Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

Structural analysis of arabinose-5-phosphate isomerase from *Bacteroides fragilis* and functional implications

Hsiu-Ju Chiu,[a,b] Joanna C. Grant,[a,c] Carol L. Farr,[a,d] Lukasz Jaroszewski,[a,e] Mark W. Knuth,[a,c] Mitchell D. Miller,[a,b] Marc-André Elsliger,[a,d] Ashley M. Deacon,[a,b] Adam Godzik,[a,e,f] Scott A. Lesley[a,c,d] and Ian A. Wilson[a,d]*

[a]Joint Center for Structural Genomics, http://www.jcsg.org, USA, [b]Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, California, USA, [c]Protein Sciences Department, Genomics Institute of the Novartis Research Foundation, San Diego, California, USA, [d]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, California, USA, [e]Program on Bioinformatics and Systems Biology, Sanford-Burnham Medical Research Institute, La Jolla, California, USA, and [f]Center for Research in Biological Systems, University of California, San Diego, La Jolla, California, USA

Correspondence e-mail: wilson@scripps.edu

Received 25 March 2014
Accepted 23 July 2014

PDB reference: arabinose-5-phosphate isomerase, 3etn

The crystal structure of arabinose-5-phosphate isomerase (API) from *Bacteroides fragilis* (bfAPI) was determined at 1.7 Å resolution and was found to be a tetramer of a single-domain sugar isomerase (SIS) with an endogenous ligand, CMP-Kdo (cytidine 5′-monophosphate-3-deoxy-D-manno-oct-2-ulosonate), bound at the active site. API catalyzes the reversible isomerization of D-ribulose 5-phosphate to D-arabinose 5-phosphate in the first step of the Kdo biosynthetic pathway. Interestingly, the bound CMP-Kdo is neither the substrate nor the product of the reaction catalyzed by API, but corresponds to the end product in the Kdo biosynthetic pathway and presumably acts as a feedback inhibitor for bfAPI. The active site of each monomer is located in a surface cleft at the tetramer interface between three monomers and consists of His79 and His186 from two different adjacent monomers and a Ser/Thr-rich region, all of which are highly conserved across APIs. Structure and sequence analyses indicate that His79 and His186 may play important catalytic roles in the isomerization reaction. CMP-Kdo mimetics could therefore serve as potent and specific inhibitors of API and provide broad protection against many different bacterial infections.

1. Introduction

Gram-negative bacteria are distinguished from Gram-positive bacteria by having an extra outer membrane outside their cell wall. This outer membrane serves as a permeability barrier for foreign molecules, thus making Gram-negative bacteria generally more resistant to antibiotics. Certain pathogenic Gram-negative bacteria, such as *Salmonella*, *Yersinia pestis*, *Klebsiella* and certain strains of *Escherichia coli*, are responsible for serious outbreaks and occurrences of infectious diseases throughout the world. Antibiotic treatment alone is often not sufficient to combat these infections, and several antibiotic-resistant strains of Enterobacteriaceae have emerged where no effective drugs are currently available for treatment.

The bacterial outer membrane is an asymmetric glycerophospholipid bilayer with lipopolysaccharides (LPS) located exclusively in the outer layer. These LPS typically consist of three distinct regions: lipid A, a core oligosaccharide and an O-antigen polysaccharide (Cipolla *et al.*, 2009). Lipid A, the hydrophobic portion of LPS, is a glucosamine-based phospholipid. The recognition of LPS by the human immune system and the subsequent immune response to Gram-negative infection is typically attributed to lipid A. The core oligosaccharide can be divided into inner and outer oligosaccharide cores. The inner oligosaccharide core directly

attaches to lipid A and is primarily made of 3-deoxy-D-manno-oct-2-ulosonate (Kdo) moieties, which together with lipid A constitute the most conserved component of LPS. In Gram-negative bacteria, Kdo is essential for outer membrane biogenesis and cell viability. The minimal LPS structure required for the growth of *E. coli* consists of two Kdo modules attached to lipid A (Kdo_2-lipid A; Raetz & Whitfield, 2002). The biosynthesis of Kdo involves four sequential steps. Arabinose-5-phosphate isomerase (API) catalyzes the reversible keto–aldol isomerization of the pentose pathway intermediate D-ribulose 5-phosphate (Ru5P) to D-arabinose 5-phosphate (A5P) in the first step of Kdo biosynthesis (Fig. 1*a*). APIs are highly conserved among Gram-negative bacteria and are often the only *de novo* source of A5P, which is not available *via* glycolysis. Therefore, API is potentially an ideal target for the development of Gram-negative antibacterial drugs, and a potent inhibitor could offer broad protection against many different bacterial infections.

APIs exist either as single-domain sugar isomerase (SIS) proteins or multi-domain proteins consisting of an SIS domain and one or two tandem cystathionine-β-synthase (CBS) domains. The SIS domain is commonly found in keto–aldol sugar isomerases and other proteins that regulate the expression of genes involved in phosphosugar biosynthesis. CBS domains are found in a variety of proteins and play regulatory roles, although their exact function is unknown. API is abundant in Gram-negative bacteria, and some bacteria even contain multiple paralogs. For example, the uropathogenic *E. coli* CFT073 (O6:K2:H1) strain contains four API paralogs: KdsD, KpsF (Meredith & Woodard, 2006), GutQ (Meredith & Woodard, 2005) and c3406 (Mosberg *et al.*, 2011). Although all of these APIs catalyze the same chemical reaction, they are involved in different specific biological functions. KdsD is involved in LPS biosynthesis. KpsF is encoded within the *kps* (capsular polysaccharide) gene cluster and is involved in capsule formation. The *gutQ* gene, on the other hand, is located in an operon involved in D-glucitol (sorbitol) metabolism. However, the involvement of GutQ in the metabolism of D-glucitol, or a link between D-glucitol metabolism and LPS biosynthesis, has not been established (Meredith & Woodard, 2005; Cech *et al.*, 2014). The biological significance of the API redundancy in *E. coli* is currently unknown, but it is speculated that they may have additional roles besides producing A5P.

In order to explore the structure and function of API, we have determined the crystal structure of arabinose-5-phosphate isomerase from *Bacteroides*

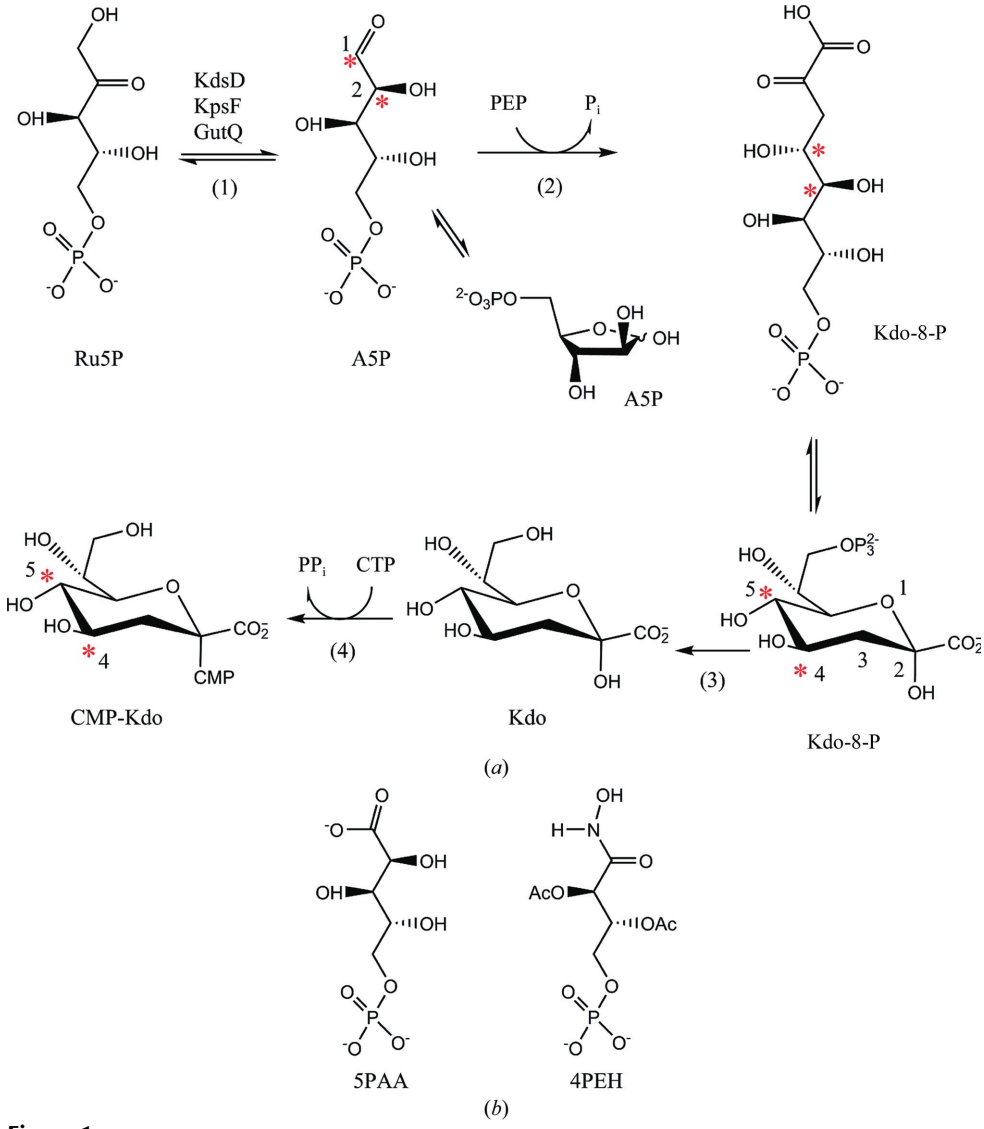


Figure 1
The Kdo biosynthetic pathway, substrate analogues and inhibitors. (*a*) The Kdo biosynthetic pathway. The enzymes involved are (1) arabinose-5-phosphate isomerase (API), (2) Kdo-8-phosphate (Kdo-8-P) synthase, (3) Kdo-8-P phosphatase and (4) cytidine 5′-monophosphate-Kdo synthetase. API catalyzes the reversible isomerization of D-ribulose 5-phosphate (Ru5P) to D-arabinose 5-phosphate (A5P) in the first step of Kdo biosynthesis. *E. coli* contains multiple paralogs of API genes including *kdsD*, *kpsF* and *gutQ*. The open-chain form and the ring form of A5P and Kdo-8-P are shown. PEP stands for phosphoenolpyruvate. The C1 and C2 atoms of A5P are marked with red asterisks. The C4 and C5 atoms of CMP-Kdo derived from C1 and C2 of A5P are also marked with red asterisks. (*b*) 5-Phospho-D-arabinonate (5PAA), a G6P substrate analog commonly used in enzyme mechanistic studies of PGI, is shown on the left. 4-Phospho-2,3-acetyl-D-erythrose hydroxamate (4PEH) was the most potent API inhibitor among the hydroxamate compounds tested in a study of *F. tularensis* KdsD (Yep *et al.*, 2011).

fragilis NCTC 9343 (bfAPI) at 1.7 Å resolution. bfAPI is a single-domain SIS protein with an endogenous CMP-Kdo molecule bound at the active site. Interestingly, CMP-Kdo is neither the substrate nor the product of the reaction catalyzed by API, but corresponds to the end product of the Kdo biosynthetic pathway; it could therefore act as a feedback inhibitor. Our conclusions are supported by a recent study that demonstrates that bfAPI has arabinose-5-phosphate isomerase activity and is inhibited by CMP-Kdo (Cech *et al.*, 2014). Here, we report the structural analysis of bfAPI and its functional implications. The crystal structure was determined using the semi-automated high-throughput pipeline of the Joint Center for Structural Genomics (JCSG; Lesley *et al.*, 2002; Elsliger *et al.*, 2010) as part of the National Institute of General Medical Sciences' Protein Structure Initiative (PSI; http://www.nigms.nih.gov/Research/SpecificAreas/PSI/Pages/default.aspx).

2. Materials and methods

2.1. Protein production and crystallization

Clones were generated using the Polymerase Incomplete Primer Extension (PIPE) cloning method (Klock *et al.*, 2008). The gene encoding bfAPI (GenBank YP_209877, UniProt Q650A5) was amplified by polymerase chain reaction (PCR) from *B. fragilis* NCTC 9343 genomic DNA (ATCC No. 25285D) using *Pfu* Turbo DNA polymerase (Stratagene) and I-PIPE (Insert) primers (forward primer, 5′-ctgtacttccagggc ATGATTGAATCTATTCAAGAACTCCTGC-3′; reverse primer, 5′-aattaagtcgcgttaCTTTACGCATAGTTTTCTTGA-TTTTTCG-3′; target sequence in upper case) that included sequences for the predicted 5′ and 3′ ends. The expression vector pSpeedET, which encodes an amino-terminal *Tobacco etch virus* protease-cleavable expression and purification tag (MGSDKIHHHHHHENLYFQ/G), was PCR-amplified with V-PIPE (Vector) primers (forward primer, 5′-taacgcgacttaat-taactcgtttaaacggtctccagc-3′; reverse primer, 5′-gccctggaagta-caggttttcgtgatgatgatgatgatg-3′). The V-PIPE and I-PIPE PCR products were mixed to anneal the amplified DNA fragments together. *E. coli* GeneHogs (Invitrogen) competent cells were transformed with the I-PIPE/V-PIPE mixture and dispensed onto selective LB–agar plates. Expression was performed in a selenomethionine-containing medium at 37°C. Selenomethionine was incorporated *via* inhibition of methionine biosynthesis (Van Duyne *et al.*, 1993). At the end of fermentation, lysozyme was added to the culture to a final concentration of 250 µg ml^{-1} and the cells were harvested and frozen. After one freeze–thaw cycle, the cells were homogenized and sonicated in lysis buffer [50 m*M* HEPES, 50 m*M* NaCl, 10 m*M* imidazole, 1 m*M* tris(2-carboxyethyl)phosphine–HCl (TCEP) pH 8.0] and the lysate was clarified by centrifugation at 32 500*g* for 30 min. The soluble fraction was passed over nickel-chelating resin (GE Healthcare) pre-equilibrated with lysis buffer, the resin was washed with wash buffer [50 m*M* HEPES, 300 m*M* NaCl, 40 m*M* imidazole, 10%(*v*/*v*) glycerol, 1 m*M* TCEP pH 8.0] and the protein was eluted with elution buffer [20 m*M* HEPES, 300 m*M* imidazole, 10%(*v*/*v*) glycerol, 1 m*M* TCEP pH 8.0]. The eluate was buffer-exchanged with HEPES crystallization buffer [20 m*M* HEPES, 200 m*M* NaCl, 40 m*M* imidazole, 1 m*M* TCEP pH 8.0] using a PD-10 column (GE Healthcare) and concentrated to 20 mg ml^{-1} using centrifugal ultrafiltration (Millipore). bfAPI was crystallized using the nanodroplet vapor-diffusion method (Santarsiero *et al.*, 2002) with standard JCSG crystallization protocols (Lesley *et al.*, 2002). Sitting drops composed of 200 nl protein solution mixed with 200 nl crystallization solution were equilibrated against 50 µl reservoir solution at 277 K for 18 d prior to harvesting. The crystallization reagent consisted of 5% PEG 6000, 0.1 *M* MES pH 6.0. Ethylene glycol was added to a final concentration of 20%(*v*/*v*) as a cryoprotectant. The oligomeric state of bfAPI in solution was determined to be a tetramer using a 1 × 30 cm Superdex 200 column (GE Healthcare) coupled with miniDAWN static light-scattering (SEC/SLS) and Optilab differential refractive-index detectors (Wyatt Technology). The mobile phase consisted of 20 m*M* Tris pH 8.0, 150 m*M* sodium chloride and 0.02%(*w*/*v*) sodium azide.

2.2. Data collection, structure solution and refinement

Multiple-wavelength anomalous diffraction (MAD) data were collected to 1.7 Å resolution on beamline 11-1 at SSRL at wavelengths corresponding to the inflection (λ_1), high-energy remote (λ_2) and peak (λ_3) of a selenium MAD experiment. The data sets were collected at 100 K using a MAR Mosaic 325 CCD detector (Rayonix). The MAD data were integrated using *MOSFLM* (Leslie, 1992) and scaled with *SCALA* from the *CCP*4 suite (Winn *et al.*, 2011). An initial substructure solution was obtained with *SHELXD* (Sheldrick, 2008) and the phases were refined using *auto-SHARP* (Vonrhein *et al.*, 2007), with a mean figure of merit of 0.50 with 28 selenium sites. Automated model building was performed with *RESOLVE* (Terwilliger, 2003). Model building and refinement were performed with *Coot* (Emsley & Cowtan, 2004) and *REFMAC*5.5 (Winn *et al.*, 2003) using the high-energy remote (λ_2) data set. The refinement included NCS restraints, phase restraints from *SHARP* and TLS parameters with four TLS groups: one group per chain. Data-collection and refinement statistics are summarized in Table 1.

Unaccounted-for electron density at the active site was observed in the experimental maps, and was later unambiguously assigned to CMP-Kdo during model building and refinement (Fig. 3*d*). Since CMP-Kdo was not added at any stage of protein production, it must have originated from the expression system and remained bound throughout purification and crystallization.

2.3. Validation and deposition

The quality of the crystal structure was analyzed using the *JCSG Quality Control* server (http://smb.slac.stanford.edu/jcsg/QC). This server verifies the stereochemical quality of the model using *AutoDepInputTool* (Yang *et al.*, 2004), *MolProbity* (Chen *et al.*, 2010) and *PHENIX* (Adams *et al.*, 2010), the agreement between the atomic model and the data

using *RESOLVE* (Terwilliger, 2000), the protein sequence using *ClustalW* (Thompson *et al.*, 1994), the ADP distribution using *PHENIX* and differences in R_{cryst}/R_{free}, expected R_{free}/R_{cryst} and various other items including atom occupancies, consistency of NCS pairs, ligand interactions, special positions *etc.* using in-house scripts to analyze the refinement log file and PDB header. Protein quaternary structure analysis was performed using the *PISA* server (Krissinel & Henrick, 2007). Atomic coordinates and experimental structure factors have been deposited in the PDB as entry 3etn.

3. Results and discussion

3.1. Sequence similarity

A sequence similarity search with the bfAPI protein sequence using *HMMER* (Finn *et al.*, 2011) against the NCBI nonredundant protein sequence database identified ∼2000 homologous protein sequences with *E*-values of $<e^{-30}$. These sequences were from several families of sugar phosphate isomerases, including KdsD, KpsF, GutQ and polysialic acid capsule expression protein. 95% of these sequences were multi-domain proteins containing an SIS domain and one or two CBS domains, and the rest were single-domain proteins containing an SIS domain only. Alignment of these top 2000 sequences revealed several highly conserved residues including Lys50, Ser64, His79, Asp81, Asp89, Ser95, Ser97, Gly98, Glu102, Glu145, Asp165 and His186 (Fig. 2). The alignment also shows that His187 and Tyr190 are only conserved in single-domain proteins which are mostly present in the genus *Bacteroides*.

API genes are abundant in the pathogenic Gram-negative bacteria within the Enterobacteriaceae family. A further sequence-similarity search with the bfAPI protein sequence using *HMMER* against the UniProtKB database limited to Enterobacteriaceae (taxid 543) found ∼5400 sequences with *E*-values of $<e^{-6}$, which were all annotated as either arabinose-5-phosphate isomerase, KdsD, GutQ or KpsF. 8437 organisms are listed in the Enterobacteriaceae family in UniProtKB; ∼93% of these Gram-negative bacteria are pathogenic and include *Klebsiella pneumoniae*, *Yersinia pestis*, *Salmonella*, *Shigella* and certain strains of *E. coli*. Overall, the 5400 sequences identified above represent ∼70% of the pathogenic Gram-negative bacteria in the Enterobacteriaceae family.

3.2. Overall structure

The final model consists of a tetramer (residues 0–191 for chains *A* and *C* and residues −4–193 for chains *B* and *D*), four CMP-Kdo, seven ethylene glycol and 725 water molecules in the asymmetric unit (Figs. 3*a* and 3*b*). Parts of the N-terminal tag (−18 to −1 in chains *A* and *C* and −18 to −5 in chains *B* and *D*) and C-terminal residues (192–201 in chains *A* and *C* and 194–201 in chains *B* and *D*) were disordered. bfAPI is an α/β protein containing an SIS domain. It adopts a flavodoxin-like, three-layered $\alpha/\beta/\alpha$ fold, in which the central five-stranded β-sheet ($\beta2$–$\beta1$–$\beta3$–$\beta4$–$\beta5$) is sandwiched between helices $\alpha4$, $\alpha5$, $\alpha6$ and $\alpha7$ on one side and $\alpha2$, $\alpha3$ and $\alpha8$ on the other side (Fig. 3*c*). In addition, bfAPI has a long helix $\alpha1$ (residues 1–17) at the N-terminus and helix $\alpha9$ (residues 188–192) at the C-terminus. bfAPI was determined to be a tetramer in solution by size-exclusion chromatography, which is consistent with the oligomerization state in the crystal lattice and the reported oligomerization states of other characterized API enzymes (Gourlay *et al.*, 2010; Meredith & Woodard, 2005, 2006). The structures of the four protein subunits of the tetramer are very similar to each other, with r.m.s.d.s ranging between 0.6 and 1.3 Å for 191 equivalent C^{α} atoms aligned. The oligomer is assembled as a dimer of dimers (*e.g.* subunits *A* and *B* form dimer 1 and subunits *C* and *D* form dimer 2) related by a twofold axis, which results in a tetramer with 222

Table 1
Data collection and refinement statistics for bfAPI.

Values in parentheses are for the highest resolution shell.

	λ_1	λ_2	λ_3
Data collection			
Space group	$P2_12_12_1$		
Unit-cell parameters (Å)	$a = 60.41$, $b = 114.22$, $c = 115.92$		
Wavelength (Å)	0.9791	0.9184	0.9785
Resolution range (Å)	29.24–1.74 (1.83–1.74)	29.22–1.70 (1.79–1.70)	29.24–1.79 (1.84–1.79)
No. of observations	369990	398085	340285
No. of unique reflections	82883	88793	76256
Completeness (%)	99.8 (99.6)	99.8 (100.0)	99.7 (99.5)
Mean $I/\sigma(I)$	10.0 (1.7)	9.9 (2.0)	9.4 (1.9)
R_{merge} on I† (%)	10.3 (80.4)	10.3 (73.2)	11.5 (72.2)
R_{meas} on I‡ (%)	11.7 (91.2)	11.7 (83.0)	13.0 (81.8)
$R_{p.i.m.}$ on I§ (%)	5.5 (42.4)	5.5 (38.7)	6.1 (38.1)
Model and refinement statistics			
Data set used in refinement	λ_2		
Resolution range (Å)	29.22–1.70		
Cutoff criterion	$\|F\| > 0$		
No. of reflections (total)	88524		
No. of reflections (test)	4437		
Completeness (%)	99.6		
R_{cryst}¶	0.172		
R_{free}††	0.197		
Stereochemical parameters			
Restraints (r.m.s.d. observed)			
Bond angles (°)	1.45		
Bond lengths (Å)	0.014		
Wilson B value (Å^2)	19.1		
Average isotropic B value‡‡ (Å^2)			
Overall, all non-H atoms	22.1		
CMP-Kdo ligand	20.8		
Waters	29.9		
ESU based on R_{free} (Å)	0.099		
Ramachandran favored§§ (%)	98.2		
Ramachandran outliers§§ (%)	0.0		
Poor rotamers§§ (%)	1.2		
No. of protein residues	780		
No. of protein atoms	5871		
No. of water molecules	725		
No. of ligands	4		

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$. ‡ R_{meas} (redundancy-independent R_{merge}) $= \sum_{hkl}\{N(hkl)/[N(hkl)-1]\}^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$ (Diederichs & Karplus, 1997). § $R_{p.i.m.}$ (precision-indicating R_{merge}) $= \sum_{hkl}\{1/[N(hkl)-1]\}^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$ (Weiss & Hilgenfeld, 1997). ¶ $R_{cryst} = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$, where F_{calc} and F_{obs} are the calculated and observed structure-factor amplitudes, respectively. †† R_{free} is calculated as for R_{cryst} but for 5.0% of the total reflections chosen at random and omitted from refinement. ‡‡ This value represents the total B that includes TLS and residual B components. §§ The percentage of residues in Ramachandran favored and outlier regions and the percentage of poor rotamers were calculated using *MolProbity*.

symmetry. Thus, subunits *A*/*B* and *C*/*D* are more tightly associated with each other compared with the *A*/*D* and *B*/*C* interactions. The buried surface area between subunits *A*/*B* and *C*/*D* is approximately 2330 Å^2 per monomer, which is about twice the buried surface area between subunits *A*/*D* and *B*/*C* (∼1100 Å^2). Interactions between dimer 1 and dimer 2 are formed from helices α1, α3 and α8 of one subunit, which make extensive hydrogen-bonding interactions with their counterparts α1′, α3′ and α8′ from the other subunit. A hydrogen bond between Lys50 from α3 of one subunit and Asp165 from α8′ of the other subunit helps to anchor the C-terminal α9 helix that contains the active-site residue His186.

3.3. The active site

As the four copies of the active site in the asymmetric unit are nearly identical, only the active site in subunit *A* will be described. The active site is located in a cleft at the tetramer interface that is composed of residues from loops $L_{\beta1-\alpha3}$ and $L_{\beta3-\alpha6}$ of subunit *A*, α9 and the C-terminal tail of subunit *B* and the region around α4 and α5 of subunit *C*. Loop $L_{\beta3-\alpha6}$ contains a Ser/Thr-rich region with the highly conserved residues Ser95 and Ser97 forming hydrogen bonds to the phosphate group of the bound CMP-Kdo molecule. In all four active sites CMP-Kdo molecules are bound on the C-terminal side of the central β-sheet, with the 3-deoxy-D-manno-oct-2-ulosonate group pointing towards the central β-sheet and

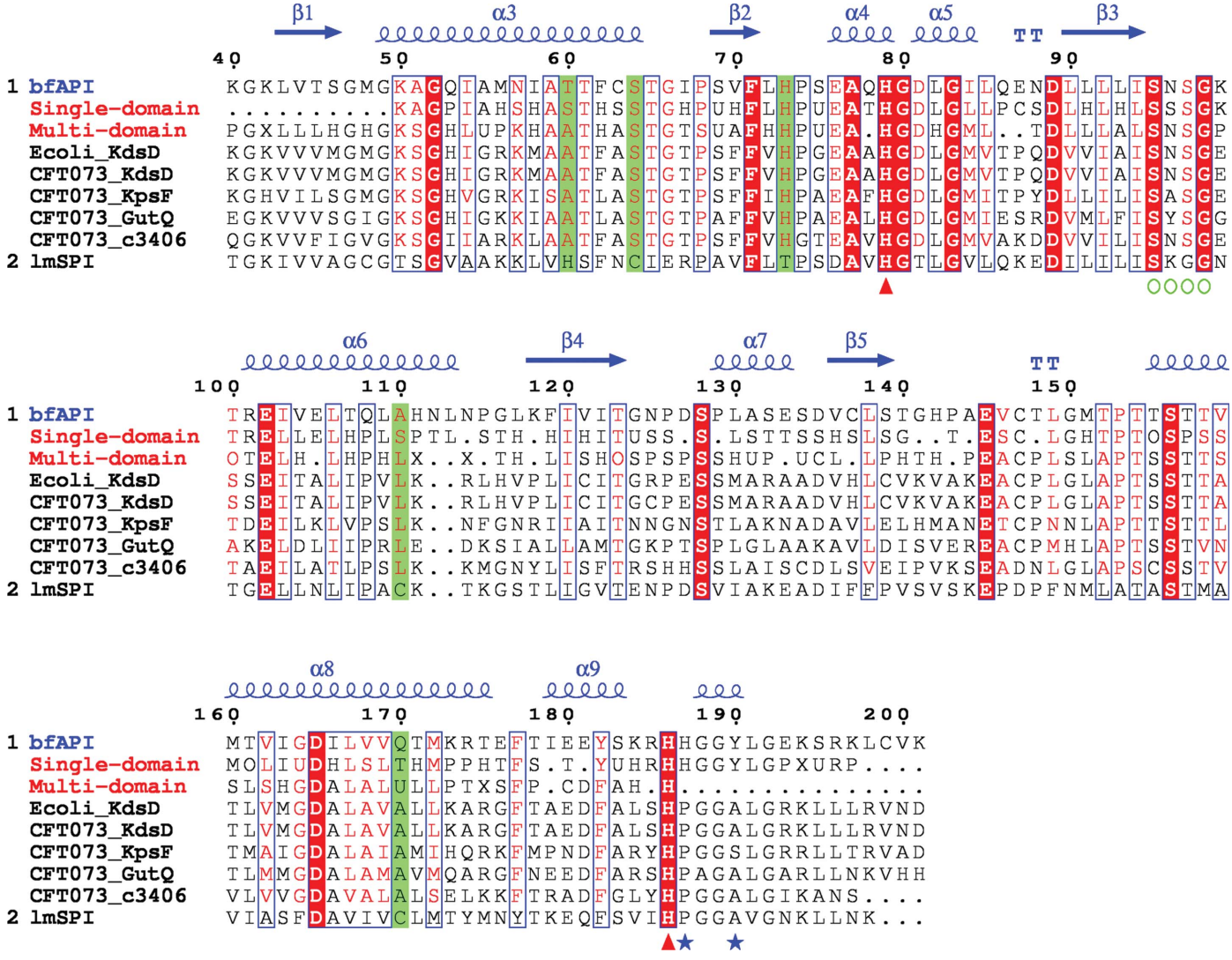


Figure 2
Comparison of API sequences. The consensus-aligned sequences of single-domain and multi-domain APIs were calculated using *Consensus* (http://coot.embl.de/Alignment/consensus.html) and rendered using *ESPript* (http://espript.ibcp.fr/ESPript/ESPript/). The invariant His79 and His186 are indicated by red triangles. The serine-rich region involved in sugar-phosphate binding is indicated by open green circles. His187 and Tyr190 are conserved among single-domain APIs and are indicated by blue stars. The following abbreviations are used. Single-domain: the consensus sequence at the 90% sequence-identity level of the top 80 single-domain API sequences identified by *HMMER*. Multi-domain: the consensus sequence at the 90% sequence-identity level of the top 500 multi-domain API sequences identified by *HMMER*. *E. coli*_KdsD: KdsD from *E. coli* K12 substrain MG1655 (ecAPI). CFT073_KdsD, CFT073_KpsF, CFT073_GutQ and CFT073_c3406: KdsD, KpsF, GutQ and c3406 from *E. coli* CFT073, respectively. lmSPI, putative sugar-phosphate isomerase from *Listeria monocytogenes* strain 4b F2365.

the CMP moiety pointing towards the protein surface. The Kdo interaction is stabilized by extensive hydrogen bonds and hydrophobic interactions with Tyr190(*B*), Ala51(*A*), Asn96(*A*) and Ser97(*A*).

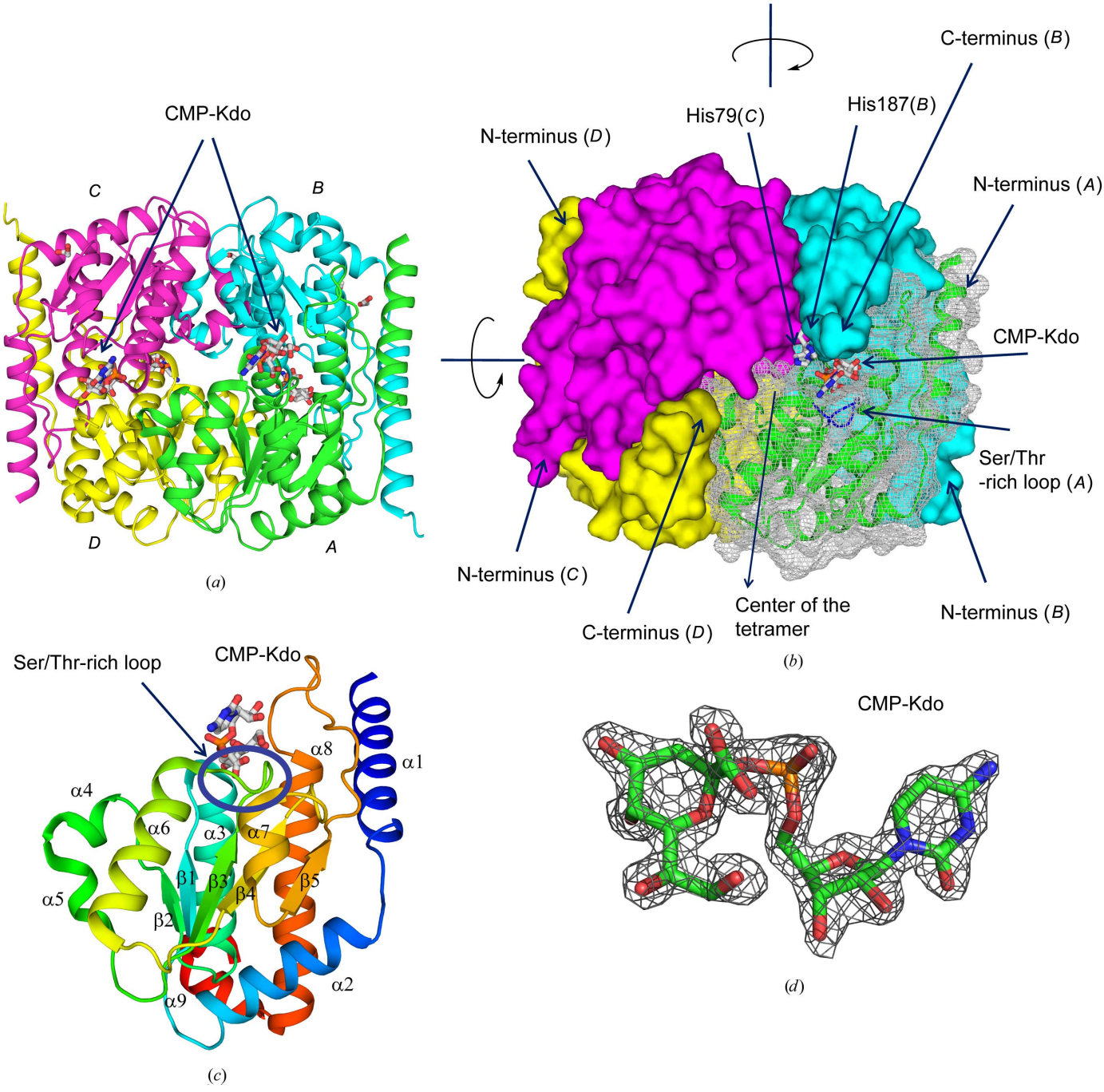


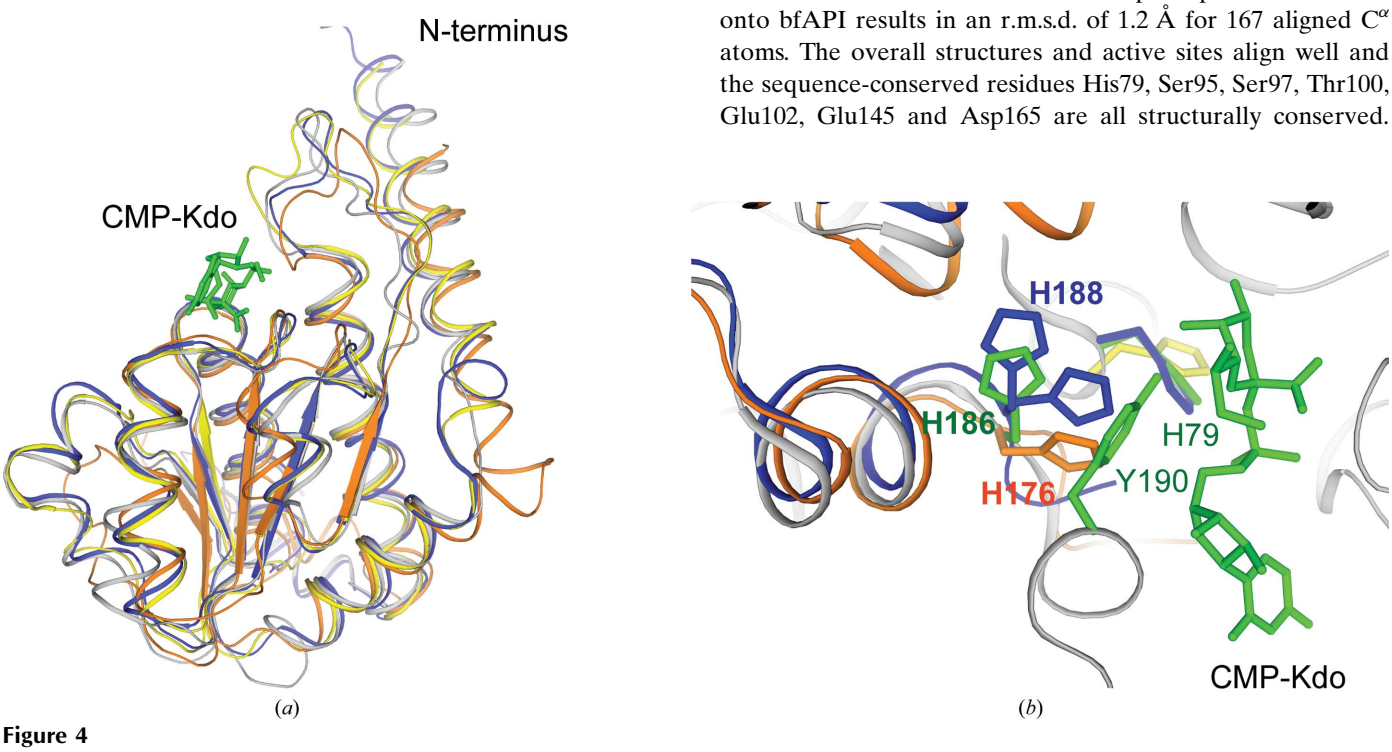
Figure 3
Crystal structure of bfAPI. (*a*) Ribbon diagram of the bfAPI tetramer with CMP-Kdo bound at the active site and each monomer rendered in a different color. Bound CMP-Kdo molecules are shown in ball-and-stick representation, with C, O, N and P atoms colored gray, red, blue and orange, respectively. (*b*) The arrangement of the subunits in an API tetramer in the same orientation as in (*a*). Subunit *A* is shown in green ribbon representation and its molecular surface is shown as a gray mesh. The molecular surfaces of subunits *B*, *C* and *D* of bfAPI are colored blue, magenta and yellow, respectively. The tetramer exhibits 222 symmetry, with three orthogonal twofold axes relating one half (dimer) of the tetramer to the other half. These twofold axes pass through the center of the tetramer. One twofold axis runs vertically and the other one runs horizontally along the view shown. The third twofold axis runs perpendicular to the molecule view, going from the outside to the inside of the view. The active site of subunit *A* is shown and is located at the interface between subunits *A*, *B* and *C*. The CMP-Kdo molecule and the active-site residues His79(*C*) and His187(*B*) are shown in ball-and-stick representation. The Ser/Thr-rich loop from subunit *A* is colored dark blue. (*c*) Ribbon diagram of a bfAPI monomer color-coded from the N-terminus (blue) to the C-terminus (red). Helices α1–α9 and strands β1–β5 are indicated. The bound CMP-Kdo is shown in a ball-and-stick representation and the Ser/Thr-rich loop is highlighted by a blue oval. (*d*) The final model of CMP-Kdo fitted in the experimental map, contoured at 1.6 times the r.m.s. of the map (0.75 e^{-} Å^{-3}), calculated with *SOLOMON* solvent-flattened phases from *autoSHARP* confirm the quality of CMP-Kdo model.

His79 in helix α4 is an invariant residue among API sequences and is equivalent to His88 in *E. coli* API/KdsD (ecAPI; PDB entry 2xhz), where biochemical and mutational experiments have confirmed its catalytic role (Gourlay *et al.*, 2010). His79 is located near the phosphate O atoms and the C4 and C5 hydroxyl groups of the bound Kdo. The C4 atom of Kdo is derived from the C1 atom of A5P during CMP-Kdo biosynthesis. The $N^{\delta1}$ atom of His79(*C*) hydrogen bonds to the Tyr190(*B*) hydroxyl, which is further hydrogen-bonded to the Kdo C5 hydroxyl. Tyr190(*B*) is located between the Kdo moiety and His186(*B*) and therefore blocks any direct interaction between His186(*B*) and CMP-Kdo.

Interestingly, the bound CMP-Kdo molecule in the active site is neither a product nor a substrate of bfAPI and is not derived from the crystallization solution. It is the end product of the reaction catalyzed by KdsB in the last step of Kdo biosynthesis. However, the primary sequences of bfAPI and KdsB are unrelated and their structures are completely different, while the known KdsB structures (*E. coli*, PDB entry 3k8d, Heyes *et al.*, 2009; *Acinetobacter baumannii*, PDB entry 3pol, Center for Structural Genomics of Infectious Diseases, unpublished work; *Yersinia pestis*, PDB entry 3jtj, Center for Structural Genomics of Infectious Diseases, unpublished work; *Haemophilus influenzae*, PDB entry 1vh3, Badger *et al.*, 2005; *Vibrio cholerae*, PDB entry 3oam, Center for Structural Genomics of Infectious Diseases, unpublished work) share significant sequence and structure similarities. Since an end product of a biosynthetic pathway often inhibits the first or early steps of the reaction pathway (Pardee, 1971), our observations suggest that the bound CMP-Kdo observed in our bfAPI structure may act as a feedback inhibitor of bfAPI. Our hypotheses are supported by recent studies that demonstrate that CMP-Kdo does effectively inhibit bfAPI (K_i = 1.91 μM; Cech *et al.*, 2014) and raises the possibility of using the natural inhibitor CMP-Kdo as a lead compound for drug discovery. An analogous study has been reported in which hydroxamates were used as enediol mimics to probe for potent inhibitors of *Francisella tularensis* KdsD (Yep *et al.*, 2011). In that study, 4-phospho-2,3-acetyl-D-erythrose hydroxamate (PDB entry 4peh; Fig. 1*b*) with an IC_{50} of 7 µm (approximate K_i of 1.93 μM calculated using the *BotDB* web server; Cer *et al.*, 2009) was the most potent inhibitor among those tested. The bfAPI and *F. tularensis* KdsDs share 40% sequence identity and the key active-site residues His79 and His186 are conserved.

3.4. Structure-similarity analyses

Other than bfAPI, *E. coli* API (ecAPI; PDB entry 2xhz) is the only other API structure available (Gourlay *et al.*, 2010). Although ecAPI is a multi-domain API containing an SIS domain and a pair of CBS domains, the protein construct used to determine the crystal structure only contains residues 1–183 corresponding to the SIS domain. The biological assembly of ecAPI is also a tetramer like bfAPI. Superimposition of ecAPI onto bfAPI results in an r.m.s.d. of 1.2 Å for 167 aligned C^{α} atoms. The overall structures and active sites align well and the sequence-conserved residues His79, Ser95, Ser97, Thr100, Glu102, Glu145 and Asp165 are all structurally conserved.



(*a*) (*b*)

Figure 4
Comparison of API and SPI structures. (*a*) Superimposition of the bfAPI structure (gray) onto KdsD from *E. coli* K-12 MG1655 (ecAPI; PDB entry 2xhz; yellow), the putative sugar-phosphate isomerase from *L. monocytogenes* strain 4b (lmSPI; PDB entry 3fxa; blue) and 3-hexulose-6-phosphate isomerase from *M. jannaschii* (mjPHI; PDB entry 1jeo; orange). CMP-Kdo in bfAPI is shown in a green ball-and-stick representation. (*b*) Comparison of the active sites in the bfAPI, ecAPI, lmSPI and mjPHI structures. His79, His186, Tyr190 and CMP-Kdo in bfAPI are shown in green ball-and-stick representation. The dual conformers of His188 in lmSPI subunit *D* are shown in blue and His176 of mjPHI is shown in an orange ball-and-stick representation. The histidine residues from ecAPI and lmSPI that are equivalent to His79 of bfAPI are shown in yellow and blue ball-and-stick representations, respectively. Residue names are labeled in green, blue and red for bfAPI, lmSPI and mjPHI, respectively.

A putative sugar-phosphate isomerase from *Listeria monocytogenes* strain 4b (lmSPI; PDB entry 3fxa; r.m.s.d. of 1.4 Å for 181 aligned C^{α} atoms; Joint Center for Structural Genomics, unpublished work) and 3-hexulose-6-phosphate isomerase from *Methanococcus jannaschii* (mjPHI; PDB entry 1jeo; r.m.s.d. of 1.9 Å for 156 aligned C^{α} atoms; Martinez-Cruz *et al.*, 2002) are the other identified top hits from a *PDBeFold* structure-similarity search (Krissinel & Henrick, 2007). lmSPI and mjPHI are both single-domain SIS-containing proteins and share sequence identities of 35% (52% similarity) and 26% (46% similarity) to bfAPI, respectively. The mjPHI structure contains a citric acid molecule bound near the equivalent Kdo binding site in bfAPI, but lmSPI does not have any bound ligand. When all four structures are superimposed (Fig. 4*a*), we observe that His186 of bfAPI is structurally equivalent to His188 of lmSPI and His176 of mjPHI (Fig. 4*b*). The equivalent residue (His193) in ecAPI is not present in the structure since this region was not included in the construct used for structure determination. ecAPI and lmSPI have a histidine residue at the position equivalent to His79 of bfAPI which is not conserved in mjPHI (Fig. 4*b*). His188 from lmSPI subunit *D* adopts two alternate side-chain conformations. The His imidazole in the first conformation points away from the active site and is observed in all subunits of lmSPI and bfAPI, whereas in the second conformation it points towards the active site in subunit *D* of lmSPI and in the mjPHI structures. If His186 of bfAPI were to adopt this second conformation, it would clash with the side chain of Tyr190. Interestingly, mutation of the equivalent residue in ecAPI, His193, to alanine yields an enzyme that is inactive both *in vivo* and *in vitro* (Gourlay *et al.*, 2010). Previous structural and sequence analyses of mjPHI have also shown that His176 has a catalytic role (Martinez-Cruz *et al.*, 2002). Based on the observations above, we speculate that His186 may play an important catalytic role in bfAPI. The fact that the His186 imidazole points away from the bound CMP-Kdo molecule could stem from the orientation of the Tyr190 side chain in this structure (Fig. 4*b*). Tyr190 is strictly conserved among single-domain APIs but not in multi-domain APIs. It is located in the flexible C-terminal tail region and could presumably function as a gate where it would move close to the bound inhibitor to block enzyme activity and move away when the actual substrate is bound to allow the approach of the catalytic His186 for enzyme activity. Therefore, Tyr190 could play a regulatory role for bfAPI activity, which is presumably carried out by the CBS domain in multi-domain API enzymes.

3.5. Insights into the catalytic mechanism through structural comparison with other sugar isomerases

API catalyzes the reversible isomerization of D-ribulose-5-phosphate to D-arabinose-5-phosphate. Although little is

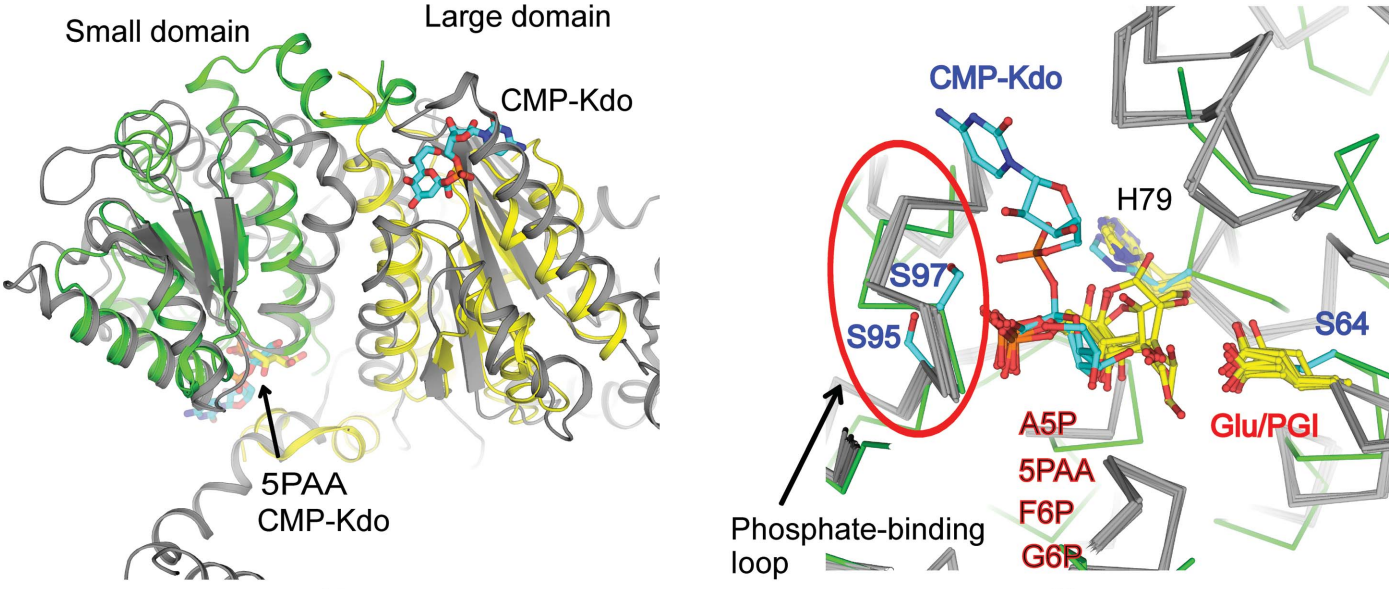
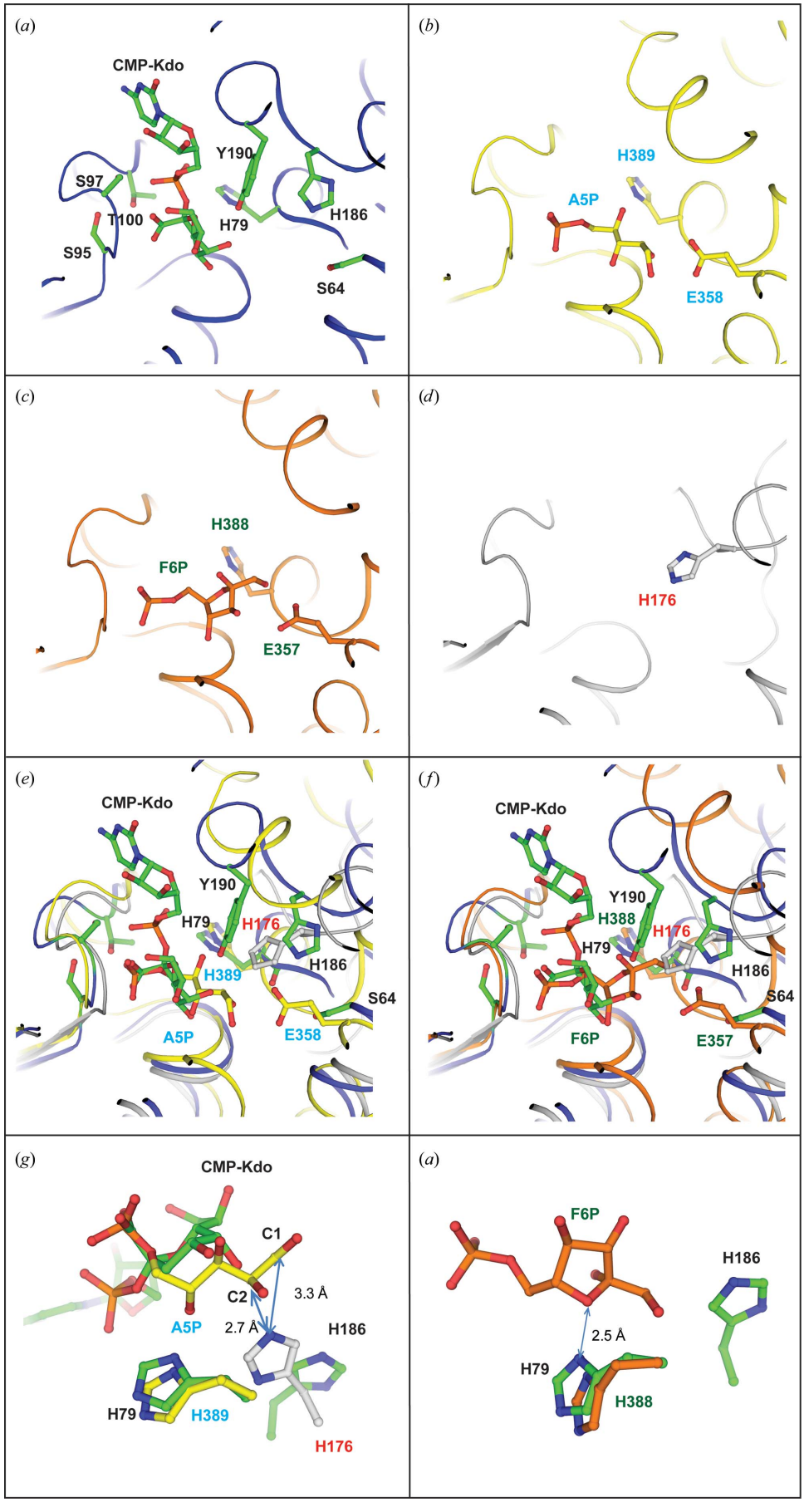


Figure 5
Comparison of API and PGI structures. (*a*) Superposition of the bfAPI and rabbit PGI–5PAA (PDB entry 1g98) structures. 1g98 is shown in gray ribbons and the bfAPI monomers superimposed onto the small and large SIS domains of 1g98 are shown as green and yellow ribbons, respectively. CMP-Kdo from the bfAPI structure and 5PAA from the 1g98 structure are shown in ball-and-stick representation. 5PAA was only bound in the small domain of 1g98. (*b*) Comparison of the bfAPI and the small domains of the PGI structures shows that His79 and the Ser/Thr-rich region are structurally conserved and Ser64 of bfAPI corresponds to a conserved glutamate in PGI. The phosphate group of each of the bound ligand in the PGI structures interacts with the Ser/Thr-rich loop and overlaps with the carboxylate group in Kdo. bfAPI is shown as green ribbons and the PGIs are shown as gray ribbons. The ligands are shown in ball-and-stick representation with N atoms in blue, O atoms in red, P atoms in orange and C atoms of CMP-Kdo and PGIs in cyan and yellow, respectively. The carbon atoms of the active-site residues in bfAPI and PGIs are colored cyan and yellow, respectively. The PDB code, the substrate/product analogs (in parentheses) and the source organisms of the PGI structures used in the superimposition are 2cxp (A5P), mouse; 1u0f (G6P, glucose 6-phosphate), mouse; 1hox (F6P, fructose 6-phosphate), rabbit; 1g98 (5PAA, 5-phospho-D-arabinonate), rabbit; 1nuh (5PAA), human; 1gzv (5PAA), *Sus scrofa*; 2o2c (G6P), *Trypanosoma brucei*; 2q8n, *Thermotoga maritima*. Note that for clarity only eight of the 16 PGI structures mentioned in §3.5 are shown.



known about the catalytic mechanism of API and the catalytic residues involved, we gleaned some insights by comparing bfAPI with other sugar isomerases, such as phosphoglucose isomerase (PGI), which catalyze an analogous reaction with similar substrates. PGI, a key enzyme in glycolysis, is a sugar isomerase that catalyzes the reversible isomerization of

Figure 6
Comparison of bfAPI with mouse PGI–A5P (PDB entry 2cxp), rabbit PGI–F6P (PDB entry 1hox) and mjPHI structures suggests a catalytic role for His79 and His186. The structures of bfAPI, 2cxp, 1hox and mjPHI are shown in ribbon representation and are colored blue, yellow, orange and gray, respectively. CMP-Kdo from bfAPI, A5P from 2cxp, F6P from 1hox and the active-site residues are shown in ball-and-stick representation with N atoms in blue, O atoms in red and P atoms in orange. The C atoms of CMP-Kdo and the active-site residues of bfAPI are in green, while the C atoms of A5P and the active-site residues of 2cxp are in yellow, the C atoms of F6P and the active-site residues of 1hox are in orange and the C atoms of His176 from mjPHI are in gray. Residue names of bfAPI, 2cxp, 1hox and mjPHI are labeled in black, blue, green and red, respectively. (*a*) The active site of bfAPI. (*b*) The active site of 2cxp. (*c*) The active site of 1hox. (*d*) The active site of the mjPHI structure. (*e*) Superimposition of the active sites of bfAPI, 2cxp and mjPHI. His79 of bfAPI and His389 of 2cxp are structurally equivalent. The His186 side chain in bfAPI points away from the active site. Ser64 of bfAPI and Glu358 of 2cxp are located in structurally equivalent positions. (*f*) Superimposition of the active-site structures of bfAPI, 1hox and mjPHI. His79 of bfAPI and His388 of 1hox are structurally equivalent. Ser64 of bfAPI and Glu357 of 1hox are located in structurally equivalent positions. (*g*) A close-up view of (*e*). The C1 and C2 atoms of A5P in 2cxp are approximately 3.3 and 2.7 Å away from the $N^{\varepsilon 2}$ atom of His176 in mjPHI. If His186 in bfAPI were to adopt a similar side-chain conformation as His176 of mjPHI, it would be placed close to the C1 and C2 atoms of A5P, a position suitable for catalysis. (*h*) A close-up view of (*f*). The O5 atom of F6P in 1hox is approximately 2.5 Å away from $N^{\delta 1}$ of His79 of bfAPI. His79 is in a position suitable for catalyzing the ring-opening step. CMP-Kdo was omitted for clarity.

glucose 6-phosphate to fructose 6-phosphate. To date, 59 structures of bacterial and eukaryotic PGIs are available in the Protein Data Bank, 52 of which have ligands bound at their active site. These PGIs range in size from 450 to 600 residues. However, all have similar structures, with differences mostly observed at the N- and C-termini and in the loop regions. PGI is active as a dimer and the active site is located at the dimer interface. Each PGI subunit consists of two SIS domains conventionally named the large domain (residues 1–98 and 294–514) and the small domain (residues 99–293), both of which share structural similarity to bfAPI. A superimposition of bfAPI onto the large and the small domains of rabbit PGI–5PAA (PDB entry 1g98; Jeffery *et al.*, 2001) results in an r.m.s.d. of 2.9 Å for 168 C^{α} atoms for the large domain and an r.m.s.d. of 2.4 Å for 132 C^{α} atoms for the small domain (Fig. 5*a*). The active-site structure of API is more similar to that of the rabbit PGI small domain. When a bfAPI subunit is superimposed onto the large and small domains of rabbit PGI separately, the Ser/Thr-rich loop in bfAPI superimposes well with the equivalent loop from the small domain of rabbit PGI, whereas the equivalent loop in the large domain is much longer. This longer loop extends through the space that is occupied by the Ser/Thr-rich loop and the CMP-Kdo molecule in bfAPI; therefore, it would block access of substrates to this site and thus prevent it from becoming an active site. Furthermore, the highly conserved histidine residue (His388 from a neighboring subunit in rabbit PGI–5PAA; PDB entry 1g98) that has been proposed to be responsible for sugar ring opening in PGI is only present in the active site of the small domain and not in that of the large domain. Therefore, PGI only has two active sites per dimer, despite it being comprised of four SIS domains as in API, which has four active sites.

Since both bfAPI and PGI catalyze isomerization reactions with similar phosphosugar substrates (a five-carbon *versus* a six-carbon phosphosugar), their active sites are expected to share some structural similarities for substrate binding and catalysis; however, each must possess unique features that define substrate specificity. For comparison of the active sites, a total of 16 PGI structures were selected from the PDB and structures containing ligands similar to A5P or Ru5P were preferably selected. The source organisms, the bound substrate/product analogues and the PDB codes of these PGI proteins are as follows: *Mus musculus*, A5P, 2cxp; *M. musculus*, glucose 6-phosphate (G6P, ring form), 1u0f; *Oryctolagus cuniculus* (rabbit), fructose 6-phosphate (F6P, ring form), 1hox; *O. cuniculus*, 5-phosphoarabinonate (5PAA; Fig. 1*b*), 1g98; *Sus scrofa*, 5PAA, 1gzv; *Homo sapiens*, 5PAA, 1nuh; *Leishmania mexicana*, F6P, 1t10; *Pyrobaculum aerophilum*, G6P, 1x9i; *Trypanosoma brucei*, G6P, 2o2c; *Francisella tularensis* subsp., F6P (ring form), 3m5p; *F. tularensis* subsp., 6-phosphogluconate (6PGA), 3q7i; *Thermotoga maritima*, 2q8n; *Mycobacterium tuberculosis*, 2wu8; *Staphylococcus aureus* subsp., 3ff1; *Vibrio cholerae*, 3hjb; *Bacillus anthracis*, 3ifs. All bound sugar substrate/product analogues are in the open-chain conformation if not otherwise specified and the last five structures do not contain any substrate/product analogues. With the exception of α8 and the C-terminal tail, the bfAPI structure superimposes well with the small domains of these PGIs. The active site, including the Ser/Thr-rich region and His79, is structurally conserved (Fig. 5*b*). In PGI, the conserved histidine residue in the active site that is equivalent to His79 of bfAPI also comes from a neighboring subunit. The phosphosugar ligands in these PGI structures all have similar binding modes, with the phosphate groups interacting with the Ser/Thr-rich region (phosphate-binding loop). In bfAPI, the phosphate moiety of CMP-Kdo also interacts with the Ser/Thr-rich region. The consensus sequence of this Ser/Thr-rich region among PGIs and APIs is S-*X*-S/T-*X*-*X*-S/T.

Previous studies of the rabbit PGI enzyme identified the key catalytic residues His388, Glu357, Lys518 and Arg272 (Jeffery *et al.*, 2001; Lee *et al.*, 2001). His388 catalyzes the ring-opening step and Glu357 is the general acid/base catalyst for the isomerization reaction; the reaction proceeds through a *cis*-enediol intermediate. Both His388 and Glu357 are highly conserved in PGI. His388 is structurally equivalent to His79 of bfAPI, but Glu357 does not have a structural equivalent in bfAPI; the equivalent residue is Ser64 and this serine is strictly conserved among APIs. Glutamate has a favorable pK_a for functioning as a general acid/base and the longer side chain of glutamate can easily access the C1 and C2 atoms of the bound phosphosugar ligand for catalysis. On the other hand, serine, with a side-chain pK_a of $\sim$13, is not a good candidate for general acid/base catalysis. Arg272 and Lys518 of rabbit PGI are absent in bfAPI, which is expected since these residues are located in the C-terminal region that differs significantly between bfAPI and rabbit PGI. These structural differences are likely to contribute to the differences in substrate specificity between API and PGI. PGI is specific for G6P and F6P, whereas API is specific for Ru5P and A5P and cannot utilize similar molecules such as D-ribose 5-phosphate, D-arabinose, D-glucose, D-glucose 6-phosphate, D-mannose 6-phosphate or D-glucosamine 6-phosphate (Meredith & Woodard, 2003, 2006).

If API does not use a glutamate as a general acid/base for catalysis, what residue could act in this role? Of the 16 PGI structures compared here, two contain ligands most relevant to API. PGI from *M. musculus* (PDB entry 2cxp) has a bound A5P in its open-chain form and rabbit PGI (PDB entry 1hox) has an F6P bound in its ring form. The active-site structures and key residues in bfAPI, 2cxp and 1hox superimpose well (Figs. 6*e* and 6*f*). Because of these similarities, once could postulate that the substrate-binding mode in bfAPI is similar. If A5P adopts the same binding mode in both 2cxp and bfAPI, C1 and C2 of A5P would be located close to His186 if His186 were to adopt a side-chain conformation similar to His176 of mjPHI, which points towards A5P (Fig. 6*g*). His186 is therefore a potential candidate for a general acid/base catalyst. This hypothesis is supported by several observations. Firstly, His186 is highly conserved among APIs and is structurally equivalent to His193 of ecAPI (Gourlay *et al.*, 2010) and His176 of mjPHI (Martinez-Cruz *et al.*, 2002), which are both critical in catalysis. Furthermore, biochemical studies have shown that the activity of APIs is inhibited by the divalent ions

Zn^{2+}, Cd^{2+} or Hg^{2+} and that this activity can be restored by the addition of EDTA (Mosberg *et al.*, 2011; Meredith & Woodard, 2006). In bfAPI, both His79 and His186 are ideal ligands for binding to d^{10} transition-metal ions such as Zn^{2+}, Cd^{2+} or Hg^{2+}. Therefore, we suggest that His186 plays an important catalytic role in the isomerization reaction catalyzed by bfAPI. Superimposition of the bfAPI and rabbit PGI–F6P structures (PDB entry 1hox) indicates that His79 of bfAPI is structurally equivalent to the catalytic His388 of 1hox (Figs. 6*f* and 6*h*), which is involved in the ring-opening step of the reaction to yield the open-chain form of the substrate. The distance between the $N^{\delta1}$ atom of His79 and the O5 atom of F6P is ∼2.5 Å (Fig. 6*h*), thus positioning His79 at a suitable distance for catalysis. Since His79 is conserved amongst API and PGI, it is possible that it adopts a similar catalytic role as His388.

4. Conclusion

The high-resolution crystal structure of bfAPI with bound CMP-Kdo enabled the identification of key active-site residues. The structural analysis presented here along with the recent enzymatic assays (Cech *et al.*, 2014) confirms that CMP-Kdo does inhibit bfAPI and is likely to serve as a feedback inhibitor. It could therefore serve as a template or lead compound for the development of potent API inhibitors. Structure and sequence comparisons suggest that two highly conserved residues, His79 and His186, play important catalytic roles in the isomerization reaction. Tyr190 could be involved in the regulation of enzyme activity in the single-domain APIs, which is the role that is played by the CBS domain in multi-domain APIs. It would therefore be interesting to further investigate the possible regulatory role of Tyr190 in bfAPI. Furthermore, the structure of a multi-domain API would provide valuable insight into the mechanism by which the CBS domain putatively regulates the activity of the SIS domain. Additional information about bfAPI is available from TOPSAN (Krishna *et al.*, 2010; Weekes *et al.*, 2010) at http://www.topsan.org/explore?PDBid=3etn.

We thank the members of the JCSG high-throughput structural biology pipeline for their contribution to this work. This work was supported by the NIH NIGMS Protein Structure Initiative (U54 GM094586). Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences (including P41GM103393). Genomic DNA from *B. fragilis* NCTC 9343 (ATCC No. 25285D) was obtained from the American Type Culture Collection (ATCC). The contents of this publication are solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Badger, J. *et al.* (2005). *Proteins*, **60**, 787–796.
Cech, D., Wang, P. F., Holler, T. P. & Woodard, R. W. (2014). *J. Bacteriol.* **196**, 2861–2868.
Cer, R. Z., Mudunuri, U., Stephens, R. & Lebeda, F. J. (2009). *Nucleic Acids Res.* **37**, W441–W445.
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.
Cipolla, L., Polissi, A., Airoldi, C., Galliani, P., Sperandeo, P. & Nicotra, F. (2009). *Curr. Drug Discov. Technol.* **6**, 19–33.
Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 269–275.
Elsliger, M.-A., Deacon, A. M., Godzik, A., Lesley, S. A., Wooley, J., Wüthrich, K. & Wilson, I. A. (2010). *Acta Cryst.* F**66**, 1137–1142.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Finn, R. D., Clements, J. & Eddy, S. R. (2011). *Nucleic Acids Res.* **39**, W29–W37.
Gourlay, L. J., Sommaruga, S., Nardini, M., Sperandeo, P., Dehò, G., Polissi, A. & Bolognesi, M. (2010). *Protein Sci.* **19**, 2430–2439.
Heyes, D. J., Levy, C., Lafite, P., Roberts, I. S., Goldrick, M., Stachulski, A. V., Rossington, S. B., Stanford, D., Rigby, S. E. J., Scrutton, N. S. & Leys, D. (2009). *J. Biol. Chem.* **284**, 35514–35523.
Jeffery, C. J., Hardré, R. & Salmon, L. (2001). *Biochemistry*, **40**, 1560–1566.
Klock, H. E., Koesema, E. J., Knuth, M. W. & Lesley, S. A. (2008). *Proteins*, **71**, 982–994.
Krishna, S. S., Weekes, D., Bakolitsa, C., Elsliger, M.-A., Wilson, I. A., Godzik, A. & Wooley, J. (2010). *Acta Cryst.* F**66**, 1143–1147.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
Lee, J. H., Chang, K. Z., Patel, V. & Jeffery, C. J. (2001). *Biochemistry*, **40**, 7799–7805.
Lesley, S. A. *et al.* (2002). *Proc. Natl Acad. Sci. USA*, **99**, 11664–11669.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
Martinez-Cruz, L. A., Dreyer, M. K., Boisvert, D. C., Yokota, H., Martinez-Chantar, M. L., Kim, R. & Kim, S.-H. (2002). *Structure*, **10**, 195–204.
Meredith, T. C. & Woodard, R. W. (2003). *J. Biol. Chem.* **278**, 32771–32777.
Meredith, T. C. & Woodard, R. W. (2005). *J. Bacteriol.* **187**, 6936–6942.
Meredith, T. C. & Woodard, R. W. (2006). *Biochem. J.* **395**, 427–432.
Mosberg, J. A., Yep, A., Meredith, T. C., Smith, S., Wang, P. F., Holler, T. P., Mobley, H. L. & Woodard, R. W. (2011). *J. Bacteriol.* **193**, 2981–2988.
Pardee, A. B. (1971). *Harvey Lect.* **65**, 59–71.
Raetz, C. R. & Whitfield, C. (2002). *Annu. Rev. Biochem.* **71**, 635–700.
Santarsiero, B. D., Yegian, D. T., Lee, C. C., Spraggon, G., Gu, J., Scheibe, D., Uber, D. C., Cornell, E. W., Nordmeyer, R. A., Kolbe, W. F., Jin, J., Jones, A. L., Jaklevic, J. M., Schultz, P. G. & Stevens, R. C. (2002). *J. Appl. Cryst.* **35**, 278–281.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.
Terwilliger, T. C. (2003). *Acta Cryst.* D**59**, 38–44.
Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.
Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. (1993). *J. Mol. Biol.* **229**, 105–124.
Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. (2007). *Methods Mol. Biol.* **364**, 215–230.
Weekes, D., Krishna, S. S., Bakolitsa, C., Wilson, I. A., Godzik, A. & Wooley, J. (2010). *BMC Bioinformatics*, **11**, 426.
Weiss, M. S. & Hilgenfeld, R. (1997). *J. Appl. Cryst.* **30**, 203–205.

Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Winn, M. D., Murshudov, G. N. & Papiz, M. Z. (2003). *Methods Enzymol.* **374**, 300–321.
Yang, H., Guranovic, V., Dutta, S., Feng, Z., Berman, H. M. & Westbrook, J. D. (2004). *Acta Cryst.* D**60**, 1833–1839.
Yep, A., Sorenson, R. J., Wilson, M. R., Showalter, H. D., Larsen, S. D., Keller, P. R. & Woodard, R. W. (2011). *Bioorg. Med. Chem. Lett.* **21**, 2679–2682.