, Tyr226 OH interacts with the carboxyl group of γ -D-Glu (Fig. 6b). This docked structure places the C ^{δ} atom close to Cys238 SH (distance ~ 3.6 Å) and could represent the chemically productive conformation. The electrostatic surface of the binding site is highly complementary to that of the substrate, indicating that the polar and charged interactions are likely to be important for substrate recognition. The basic His290 and Arg306 residues could interact with carboxyl groups. It has previously been reported that the catalytic efficiency of *B. subtilis* Ykfc for L-Ala- γ -D-Glu-Lys is significantly less than that for tetrapeptides or pentapeptides (Schmidt *et al.*, 2001). This observation could be explained by the unfavorable basic environment in the S' site. Two residues (Arg306/Glu308) at the BcYkfc S1' site are substituted by Lys/Gly in *B. subtilis* Ykfc (Fig. 5), resulting in a subsite with a positive charge but no negative charge that would be less likely to accommodate the positively charged Lys of L-Ala- γ -D-Glu-Lys (Fig. 6a).

3.8. Recognition of γ -D-Glu by NlpC/P60 cysteine peptidases

NlpC/P60 cysteine peptidases are ubiquitous in bacteria, with ~ 5000 NlpC/P60 domains in the current Pfam database when metagenomics data are included (Bateman *et al.*, 2004). However, only a

few of these domains have been biochemically characterized. In the light of the much clearer picture that we now have of substrate binding in BcYkfc, we examined the sequence conservation of NlpC/P60-family proteins around the active site in order to obtain further insights into substrate specificity across the entire family.

The nonredundant (nr) database of protein sequences was searched using the *PSI-BLAST* program (Altschul *et al.*, 1997) with the NlpC/P60 domain of BcYkfc as a search probe. We extracted a total of 2599 protein sequences using the criteria of an *E* value of >0.02 and an alignment length of >70 residues from 3985 total hits, representing a significant subset of NlpC/P60 proteins. Among these proteins, only $\sim 1\%$ do not possess the conserved Cys/His dyad and an additional 9% (with catalytic dyad intact) do not contain the conserved tyrosine that is believed to be essential for catalysis. These proteins are likely to have the NlpC/P60 fold, but either have lost cysteine peptidase activity or are remote homologs that hydrolyze different substrates (e.g. the CHAP family; Anantharaman & Aravind, 2003; Bateman & Rawlings, 2003; Rigden *et al.*, 2003). The sequence-conservation pattern of active-site residues of the remaining 2277 proteins (average sequence identity of 19%) is shown in Fig. 6(c). The S sites (S2 and S1) are significantly more conserved than the S' sites in these proteins. The most conserved site is S1, which is tailored to recognize γ -D-Glu (Fig. 6b). The three most highly conserved residues (Asp237, Ser239 and Tyr226), aside from the catalytic dyad, are involved in hydrogen bonding with γ -D-Glu (Fig. 6b). At position 257, residues with small side chains (Ala/Ser/Thr) are observed, creating a cavity for the carboxyl group of γ -D-Glu. Trp228 contributes to both the S1 and S2 sites and a conserved hydrophobic residue is usually found at this position, which may facilitate interaction with the substrate Ala C ^{β} . Therefore, the conservation of active-site residues indicates that a significant percentage of the NlpC/P60 family is likely to be specific for γ -D-Glu. More than 1300 proteins which contain the strictly conserved Trp228, Asp237, Ser239 and Tyr226 residues (and catalytic triads) are most likely to bind X-L-Ala- γ -D-Glu moieties (where X is H or other moieties).

3.9. Structural comparison to the cyanobacterial NlpC/P60 L-Ala- γ -D-Glu endopeptidases

We have previously determined the crystal structures of two closely related NlpC/P60 L-Ala- γ -D-Glu endopeptidases from the cyanobacteria *A. variabilis* (AvPCP) and *N. punctiforme* (NpPCP). These enzymes contain an N-terminal SH3b domain and a C-terminal NlpC/P60 catalytic domain (Xu *et al.*, 2009). Surprisingly, the structures of the cyanobacterial enzymes are essentially a substructure of Ykfc: the two domains of the cyanobacterial enzymes are structurally equivalent to the first and third domains of BcYkfc (Figs. 5, 7a and 7b). The full-length AvPCP can be superimposed onto BcYkfc with an r.m.s.d. of 2.3 Å and 29% sequence identity for 193 aligned C ^{α} atoms (Table 2). Thus, the individual SH3b and NlpC/P60 domains, as well as their relative arrangements, are highly similarly in AvPCP and BcYkfc. Furthermore, the residues in the S2 and S1 pockets are nearly identical (except for S257A; Fig. 7c). The striking similarity between the cyanobacterial enzymes and Ykfc was not previously detected by sequence analysis owing to the presence of two insertions: the entire SH3b2 domain and the $\alpha 1$ – $\alpha 3$ region of SH3b1.

The βC – βD loop and the $\beta A1$ – $\beta A2$ loop of SH3b1 are longer compared with SH3b of AvPCP. The βC – βD loop of Ykfc is located upstream of the S2 site. The $\beta A1$ – $\beta A2$ loop of Ykfc contains three helices ($\alpha 1$ – $\alpha 3$) and is located on the S' side of the binding groove in the catalytic domain. This insertion packs against the surface of the NlpC/P60 domain and contributes to stabilizing the interface between

the SH3b1 and NlpC/P60 domains. Furthermore, it also defines the shape of the S' binding site, which appears to be more restricted for BcYkfC than for AvPCP (Fig. 7*d*). Without these insertions to stabilize this domain interface, AvPCP uses a different strategy to achieve a similar purpose: a longer C-terminal loop in NlpC/P60 extends to interact with the SH3b domain.

We previously proposed that the cyanobacterial enzymes have the same substrate requirement as DPP VI based on modeling studies (Xu *et al.*, 2009). The crystal structure of YkfC reported here further supports DPP VI, YkfC and the cyanobacterial enzymes belonging to a subset of NlpC/P60 γ -D-Glu-DAP endopeptidases whose substrates possess a free L-Ala at the N-terminus. The SH3b domain helps define the S2 binding site (Tyr118 in BcYkfC and Tyr64 in AvPCP). Additionally, it sterically hinders the docking of a large moiety beyond the S2 site and directly contributes to specificity. This role constitutes a new function for SH3b domains.

3.10. Identification and distribution of YkfC homologs

BcYkfC and cyanobacterial γ -D-Glu-DAP endopeptidases have the same specificity. Other NlpC/P60 enzymes with similar properties could potentially be detected by analyzing sequence similarities in both the SH3b and NlpC/P60 domains. However, owing to the presence of long insertions/deletions and highly divergent sequences in the SH3b domains, it is often difficult to detect similar enzymes using sequence searches and the NlpC/P60 region tends to dominate the

hits. We examined an alternate method by examining specific residues in the NlpC/P60 domain only. An acidic residue (Asp256) is essential for YkfC specificity by neutralizing the positive charge on the free amine of L-Ala. Among the 2599 sequences obtained above, 282 proteins were identified that contained an aspartate equivalent to Asp256 and a conserved catalytic dyad. In 239 of the 282 proteins (85%), the conserved tyrosine was maintained in their SH3b domains (Tyr118 of BcYkfC), while this tyrosine was not conserved in enzymes that lacked Asp256. Thus, the presence of an aspartate residue at position 256 and a conserved tyrosine in SH3b is highly correlated (Fig. 6*d*). Interestingly, NlpC/P60 domains with the conserved aspartate are also found in a few single-domain or multiple-domain proteins that do not contain any detectable SH3b domain. The biochemical properties of these proteins are currently unknown, but they could represent novel variants of a YkfC-type enzyme.

Nevertheless, the 239 proteins above with the conserved aspartate (Asp256) and tyrosine (Tyr118) are most likely to define a YkfC-like subfamily of NlpC/P60 proteins. These NlpC/P60 enzymes are predominately distributed in four phyla of bacteria: bacteroidetes, cyanobacteria, firmicutes and proteobacteria (α -proteobacteria). They include all currently known enzymes with the same requirement for a free N-terminal L-Ala, *i.e.* DPP VI, YkfC and AvPCP/NpPCP. The cyanobacterial enzymes contain one SH3b domain, while the homologs in other bacteria contain two SH3b domains. As expected, the SH3b domains are highly divergent. The β D strand, in which the conserved tyrosine is located, is more conserved (Fig. 6*d*). This strand

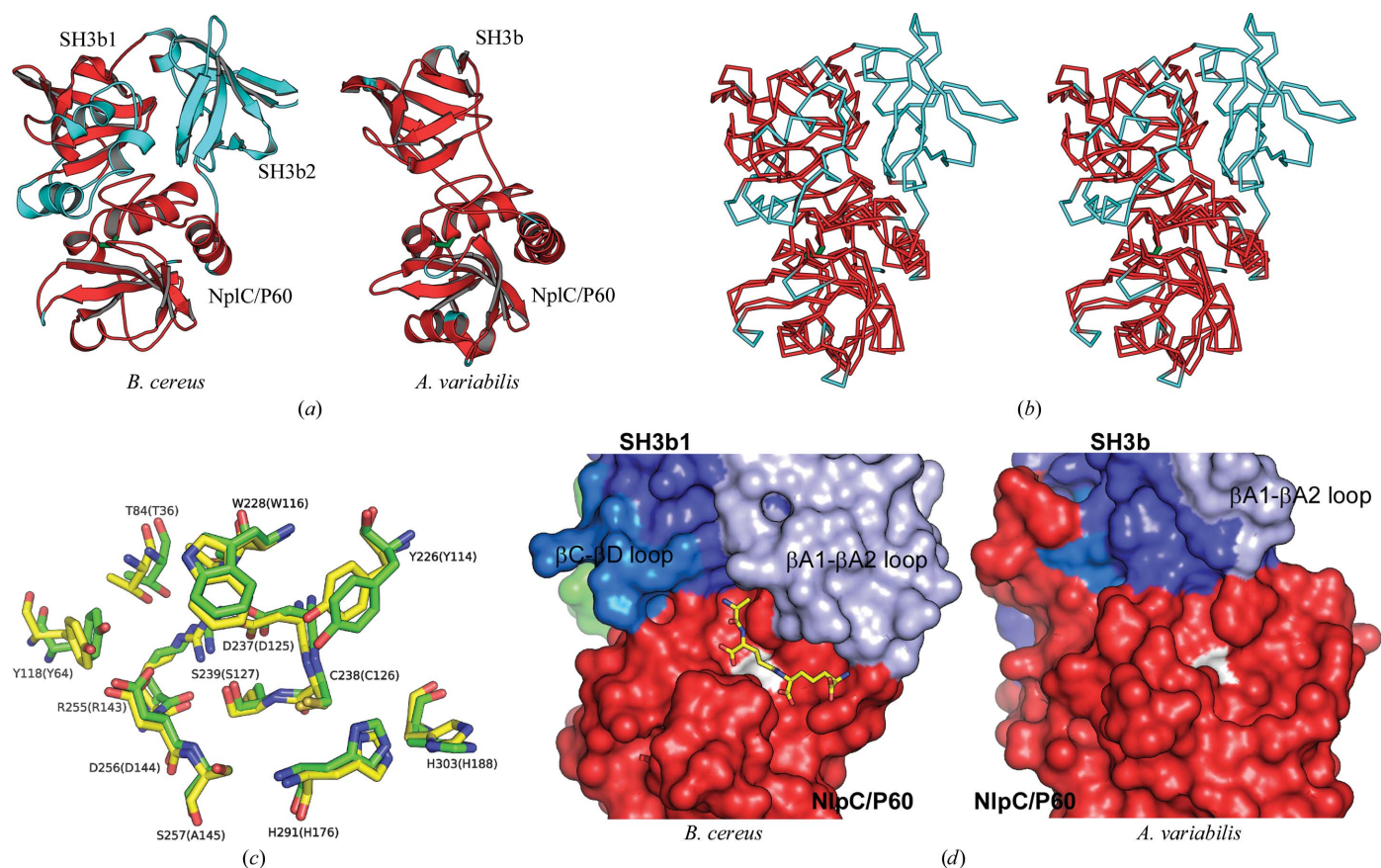


Figure 7

Structural comparisons between BcYkfC and the cyanobacterial NlpC/P60 endopeptidase AvPCP. (a) BcYkfC and AvPCP (PDB code 2hbw; Xu *et al.*, 2009) are shown with the same orientation of their common SH3b1 and NlpC/P60 domains. The structurally equivalent residues in each are shown in red. (b) Stereoview of the C α traces of the two proteins shown in (a) [the coloring is the same as in (a)]. (c) The S binding sites of the two proteins are nearly identical. The corresponding residues of AvPCP are labeled in parentheses. (d) Comparison of the active-site cavities and their environments. The catalytic cysteine is shown in white. A stick model of a docked murein tripeptide is shown in the active site of BcYkfC.

is also responsible for the interface with the catalytic domain. BcYkfc is a member of a cluster of highly conserved homologs that are present in closely related species such as *B. anthracis*, *B. thuringiensis* and *B. weihenstephanensis* (sequence identity of $\geq 95\%$). All members of this group contain signal peptides. Signal peptides are also present in some members of the bacteroidetes group, but not in proteobacterial and cyanobacterial homologs.

It is thus likely that the Ykfc subfamily of highly specialized enzymes has evolved from a common ancestor, which was likely to have been a P60-like general-purpose enzyme with two or more SH3b domains connected to a C-terminal NlpC/P60 domain. SH3b domains may have lost the function of targeting domains and over time the nonessential SH3b domain would have been lost, as seen in the cyanobacterial enzymes.

4. Conclusions

The structure of BcYkfc in complex with L-Ala- γ -D-Glu is the first structural representative of an NlpC/P60 enzyme with a bound ligand. This structure allowed us to identify the determinants of substrate specificity, which further led to the classification of a subfamily of highly specialized NlpC/P60 enzymes. The studies here also provide structural and functional insights into the NlpC/P60 family of enzymes. Additional information about BcYkfc is available from TOPSAN (Krishna *et al.*, 2010) at <http://www.topsan.org/explore?PDBid=3h41>.

This work was supported by the NIH, National Institute of General Medical Sciences, Protein Structure Initiative grant U54 GM074898. Portions of this research were carried out at the Stanford Synchrotron Radiation Lightsource (SSRL). The SSRL is a national user facility operated by Stanford University on behalf of the US Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research and by the National Institutes of Health (National Center for Research Resources, Biomedical Technology Program and the National Institute of General Medical Sciences). Genomic DNA from *B. cereus* NRS248 (ATCC No. 10987D) was obtained from the American Type Culture Collection (ATCC). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health.

References

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.
 Anantharaman, V. & Aravind, L. (2003). *Genome Biol.* **4**, R11.
 Aramini, J. M., Rossi, P., Huang, Y. J., Zhao, L., Jiang, M., Maglaqui, M., Xiao, R., Locke, J., Nair, R., Rost, B., Acton, T. B., Inouye, M. & Montelione, G. T. (2008). *Biochemistry*, **47**, 9715–9717.
 Asano, S. I., Nukumizu, Y., Bando, H., Iizuka, T. & Yamamoto, T. (1997). *Appl. Environ. Microbiol.* **63**, 1054–1057.
 Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L., Studholme, D. J., Yeats, C. & Eddy, S. R. (2004). *Nucleic Acids Res.* **32**, D138–D141.
 Bateman, A. & Rawlings, N. D. (2003). *Trends Biochem. Sci.* **28**, 234–237.

Bricogne, G., Vornrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. (2003). *Acta Cryst. D59*, 2023–2030.
 Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst. D66*, 12–21.
 Cohen, A. E., Ellis, P. J., Miller, M. D., Deacon, A. M. & Phizackerley, R. P. (2002). *J. Appl. Cryst.* **35**, 720–726.
 Cohen, S. X., Morris, R. J., Fernandez, F. J., Ben Jelloul, M., Kakaris, M., Parthasarathy, V., Lamzin, V. S., Kleywegt, G. J. & Perrakis, A. (2004). *Acta Cryst. D60*, 2222–2229.
 Collaborative Computational Project, Number 4 (1994). *Acta Cryst. D50*, 760–763.
 Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. (2004). *Genome Res.* **14**, 1188–1190.
 Cruickshank, D. W. J. (1999). *Acta Cryst. D55*, 583–601.
 Doyle, R. J., Chaloupka, J. & Vinter, V. (1988). *Microbiol. Rev.* **52**, 554–567.
 Eddy, S. R. (1998). *Bioinformatics*, **14**, 755–763.
 Emsley, P. & Cowtan, K. (2004). *Acta Cryst. D60*, 2126–2132.
 Holm, L. & Sander, C. (1995). *Trends Biochem. Sci.* **20**, 478–480.
 Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.
 Kabsch, W. (2010). *Acta Cryst. D66*, 125–132.
 Kall, L., Krogh, A. & Sonnhammer, E. L. (2007). *Nucleic Acids Res.* **35**, W429–W432.
 Klock, H. E., Koesema, E. J., Knuth, M. W. & Lesley, S. A. (2008). *Proteins*, **71**, 982–994.
 Korndorfer, I. P., Danzer, J., Schmelcher, M., Zimmer, M., Skerra, A. & Loessner, M. J. (2006). *J. Mol. Biol.* **364**, 678–689.
 Krishna, S. S., Weekes, D., Bakolitsa, C., Elsliger, M.-A., Wilson, I. A., Godzik, A. & Wooley, J. (2010). *Acta Cryst. F66*, 1143–1147.
 Kuhn, M. & Goebel, W. (1989). *Infect. Immun.* **57**, 55–61.
 Lesley, S. A. *et al.* (2002). *Proc. Natl Acad. Sci. USA*, **99**, 11664–11669.
 Lu, J. Z., Fujiwara, T., Komatsuzawa, H., Sugai, M. & Sakon, J. (2006). *J. Biol. Chem.* **281**, 549–558.
 Marino, M., Banerjee, M., Jonquieres, R., Cossart, P. & Ghosh, P. (2002). *EMBO J.* **21**, 5623–5634.
 Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst. D53*, 240–255.
 Nasertorabi, F., Tars, K., Becherer, K., Kodandapani, R., Liljas, L., Vuori, K. & Ely, K. R. (2006). *J. Mol. Recognit.* **19**, 30–38.
 Pai, C.-H., Chiang, B.-Y., Ko, T.-P., Chou, C.-C., Chong, C.-M., Yen, F.-J., Chen, S., Coward, J. K., Wang, A. H. & Lin, C.-H. (2006). *EMBO J.* **25**, 5970–5982.
 Park, J. T. & Uehara, T. (2008). *Microbiol. Mol. Biol. Rev.* **72**, 211–227.
 Plewniak, F. *et al.* (2003). *Nucleic Acids Res.* **31**, 3829–3832.
 Pohl, E., Holmes, R. K. & Hol, W. G. (1999). *J. Mol. Biol.* **292**, 653–667.
 Rigden, D. J., Jedrzejas, M. J. & Galperin, M. Y. (2003). *Trends Biochem. Sci.* **28**, 230–234.
 Santarsiero, B. D., Yegian, D. T., Lee, C. C., Spraggon, G., Gu, J., Scheibe, D., Uber, D. C., Cornell, E. W., Nordmeyer, R. A., Kolbe, W. F., Jin, J., Jones, A. L., Jaklevic, J. M., Schultz, P. G. & Stevens, R. C. (2002). *J. Appl. Cryst.* **35**, 278–281.
 Schmidt, D. M., Hubbard, B. K. & Gerlt, J. A. (2001). *Biochemistry*, **40**, 15707–15715.
 Sheldrick, G. M. (2008). *Acta Cryst. A64*, 112–122.
 Smith, T. J., Blackman, S. A. & Foster, S. J. (2000). *Microbiology*, **146**, 249–262.
 Srisailam, S., Lukin, J. A., Lemak, A., Yee, A. & Arrowsmith, C. H. (2006). *J. Biomol. NMR*, **36**, Suppl. 1, 27.
 Storer, A. C. & Ménard, R. (1994). *Methods Enzymol.* **244**, 486–500.
 Teng, F., Kawalec, M., Weinstock, G. M., Hryniewicz, W. & Murray, B. E. (2003). *Infect. Immun.* **71**, 5033–5041.
 Uehara, T. & Park, J. T. (2003). *J. Bacteriol.* **185**, 679–682.
 Uehara, T. & Park, J. T. (2004). *J. Bacteriol.* **186**, 7273–7279.
 Uehara, T. & Park, J. T. (2007). *J. Bacteriol.* **189**, 5634–5641.
 Uehara, T., Suefujii, K., Valbuena, N., Meehan, B., Donegan, M. & Park, J. T. (2005). *J. Bacteriol.* **187**, 3643–3649.
 Vacheron, M. J., Guinand, M., Francon, A. & Michel, G. (1979). *Eur. J. Biochem.* **100**, 189–196.
 Wylie, G. P., Rangachari, V., Bienkiewicz, E. A., Marin, V., Bhattacharya, N., Love, J. F., Murphy, J. R. & Logan, T. M. (2005). *Biochemistry*, **44**, 40–51.
 Xu, Q. *et al.* (2009). *Structure*, **17**, 303–313.