Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

A new crystal form of MshB from *Mycobacterium tuberculosis* with glycerol and acetate in the active site suggests the catalytic mechanism

Simon Gareth Broadley,[a] James Conrad Gumbart,[b] Brandon William Weber,[c] Mohlopheni Jackson Marakalala,[c] Daniel Jacobus Steenkamp[c] and Bryan Trevor Sewell[c,d]*

[a]Department of Molecular and Cell Biology, University of Cape Town, University Avenue, Rondebosch, Western Cape 7700, South Africa, [b]Biosciences Division, Argonne National Laboratory, Argonne, IL 60439, USA, [c]Department of Clinical Laboratory Sciences, University of Cape Town, Observatory, Western Cape 7700, South Africa, and [d]Institute of Infectious Diseases and Molecular Medicine, University of Cape Town, Observatory, Western Cape 7700, South Africa

Correspondence e-mail: trevor.sewell@uct.ac.za

Received 7 March 2012
Accepted 2 August 2012

PDB References: MshB, 4ewl

MshB, a zinc-based deacetylase, catalyses a step in the mycothiol biosynthetic pathway that involves the deacetylation of 1-*O*-(2-acetamido-2-deoxy-α-D-glucopyranosyl)-D-*myo*-inositol (GlcNAc-Ins), *via* cleavage of an amide bond, to 1-*O*-(2-amino-2-deoxy-α-D-glucopyranosyl)-D-*myo*-inositol (GlcN-Ins) and acetate. In this study, MshB was expressed, purified and crystallized. A new crystal form was encountered in 0.1 *M* sodium acetate, 0.2 *M* ammonium sulfate, 25% PEG 4000 pH 4.6. The crystals diffracted to 1.95 Å resolution and the resulting electron-density map revealed glycerol and the reaction product, acetate, in the active site. These ligands enabled the natural substrate GlcNAc-Ins to be modelled in the active site with some certainty. One acetate O atom is hydrogen bonded to Tyr142 and is located 2.5 Å from the catalytic zinc. The other acetate O atom is located 2.7 Å from a carboxylate O atom of Asp15. This configuration strongly suggests that Asp15 acts both as a general base catalyst in the nucleophilic attack of water on the amide carbonyl C atom and in its protonated form acts as a general acid to protonate the amide N atom. The configuration of Tyr142 differs from that observed previously in crystal structures of MshB (PDB entries 1q74 and 1q7t) and its location provides direct structural support for recently published biochemical and mutational studies suggesting that this residue is involved in a conformational change on substrate binding and contributes to the oxyanion hole that stabilizes the tetrahedral intermediate.

1. Introduction

The enzymes of the pathway leading to the synthesis of mycothiol (MSH; Fan *et al.*, 2009) and the enzymes involved in its recycling are potential drug targets, since mycothiol, which is used by Actinomycetes for defence against electrophilic toxins and oxidative stress, is not found in humans (Newton *et al.*, 2000).

The zinc-dependent enzyme *N*-acetyl-1-D-*myo*-inosityl-2-amino-2-deoxy-D-glucopyranoside deacetylase (MshB) catalyses a step in the mycothiol biosynthetic pathway that involves the deacetylation of 1-*O*-(2-acetamido-2-deoxy-α-D-glucopyranosyl)-D-*myo*-inositol (GlcNAc-Ins) to 1-*O*-(2-amino-2-deoxy-α-D-glucopyranosyl)-D-*myo*-inositol (GlcN-Ins) and acetate. The enzyme has a partial Rossmann fold (CATH node 3.40.50.10320) and an active site that resembles those of carboxypeptidase A and other metalloproteases (Hernick & Fierke, 2005). After mycothiol has

bound to an electrophile, GlcN-Ins is recycled through the cleavage of a similarly located amide bond, leading to the release of a mercapturic acid by the homologous enzyme mycothiol S-conjugate amidase (Mca; Newton *et al.*, 2008). The details of the detoxification system, its role in *Mycobacterium tuberculosis* and the possibility of intervention in the system leading to the creation of useful drugs are controversial and are the subject of current research (Dosanjh *et al.*, 2008; Hassan *et al.*, 2006; Xu *et al.*, 2011).

The structure of MshB was first described by Maynes *et al.* (2003) (PDB entry 1q74). The pentacoordinate zinc(II) in the active site was found to be liganded by His13, Asp16, His147 and two water molecules. A mechanism was proposed in which the tetrahedral transition state formed during amide hydrolysis would be stabilized by the positively charged Zn^{2+} and by the imidazolium side chain of His144. In addition, the homology of MshB to Mca (33% sequence identity) enabled a model of the first 180 residues of Mca to be built. This allowed the conservation of all of the active-site residues to be recognized, with the exception of Ser20 in MshB, which was replaced by a lysine in Mca. The structure of MshB was determined almost simultaneously by McCarthy *et al.* (2004) (PDB entry 1q7t), who crystallized the enzyme in the presence of β-octylglucoside (BOG), which was found to occupy a location near the catalytic zinc. Hydrogen bonding to the glucopyranoside ring of BOG was interpreted as being analogous to that occurring with the natural substrate. In particular, three conserved residues, Arg68, Asp95 and His144, were hydrogen bonded *via* their side chains to the 3-OH, 4-OH and 6-OH hydroxyl groups of BOG, respectively.

Huang & Hernick (2012) have recently explored the kinetics of MshB, in particular the pH dependence of MshB and the effect of the presence of the viscogens glycerol, sucrose and Ficoll. Measurement of the reaction rate as a function of pH showed a point of inflection at pH 10.5, indicating the involvement of an amino-acid side chain with this pK_a in the reaction. It was shown by mutagenesis to reflect the ionization of Tyr142, a residue that was observed in various locations in the six MshB chains that were visualized in the previous two crystal structures. Modelling suggested that the side chain of this residue could swivel around the $C^{\alpha}-C^{\beta}$ bond to interact with the amide carbonyl group of the substrate. The conclusion of the kinetics work was that Tyr142, a residue that had not previously been implicated in the mechanism, plays a major role in the catalysis. The suggested model is that Tyr142 is hydrogen bonded *via* its phenolic hydroxyl to the substrate during the hydrolysis reaction, thus aiding the polarization of the carbonyl bond of the acetyl group of the substrate. After the general base-assisted attack of a water molecule on the carbonyl C atom, it forms part of the oxyanion hole that stabilizes the negative charge on the O atom of the tetrahedral intermediate. Finally, restoration of the side chain to one of the positions observed in the crystal structures is proposed to aid product release. It is proposed that His144 acts as the general acid catalyst that protonates the amide N atom as it leaves the acetate (Huang & Hernick, 2012).

This work, in which acetate (one of the normal reaction products) is visualized in the active site, provides necessary structural evidence for the proposal that Tyr142 is indeed located in the predicted position (Huang & Hernick, 2012). This structure also rules out the possibility of His144 acting as a component of the oxyanion hole (Maynes *et al.*, 2003). Furthermore, the geometry that we observe in the active site strongly suggests that the protonated form of Asp15 fulfils the role of a general acid as originally suggested by Maynes *et al.* (2003) and rules out the proposal of Huang & Hernick (2012) that the general acid is His144. Furthermore, the structure reported here enables a detailed analysis of the binding of the natural substrate *via* hydrogen bonding to Arg68, Asp95 and His144 and identifies a change in the conformation of the side chain of Asp95 that stabilizes the substrate-binding loop in the absence of substrate. It also suggests that glycerol, which is found in well defined locations in the active site, may act as a competitive inhibitor and not primarily as a viscogen as was previously suggested.

2. Materials and methods

2.1. Assay

The assay for MshB activity was performed as described by Gammon *et al.* (2010).

2.2. Expression and purification of MshB

The gene encoding MshB, Rv1170, was amplified from *M. tuberculosis* genomic DNA and ligated into the pET17b (Novagen) expression plasmid using the *Nde*I and *Bam*HI restriction sites. The forward primer was 5′-GCCCATATG-GTGTCTGAGACGCCGCG-3′ and the reverse primer was 5′-GTTGGGATCCTACGTGCCGGACGCG-3′. The recombinant plasmid expressed residues 1–303 of the wild-type sequence without modifications.

Escherichia coli BL21 (DE3) pLysS cells were transformed with the plasmid and induced with 0.4 m*M* IPTG at an OD_{600} of 0.6. The cells were incubated for 3 h at 291 K, harvested by centrifugation and the pellet was stored at 193 K. Thawed cells (3.5 g) were lysed by sonication in lysis buffer (20 ml) consisting of 50 m*M* NaCl and 1.0 m*M* each of the protease inhibitors TPCK, TLCK and PMSF in 50 m*M* HEPES pH 7.5. The sonicate was clarified by centrifugation at 15 000*g* for 30 min at 277 K and the supernatant was diluted by the addition of one volume of 25 m*M* Tris–HCl pH 8.0. It was then applied onto a DEAE-Sepharose (GE Healthcare) column that had been equilibrated with 25 m*M* NaCl, 25 m*M* Tris–HCl pH 8.0 and was eluted with a salt gradient ranging from 25 to 600 m*M*. The active fractions eluted as a sharp well separated peak at approximately 300 m*M* NaCl. Fractions that displayed activity were pooled, concentrated fourfold by ultrafiltration and dialysed overnight against 50 m*M* potassium phosphate buffer pH 7.0 containing 500 m*M* NaCl. The enzyme preparation was then applied onto a Zn-IMAC column prepared by equilibrating iminodiacetic acid agarose (Sigma) resin with 50 m*M* $ZnSO_4$. The column was

pre-equilibrated with 1 m*M* imidazole in 50 m*M* potassium phosphate buffer pH 7.0 containing 500 m*M* NaCl. The column was eluted with a 1–20 m*M* imidazole gradient in this buffer. The purpose of this procedure was to remove any inactive protein that did not contain Zn^{2+}. The active fractions did not adhere to the column and eluted immediately. The sample was concentrated fourfold by ultrafiltration. It was then applied onto a Sephacryl 300 (GE Healthcare) column and eluted with 50 m*M* potassium phosphate buffer pH 7.0. Glycerol was added to the fractions containing MshB to a concentration of 20% and these were stored at 253 K.

The three chromatographic steps resulted in a 28-fold increase in specific activity. The final eluent peak was verified by SDS–PAGE and Coomassie Blue staining to contain only a single band with M_r = 33 000.

2.3. Crystallization

The protein was concentrated to 19 g l^{-1} and used to set up a sitting-drop vapour-diffusion screen with the Crystal Screen kit from Hampton Research. The screen was set up using 19, 10 and 5 g l^{-1} protein for each condition and was incubated at 295 KC. Initial crystallization hits were identified and were used to streak-seed hanging-drop vapour-diffusion crystal-optimization experiments.

Selected crystals were removed from the hanging drops with a nylon mounting loop (Hampton Research) and were suspended in precipitant solution containing 20% glycerol for 10 s prior to being vitrified by immersion in liquid nitrogen.

2.4. Data collection, structure solution and refinement

X-ray diffraction data were collected from a single crystal using the diffractometer system at the BM14 beamline at the European Synchrotron Radiation Facility. 180 frames were collected with 1° oscillations at a wavelength of 0.8856 Å. The data were processed, integrated and scaled using *XDS* (Kabsch, 2010). The structure was solved by molecular replacement with *Phaser* (McCoy *et al.*, 2007) using the coordinates of the *A* chain of MshB (McCarthy *et al.*, 2004) retrieved from the PDB (entry 1q7t) as a template. Model building was performed with *Coot* (Emsley *et al.*, 2010) and the structure was refined using *PHENIX* (Adams *et al.*, 2010). The two polypeptide chains were initially refined and interpreted completely independently, but once convergence was reached the noncrystallographic symmetry was exploited to further refine the similar residues in each chain using *REFMAC*5 (Murshudov *et al.*, 2011). Finally, six cycles of refinement without the imposition of noncrystallographic symmetry were performed with *PHENIX* using simulated annealing.

2.5. Molecular-dynamics simulation and substrate modelling

A model of GlcNAc-Ins was built using *Accelrys Discovery Studio Visualizer* 2.5 following Nicholas *et al.* (2002). The atoms were named according to the CHARMM36 force field (Guvench *et al.*, 2008; Hatcher *et al.*, 2009) corresponding to the numbering for D-*myo*-inositol prescribed by the Nomenclature Committee of the IUB (1989). The initial docking of GlcNAc-Ins into the MshB model was performed in *UCSF Chimera* (Pettersen *et al.*, 2004) by superimposing it onto the spatially corresponding atoms of glycerol and acetate. The model used in the simulation was prepared using *VMD* (Humphrey *et al.*, 1996) and comprised the complete monomer polypeptide chain, Zn^{2+}, GlcNAc-Ins, all crystallographic waters and a cubic water box that extended 10 Å beyond the protein in all directions. Simulations using a variety of protocols were performed with *NAMD* (Phillips *et al.*, 2005)

Table 1
Data-collection and structure-solution statistics.

Values in parentheses are for the outer shell.

Data collection	
X-ray source	BM14, ESRF
Wavelength (Å)	0.8856
Crystal dimensions (mm)	0.2 × 0.25 × 0.1
Temperature (K)	100
Space group	*C*2
Oscillation angle (°)	1
No. of frames	180
Unit-cell parameters (Å, °)	*a* = 147.42, *b* = 71.90, *c* = 97.69, β = 128.29
Resolution range (Å)	76.67–1.95 (2.07–1.95)
Mosaicity (°)	0.21
No. of observed reflections	221610 (35224)
No. of unique reflections	58226 (9287)
No. of unique reflections in test set	2911 (464)
Completeness (%)	99.3 (98.8)
Multiplicity	3.8 (3.8)
$\langle I/\sigma(I)\rangle$	16.66 (3.68)
Wilson *B* factor (Å^2)	29.9
R_{int}†	0.07 (0.42)
R_{meas}‡	0.08 (0.49)
No. of protein molecules in asymmetric unit	2
Matthews coefficient V_M (Å^3 Da^{-1})	3.20
Solvent content (%)	61.53
Refinement	
Resolution range (Å)	76.67–1.95
Contents of asymmetric unit (excluding H atoms)	
No. of protein atoms	4382
No. of Zn atoms	2
No. of acetate atoms	8
No. of glycerol atoms	54
No. of water atoms	570
No. of PEG atoms	48
Average *B* factor (Å^2)	
Protein atoms	25.9
Zn atoms	17.0
Acetate atoms	25.1
Glycerol atoms	35.5
Water atoms	34.0
PEG atoms	33.7
R.m.s. deviations from ideal	
Bond lengths (Å)	0.019
Bond angles (°)	1.413
Ramachandran outliers	9
Rotamer outliers	22
Ramachandran favoured	549
R_{cryst}§	0.17 (0.23)
R_{free}¶	0.20 (0.27)

† $R_{int} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the *i*th measurement of the intensity of reflection *hkl* and $\langle I(hkl)\rangle$ is the mean intensity of reflection *hkl*. ‡ $R_{meas} = \sum_{hkl}\{N(hkl)/[N(hkl)-1]\}^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the *i*th measurement of the intensity of reflection *hkl*, $\langle I(hkl)\rangle$ is the mean intensity of reflection *hkl* and *N* is the number of observations of intensity *I*(*hkl*) (Diederichs & Karplus, 1997). § $R = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$, where F_{obs} and F_{calc} are the observed and the calculated structure factors, respectively. ¶ R_{free} was calculated analogously for a randomly selected 5% of the reflections.

using the CHARMM27 (MacKerell *et al.*, 1998) and CHARMM36 (Guvench *et al.*, 2008, 2009; Hatcher *et al.*, 2009) force fields. A constant temperature of 310 K was enforced by Langevin dynamics, with a damping constant of 1 ps^{-1} and a constant pressure of 101.3 kPa, using a Nosé–Hoover Langevin piston (Phillips *et al.*, 2005). A multiple time-stepping algorithm was employed, with bonded interactions evaluated every 1 fs, short-range interactions evaluated every 2 fs (cutoff of 12 Å) and long-range interactions evaluated every 4 fs. Long-range electrostatics were calculated using the particle-mesh Ewald method. The final structure used for analysis was generated by minimizing the equilibrated structure while restraining the protein backbone and selected interactions to their crystallographic values.

Table 2
Deposited crystal structures of MshB.

PDB code	Reference	Chain ID	Ligands in active site†	Residues visible
1q74	Maynes *et al.* (2003)	*A*	Zn^{2+}	3–99, 104–163, 170–299
		B	Zn^{2+}	2–97, 104–163, 168–206, 216–298
		C	Zn^{2+}	2–96, 104–162, 168–210, 217–298
		D	Zn^{2+}	2–99, 103–163, 168–204, 217–298
1q7t	McCarthy *et al.* (2004)	*A*	BOG, Hg	−12–164, 171–303
		B	BOG‡, Hg	2–164, 171–299
4ewl	This work	*A*	Zn^{2+}, GOL, Ac	1–299
		B	Zn^{2+}, GOL, Ac	1–164, 173–298

† Abbreviations: BOG, β-octylglucoside; GOL, glycerol; Ac, acetate. ‡ The β-octylglucoside associated with the *B* chain is not in the substrate-binding location.

3. Results

Crystals measuring approximately 0.2 × 0.25 × 0.1 mm formed overnight in mother liquor consisting of 0.1 *M* sodium acetate, 0.2 *M* ammonium sulfate, 25% PEG 4000, pH 4.6, that had been diluted 1:1 with 4 g l^{-1} MshB in 4% glycerol and 50 m*M* potassium phosphate buffer pH 7.0 after streak-seeding with microcrystals. The crystals diffracted to a resolution of better than 1.95 Å. The diffraction patterns were indexed in space group *C*2, with unit-cell parameters $a = 147.42$, $b = 71.90$, $c = 97.69$ Å, $\beta = 128.29°$. Data-collection and refinement statistics are shown in Table 1.

There were two molecules per asymmetric unit, designated chains *A* and *B*. Residues 1–99, 106–164, 173–204 and 216–298 were related by noncrystallographic twofold symmetry, but the density for residues 100–105, 165–172 and 205–215 was substantially different in the two chains. A model for residues 1–299 was built for chain *A* and residues 1–164 and 173–298 were built for chain *B*. In general, the density for chain *A* was better resolved than that for chain *B*. Residues 213–215 were poorly resolved in both chains. The density in this region suggests the possibility of a number of different conformations of the polypeptide backbone, but is not sufficiently well defined to model unambiguously. Several residues in poorly defined density, including Ala215*A*, Ala215*B*, Ser2*A* and Asp102*B*, refined to torsion angles outside the allowed region of the Ramachandran plot. His60, which is located in a distorted helical segment, also has a strained conformation, but this assignment is supported by good electron density. The loop 165–172 in the *A* chain was visualized in its entirety for the first time, possibly because it is stabilized by interactions with Glu231*A* and Gln232*A* of symmetry-related molecules. The poorly resolved loop 99–106 is adjacent to the active site and its high mobility may

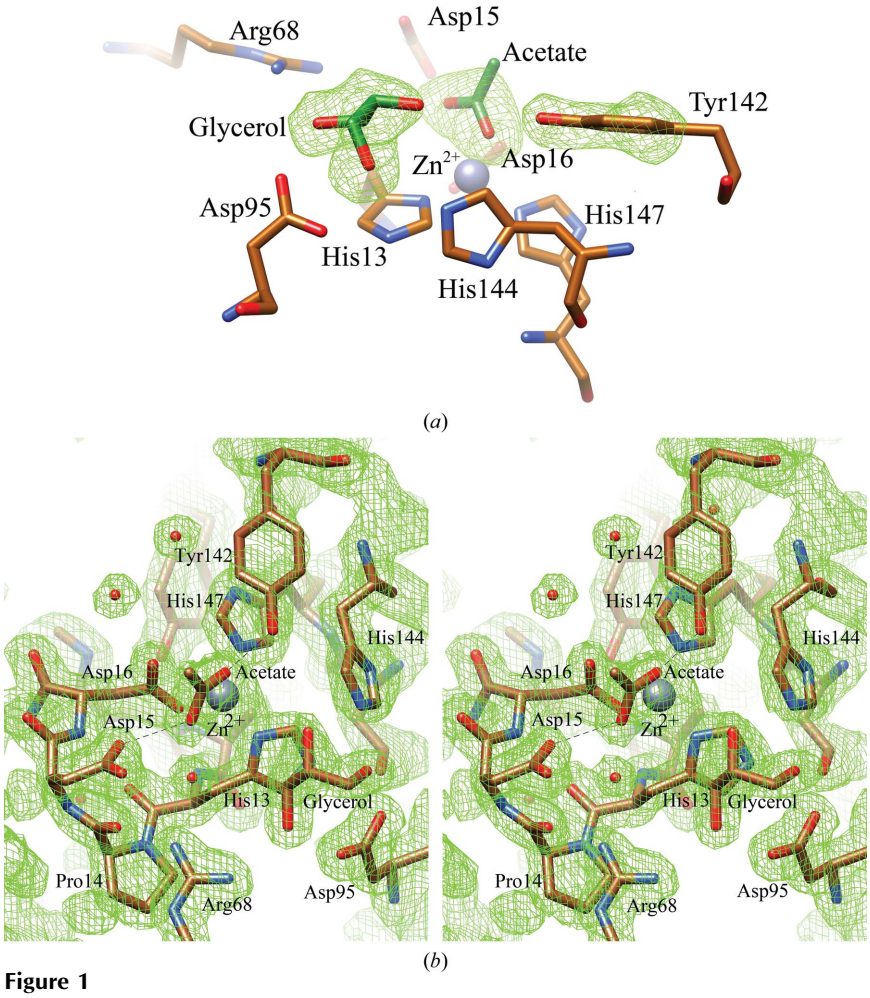


Figure 1
(*a*) The green mesh shows difference electron density ($F_o - F_c$) contoured at 1σ calculated for a model in which the glycerol and the acetate are omitted and Tyr142 is modelled as an alanine. The image is centred on the catalytic Zn^{2+}, which is depicted as a silver sphere. The locations of the glycerol and acetate relative to the residues lining the active-site pocket are unambiguously determined. Tyr142 has rotated relative to the positions previously determined and is within hydrogen-bonding distance of the acetate. (*b*) Stereoview of an electron-density map ($2F_o - F_c$) contoured at 2σ showing the region around the catalytic Zn^{2+}.

Table 3
Interactions with bound ligands in the active site.

Abbreviation: NAG, *N*-acetyl-D-glucosamine (GlcNAc).

	Corresponding atom in the GlcNAc-Ins model	Residues	Distance (Å) Chain *A*	Chain *B*
Glycerol				
3-OH	NAG1 O6	His144 $N^{\varepsilon 2}$	2.9	3.1
3-OH	NAG1 O6	Asp95 $O^{\delta 1}$	2.8	3.0
1-OH	NAG1 O4	Asp95 $O^{\delta 2}$	2.5	2.7
1-OH	NAG1 O4	Arg68 $N^{\eta 1}$	3.0	2.5
Acetate				
OXT	NAG1 O	Zn^{2+}	2.6	2.4
OXT	NAG1 O	Asp16 $O^{\delta 1}$	2.4	2.6
OXT	NAG1 O	Asp15 $O^{\delta 1}$	2.9	2.9
C	NAG1 C	Asp15 $O^{\delta 1}$	3.6	3.6
O		Tyr142 O^{η}	3.1	2.5
O		Zn^{2+}	1.7	1.8
Zn^{2+}				
Zn^{2+}		His13 $N^{\delta 1}$	2.0	2.1
Zn^{2+}		Asp16 $O^{\delta 1}$	2.5	2.5
Zn^{2+}		Asp16 $O^{\delta 2}$	2.1	2.1
Zn^{2+}		His147 $N^{\varepsilon 2}$	2.1	2.1
Additional interactions mentioned in the text				
Asp146 $O^{\delta 1}$		Ser96 O^{γ}	2.5	2.6
Asp146 $O^{\delta 2}$		Ser96 N	2.7	2.7
Asp146 $O^{\delta 2}$		His13 $N^{\varepsilon 2}$	2.9	2.8

facilitate the entry of the substrate into the active-site pocket. Table 2 lists the different components of MshB visualized in the structure described here and compares them with the parts visualized in the two previous structure determinations of this enzyme.

The metal-ion cofactor was present in both chains and was modelled as Zn^{2+}. This appeared to be consistent with the density, but there was no explicit evidence for the ion being zinc except that it seems likely as the protein was exposed to Zn^{2+} ions during purification. No clearly defined water molecules could be observed in the vicinity of the Zn^{2+} ion, whereas two water molecules were found to be coordinated to the zinc by Maynes *et al.* (2003).

A clearly defined acetate ion that was visible in both the *A* and *B* chains was coordinated to the metal ion (Fig. 1 and Table 3). Fig. 1(*a*) depicts the difference density for a map in which the acetate and glycerol were omitted and Tyr142 was modelled as an alanine. Fig. 1(*b*) shows the electron density ($\sigma = 1.0$) surrounding the Zn^{2+} ion and shows the relationships between the acetate, the zinc ion and Asp15. Tyr142 is within hydrogen-bonding distance of an acetate oxygen (O) and is in a different position to that found in any of the six chains of MshB previously modelled (Maynes *et al.*, 2003; McCarthy *et al.*, 2004). The other carboxylate O atom (OXT) is within hydrogen-bonding distance of Asp15 $O^{\delta 1}$ (Fig. 2*a*). In turn, the location of the carboxylate of Asp15 is stabilized by a hydrogen-bonding network comprising the backbone N atom of Glu47, a water, Gln247 $O^{\varepsilon 1}$, a water and Asp15 $O^{\delta 2}$. The Ramachandran angles of Asp15 ($\varphi = 117.6^{\circ}$, $\psi = -27.5^{\circ}$) are outside the normally allowed region, giving rise to a close approach between the carbonyl O atom of Pro14 and Asp15 C^{β} (2.7 Å). The density clearly supports our assignment of the atomic positions and the strained conformation is maintained as a result of the hydrogen bond between the carbonyl O atom and Arg68 $N^{\eta 2}$ (2.6 Å). The conformation is similar in both previously determined MshB structures (PDB entries 1q74 and 1q7t) despite the absence of the acetate in these structures.

Also present in the active site is a glycerol molecule that is hydrogen-bonded to Arg68, Asp95 and His144 (Fig. 2*a* and Table 3). The hydrogen bonding to the glycerol in both chains is similar to that described by McCarthy *et al.* (2004) for the β-octylglucoside (BOG) that was reported to be associated with chain *A* in their structure. Seven other glycerol molecules

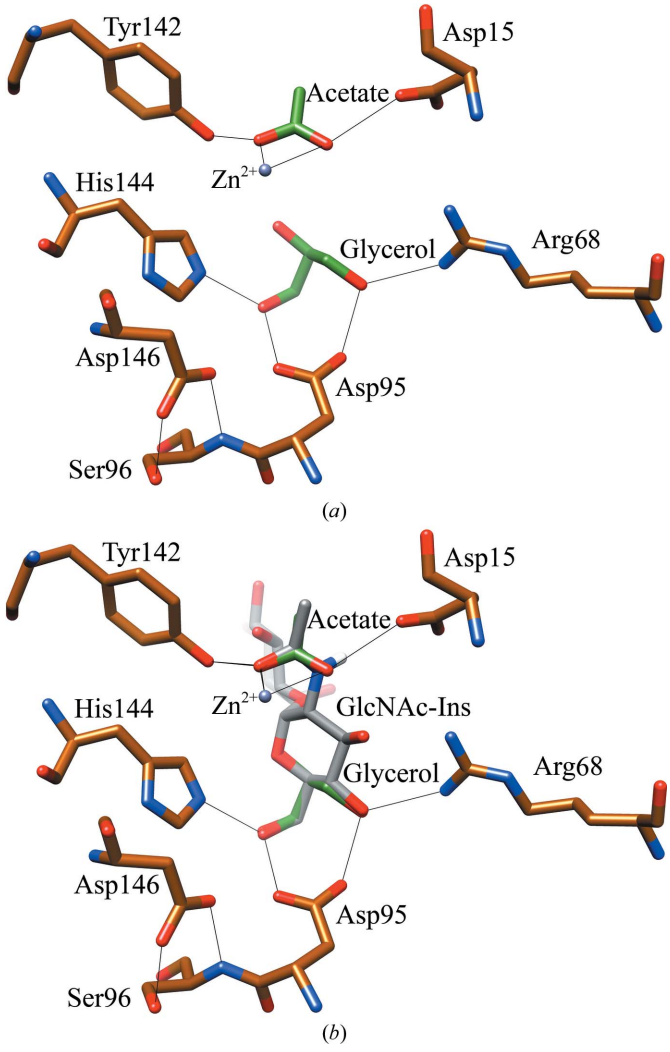


Figure 2
Hydrogen bonding to the substrate. (*a*) Acetate is positioned by interactions with the catalytic zinc ion, Asp15 and Tyr142. Glycerol is positioned by interactions with the side chains of His144, Asp95 and Arg68. Asp146 contributes to stabilizing the position of Asp95 through its interactions with Ser96. (*b*) The location of the natural substrate GlcNAc-Ins can be modelled in the active-site pocket by superimposing the atoms of the glucopyranoside ring onto the corresponding atoms of glycerol and the C and O atoms of the amide carbonyl onto the corresponding atoms of the acetate. This locates the amide bond such that Asp15, assisted by the electron-withdrawing effect of Zn^{2+} and Tyr142, can catalyse the nucleophilic attack of the water that is coordinated to the Zn^{2+} on the carbonyl carbon. The inositol ring is accommodated in a pocket behind the plane depicted.

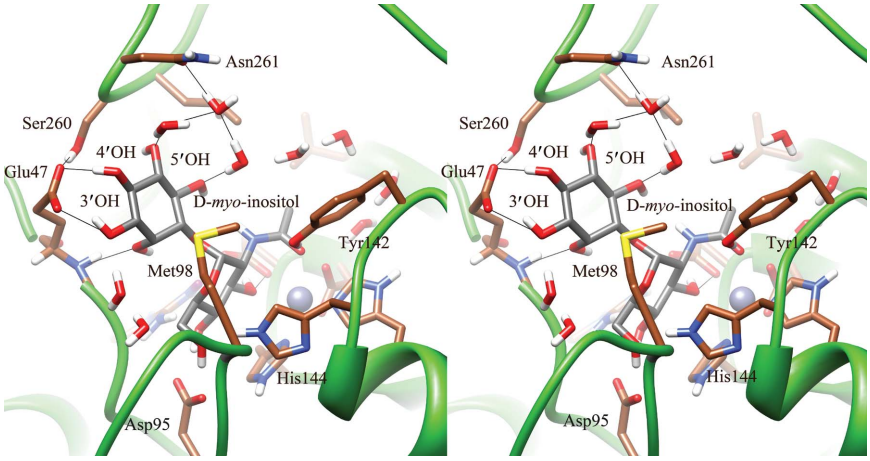


Figure 3
Molecular-dynamics simulations of the natural substrate (shown in grey) in the active site of MshB. The D-*myo*-inositol is modelled as being located in a pocket lined by Glu47, Ser260, Leu259, Asn261 and Met98. Two hydrogen bonds are formed between Glu47 and the D-*myo*-inositol ring, *i.e.* $O^{\varepsilon 2}$ to 3′-OH and $O^{\varepsilon 1}$ to 4′-OH, and the backbone N atom of Glu47 forms a hydrogen bond to 2′-OH. Met98, which is located in a mobile externally located loop, could stabilize this loop through hydrophobic interactions with the D-*myo*-inositol ring. Also suggested by the simulations are hydrogen bonds between the 3-OH of GlcNAc and both Arg68 $N^{\eta 2}$ and Asp15 $O^{\delta 2}$ (Fig. 2).

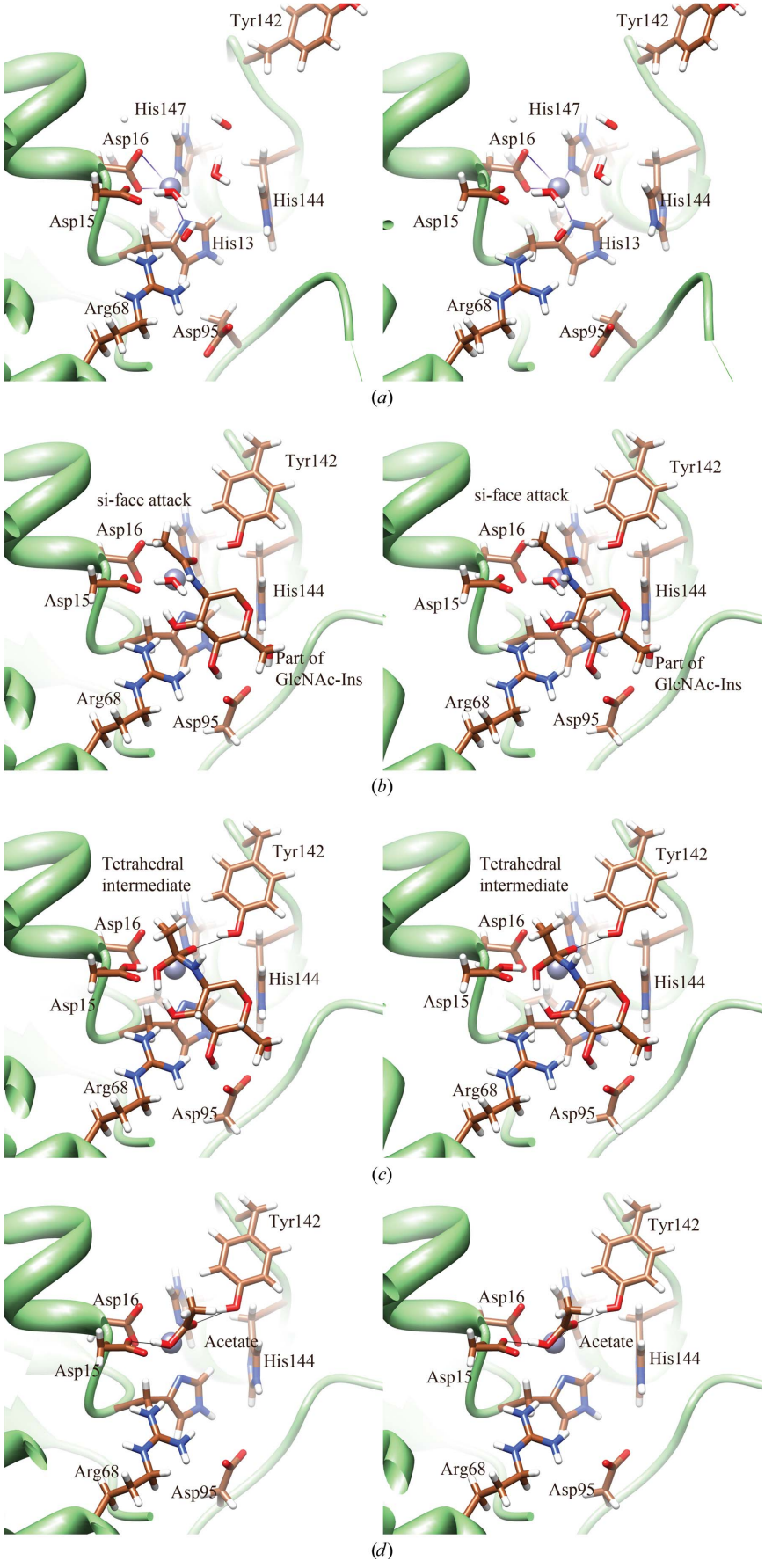
were found to be associated with the structure reported here; for example, in one of these the 1-OH, 2-OH and 3-OH of a glycerol molecule hydrogen-bonded to Ser82*A* O^{γ}, the main-chain carbonyl O atom of Val35*A* and the main-chain carbonyl O atom of Ala33*A*, respectively. These glycerol molecules probably lack functional significance in all cases. Two polyethylene glycol molecules were also found in fixed positions in the crystal structure (Supplementary Fig. 1[1]).

GlcNAc-Ins was modelled in the active site by superimposing the atoms of glycerol onto the corresponding atoms of the glucopyranoside ring and by superimposing the O atom of the carbonyl group of the amide bond onto the acetate O atom located between the Zn^{2+} ion and Tyr142 O^{ε}. Remarkably, the fit was unambiguous (Fig. 2*b*) and this simple docking suggests the possibility of an additional hydrogen bond being formed between GlcNAc 3-OH and Arg68 $N^{\eta 2}$. Furthermore, the D-*myo*-inositol ring of the natural substrate could be accommodated in a pocket lined by Met98, Glu47, Ser260, Asn261 and Leu259, with the possibility of hydrogen bonds forming between 3′-OH and Glu47 and between 4′-OH and Asn261. To further investigate this possibility, molecular-dynamics simulations were performed to optimize the position of GlcNAc-Ins while restraining atoms and distances known from the crystallographic structure. The simulations suggest that the location of the cyclohexyl ring is stabilized by hydrophobic interactions with Met98 and that two hydrogen bonds are formed between Glu47 and the D-*myo*-inositol ring, *i.e.* $O^{\varepsilon 2}$ to 3′-OH and $O^{\varepsilon 1}$ to 4′-OH. Additionally, the backbone N atom of Glu47 forms a hydrogen bond to 2′-OH, replacing the water that was observed at this location in the crystal structure. Asn261 does not directly interact with GlcNAc-Ins; rather, Asn261 $O^{\delta 1}$ interacts with 5′-OH and 6′-OH *via* a water-mediated hydrogen-bond network. Fig. 3 shows the docked structure of the D-*myo*-inositol ring. Further persistent interactions are observed in the molecular-dynamics simulations between Asp15 $O^{\delta 1}$ and GlcNAc 3-OH, and between Asp15 $O^{\delta 2}$ and Arg68 $N^{\eta 2}$. This has implications for the mechanism, as discussed below.

4. Discussion

The structure presented here shows MshB with the normal reaction product, acetate, coordinated to the metal-ion cofactor in the active site. The structure suggests a catalytic mechanism which has been modelled in detail in Fig. 4. The side chain of Tyr142 has rotated about the $C^{\alpha}-C^{\beta}$ bond relative to its previously observed positions in crystal structures of MshB (PDB entries 1q74 and 1q7t) to make a hydrogen bond with one of the carboxylate O atoms of the acetate. The position of Tyr142 corresponds closely to that predicted by Huang & Hernick (2012), who have shown that mutagenesis of this residue to phenylalanine, alanine or glutamine results in an almost 20-fold loss of activity relative to the wild-type enzyme. This configuration is suggestive of a conformational change occurring on substrate binding in which Tyr142 hydrogen bonds to the amide carbonyl of the substrate, thereby withdrawing electrons and making the carbonyl C atom more electrophilic. This increases the likelihood of nucleophilic attack by a water molecule catalysed by Asp15 acting as a general base. Maynes *et al.* (2003) have previously reported oxygen locations attributed to water molecules in locations corresponding to those assigned to the oxygen positions of the acetate in the present study. The geometry that we observe strongly suggests that one of these water molecules is displaced by the substrate and that the other acts as a nucleophile attacking the *si* face of the amide C atom and losing its proton to Asp15 $O^{\delta 1}$. The tetrahedral intermediate is stabilized by the oxyanion hole formed by the hydroxyl group of Tyr142 and the zinc ion. In the altered geometry of the transition state, the lone pair of the amide N atom is directed towards the newly acquired proton on Asp15 $O^{\delta 1}$. We therefore suggest that the proton that enables GlcN-Ins to leave arises from the water involved in the initial nucleophilic attack and that in the second stage of the reaction Asp15 acts as a general acid (Fig. 5). The stereoelectronic alignment required for the transfer of the proton, as suggested in Fig. 5(*b*), is imperfect in our naïve model, which is shown in Fig. 4(*c*). For this to occur, a small movement of the carboxylate of Asp15 from its crystallographically determined position is necessary. The molecular-dynamics simulations suggest that alternative hydrogen-bonding arrangements can arise in the presence of substrate, notably between Asp15 $O^{\delta 1}$ and

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: SX5106). Services for accessing this material are described at the back of the journal.



GlcNAc 3-OH and between Asp15 $O^{\delta2}$ and Arg68 $N^{\eta2}$. These suggested interactions could result in the necessary movement, but it should also be noted that the electronic environment in the vicinity of the Asp15 carboxylate will change further on formation of the tetrahedral intermediate and that this was not modelled.

There are two differences between the scenario that we propose and those proposed previously. Maynes *et al.* (2003) proposed that the imidazolium side chain of His144 forms part of the oxyanion hole that stabilizes the negative charge on the oxygen of the transition state and Huang & Hernick (2012) proposed that His144 acts as the general acid. The model presented here proposes that His144 $N^{\varepsilon2}$ is involved in making a hydrogen bond to the 3-OH of glycerol (which corresponds to the 6-OH of the glucopyranoside ring of the substrate GlcNAc-Ins), leaving His144 $C^{\delta1}$ directed towards the acetate. Furthermore, the distances from His144 $C^{\delta1}$ to the acetate O atom and from His144 $C^{\delta1}$ to the amide N atom of the scissile bond are 4.4 and 5.5 Å, respectively. The protonated $N^{\varepsilon2}$ of His144 is both incorrectly orientated and too far away to play the role of a general acid or to stabilize the oxyanion of the tetrahedral intermediate. Our interpretation agrees with that of McCarthy *et al.* (2004), who further comment that the backbone NH of Asp146

Figure 4
Structures and models in support of the proposed mechanism. (*a*) The active site prior to substrate binding (Maynes *et al.*, 2003; PDB entry 1q74, chain *A*): two water molecules coordinate to the Zn^{2+} ion. Both Tyr142 and Asp95 point away from the active site. (*b*) Part of the GlcNAc-Ins as modelled into the active site: the D-*myo*-inositol has been omitted from the drawing as it obscures Tyr142. The GlcNAc-Ins location is informed by the observed positions of glycerol and acetate in the structure reported here. Conformational changes involving Tyr142 and Asp95 have occurred. One water has been displaced, while the other water is poised to perform a *si*-face nucleophilic attack on the amide carbonyl C atom assisted by Asp15. (*c*) The modelled tetrahedral intermediate stabilized by the oxyanion hole formed by Zn^{2+} and by Tyr142. The carboxylate of Asp15 is protonated. In order achieve the stereoelectronic alignment for the amino group to acquire the proton it is necessary for the carboxylate to move slightly. We speculate that this movement may be induced by interaction between Asp15 $O^{\delta2}$ and 3-OH of GlcNAc. (*d*) The product acetate with one O atom hydrogen-bonded to Asp15 and the other located in the oxyanion hole based on the structure described here.

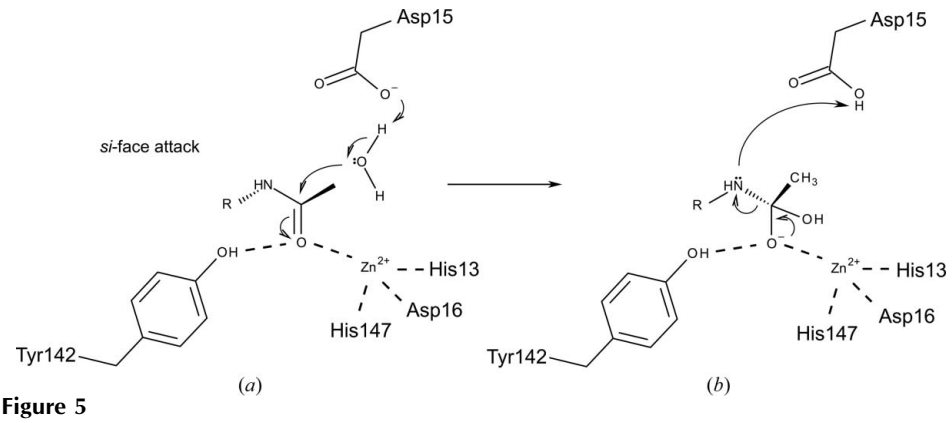


Figure 5
Catalytic mechanism. (*a*) The Asp15 carboxylate fulfils the role of a general base catalyst that activates a water for nucleophilic attack on the carbonyl C atom of the scissile amide bond. (*b*) Tyr142 forms part of the oxyanion hole that stabilizes the negative charge on the O atom of the tetrahedral intermediate. The newly protonated carboxylate of Asp15 acts as a general acid to complete the hydrolysis.

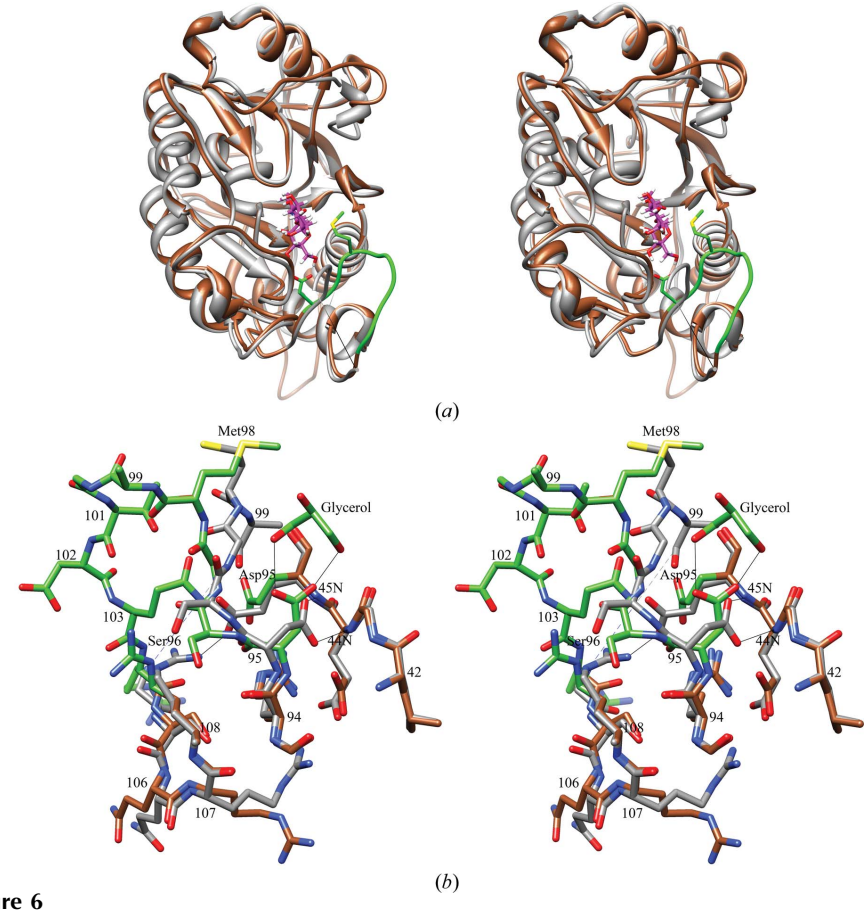


Figure 6
Conformational changes upon the binding of substrate. (*a*) A ribbon diagram of MshB as visualized in chain *A* of the structure of Maynes *et al.* (2003) (grey) is superimposed onto the glycerol-bound form described here (brown and green). The modelled substrate is depicted with magenta C atoms. The comparison highlights the functionally important conformational change in the residues that form the substrate-binding pocket. The rigidity of the loop comprising residues 95–107 shown in green is increased as a result of the interaction of the substrate with Asp95. This initiates changes which create the binding pocket for the D-*myo*-inositol ring. In particular, according to our model, Met98 moves so as to interact with the cyclohexyl ring. (*b*) In the form of MshB in which no ligands are bound in the active-site pocket, Asp95 forms two hydrogen bonds from each of the carboxylate O atoms to the main-chain NH groups of Glu44 and Glu45. In the glycerol-bound structure two new hydrogen bonds are formed to the 1-OH and 3-OH of glycerol, respectively. This results in a number of changes that result in the movement of the parts of the molecule in which the C atoms are coloured green. It can be clearly seen that the trajectory of the backbone between residues 95 and 99 has altered as a result of glycerol binding. Residues 100–103 are invisible in the MshB structure in which no ligands are bound in the pocket, indicating the high flexibility of the region.

forms a hydrogen bond to His144 $N^{\delta1}$ in order to ensure that His144 $N^{\varepsilon2}$ can act as a proton donor in its hydrogen bond to GlcNAc 6-OH. The geometry therefore strongly suggests that the protonated $O^{\delta1}$ of Asp15 plays the role of general acid in the catalysis and that His144 does not, which is in agreement with the proposal of Maynes *et al.* (2003).

Huang & Hernick (2012) further suggest that Tyr142 could play a role in product release as it is mobile and occupies different locations during different phases of the catalytic cycle. Indeed, restoring Tyr142 from the position found in this study to one of the locations observed in the MshB crystal structures with PDB entries 1q74 or 1q7t would assist in removing the acetate from the active site. This conformational change could be brought about by the increased mobility of the externally located loops following the departure of GlcN-Ins from the active site.

Huang & Hernick (2012) also demonstrate a loss of activity in the presence of concentrations of glycerol and sucrose that result in an increase in the viscosity of the solution. The structure reported here contains seven molecules of glycerol per asymmetric unit in specific locations determined by well defined hydrogen-bonding interactions. Two of these glycerol molecules are located in the active sites of chains *A* and *B*. These interactions are analogous to the hydrogen-bonding interactions with the glucopyranoside ring of BOG (McCarthy *et al.*, 2004) and suggest that analogous interactions position the natural substrate in the active site of MshB. Structural evidence thus suggests that the observed loss of activity should be attributed to competitive inhibition in addition to possible viscosity effects (Brouwer & Kirsch, 1982).

Our structure and that of McCarthy *et al.* (2004) allow the remarkable conformational changes (Fig. 6) in the flexible externally located loop comprising residues 95–107 to be visualized. The change occurs on the binding of either glycerol or BOG and can be seen by comparison with chains *A* and *D* of the MshB structure visualized by Maynes *et al.* (2003) which contained no ligand in the active site. This is suggestive of an induced-fit model in which formation of the hydrogen bonds between the substrate and Asp95 triggers a series of changes that result in formation of the side of the D-*myo*-inositol pocket especially through repositioning of Met98.

Mutation of Asp146 to alanine was also found to significantly decrease the activity and to increase K_m by a factor of more than ten (Huang & Hernick, 2012). The carboxyl group of Asp146 is within hydrogen-bonding distance of both the backbone amide (2.6–2.7 Å) and the O^{γ} (2.5–2.8 Å) of Ser96, as well as of His13 $N^{\varepsilon 2}$ (2.8–3.0 Å) (Fig. 2*a*). Huang & Hernick (2012) propose that the effects of mutating Asp146 are a consequence of its interaction with His13 and/or His144. While our structure supports the suggested interaction with His13, an interaction with His144 is unlikely as the closest approach of Asp146 $O^{\delta 1}$ to His144 is 3.3 Å and this is to the $C^{\varepsilon 1}$ atom. These interactions are similar to those observed in the *A* chain of MshB without a ligand in the active site (Maynes *et al.*, 2003) except that the direct interaction between Asp146 $O^{\delta 2}$ and Ser96 O^{γ} is replaced by a water-mediated hydrogen bond in order to accommodate the conformational change on substrate binding. The interactions between Asp146, Ser96 and His13 stabilize this region of the enzyme both before and after the conformational change and probably play a role in propagating the conformational change in response to the movement of Asp95. Thus, the mutation of Asp146 may affect substrate binding and this could potentially explain the observed loss of activity. The pK_a values of the functional groups in enzymes in which either His144 or Asp146 were mutated were the same as those of the wild-type protein, suggesting that neither of these residues were involved in ionizations that affected the catalytic events (Huang & Hernick, 2012). This strengthens the evidence for the view that both of these residues contribute primarily to substrate binding rather than catalysis.

Our findings provide direct structural evidence for the proposed binding of GlcNAc-Ins and we propose a new mechanism by which the amide bond is cleaved. The predicted location and role of Tyr142 in MshB during catalytic hydrolysis is confirmed. Historically, there have been two different misconceptions of the catalytic role of His144. The role of this residue is shown to be entirely concerned with substrate binding. We have speculated that the role of Asp15 as a general acid catalyst requires a positional change and that this may be achieved by interaction with the 3-OH of GlcNAc. This speculation could be confirmed by quantum-mechanical modelling. Further experimental work is also necessary in order to confirm the details of our modelling of the binding of the D-*myo*-inositol ring and the accompanying conformational changes.

We would like to thank Dr Hassan Belrhali of EMBL Grenoble for making beam time available on BM14. SGB would like to thank the National Research Foundation (NRF) and the UCT Drug Discovery Signature Theme for financial support. JCG is supported by a Director's Postdoctoral Fellowship from Argonne National Laboratory. MJM acknowledges funding from the Canon Collins Trust and NRF. We would like to thank Professor David Gammon for his encouragement and useful comments on the manuscript and Dr Andrew McCarthy for providing the structure factors for the MshB structure with PDB code 1q7t. We would like to thank the co-editor and the referees for their comments and suggestions. In particular, we thank Professor Mike James, who suggested that we include the mechanism diagrams and perform the molecular-dynamics simulation and contributed the insight concerning the movement of Asp95 on substrate binding.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Brouwer, A. C. & Kirsch, J. F. (1982). *Biochemistry*, **21**, 1302–1307.
Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 269–275.
Dosanjh, M., Newton, G. L. & Davies, J. (2008). *Res. Microbiol.* **159**, 643–650.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Fan, F., Vetting, M. W., Frantom, P. A. & Blanchard, J. S. (2009). *Curr. Opin. Chem. Biol.* **13**, 451–459.
Gammon, D. W., Steenkamp, D. J., Mavumengwana, V., Marakalala, M. J., Mudzunga, T. T., Hunter, R. & Munyololo, M. (2010). *Bioorg. Med. Chem.* **18**, 2501–2514.
Guvench, O., Greene, S. N., Kamath, G., Brady, J. W., Venable, R. M., Pastor, R. W. & MacKerell, A. D. Jr (2008). *J. Comput. Chem.* **29**, 2543–2564.
Guvench, O., Hatcher, E. R., Venable, R. M., Pastor, R. W. & MacKerell, A. D. Jr (2009). *J. Chem. Theor. Comput.* **5**, 2353–2370.
Hassan, S., Daugelat, S., Rao, P. S. S. & Schreiber, M. (2006). *PloS Comput. Biol.* **2**, e61.
Hernick, M. & Fierke, C. A. (2005). *Arch. Biochem. Biophys.* **433**, 71–84.
Hatcher, E., Guvench, O. & MacKerell, A. D. Jr (2009). *J. Chem. Theor. Comput.* **5**, 1315–1327.
Huang, X. & Hernick, M. (2012). *J. Biol. Chem.* **287**, 10424–10434.
Humphrey, W., Dalke, A. & Schulten, K. (1996). *J. Mol. Graph.* **14**, 33–38.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 133–144.
MacKerell, A. D. Jr *et al.* (1998). *J. Phys. Chem. B*, **102**, 3586–3616.
Maynes, J. T., Garen, C., Cherney, M. M., Newton, G., Arad, D., Av-Gay, Y., Fahey, R. C. & James, M. N. G. (2003). *J. Biol. Chem.* **278**, 47166–47170.
McCarthy, A. A., Peterson, N. A., Knijff, R. & Baker, E. N. (2004). *J. Mol. Biol.* **335**, 1131–1141.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
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.
Newton, G. L., Av-Gay, Y. & Fahey, R. C. (2000). *Biochemistry*, **39**, 10739–10746.
Newton, G. L., Buchmeier, N. & Fahey, R. C. (2008). *Microbiol. Mol. Biol. Rev.* **72**, 471–494.
Nicholas, G. M., Kovác, P. & Bewley, C. A. (2002). *J. Am. Chem. Soc.* **124**, 3492–3493.

Nomenclature Committee of the IUB (1989). *Biochem. J.* **258**, 1–2.
Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.
Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L. & Schulten, K. (2005). *J. Comput. Chem.* **26**, 1781–1802.
Xu, X., Vilchèze, C., Av-Gay, Y., Gómez-Velasco, A. & Jacobs, W. R. (2011). *Antimicrob. Agents Chemother.* **55**, 3133–3139.