



Published in final edited form as:

Proteins. 2006 January 1; 62(1): 144–151. doi:10.1002/prot.20702.

## ***B. subtilis* ykuD Protein at 2.0 Å Resolution: Insights into the Structure and Function of a Novel, Ubiquitous Family of Bacterial Enzymes**

Jakub Bielnicki<sup>1</sup>, Yancho Devedjiev<sup>1</sup>, Urszula Derewenda<sup>1</sup>, Zbigniew Dauter<sup>2</sup>, Andrzej Joachimiak<sup>3</sup>, and Zygmunt S. Derewenda<sup>1,\*</sup>

<sup>1</sup>Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, Virginia

<sup>2</sup>Synchrotron Radiation Research Section, National Cancer Institute, Argonne National Laboratory, Argonne, Illinois

<sup>3</sup>Biosciences Division and Structural Biology Center, Argonne National Laboratory, Argonne, Illinois

### **Abstract**

The crystal structure of the product of the *Bacillus subtilis* *ykuD* gene was solved by the multiwavelength anomalous dispersion (MAD) method and refined using data to 2.0 Å resolution. The *ykuD* protein is a representative of a distinctly prokaryotic and ubiquitous family found among both pathogenic and nonpathogenic Gram-positive and Gram-negative bacteria. The deduced amino acid sequence reveals the presence of an N-terminal LysM domain, which occurs among enzymes involved in cell wall metabolism, and a novel, putative catalytic domain with a highly conserved His/Cys-containing motif of hitherto unknown structure. As the wild-type protein did not crystallize, a double mutant was designed (Lys117Ala/Gln118Ala) to reduce excess surface conformational entropy. As expected, the structure of the LysM domain is similar to the NMR structure reported for an analogous domain from *Escherichia coli* murein transglycosylase MltD. The molecular model also shows that the 112-residue-long C-terminal domain has a novel tertiary fold consisting of a  $\beta$ -sandwich with two mixed sheets, one containing five strands and the other, six strands. The two  $\beta$ -sheets form a cradle capped by an  $\alpha$ -helix. This domain contains a putative catalytic site with a tetrad of invariant His123, Gly124, Cys139, and Arg141. The stereochemistry of this active site shows similarities to peptidotransferases and sortases, and suggests that the enzymes of the *ykuD* family may play an important role in cell wall biology.

### **Keywords**

hydrolase; crystal structure; catalytic triad; structural genomics; protein crystallization

### **INTRODUCTION**

In spite of rapid advances in microbial genomics and an expanding set of complete prokaryotic genomes, a significant portion of bacterial genes code for nonannotated proteins of unknown structure and function. As the search for novel antibiotic targets intensifies in the face of an

advancing threat of revival in infectious diseases, structural characterization of such orphan proteins or entire families of bacterial proteins is of particular importance.

One such structurally uncharacterized, ubiquitous family of exclusively prokaryotic proteins comprises putative enzymes containing the so-called ErfK/YbiS/YhnG domain (Pfam03734.8) as identified by the SMART database.<sup>1</sup> These proteins are found in a vast variety of bacteria, ranging from pathogenic *Bacillus anthracis*, *Vibrio cholera*, *Clostridium tetani*, and others, to nonpathogenic *Escherichia coli*, *Thermotoga maritima*, *Bacillus subtilis*, and so forth. All ErfK/YbiS/YhnG domains contain a highly conserved stretch of amino acids ( $\Phi\text{G}\Phi\text{HGTX}_{10}(\text{S/T})\text{XGC}\Phi\text{R}(\text{M/L})$ , where  $\Phi$  is a hydrophobic residue and  $X$  denotes any amino acid). The presence of absolutely conserved His and Cys residues is suggestive of a catalytic function, probably as hydrolases or transferases (Fig. 1). Many of these proteins also contain the so-called LysM domain, which appears to have a peptidoglycan-binding function and is found in enzymes involved in cell wall degradation, and other proteins that could be associated with bacterial cell walls.<sup>2</sup> Interestingly, homologous domains have also been discovered in a family of plant transmembrane receptor kinases, which are involved in the recognition of bacteria secreting lipochitin oligosaccharide signal molecules.<sup>3</sup> Only one structural study of a LysM domain has been reported to date.<sup>2</sup> The association of the ErfK/YbiS/YhnG domain with the LysM module suggests that the former may also play a role in cell wall degradation or, more specifically, in peptidoglycan metabolism. This is interesting because such unique prokaryotic proteins constitute a particularly suitable target for antibiotics.

The *B. subtilis* *ykuD* protein was selected as one of the targets for high-throughput structure determination by the Midwest Center for Structural Genomics (MCSG). It expressed well in *E. coli* but failed to crystallize despite extensive screening. We have therefore applied the recently proposed approach of surface conformational entropy reduction to generate on the protein's surface patches that would be able to mediate crystal contacts.<sup>4</sup> Accordingly, Lys117 and Gln118 were mutated to alanines, and the purified double mutant yielded high-quality crystals.

In this article we describe the details of the *ykuD* structure. Aside from the expected LysM module, the structure reveals a novel tertiary fold for the C-terminal ErfK/YbiS/YhnG domain. The stereochemical features of the amino acids in the fingerprint, conserved stretch (i.e., of His123, Gly124, Cys139, and Arg131), confirm that the ErfK domain is very likely to have catalytic properties. Based on the crystal structure, we propose a possible catalytic mechanism for the enzyme and speculate on the nature of the substrate. Our work sets the stage for functional and biochemical characterization of a novel and probably important family of ubiquitous bacterial enzymes.

## MATERIALS AND METHODS

### Expression, Purification, and Crystallization Screens of *ykuD*

The *ykuD* gene was cloned by polymerase chain reaction (PCR) into the pMCSG7 vector.<sup>5</sup> The vector encodes a leader sequence consisting of a hexahistidine affinity tag followed by an eight-amino acid spacer and the tobacco etch virus (TEV) protease recognition site. The protein was expressed in the *E. coli* strain BL21 (DE3) and purified by Ni(2<sup>+</sup>)-nitrilotriacetic acid (Ni-NTA) metal-affinity chromatography. The affinity tag was removed during incubation with recombinant tobacco etch virus (rTEV) protease for 24 h. Protein was dialyzed against 20 mM 2-morpholinoethanesulfonic acid (MES) (pH 6.5), 300 mM NaCl, and 1.4 mM  $\beta$ -mercaptoethanol; and concentrated to 15 mg/mL for crystallization. Extensive screens yielded no results. Consequently, suitable sites for surface mutagenesis to enhance crystallizability were identified based on secondary structure prediction (<http://www.bioinfo.pl>). The dipeptide Lys-Gln (residues 117 and 118) was identified as one such potentially solvent exposed site and

both residues were mutated to Ala with use of the QuikChange™ mutagenesis kit (Stratagene). Crystals of this mutant were obtained at 21°C from 0.1 M acetate (pH 4.5), 0.2 M Li<sub>2</sub>SO<sub>4</sub>, 30% polyethylene glycol (PEG) 8000, and 10 mM CdCl<sub>2</sub>. The selenomethionine (Se-Met)-labeled protein was expressed in *E. coli* strain B843 and purified in a manner identical to that described for the native protein. Crystals of the labeled protein were obtained from 0.1 M acetate (pH 5.1), 0.2 M Li<sub>2</sub>SO<sub>4</sub>, 30% PEG 3350, and 10 mM CdCl<sub>2</sub>.

### Data Collection and Structure Solution

Data collection was performed at the National Synchrotron Light Source (NSLS), beamline X9B, at the Brookhaven National Laboratory. All data were processed and scaled using HKL2000 software package.<sup>6</sup> The ykuD structure was solved at 2.05 Å resolution by multiwavelength anomalous dispersion (MAD) using data collected at the absorption peak, inflection point, and high-energy remote wavelengths. Data statistics are shown in Table I. The crystal belongs to space group P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>; the unit cell parameters are as follows:  $a = 56.2$  Å,  $b = 63.9$  Å,  $c = 93.6$  Å. The Matthews coefficient is 2.4 Å<sup>3</sup>/Da assuming two molecules per asymmetric unit. SHELXD<sup>7</sup> was used to solve the anomalous substructure. Four selenium sites were used for phase calculation in SHARP.<sup>8</sup> The phases were used as an input to RESOLVE,<sup>9</sup> which automatically built 297 out of 328 residues present in the asymmetric unit. The remaining portions of the model were built manually with O<sup>10</sup> and COOT.<sup>11</sup> The refinement of the ykuD model was carried out with REFMAC5<sup>12</sup> using the high-energy remote data to 2.0 Å resolution and individual isotropic displacement parameters (B factors).

### Accession Code of the Structure

The atomic coordinates and structure factors of the ykuD protein have been deposited in the Protein Data Bank (PDB) under the accession code 1Y7M.

## RESULTS AND DISCUSSION

### Quality of the Model

The refined crystallographic model consists of two ykuD molecules in the asymmetric unit, each made up of 164 residues in a single polypeptide chain, 2 Cd<sup>2+</sup> ions, 4 sulfate ions, and 150 water molecules. The stereochemical quality of the model was analyzed with PROCHECK<sup>13</sup>: 88.6% of nonglycine residues are in the favored regions, and 11.4% are in the allowed regions of the Ramachandran plot. The side-chains of 18 amino acids (11 from molecule A, and 7 from molecule B) are not visible in the  $\sigma_A$  weighted  $2mF_{obs} - DF_{calc}$  electron density map contoured at 1  $\sigma$ , and were consequently excluded from refinement. The average B factors for the two LysM domains are consistently about 15% higher than those for the catalytic domains, consistent with a degree of dynamic disorder in these modules. The final values of  $R$  and  $R_{free}$  factors were 21.1 and 26.9 %, respectively. The refinement statistics are shown in Table I.

### Overall Structure

The ykuD protein is made up of two distinct domains: the 52-amino acid-long N-terminal LysM domain and the 112-amino acid C-terminal putative catalytic domain. The two ykuD molecules form a noncrystallographic, head-to-tail dimer with the local two fold axis nearly parallel to the crystallographic  $c$  axis [Fig. 2(A)]. Gel filtration indicated a monomer as a dominant species in solution; consequently, the noncrystallographic dimer has no functional implications.

The LysM domain was identified in ykuD based on sequence analysis, and indeed its structure is similar to the solution (NMR) structure of the isolated LysM domain from the *E. coli* membrane-bound lytic murein transglycosylase D.<sup>2</sup> Ours is the first reported crystal structure

of any protein containing LysM. The domain is known to have a general peptidoglycan binding function and is characterized by a  $\beta\alpha\alpha\beta$  tertiary structure with two  $\alpha$ -helices located on the same side of a two stranded, antiparallel  $\beta$ -sheet. Figure 2(B) represents the LysM domain of ykuD superimposed on its structural homolog. Although both structures are very similar, there are some discrepancies, including part of helix B and the downstream loop (residues 22 to 36). The average B factors for that part of the structure are higher than for the rest of the domain ( $\sim 25 \text{ \AA}^2$  vs  $\sim 22 \text{ \AA}^2$ ), consistent with higher intrinsic mobility of the loop. Moreover, the LysM domain of ykuD does not form any significant crystal contacts, which may be a reason for its reduced stability.

The 112-amino acid ErfK/YbiS/YhnG domain represents a novel fold. No structure similar to this domain was detected using DALI.<sup>14</sup> It is a  $\beta$ -sandwich, with two mixed  $\beta$ -sheets, one containing five, and the other, six  $\beta$ -strands [Fig. 2(C and D)]. Strand 3 is involved in both sheets, while strands 4 and 5 are almost contiguous, except for Ile 98, which has atypical secondary conformation ( $\phi = -100.6$ ,  $\psi = -50.9$ ) and creates a kink between them. A single  $\alpha$ -helix is found between strands 9 and 10. The most conserved residues of the fingerprint sequence [i.e., <sup>120</sup> $\Phi G \Phi H G T$ <sup>125</sup> and <sup>136</sup> $(S/T) X G C \Phi R (M/L)$ <sup>142</sup>] are located on the solvent-exposed concave face within strands 7 and 9 [Fig. 2(C)].

### The Putative Catalytic Site

Because the fingerprint of the ErfK/YbiS/YhnG domain family (Fig. 1) contains strictly invariant His and Cys residues, often found in hydrolytic enzymes and transferases, the SMART database<sup>1</sup> suggests that the domain has a catalytic function; consequently, we will henceforth refer to it as such. As pointed out above, the most conserved residues cluster within the seventh and the ninth strands of the  $\beta$ -sheet. We note that the disposition of three residues—His123 and Gly124, located on  $\beta$ -strand 7, and Cys 139, located on  $\beta$ -strand 9—has distinct similarities to so-called catalytic triads present in a wide range of enzymes, including serine proteases, cysteines proteases, esterases, lipases, and a number of more esoteric, nonhydrolytic enzymes.<sup>15</sup> The role of a nucleophile is typically reserved for a Ser or a Cys, which interacts with the most invariant residue in a triad, a histidine, which in turn is hydrogen-bonded to a neutral or negatively charged H-bond acceptor. The details vary from family to family.<sup>15</sup> The interaction of Ser with His is in most cases mediated in the ground state by a weak H-bond between the hydroxyl and imidazol's unprotonated N $\epsilon$ 2. During the acylation step, the hydrogen released by Ser is captured by His, which adopts the protonated, imidazolium form. In papainlike cysteine proteinases, the ground state enzyme contains an ion pair in which both Cys and His are found in the ionized forms as a thiolate and imidazolium, enhancing the nucleophilic potential of cysteines.<sup>16</sup> This pattern is not universal though, because viral 3C proteases operate by a general base mechanism, similar to serine proteases.<sup>17</sup> The conformation of the catalytic histidine is stabilized in all cases by additional interactions. In serine hydrolases, the N $\delta$ 1 group of the imidazole is suitably positioned to donate an H-bond to Asp or Glu,<sup>18,19</sup> and in two notable cases the accepting group is a backbone carbonyl<sup>20,21</sup>; in the papain family of cysteine proteases, the N $\epsilon$ 2 of the catalytic histidine donates an H-bond to Asn, although examples of enzymes with aspartates in that position are also known.<sup>22,23</sup>

In ykuD, the three amino acids of the putative triad are capable of forming similar interactions, albeit in the crystal structure, the side-chains of Cys139 and His123 in both molecules are perturbed by an intimate symmetric crystal contact that traps two Cd<sup>2+</sup> ions and two sulfates at the intermolecular interface (Fig. 3). Cys139 directly coordinates Cd<sup>2+</sup> (sulfur–metal distance of 2.33  $\text{\AA}$ ), which is also coordinated by N $\epsilon$ 2 of His119 from an adjacent protein molecule and two oxygen atoms of a sulfate ion, to complete a tetrahedral complex. The third oxygen of the sulfate accepts an H-bond from N $\epsilon$ 2 of the catalytic His123. Thus, the thiolate of Cys139 actually points away from the imidazole of His123, not as expected for a catalytic

triad. However, we believe that this is a crystallization artifact as a rotation around the  $\chi_1$  torsion angle of Cys139 is sufficient to generate a canonical interaction with N $\epsilon$ 2 of His123. The N $\delta$ 1 of His123 donates an H-bond to the main chain carbonyl of Gly124, in a manner similar to other triads. Overall, the mutual disposition of Cys139, His123 and Gly124 strongly suggest that this group constitutes a variant of the ubiquitous catalytic triad [Fig. 4(A)].

A triad is not sufficient to account for all the catalytic potential of serine- and cysteines-dependent hydrolases and transferases. The stabilization of the incipient oxyanion following the acylation step is accomplished within a so-called oxyanion hole, a positively charged surface crevice with two or three strong H-bond donors.<sup>24,25</sup> Our atomic model reveals that a similar structure is also found within ykuD. The aforementioned sulfate molecule coordinated by the Cd<sup>2+</sup> ion fortuitously serves as an approximate model for the tetrahedral intermediate. One of its oxygens forms H-bonds with backbone amides of Lys137, the absolutely conserved Gly138, and Cys139. The distances are 3.19 Å, 3.24 Å, and 3.03 Å, respectively. Two or three of these amides could act as the oxyanion hole, forming hydrogen bonds with the negatively charged oxygen of the tetrahedral intermediate of proteolytic reaction [Fig. 4(A)]. This observation further supports the notion that the putative active site's stereochemistry in ykuD is consistent with canonical general base mechanism for serine/cysteine hydrolases.

### Similarity to Sortases

Although the stereochemistry of the putative active site in ykuD is consistent with it having a hydrolytic or transferase function, akin to that of a protease or an esterase, the tertiary fold of the catalytic domain is not similar to any of the triad-containing families, such as chymotrypsin, subtilisin, papain or  $\alpha/\beta$  hydrolase proteins. However, we note some limited topological similarities to the recently characterized structures of sortases (i.e., transpeptidases located in the cell envelope of Gram-positive bacteria) that catalyze the anchoring of specific cell-surface proteins via their unique LPXTG motif to a pentaglycine cross-bridge of peptidoglycans.<sup>26–29</sup> In either sortase, the putative active site is located in a shallow depression on a concave side of a  $\beta$ -sheet, with the Cys residue found at the end of a  $\beta$ -strand. Interestingly, even though an involvement of a His residue has been postulated in sortases, the crystal structures suggest that it is a dyad of Cys and an invariant Arg that constitutes the active ensemble of residues.<sup>30</sup> This has been corroborated by biochemical studies of the *Staphylococcus aureus* sortase A, which showed that replacement of the relevant Arg197 with other amino acids severely impedes catalysis.<sup>31</sup> It has been postulated that Arg197 may facilitate the formation of a thiolate during the transpeptidation reaction catalyzed by the sortases.<sup>31</sup>

Close inspection of the ykuD active site also reveals the presence of the side-chain of an arginine, Arg141, in the active site, in an immediate proximity of Cys139. The mutual disposition of the two amino acids is very similar to that seen in sortases [Fig. 4(B)]. Arg141 is conserved in all sequences of the ErfK/YbiS/YhnG domains (Fig. 1). Thus, ykuD and its homologs appear to have within a short stretch of amino acids all the active site components found in cysteines-dependent proteinases and transpeptidases: a triad of Cys, His, and an H-bond acceptor stabilizing the imidazol, a well-formed oxyanion hole stabilizing the tetrahedral intermediate, as well as a positively charged residue enhancing the nucleophilicity of the Cys. It is tempting to speculate that the structural similarities between ykuD constitute a reflection of functional parallels.

### Substrate Specificity and Biological Function of ykuD

The biological function of ykuD, or indeed of any of its homologs containing the ErfK/YbiS/YhnG domain, remains unknown. However, the presence of the LysM domain in ykuD indicates that the enzyme may be active when bound to peptidoglycan, although we note that in a number of proteins containing the ErfK/YbiS/YhnG domains the LysM domain is absent.



It is known that *ykuD* is expressed in *B. subtilis* among spore proteins, and that its transcription is regulated by the  $\sigma^K$  factor, which is active primarily during bacterial sporulation.<sup>32</sup> However, the function of the protein cannot be exclusive to sporulation, because it is found in a variety of Gram-negative bacteria as well.

To our knowledge, there are no structures known for Cys-dependent, triad-containing esterases, thioesterases, or lipases. It appears that cysteines only function in proteases and transpeptidases. The stereochemistry of the putative active site in *ykuD*, and in particular the presence of a well-developed oxyanion hole, supports the notion that the enzyme utilizes a general base mechanism and undergoes acylation. This is consistent with either a peptidase (protease) activity or a transpeptidase activity. Although we were unable to confidently model an acceptor peptide (such as pentaglycine) into the active site, there appears room for such an acceptor, so we cannot rule out that the enzyme in fact is a transpeptidase similar to sortases, anchoring proteins in the spore's coat, and performing similar function in the cell wall of Gram-negative bacteria.

Further studies are necessary to answer all the pending questions regarding the function and activity of the ErfK/YbiS/YhnG domains. The determination of the *ykuD* crystal structure is an important step in this process and paves the way for the future studies of this family of proteins. Once a substrate is identified and a functional assay is available, it will be possible to dissect the structure–function relationships in the putative active site using site-directed mutagenesis.

### Crystal Contacts

The determination of the crystal structure of *ykuD* has been possible due to the application of the conformational surface entropy reduction concept, because the wild-type protein did not crystallize. It is therefore instructive to analyze the crystal contacts and assess the impact of surface mutagenesis on the lattice-forming interactions.

There are extensive contacts crystallographically related to dimers in the crystal structure, but there are specifically two crystal contacts that appear to stem directly from the surface mutations and crystallization conditions. The first contact is formed between two polypeptide chains related by noncrystallographic symmetry that belong to two adjacent noncrystallographic dimers (Fig. 3). Not surprisingly, it involves the two mutated residues—Ala117 and Ala118. These amino acids are located, as predicted, on a solvent-exposed loop. Although the two alanines are not involved in any direct intermolecular interactions, the wild-type side-chains of Lys and Gln within this loop would have almost certainly caused a steric obstacle, preventing two *ykuD* molecules from interacting. The third residue involved in this contact is His119, which coordinates a cadmium ion ( $N\epsilon_2$ – $Cd^{2+}$  distance of 2.27 Å) trapped between two protein molecules at the active site. The second crystal contact is mediated by sulfate ion sequestered between two arginines (Arg159) in the center of the noncrystallographic dimer. The distances between sulfate oxygen and  $N\eta_1$  of arginines from molecule A and molecule B are 2.78 Å and 2.65 Å, respectively. Moreover,  $N\eta_2$  of Arg159 from molecule A creates one more hydrogen bond (3.40 Å) with sulfate oxygen. This interaction is additionally stabilized by His57 from molecule A and by Tyr96 from molecule B.

Successful crystallization of *ykuD* depends on the mutations (wild-type protein does not crystallize), as well as the presence of both sulfate and  $Cd^{2+}$  ions in the precipitant. The structure shows that the sulfate ions help stabilize a noncrystallographic dimer, while the  $Cd^{2+}$  ions and mutations are critical in forming cross bridges between the dimers generating layers of molecules perpendicular to the *c* axis. We note that this crystal architecture has similarities to other successful cases of surface mutagenesis, where noncrystallographic dimers are found in the asymmetric unit.<sup>33</sup>

## CONCLUSION

This is the first crystallographic structural study of a representative of a ubiquitous family of exclusively prokaryotic proteins that contain both the LysM module, which mediates interactions with cell wall peptidoglycan, and the ErfK/YbiS/YhnG domain. The structure of the *B. subtilis* ykuD reveals a well-formed putative catalytic site, identifying potential functional roles for specific conserved residues such as His123, Cys139, and Arg141. The stereochemistry of the active site strongly suggests that the enzyme is a peptidase or transpeptidase, although we cannot be certain of the substrate. The presence of ykuD homologs in a variety of pathogenic bacteria, and the likelihood that these proteins play a role in the maintenance of cell wall, will hopefully lead to a thorough functional characterization of these interesting enzymes.

## Acknowledgments

Grant sponsor: National Institute of General Medical Sciences; Grant numbers: GM62615 (to Z. S. Derewenda) and P-50-GM62414 (to A. Joachimiak).

## REFERENCES

1. Letunic I, Copley RR, Schmidt S, Ciccarelli FD, Doerks T, Schultz J, Ponting CP, Bork P. SMART 4.0: towards genomic data integration. *Nucleic Acids Res* 2004;32:D142–D144. [PubMed: 14681379]
2. Bateman A, Bycroft M. The structure of a LysM domain from *E. coli* membrane-bound lytic murein transglycosylase D (MltD). *J Mol Biol* 2000;299:1113–1119. [PubMed: 10843862]
3. Spaik HP. Specific recognition of bacteria by plant LysM domain receptor kinases. *Trends Microbiol* 2004;12:201–204. [PubMed: 15120137]
4. Derewenda ZS. Rational protein crystallization by mutational surface engineering. *Structure* 2004;12:529–535. [PubMed: 15062076]
5. Stols L, Gu M, Dieckman L, Raffin R, Collart FR, Donnelly MI. A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. *Protein Expr Purif* 2002;25:8–15. [PubMed: 12071693]
6. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 1997;276:307–326.
7. Schneider TR, Sheldrick GM. Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr* 2002;58:1772–1779. [PubMed: 12351820]
8. de la Fortelle E, Bricogne G. Maximum-likelihood heavy atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol* 1997;276:472–494.
9. Terwilliger TC. Automated side-chain model building and sequence assignment by template matching. *Acta Crystallogr D Biol Crystallogr* 2003;59:45–49. [PubMed: 12499538]
10. Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* 1991;47:110–119. [PubMed: 2025413]
11. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 2004;60:2126–2132. [PubMed: 15572765]
12. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 1997;53:240–255. [PubMed: 15299926]
13. Laskowski RA, McArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 1993;26:282–291.
14. Holm L, Sander C. Dali/FSSP classification of three-dimensional protein folds. *Nucleic Acids Res* 1997;25:231–234. [PubMed: 9016542]
15. Dodson G, Wlodawer A. Catalytic triads and their relatives. *Trends Biochem Sci* 1998;23:347–352. [PubMed: 9787641]

16. Lewis SD, Johnson FA, Shafer JA. Effect of cysteine-25 on the ionization of histidine-159 in papain as determined by proton nuclear magnetic resonance spectroscopy: evidence for a his-159–Cys-25 ion pair and its possible role in catalysis. *Biochemistry* 1981;20:48–51. [PubMed: 7470479]
17. Huang C, Wei P, Fan K, Liu Y, Lai L. 3C-like proteinase from SARS coronavirus catalyzes substrate hydrolysis by a general base mechanism. *Biochemistry* 2004;43:4568–4574. [PubMed: 15078103]
18. Sprang S, Standing T, Fletterick RJ, Stroud RM, Finer-Moore J, Xuong NH, Hamlin R, Rutter WJ, Craik CS. The three-dimensional structure of Asn102 mutant of trypsin: role of Asp102 in serine protease catalysis. *Science* 1987;237:905–909. [PubMed: 3112942]
19. Schrag JD, Li Y, Wu S, Cygler M. Ser-His-Glu triad forms the catalytic site of the lipase from *Geotrichum candidum*. *Nature* 1991;351:761–764. [PubMed: 2062369]
20. Wei Y, Contreras JA, Sheffield P, Osterlund T, Derewenda U, Kneusel RE, Matern U, Holm C, Derewenda ZS. Crystal structure of brefeldin A esterase, a bacterial homolog of the mammalian hormone-sensitive lipase. *Nat Struct Biol* 1999;6:340–345. [PubMed: 10201402]
21. Serre L, Verbree EC, Dauter Z, Stuitje AR, Derewenda ZS. The *Escherichia coli* malonyl-CoA:ACP transacylase at 1.5 Å resolution. *J Biol Chem* 1995;270:12961–12964. [PubMed: 7768883]
22. Wenig K, Chatwell L, von Pawel-Rammingen U, Bjorck L, Huber R, Sondermann P. Structure of the streptococcal endopeptidase IdeS, a cysteine proteinase with strict specificity for IgG. *Proc Natl Acad Sci USA* 2004;101:17371–17376. [PubMed: 15574492]
23. Johnston SC, Larsen CN, Cook WJ, Wilkinson KD, Hill CP. Crystal structure of a deubiquitinating enzyme (human UCH-L3) at 1.8 Å resolution. *EMBO J* 1997;16:3787–3796. [PubMed: 9233788]
24. Matthews DA, Alden RA, Birktoft JJ, Freer ST, Kraut J. X-ray crystallographic study of boronic acid adducts with subtilisin BPN' (Novo): a model for the catalytic transition state. *J Biol Chem* 1975;250:7120–7126. [PubMed: 1165237]
25. Martinez C, Nicolas A, van Tilbeurgh H, Egloff MP, Cudrey C, Verger R, Cambillau C. Cutinase, a lipolytic enzyme with a preformed oxyanion hole. *Biochemistry* 1994;33:83–89. [PubMed: 8286366]
26. Novick RP. Sortase: the surface protein anchoring transpeptidase and the LPXTG motif. *Trends Microbiol* 2000;8:148–151. [PubMed: 10754567]
27. Zhang R, Wu R, Joachimiak G, Mazmanian SK, Missiakas DM, Gornicki P, Schneewind O, Joachimiak A. Structures of sortase B from *Staphylococcus aureus* and *Bacillus anthracis* reveal catalytic amino acid triad in the active site. *Structure* 2004;12:1147–1156. [PubMed: 15242591]
28. Zong Y, Bice TW, Ton-That H, Schneewind O, Narayana SV. Crystal structures of *Staphylococcus aureus* sortase A and its substrate complex. *J Biol Chem* 2004;279:31383–31389. [PubMed: 15117963]
29. Zong Y, Mazmanian SK, Schneewind O, Narayana SV. The structure of sortase B, a cysteine transpeptidase that tethers surface protein to the *Staphylococcus aureus* cell wall. *Structure* 2004;12:105–112. [PubMed: 14725770]
30. Dessen A. A new catalytic dyad regulates anchoring of molecules to the Gram-positive cell wall by sortases. *Structure* 2004;12:6–7. [PubMed: 14725759]
31. Marraffini LA, Ton-That H, Zong Y, Narayana SV, Schneewind O. Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*: a conserved arginine residue is required for efficient catalysis of sortase A. *J Biol Chem* 2004;279:37763–37770. [PubMed: 15247224]
32. Kodama T, Takamatsu H, Asai K, Ogasawara N, Sadaie Y, Watabe K. Synthesis and characterization of the spore proteins of *Bacillus subtilis* YdhD, YkuD, and YkvP, which carry a motif conserved among cell wall binding proteins. *J Biochem* 2000;128:655–663. [PubMed: 11011148]
33. Devedjiev Y, Surendranath Y, Derewenda U, Gabrys A, Cooper DR, Zhang RG, Lezondra L, Joachimiak A, Derewenda ZS. The structure and ligand binding properties of the B.subtilis YkoF gene product, a member of a novel family of thiamin/HMP-binding proteins. *J Mol Biol* 2004;343:395–406. [PubMed: 15451668]



**Consensus**
 $\Phi$ G $\Phi$ HG<sup>S/T</sup> (X)<sub>10</sub> <sup>S/T</sup>XGC $\Phi$ R<sup>M/L</sup>
*Bacillus subtilis* YkuD<sup>119</sup>HYGIHGTNNPASI<sup>143</sup>GKAVSKGCIRMH**Pathogenic**

Gram (+)

*Clostridium tetani*MYGIHGSNNPWSIGT<sup>143</sup>NVSLGCIRMF*Bacillus anthracis*

KYGIHGTNQPSSIGKYISQGCIRMK

*Clostridium perfringens*

NYLFHSIPFDS--GVPSSHGCIRLS

Gram (-)

*Neisseria meningitidis* AGLGIHGTNAPASVPGV<sup>143</sup>RS<sup>143</sup>HGCVRMK*Neisseria gonorrhoeae*

GLGIHGTNAPASVPGIRSHGCVRMK

*Yersinia pestis*

EYLIHGTNADFGIGLRVSSGCIRLR

*Salmonella typhimurium*

EYLIHGTSAPDSVGLRVSSGCIRMN

*Shigella flexneri*

EYLIHGTSAPDSVGLRVSSGCIRMN

*Vibrio parahaemolyticus*

EYLIHGTNKDFGIGMRVSAGCIRMD

*Vibrio cholerae*

EYLIHGTNKDFGVGMRVSAGCIRMD

*Brucella melitensis*GYRIHGSP<sup>143</sup>EWWSIGQSMSSGCIRLM**Non-Pathogenic**

Gram (+)

*Bacillus licheniformis*

HYGIHGTNNPSSIGKAVSRGCIRMH

*Geobacillus kaustophilus*

HYGIHGTNDPSSIGKYVSKGCIRMH

*Clostridium thermocellum*HYGIHGTNNPASI<sup>143</sup>GTAASNGCIRMY*Moorella thermoacetica*NYGIHGTNRPWSIGWAASAGCFR<sup>143</sup>MF

Gram (-)

*Thermotoga maritima*

TYAIHGTSKPWEIGKRISHGCIRMF

*Methylococcus capsulatus*

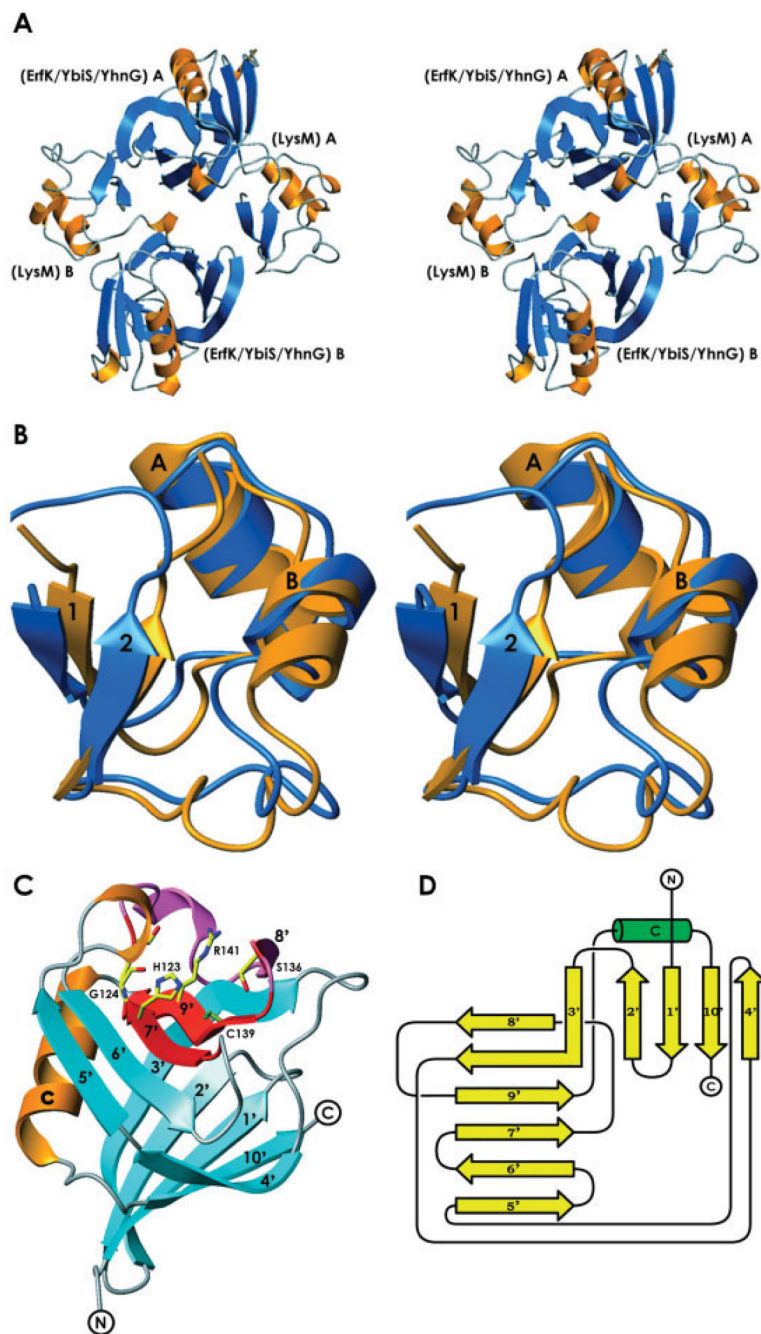
QYLIHGTNKPYSIGFRASNGCIRLY

*Xanthomonas axonopodis*HYGLHGTP<sup>143</sup>EPGHV<sup>143</sup>GKTESHGCVRLT*Sinorhizobium meliloti*

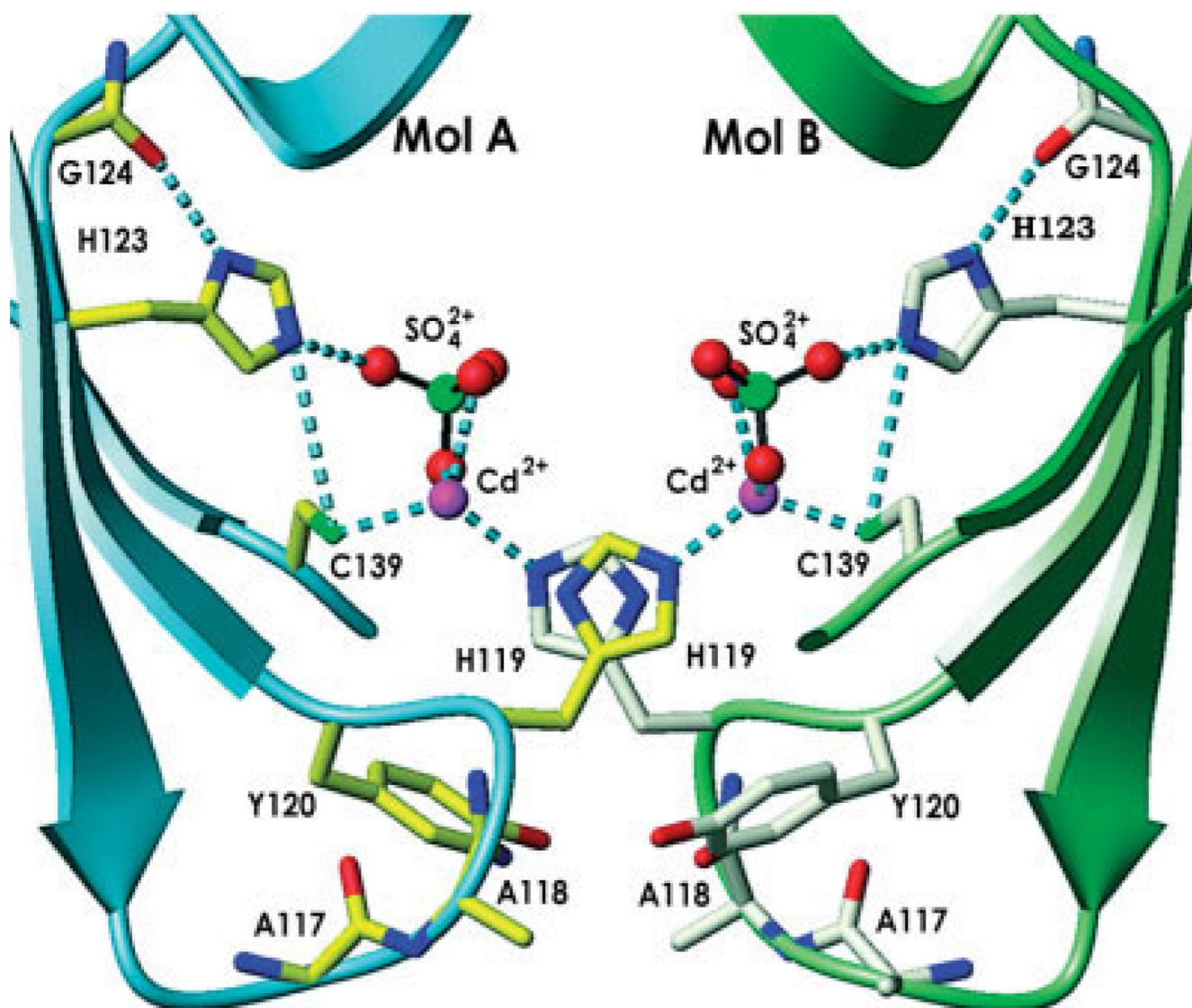
TYGIHGTPEPSKIGKTQSHGCVRLT

**Fig. 1.**

Sequence alignment of the fingerprint motif in pathogenic and nonpathogenic bacteria ( $\Phi$  is a hydrophobic residue and X denote any amino acid in consensus sequence).

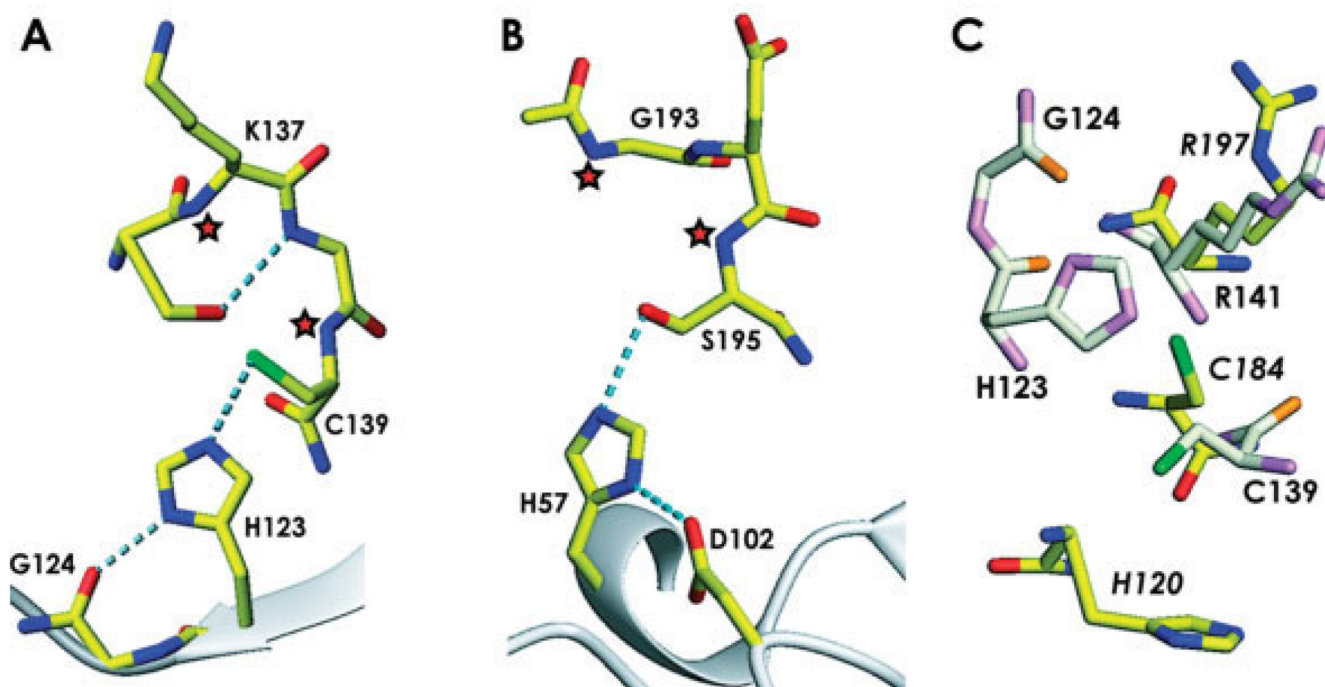


**Fig. 2.** The molecular architecture of ykuD. **(A)** A stereodiagram showing the head-to-tail noncrystallographic dimer of ykuD molecules; the two molecules are labeled A and B. **(B)** A comparison of the NMR structure of the LysM domain from murein transglycosylase D (PDB entry: 1E0G), orange, and the LysM domain of ykuD, blue; for details see text. **(C)** The tertiary structure of the ErfK/YbiS/YhnG domain of ykuD with the fingerprint conserved stretch shown in red; the side-chains of His123, Gly124, Ser136, Cys139, and Arg141 are shown in full. **(D)** A schematic representation of the tertiary fold of the ErfK/YbiS/YhnG domain.



**Fig. 3.**  
The ykuD putative active site involved in a crystal contact. Side-chains of select residues, as well as Gly124, are shown in full; sulfate ions are represented by spheres: red (oxygen) and green (sulfur).





**Fig. 4.**

A comparison of the catalytic site of *ykuD* with other enzymes. **(A)** The *ykuD* catalytic triad with the side-chain of Cys139 rotated to achieve the best stereochemistry; H-bonds are shown by dotted lines, putative oxyanion hole forming backbone amides are denoted by asterisks. **(B)** Trypsin (PDB entry: 1YYY), with oxyanion hole indicated by asterisks. **(C)** Comparison of the *ykuD* and sortase A (PDB entry: 1T2O) putative Cys-Arg dyads; *ykuD* residues are shown in light colors and the labels indicating the sortase A amino acids are in italics.

TABLE I

## Data Collection and Refinement Statistics

Data collection <sup>d</sup>							
Data type	Wavelength (Å)	Resolution (Å)	No. of unique reflections	Redundancy	Completeness (%)	$R_{\text{merge}}^b$ (%)	$I/\sigma(I)$
$\lambda_1$ -remote	0.97178	2.04	45,851	2.3	99.1 (99.0)	5.7 (34.2)	15.5 (2.8)
$\lambda_2$ -peak	0.97919	2.18	34,329	2.2	98.9 (93.7)	4.3 (26.6)	19.0 (2.8)
$\lambda_3$ -inflection	0.97946	2.07	43,799	2.1	98.4 (91.1)	4.9 (34.4)	16.0 (2.1)
Phasing statistics							
Parameter (unit)	Value						
F.O.M. SHARP	0.42						
F.O.M. Resolve	0.66						
Phasing power, isomorphous/anomalous <sup>c</sup>							
$\lambda_1$	-0.85						
$\lambda_2$	0.72/1.55						
$\lambda_3$	1.0/0.55						
$R_{\text{cullis}}$ , isomorphous/anomalous <sup>d</sup>							
$\lambda_1$	-0.89						
$\lambda_2$	0.75/0.71						
$\lambda_3$	0.68/0.94						
Refinement statistics							
Parameter (unit)	Value						
Resolution range (Å)	30.0-2.05						
Reflections (working)	20,691						
Reflections (test)	1114						
$R_{\text{work}}$ (%) / $R_{\text{free}}$ (%) / $e$	21.1/26.9						
No. of nonhydrogen atoms	2469/150						
Protein/waters	23.6						
Mean B value (Å <sup>2</sup> )							
Root-mean-square (r.m.s.) deviations							
Bond length (Å)/bond angle (°)	0.012/1.441						

<sup>a</sup>Values in parentheses correspond to the last shell.

<sup>b</sup> $R_{\text{merge}} = \sum |I_i - \langle I \rangle| / \sum I_i$ , where  $I_i$  is the intensity of the  $i$ th observation, and  $\langle I \rangle$  is the mean intensity of the reflections. The values are for unmerged Friedel pairs.

<sup>c</sup>Phasing power = r.m.s.  $(|F_h|/E)$ , where  $|F_h|$  is the heavy atom structure factor amplitude, and  $E$  is residual lack of closure error.

<sup>d</sup> $R_{\text{cullis}} = \sum |F_{h,\text{obs}}| - |F_{h,\text{calc}}| / \sum |F_h|$  for acentric reflections, where  $|F_{h,\text{obs}}|$  is the observed heavy atom structure factor amplitude, and  $|F_{h,\text{calc}}|$  is the calculated heavy atom structure factor amplitude.

<sup>e</sup> $R_{\text{work}} = \sum |F_{\text{obs}}| - |F_{\text{calc}}| / \sum |F_{\text{obs}}|$ , crystallographic  $R$  factor, and  $R_{\text{free}} = \sum |F_{\text{obs}}| - |F_{\text{calc}}| / \sum |F_{\text{obs}}|$ , where all reflections belong to a test set of randomly selected data.