Ligand-bound structures of 3-deoxy-D-*manno*-octulosonate 8-phosphate phosphatase from *Moraxella catarrhalis* reveal a water channel connecting to the active site for the second step of catalysis

Sonali Dhindwal,[a] Priyanka Priyadarshini,[b] Dipak N. Patil,[b] Satya Tapas,[b] Pramod Kumar,[b] Shailly Tomar[b] and Pravindra Kumar[b]*

[a]Department of Biotechnology, Indian Institute of Technology Roorkee, Roorkee, Uttarakhand 247 667, India, and [b]Department of Biotechnology, Indian Institute of Technology Roorkee, Roorkee, Haridwar, Uttarakhand 247 667, India

Correspondence e-mail: pravshai@gmail.com

KdsC, the third enzyme of the 3-deoxy-D-*manno*-octulosonic acid (KDO) biosynthetic pathway, catalyzes a substrate-specific reaction to hydrolyze 3-deoxy-D-*manno*-octulosonate 8-phosphate to generate a molecule of KDO and phosphate. KdsC is a phosphatase that belongs to the C0 subfamily of the HAD superfamily. To understand the molecular basis for the substrate specificity of this tetrameric enzyme, the crystal structures of KdsC from *Moraxella catarrhalis* (*Mc*-KdsC) with several combinations of ligands, namely metal ion, citrate and products, were determined. Various transition states of the enzyme have been captured in these crystal forms. The ligand-free and ligand-bound crystal forms reveal that the binding of ligands does not cause any specific conformational changes in the active site. However, the electron-density maps clearly showed that the conformation of KDO as a substrate is different from the conformation adopted by KDO when it binds as a cleaved product. Furthermore, structural evidence for the existence of an intersubunit tunnel has been reported for the first time in the C0 subfamily of enzymes. A role for this tunnel in transferring water molecules from the interior of the tetrameric structure to the active-site cleft has been proposed. At the active site, water molecules are required for the formation of a water bridge that participates as a proton shuttle during the second step of the two-step phosphoryl-transfer reaction. In addition, as the KDO biosynthesis pathway is a potential antibacterial target, pharmacophore-based virtual screening was employed to identify inhibitor molecules for the *Mc*-KdsC enzyme.

1. Introduction

3-Deoxy-D-*manno*-octulosonate (KDO), an eight-carbon carbohydrate molecule, is an essential component of the lipid polysaccharide (LPS) that constitutes the outer membrane in Gram-negative bacteria (Wang & Quinn, 2010). It is crucial for the maintenance of membrane integrity and cell viability (Raetz & Whitfield, 2002; Raetz, 1990; Belunis *et al.*, 1995). Any blockage of or damage to the KDO biosynthetic pathway would affect the outer-membrane assembly process owing to an incomplete LPS structure and will result in the formation of a highly permeable outer membrane. The enzymes involved in the KDO biosynthetic pathway have recently gained attention as potential antibiotic targets for the development of novel drug molecules (Baasov & Belakhov, 2000).

The enzyme KdsC (3-deoxy-D-*manno*-octulosonate 8-phosphate phosphatase) is the third enzyme in the KDO

research papers

biosynthesis process. This enzyme catalyzes a very specific reaction: the hydrolysis of 3-deoxy-D-*manno*-octulosonate 8-phosphate (KDO8P) to release KDO and inorganic phosphate (Ray & Benedict, 1980). The KdsC enzyme belongs to the HAD (haloacid dehalogenase) superfamily, which is one of the largest and most ubiquitous superfamilies. It is a magnesium-dependent phosphatase characterized by an amino-terminal conserved Asp as the nucleophile (Selengut & Levine, 2000). The HAD catalytic scaffold has evolved to catalyze diverse phosphoryl-transfer reactions that are involved in various cellular functions that range from basic metabolic functions such as signal transduction and membrane transport to highly specialized functions such as nucleic acid repair (Allen & Dunaway-Mariano, 2004). These enzymes exhibit the Rossmann fold consisting of an α/β core catalytic domain formed by central parallel β-sheets flanked by α-helices on both sides. The active site is formed by four loops containing highly conserved sequence motifs (I–IV) arranged to form a cleft for substrate and metal binding (Burroughs *et al.*, 2006). The core catalytic domain is unable to ward off solvent from entering the active site and also lacks substrate-recognition capability. To seal off the active site from solvent, the HAD superfamily enzymes have evolved another domain known as the cap domain that is inserted into the core Rossmann fold. Depending on the absence or presence as well as the location of the cap domain, HAD superfamily enzymes have been classified into three subfamilies: C0, C1 and C2 (Burroughs *et al.*, 2006). The KdsC enzyme belongs to the C0 subfamily that lacks the cap domain. The other enzymes which belong to this subfamily are T4 polynucleotide 5′-kinase/3′-phosphatase, magnesium-dependent phosphatase 1 (MDP-1) and bacterial class B phosphatase (Galburt *et al.*, 2002; Selengut, 2001; Calderone *et al.*, 2004). To exclude solvent, KdsC enzymes exist in a tetrameric form held together by the extended β-sheets that form a β-barrel structure at the centre. Each monomer forms a cap over the active site of the adjacent monomer to ward off water and also to assist in substrate recognition.

The catalytic reaction takes place in two steps *via* the formation of a covalently bound enzyme–substrate intermediate linked to the invariant Asp in the active site. In the first step, the phosphate group attacks the active-site residue Asp, resulting in the formation of a phosphoaspartate intermediate and generation of the KDO molecule. In the second step, an activated water molecule hydrolyzes this aspartyl-phosphate intermediate, regenerating the enzyme and releasing the phosphate group, as shown in Fig. 1. To date, several crystal structures of KdsC enzymes have been determined from various species such as *Escherichia coli*, *Haemophilus influenzae*, *Aquifex aeolicus*, *Yersinia pestis*, *Vibrio cholerae*, *Legionella pneumophila*, *Bacteroides thetaiotaomicron etc.* (Biswas *et al.*, 2009; Parsons *et al.*, 2002; Daughtry *et al.*, 2013). However, the migration of the solvent molecule to the active site during the second step of the catalytic reaction is largely unknown and to our knowledge has never been discussed before for the C0 class of the HAD superfamily.

Moraxella catarrhalis is a Gram-negative, aerobic, diplococcus bacterium formerly known as *Neisseria catarrhalis*. It is an emerging human respiratory-tract pathogen and causes acute otitis media in children. It is the second most common cause of exacerbations in adults with chronic obstructive pulmonary disease (COPD), which is associated with enormous morbidity and healthcare costs (Verduin *et al.*, 2002; Murphy *et al.*, 2005). Over the last two decades, it has emerged as a significant bacterial pathogen towards humans (Verduin *et al.*, 2002). Owing to an increase in antibiotic resistance among various species of *Moraxella*, there is a need to study alternative pathways for the development of new antibiotics against *M. catarrhalis*. As the KDO biosynthesis pathway is essential for the synthesis of the outer membrane in Gram-negative bacteria, a structure-based approach for the rational design of inhibitor molecules targeting the enzymes of this
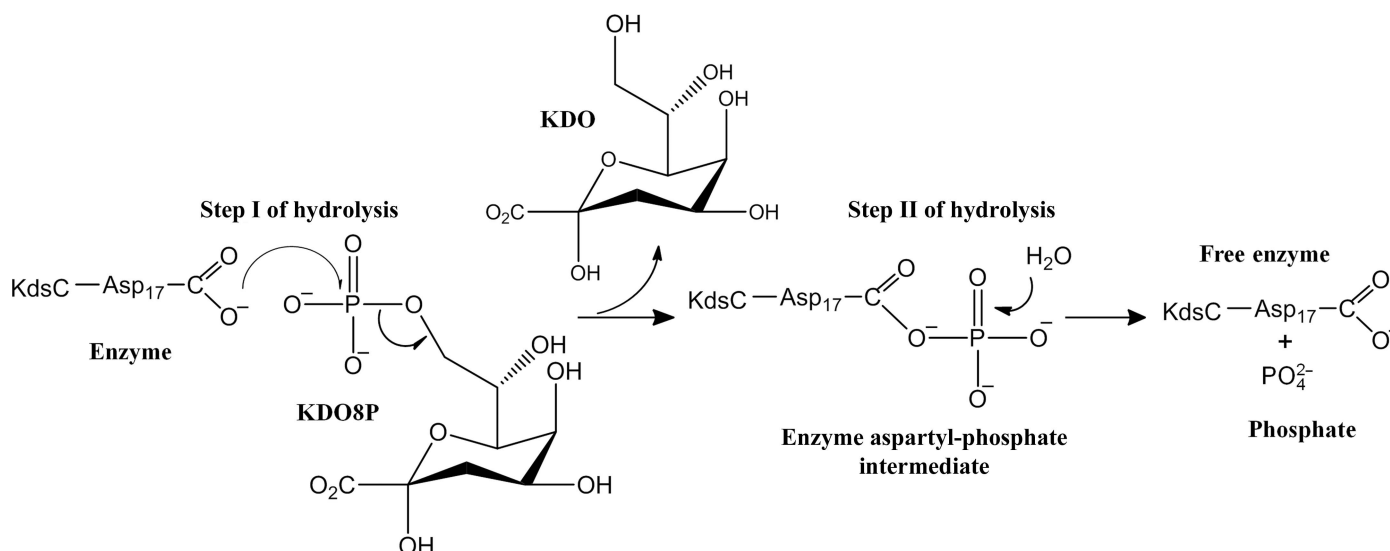

Figure 1
Schematic representation of the two-step reaction catalyzed by the enzyme KdsC.

pathway would provide novel antibiotics and a way to prevent *M. catarrhalis* infections (Cipolla *et al.*, 2009, 2010).

In the work presented here, we have successfully cloned, expressed, purified and crystallized 3-deoxy-D-*manno*-octulosonate 8-phosphate phosphatase (KdsC) from *M. catarrhalis* (*Mc*-KdsC). We have determined the three-dimensional structure of KdsC with various combinations of metal ligand, citrate and product molecules. A detailed structural characterization provides insight into the binding mode of these ligand molecules at the active site. We have identified the conformational differences in the KDO molecule during its binding as a substrate (KDO8P) and in the cleaved form. In addition, we have structurally elucidated for the first time the existence of a tunnel at the subunit interface that connects the bulk solvent in the interior of the tetrameric enzyme to the active site. We propose here that this tunnel could be involved in the process of transferring the water molecule to the active site which is required in the second step of phosphoaspartate hydrolysis.

Finally, as *Mc*-KdsC is a potential drug target and with the availability of the crystal structures of various enzyme states determined in this study, pharmacophore mapping and virtual screening was carried out to identify putative inhibitors of *Mc*-KdsC.

2. Materials and methods

2.1. General reagents

An Axygen Biosciences DNA-isolation kit was used for isolating genomic DNA. A MiniPrep plasmid-isolation kit from Qiagen, USA was used for plasmid isolation. Ni–NTA beads were purchased from Qiagen, USA. The HiLoad Superdex 75 16/60 column and ÄKTApurifier system for protein purification was purchased from GE Healthcare. The product analogue 3-deoxy-D-*manno*-octulosonic acid ammonium salt (KDO) was purchased from Sigma–Aldrich, India.

2.2. Amplification and cloning of the KdsC gene

M. catarrhalis was obtained from MTCC, IMTECH, Chandigarh, India and revived in nutrient broth media at 25°C. The genomic DNA was isolated and the gene for KdsC was cloned from the genomic DNA using the forward primer AGCTGCGC**CATATG**AATGAAATTTACCAAAAAGCC containing an NdeI restriction-enzyme site (bold) and the reverse primer GAATCAGTG**CTCGAG**CTACTGAAAT-GTTGCAATG containing a stop codon and an XhoI restriction-enzyme site (bold). The primers were designed on the basis of GenBank accession No. EGE20456.1. PCR amplification was performed using 35 cycles of 1 min at 367 K, 1.5 min at 331 K and 1.3 min at 345 K. The resulting PCR product of 500 bp was purified using a PCR purification kit according to the manufacturer's instructions (Qiagen, USA) to remove unincorporated nucleotides and primers. The purified PCR fragment and pET-28c vector were digested with NdeI and XhoI restriction enzymes. The digested DNA fragments were separated by 1% agarose gel electrophoresis,

purified using a DNA gel-extraction kit (Qiagen, USA) and were ligated with T4 DNA ligase. The ligated product was used for transformation of competent DH5α cells by the heat-shock method (Inoue *et al.*, 1990). Some of the colonies obtained by plating the transformed cells on Luria–Bertani (LB) agar plates containing 50 µg ml^{-1} kanamycin were picked and grown overnight. Plasmids were isolated and screened for the presence of the KdsC gene by gene amplification and restriction-enzyme digestion. The plasmid containing the correct-sized insert was further confirmed by sequencing in both directions using T7 promoter and T7 terminator primers.

2.3. Expression of the KdsC protein

The cloned plasmid containing the KdsC-encoding DNA fragment was transformed into *E. coli* BL21 (DE3) cells for the production of N-terminally His-tagged recombinant protein. LB broth supplemented with 50 µg ml^{-1} kanamycin at 310 K was used to grow the bacterial culture to an optical density of 0.8 at 600 nm (OD$_{600}$). Expression was then induced using 0.2 m*M* isopropyl β-D-1-thiogalactopyranoside (IPTG) and the culture was allowed to grow for 4 h at 310 K after induction. Finally, the cells were harvested by centrifugation at 5000*g* and 277 K and the pellets were stored at 253 K until further use.

2.4. Purification of *Mc*-KdsC

The cell pellet from 500 ml culture was resuspended in lysis buffer (100 m*M* Tris buffer pH 7.5, 1.2 *M* NaCl, 7% glycerol) and lysed using a French press (Constant Systems Ltd, Daventry, England). The cell lysate was centrifuged at 15 000*g* for 45 min at 277 K. The supernatant was incubated with Ni–NTA beads pre-equilibrated with purification buffer (100 m*M* Tris buffer pH 7.5, 1.2 *M* NaCl, 7% glycerol) at 277 K. The incubated beads were then loaded onto a gravity-flow column and the flowthrough was collected. The incubated beads were then washed with a step gradient of 40, 80 and 100 m*M* imidazole in purification buffer to remove any loosely bound protein impurities. The recombinant KdsC protein was finally eluted with 250 m*M* imidazole in purification buffer (100 m*M* Tris–HCl pH 7.5, 1.2 *M* NaCl). All fractions were run on SDS–PAGE. Fractions containing purified protein were pooled and concentrated to ∼45 mg ml^{-1}. The protein was further purified to remove any impurities by gel-filtration chromatography using a pre-equilibrated HiLoad Superdex 75 16/60 column (GE Healthcare) and an ÄKTApurifier system (GE Healthcare) operated at 277 K at a flow rate of 0.5 ml min^{-1} in 100 m*M* Tris–HCl pH 7.5. The size-exclusion column was calibrated using a low-molecular-weight calibration kit consisting of bovine serum albumin (66 kDa), ovalbumin (45 kDa), trypsin (23 kDa) and lysozyme (14 kDa) for determination of the void volume, construction of the standard curve and estimation of the molecular weight of the purified protein. The gel-filtration eluate was collected in 2 ml fractions and the purity of the fractions was analyzed by Coomassie Blue-stained SDS–PAGE. The fractions containing

pure protein were pooled and concentrated to ∼42 mg ml^{-1} using an Amicon Ultra-15 concentrator with a cutoff of 10 kDa (Millipore, Bedford, Massachusetts, USA). The concentration of the purified protein was determined by UV absorbance spectroscopy at 280 nm using a calculated extinction coefficient of 9065 M^{-1} cm^{-1}.

2.5. Crystallization

Crystals were grown by the sitting-drop vapour-diffusion method using crystallization screens from Hampton Research, California, USA. A protein concentration of 20 mg ml^{-1} in 100 mM Tris–HCl pH 7.5 was used for crystallization. The protein and the well solution were used in a 1:1 ratio and were equilibrated against 50 µl reservoir buffer. The crystal-growth conditions were monitored initially and the hits obtained in the screens were further optimized. We were able to obtain good-quality diffracting crystals in a condition consisting of 100 mM magnesium formate, 100 mM bis-tris propane pH 5.0, 28%(w/v) PEG 3350 (crystal form I). Another crystal form (crystal form III) was grown in a condition consisting of 100 mM citric acid, 100 mM bis-tris propane pH 5.0, 28%(w/v) PEG 3350 at 293 K.

2.6. Data collection and structure determination

Diffraction data were collected from various crystal forms at the home source with a MAR345 imaging-plate detector at a wavelength of 1.54 Å using Cu $K\alpha$ radiation generated by a Bruker–Nonius Microstar H rotating-anode generator operated at 45 kV and 60 mA. Prior to data collection, the protein crystals were cryoprotected by direct transfer into a mother-liquor drop containing 20% glycerol as a cryoprotectant. Crystals were then mounted on cryo-loops over the goniometer. The data were collected under cryogenic conditions (∼100 K) with the crystal dipped in the cryostream.

The data for all crystal forms were indexed, integrated and scaled using the *HKL*-2000 program suite (Otwinowski & Minor, 1997). Initial phases were obtained by the molecular-replacement method using *MOLREP* from the *CCP*4 suite v.6.3 (Vagin & Teplyakov, 2010; Winn *et al.*, 2011). The crystal structure of KdsC from *E. coli* (PDB entry 2r8y; Biswas *et al.*, 2009) was used as the search model. *REFMAC*5.2 was utilized for rigid-body refinement, which was followed by iterative cycles of restrained atomic parameter refinement (Murshudov *et al.*, 2011). *Coot* was utilized for model building and electron-density map analysis (Emsley & Cowtan, 2004). Several cycles of rigid-body refinement and restrained refinement were used to achieve acceptable R_{cryst} and R_{free} values. *CNS* was utilized for the generation of a composite OMIT map of the ligand molecules (Brünger *et al.*, 1998). The stereochemistry of the models was assessed using *MolProbity* (Chen *et al.*, 2010).

2.7. Pharmacophore model generation and virtual screening

The docked pose of KDO8P in the active site was considered to be the biological conformation of the substrate. It was utilized for building the pharmacophore model using chemical feature-based and shape-based hypotheses.

For the chemical feature-based hypothesis, *LigandScout*, an automated tool for pharmacophore generation, was utilized. *LigandScout* identifies the interaction between the protein and the ligand. It extracts and interprets interactions such as hydrogen bonds, hydrophobic bonds and charge transfer between the protein and the ligand (Wolber & Langer, 2005). It also identifies inaccessible areas for the potential ligand by defining the excluded volume spheres. The docked complex of the enzyme and the KDO8P was provided as the input to the *LigandScout* software and the interaction between the ligand and the amino acid provided sufficient input for generation of the pharmacophore hypothesis. The generated pharmacophore was exported in the 'hypoedit' format and then converted to 'chm' format using the hypoedit tool in the *Accelrys Discovery Studio* software (Accelrys Software Inc.).

The shape-based query was generated with the shape query module of *Accelrys Discovery Studio* utilizing the docked conformation of KDO8P. The chemical feature-based hypothesis and the shape-based query were then merged to give a single three-dimensional pharmacophore model. This pharmacophore model was utilized as a three-dimensional query to screen the NCI database of diverse chemical compounds (consisting of 87 373 commercially available compounds). For this purpose, the virtual screening program of *Accelrys Discovery Studio* was used. Virtual screening was performed using the *BEST* method for generation of the ligand conformations. The maximum omitted features was set to '−1' to screen the database. The best mapped compounds were sorted according to their fitness value and the shape-similarity scores, and were further evaluated for their drug-like properties using Lipinski's rule of five.

2.8. Molecular docking

The molecular docking of the substrate KDO8P and the ligands obtained as hits from the virtual screening was carried out using the *Maestro* suite (v.9.1) from Schrödinger (Friesner *et al.*, 2004). For this purpose, the crystal structure of *Mc*-KdsC in complex with KDO and Mg^{2+} was utilized. In the first step, the protein was prepared using the *Maestro* protein-preparation wizard by addition of hydrogen, assignment of bond orders and minimization of the protein using the OPLS2001 force field until the root-mean-square (r.m.s.) deviation between the minimized structure and the starting structure reached 0.3 Å. In the second step, the substrate KDO8P and all of the ligands obtained by virtual screening were prepared using the *LigPrep* module in *Maestro*. Next, a receptor grid for docking was generated using the bound KDO molecule as the centroid. Finally, *Glide* (v.2.6; Schrödinger) was utilized for docking of the substrate and the screened ligand molecules using the extra precision (XP) mode. The XP mode avoids false positives from the set of screened compounds as it exacts severe penalties on poses that violate established physical chemistry principles (Friesner *et al.*, 2004; Halgren *et al.*, 2004). All of the docked compounds were screened on the basis of their dock score, binding mode and molecular interactions in the active site.

2.9. Protein analysis and visualization

PyMOL was used for structure visualization and to generate the figures (DeLano, 2002). *CAVER* was used for the analysis and visualization of tunnels in *Mc*-KdsC and other KdsC structures. *CAVER* incorporates an algorithm to find and characterize the potential paths that can be utilized by a small molecule to travel to and from a buried site of a protein. These tunnels connect a buried cavity of a protein to the bulk solvent (Chovancova *et al.*, 2012). For this analysis, the two subunits of KdsC were selected and given as input to find any probable path or tunnel. The tunnel search was performed utilizing a probe of 0.9 Å radius; all of the other parameters were kept at default values.

3. Results and discussion

3.1. Overall structure of *Mc*-KdsC

To understand the substrate-specific interactions, various enzyme–ligand complexes of *Mc*-KdsC have been determined and analyzed. The overall structure remains the same in all of the crystal forms. *Mc*-KdsC exists as a tetramer as shown in Fig. 2(*a*) and as also confirmed by gel-filtration chromatography. Each monomeric unit exhibits the classic Rossmann fold that forms a core catalytic domain in which six-stranded parallel β-sheets are flanked by six α-helices, three on each side. This architecture creates four loops that constitute four highly conserved motifs characteristic of the HAD super-family (Fig. 2*b*). Motif I consists of two aspartate residues: Asp17 and Asp19 in *Mc*-KdsC. Asp17 is involved in the formation of the aspartyl-phosphate intermediate during the first step of catalysis, whereas Asp19 acts as an acid that helps in protonation of the substrate leaving group in the first step

and acts as a base by deprotonating the nucleophile in the second step (Burroughs *et al.*, 2006; Herschlag & Jencks, 1989). Motif II has a conserved Thr residue at position 61 that binds the phosphoryl group of the substrate. Motif III has a conserved Lys at position 87 that interacts with Asp17 and a phosphate group. Lys87 shields the negative charge of the reaction intermediates during catalysis (Burroughs *et al.*, 2006). Both motifs II and III provide stability to the reaction intermediates formed during catalysis. Asp110 from motif IV along with motif I is involved in coordination of the Mg^{2+} ion (Burroughs *et al.*, 2006; Herschlag & Jencks, 1989).

As mentioned earlier, KdsC belongs to the C0 subfamily of the HAD superfamily owing to the absence of the cap domain. The cap domain evolved in the HAD superfamily to protect the active site from bulk solvent. In the absence of the cap domain, the C0 subfamily of enzymes have evolved a strategy for solvent exclusion at the active site by oligomerizing the four subunits into a tetramer *via* a β-hairpin formed by two elongated β-strands that project from the core domain (Burroughs *et al.*, 2006). This results in the formation of a central eight-stranded intramolecular β-barrel. The active site is present at the interface of the two neighbouring subunits, resulting in the formation of four active sites per tetramer (Fig. 2*a*). The catalytic scaffold is formed by the core domain of one monomer, and a cap-like structure (similar to the cap domain in the other subfamilies of the HAD superfamily) is formed by the β-hairpin structure of the adjacent monomer. This cap-like structure prevents bulk solvent from entering the active site.

3.2. Different crystal forms of KdsC

Three-dimensional structures of the following crystal forms of *Mc*-KdsC with several combinations of metal ion, KDO and
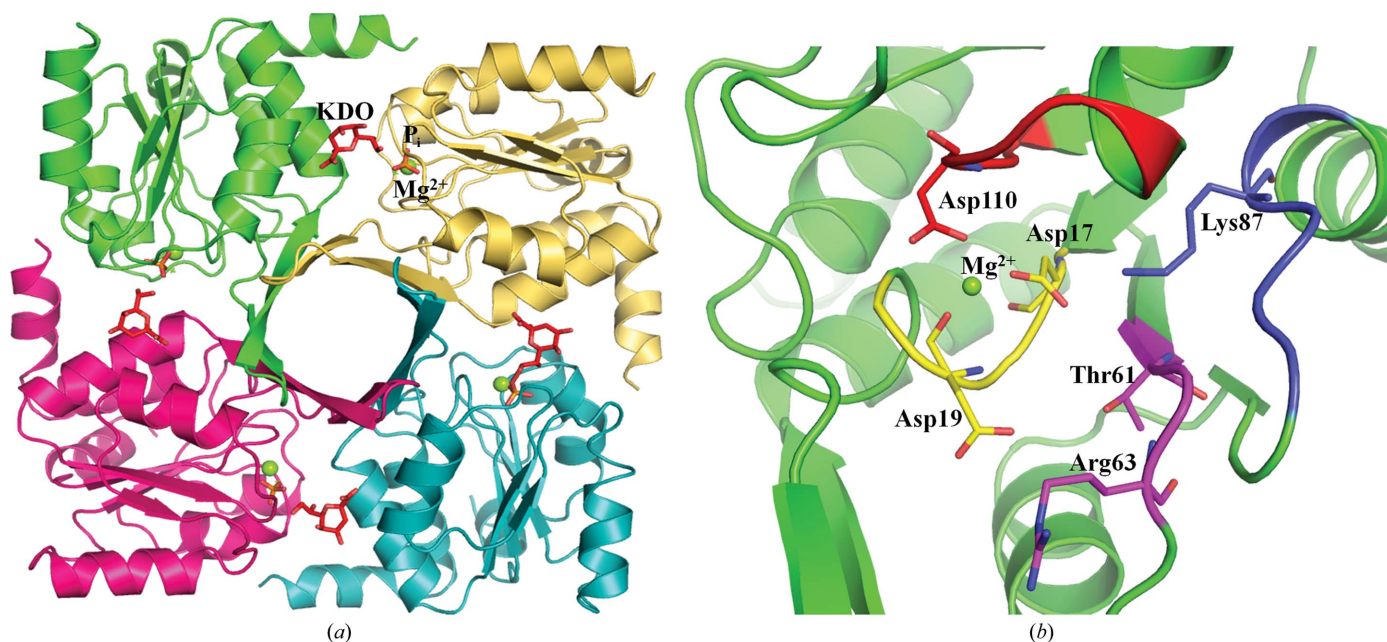

Figure 2
(*a*) Tetrameric view of *Mc*-KdsC (crystal form V) displaying KDO, phosphate ion and Mg^{2+} ion bound in the active site at the interface of two subunits. (*b*) Active-site residues of *Mc*-KdsC highlighting four motifs in different colours. The green sphere represents the Mg^{2+} ion.

Table 1
Data-collection and refinement statistics for *Mc*-KdsC in crystal forms I, II, III, IV and V.

Values in parentheses are for the last shell.

	Form I (KdsC + Mg^{2+})	Form II (KdsC + Mg^{2+} + phosphate)	Form III (KdsC + citrate)	Form IV (KdsC + Mg^{2+} + KDO)	Form V (KdsC + Mg^{2+} + KDO + phosphate)
PDB code	4um7	4um5	4umd	4ume	4umf
Data collection					
Space group	$P2_1$	$P2_1$	$I4$	$I4$	$P2_1$
Unit-cell parameters					
a (Å)	86.26	87.80	86.39	86.61	86.91
b (Å)	43.56	45.77	86.39	86.61	46.05
c (Å)	87.55	87.80	40.80	41.40	87.22
β (°)	91.18	90.32	90.00	90.00	90.25
Wavelength (Å)	1.54	1.54	1.54	1.54	1.54
Resolution (Å)	1.64	2.34	2.30	2.10	2.30
R_{merge}† (%)	4.8 (43.0)	5.8 (37.9)	8.9 (36.0)	5.2 (41.0)	5.0 (43.0)
$\langle I/\sigma(I)\rangle$	22.3 (2.3)	23.0 (2.1)	9.4 (2.1)	7.6 (4.4)	17.1 (2.0)
Completeness (%)	99.8 (99.5)	91.6 (70.0)	9.0 (60.0)	96.2 (77.4)	90.7 (65.0)
Unique reflections	92882	26909	6587	9166	33990
Multiplicity	3.9 (3.6)	3.2 (2.7)	3.2 (2.3)	4.2 (3.8)	3.6 (3.2)
Wilson B factor (Å^2)	15.6	41.9	11.9	8.0	43.9
Refinement					
Resolution range (Å)	27.16–1.64	31.68–2.34	43.23–2.29	37.38–2.09	31.69–2.28
No. of reflections	76193	25542	6019	8733	27641
R_{cryst} (%)	19.6	18.8	17.6	19.0	21.4
R_{free} (%)	22.8	24.5	26.7	25.8	28.0
No. of residues	696	691	173	174	693
No. of molecules in asymmetric unit	4	4	1	1	4
No. of water molecules	731	86	66	85	170
Average B factors (Å^2)					
Chain A	19.31	55.16	36.16	18.80	66.40
Chain B	19.62	53.12	—	—	61.65
Chain C	21.07	59.48	—	—	66.10
Chain D	18.73	59.22	—	—	70.40
Water atoms	32.16	48.58	32.77	25.20	49.71
Ligands					
Mg^{2+}	A, 17.9; B, 15.6; C, 19.5; D, 19.3	A, 29.8; B, 39.6; C, 38.2; D, 42.2		17.5	A, 40.0; B, 51.1; C, 42.2; D, 39.8
Phosphate ion		A, 61.4; B, 67.2; C, 64.4; D, 87.7			A, 49.8; B, 57.8; C, 60.4; D, 61.0
KDO				33.5	A, 52.5; B, 57.18; C, 60.6; D, 60.9
Citrate			43.7		
All atoms	21.40	56.7	36.08	19.36	65.5
R.m.s. deviations					
Bond lengths (Å)	0.004	0.007	0.008	0.07	0.009
Bond angles (°)	0.91	1.51	1.32	1.30	1.62
Ramachandran plot (%)					
Favoured	99.2	96.6	98.2	98.8	96.5
Allowed	100.0	99.6	100.0	100.0	99.4
Outliers	0.0	3	0.0	0.0	4

† $R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl} \sum_i I_i(hkl)$.

phosphate ion were determined: (I) KdsC bound to Mg^{2+} ion, (II) KdsC bound to Mg^{2+} and phosphate ion, (III) KdsC with citrate bound in the active site, (IV) KdsC bound to Mg^{2+} and a KDO molecule and (V) KdsC bound to Mg^{2+} ion, phosphate ion and a KDO molecule. The crystallographic data-collection and refinement statistics for all of the crystal forms are summarized in Table 1.

The structure of crystal form I was determined at a resolution of 1.64 Å. It belonged to space group $P2_1$, with unit-cell parameters $a = 86.26$, $b = 43.56$, $c = 87.55$ Å, $\beta = 91.18°$. It has four monomers in the asymmetric unit and each active site contains an Mg^{2+} ion. The Mg^{2+} ion forms an octahedral geometry by coordination to the side chains of Asp17 and

Asp110, the main-chain O atom of Asp19 and three water molecules, as shown in Fig. 3(a).

The structure of crystal form II was determined at 2.34 Å resolution. It belonged to space group $P2_1$, with unit-cell parameters $a = 87.80$, $b = 45.77$, $c = 87.80$ Å, $\beta = 90.32°$. It was obtained by soaking crystal form I with 200 mM sodium phosphate buffer. In this crystal structure, electron density for Mg^{2+} ion and phosphate ion were present in all of the active sites of the tetramer. Comparing crystal form II with crystal form I, the Mg^{2+} ion coordinates to the enzyme in a similar fashion as in crystal form I, except that one of the water molecules which was coordinated to the Mg^{2+} ion in crystal form I was replaced by the O atom of the phosphate ion, thus

maintaining its overall geometry, as shown in Fig. 3(b). The binding of the phosphate ion replaces three other water molecules which were initially bound in the active site in crystal form I. The O atoms of the phosphate ion make hydrogen bonds to the side chains of Asp17, Thr61 and Lys87, the side chain and main chain of Asp19, the main chain of Gly62 and two water molecules that are coordinated to the Mg^{2+} ion.

The structure of crystal form III was determined at 2.3 Å resolution. It belonged to space group $I4$, with unit-cell parameters $a = 86.39$, $b = 86.39$, $c = 40.80$ Å. It was obtained by growing Mc-KdsC crystals in the presence of citrate in the crystallization buffer. In this crystal form, only one monomer was present in the asymmetric unit and it forms a tetramer with symmetry-related molecules. There is clear electron density for the citrate molecule in the active site. The carboxyl group of one arm of the citrate fills the binding site for the scissile phosphate (the position of phosphate in crystal form II), in which it hydrogen-bonds to the side chain of Thr61 and

Lys87 and the main chain of Val18, Aps19 and Gly62, as shown in Fig. 3(c). The O6 and C4 atoms of the citrate occupy the position that was occupied by the two water molecules coordinated to the Mg^{2+} ion in crystal form I.

The structure of crystal form IV was determined at 2.1 Å resolution. It belonged to space group $I4$, with unit-cell parameters $a = 86.61$, $b = 81.61$, $c = 41.40$ Å. The complex was obtained by soaking Mc-KdsC crystal form I with the reaction product KDO in the cryoprotectant solution. This crystal form has one molecule in the asymmetric unit and forms a tetramer with other symmetry-related molecules. The presence of KDO was distinctly visible in the difference Fourier map along with density for the magnesium ion bound in the active site. In another experiment, crystal form I was soaked in reservoir solution with the reaction product KDO and sodium phosphate buffer to give crystal form V. The structure of crystal form V was determined at 2.3 Å resolution. It belonged to space group $P2_1$, with unit-cell parameters $a = 86.91$, $b = 46.05$, $c = 87.22$ Å, $\beta = 90.25°$. Crystal form V clearly showed electron

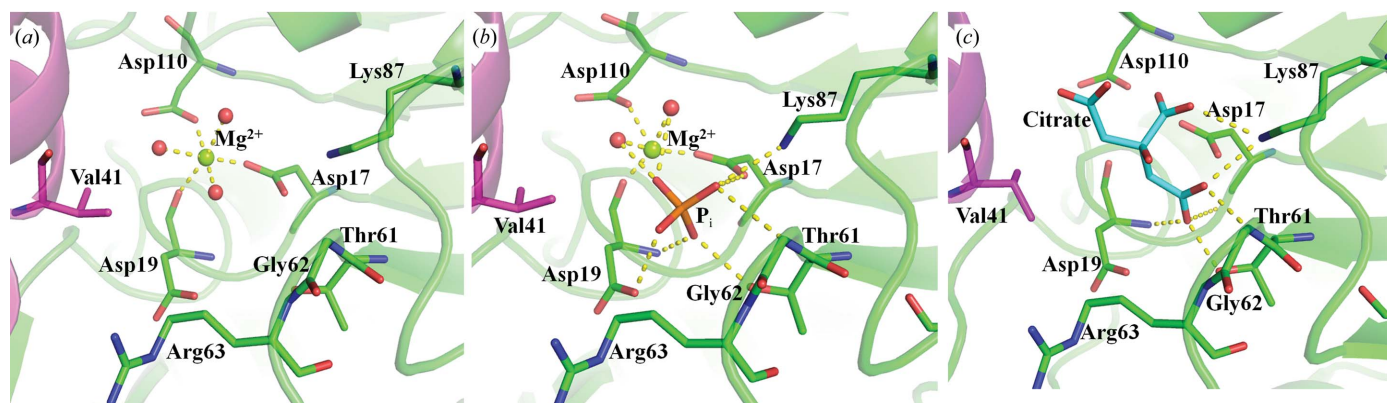**Figure 3**
Structures of Mc-KdsC: (a) crystal form I with Mg^{2+} ion, (b) crystal form II with phosphate and Mg^{2+} ion and (c) crystal form III with a citrate molecule bound in the active site. The green sphere represents the Mg^{2+} ion and the red spheres represent water molecules. Hydrogen bonds are shown as yellow dashed lines. The two subunits are coloured green and magenta.
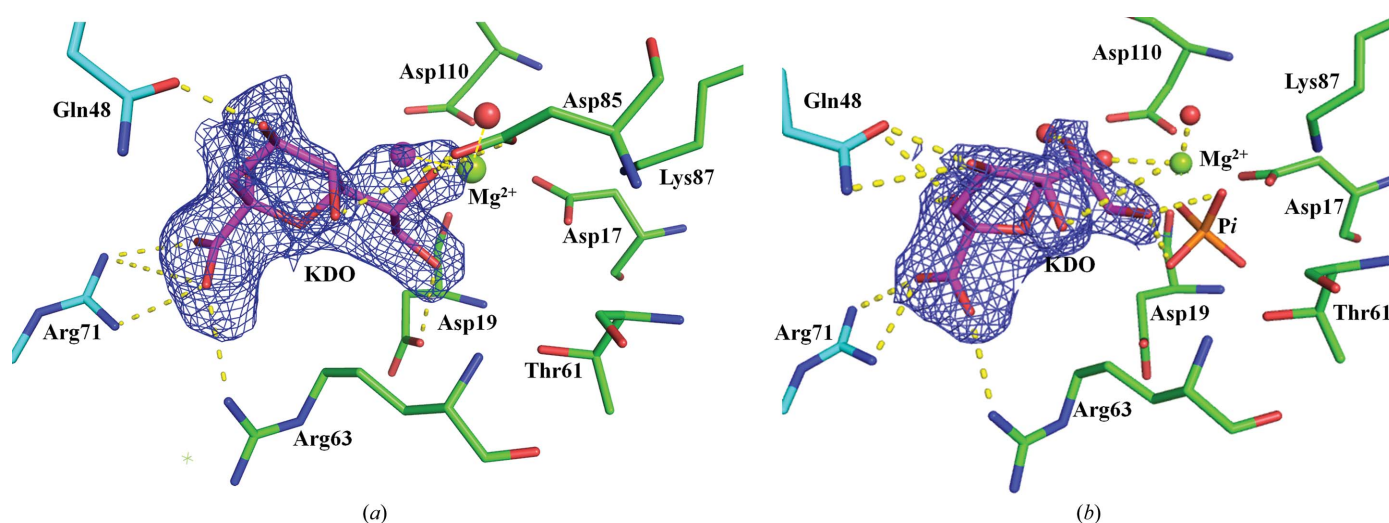

Figure 4
Structure of Mc-KdsC: calculated composite OMIT map showing $2F_o - F_c$ electron-density maps contoured at 1σ for (a) KDO in crystal form IV and (b) KDO with phosphate bound in crystal form V. The green sphere represents Mg^{2+} ion and the red spheres represent water molecules coordinated to the Mg^{2+} ion. Hydrogen bonds are shown as yellow dashed lines. The residues from two monomers are differentiated in green and cyan.

density for KDO, phosphate ion and Mg^{2+} ion in the active site in all of the monomers. The presence of KDO in crystal forms IV and V was confirmed by generating an unbiased composite OMIT map using *CNS* (Figs. 4*a* and 4*b*; Brünger *et al.*, 1998).

In the tetrameric state of crystal form IV, the KDO is coordinated by three water molecules and forms nine other hydrogen bonds to residues from both core- and cap-forming monomers. The side chain of Asp19 interacts with the O8 atom of KDO, the side chain of Asp85 interacts with the O7 and O5 atoms of KDO, and the side chain of Arg63 forms a hydrogen bond to the O1A atom of the KDO molecule. From the adjacent monomer, the side chain of Gln48 interacts with the O4 atom of KDO and Arg71 interacts with the O1A and O1B atoms of KDO, as shown in Fig. 4(*a*). The O2 atom of KDO interacts with the main chain of Val41. Comparing crystal form IV with crystal form I shows that the overall structure does not display a major change owing to KDO binding. KDO binds well at the active site, replacing a few water molecules which were present in crystal form I. There are only a few local changes observed in the side-chain conformations of a few amino acids such as Gln48, Arg63 and Asp85 to bind the KDO molecule at the active site. Superposition of *Mc*-KdsC crystal form IV (KdsC + Mg^{2+} + KDO) on chain *A* of crystal form II (KdsC + Mg^{2+} + phosphate) gives an r.m.s.d. of 0.38 Å. The superposed crystal forms show that the O8 atom of KDO (from crystal form IV) overlaps with the O1 atom of the phosphate ion (from crystal form II). These together mimic the substrate binding in the transition state during the first step of the phosphoryl-transfer reaction (Fig. 5*a*).

Superposing crystal form IV on chain *A* of crystal form V gives an r.m.s. value of 0.30 Å. Both crystal forms display a similar overall structure. The KDO molecule in both structures superposes well at all atoms except for the alcoholic moiety at the C6 atom of KDO, which adopts a different conformation. In the superposed view of these structures, as shown in Fig. 5(*b*), the O8 atom of KDO in crystal form IV

is present at a distance of 0.6 Å from the O1 atom of the phosphate ion in crystal form V. Therefore, to prevent a short contact between the C8 and O8 atoms of KDO with the O atom of the phosphate ion, the alcoholic substituent at the C6 atom of KDO changes its conformation in crystal form V compared with crystal form IV. The O8 atom of KDO in crystal form V now makes hydrogen bonds to two O atoms of the phosphate ion (Fig. 4*b*). Crystal form V apparently represents the structure with both of the product molecules formed after the second step of hydrolysis, KDO and phosphate ion (leaving groups), bound at the active site.

3.3. Presence of the tunnel

As mentioned above, the four monomers in *Mc*-KdsC assemble together by their elongated β-hairpins, forming an eight-stranded β-barrel at the centre of the tetramer. All of the crystal forms of *Mc*-KdsC exhibit a narrowing of the aperture at the top of the centrally placed β-barrel, where the side chain of Glu32 seems to govern its closing and opening, as shown in Fig. 6(*a*). Superposing these crystal forms shows a difference in the side-chain orientation of Glu32 (Fig. 6*a*). In crystal form III (KdsC + citrate) the side chain of Glu32 is oriented in such a fashion as to completely seal off the opening of the barrel (as seen in the surface view of the tetramer formed by symmetry-related molecules; Figs. 6*b* and 6*c*), whereas in the phosphate-, Mg^{2+}- and KDO-bound crystal forms I, II, IV and V the β-barrel does not seem to be closed and the side chain of Glu32 provides an opening at the top of the barrel (Figs. 6*d* and 6*e*). The size of the opening varies in all of these crystal forms, as seen in the surface view. It is largest in crystal forms II and V, with distances of 8.5 and 9.5 Å (when measured from the side chains of Glu32 of the monomers placed diagonally in the tetramer), compared with 6.5 and 6.3 Å in crystal forms I and IV, respectively. In crystal form I the side chain of Glu32 is very flexible and shows alternate conformations. It appears that the β-barrel is most
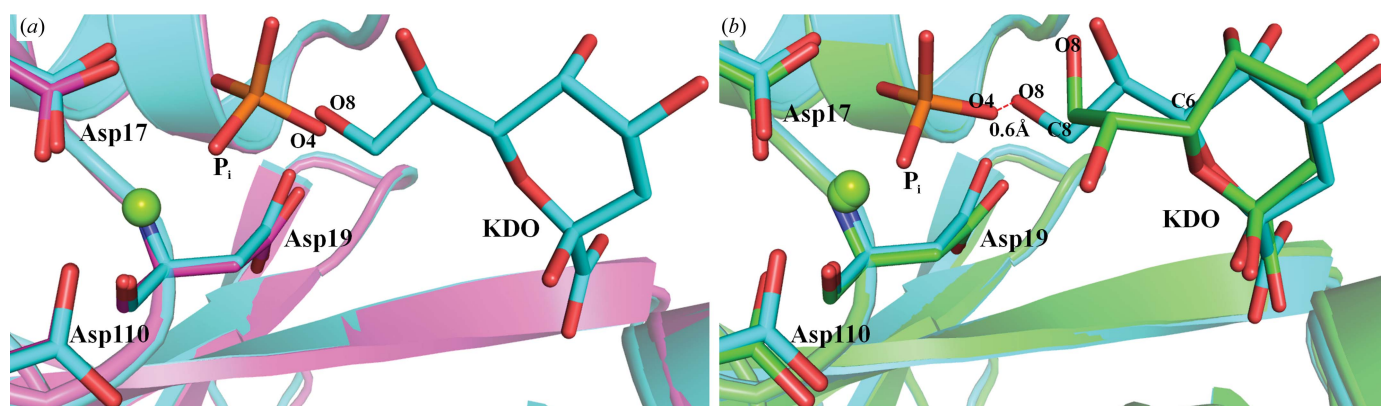

Figure 5
(*a*) Superposed view of *Mc*-KdsC crystal forms II (magenta) and IV (cyan), showing the arrangement of KDO and phosphate ion in the active site mimicking substrate binding in the transition state during the first step of the phosphoryl-transfer reaction. (*b*) Superposed view of *Mc*-KdsC crystal forms IV (cyan) and V (green) to highlight the difference in the conformation of the alcoholic moiety of KDO in the crystal forms. The green sphere represents Mg^{2+} ion.

open in the presence of phosphate, as in crystal forms I, II, IV and V, compared with crystal form III in which the nonphysiological substrate citrate is present at the active site.

The bottom of the β-barrel opens to a wide space formed by the core domains of the four monomers. Overall, the enzyme looks like a shallow inverted funnel in which the tetrameric arrangement forms the wide neck and the β-barrel structure at the centre forms the small stem of the funnel. As in a funnel, water can pass from the interior surface of the protein through the β-barrel shielding the active site located at the interface of two subunits from the bulk solvent. In all of the crystal forms of Mc-KdsC electron density for water molecules was found to be present throughout the barrel. In crystal forms IV and V, Mg^{2+} ion and phosphate ion, respectively, were also found to occupy the centre of the β-barrel. These ligand molecules are present at a position roughly equivalent to those of MES [2-(N-morpholino)ethanesulfonic acid] buffer and a glycerol molecule in crystal structures of $H.$ $influenzae$ KdsC (Hi-KdsC; Parsons et $al.$, 2002). However, the presence of these molecules was proposed to be insignificant biologically as no tunnel connecting the interior of the barrel to the active site was observed (Parsons et $al.$, 2002).

Interestingly, in the surface view of crystal form V (Mc-KdsC + phosphate + Mg^{2+} + KDO), a tunnel-like structure is evident at the interface of the two monomers (Fig. 7a). Further investigation of the presence of a channel was performed using $CAVER$ (Chovancova et $al.$, 2012). Interestingly, the $CAVER$ results showed the presence of a tunnel at the interface of the two subunits that connects the water-rich cavity of the inverted funnel-shaped KdsC structure to the active-site cleft, as shown in Fig. 7(a). The tunnel starts from the bottom of the β-barrel, spans a distance of up to 8.4 Å inside the protein and opens near the metal ion, the main chain of Asp19 and the side chain of Asp110 in the active site (Fig. 7b). The tunnel is lined by several hydrophilic residues that provide a polar environment suitable for the movement of water molecules. These residues are Asp24, Gly25, Asn129 and Gly146 from one subunit and Tyr40, Val41 and Gln42 from the neighbouring subunit (Fig. 8). Additionally, these residues are mostly conserved among KdsCs from other species, suggesting a potential role of these residues in facilitating the approach of solvent to the active site (Fig. 8).

This tunnel is present in all of the Mc-KdsC crystal forms, but the largest sized tunnel is observed in crystal form V as
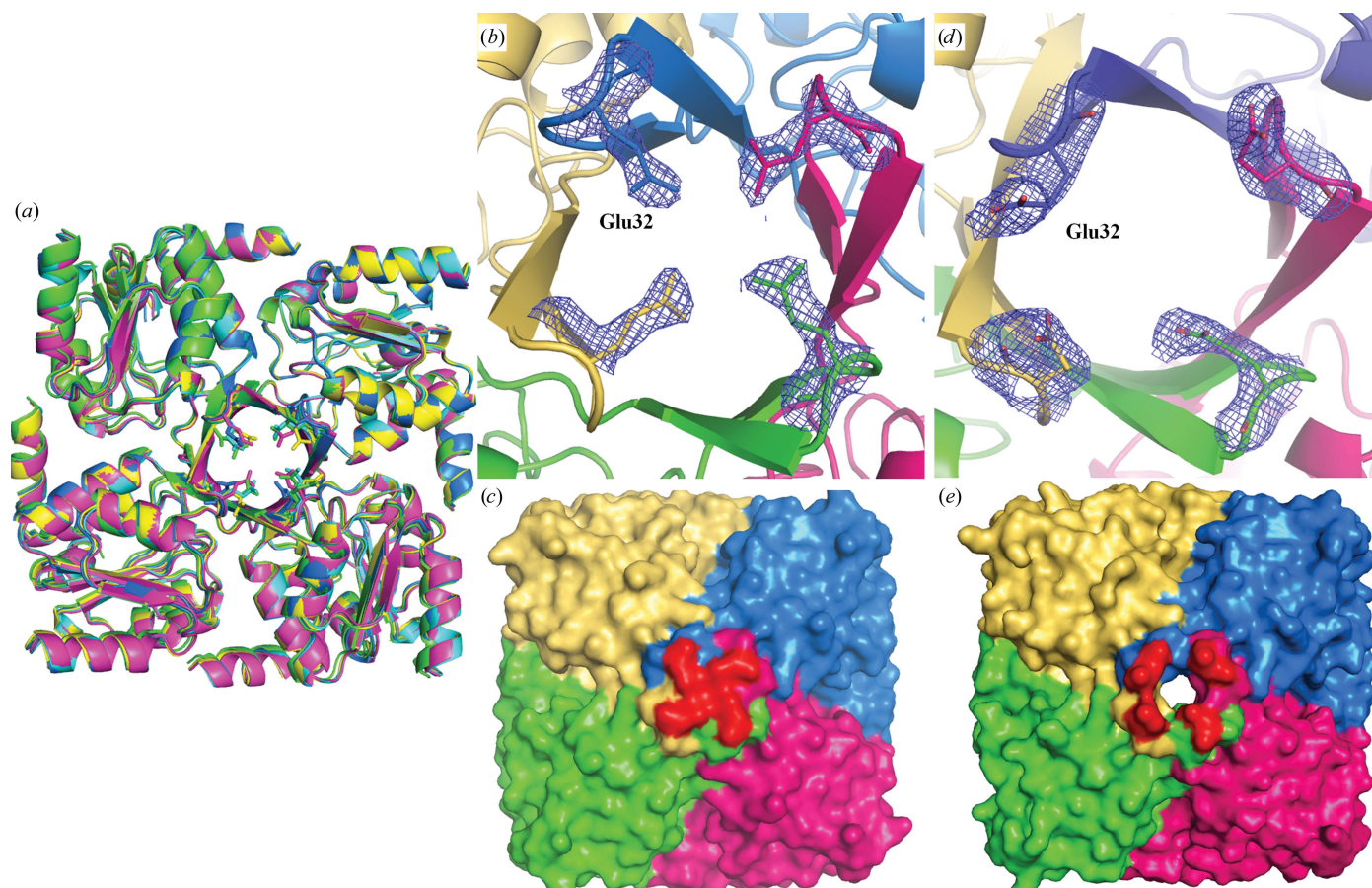

Figure 6
(a) Cartoon view of the tetrameric form of Mc-KdsC crystal forms I, II, III, IV and V in green, blue, pink, cyan and yellow, respectively, superposed on each other. The difference in conformation of Glu32 (shown in stick form) in the different crystal forms is shown. (b, d) An enlarged view of the top of the β-barrel in Mc-KdsC for (b) crystal form III and (d) crystal form V, showing the electron-density map for the residue Glu32. (c, e) Surface view of the tetramer in (c) crystal form III and (e) crystal form V, highlighting Glu32 in red at the top of the β-barrel. All subunits are coloured differently.
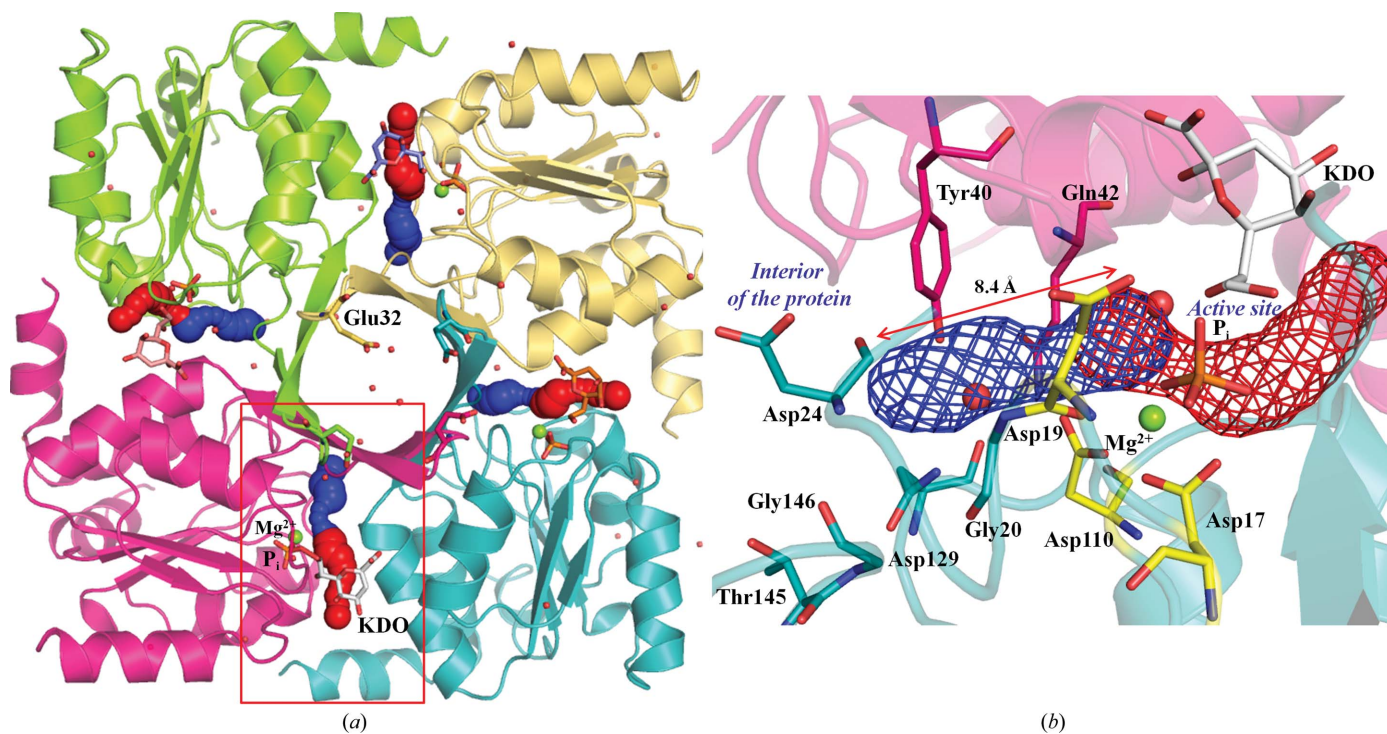
Figure 7
(*a*) Cartoon view of the tetrameric form of crystal form V, displaying the tunnel as connected spheres at the interface of two subunits. The tunnel in blue starts from the interior of the protein, connecting to the active-site region with red spheres. The square box highlights one of the subunit interfaces. (*b*) An enlarged view of the highlighted subunit interface displaying the tunnel as a mesh. The residues that line the tunnel are shown in stick form. The residues from the active site are coloured yellow and KDO is coloured white and shown in stick form. The arrow represents the distance of the tunnel from the interior of the protein to the distal end of the active site. The green sphere represents an Mg^{2+} ion and the red spheres represent water molecules.
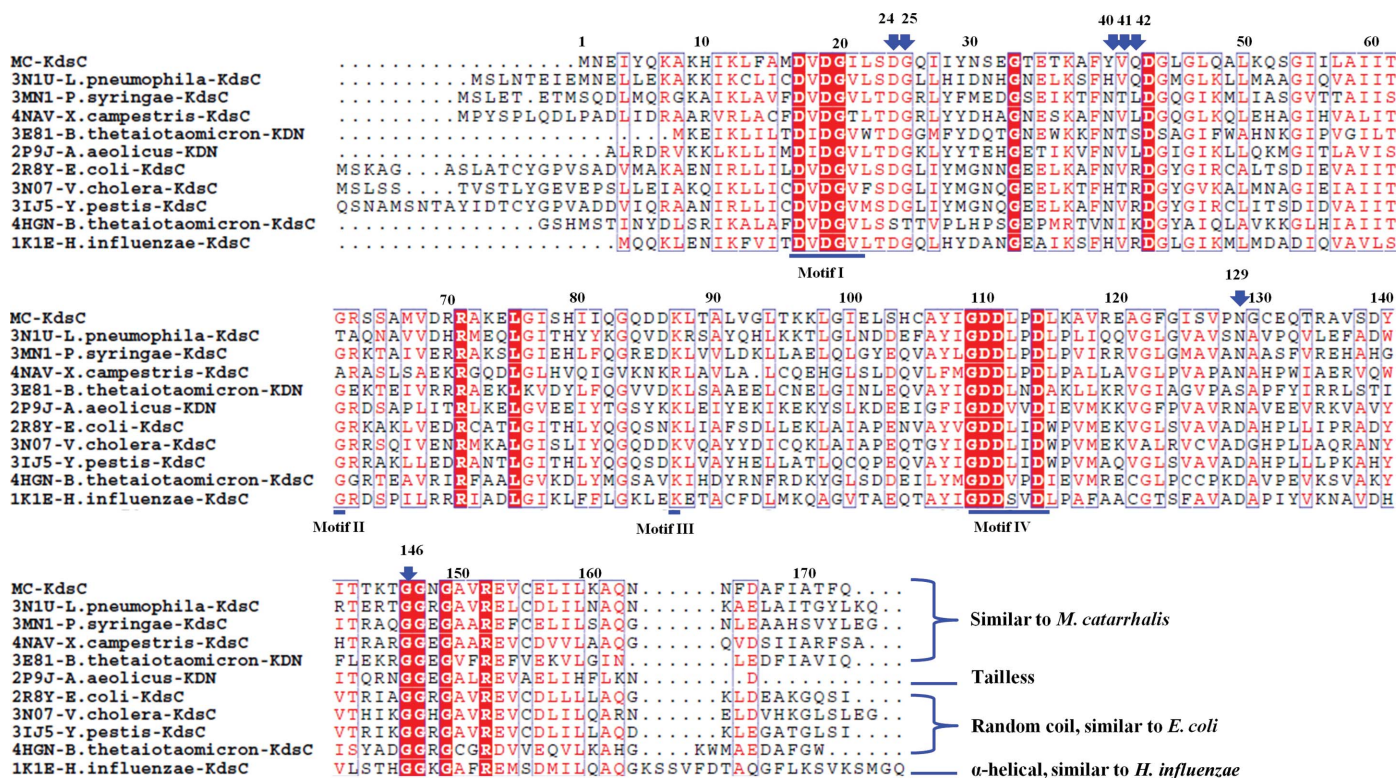
Figure 8
A multiple sequence alignment of *Mc*-KdsC and homologous KdsCs for which three-dimensional crystal structures are known. The blue arrows highlight the residues that line the tunnel. The four motifs are indicated by blue lines. The types of C-terminal conformation formed by these KdsCs are shown at the end of the alignment.

quantified by *CAVER*. Among the different *Mc*-KdsC crystal forms, a difference in conformation was found for the side chain of Asn129 that lies at the entrance of this tunnel towards the interior of the protein. In crystal forms I, III and IV the side chain of Asn129 points towards the interface of the two monomers, but in crystal forms II and V it has flipped 135° inwards and binds a water molecule.

The existence of a similar tunnel in the crystal structures of all of the other KdsC structures in the PDB was also explored. Interestingly, a tunnel similar to the intersubunit tunnel present in the *Mc*-KdsC structure was found to be present in all of the homologous structures, as shown in Fig. 9. The presence of this tunnel in all of the crystal structures of KdsC indicates that it has a significant role.
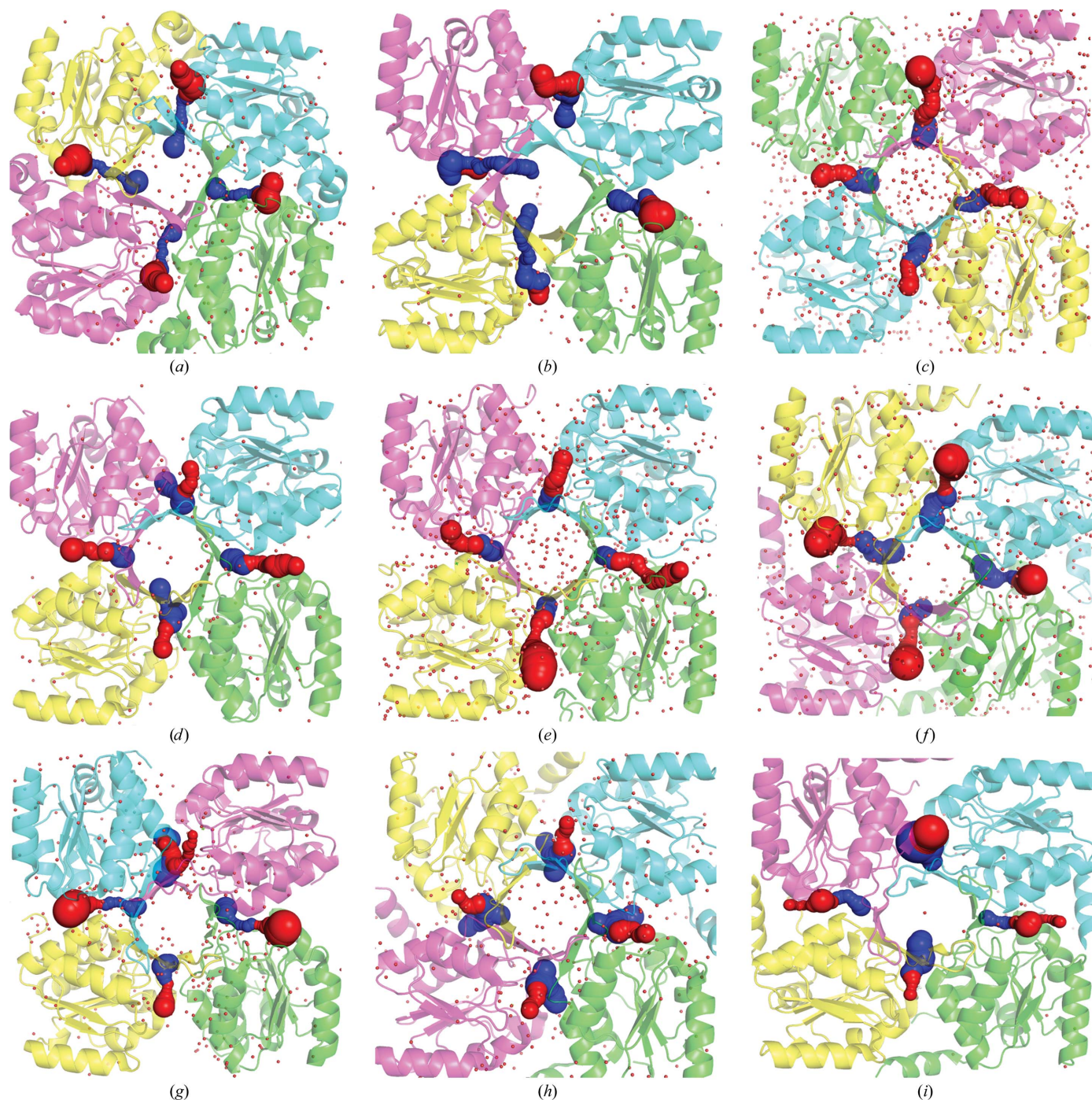
Figure 9
The tunnel at the interface of two subunits in KdsC from (*a*) *H. influenzae* (PDB entry 1j8d), (*b*) *A. aeolicus* (PDB entry 2p9j), (*c*) *E. coli* (PDB entry 2r8e), (*d*) *E. coli* (PDB entry 3i6b), (*e*) *Y. pestis* (PDB entry 3ij5), (*f*) *P. syringae* (PDB entry 3mn1), (*g*) *V. cholerae* (PDB entry 3n07), (*h*) *B. thetaiotaomicron* (KDN9PP; PDB entry 4hgq) and (*i*) *X. campestris* (PDB entry 4nav). The tunnel is shown as connecting spheres. The blue-coloured tunnel starts from the interior of the protein, connecting to the active-site region by red-coloured spheres. The small red spheres represent water molecules.

Dhindwal *et al.* • 3-Deoxy-D-*manno*-octulosonate 8-phosphate phosphatase

As mentioned earlier, together the tetrameric assembly of KdsC and the bound substrate prevent the solvent from entering the active site. However, the second step of catalysis requires a water molecule for the hydrolysis of the phospho-aspartyl intermediate *via* nucleophilic attack of a water molecule at the scissile P atom (Fig. 1; Burroughs *et al.*, 2006). Based on the presence of a tunnel in the structure of *Mc*-KdsC, we have hypothesized that the water molecule that is needed during the second step of the catalytic reaction could be accessed through this tunnel.

Previously, Parsons and coworkers proposed that a water molecule hydrogen-bonded to the second Asp of motif I and Ser/Thr of motif II in the *Hi*-KdsC crystal structure (Asp19 and Thr61 in *Mc*-KdsC) could be the water molecule that is utilized during catalysis (Parsons *et al.*, 2002). However, this water molecule was found to be displaced by the O atom of the phosphate ion in the *Mc*-KdsC crystal structure (crystal forms II and V). In these crystal structures, Thr61 in motif II makes interactions with the phosphate ion and thus contributes to the stability of the reaction intermediate during hydrolysis (Burroughs *et al.*, 2006).

Here, we have proposed that the tunnel present at the interface of the two monomeric units in *Mc*-KdsC could pave the way for mobilization of the water molecule to the active site. Additionally, our hypothesis for the role of this tunnel in catalysis is supported by the following evidence.

The presence of water molecules inside the tunnel in the *Mc*-KdsC crystal structures is the first indication which corroborates our hypothesis for the movement of solvent molecules in the tunnel (Fig. 7b).

Secondly, De Vivo and coworkers have performed computational studies that include classical molecular-dynamics and mixed quantum-mechanical/molecular-mechanics (QM/MM) calculations on an Mg^{2+}-dependent phosphatase from the HAD superfamily (De Vivo *et al.*, 2007). They have suggested that during the second step of catalysis, in the transition state, the enzyme promotes water deprotonation and facilitates a proton shuttle from an Mg^{2+}-coordinated water molecule to the leaving-group oxygen *via* a bridging solvent water molecule. These water-bridge-mediated proton shuttles are said to be essential for activation of the solvent nucleophile and for stabilization of the leaving group.

Interestingly, in *Mc*-KdsC crystal structure V a water molecule is evident at the mouth of the tunnel towards the active site, as shown in Fig. 10. This water molecule is equidistant (~2.6 Å) from the side chain of Asp111 (from motif IV) and the Mg^{2+}-coordinated water molecule. In the presence of a leaving group bound in the active site blocking the entry of solvent, the water molecule might have migrated to the active site through the intersubunit tunnel. This water can act as the bridging solvent molecule to take part in a proton shuttle from an Mg^{2+}-coordinated water molecule to the leaving-group oxygen and for the activation of the solvent nucleophile along with the metal ion and active-site residues Asp19, Lys87 and Asp110 (Fig. 10). Thus, the presence of a tunnel that would provide a water molecule in the active site near the metal ion and Asp19 for solvent activation is in

accordance with the structural framework that would be needed for the proton-shuttle mechanism.

Furthermore, there are several other reports that suggest and confirm the presence of channels and tunnels in enzymes that belong to the HAD superfamily, such as the crystal structure of L-2-haloacid dehalogenase from *Xanthobacter autotrophicus* (Ridder *et al.*, 1997). Ridder and coworkers proposed that this unidentified water molecule could plausibly be imported into the active site *via* the small tunnel (Ridder *et al.*, 1997). Hamid and coworkers have described the presence of a tunnel for the substrate to access the binding site in the haloalkanoic acid dehalogenase DehE from *Rhizobium* sp. RC1 (Hamid *et al.*, 2013).

Thus, structural studies of *Mc*-KdsC along with the above evidence support the hypothesis that the tunnel which is structurally evident in all of the crystal structures of KdsC available to date could provide accessibility to solvent during enzyme catalysis.

3.4. Structural comparison with other KdsC structures

In previously determined structures of KdsC from *E. coli* and *H. influenzae*, it has been reported that the C-terminus is flexible and that its extent of disorder in the 'open conformation' governs the accessibility of substrate, and that once catalysis is over it facilitates the displacement of the product from the active site (Parsons *et al.*, 2002; Biswas *et al.*, 2009; Daughtry *et al.*, 2013).

In *Mc*-KdsC, the C-terminal amino acids were not found to interact with the active-site residues in any of the crystal forms and thus are not involved in any catalytic function. The residues at the C-terminus form a helix that extends away from the main structure towards the adjacent monomer. Of the eight residues at the C-terminus, six are hydrophobic in
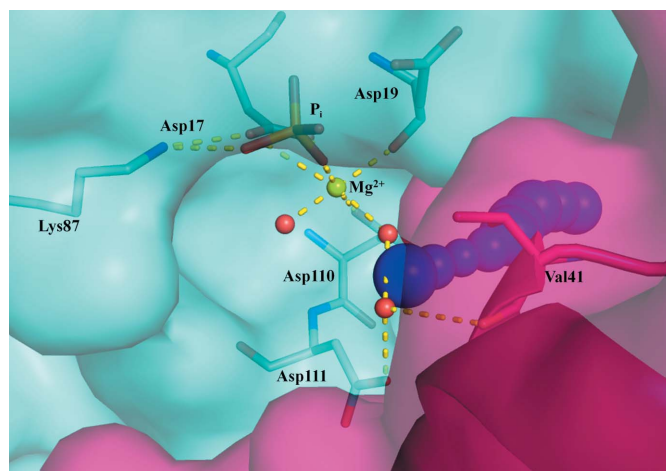

Figure 10
Figure showing the presence of a water molecule at the opening of the tunnel (blue spheres) near the active site. This water molecule is at a hydrogen-bond distance from Asp111 and from water molecules coordinated to the Mg^{2+} ion. The two subunits are shown as a surface view in cyan and magenta. Active-site residues are shown in stick form. The hydrogen bonds are shown as yellow dashed lines. The green sphere represents the Mg^{2+} ion and the red spheres represent water molecules.

nature. The main chain of Gln173 from one monomer makes a hydrogen bond to the side chain of Arg119 from the adjacent monomer. Phe172 makes hydrophobic interactions with Leu112, Pro113, Leu115 and Lys116. These interactions are present in all of the subunits in all of the crystal forms of *Mc*-KdsC.

A *BLASTP* search using the *Mc*-KdsC sequence was performed against the PDB and several homologous structures to KdsC were retrieved (Altschul *et al.*, 1990). From alignment of all of the KdsC crystal structures, it was observed that some of the structures have a C-terminal conformation similar to that of *Mc*-KdsC. The putative hydrolase from *L. pneumophila* (PDB entry 3n1u; New York SGX Research Center for Structural Genomics, unpublished work), the probable YrbI family phosphatase from *Pseudomonas syringae* pv. *phaseolica* (PDB entry 3mn1; New York SGX Research Center for Structural Genomics, unpublished work) and the hypothetical protein Xcc2798 from *Xanthomonas campestris* (PDB entry 4nav; Enzyme Function Initiative, unpublished work) display 48, 46 and 38% sequence identity to *Mc*-KdsC, respectively. Among these, the hydrolase from *L. pneumophila* and the phosphatase from *P. syringae* have been assigned as KdsCs owing to their higher k_{cat} and k_{cat}/K_m values for catalysis of KDO8P hydrolysis compared with KDN9P hydrolysis (Daughtry *et al.*, 2013). In addition, the 2-keto-3-deoxy-D-glycero-D-galacto-nonic acid 9-phosphate (KDN9P) phosphatase (KDN9PP) from *B. thetaiotaomicron* (PDB entries 3e81 and 4hgq; Lu *et al.*, 2009; Daughtry *et al.*, 2013) also displays a similar C-terminal conformation and 38% sequence identity to *Mc*-KdsC.

Yi proposed three types of structural distributions for the C-terminal region in KdsC enzymes (Yi, 2009). These are KdsCs with a short C-terminal extension forming a random coil similar to that of *Ec*-KdsC (PDB entry 2r8y; Biswas *et al.*, 2009), KdsCs with a long C-terminal extension that adopts an α-helical structure such as *Hi*-KdsC (PDB enty 1k1e; Parsons *et al.*, 2002) and a third type with no C-terminal extension such as *A. aeolicus* KdsC (PDB entry 2p9j; Southeast Collaboratory for Structural Genomics and RIKEN Structural Genomics/Proteomics Initiative, unpublished work) and *Coxiella burnetii*

KdsC (Yi, 2009), as shown in Figs. 8, 11(*a*), 11(*b*) and 11(*c*). In first and second types of KdsCs, the C-terminal region binds at the active site and plays an important role in carrying out catalysis. KdsC structures similar to that from *M. catarrhalis* are neither tailless nor does their C-terminal region interact with the active site, and they form a fourth category of this enzyme (Fig. 11*d*). This suggests that the C-terminus, which has previously been reported to play an important role in product release, thus may not be necessary for catalytic efficiency in *Mc*-KdsC.

3.5. Virtual screening

As KDO is an essential component of the lipopolysaccharide layer in Gram-negative bacteria, the enzymes involved in KDO biosynthesis present possible targets for the development of antibacterial agents (Raetz, 1990; Baasov & Belakhov, 2000; Cipolla *et al.*, 2010). With the availability of various ligand-bound crystal structures of *Mc*-KdsC, virtual screening was carried out to identify putative inhibitors. A library of molecules was screened by the pharmacophore map generated on the basis of shape and electronic features of the ligand molecules.

The substrate KDO8P was docked into the active site of Mg^{2+}-bound *Mc*-KdsC using the *Glide* program from the *Maestro* suite v.9.1 (Schrödinger; Friesner *et al.*, 2004). The docked pose of the substrate in which the phosphate and KDO of KDO8P orient in a similar conformation to the transition state obtained by superposition of crystal form II (KdsC + Mg^{2+} + phosphate) and crystal form IV (KdsC + Mg^{2+} + KDO) was taken as the biological conformation of KDO8P. This conformation was utilized to build chemical feature-based and shape-based pharmacophore models. The docked KDO8P shows the following interactions with the active-site residues: the hydroxyl groups on the ring make hydrogen bonds to the main chains of Val41 and Leu45 and to the side chains of Gln48, Arg63, Arg71 and Asp85, and the phosphate moiety interacts with the side chains of residues Asp17, Asp19, Gly62, Asp85 and Lys87.
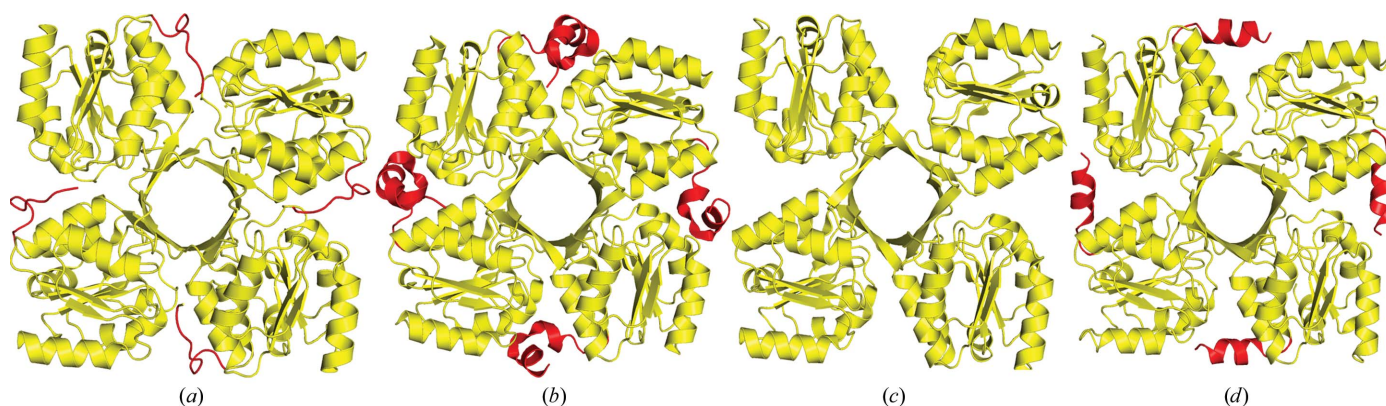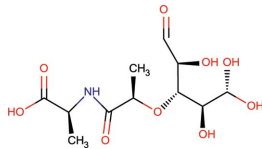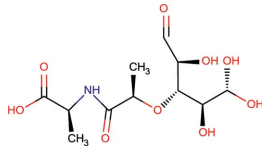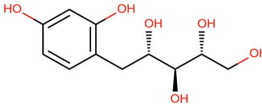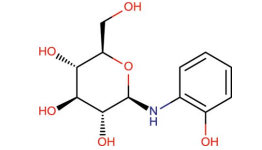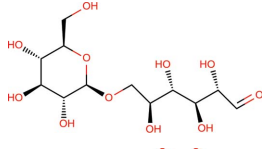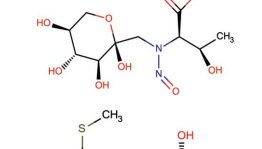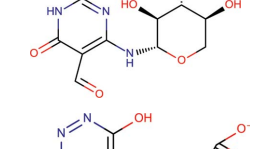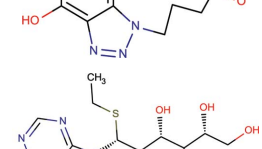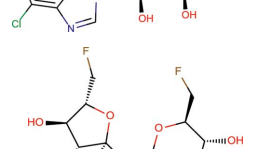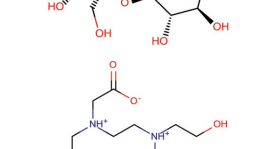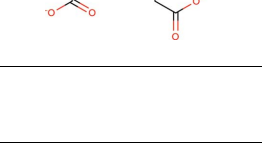

Figure 11
Cartoon view of the tetrameric form of KdsC from (*a*) *E. coli* (PDB entry 2r8e), (*b*) *H. influenzae* (PDB entry 1k1e), (*c*) *A. aeolicus* (PDB entry 2p9j) and (*d*) *M. catarrhalis* (crystal form I). The difference in the conformation of the residues at the C-terminus in these structures is highlighted in red. The green sphere represents Mg^{2+} ion and the red spheres represent water molecules.

Table 2
The list of ligand molecules which were obtained as hits after virtual screening along with their FitValue, shape similarity, docking score and Emodel values.

Zinc ID	FitValue	Shape similarity	Docking score	*Glide* Emodel	Chemical structure
ZINC16941528	3.50153	0.51	−13.76	−93.89	
ZINC05374450	3.45119	0.6	−10.82	−83.43	
ZINC16450480	3.52948	0.57	−9.80	−76.14	
ZINC16889850	3.22509	0.54	−9.70	−108.25	
ZINC17002694	3.04667	0.576	−9.45	−84.17	
ZINC16958004	3.70632	0.52	−9.30	−85.60	
ZINC16952454	3.31269	0.6	−9.17	−72.62	
ZINC16958837	3.33141	0.53	−9.12	−60.81	
ZINC06001431	3.51916	0.6	−9.00	−84.31	
ZINC16941397	3.12219	0.54	−8.74	−71.43	
ZINC19364242	3.12282	0.6	−7.80	−73.32	

The substrate-binding site has more electronegative charge where the phosphate moiety binds, whereas it has electropositive charge around the KDO-binding site. The area and volume of the substrate-binding cavity as calculated by the *CASTp* server (Dundas *et al.*, 2006) were estimated to be around 326.8 Å^2 and 500.0 Å^3, respectively.

The built pharmacophore model consisted of a spatial arrangement of eight chemical features and a shape feature for the bound conformation of KDO8P. The eight chemical features are two hydrogen bonds donated by two hydroxyl groups of the KDO moiety pointing towards Asp85 and two negatively ionizable groups formed by the carboxyl group of the KDO moiety and the phosphoryl group, which also contribute to four hydrogen-bond acceptors pointing towards Asp19, Gly62, Arg63 and Lys87 from one subunit and Arg71 from the neighbouring subunit. This is shown in Supplementary Figs. S1(*a*) and S1(*b*).

The generated three-dimensional model was used as a query to search the National Cancer Institute (NCI) database, which is comprised of 87 373 commercially available compounds. A total of 538 molecules were mapped and these were reduced to 520 molecules when screened for drug-like properties by applying Lipinski's rule. The screened molecules were then sorted on the basis of FitValue and similarity index. All molecules showing a FitValue of greater than 3.0 were selected. The hits were found to superimpose well with the three-dimensional pharmacophore model as shown by their fitness values, as shown in Table 2. Furthermore, molecular docking of these selected molecules was performed to remove any false-positive hits.

The identified top-scoring potential inhibitors are given in Table 2 along with their corresponding ZINC IDs and the scoring values. The docked conformations of these molecules, showing the hydrogen bonding to the active-site residues, are shown in Fig. 12. A molecular overlay of these hits with the biological substrate KDO8P revealed their similar binding orientations at the active site. Almost all of the reported hits show a hydrogen-bond acceptor feature pointing towards Asp19, Gly62, Arg63 and Lys87, which

suggests that this feature is important for interaction with the enzyme. Similarly, all of the hits have a hydrogen-bond donor feature pointing towards Asp85. This suggest that the binding interaction patterns observed during docking studies were complementary to that of the pharmacophoric hydrogen-bond acceptor, hydrogen-bond donor and negatively ionizable features.

It is noteworthy that the citrate molecule bound at the active site of crystal form III occupies the corresponding position to the scissile phosphate group and the metal ion.

Interestingly, it occupies the position of the Mg^{2+} ion which is essential for catalytic activity. This indicates that the citrate molecule may act as an inhibitor of this enzyme. Citrate has also been reported in the literature to act as an inhibitor of a number of enzymes such as pyrophosphatases, phosphatases and phosphotransferases (Evered & Steenson, 1964; Nordlie & Lygre, 1966). Additionally, the pharmacophore map was generated for the bound citrate in order to compare it with that of KDO8P. Similar to KDO8P, it exhibits two negatively ionizable groups formed by its two carboxylic arms pointing
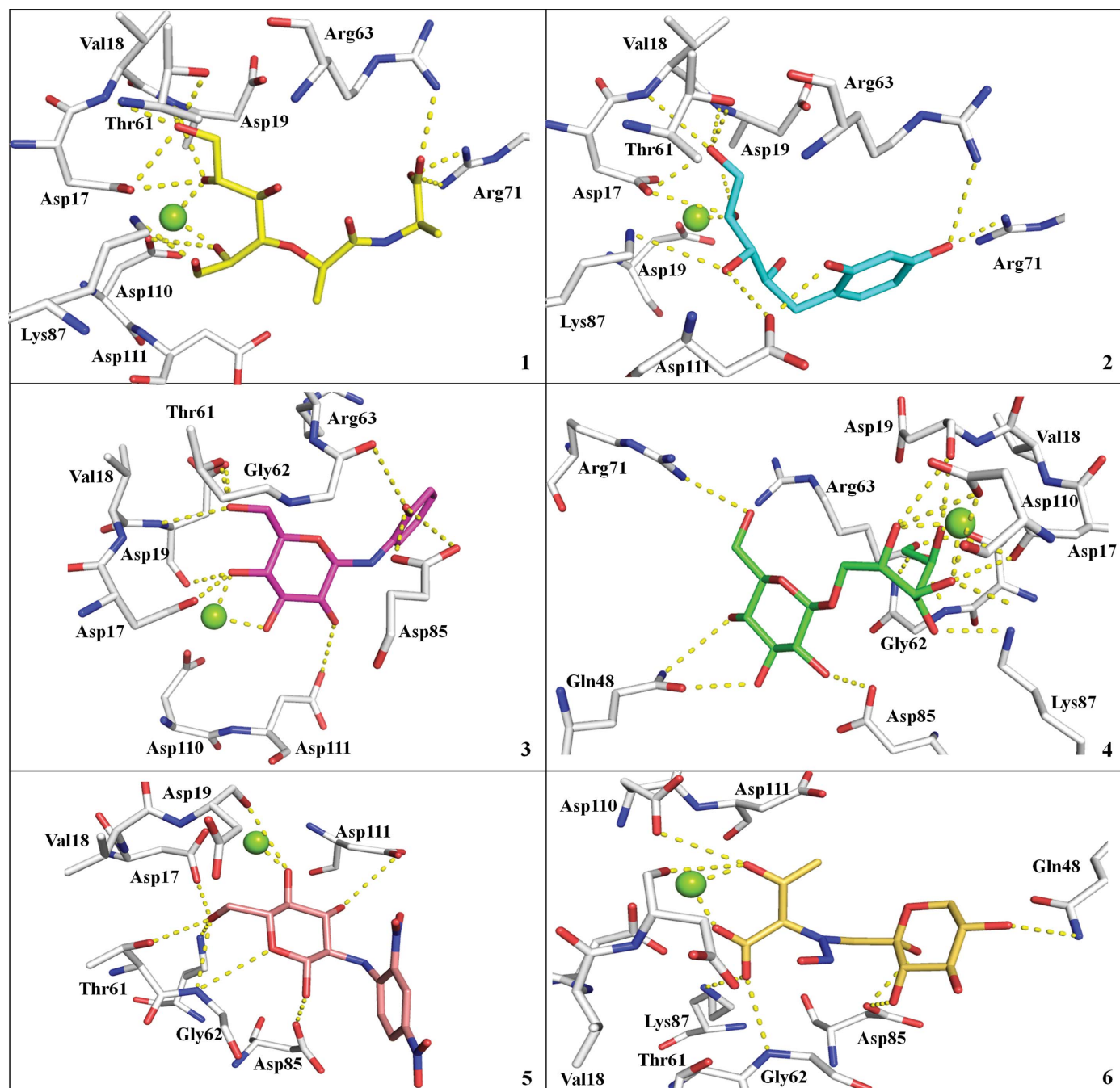

Figure 12
Docked views of the top six ligands obtained by virtual screening are shown. The yellow dashed lines highlight the hydrogen bond that the ligand makes to the enzyme. The green sphere represents the Mg^{2+} ion.

towards Thr61 and Asp114 and four hydrogen-bond acceptor features pointing towards Val18, Asp19, Gly62 and Lys87. Thus, the citrate molecule could also be further modified to add other pharmacophoric features as observed in KDO8P to generate potential inhibitors of this enzyme.

4. Conclusion

In conclusion, the crystal structures presented here provide molecular insight into the binding of various ligands in the active site of *Mc*-KdsC. For the first time, we have provided structural evidence for the presence of a tunnel at the interface of the two subunits in KdsC. Here, a role of this tunnel in the migration of water molecules into the active site to take part in the catalytic reaction has been proposed, which is further supported by other evidence. The residues that line the tunnel are mostly conserved and could be targeted for mutagenesis studies to confirm the role of the identified tunnel in transferring water to the active site. We have hypothesized the presence of a water channel through the β-barrel. At this time, we cannot support the role of Glu32 in the opening and closing of the β-barrel with evidence, but there may be a possibility that the presence of ligand may affect the conformation of this residue in assisting the transfer of water from the β-barrel through the proposed water channel. Differences in the conformation of the C-terminal loop within different KdsC crystal structures indicate that *Mc*-KdsC might have evolved much earlier and that the role of the C-terminus evolved later for improved efficiency of the enzyme but is not absolutely essential for the catalytic reaction.

In addition, as the KDO biosynthesis pathway is essential for the survival of Gram-negative bacteria, we have reported 11 ligands with the help of pharmacophore mapping for their ability to act as potential inhibitors of this enzyme. The docking results of the screened ligands show similar pharmacophoric features and binding patterns to that of the natural substrate of the enzyme, KDO8P. We are further working towards *in vitro* screening of the proposed inhibitors to assess the binding potential of these ligands at the active site.

References

Allen, K. N. & Dunaway-Mariano, D. (2004). *Trends Biochem. Sci.* **29**, 495–503.
Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). *J. Mol. Biol.* **215**, 403–410.
Baasov, T. & Belakhov, V. (2000). *Drug Dev. Res.* **50**, 416–424.
Belunis, C. J., Clementz, T., Carty, S. M. & Raetz, C. R. H. (1995). *J. Biol. Chem.* **270**, 27646–27652.
Biswas, T., Yi, L., Aggarwal, P., Wu, J., Rubin, J. R., Stuckey, J. A., Woodard, R. W. & Tsodikov, O. V. (2009). *J. Biol. Chem.* **284**, 30594–30603.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Burroughs, A. M., Allen, K. N., Dunaway-Mariano, D. & Aravind, L. (2006). *J. Mol. Biol.* **361**, 1003–1034.
Calderone, V., Forleo, C., Benvenuti, M., Thaller, M. C., Rossolini, G. M. & Mangani, S. (2004). *J. Mol. Biol.* **335**, 761–773.
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.* D**66**, 12–21.
Chovancova, E., Pavelka, A., Benes, P., Strnad, O., Brezovsky, J., Kozlikova, B., Gora, A., Sustr, V., Klvana, M., Medek, P., Biedermannova, L., Sochor, J. & Damborsky, J. (2012). *PLoS Comput. Biol.* **8**, e1002708.
Cipolla, L., Gabrielli, L., Bini, D., Russo, L. & Shaikh, N. (2010). *Nat. Prod. Rep.* **27**, 1618–1629.
Cipolla, L., Polissi, A., Airoldi, C., Galliani, P., Sperandeo, P. & Nicotra, F. (2009). *Curr. Drug Discov. Technol.* **6**, 19–33.
Daughtry, K. D., Huang, H., Malashkevich, V., Patskovsky, Y., Liu, W., Ramagopal, U., Sauder, J. M., Burley, S. K., Almo, S. C., Dunaway-Mariano, D. & Allen, K. N. (2013). *Biochemistry*, **52**, 5372–5386.
DeLano, W. L. (2002). *PyMOL.* http://www.pymol.org.
De Vivo, M., Ensing, B., Dal Peraro, M., Gomez, G. A., Christianson, D. W. & Klein, M. L. (2007). *J. Am. Chem. Soc.* **129**, 387–394.
Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y. & Liang, J. (2006). *Nucleic Acids Res.* **34**, W116–W118.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Evered, D. F. & Steenson, T. I. (1964). *Nature (London)*, **202**, 491–492.
Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M. & Perry, J. K. (2004). *J. Med. Chem.* **47**, 1739–1749.
Galburt, E. A., Pelletier, J., Wilson, G. & Stoddard, B. L. (2002). *Structure*, **10**, 1249–1260.
Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T. & Banks, J. L. (2004). *J. Med. Chem.* **47**, 1750–1759.
Hamid, A. A. A., Wong, E. L., Joyce-Tan, K. H., Shamsir, M. S., Hamid, T. & Huyop, F. (2013). *Biotechnol. Biotechnol. Equip.* **27**, 3725–3736.
Herschlag, D. & Jencks, W. P. (1989). *J. Am. Chem. Soc.* **111**, 7587–7596.
Inoue, H., Nojima, H. & Okayama, H. (1990). *Gene*, **96**, 23–28.
Lu, Z., Wang, L., Dunaway-Mariano, D. & Allen, K. N. (2009). *J. Biol. Chem.* **284**, 1224–1233.
Murphy, T. F., Brauer, A. L., Grant, B. J. B. & Sethi, S. (2005). *Am. J. Respir. Crit. Care Med.* **172**, 195–199.
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.* D**67**, 355–367.
Nordlie, R. C. & Lygre, D. G. (1966). *J. Biol. Chem.* **241**, 3136–3141.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Parsons, J. F., Lim, K., Tempczyk, A., Krajewski, W., Eisenstein, E. & Herzberg, O. (2002). *Proteins*, **46**, 393–404.
Raetz, C. R. H. (1990). *Annu. Rev. Biochem.* **59**, 129–170.
Raetz, C. R. H. & Whitfield, C. (2002). *Annu. Rev. Biochem.* **71**, 635–700.
Ray, P. H. & Benedict, C. D. (1980). *J. Bacteriol.* **142**, 60–68.
Ridder, I. S., Rozeboom, H., Kalk, K. H., Janssen, D. B. & Dijkstra, B. W. (1997). *J. Biol. Chem.* **272**, 33015–33022.
Selengut, J. D. (2001). *Biochemistry*, **40**, 12704–12711.
Selengut, J. D. & Levine, R. L. (2000). *Biochemistry*, **39**, 8315–8324.

Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.

Verduin, C. M., Hol, C., Fleer, A., van Dijk, H. & van Belkum, A. (2002). *Clin. Microbiol. Rev.* **15**, 125–144.

Wang, X. & Quinn, P. J. (2010). *Prog. Lipid Res.* **49**, 97–107.

Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.

Wolber, G. & Langer, T. (2005). *J. Chem. Inf. Model.* **45**, 160–169.

Yi, L. (2009). PhD thesis. University of Michigan, USA. http://deepblue.lib.umich.edu/handle/2027.42/62252.