

Structure of Csd3 from *Helicobacter pylori*, a cell shape-determining metallopeptidase

Doo Ri An,^a Hyoun Sook Kim,^{b,c} Jieun Kim,^a Ha Na Im,^a Hye Jin Yoon,^b Ji Young Yoon,^b Jun Young Jang,^b Dusan Heseck,^d Mijoon Lee,^d Shahriar Mobashery,^d Soon-Jong Kim,^e Byung Il Lee^f and Se Won Suh^{a,b,*}

Received 30 October 2014

Accepted 6 January 2015

Keywords: *Helicobacter pylori*; *csd3*; HP0506; peptidoglycan hydrolase; cell-shape determinant; D,D-endopeptidase; D,D-carboxypeptidase; M23B family metallopeptidase; LytM.

PDB references: Csd3, 4rny; 4rnz

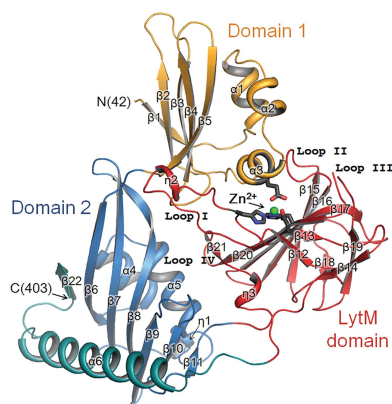
Supporting information: this article has supporting information at journals.iucr.org/d

^aDepartment of Biophysics and Chemical Biology, College of Natural Sciences, Seoul National University, Seoul 151-742, Republic of Korea, ^bDepartment of Chemistry, College of Natural Sciences, Seoul National University, Seoul 151-742, Republic of Korea, ^cResearch Institute of Pharmaceutical Sciences, College of Pharmacy, Seoul National University, Seoul 151 742, Republic of Korea, ^dDepartment of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, IN 46556, USA, ^eDepartment of Chemistry, Mokpo National University, Chonnam 534-729, Republic of Korea, and ^fBiomolecular Function Research Branch, Division of Convergence Technology, Research Institute, National Cancer Center, Gyeonggi 410-769, Republic of Korea. *Correspondence e-mail: sewonsuh@snu.ac.kr

Helicobacter pylori is associated with various gastrointestinal diseases such as gastritis, ulcers and gastric cancer. Its colonization of the human gastric mucosa requires high motility, which depends on its helical cell shape. Seven cell shape-determining genes (*csd1*, *csd2*, *csd3/hdpA*, *ccmA*, *csd4*, *csd5* and *csd6*) have been identified in *H. pylori*. Their proteins play key roles in determining the cell shape through modifications of the cell-wall peptidoglycan by the alteration of cross-linking or by the trimming of peptidoglycan muropeptides. Among them, Csd3 (also known as HdpA) is a bifunctional enzyme. Its D,D-endopeptidase activity cleaves the D-Ala⁴-mDAP³ peptide bond between cross-linked muramyl tetrapeptides and pentapeptides. It is also a D,D-carboxypeptidase that cleaves off the terminal D-Ala⁵ from the muramyl pentapeptide. Here, the crystal structure of this protein has been determined, revealing the organization of its three domains in a latent and inactive state. The N-terminal domain 1 and the core of domain 2 share the same fold despite a very low level of sequence identity, and their surface-charge distributions are different. The C-terminal LytM domain contains the catalytic site with a Zn²⁺ ion, like the similar domains of other M23 metallopeptidases. Domain 1 occludes the active site of the LytM domain. The core of domain 2 is held against the LytM domain by the C-terminal tail region that protrudes from the LytM domain.

1. Introduction

Helicobacter pylori is a spiral-shaped Gram-negative bacterium that colonizes the human stomach. It colonizes approximately half of the world's population and its infection of the gastric mucosa has been associated with various diseases of the upper gastrointestinal tract, such as chronic gastritis, peptic ulcer, mucosa-associated lymphoid tissue (MALT) lymphoma and gastric adenocarcinoma (Roesler *et al.*, 2014; Kusters *et al.*, 2006). *H. pylori* has been classified as a group I carcinogen by The International Agency for Research on Cancer and is regarded as a primary factor in the development of gastric cancer (International Agency for Research on Cancer, 1994). In recent years, *H. pylori* infection has also been implicated in some extra-digestive diseases (Roubaud Baudron *et al.*, 2013). Relatively effective treatment regimens are available for *H. pylori* infection that typically consist of a proton-pump inhibitor such as omeprazole and the antibiotics clarithromycin and amoxicillin (or metronidazole). However, increasing antibiotic resistance requires new therapies and the discovery of new antibiotics (Malfertheiner *et al.*, 2012).



OPEN ACCESS

High motility of *H. pylori* is important for its colonization of the human stomach and its survival in the gastric mucosa (Ottemann & Lowenthal, 2002; Schreiber *et al.*, 2004; Lertseththakarn *et al.*, 2011). The spiral or helical cell shape of *H. pylori* is believed to facilitate efficient colonization of the viscous epithelial mucus layer *via* a corkscrewing mechanism (Berg & Turner, 1979; Hazell *et al.*, 1986; Worku *et al.*, 1999). *H. pylori* mutants with altered cell shapes exhibit attenuated colonization (Bonis *et al.*, 2010; Sycuro *et al.*, 2010, 2012; Fridrich *et al.*, 2012; Wyckoff *et al.*, 2012). The peptidoglycan layer of the bacterial cell wall plays a role not only in withstanding the turgor pressure but also in maintaining the cell shape (Scheffers & Pinho, 2005; Vollmer & Bertsche, 2008). An essential component of bacterial peptidoglycan is a linear polysaccharide chain consisting of alternating β -1,4-linked *N*-acetylglucosamine (NAG) and *N*-acetylmuramic acid (NAM) disaccharides, with a pentapeptide linked to NAM (Vollmer, Blanot *et al.*, 2008). In helical-shaped *H. pylori*, the pentapeptide sequence is L-Ala¹- γ -D-Glu²-*m*DAP³-D-Ala⁴-D-Ala⁵, where *m*DAP refers to *meso*-2,6-diaminopimelate and the neighbouring peptides are cross-linked exclusively by the 4 \rightarrow 3 linkage between the main chain of D-Ala⁴ from one strand and the side chain of *m*DAP³ from another strand (Costa *et al.*, 1999) to form a mesh-like structure (Meroueh *et al.*, 2006). In many bacteria, the peptidoglycan layer is remodelled by a number of cell-wall hydrolases as well as synthetases for peptidoglycan maturation, regulation of cell-wall growth, cell division, peptidoglycan turnover and recycling, cell lysis and the release of peptidoglycan fragments for host–pathogen interactions (Vollmer, Joris *et al.*, 2008).

At least seven *H. pylori* genes have been identified to be required for determining the helical cell shape: *csd1*, *csd2*, *csd3/hdpA*, *ccmA*, *csd4*, *csd5* and *csd6* (Sycuro *et al.*, 2010, 2012, 2013; Bonis *et al.*, 2010). They play important roles in determining the helical cell shape of *H. pylori* by the relaxation of peptidoglycan cross-linking or by the trimming of pentapeptides to shorter peptides in peptidoglycan. Among them, the Csd3/HdpA protein as well as Csd1 and Csd2 belong to the MEROPS M23B metallopeptidase family (Sycuro *et al.*, 2010; Bonis *et al.*, 2010). Deletion of the *csd1*, *csd2* and *csd3* genes reduced the D,D-endopeptidase (D,D-EPase) activity, which cleaves the D-Ala⁴-*m*DAP³ peptide bond in cross-linked dimers of muramyl tetrapeptides and pentapeptides (muramyl-L-Ala¹- γ -D-Glu²-*m*DAP³-D-Ala⁴ and muramyl-L-Ala¹- γ -D-Glu²-*m*DAP³-D-Ala⁴-D-Ala⁵, respectively; Bonis *et al.*, 2010; Sycuro *et al.*, 2010).

Interestingly, the muropeptide composition of the deletion mutant of the *csd3* gene (Δ *csd3*) indicated that Csd3 has an additional D,D-carboxypeptidase (D,D-CPase) activity that cleaves the D-Ala⁴-D-Ala⁵ bond of the muramyl pentapeptide to produce the muramyl tetrapeptide and D-Ala (Bonis *et al.*, 2010; Sycuro *et al.*, 2010). In accordance with this observation, Δ *csd3* showed irregular C-shaped or stocky branched cells, which are distinct from the curved rod morphology possessed by Δ *csd1* and Δ *csd2* cells (Bonis *et al.*, 2010; Sycuro *et al.*, 2010). Inactivation of Csd3 by an H259A mutation, which is predicted to affect metal coordination in the active site, also

resulted in the same degree of morphological abnormality as in Δ *csd3* (Bonis *et al.*, 2010). *H. pylori* contains high levels of non-cross-linked pentapeptide in the peptidoglycan sacculus (Costa *et al.*, 1999). The bifunctional peptidase activities of Csd3, together with this observation, makes Csd3 an important regulator of *H. pylori* morphology.

Despite the important roles played by the helical cell shape-determining proteins of *H. pylori* in facilitating stomach colonization, structural reports on them have been very limited. We have recently determined the structure of *H. pylori* Csd4, a Zn²⁺-dependent D,L-CPase and a unique member of the M14 metallopeptidase family (Kim *et al.*, 2014). It cleaves the bond between γ -D-Glu² and *m*DAP³ of the non-cross-linked muramyl tripeptide (muramyl-L-Ala¹- γ -D-Glu²-*m*DAP³) of the peptidoglycan to produce the muramyl dipeptide (muramyl-L-Ala¹- γ -D-Glu²) and *m*DAP. Although both Csd3 and Csd4 play essential roles in determining the helical cell shape, their amino-acid sequences are unrelated. Genetic interaction studies between the *csd3* and *csd4* genes revealed that Csd4 D,L-CPase activity does not depend on Csd3 CPase/EPase activity and *vice versa* (Sycuro *et al.*, 2012).

To provide a structural framework for a better understanding of the molecular function of *H. pylori* Csd3, we report here the crystal structure of N-terminally truncated Csd3 encompassing residues 42–403 (Csd3 $_{\Delta 41}$). It consists of three domains: domain 1 (residues Glu42–Ile124), domain 2 (residues Ile125–Gly228 and Ala360–Phe403) and the C-terminal LytM domain (residues Phe229–Thr359). Csd3 domain 1 and the core of domain 2 (residues Ile125–Gly228) share a common fold despite a very low level of sequence identity. The LytM domain of Csd3 is structurally similar to the corresponding domains of other MEROPS M23 family metallopeptidases. Substrate binding to the active site of the LytM domain is blocked by domain 1 in our structure, suggesting that domain 1 is the inhibitory domain and that our Csd3 structure is in the latent state. The core of domain 2 is held stably against the LytM domain by the C-terminal extended tail region that protrudes from the LytM domain. This work could serve as a foundation for the discovery of novel inhibitors that could prove to be helpful in fighting infections by the major human pathogen *H. pylori*.

2. Materials and methods

2.1. Protein expression and purification

The gene encoding the N-terminally truncated form (residues 42–403; Csd3 $_{\Delta 41}$) of *H. pylori* Csd3 (HP0506 from strain 26695) was PCR-amplified and cloned into the expression vector pET-21a(+) (Novagen) using the NdeI and XhoI restriction-enzyme sites. The recombinant protein, which was fused to a hexahistidine-containing tag (LEHHHHHH) at its C-terminus, was overexpressed in *Escherichia coli* Rosetta 2(DE3)pLysS cells. The cells were grown at 37°C in Terrific Broth culture medium containing 50 μ g ml⁻¹ ampicillin. Protein expression was induced by 0.5 mM isopropyl β -D-1-thiogalactopyranoside and the cells were incubated for an additional 15 h at 30°C. The cells were harvested by centri-

fugation at 5600g for 15 min at 4°C and subsequently lysed by sonication in ice-cold buffer A [20 mM Tris–HCl pH 7.9, 500 mM sodium chloride, 50 mM imidazole, 10%(v/v) glycerol] which was supplemented with 1 mM phenylmethylsulfonyl fluoride, 60 mM ammonium chloride and 15 mM magnesium acetate. The lysate was centrifuged at 36 000g for 1 h at 4°C to discard the cell debris. The supernatant was applied onto a HiTrap Chelating HP affinity-chromatography column (GE Healthcare) which was previously equilibrated with buffer A. The column was eluted with a linear gradient of imidazole from 50 mM in buffer A to 500 mM in buffer B, which consists of 20 mM Tris–HCl pH 7.9, 500 mM NaCl, 500 mM imidazole, 10%(v/v) glycerol. The Csd3 protein eluted at 125–150 mM imidazole concentration was further purified by gel filtration on a HiLoad 16/60 Superdex 200 prep-grade column (GE Healthcare) at two different salt conditions, either with buffer C (20 mM Tris–HCl pH 7.9, 400 mM sodium chloride) or buffer D (20 mM Tris–HCl pH 7.9, 200 mM sodium chloride, 0.1 mM zinc chloride). Two different batches of Csd3 yielded different crystal forms, as described below. The purified protein was homogeneous as analyzed by SDS–PAGE. Fractions containing recombinant Csd3 were pooled and concentrated to 10 mg ml⁻¹ (0.24 mM) for crystallization using an Amicon Ultra-15 Centrifugal Filter Unit (Millipore).

2.2. Crystallization

Crystals were grown at 23°C by the sitting-drop vapour-diffusion method using a Mosquito robotic system (TTP Labtech). We obtained two types of native crystals (form 1 and form 2) under different crystallization conditions. For form 1 crystals, the sitting drops were prepared by mixing 0.4 µl reservoir solution [160 mM ammonium sulfate, 80 mM sodium acetate pH 4.6, 20%(v/v) PEG 4000, 20%(v/v) glycerol] and 0.4 µl purified protein in buffer C. Elongated rectangular crystals grew to approximate dimensions of 0.3 × 0.3 × 0.3 mm within a few days. For form 2 crystals, the recombinant Csd3_{Δ41} protein purified in buffer D was pre-incubated with buffer D supplemented with 1 mM zinc chloride under ice for 30 min prior to crystallization setup. Sitting drops were prepared by mixing 0.3 µl reservoir solution [200 mM diammonium hydrogen phosphate pH 7.9, 20%(v/v) PEG 3350] and 0.3 µl protein solution. Hexagonal bipyramidal crystals grew to approximate dimensions of 0.2 × 0.2 × 0.3 mm within a few days.

2.3. X-ray data collection and structure determination

Native X-ray diffraction data were collected on beamline BL-5A at Photon Factory, Japan using an ADSC Q315 CCD detector and on beamline BL-44XU of SPring-8, Japan using a DIP6040 imaging-plate/CCD hybrid detector (Table 1). Form 1 crystals were flash-cooled in a nitrogen-gas stream at 100 K. They diffracted to 2.0 Å resolution and belonged to space group *P*₂₁₂₁, with unit-cell parameters *a* = 62.6, *b* = 112.1, *c* = 112.9 Å. Assuming the presence of two Csd3 monomers in the asymmetric unit, the Matthews coefficient and solvent

content are 2.33 Å³ Da⁻¹ and 47.2%, respectively. Form 2 crystals were soaked for several seconds in a cryoprotectant solution consisting of the reservoir solution supplemented with 25%(v/v) glycerol and were flash-cooled in a nitrogen-gas stream at 100 K. Form 2 crystals diffracted to 1.98 Å resolution and belonged to space group *P*₆₅₂₂, with unit-cell parameters *a* = *b* = 91.5, *c* = 187.0 Å. Assuming the presence of one Csd3 monomer in the asymmetric unit, the Matthews coefficient and solvent content are 2.65 Å³ Da⁻¹ and 53.6%, respectively.

Selenomethionine-substituted Csd3_{Δ41} protein was not expressed in *E. coli*. Therefore, we prepared a Pt derivative of form 2 crystals by soaking them for 3 min in 2 µl of a heavy-atom-containing cryoprotectant solution, which consisted of the reservoir solution supplemented with 25%(v/v) glycerol and 30 mM K₂PtCl₄. Single-wavelength anomalous diffraction (SAD) data were collected at 100 K from the Pt-derivative crystal at a wavelength of 1.0720 Å using an ADSC Q270 CCD detector at the BL-7A experimental station of Pohang Light Source, Republic of Korea. The raw data were processed and scaled using the *HKL-2000* program suite (Otwinowski & Minor, 1997). SAD phases were calculated using *AutoSol* from the *PHENIX* software package (Adams *et al.*, 2010) and were further improved by *RESOLVE* (Terwilliger, 2003), yielding an interpretable electron-density map at 2.95 Å resolution. Phasing statistics are presented in Table 1.

2.4. Model building and refinement

RESOLVE was also used for autobuilding, which resulted in an initial model accounting for ~38% of the residues in the recombinant polypeptide chain with much of the sequence assigned. The refined model of the form 2 crystal was used as a search model to determine the Csd3_{Δ41} structure in the form 1 crystal by molecular replacement utilizing *MOLREP* (Vagin & Teplyakov, 2010). Manual model building was performed using *Coot* (Emsley *et al.*, 2010) and the models were refined with *REFMAC5* (Murshudov *et al.*, 2011), including bulk-solvent correction. A total of 5% of the data was randomly set aside as test data for the calculation of *R*_{free} (Brünger, 1992). The stereochemistry of the refined models was assessed by *MolProbity* (Chen *et al.*, 2010). Refinement statistics are presented in Table 1.

2.5. Identification of the bound metal ion by anomalous scattering

To test whether the metal-binding site is occupied by a Zn²⁺ ion, SAD data were collected from both form 1 and form 2 crystals at 100 K at an X-ray wavelength of 1.2820 Å on beamline 5C of Pohang Light Source (Supplementary Table S1). Anomalous difference maps were calculated using *FFT* (Immirzi, 1966; Ten Eyck, 1973) from the *CCP4* software package.

2.6. Equilibrium sedimentation

Equilibrium sedimentation experiments were performed using a Beckman ProteomeLab XL-A analytical ultra-

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

Data set	Form 1	Form 2	Pt (peak)
Data collection			
Beamline and source†	BL-5A, PF	BL-44XU, SPring-8	BL-7A, PLS
Space group	$P2_12_12_1$	$P6_522$	$P6_522$
Unit-cell parameters			
a (Å)	62.6	91.5	92.0
b (Å)	112.1	91.5	92.0
c (Å)	112.9	187.0	186.5
$\alpha = \beta$ (°)	90	90	90
γ (°)	90	120	120
X-ray wavelength (Å)	1.0000	0.9000	1.0720
Resolution range (Å)	50.0–2.00 (2.03–2.00)	50.0–1.98 (2.01–1.98)	30.0–2.95 (3.00–2.95)
Total/unique reflections	264154/54371	350996/33042	480902/18708‡
Completeness (%)	99.5 (100.0)	99.9 (100.0)	100.0 (100.0)‡
$\langle I \rangle / \langle \sigma(I) \rangle$	27.5 (2.9)	41.9 (3.5)	64.0 (13.0)‡
R_{merge}^{\S} (%)	8.3 (66.9)	8.8 (87.7)	12.9 (58.2)‡
$CC_{1/2}^{\parallel}$ (%)	99.7 (75.9)	99.9 (91.6)	99.9 (92.6)
SAD phasing			
Figure of merit (before/after density modification)			0.39/0.73
Model refinement			
PDB code	4rny	4rnz	
Resolution range (Å)	50.0–2.00	50.0–1.98	
$R_{\text{work}}/R_{\text{free}}^{\dagger\dagger}$ (%)	20.3/25.6	20.8/23.9	
No. of non-H atoms			
Protein	5822	2919	
Metal ion‡‡	2	3	
Water oxygen	225	148	
Glycerol	30	18	
Sulfate ion	45	—	
Phosphate ion	—	15	
Average B factor (Å ²)			
Protein	37.0	40.4	
Metal ion‡‡	26.1	33.7	
Water oxygen	34.6	40.8	
Glycerol	49.3	55.0	
Sulfate ion	2.3	—	
Phosphate ion	—	65.0	
R.m.s. deviations from ideal geometry			
Bond lengths (Å)	0.013	0.009	
Bond angles (°)	1.58	1.42	
R.m.s. Z -scores§§			
Bond lengths	0.64	0.46	
Bond angles	0.73	0.64	
Ramachandran plot¶¶			
Favoured/outliers (%)	96.9/0.0	96.3/0.0	
Poor rotamers (%)	0.16	0.00	

† PF, Photon Factory, Japan; PLS, Pohang Light Source, Republic of Korea. ‡ Friedel pairs were treated as separate observations. § $R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, where $I(hkl)$ is the intensity of reflection hkl , \sum_{hkl} is the sum over all reflections and \sum_i is the sum over i measurements of reflection hkl . ¶ $CC_{1/2}$ is the correlation coefficient of the mean intensities between two random half-sets of data. †† $R_{\text{work}} = \frac{\sum_{hkl} ||F_{\text{obs}}| - |F_{\text{calc}}||}{\sum_{hkl} |F_{\text{obs}}|}$, where R_{free} is calculated for a randomly chosen 5% of reflections which were not used for structure refinement and R_{work} is calculated for the remaining reflections. ‡‡ The two metal ions in form 1 are Zn^{2+} ions in the active site of two chains of Csd3 $_{\Delta 41}$. Form 2 contains a Zn^{2+} ion in the active site and is likely to contain two Ni^{2+} ions bound to the C-terminal hexahistidine tag. §§ Values obtained using *REFMAC5*. ¶¶ Values obtained using *MolProbity*.

centrifuge in buffer *C* at 4°C. The recombinant Csd3 $_{\Delta 41}$ protein samples were monitored by measuring the absorbance at 280 nm using six-sector cells at three rotor speeds (12 000, 14 000 and 18 000 rev min⁻¹, corresponding to 9660, 13 148 and 21 734g, respectively, at 6.0 cm radius) and three different protein concentrations (3.77, 5.03 and 6.29 μM). The protein concentration of the recombinant Csd3 $_{\Delta 41}$ was estimated using $\epsilon_{280 \text{ nm}} = 39\,770 \text{ M}^{-1} \text{ cm}^{-1}$ and a molecular mass of 42 562 Da, which includes the C-terminal hexahistidine-containing tag. The partial specific volume of the protein and

the buffer density were calculated using *SEDNTERP* (Laue *et al.*, 1992). Further data manipulation and data analysis by mathematical modelling were performed using *MLAB* (Knott, 1979).

2.7. Accession codes

The coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 4rny and 4rnz for form 1 and form 2 crystals, respectively.

3. Results and discussion

3.1. Structure determination of Csd3

H. pylori Csd3 was predicted to have a single transmembrane helix between Lys7 and Leu26 when its sequence was analyzed by the *TMHMM* server v.2.0 (Krogh *et al.*, 2001). Therefore, we initially overexpressed the construct comprised of residues 29–403 fused to a C-terminal hexahistidine-containing tag in *E. coli* and the expressed protein was crystallized. However, the crystals diffracted poorly to low resolution (~4 Å) despite extensive screening of crystallization conditions. This prompted us to try overexpressing a number of shorter constructs encompassing residues 40–403, 40–397, 40–393, 40–390, 42–403, 42–397, 42–393, 42–390, 45–403, 45–397, 45–393 and 45–390. Among these, only the N-terminally 41-residue truncated Csd3 (Csd3 $_{\Delta 41}$) comprised of residues 42–403 was expressed in soluble form, and it yielded two different forms (the orthorhombic form 1 and the hexagonal form 2) of well diffracting crystals (Table 1).

We have determined the structure of Csd3 $_{\Delta 41}$ (Fig. 1) using SAD data from a Pt-derivatized form 2 crystal. The model of the form 2 crystal was refined at 1.98 Å resolution to an R_{work} and R_{free}

of 20.8 and 23.9%, respectively (Table 1). The form 2 crystal contains one monomer of Csd3 $_{\Delta 41}$ in the asymmetric unit. This model of Csd3 $_{\Delta 41}$ excludes six residues (Pro251–Arg256) in a disordered loop near the metal-binding site as well as four residues Gly333–Thr336 and two histidines at the end of the C-terminal affinity tag. The model of the form 1 crystal was refined at 2.00 Å resolution to an R_{work} and R_{free} of 20.3 and 25.6%, respectively (Table 1). The form 1 crystal contains two monomers (chains *A* and *B*) of Csd3 $_{\Delta 41}$ in the asymmetric unit. Chains *A* and *B* are related by noncrystallographic

twofold symmetry. In both chains *A* and *B*, four residues Gly333–Thr336 and the C-terminal affinity tag (LEHHHHHH) are disordered. Chains *A* and *B* are highly similar to each other, with an r.m.s. deviation of 0.67 Å for 359 C α atoms. However, they show larger structural deviations from the chain of the form 2 crystal, with r.m.s. deviations of 1.61 and 1.85 Å for 353 C α atoms in chains *A* and *B*, respectively. The largest deviations occur in the η 1 and α 6 helices, with deviations of 6.65 and 5.58 Å at the C α atoms of Pro167 and Gly366, respectively (Supplementary Fig. S1). The observed structural variation is likely to be owing to the inherent flexibility of these regions and also owing to different crystal contacts.

3.2. Oligomeric state of Csd3 $_{\Delta 41}$ in solution

In the form 1 crystal, the two monomers in the asymmetric unit bury a relatively large surface area of 1130 Å² per monomer (6.0% of the monomer surface area), whereas the largest buried surface area in the form 2 crystal is 710 Å² per monomer (3.8% of the monomer surface area). The bulk of crystal interfaces have areas below 1000 Å², with very few representatives above this value (Duarte *et al.*, 2012). It is also well known that biological interfaces tend to exhibit large

areas, with a majority of cases having areas of 1000 Å² and above (Duarte *et al.*, 2012). This raised a question about the oligomeric state of Csd3 $_{\Delta 41}$ in solution. Therefore, we carried out equilibrium sedimentation experiments. All the measured data fit well to a homogeneous monomer model, indicating that Csd3 $_{\Delta 41}$ exists as monomers in solution at concentrations of up to 6.29 μ M. A representative result measured at 18 000 rev min⁻¹ using a protein concentration of 3.77 μ M is presented in Supplementary Fig. S2.

3.3. Three-domain structure of Csd3 $_{\Delta 41}$ and structural similarity searches

The structure of Csd3 $_{\Delta 41}$ can be divided into three domains: domain 1 (residues Glu42–Ile124), domain 2 (residues Ile125–Gly228 and Ala360–Phe403) and the LytM domain (residues Phe229–Thr359) (Fig. 1). A structural similarity search using the DALI server (Holm & Rosenström, 2010) revealed that the entire structure (form 1 crystal, chain *A*) of Csd3 $_{\Delta 41}$ resembles an outer-membrane protein from *Neisseria meningitidis* (NMB0315; PDB entry 3slu; Wang *et al.*, 2011; r.m.s. deviation of 5.2 Å for 315 equivalent C α positions, Z-score of 21.7 and sequence identity of 26%) and a putative lysostaphin peptidase from *Vibrio cholerae* (VC0503; PDB entry 2gu1;

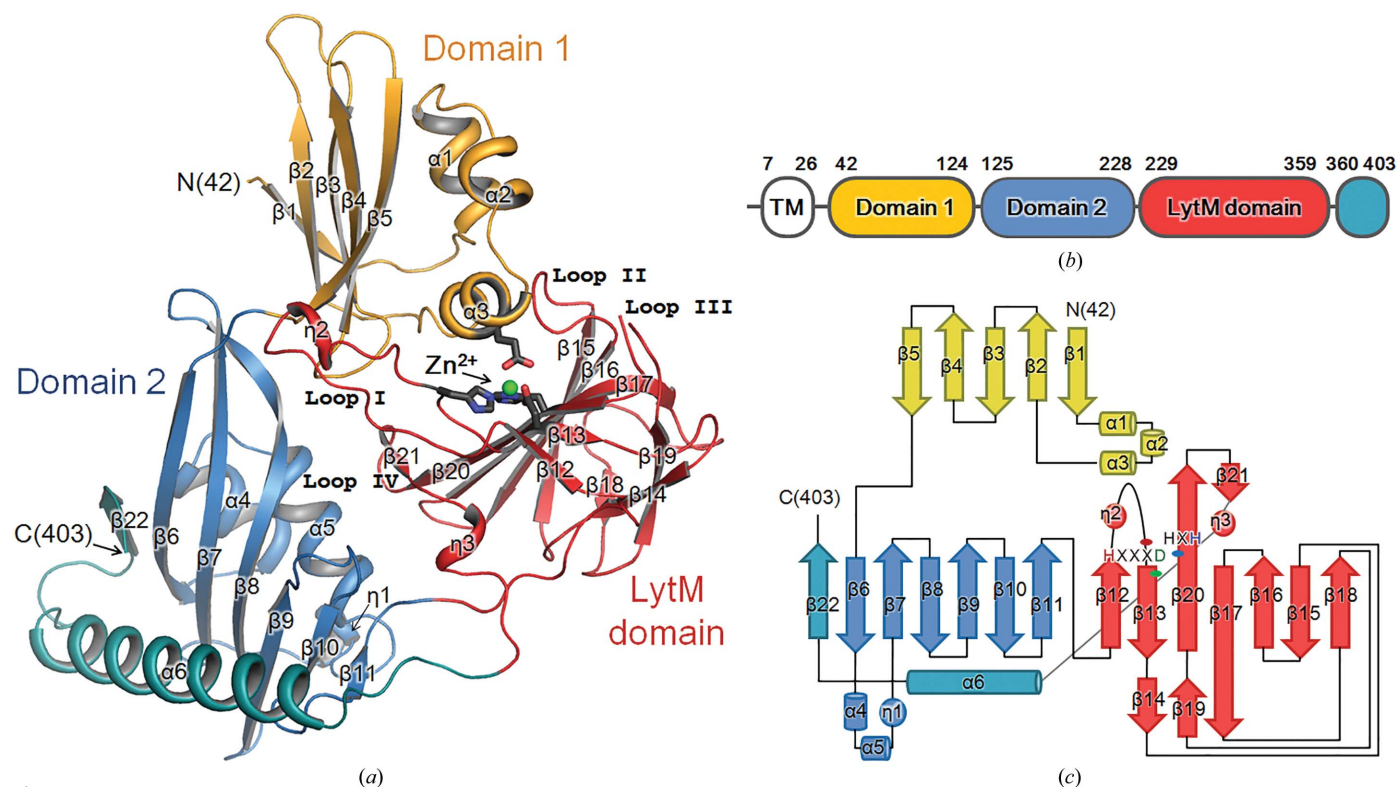


Figure 1 Overall monomer structure and topology of *H. pylori* Csd3 $_{\Delta 41}$. (a) Ribbon diagram of the Csd3 $_{\Delta 41}$ monomer (chain *A* of form 1), with the secondary-structure elements labelled. Domain 1, the core of domain 2 and the LytM domain are shown in bright orange, sky blue and red, respectively. The C-terminal α -helix (α 6) and β -strand (β 22) are coloured teal. The green sphere is a Zn²⁺ ion. Side chains of the metal-coordinating residues (Glu74, His259, Asp263 and His341) are shown in stick models (dark grey). The secondary structures were defined by STRIDE (Heinig & Frishman, 2004). The walls of the active site in the LytM domain are made up of four loops: loop I (the β 12– β 13 loop), loop II (the β 15– β 16 loop), loop III (the β 19– β 20 loop) and loop IV (the β 20– β 21 loop). (b) Domains of *H. pylori* Csd3 coloured as in (a). TM, transmembrane helix. Residue numbers for each domain are indicated. (c) Topology diagram of Csd3 $_{\Delta 41}$ coloured as in (a). α -Helices, β -strands, $_{310}$ -helices and loops are shown as cylinders, arrows, circles and solid lines, respectively. Structure figures were drawn using PyMOL (DeLano, 2002).

Ragumani *et al.*, 2008; r.m.s. deviation of 4.5 Å for 219 equivalent C α positions, Z-score of 18.8 and sequence identity of 23%) (Fig. 2 and Supplementary Fig. S3). The above two proteins are three-domain proteins, but domain 2 of VC0503 is not included in the structural overlap owing to a large difference in domain arrangements. They belong to the M23B metallopeptidase family (Wang *et al.*, 2011; Ragumani *et al.*, 2008).

Domain 1 of Csd3 has an α/β fold consisting of a five-stranded antiparallel β -sheet ($\beta 1\downarrow-\beta 2\uparrow-\beta 3\downarrow-\beta 4\uparrow-\beta 5\downarrow$) and

three short α -helices (Fig. 1). A cluster of continuously linked helices ($\alpha 1-\alpha 2-\alpha 3$) is inserted between strands $\beta 1$ and $\beta 2$ and is packed on one side of the β -sheet. The core of domain 2 (Ile125–Gly228) also has an α/β fold consisting of a six-stranded antiparallel β -sheet ($\beta 6\downarrow-\beta 7\uparrow-\beta 8\downarrow-\beta 9\uparrow-\beta 10\downarrow-\beta 11\uparrow$) and three short helices (Fig. 1). As in domain 1, a cluster of consecutively linked helices ($\alpha 4-\alpha 5-\eta 1$) is inserted between strands $\beta 6$ and $\beta 7$ and lies on the concave side of the β -sheet. Domain 1 and the core of domain 2 appear to share a common fold despite a very low level of amino-acid sequence identity

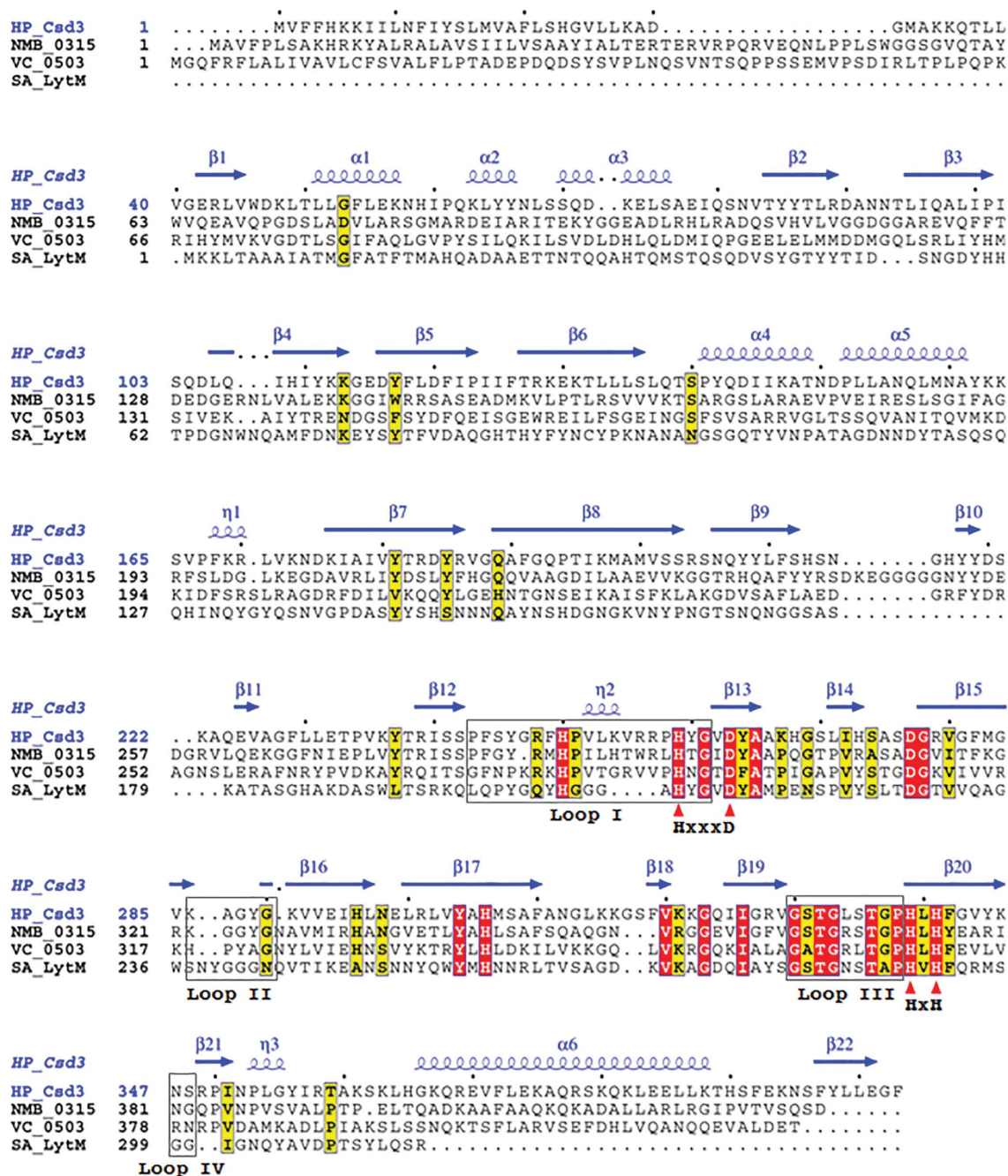


Figure 2 Sequence alignment of four M23B metallopeptidase proteins. Sequence alignment of Csd3 from *H. pylori* strain 26695 (HP0506; HP_Csd3; SWISS-PROT accession code O25247), NMB0315 from *N. meningitidis* (NMB_0315; Q9K163), VC0503 from *Vibrio cholerae* (VC_0503; Q9KUL5) and LytM from *S. aureus* (SA_LytM; O33599). Red triangles indicate the conserved residues in the HxxxD and HxH motifs (H259-xxx-D263 and H339-x-H341 in *H. pylori* Csd3) that are important for the metallopeptidase activity. Four loops are indicated by grey boxes.

(Supplementary Fig. S4a). However, their charge distributions on the surface are distinct from each other (Supplementary Fig. S4b). The C-terminal helix (α_6) and β -strand (β_{22}) protruding from the LytM domain are tightly associated with the core of domain 2, covering another face of the β -sheet and extending the β -sheet, respectively.

When a *DALI* search was performed separately against domain 1 (form 1 crystal, chain A), the highest structural similarity is observed to the corresponding domains from VC0503 and NMB0315, with *Z*-scores of 7.6 and 5.0, respectively. Besides these, domain 1 displays a very remote structural similarity to single-chain monellin (PDB entry 1mnl; Lee *et al.*, 1999), with a *Z*-score of 4.2, and human cystatin A (also called stefin A; PDB entry 3kse, chain D; M. Renko & D. Turk, unpublished work), with a *Z*-score of 3.5. When the *DALI* search was performed separately against the core of domain 2 (form 1 crystal, chain A), the highest structural similarity is observed to the corresponding domains from VC0503 and NMB0315, with *Z*-scores of 14.6 and 14.5, respectively. Besides these, domain 2 exhibits a very remote structural similarity to the core of a single-stranded DNA-binding protein from *Aeropyrum pernix* K1 (PDB entry 4pso; Ghalei *et al.*, 2014) and perfringolysin O from *Clostridium perfringens* (PDB entry 1m3i; J. Rossjohn, M. Parker, G. Polekhina, S. Feil

& R. Tweten, unpublished work), with *Z*-scores of 4.5 and 4.4, respectively.

The bulk of the Csd3 LytM domain is folded as a two-layered sandwich consisting of a larger, seven-stranded antiparallel β -sheet ($\beta_{12}\uparrow$ – $\beta_{13}\downarrow$ – $\beta_{20}\uparrow$ – $\beta_{17}\downarrow$ – $\beta_{16}\uparrow$ – $\beta_{15}\downarrow$ – $\beta_{18}\uparrow$) and a smaller, three-stranded antiparallel β -sheet ($\beta_{14}\downarrow$ – $\beta_{19}\uparrow$ – $\beta_{17}\downarrow$). These β -sheets share a long strand (β_{17}), which is bent like the letter J. Another long strand (β_{20}) associates with a short strand (β_{21}) to form a mini, antiparallel β -sheet ($\beta_{20}\uparrow$ – $\beta_{21}\downarrow$), which is followed by a 3_{10} -helix (η_3) (Fig. 1). As expected, the LytM domain (form 1 crystal, chain A) of Csd3 $_{\Delta 41}$ exhibits significant structural similarity to the corresponding domains of the M23B metallopeptidases, such as NMB0315 (PDB entry 3slu; r.m.s. deviation of 2.8 Å for 123 equivalent C $^{\alpha}$ positions, *Z*-score of 19.7 and sequence identity of 44%) and VC0503 (PDB entry 2gu1; r.m.s. deviation of 2.7 Å for 120 equivalent C $^{\alpha}$ positions, *Z*-score of 19.0 and sequence identity of 35%). It is also structurally similar to the LytM domains of *Staphylococcus aureus* glycylglycine endopeptidase, another M23B metallopeptidase (PDB entry 2b13; Firczuk *et al.*, 2005; r.m.s. deviation of 1.9 Å for 118 equivalent C $^{\alpha}$ positions, *Z*-score of 18.3 and sequence identity of 30%), and the virulence factor LasA from *Pseudomonas aeruginosa*, an M23A metallopeptidase (PDB entry 3it7; Spencer *et al.*,

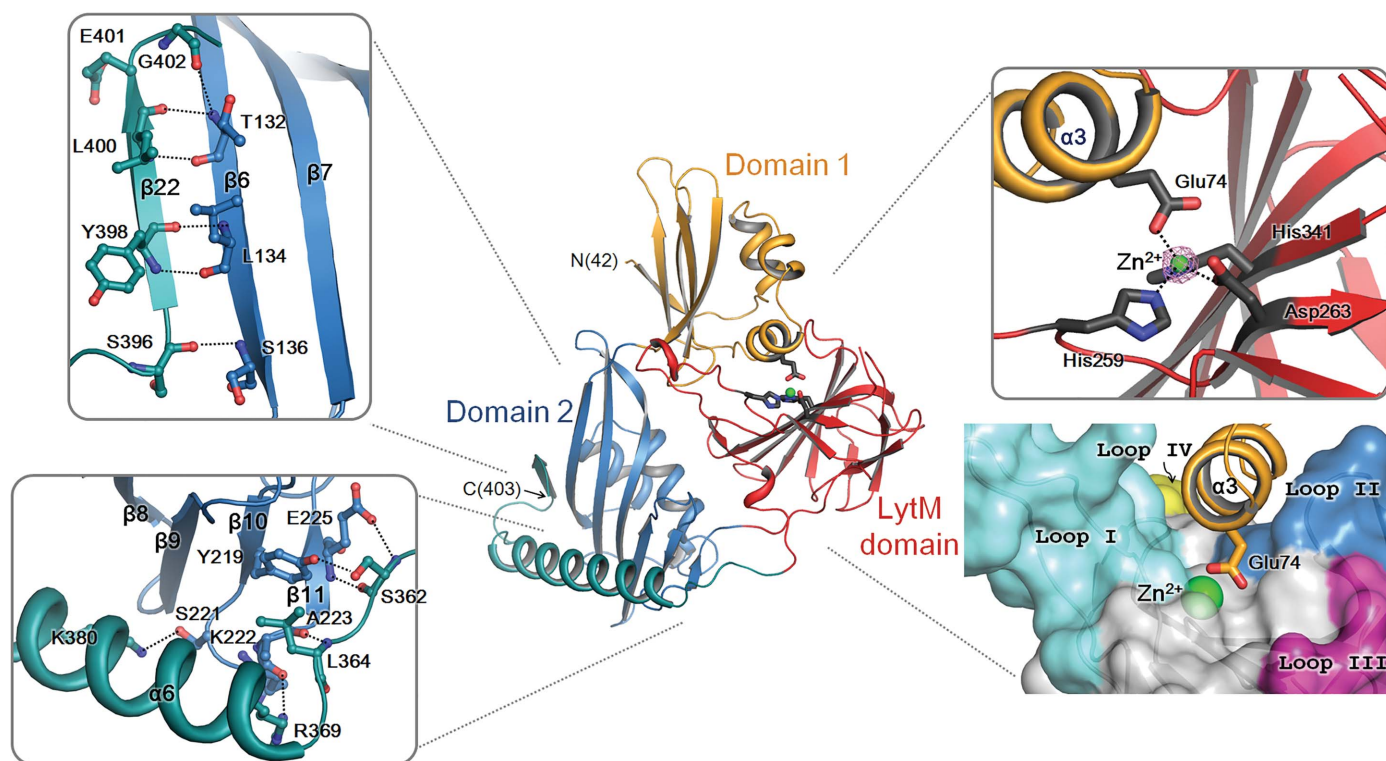


Figure 3

Metal coordination in Csd3 $_{\Delta 41}$ and interactions of the C-terminal α -helix (α_6) and β -strand (β_{22}) with the core of domain 2. A ribbon diagram of the Csd3 $_{\Delta 41}$ monomer, coloured as in Fig. 1(a), is shown in the centre. The close-up views on the left represent interactions of the C-terminal strand (β_{22}) with the β_6 strand in the core of domain 2 (top) and of the C-terminal helix (α_6) with the β -sheet in the core of domain 2 (bottom). Hydrogen-bond interactions are shown as black dotted lines. The close-up views on the right represent the ribbon diagram of the Zn $^{2+}$ -binding motif (top) and the surface representation of the substrate-binding groove formed by four loops of the LytM domain (bottom). The electron density for the Zn $^{2+}$ ion in the OMIT $mF_o - DF_c$ map is shown as a light pink mesh (contoured at 10σ). To show the detailed interactions more clearly, the close-up views have slightly different orientations.

2010; r.m.s. deviation of 2.1 Å for 107 equivalent C α positions, Z-score of 11.4 and sequence identity of 21%). M23A metallopeptidases are distinguished from the more numerous M23B enzymes by structural features that include disulfide bridges and the possession of an additional C-terminal subdomain as well as alterations to the active-site region that are manifested in differences in the sequence spacing between the His and Asp residues of the HxxxD motif of the M23B family (Spencer *et al.*, 2010). In M23A metallopeptidases the intervening sequence is variable and several times longer than that of M23B.

After leaving the LytM domain, the polypeptide chain is connected to a C-terminal α -helix ($\alpha 6$) and a short β -strand ($\beta 22$) that fold back onto the core of domain 2, making a tight interaction with the core of domain 2 (Fig. 3). The $\alpha 6$ helix sits on the opposite side of the β -sheet of domain 2 from the cluster of three helices ($\alpha 4$ – $\alpha 5$ – $\eta 1$) and is tightly anchored to the β -sheet through hydrophobic and hydrogen-bond interactions (Fig. 3). The $\beta 22$ strand lies antiparallel to the $\beta 6$ strand of domain 2 and forms the seventh strand of the β -sheet in the core of domain 2 (Fig. 3). The C-terminal $\alpha 6$ – $\beta 22$ region may play an important role in stabilizing the interdomain orientation of the core of domain 2 and the LytM domain of Csd3.

3.4. The active-site structure of the LytM domain

The Csd3 LytM domain possesses characteristic features of the MEROPS M23 family of metallopeptidases. In these enzymes, the catalytic residues are anchored by the larger antiparallel β -sheet and are grouped around a metal ion (Figs. 3, 4 and 5). We confirmed that this metal-binding site is indeed occupied by a Zn $^{2+}$ ion by calculating anomalous difference maps using anomalous data collected at the zinc absorption edge of 1.2820 Å (Supplementary Table S1 and Supplementary Fig. S5). In all Csd3 $_{\Delta 41}$ structures we observed a tetracoordinated Zn $^{2+}$ ion in the active site, with the expected three amino-acid ligands (His259 e2 , Asp263 $^{\delta 1}$ and His341 $^{\delta 1}$) and one amino-acid ligand (Glu74 e2) from the $\alpha 3$ helix of domain 1 (Fig. 3). His259 and Asp263 belong to the characteristic HxxxD motif, while His341 is the second histidine of the HxH motif. The metal–ligand atom distances are in the range 1.95–2.11 Å, consistent with typical Zn $^{2+}$ ion–ligand atom distances. The Zn $^{2+}$ -coordinating groups are stabilized by making hydrogen bonds to neighbouring amino-acid residues or water molecules. His259 $^{\delta 1}$ and His341 e2 are hydrogen-bonded to the main-chain O atom of Pro243 and a water molecule, respectively. Glu74 e1 is hydrogen-bonded to His306 e2 . His306 of Csd3 is conserved among M23B peptidase proteins (Fig. 2) and corresponding residues similarly interact with the fourth ligand of the metal ion. For example, His343 in NMB0315 makes a hydrogen bond to a water molecule (Fig. 5), while His260 in the active LytM from *S. aureus* makes a hydrogen bond to a phosphate ion. If Csd3 were in the active state for peptidase activity, two water molecules should occupy the positions close to the side-chain O atoms of Glu74, as

pentacoordination is believed to be consistent with the proposals for the catalytic mechanism (Sabala *et al.*, 2014).

In Csd3, the distances between the Zn $^{2+}$ and two O atoms from the side chain of Glu74 (Glu74 e1 and Glu74 e2) are 2.9 and 2.0 Å, respectively. This coordination differs from that in the uncomplexed structure of LasA from *P. aeruginosa* (Spencer *et al.*, 2010). The Zn $^{2+}$ ion in the uncomplexed structure of LasA is described as pentacoordinated (with a slightly distorted trigonal bipyramidal geometry) with three conserved metal ligands and two conserved water molecules. The distances between Zn $^{2+}$ and the water O atoms (Wat-1 and Wat-2) are 2.1 and 2.7 Å, respectively. In the tartrate-complexed structure of LasA, tartrate O atoms occupy nearly identical positions as both Zn $^{2+}$ -coordinated water molecules, with the distances between Zn $^{2+}$ and tartrate O atoms being 2.5 and 2.1 Å, respectively. However, the Zn $^{2+}$ coordination in the tartrate-complexed structure has been described as a tetrahedral geometry (Spencer *et al.*, 2010). It has been suggested that pentacoordination is consistent with many of

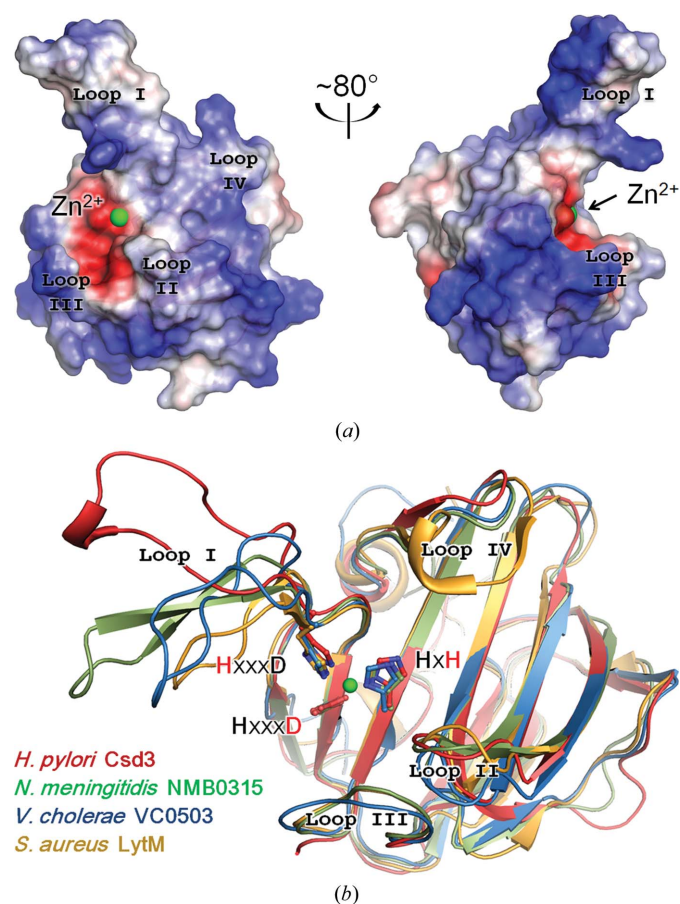


Figure 4
LytM domain of Csd3. (a) Two different views of the electrostatic potential surface of the LytM domain of Csd3. The positive and negative electrostatic potentials on the surface are coloured blue and red, respectively. Four loops that form the substrate-binding groove around the Zn $^{2+}$ ion (green sphere) are denoted by loops I–IV. (b) Superposition of LytM domains in four M23B metallopeptidases. The LytM domains of *H. pylori* Csd3 (red), *N. meningitidis* NMB0315 (pale green; PDB entry 3slu), *V. cholerae* VC0503 (sky blue; PDB entry 2gu1) and *S. aureus* LytM (yellow/orange; PDB entry 1qwy) are superimposed and shown as ribbon diagrams.

the proposals for the catalytic mechanism and that the observations of variable Zn^{2+} coordination among the characterized M23 family metallopeptidases may reflect low energy barriers for changes in Zn^{2+} coordination for catalysis (Sabala *et al.*, 2014; Spencer *et al.*, 2010).

The floor of the substrate-binding groove of the LytM domain is built by the larger β -sheet and the walls of the active site are made up of four loops: loop I (the β_{12} – β_{13} loop), loop II (the β_{15} – β_{16} loop), loop III (the β_{19} – β_{20} loop) and loop IV (the β_{20} – β_{21} loop) (Figs. 2, 3 and 4). A superimposition of the LytM domain of Csd3 with those of NMB0315, VC0503 and *S. aureus* LytM reveals that the larger β -sheets deviate little from each other, but that the difference is larger for loops I–IV (Fig. 4*b*). Loop I of Csd3 shows the most significant deviation among the homologues. In form 1 crystals, it is involved in crystal-packing interactions by forming sulfate-mediated salt bridges (Supplementary Fig. S6). In form 2 crystals, it is not involved in crystal packing and is disordered. Therefore, we conclude that the observed structural difference of loop I is largely owing to differences in crystal packing and also to its inherent flexibility. In all structures of Csd3 $_{\Delta 41}$, three residues (Gly333–Leu334–Ser335) of loop III are disordered. Loop III has significant sequence conservation and immediately precedes the HxH motif (Fig. 2). Gly333 and Ser335 of Csd3 are conserved, whereas Leu334 in Csd3 is substituted by Arg368 and Arg365 in NMB0315 and VC0503, respectively, and Asn286 in *S. aureus* LytM (Fig. 2). In the structure of the active LytM from *S. aureus* (PDB entry 2b44), a phosphate ion, as a substrate candidate, is fixed in space in the active site by multiple hydrogen bonds to Asn286 (Firczuk *et al.*, 2005). Ser369 of NMB0315 and Ser287 of the active LytM from

S. aureus, corresponding to the disordered Ser335 in loop III of Csd3, interact with the first histidine of the HxH motif. This histidine is considered to be a catalytic residue by coordinating a Zn^{2+} -bound water molecule that acts as the nucleophile in the hydrolytic reaction (Spencer *et al.*, 2010). Interestingly, the side chain of His339 of Csd3, the first histidine of the HxH motif, shows two different orientations. Its orientation in the form 2 crystals is similar to those in other homologous proteins (Fig. 5). However, it is flipped about 104° towards the solvent in the form 1 crystals (Fig. 5).

3.5. Domain 1 occludes the active site of the LytM domain

In both the form 1 and the form 2 crystals, domain 1 blocks the active-site cleft of the LytM domain (Fig. 6), with the protruding helix α_3 contributing to the Zn^{2+} coordination sphere. Several negatively charged residues from domain 1 (Asp72, Glu74, Glu78 and Asp105) and positively charged residues (Arg257, Arg301 and Arg349) as well as the Zn^{2+} ion from the LytM domain form strong salt bridges at the interface between these domains. As described above, the Zn^{2+} ion in the active site of the LytM domain is tetrahedrally coordinated by the three conserved residues (His259 and Asp263 of the HxxxD motif and the second histidine His341 of the HxH motif) of the LytM domain as well as the nonconserved Glu74 in helix α_3 of domain 1. There is a hydrogen bond between Glu74 and the conserved His306 of the LytM domain. Asp72 interacts with Arg257 in loop I of the LytM domain by forming a salt bridge. Glu78 forms a salt bridge with Arg301 in the β_{17} strand of the LytM domain; it also makes a hydrogen bond to Tyr289 in loop II of the LytM domain. In addition, Asp105 in

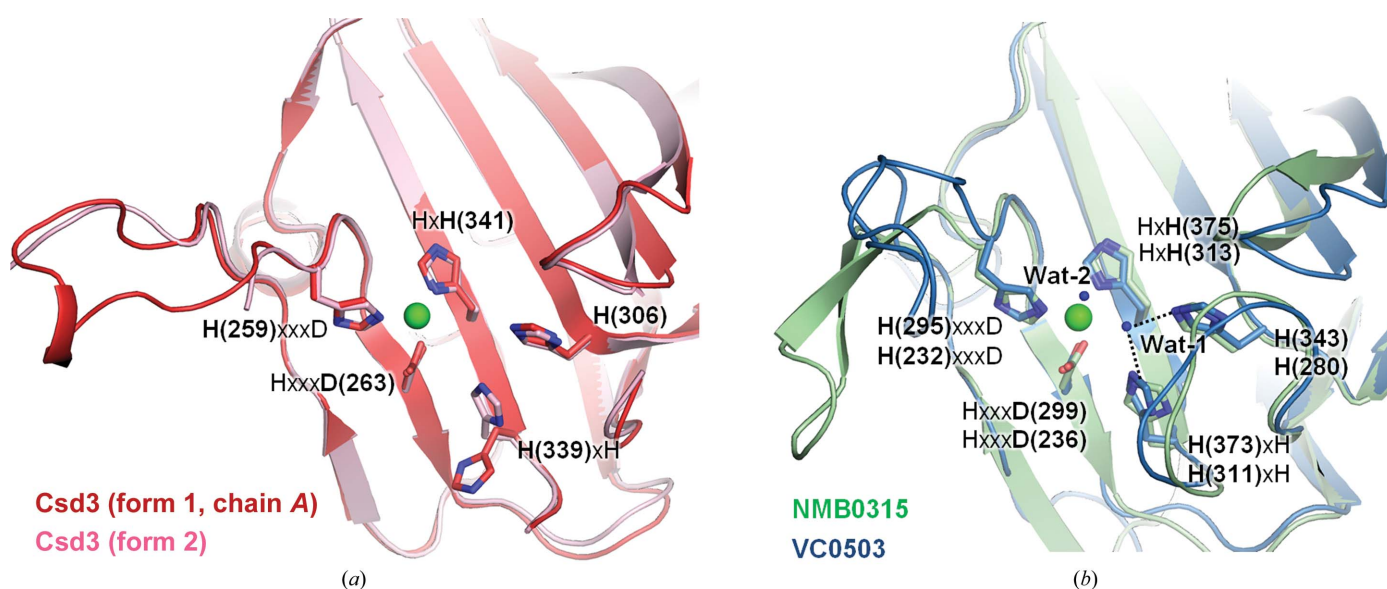


Figure 5

Conserved residues in the active site of M23B metallopeptidases. (a) Superposition of LytM domains in two crystal forms of *H. pylori* Csd3. Chain A of the form 1 crystal (red) and the form 2 crystal (pink) are shown as ribbon diagrams. (b) Superposition of the LytM domains in two M23B family members (NMB0315 and VC0503) coloured as in Fig. 4(b). Conserved residues are shown as stick models and are labelled (NMB0315 at the top and VC0503 at the bottom). Metal ions and water molecules are shown as green spheres and blue dots, respectively. In NMB0315, the Zn^{2+} ion was replaced by an Ni^{2+} ion during affinity chromatography (Wang *et al.*, 2011). Wat-1 and Wat-2 are present in NMB0315, where the Ni^{2+} ion is pentacoordinated. Black dotted lines denote hydrogen bonds to Wat-1 in NMB0315. In VC0503, the Zn^{2+} ion is tetracoordinated, with one water molecule (omitted for clarity) located between Wat-1 and Wat-2.

the $\beta 3$ – $\beta 4$ loop of domain 1 forms a salt bridge with Arg349 in the $\beta 21$ strand of the LytM domain. Tyr65 in the $\alpha 2$ helix of domain 1 forms a hydrogen bond to the main-chain N atom of Gly288 in loop II of the LytM domain. Ser103 in the $\beta 3$ – $\beta 4$ loop of domain 1 is hydrogen-bonded to Tyr260 in the $\beta 13$ strand of the LytM domain. On the basis of these extensive interactions between domain 1 and the LytM domain, we suggest that Csd3 $_{\Delta 41}$ is in an auto-inhibited state in the crystal. This is similar to the occlusion of the active site by an N-terminal segment in structurally related three-domain proteins in the inhibited conformational state (Supplementary Fig. S7). In NMB0315, the N-terminal short $\beta 3$ – $\beta 4$ loop stretches into the active site and tightly associates with the catalytic domain (Wang *et al.*, 2011). In VC0503, an N-terminal helix ($\alpha 2$) occupies the active-site cleft (Ragumani *et al.*, 2008). In both

cases the loop or helix does not participate in metal coordination.

The Zn²⁺ ion bound to the active site of Csd3 is tetrahedrally coordinated by the side chains of four residues (Glu74, His259, Asp263 and His341) without any ordered water molecules. Glu74 is not conserved, whereas His259, Asp263 and His341 are conserved (Fig. 2). The ligand position corresponding to the nonconserved Glu74 is usually occupied by a catalytic water molecule or various anions in the active forms of other M23 metallopeptidases (Supplementary Fig. S7). In the case of LytM from *S. aureus*, a truncated version that lacks Asn117 has a much higher specific activity than the full-length enzyme, in which the poorly conserved Asn117 of the inhibitory domain occupies one of the ligand sites of Zn²⁺ (Odintsov *et al.*, 2004). In the active, truncated LytM from

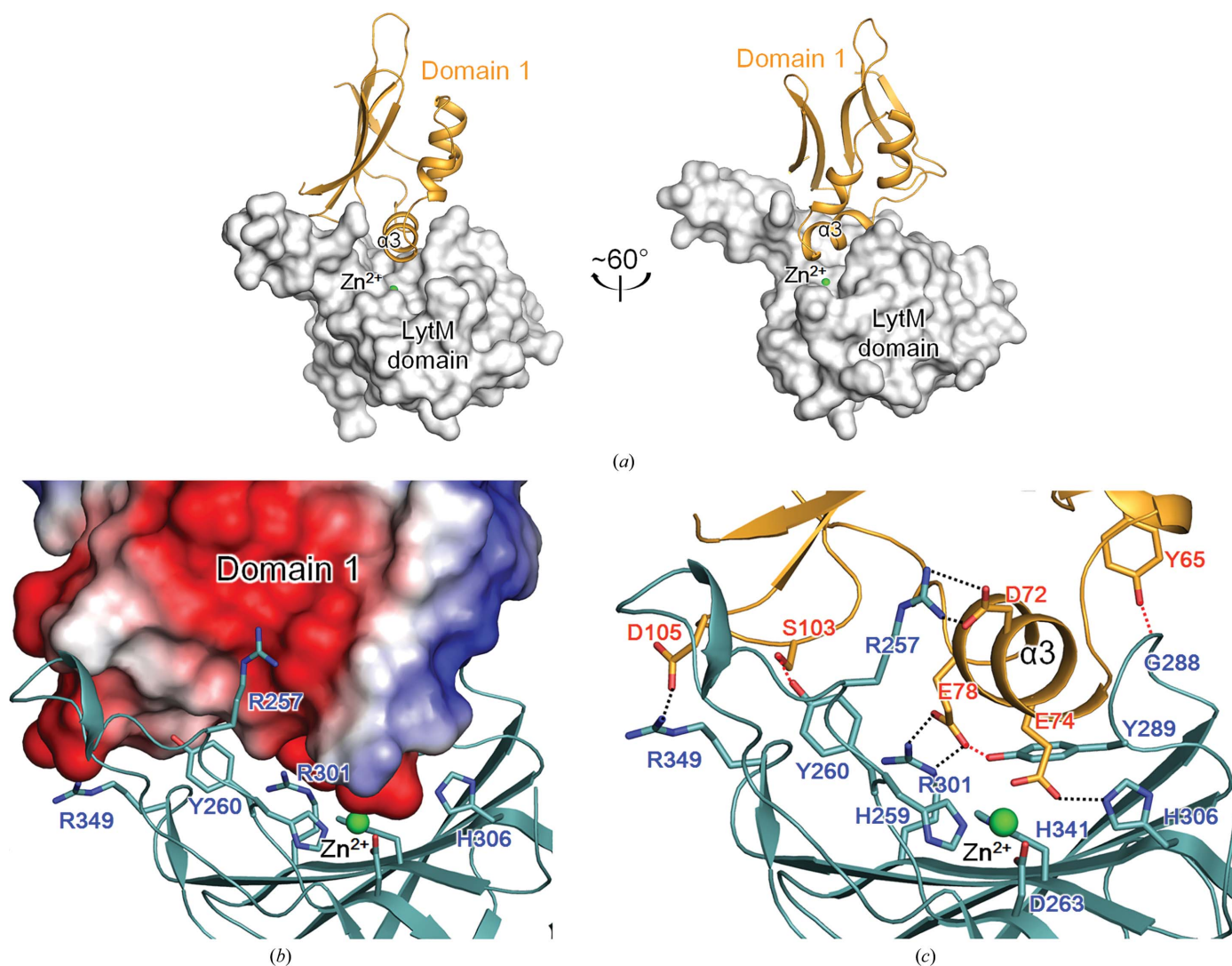


Figure 6 The active site of the Csd3 LytM domain is blocked by domain 1. In this figure, domain 2 is omitted for clarity. (a) Two different views of the interaction between domain 1 (shown as bright orange ribbons) and the LytM domain (shown as a grey surface diagram) of Csd3. (b) Electrostatic surface diagram of domain 1 and a ribbon diagram of the LytM domain (coloured deep teal). Residues of the LytM domain located at the domain interface are shown as stick models and are labelled. (c) Ribbon diagram of domain 1 (bright orange) and the LytM domain (coloured deep teal). Residues at the domain interface are shown as stick models. Hydrogen-bond interactions and salt-bridge interactions are shown as red and black dotted lines, respectively.

S. aureus, Asn117 is replaced by phosphate, cacodylate or tartrate from the crystallization buffer (Firczuk *et al.*, 2005).

4. Discussion

We have determined the first crystal structure of Csd3 from *H. pylori*, a protein that influences the helical cell shape crucial for the survival of *H. pylori* in the stomach. We note that the fold of Csd3 domain 1 is very remotely related to monellin/cystatin superfamily proteins, some of which act as inhibitors of cysteine peptidases. The active site of the C-terminal LytM domain is blocked by the inhibitory domain 1, in particular helix $\alpha 3$ with the Zn^{2+} -coordinating Glu74. The structure of an AmiB orthologue from *Bartonella henselae* revealed that the active site of AmiB is similarly occluded by a conserved α -helix with a Zn^{2+} -coordinating glutamate residue (Glu290) and it was suggested that auto-inhibition is a critical feature of the regulation of peptidoglycan amidases required for cell division in Gram-negative bacteria (Yang *et al.*, 2012). We measured the CPase activity of *H. pylori* Csd3 against the synthetic muramyl pentapeptide (the substrate for the CPase activity of Csd3) using mass analysis. The muramyl pentapeptide (5 mM) was incubated at 37°C for 150 min with recombinant Csd3 $_{\Delta 41}$ (5 μ M) in the presence of 2.5 mM metal ion (either Zn^{2+} or Mg^{2+}) or 2.5 mM EDTA. We used two different buffers for dissolving the reaction mixture: (i) 20 mM HEPES pH 7.9, 100 mM NaCl or (ii) 20 mM sodium phosphate pH 6.0. No reaction product was observed in all of the reaction conditions that we tested. As a positive control, we could detect the muramyl tetrapeptide product when the muramyl pentapeptide was incubated with *Enterococcus faecium* VanY. This result is consistent with a latent, inactive state of Csd3 $_{\Delta 41}$ in the crystal.

Our work suggests that the inhibitory domain 1, including the $\alpha 3$ helix, should be displaced from the active site of the LytM domain for activation of the latent Csd3. The activation of *H. pylori* Csd3 may occur by autoproteolysis or may require proteolytic cleavage by other endopeptidases to free the catalytic LytM domain from the inhibitory domain 1. When we tried limited proteolysis with trypsin, chymotrypsin and pepsin, no discrete cleavage at specific sites took place. To physically remove the inhibitory domain 1 from Csd3, we tried to overexpress various constructs (residues 80–403, 81–403, 83–403, 124–403, 125–403, 127–403, 130–403, 227–403, 230–403, 233–403, 125–359, 127–359, 227–359, 230–359 and 233–359) that lacked helix $\alpha 3$. However, none of them was expressed in *E. coli* in a soluble form. Alternatively, a large conformational change induced by an allosteric regulator may lead to opening of the active site. For instance, a large conformational change culminates in opening of the active site to permit substrate entry when an allosteric site ~ 60 Å distant from the D,D-transpeptidase active site is occupied in penicillin-binding protein 2a (PBP2a) from *S. aureus* (Otero *et al.*, 2013).

To provide details of the interaction of the substrate with Csd3, we soaked the form 1 and form 2 crystals in a cryo-protectant solution that was supplemented with the muramyl

pentapeptide. No electron density for the substrate peptide was observed. We also tried to co-crystallize Csd3 $_{\Delta 41}$ in the presence of excess muramyl pentapeptide under conditions similar to the crystallization conditions for form 1 and form 2 crystals, but this attempt did not produce any crystals. Owing to a limited supply of the muramyl pentapeptide, it was not possible to perform more extensive screening of the crystallization conditions in its presence. Co-crystallization of Csd3 $_{\Delta 41}$ in the presence of cross-linked dimers of muramyl pentapeptides and tetrapeptides (the substrate for the EPase activity of Csd3) was not possible because this substrate is neither commercially available nor easy to synthesize.

H. pylori Csd1 (HP1543), Csd2 (HP1544) and Csd3 (HP0506) contain a LytM domain of the M23B metallo-peptidase family; they all catalyze the same reaction as EPases. Csd2 could be a nonpeptidase homologue of the M23B family, as the first histidines of the characteristic HxxxD and HxH motifs in its LytM domain are mutated to Glu at residue 165 and Lys at residue 246, respectively. Csd3 exhibits CPase activity, while Csd1 may or may not have such an activity. *H. pylori* Csd3 (with 403 residues) is considerably longer in sequence compared with Csd1 (312 residues) and Csd2 (308 residues). Sequence alignment of Csd1 and Csd2 with either domain 1 or the core of domain 2 of Csd3 reveals that the proregions of both Csd1 and Csd2 have higher levels of sequence similarity to domain 1. The significantly shorter lengths of Csd1 and Csd2 indicates the presence of only one domain in their proregions. This may affect their CPase activity. However, no information on the three-dimensional structures of Csd1 and Csd2 is available at present and thus structural studies on them are required to compare the structures of Csd1 and Csd2 with Csd3 and to understand the functional differences in structural terms.

Bacteria may change their morphology to fit the circumstances for survival (Young, 2007). Under stressed conditions such as subinhibitory concentrations of antibiotics, *H. pylori* is able to enter a viable but nonculturable state, in which the microorganism modifies its morphology from a spiral to a coccoid form (Cellini, 2014). The viable coccoid form is more persistent in the host and environment (Cellini, 2014). It has been reported that overproduction of Csd3/HdpA in *H. pylori* strain N6 led to a transformation from rod-shaped to viable cocci-shaped bacteria (Bonis *et al.*, 2010). This raises the possibility that transition of the helical shape to a coccoid shape may be associated with the overexpression of Csd3. If this is the case, Csd3 could be an attractive drug target not only for eradicating helical-shaped *H. pylori* but also for inhibiting the morphological transformation into the persistent coccoid form.

Acknowledgements

We thank the beamline staff for assistance during X-ray diffraction experiments at Pohang Light Source (beamlines BL-5C and BL-7A), SPring-8 (beamline BL-44XU; proposal No. 2012B6741) and Photon Factory (beamlines BL-1A, BL-5A, BL-17A, NE3A and NW12A). The work in the laboratory of SWS was supported by the Korea Ministry of Science, ICT

and Future Planning, National Research Foundation (NRF) of Korea (2013R1A2A1A05067303) and the Innovative Drug Research Center for Metabolic and Inflammatory Disease, the Korea Ministry of Health, Welfare and Family Affairs (Korea Healthcare Technology R&D Project A092006). The work in the laboratory of BIL was supported by the Basic Science Research Program (NRF-2010-0020993) through the NRF of Korea funded by the Ministry of Education. HSK was supported by the NRF of Korea funded by the Ministry of Education (2012-039930). HJY was supported by the NRF of Korea funded by the Ministry of Education (2014R1A1A3A04050250). The work in the US was supported by a grant from the National Institutes of Health (AI090348).

References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
- Berg, H. C. & Turner, L. (1979). *Nature (London)*, **278**, 349–351.
- Bonis, M., Ecobichon, C., Guadagnini, S., Prévost, M.-C. & Boneca, I. G. (2010). *Mol. Microbiol.* **78**, 809–819.
- Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
- Cellini, L. (2014). *World J. Gastroenterol.* **20**, 5575–5582.
- 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.
- Costa, K., Bacher, G., Allmaier, G., Dominguez-Bello, M. G., Engstrand, L., Falk, P., de Pedro, M. A. & García-del Portillo, F. (1999). *J. Bacteriol.* **181**, 3710–3715.
- DeLano, W. L. (2002). *PyMOL*. <http://www.pymol.org>.
- Duarte, J. M., Srebnik, A., Schärer, M. A. & Capitani, G. (2012). *BMC Bioinformatics*, **13**, 334.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Firczuk, M., Mucha, A. & Bochtler, M. (2005). *J. Mol. Biol.* **354**, 578–590.
- Frirdich, E., Biboy, J., Adams, C., Lee, J., Ellermeier, J., Giolda, L. D., DiRita, V. J., Girardin, S. E., Vollmer, W. & Gaynor, E. C. (2012). *PLoS Pathog.* **8**, e1002602.
- Ghalei, H., von Moeller, H., Eppers, D., Sohmen, D., Wilson, D. N., Loll, B. & Wahl, M. C. (2014). *Nucleic Acids Res.* **42**, 6698–6708.
- Hazell, S. L., Lee, A., Brady, L. & Hennessy, W. (1986). *J. Infect. Dis.* **153**, 658–663.
- Heinig, M. & Frishman, D. (2004). *Nucleic Acids Res.* **32**, W500–W502.
- Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.
- Immirzi, A. (1966). *Crystallographic Computing Techniques*, edited by F. R. Ahmed, p. 399. Copenhagen: Munksgaard.
- International Agency for Research on Cancer (1994). *IARC Monogr. Eval. Carcinog. Risks Hum.* **61**, 1–241.
- Kim, H. S., Kim, J., Im, H. N., An, D. R., Lee, M., Heseck, D., Mobashery, S., Kim, J. Y., Cho, K., Yoon, H. J., Han, B. W., Lee, B. I. & Suh, S. W. (2014). *Acta Cryst.* **D70**, 2800–2812.
- Knott, G. D. (1979). *Comput. Programs Biomed.* **10**, 271–280.
- Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. (2001). *J. Mol. Biol.* **305**, 567–580.
- Kusters, J. G., van Vliet, A. H. & Kuipers, E. J. (2006). *Clin. Microbiol. Rev.* **19**, 449–490.
- Laue, T. M., Shah, B., Ridgeway, T. M. & Pelletier, S. L. (1992). *Analytical Ultracentrifugation in Biochemistry and Polymer Science*, edited by S. E. Harding, A. J. Rowe & J. C. Horton, pp. 90–125. Cambridge: The Royal Society of Chemistry.
- Lee, S.-Y., Lee, J.-H., Chang, H.-J., Cho, J. M., Jung, J.-W. & Lee, W. (1999). *Biochemistry*, **38**, 2340–2346.
- Lertsethtakarn, P., Ottemann, K. M. & Hendrixson, D. R. (2011). *Annu. Rev. Microbiol.* **65**, 389–410.
- Malfetheriner, P., Selgrad, M. & Bornschein, J. (2012). *Curr. Opin. Gastroenterol.* **28**, 608–614.
- Meroueh, S. O., Bencze, K. Z., Heseck, D., Lee, M., Fisher, J. F., Stemmler, T. L. & Mobashery, S. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 4404–4409.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.
- Odintsov, S. G., Sabala, I., Marcyjaniak, M. & Bochtler, M. (2004). *J. Mol. Biol.* **335**, 775–785.
- Otero, L. H., Rojas-Altuve, A., Llarrull, L. I., Carrasco-López, C., Kumarasiri, M., Lastochkin, E., Fishovitz, J., Dawley, M., Heseck, D., Lee, M., Johnson, J. W., Fisher, J. F., Chang, M., Mobashery, S. & Hermoso, J. A. (2013). *Proc. Natl Acad. Sci. USA*, **110**, 16808–16813.
- Ottemann, K. M. & Lowenthal, A. C. (2002). *Infect. Immun.* **70**, 1984–1990.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Ragumani, S., Kumaran, D., Burley, S. K. & Swaminathan, S. (2008). *Proteins*, **72**, 1096–1103.
- Roesler, B. M., Rabelo-Gonçalves, E. M. & Zeitune, J. M. (2014). *Clin. Med. Insights Gastroenterol.* **7**, 9–17.
- Roubaud Baudron, C., Franceschi, F., Salles, N. & Gasbarrini, A. (2013). *Helicobacter*, **18**, 44–51.
- Sabala, I., Jagielska, E., Bardelang, P. T., Czapinska, H., Dahms, S. O., Sharpe, J. A., James, R., Than, M. E., Thomas, N. R. & Bochtler, M. (2014). *FEBS J.* **18**, 4112–4122.
- Scheffers, D. J. & Pinho, M. G. (2005). *Microbiol. Mol. Biol. Rev.* **69**, 585–607.
- Schreiber, S., Konradt, M., Groll, C., Scheid, P., Hanauer, G., Werling, H. O., Josenhans, C. & Suerbaum, S. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 5024–5029.
- Spencer, J., Murphy, L. M., Connors, R., Sessions, R. B. & Gamblin, S. J. (2010). *J. Mol. Biol.* **396**, 908–923.
- Sycuro, L. K., Pincus, Z., Gutierrez, K. D., Biboy, J., Stern, C. A., Vollmer, W. & Salama, N. R. (2010). *Cell*, **141**, 822–833.
- Sycuro, L. K., Rule, C. S., Petersen, T. W., Wyckoff, T. J., Sessler, T., Nagarkar, D. B., Khalid, F., Pincus, Z., Biboy, J., Vollmer, W. & Salama, N. R. (2013). *Mol. Microbiol.* **90**, 869–883.
- Sycuro, L. K., Wyckoff, T. J., Biboy, J., Born, P., Pincus, Z., Vollmer, W. & Salama, N. R. (2012). *PLoS Pathog.* **8**, e1002603.
- Ten Eyck, L. F. (1973). *Acta Cryst.* **A29**, 183–191.
- Terwilliger, T. C. (2003). *Acta Cryst.* **D59**, 38–44.
- Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* **D66**, 22–25.
- Vollmer, W. & Bertsche, U. (2008). *Biochim. Biophys. Acta*, **1778**, 1714–1734.
- Vollmer, W., Blanot, D. & De Pedro, M. A. (2008). *FEMS Microbiol. Rev.* **32**, 149–167.
- Vollmer, W., Joris, B., Charlier, P. & Foster, S. (2008). *FEMS Microbiol. Rev.* **32**, 259–286.
- Wang, X., Yang, X., Yang, C., Wu, Z., Xu, H. & Shen, Y. (2011). *PLoS One*, **6**, e26845.
- Worku, M. L., Sidebotham, R. L., Walker, M. M., Keshavarz, T. & Karim, Q. N. (1999). *Microbiology*, **145**, 2803–2811.
- Wyckoff, T. J., Taylor, J. A. & Salama, N. R. (2012). *Trends Microbiol.* **20**, 540–547.
- Yang, D. C., Tan, K., Joachimiak, A. & Bernhardt, T. G. (2012). *Mol. Microbiol.* **85**, 768–781.
- Young, K. D. (2007). *Curr. Opin. Microbiol.* **10**, 596–600.