Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure and function of GlmU from *Mycobacterium tuberculosis*

Zhening Zhang,‡ Esther M. M. Bulloch, Richard D. Bunker, Edward N. Baker and Christopher J. Squire*

School of Biological Sciences and Maurice Wilkins Centre for Molecular Biodiscovery, University of Auckland, New Zealand

‡ Current affiliation: Biology Department, Brookhaven National Laboratory, Upton, NY, USA.

Correspondence e-mail: c.squire@auckland.ac.nz

Received 12 December 2008
Accepted 9 January 2009

PDB References: GlmU, GlcN-1-P complex, 2qkx, r2qkxsf; UDP-GlcNAc complex, 3d8v, r3d8vsf; unliganded, 3d98, r3d98sf.

Antibiotic resistance is a major issue in the treatment of infectious diseases such as tuberculosis. Existing antibiotics target only a few cellular pathways and there is an urgent need for antibiotics that have novel molecular mechanisms. The *glmU* gene is essential in *Mycobacterium tuberculosis*, being required for optimal bacterial growth, and has been selected as a possible drug target for structural and functional investigation. GlmU is a bifunctional acetyltransferase/uridyltransferase that catalyses the formation of UDP-GlcNAc from GlcN-1-P. UDP-GlcNAc is a substrate for two important biosynthetic pathways: lipopolysaccharide and peptidoglycan synthesis. The crystal structure of *M. tuberculosis* GlmU has been determined in an unliganded form and in complex with GlcNAc-1-P or UDP-GlcNAc. The structures reveal the residues that are responsible for substrate binding. Enzyme activities were characterized by ^{1}H NMR and suggest that the presence of acetyl-coenzyme A has an inhibitory effect on uridyltransferase activity.

1. Introduction

The general metabolic route for the incorporation of sugar units into structural polysaccharides involves the conversion of sugars to sugar nucleotides (Frey, 1996; Slawomir *et al.*, 2006). In prokaryotic cells, the sugar nucleotide UDP-*N*-acetyl-D-glucosamine (UDP-GlcNAc) is synthesized from fructose-6-phosphate (Fru-6-P) in four steps by three different enzymes (Slawomir *et al.*, 2006; Mengin-Lecreulx & van Heijenoort, 1993). Fructose-6-phophate is first converted to glucosamine-6-phosphate (GlcN-6-P) by D-fructose-6-phosphate aminotransferase (GlmS), utilizing L-glutamine. GlcN-6-P is then isomerized to glucosamine-1-phosphate (GlcN-1-P) by the enzyme phosphoglucosamine mutase. The last two steps of the reaction are catalyzed by the bifunctional enzyme GlmU, which has acetyltransferase and uridyltransferase activity, to form UDP-GlcNAc (Mengin-Lecreulx & van Heijenoort, 1994). In contrast, in eukaryotic cells each step of the pathway is catalyzed by a separate enzyme (Slawomir *et al.*, 2006).

The final product of GlmU, UDP-GlcNAc, is a substrate for two important biosynthetic pathways: peptidoglycan and lipid A synthesis (Slawomir *et al.*, 2006; Mengin-Lecreulx & van Heijenoort, 1993). Lipid A is the major component of the outer monolayer of the outer membranes of most Gram-negative bacteria and provides hydrophobic anchoring for lipopolysaccharides (Anderson & Raetz, 1987). Peptidoglycan polymers are cross-linked through short peptides and penicillin-binding proteins (PBPs), generating the net-like

three-dimensional structure of the cell wall (Walsh & Bugg, 1992; Brennan & Nikaido, 1995). Many antibiotics target this cell-wall synthesis machinery, including the β-lactams, vancomycin, fosfomycin, nisin and bacitracin (Walsh & Bugg, 1992). GlmU is also involved in the synthesis of the β-1,6-*N*-acetyl-D-glucosamine polysaccharide adhesin, which is required for biofilm formation in *Escherichia coli* and *Staphylococcus epidermidis*, and GlmU inhibitors have shown strong anti-biofilm activity on urinary catheters (Burton *et al.*, 2006). *E. coli* GlmU (EcGlmU) loses acetyltransferase activity in the absence of thiol-reducing agents (Pompeo *et al.*, 1998) and the best-known inhibitors of GlmU are thiol-specific reagents such as iodoacetamide (IDA) and N-substituted maleimides, which include *N*-ethylmaleimide (NEM), *N*-phenylmaleimide, *N*,*N*′-(1,2-phenylene)dimaleimide (oPDM) and *N*-(1-pyrenyl)-maleimide (PyrM) (Burton *et al.*, 2006).

The gene *glmU* has been identified as essential for optimal growth of *M. tuberculosis* (Sassetti *et al.*, 2003) and is not present in humans; hence, it is of interest as a drug-design target. To this end, the crystal structures of GlmU from *E. coli*, *Streptococcus pneumoniae* and *Haemophilus influenzae* have been determined (Brown *et al.*, 1999; Kostrewa *et al.*, 2001; Mochalkin *et al.*, 2007). These trimeric protein structures show a C-terminal left-handed β-helix domain that catalyzes the synthesis of *N*-acetylglucosamine-1-phosphate (GlcNAc-1-P) from GlcN-1-P and acetyl coenzyme A (acetyl CoA),

$$\text{GlcN-1-P} + \text{acetyl-CoA} \rightarrow \text{GlcNAc-1-P} + \text{CoA}.$$

The N-terminal pyrophosphatase (PPase) domain displays a Rossmann fold and catalyzes the formation of UDP-GlcNAc from GlcNAc-1-P (produced by the C-terminal domain) and UTP,

$$\text{GlcNAc-1-P} + \text{UTP} \rightarrow \text{UDP-GlcNAc} + \text{PP}_\text{i}.$$

Although GlmU has been crystallized from several species and in a variety of forms, no crystal structure is available with the substrate GlcNAc-1-P in the uridyltransferase active site, leading to uncertainty concerning the protein–substrate contacts and the chemical mechanism of the uridyltransferase reaction. Here, we report the structures of *M. tuberculosis* GlmU (MtGlmU) in an unliganded form and in complexes in which either GlcNAc-1-P or UDP-GlcNAc occupies the uridyltransferase active site. These structures identify the active-site contacts between protein and ligands and suggest a ternary-complex mechanism of action for the GlmU uridyltransferase reaction. Enzyme assays for both GlmU reactions commonly use radio-labelled substrates: [^{14}C]-acetyl-CoA for study of acetyltransferase activity and [^{14}C]-GlcNAc-1-P for study of uridyltransferase activity. Here, we describe a novel ^{1}H NMR methodology for characterizing the enzymatic activities of MtGlmU. The NMR assays suggest an inhibitory effect for acetyl-CoA on the acetyltransferase reaction.

2. Materials and methods

2.1. Expression and purification

The *glmU* gene was amplified from *M. tuberculosis* genomic DNA by PCR using the primers 5′-GGCAGCGGCGCGATGACGTTTCCTGGTGACAC-3′ and 5′-GAAAGCTGGGTGTCACGGTGTCTGATCAGC-3′. The construct was cloned into the pDEST17 N-terminal His_6-tag vector (Invitrogen) using Gateway cloning. Protein was overexpressed in *E. coli* BL21(DE3) cell strain using an autoinduction protocol (Studier, 2005). An overnight culture was prepared by inoculating 5 ml PA-0.5G medium supplemented with ampicillin (100 mg l^{-1}) and incubating at 310 K with shaking (160 rev min^{-1}). Overnight cultures were transferred into 1 l ZYP-5052 medium and incubated at 310 K with shaking at 160 rev min^{-1}. On reaching an OD_{600} of 0.7 (approximately 4 h), the cell culture was cooled to 277 K for 1 h. The culture was then incubated for approximately 18 h at 291 K with shaking at 160 rev min^{-1}. Cells were harvested by centrifugation at 5000*g* for 10 min at 277 K and resuspended in buffer *A* [25 m*M* NaH_2PO_4/Na_2HPO_4 pH 7.5, 100 m*M* NaCl, 14 m*M* β-mercaptoethanol (βME), 1 m*M* EDTA]. Cells were lysed by cell disruption (Constant Cell Disruption System) and the lysate was centrifuged at 16 000 rev min^{-1} for 25 min. The supernatant was applied onto an immobilized metal (Ni^{2+}) affinity column (IMAC) pre-equilibrated with buffer *A*. A linear imidazole gradient (0–1 *M*) was used to elute the fusion protein. The His_6 tag was cleaved by incubation with recombinant tobacco etch virus (rTEV) protease (overnight at 277 K), after which the rTEV protease and uncleaved protein were removed by a second IMAC step. The cleaved protein retains a Gly-Ser sequence (from the His_6 tag) at the N-terminus of the native GlmU sequence. For final purification, GlmU was concentrated to 20 mg ml^{-1} and then applied onto a size-exclusion column (Superdex S200 10/30, Pharmacia) equilibrated with buffer *A*, resulting in a single peak.

2.2. Crystallization and data collection

Purified GlmU at a concentration of 10 mg ml^{-1} in 25 m*M* NaH_2PO_4/Na_2HPO_4 pH 7.5, 100 m*M* NaCl, 14 m*M* βME and 1 m*M* EDTA was mixed with equal volumes of crystallization buffer from the Top67 crystallization screen (Page *et al.*, 2003) using a Cartesian nanolitre dispensing robot. The sitting-drop vapour-diffusion method was used for crystallizations with 0.2 µl drop volume and 50 µl reservoir volume. Native GlmU crystals grew overnight from 0.2 *M* magnesium formate and 20% PEG 3350. Crystals of GlmU complexed with UDP-GlcNAc were grown by cocrystallization with 10 m*M* UDP-GlcNAc, which was included in the protein solution. The crystals were obtained using 9% MPD, 16% ethylene glycol, 5% PEG 8000 and 0.1 *M* sodium cacodylate as precipitant solution. Crystals of GlmU complexed with GlcNAc-1-P were similarly grown by cocrystallization, with 10 m*M* GlcN-1-P and 10 m*M* acetyl-CoA included in the protein solution and 20% PEG 3350 and 0.2 *M* lithium nitrate as precipitant. The unliganded and GlcNAc-1-P complex crystals grew to maximum dimensions of 50 × 50 × 50 µm over three weeks and

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

	UDP-GlcNAc complex	GlcNAc-1-P complex	Unliganded
Data collection			
Space group	*H*32	*H*32	*H*32
Unit-cell parameters (Å)	$a = b = 113.99$, $c = 362.36$	$a = b = 94.30$, $c = 288.04$	$a = b = 94.28$, $c = 284.31$
Source	MM007-HR	MM007-HR	SSRL BL9-2
Wavelength	1.5418	1.5418	0.8500
Resolution	2.50 (2.64–2.50)	2.75 (2.90–2.75)	2.50 (2.64–2.50)
$R_{\rm merge}$†	0.104 (0.608)	0.147 (0.568)	0.088 (0.520)
Completeness	100 (100)	100 (100)	100 (100)
Unique reflections	31883	13225	16733
$\langle I/\sigma(I)\rangle$	15.1 (3.3)	11.9 (3.3)	15.3 (3.4)
Multiplicity	12.2 (11.7)	7.0 (7.0)	6.1 (6.2)
Wilson *B* factor (Å^2)	58.3	54.7	50.4
Refinement and model quality			
Resolution range	34.0–2.50	42.3–2.75	42.3–2.50
No. of reflections	28590	12588	15784
$R_{\rm cryst}$‡	0.195	0.197	0.195
$R_{\rm free}$‡	0.251	0.259	0.243
Total protein atoms	3491	2841	2883
Total ligand atoms	43	19	0
Total water atoms	163	39	62
Mean *B* factor (Å^2)			
Protein atoms	47.3	27.6	37.0
Ligand atoms	46.4	37.3	—
Water atoms	44.7	27.6	39.3
R.m.s. deviation from ideal values			
Bond lengths (Å)	0.017	0.016	0.018
Bond angles (°)	1.817	1.673	1.77
Ramachandran plot			
Favoured regions (%)	96.7	94.0	96.4
Outliers (%)	0	0.8	0
PDB code	2qkx	3d8v	3d98

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$. ‡ $R_{\rm cryst}$ and $R_{\rm free} = \sum ||F_{\rm obs}| - |F_{\rm calc}|| / \sum |F_{\rm obs}|$, where $R_{\rm free}$ was calculated over 5% of amplitudes that were chosen at random and not used in refinement.

were difficult to reproduce. In contrast, the UDP-GlcNAc complex crystals grew overnight to maximum dimensions of 500 × 250 × 250 µm and were easy to reproduce. The unliganded and GlcNAc-1-P complex crystals were dipped briefly in cryoprotectant (20% ethylene glycol in the reservoir solution) and were flash-frozen in liquid nitrogen prior to data collection. The UDP-GlcNAc complex crystals were directly flash-frozen in liquid nitrogen without any additional cryoprotection prior to data collection. X-ray diffraction data were collected at 110 K on a Rigaku MicroMax007-HR rotating-anode generator with Cu *Kα* radiation and MAR345 image-plate detectors, Osmic mirror focusing optics and an Oxford Cryostream system or on Stanford Synchrotron Radiation Laboratory beamline 9-2 equipped with a MAR325 CCD detector. Data-collection statistics are summarized in Table 1.

2.3. Structure determination and refinement

Data were processed in space group *H*32 using *MOSFLM* and *SCALA* (Collaborative Computational Project, Number 4, 1994). The UDP-GlcNAc complex structure was determined first by molecular replacement using *Phaser* (Read, 2001; Storoni *et al.*, 2004; McCoy *et al.*, 2005) with *S. pneumoniae* GlmU (PDB code 1hm9) as the search model. The top solution placed one molecule in the asymmetric unit and this model was completed by several cycles of manual building with *Coot* (Emsley & Cowtan, 2004) and refinement with *REFMAC* (Murshudov *et al.*, 1997). The UDP-GlcNAc molecular structure and geometrical restraints were derived using the Dundee *PRODRG*2 server (Schüttelkopf & van Aalten, 2004). A cacodylate ion was located on a threefold crystallographic rotation axis and refined at 33% occupancy. Solvent molecules were added by automatic peak picking from $F_{\rm o} - F_{\rm c}$ electron-density maps using *Coot*. Potential water molecules with electron-density peaks above 3σ were selected and were manually checked for appropriate hydrogen-bond geometry in *Coot*. The final structure was refined using data to 2.5 Å resolution and had an $R_{\rm cryst}$ of 19.5%, an $R_{\rm free}$ of 25.1% and an average *B* factor of 36.1 Å^2. The stereochemistry of the final structure was evaluated using *MOLPROBITY* (Davis *et al.*, 2007). The Ramachandran plot showed that 96.7% of residues are in the most favoured regions. The final model includes 479 residues (residues 1–479) of the 495 residues in the complete GlmU sequence. The last 16 C-terminal residues were not modelled owing to lack of density. Refinement statistics are summarized in Table 1. The other two GlmU structures were determined by molecular replacement using the fully refined GlmU–UDP-GlcNAc complex structure as a search model and were refined as above. Electron density was not observed for approximately 100 C-terminal residues in both structures and only residues 2–391 and 2–388 were present in interpretable density for the GlcNAc-1-P complex and unliganded structures, respectively. Refinement statistics are summarized in Table 1. Figures were prepared using *PyMOL* (DeLano, 2002), calculations of root-mean-square differences in atomic positions between different structures were made with *SSM* (Krissinel & Henrick, 2004) and interfaces were analysed with the EMBL–EBI *PISA* server (Krissinel & Henrick, 2007).

2.4. Enzyme-activity assays by ^{1}H NMR

All chemicals were purchased from Sigma. Reactions were carried out in 25 m*M* potassium phosphate buffer pH 7.5, 50 m*M* NaCl, 0.5 m*M* $MgCl_2$ and 3 m*M* βME. The acetyltransferase reactions (50 µl) contained 40 m*M* glucosamine-1-phosphate, 40 m*M* acetyl-CoA and 4 µ*M* GlmU. The uridyltransferase reactions (50 µl) contained 40 m*M* *N*-acetylglucosamine, 40 m*M* UTP and 4 µ*M* GlmU. Reaction mixtures

were incubated at 310 K for 45 min, diluted to 1 ml with NMR buffer (25 m*M* potassium phosphate pH 7.5, 50% D_2O) and were stored on ice before being brought to room temperature 30 min before NMR spectra were recorded. For assays in the absence of a reducing agent or the presence of a thiol-modifying agent, βME was first removed from GlmU by dialysis. To determine the effect of a thiol-modifying agent, GlmU was incubated for 30 min with 1 m*M* *N*-ethylmaleimide prior to the assay. GlmU activity was monitored using one-dimensional ^{1}H NMR spectroscopy. Spectra of reaction solutions were acquired on a DRX 400 spectrometer (Bruker) at 300 K using presaturation of the water resonance (Bax, 1985). Data were processed using the software *XWINNMR* 3.0 and *TopSpin* 1.3 (Bruker).

3. Results and discussion

3.1. Assays of MtGlmU enzymatic activity by ^{1}H NMR

NMR spectroscopy was used to assay the enzyme activity of MtGlmU. The substrates GlcN-1-P and GlcNAc-1-P and the product UDP-GlcNAc can be readily distinguished from one-dimensional ^{1}H NMR spectra in the region 5.1–5.7 p.p.m. The chemical shifts for many of the protons of the substrates and product are similar. However, the doublet of doublets corresponding to the proton on the anomeric C atom of the glucose moiety is distinctive for each compound. Commercial preparations of the compounds were used to determine that the chemical shift of this proton is 5.55, 5.27 and 5.42 p.p.m. for GlcN-1-P, GlcNAc-1-P and UDP-GlcNAc, respectively, under the conditions of the NMR assay. This signal was used to detect the presence and the relative amounts of substrate and product for each reaction as described below.

3.1.1. Acetyltransferase activity. On incubation of GlcN-1-P and acetyl-CoA with MtGlmU there was a decrease in the signal from the anomeric proton of GlcN-1-P at 5.55 p.p.m. and the appearance of a peak at 5.27 p.p.m. corresponding to GlcNAc-1-P (Fig. 1*a*), confirming acetyltransferase activity. In a control reaction incubated under the same conditions but without MtGlmU, no changes were seen in the spectrum. The acetyltransferase reaction was repeated without $MgCl_2$ and βME in the enzyme buffer and reaction solution. Under these conditions the enzyme is still able to convert GlcN-1-P to GlcNAc-1-P. The reaction was also carried out in the presence of the thiol-modifying reagent *N*-ethylmaleimide (NEM) and again no affect on enzyme activity was observed.

We conclude that MtGlmU has acetyltransferase activity in the absence of reducing agent and in the presence of a thiol-reactive reagent. This is in marked contrast to EcGlmU, for which the presence of a reducing agent is required for acetyltransferase activity (Pompeo *et al.*, 1998). Spontaneous inactivation in the absence of reducing agent or inactivation by thiol-specific agents, particularly those with bulky substituents, could occur in the acetyltransferase active site of EcGlmU, specifically at Cys385, which is located within 6 Å of the acetyl-CoA sulfur. Incorporation of bulky thiol-reactive reagents at this site would prevent binding of acetyl-CoA. It is also conceivable that in the absence of reducing agent *in vitro* acetyl transfer may occur from acetyl-CoA to Cys385, blocking further cofactor binding and inhibiting acetyltransferase function. The three other cysteines within the EcGlmU structure are inaccessible to solvent. Uropathogenic bacteria on urinary catheters are inhibited by various NEM analogues (Burton *et al.*, 2006) and in each case the GlmU acetyltransferase active site contains at least one free cysteine located near the acetyl-CoA sulfur. The MtGlmU structure does not contain a free cysteine in the acetyl-CoA binding site nor any solvent-accessible cysteines elsewhere in the structure.

3.1.2. Uridyltransferase activity. On incubating GlcNAc-1-P and UTP with MtGlmU, a doublet of doublets appeared at 5.42 p.p.m. corresponding to the anomeric proton of UDP-GlcNAc and there was a decrease in the signal from GlcNAc-1-P (Fig. 1*b*). This indicates the turnover of GlcNAc-1-P to UDP-GlcNAc and confirms the uridyltransferase activity of MtGlmU. In a control reaction incubated without MtGlmU no change in the spectrum was observed. The uridyltransferase reaction was repeated under the same conditions but without $MgCl_2$ in the enzyme buffer and reaction solution. In this case turnover to the UDP-GlcNAc product did not occur, confirming that MtGlmU has an absolute requirement for Mg^{2+} for uridyltransferase activity.

3.1.3. Overall reaction. In an attempt to couple together the acetyltransferase and uridyltransferase activities of MtGlmU, GlcN-1-P, acetyl-CoA and UTP were incubated with the enzyme. A peak at 5.27 p.p.m. corresponding to the anomeric proton of GlcNAc-1-P was observed, but there was no signal evident at the chemical shift associated with UDP-GlcNAc. Hence, only the initial acetyltransferase reaction was detected and not the second uridyltransferase reaction.

The system was further characterized by incubating the enzyme with acetyl-CoA or GlcN-1-P before addition to the solution containing the uridyltransferase substrates (GlcNAc-1-P and UTP). The uridyltransferase activity was not affected by the presence of GlcN-1-P, whereas in the presence of acetyl-CoA no UDP-GlcNAc was detected. This apparent inhibition of uridyltransferase activity by acetyl-CoA could be the consequence of an allosteric interaction between the active sites. The structural basis of this potential allostery is not apparent from previous crystal structure information nor from the MtGlmU crystal structures described in full below.

3.2. Overall structures

Three crystal structures of *M. tuberculosis* GlmU were determined: an unliganded structure at 2.5 Å resolution, a complex with UDP-GlcNAc at 2.5 Å resolution and a complex with GlcNAc-1-P at 2.75 Å resolution. All three structures belong to space group *H*32 but with varying unit-cell parameters. The UDP-GlcNAc structure was determined first using molecular replacement and displays unit-cell parameters $a = b = 113.99$, $c = 362.26$ Å. The asymmetric unit contains one molecule comprising residues 1–475, with only the final 16 residues (480–495) missing, giving a solvent content of 71% and a V_M value of 4.20 Å^3 Da^{-1} (Matthews, 1968). The

structure shows a GlmU monomer folded into two distinct domains (Fig. 2*a*): an N-terminal domain (residues 2–262) which has a typical uridyltransferase fold based on a dinucleotide-binding Rossmann fold and a C-terminal domain (residues 263–478) which has a left-handed β-helix (LβH) fold that is characteristic of a number of acetyltransferase enzymes. The unliganded and GlcNAc-1-P structures were determined by molecular replacement using the UDP-GlcNAc coordinates as a search model. Both of these structures contained a single truncated molecule in the asymmetric unit: the final ~100 C-terminal residues of each molecule are missing, presumably owing to proteolysis during crystallization. The truncations result in reduced unit-cell parameters, with the unliganded structure having $a = b = 94.28$, $c = 284.31$ Å and the GlcNAc-1-P complex structure having $a = b = 94.30$, $c = 288.04$ Å. The solvent content and V_M of these truncated structures are 55% and 2.76 Å^3 Da^{-1}, respectively, for the unliganded structure and 56% and 2.80 Å^3 Da^{-1}, respectively, for the GlcNAc-1-P structure.

Dynamic light scattering (data not shown) and analytical gel filtration show that MtGlmU is trimeric in solution and in the three crystal structures this trimer is produced by a crystallographic threefold axis parallel to the long axis of the monomer (Fig. 2*b*). The *S. pneumoniae*, *E. coli* and *H. influenzae* GlmU structures display the same trimer and show root-mean-square differences (r.m.s.d.) in C$^\alpha$ atomic positions when compared with MtGlmU in the range 1.7–2.5 Å for about 1200 C$^\alpha$ atoms. The trimer is the biologically functional unit, with the C-terminal acetyltransferase active site containing contributions from all three subunits. The trimer interface is dominated by contacts between the C-terminal β-helical domains of the three molecules and buries a total surface of 6320 Å^2 in the truncated unliganded and GlcNAc-1-P structures and of 12 600 Å^2 in the full-length UDP-GlcNAc structure. Although there are some local differences between the three MtGlmU structures, the overall r.m.s.d. for the trimers of the three structures, as calculated by *SSM*, ranges from 0.6 to 0.8 Å over 1162 C$^\alpha$ atomic positions and in our discussion below the three structures are treated as equivalent; only significant and unique features of each structure are highlighted.

3.3. The left-handed β-helical (acetyltransferase) domain

The C-terminal domain comprises ten repeats (ten coils) of about 18 residues, forming a regular left-handed β-helix. Each turn of this domain displays a triangular cross-section with a

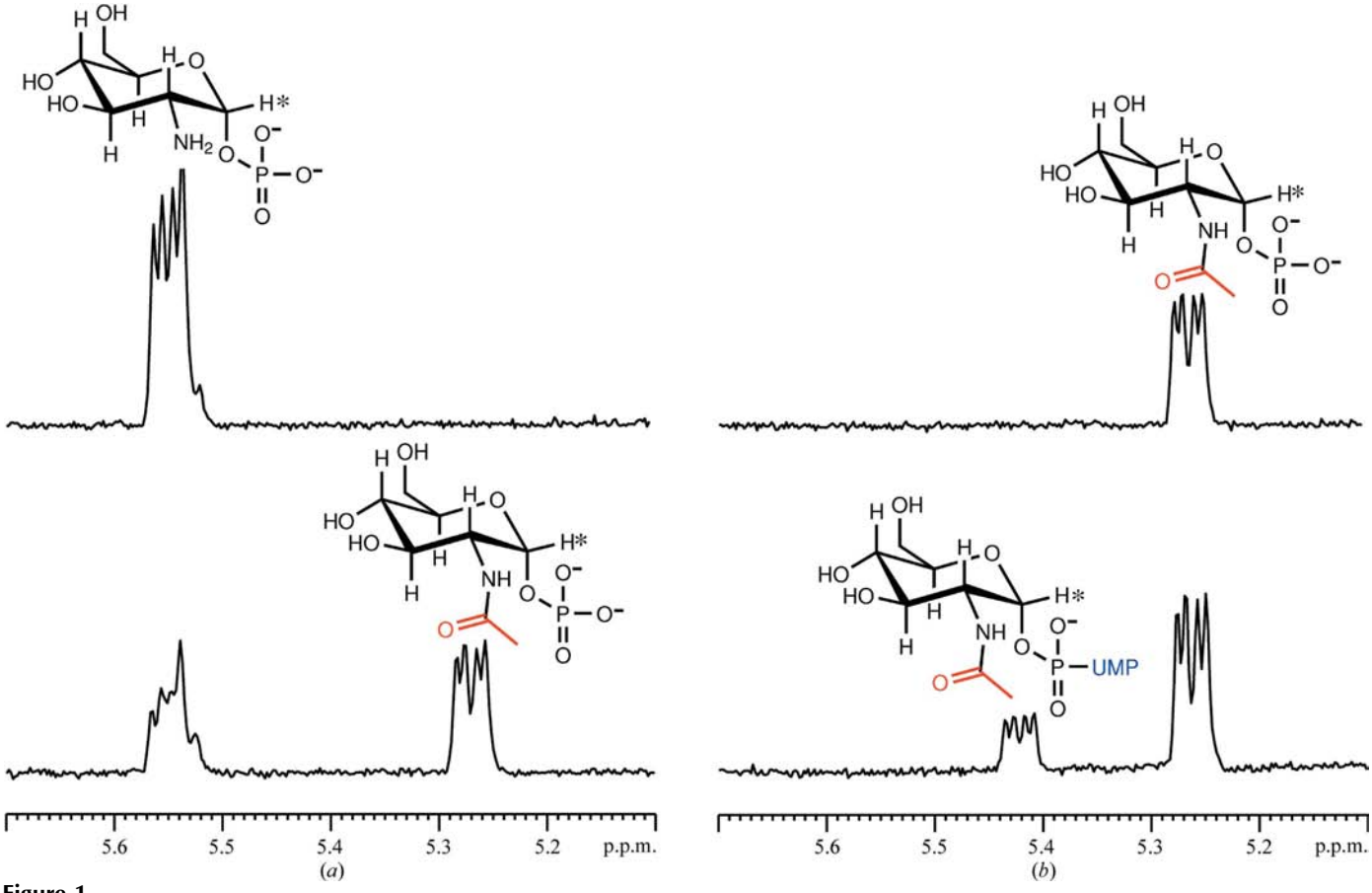

Figure 1
Analysis of the reactions catalysed by MtGlmU using ^{1}H NMR spectroscopy. The change in the chemical shift of the peak corresponding to the anomeric proton of GlcN-1-P (indicated by the asterisks) upon acetylation and uridylation was followed. For GlcN-1-P, GlcNAc-1-P and UDP-GlcNAc the chemical shift of this proton was 5.55, 5.27 and 5.42 p.p.m., respectively. (*a*) Acetyltransferase reaction: GlcN-1-P and acetyl CoA were incubated alone (top) and in the presence of MtGlmU (bottom). (*b*) Uridyltransferase reaction: GlcNAc-1-P and UTP were incubated alone (top) and in the presence of MtGlmU (bottom).

diameter of about 17 Å and is formed by three β-strands with a distinctive repetitive pattern of hydrophobic residues. The interior of the helix is predominantly hydrophobic, with stacks of aliphatic or aromatic residues. The result is a long narrow channel coincident with the helix axis, measuring 2–3 Å in diameter from which solvent is excluded. The UDP-GlcNAc structure contains a cacodylate ion located on the threefold trimer axis and coordinated by Arg323 and its two symmetry-related equivalents.

All attempts to cocrystallize or soak crystals of MtGlmU with acetyl-CoA, CoA or analogues were unsuccessful, but given the high degree of sequence and structural similarity of known GlmU proteins these molecules should bind in a similar way to that seen in the other GlmU structures, *i.e.*

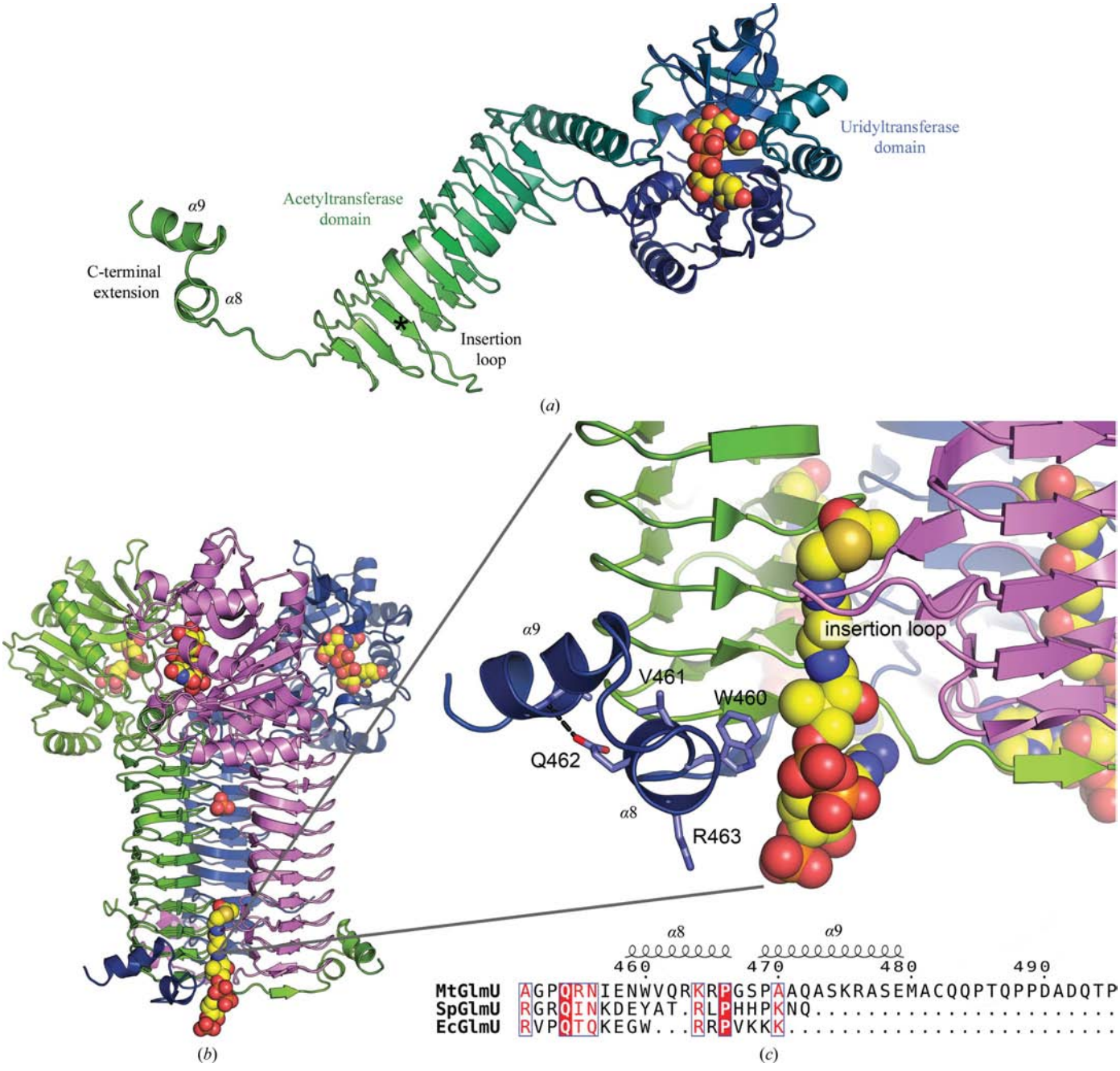

Figure 2
Structure of *M. tuberculosis* GlmU. (*a*) Monomer of the UDP-GlcNAc complex. The protein is shown as a cartoon model and is colour-ramped from blue (N-terminus) to green (C-terminus). The UDP-GlcNAc and cacodylate molecules are drawn as space-filling models. The site of C-terminal truncation in the unliganded and GlcNAc-1-P structures is indicated by an asterisk. (*b*) The trimeric structure of MtGlmU as produced by applying threefold crystallographic symmetry. Acetyl-CoA is modelled into the acetyltransferase active site by comparison to an *E. coli* GlmU structure and is drawn as a space-filling model. (*c*) Close-up view of the acetyltransferase active site. The active site is formed between two adjacent LβH domains, the insertion loop and the domain-swapping C-terminal extension from the third molecule of the trimer. A structure-based sequence alignment is shown for *M. tuberculosis*, *S. pneumonia* and *E. coli* GlmU proteins. MtGlmU has an additional 24 amino acids in its C-terminal extension, some of which form helix α9 and additional interactions with the LβH domain of one monomer.

within the trimeric LβH assembly between pairs of adjacent domains (Olsen & Roderick, 2001; Sulzenbacher *et al.*, 2001; Olsen *et al.*, 2007). The cofactor-binding site is delineated by the outer faces of two adjacent LβH domains, including an external loop that interrupts the eighth helical coil of one domain and by the domain-swapped C-terminal tail of the third molecule of the trimer (Fig. 2*c*). The external loop of the eighth helical repeat (residues 395–405) is the site of truncation in the unliganded and GlcNAc-1-P structures. The propensity of this site for proteolytic cleavage may be suggestive of a regulatory role; MtGlmU acetyltransferase activity may be downregulated by protein degradation under proteolytic conditions or in response to a cellular signal.

The final part of the MtGlmU C-terminal extension comprises two α-helices, α8 (459–466) and α9 (469–479), which are linked by a series of β-turns (Fig. 2*c*). This arrangement buries residues Trp460, Val461, Ala470 and Ala473 both internally within the helix–turn–helix section and within the C-terminal extension–LβH interface. An additional interaction between Gln462 OE1 and the peptide N atom of Ser474 further links and stabilizes the topology of these final two helices. As revealed in a multiple structure-based sequence alignment (Fig. 2*c*), the domain-swapping C-terminal extension is longer than in the *E. coli* and *S. pneumoniae* GlmU structures and provides additional stability to the biological trimer and the acetyl-CoA-binding site. In the α8 region of the domain-swapping arms, the structures of all three species contain arginine or lysine residues equivalent to MtGlmU Arg463 that bind the 3′ ribosyl phosphate of acetyl-CoA; beyond this the structures diverge.

3.4. Uridyltransferase domain

The MtGlmU uridyltransferase-domain structure is very similar to that reported for *S. pneumoniae* (SpGlmU), with an overall r.m.s.d. in 228 C^{α} positions of 1.1 Å between the pair of UDP-GlcNAc structures. The unliganded MtGlmU structure displays a more closed conformation in the GlcNAc-binding site than in any other known unliganded GlmU structure. Analysis of the crystal packing in the MtGlmU structure suggests that an open structure in the absence of substrate is disfavoured by intermolecular crystal contacts. The binding mode for UDP-GlcNAc in the active site of MtGlmU matches that seen for SpGlmU very closely, with the UDP-GlcNAc molecule, the final product of the reaction, making the same contacts to conserved amino-acid residues. When the two protein structures are superimposed, the UDP-GlcNAc molecules superimpose equally well, with an r.m.s.d. of 0.84 Å for the 39 ligand atoms. The only notable difference is that no Mg^{2+} ion is observed in the MtGlmU active site.

The novel GlcNAc-1-P structure unambiguously defines the binding site for this ligand and identifies an extensive network of hydrogen bonds between its various oxygen groups and residues of the active-site cavity (Fig. 3*a*). Thus, the acetyl O atom (O7′) of the *N*-acetyl group forms a hydrogen bond to Thr89 OG1 (2.9 Å) and the glucose O3 hydroxyl group forms hydrogen bonds to Glu166 OE2 (2.4 Å) and the side-chain amido N atom (ND2) of Asn181 (2.8 Å). The O4 hydroxyl also forms two hydrogen bonds to the peptide NH of Gly151 (2.9 Å) and the peptide O of Asn181 (2.6 Å). A phosphate O atom forms a water-mediated hydrogen bond to Asp114. It is possible that the water molecule involved in this interaction is actually an Mg^{2+} cation that would effectively balance the negative phosphate charge; however, the bond distances and geometry do not allow us to distinguish between a water and Mg^{2+} in this site. The carboxyl O atoms (OE1 and OE2) of Asp114 also hydrogen bond to Ser112 OG and Lys22 NZ, respectively. All these residues are highly conserved within the *M. tuberculosis*, *E. coli*, *H. influenzae* and *S. pneumoniae* GlmU sequences.

The GlcNAc-1-P substrate clearly binds in a very similar way to the UDP-GlcNAc

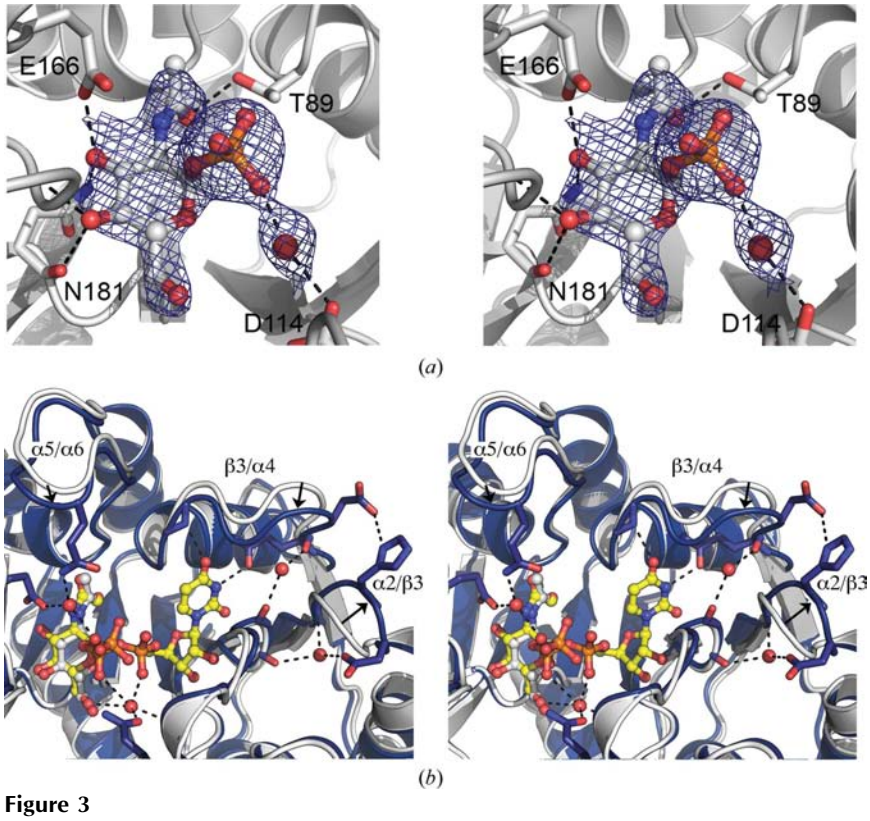

Figure 3
The uridyltransferase active site. (*a*) Binding of GlcNAc-1-P (ball-and-stick model). Hydrogen bonds are shown as dashed lines and the $2F_o - F_c$ electron density is contoured at 1.0σ. (*b*) Comparison of GlcNAc-1-P (white outlined protein) and UDP-GlcNAc (blue protein) binding and associated loop movement within the uracil site. GlcNAc-1-P and UDP-GlcNAc molecules are drawn as ball-and-stick models with white- and yellow-coloured C atoms, respectively. Water molecules are drawn as red spheres and are associated with the UDP-GlcNAc protein model.

product (Fig. 3*b*), implying that it is the interactions made by the GlcNAc moiety that determine the binding mode. The uracil-binding site displays local differences in the loop regions of the three MtGlmU structures. The GlcNAc-1-P structure is very similar to the unliganded structure in this region, but movements are seen to accompany UDP binding in the UDP-GlcNAc complex (Fig. 3*b*). The $\beta3/\alpha4$ loop moves approximately 1 Å towards the uracil base, resulting in the formation of hydrogen bonds between Gln83 OE1 and uracil N3 and between Gly88 NH and the exocyclic O atom O4 of uracil. His58, which is located on the turn linking $\alpha2$ to $\beta3$, moves 3.4 Å away from the active site and forms a new hydrogen bond to Asp84 of the $\beta3/\alpha4$ loop. The $\alpha5/\alpha6$ loop on top of the GlcNAc site moves towards the ligand, driven by the formation of a water-mediated interaction between Gln205 NE2 and the β-phosphate of UDP-GlcNAc. Arg19, which is located at the tip of the $\beta1/\alpha1$ loop near the entrance of the active site, is highly conserved among dinucleotide sugar transferases and differences are seen in conformation between the two complex structures: it projects into the external solvent space in the GlcNAc-1-P structure but flips into the active site of the UDP-GlcNAc structure. Here, it interacts with the Asn239 side chain, which in turn directly binds both phosphates of UDP-GlcNAc.

The GlmU–GlcNAc-1-P complex structure is the final piece of information required to confirm the proposed uridyltransferase catalytic mechanism (Kostrewa *et al.*, 2001; Mochalkin *et al.*, 2008). During our cocrystallization of MtGlmU with GlcN-1-P and acetyl CoA, the two substrates of the first reaction, acetylation of GlcN-1-P occurred and the product (GlcNAc-1-P) was found to be located in the uridyltransferase active site. Since there is no obvious channel present between the two GlmU active sites, this implies that GlmU releases GlcNAc-1-P into solution for transfer to the other active site. When attempts were made to soak both UDP and GlcNAc-1-P into *H. influenzae* GlmU crystals, it was found that only UDP binds (Mochalkin *et al.*, 2007). This suggests that the order of substrate binding is UTP first and then GlcNAc-1-P.

To further investigate the structural basis of uridyltransferase catalysis by MtGlmU, UTP was modelled into the active site of the GlcNAc-1-P-bound structure. The phosphate group of GlcNAc-1-P closely matches the position of the β-phosphate in the UDP-GlcNAc structure and the α-phosphate of UTP mimics the α-phosphate of UDP-GlcNAc. Given the previous suggestion that the nucleotide triphosphate binds to the enzyme first, we propose similarly that UTP and Mg^{2+} bind to the uridyltransferase active site prior to GlcNAc-1-P binding in the adjacent pocket. When GlcNAc-1-P binds, its phosphate O atom initiates a nucleophilic attack on the α-phosphate of UTP and catalysis proceeds *via* a pentacoordinate intermediate at the α-phosphate of UTP stabilized by the Mg^{2+} ion. The final steps include inversion of stereochemistry at the α-phosphate, UTP hydrolysis and release of PP_i. The Mg^{2+} fulfils a number of roles in catalysis: to enforce appropriate geometry for catalysis through its octahedral coordination, to polarize the P—O bond of the α-phosphate, activating the phosphorus for nucleophilic attack, and to balance the negative charge of the transition state. Arg19 is also important in this mechanism, as shown by the fact that site-directed mutagenesis of the equivalent arginine in EcGlmU dramatically impairs uridyltransferase activity (Brown *et al.*, 1999). Our modelling suggests that Arg19 binds the β-phosphate and/or γ-phosphate of UTP prior to the reaction, stabilizing the phosphate charge and directing it away from the site to be occupied by GlcNAc-1-P, leaving the α-phosphate open to direct attack from GlcNAc-1-P. It could also aid in polarizing the UTP α–β phosphate bond for hydrolysis and the subsequent release of PP_i.

This work was funded by the Health Research Council of New Zealand and the Maurice Wilkins Centre for Molecular Biodiscovery through the Tertiary Education Commission of New Zealand. Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory, a national user facility operated by Stanford University on behalf of the US Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research and by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program and the National Institute of General Medical Sciences. The authors thank the beamline 9-2 staff at SSRL, particularly Dr Clyde Smith, for their support. The authors would like to thank Drs Victoria Money, Tom Caradoc-Davies and Jodie Johnston for helpful discussions.

References

Anderson, M. S. & Raetz, C. R. (1987). *J. Biol. Chem.* **262**, 5159–5169.
Bax, A. (1985). *J. Magn. Reson.* **65**, 142–145.
Brennan, P. J. & Nikaido, H. (1995). *Annu. Rev. Biochem.* **64**, 29–63.
Brown, K., Pompeo, F., Dixon, S., Mengin-Lecreulx, D., Cambillau, C. & Bourne, Y. (1999). *EMBO J.* **18**, 4096–4107.
Burton, E., Gawande, P. V., Yakandawala, N., Lovetri, K., Zhanel, G. G., Romeo, T., Friesen, A. D. & Madhyastha, S. (2006). *Antimicrob. Agents Chemother.* **50**, 1835–1840.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., Snoeyink, J., Richardson, J. S. & Richardson, D. C. (2007). *Nucleic Acids Res.* **35**, W375–W383.
DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* http://www.pymol.org.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Frey, P. A. (1996). *FASEB J.* **10**, 461–470.
Kostrewa, D., D'Arcy, A., Takacs, B. & Kamber, M. (2001). *J. Mol. Biol.* **305**, 279–289.
Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). *Acta Cryst.* D**61**, 458–464.
Mengin-Lecreulx, D. & van Heijenoort, J. (1993). *J. Bacteriol.* **175**, 6150–6157.
Mengin-Lecreulx, D. & van Heijenoort, J. (1994). *J. Bacteriol.* **176**, 5788–5795.

Mochalkin, I., Lightle, S., Narasimhan, L., Bornemeier, D., Melnick, M., Vanderroest, S. & McDowell, L. (2008). *Protein Sci.* **17**, 577–582.
Mochalkin, I., Lightle, S., Zhu, Y., Ohren, J. F., Spessard, C., Chirgadze, N. Y., Banotai, C., Melnick, M. & McDowell, L. (2007). *Protein Sci.* **16**, 2657–2666.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Olsen, L. R. & Roderick, S. L. (2001). *Biochemistry*, **40**, 1913–1921.
Olsen, L. R., Vetting, M. W. & Roderick, S. L. (2007). *Protein Sci.* **16**, 1230–1235.
Page, R., Grzechnik, S. K., Canaves, J. M., Spraggon, G., Kreusch, A., Kuhn, P., Stevens, R. C. & Lesley, S. A. (2003). *Acta Cryst.* D**59**, 1028–1037.
Pompeo, F., van Heijenoort, J. & Mengin-Lecreulx, D. (1998). *J. Bacteriol.* **180**, 4799–4803.
Read, R. J. (2001). *Acta Cryst.* D**57**, 1373–1382.
Sassetti, C. M., Boyd, D. H. & Rubin, E. J. (2003). *Mol. Microbiol.* **48**, 77–84.
Schüttelkopf, A. W. & van Aalten, D. M. F. (2004). *Acta Cryst.* D**60**, 1355–1363.
Slawomir, M., Gabriel, I. & Olchowy, J. (2006). *Yeast*, **23**, 1–14.
Storoni, L. C., McCoy, A. J. & Read, R. J. (2004). *Acta Cryst.* D**60**, 432–438.
Studier, F. W. (2005). *Protein Expr. Purif.* **41**, 207–234.
Sulzenbacher, G., Gal, L., Peneff, C., Fassy, F. & Bourne, Y. (2001). *J. Biol. Chem.* **276**, 11844–11851.
Walsh, T. C. & Bugg, T. D. H. (1992). *Nature Prod. Rep.* **2**, 199–215.