SUBSTRATE BINDING TO MshB, A ZINC-DEPENDENT DEACETYLASE IN THE MYCOTHIOL BIOSYNTHETIC PATHWAY OF *Mycobacterium tuberculosis*.

Simon Gareth Broadley

Dissertation submitted in fulfilment of the requirements for the degree of Magister Scientiae in the department of Molecular and Cell Biology and Electron Microscope Unit, University of Cape Town

Supervisor:

Professor Trevor Sewell

February 2012

Keywords

MshB

Structure

Mycobacterium tuberculosis

X-ray crystallography

Protein

Mycothiol

Abstract

New drugs are needed to combat the ever-increasing incidence of drug resistant tuberculosis. Eukaryotes and Gram negative bacteria make use of a low molecular weight thiol-containing peptide, glutathione (GSH), in their defense against toxins and oxidative stress. In mycobacteria a carbohydrate-based thiol, mycothiol (MSH), is used instead.

MshB catalyses the third step in the mycothiol biosynthetic pathway which 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. MshB is a homolog of another zinc-dependent amidase, mycothiol S-conjugate amidase (Mca), which recycles GlcNAc-Ins by cleaving the similarly located amide bond after mycothiol has reacted with an electrophile. The enzymes have overlapping enzyme activity and their products feed into the same step of the mycothiol synthetic pathway. Structure-based drug design could potentially be used to develop leads that could inhibit both enzymes simultaneously.

This study was motivated by the need to visualise a putative inhibitor bound to MshB. An attempt was made to co-crystallise VU5 and MshB. VU5, which comprises plumbagin tethered to a thiophenylglycoside by five methylene carbons, was found by Gammon *et al.* (2010) to be an inhibitor of MshB and Mca.

Diffraction quality crystals were grown and data were collected at the ESRF synchrotron facility. The crystal structure was solved by molecular replacement and most notably, glycerol and acetate (a product of the natural reaction) was found bound in the active site. This allowed a greater understanding regarding important residues involved in substrate binding and the catalytic mechanism.

The findings presented here provide some insight into the substrate binding mechanism and support a recent publication that infers that Tyr-142 plays an important role in the catalytic mechanism, as well as suggesting that VU5 may behave as a substrate and that DTT may act as an inhibitor.

Declaration

I declare that "SUBSTRATE BINDING TO MshB, A ZINC-DEPENDENT DEACETYLASE IN THE MYCOTHIOL BIOSYNTHETIC PATHWAY OF *Mycobacterium tuberculosis*" is my own work, that it has not been submitted for any degree or examination in any other university, and that all sources I have used or quoted have been indicated and acknowledged by complete references.

Simon Gareth Broadley February 2012

Signed:.....

Acknowledgements

I would like to thank my supervisor Professor Trevor Sewell for his belief in me, his guidance and most of all for his patience. I am grateful to all the staff and students who advised and assisted me.

I thank my family for their support and encouragement.

Finally, I gratefully acknowledge the financial assistance of UCT, the National Research Foundation, the Council for Scientific and Industrial Research and the Drug Discovery Signature Theme.

List of abbreviations

ACE	angiotensin converting enzyme
AcCys	*N*-acetylcysteine
ASU	asymmetric unit
BOG	β-octyl glucoside
BSA	bovine serum albumin
CCD	charge coupled device
CCP4	Collaborative Computational Project number four
C. glut	*Corynebacterium glutamicum*
CNS	central nervous system
CV	column volume
DMSO	dimethyl sulfoxide
DTT	dithiothreitol
EDTA	ethylenediamine tetra-acetic acid
ESRF	European Synchrotron Radiation Facility
FSA	fluorescamine
FT	flow-through
GI	gastrointestinal
GlcN	glucosamine
GlcNAc	*N*-acetyl glucosamine
GlcNAc-Ins	inositol *N*-acetyl glucosamine
GSH	glutathione
HIV	Human Immune Virus
IPTG	isopropyl-1-thio-β-D-galactopyranoside
Mca	Mycothiol S-conjugate Amidase
MDR-TB	multi drug resistant TB
MSH	mycothiol
M. bovis	*Mycobacterium bovis*
M. smegmatis	*Mycobacterium smegmatis*

M. tb	*Mycobacterium tuberculosis*
MW	molecular weight
PDB	protein databank
PEG	polyethylene glycol
Q-HP	Q-sepharose high performance anion exchange chromatography
r.m.s	root mean squared
r.m.s.d.	root mean squared deviation
SAB	sample application buffer
SEC	size exclusion chromatography
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate – polyacrylamide gel electrophoresis
SNF	supernatant fraction
spp.	species
TB	tuberculosis
TCA	trichloroacetic acid
TMV	Tobacco Mosaic Virus
UDP	uridine-diphosphate
Ve	elution volume
Vi	initial velocity
Vin	internal volume
Vo	outer volume
WHO	World Health Organization
XDR-TB	extremely drug resistant TB
zDOPE	normalised discrete optimised protein energy

Table of Contents

Keywords 1
Abstract 2
Declaration 4
Acknowledgements 5
List of abbreviations 6
Table of Contents 8
List of Figures 12
List of Tables 15
1 Introduction 16
1.1 Tuberculosis 16
1.2 Mycothiol 21
1.3 MSH biosynthetic pathway 23
1.4 MSH-dependent enzymes 25
1.5 MSH-related drug targets 26
1.6 MshB and Mca 29
1.7 Structure-based drug design 31
1.8 Structure and mechanism of MshB 32
1.9 Homology modeling of Mca 40
1.10 Co-crystallisation of MshB with an inhibitor 41
1.11 Derivitisation of the product of the MshB activity assay 43
1.12 Aims and objectives 45
2 Materials and Methods 46

2.1 Expression of recombinant MshB protein in E. coli 46

2.2 Purification of MshB 47

2.2.1 Lysis of MshB-expressing *E. coli* BL21 cells 47

2.2.2 Ammonium sulphate precipitation 47

2.2.3 Anion exchange chromatography 48

2.2.4 Removal of contaminating proteins by size exclusion chromatography 49

2.2.5 Molecular weight estimation by size exclusion chromatography 50

2.2.6 Concentration of protein samples 52

2.2.7 Protein concentration determination 52

2.2.8 SDS-polyacrylamide gel electrophoresis (SDS-PAGE) 53

2.2.9 Protein Purity Estimation 53

2.3 X-ray Crystallography 54

2.3.1 Crystallisation 54

2.3.2 Identifying crystallising conditions for MshB 57

2.3.3 Optimisation of crystallisation conditions for MshB 57

2.3.4 Optimising crystal growth for MshB crystals 59

2.3.5 Preparation of MshB crystals for transport and X-ray diffraction 59

2.3.6 X-ray diffraction, model building and model refinement of MshB 60

2.4 Modeling of Mca 63

2.5 Characterisation of MshB Activity 64

2.5.1 MshB activity assays 64

2.5.2 Effect of DTT and VU5 on MshB activity assayed with GlcNAc substrate 65

2.5.3 Graphics and structural analysis 65

3 Results 66

3.1 Protein expression 66

3.2 Protein purification scheme 67

3.2.1 Purification scheme overview 67
3.2.2 Ammonium sulphate precipitation 68
3.2.3 Anion exchange chromatography 69
3.2.4 Sephacryl™ S300 calibration 75
3.2.5 Sephacryl™ S300 Size exclusion chromatography of the MshB containing fractions 77
3.2.6 Purification table 80
3.3 X-ray Crystallography of MshB crystals 82
3.3.1 Identifying and optimising crystallizing conditions 82
3.3.2 Crystal quality, space group and unit cell determination 84
3.3.3 Molecular replacement and structure refinement 85
3.3.4 Structural analysis of MshB 88
3.3.5 GlcNAc-Ins modelled into the active site of MshB solved here 95
3.4 Modeling of Mca 96
3.5 MshB activity 105
3.5.1 MshB assays 105
3.5.2 Effect of DTT 108
3.5.3 Effect of VU5 108
4 Discussion 111
4.1.1 MshB purification 111
4.2 Biological unit of MshB 114
4.3 X-ray Crystallography 116
4.3.1 Co-crystallisation of inhibitor 116
4.3.2 MshB Structure 118
4.3.3 Catalytic mechanism 124
4.4 Modeling of Mca 132

4.5 MshB Activity 135

4.5.1 MshB assays 135

4.5.2 Effect of Glycerol on MshB activity 137

4.5.3 Effect of DTT and VU5 on MshB activity 138

5 Concluding remarks 140

6 References 142

7 Appendix A: Densitometry analysis 150

List of Figures

Figure 1-1: Transmission and progression of TB. 17

Figure 1-2: Structure of mycothiol 21

Figure 1-3: MSH biosynthetic pathway. 24

Figure 1-4: The role of Mca in recycling GlcNAc-Ins. 25

Figure 1-5: Cartoon representation of MshB (PDB code 1Q7T, chain A). 32

Figure 1-6: Stereo view of MshB (PDB code 1Q74) showing the zinc ion (grey sphere) and water molecules (red spheres) in the putative active site. 33

Figure 1-7: Stereo view of the superimposition of the active sites of MshB (PDB code 1Q7T) showing β-octyl glucoside. 34

Figure 1-8: Schematic of the catalytic mechanism for MshB proposed by Maynes *et al.* (2003). 35

Figure 1-9: Comparison between the orientation of His-144 from McCarthy *et al.* (2004) (a,b,d) and Maynes *et al* (2003) (c) 36

Figure 1-10: pH rate profiles figure take from Huang and Hernick (2012). 38

Figure 1-11: Schematic of the catalytic mechanism proposed by Huang and Hernick (2012). 39

Figure 1-12: General structure of VU inhibitors. 41

Figure 1-13: Chemical structure of fluorescamine and conversion to fluorescent product once the fluorescamine molecule had reacted with a primary amine. 43

Figure 2-1: BSA standard curve. 52

Figure 2-2: Simplified phase diagram for protein crystallisation. 55

Figure 2-3: Hanging drop vapour diffusion 55

Figure 2-4: Crystals forming along a streak line after streak seeding. 58

Figure 2-5: This chart describes the steps required to solve a structure from a crystal. 60

Figure 3-1: 20 % SDS-PAGE showing expression of MshB 66

Figure 3-2: Purification scheme overview 67

Figure 3-3: 16 % SDS-PAGE of ammonium sulphate precipitation steps 69

Figure 3-4: Analysis of soluble cell content eluted off an anion exchange column. 71

Figure 3-5: Chromatogram of the FT from the previous run, eluted from Q-HP. 72

Figure 3-6: Representative chromatogram of anion exchange chromatography of the soluble cell content after ammonium sulphate precipitation. 73

Figure 3-7: SDS-PAGE analysis of MshB purification by ammonium sulphate precipitation followed by anion exchange chromatography. 74

Figure 3-8: Calibration of the Sephacryl^{TM} S300 column 76

Figure 3-9: Sephacryl^{TM} S300 Calibration curve 77

Figure 3-10: Sephacryl^{TM} S300 Size exclusion analysis 78

Figure 3-11: MshB crystals grown after streak seeding 83

Figure 3-12: Ramachandran analysis output from MolProbity (Lovell *et al.* 2003). 87

Figure 3-13: Cartoon representation of solved crystal structure of MshB 88

Figure 3-14: Electron density maps of both active sites in the MshB structure solved here. 89

Figure 3-15: The loop in chain A of MshB which is described here for the first time 90

Figure 3-16: MshB showing the B factors of the modelled residues in cartoon representation. 91

Figure 3-17: Cartoon representation of Chain A of MshB. 92

Figure 3-18: Stereo view of glycerol and acetate in the active site. 94

Figure 3-19: GlcNAc-Ins modeled into the active site of MshB solved here 95

Figure 3-20: Structural alignment reported by GenThreader. 98

Figure 3-21: Secondary structural elements of Mca predicted by Psipred. 99

Figure 3-22: Multiple structural alignment of Mca model and MshB structure solved here. 100

Figure 3-23: Active sites of MshB structure solved here and Mca model made here superimposed. 101

Figure 3-24: GlcNAc-Ins modeled into the active site of MshB (PDB code 1Q7T), superimposed onto the Mca model made here. 102

Figure 3-25: Local model quality plots of Mca modeled here and MshB solved here. 104

Figure 3-26: Overall model quality plots of Mca modeled here and MshB solved here. 104

Figure 3-27: MshB assay showing increase of fluorescence as a function of time. 106

Figure 3-28: Initial velocity of MshB assays showing increase of fluorescence as a function of time. 106

Figure 3-29: Plot showing the effect of DTT on MshB activity, showing the rate of increasing fluorescence. 108

Figure 3-30: Plots showing the effect of VU5 on MshB activity, showing the rate of fluorescence. 110

Figure 4-1: Chain A (green) and B (gold) of the MshB structure solved here superimposed onto the corresponding chains in MshB PDB code 1Q74 (yellow). 114

Figure 4-2: Stereo view comparison of VU5 and GlcNAc-Ins modelled into the active site of chain A of the MshB structure solved in this study. 117

Figure 4-3: Conformational changes in MshB upon binding of substrate. 120

Figure 4-4: Stereo view showing the movement observed for Tyr-142 in all known crystal structures of MshB, and the formation of the oxyanion hole. 121

Figure 4-5: Stereo view of ligands found in the active sites of MshB structures with PDB codes 1Q7T and 4EWL, superimposed upon one another. 122

Figure 4-6: Schematic of the catalytic mechanism for MshB that is proposed here. 125

Figure 4-7: Models supporting the catalytic mechanism proposed here. 127

Figure 4-8: pH rate profiles figure take from Huang and Hernick (2012). 128

Figure 7-1: Densitometric analysis of SDS-PAGE from samples taken from S300 chromatogrphy 151

List of Tables

Table 1-1: Adverse side effects of current first and second line drugs used to cure TB 20

Table 2-1: Table of gel filtration standards.....50

Table3-1: MshB purification table.....81

Table 3-2: Data collection statistics.....85

Table 3-3: Refinement and model validation statistics.....86

Table 3-4: Distances between relevant atoms in each chain of the structure solved here 93

Table 3-5: Equations of each graphical trend line from the initial velocity of the MshB assays and comparison of the resulting rates of fluorescence with previously published data.....107

Table 4-1: MshB Purification table drawn up by Dr Marakalala (2008).....113

1 Introduction

1.1 Tuberculosis

A third of the world's population is estimated to be carrying latent *Mycobacterium tuberculosis* (*M. tb*), the bacteria responsible for causing tuberculosis (TB). *Mycobacterium tuberculosis* is an aerobic, non-mobile bacteria that is characterised by acid fast detection, as gram staining is made difficult by a high mycolic acid content in the cell wall. TB was responsible for 1.7 million deaths worldwide in 2009, not taking into account 22 African countries that have no data reported (WHO 2011). TB has a high incidence of antibacterial resistance which is increased by patient non-compliance. South Africa is regarded by the World Health Organization (WHO) as one of twenty seven high burden countries for multi drug resistant tuberculosis (MDR-TB). MDR-TB is defined as infection by bacteria that are resistant to the most effective anti-TB drugs (isoniazid and rifampicin). South Africa is also one of the countries affected by the occurrence of extremely drug resistant tuberculosis (XDR-TB). XDR-TB is a form of tuberculosis caused by bacteria that are resistant to isoniazid and rifampicin, as well as any fluoroquinolone and any of the second-line anti-TB injectable drugs (amikacin, kanamycin or capreomycin) (WHO 2011). The drugs used to combat MDR- and XDR-TB are more expensive, less effective and more toxic than the first line drugs.

M. tb most often infects the lungs (pulmonary tuberculosis), but there have been cases where it has affected almost every organ in the body. It is primarily spread by the transfer of water droplets carrying the bacteria in the air from one person to another. In a healthy person, the immune system is most often able to overcome the primary infection. However the bacteria can lie dormant in the human host for years before an opportunity arises where the bacteria can cause an active infection again. On average it is estimated that 5 % of people carrying

the latent bacteria will develop the active disease again within two years, and 5 % will develop the active disease after two years. In patients that are immune deficient, as in the case of people infected with Human Immune Virus (HIV), this estimate increases to 10 % re-activation every year (Figure 1-1) (Small, Fujiwara 2001).

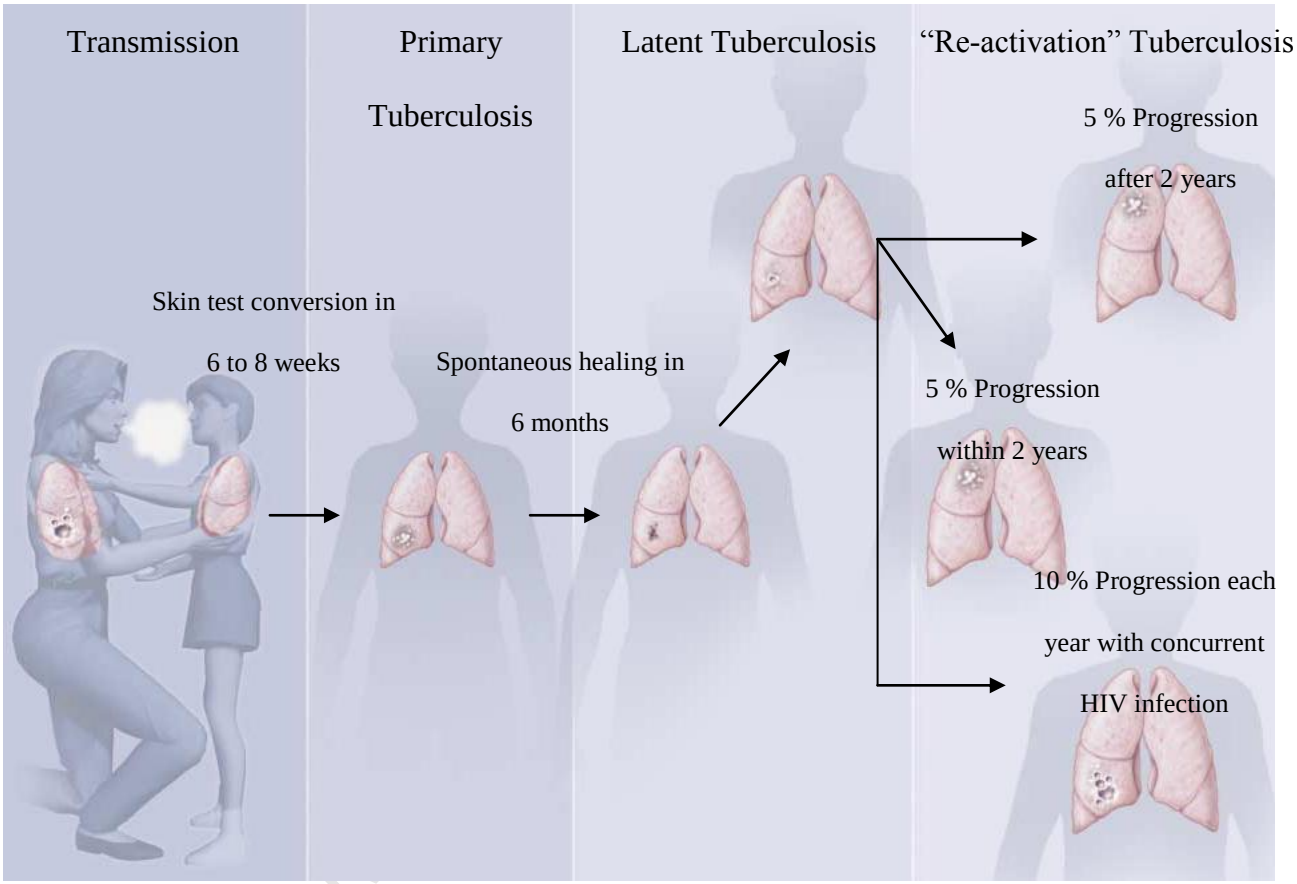


Figure 1-1: Transmission and progression of TB.

Adapted from (Small, Fujiwara 2001)

TB is transmitted primarily by airborne droplets of water carrying the bacteria. The skin test for TB will show positive for infection in six to eight weeks. In immuno-competent individuals, the body's immune system can often overcome the primary infection but can also lead to latent TB. The latent bacteria progress to full TB again in 5 % of cases within two years, and in 5 % of cases after two years. This increases to 10 % for each year in patients concurrently infected with HIV.

There are many compounding challenges that face the treatment of TB. In the late 1980's it was noticed that many patients infected with TB were defaulting on taking their medication due to the adverse side effects that the drugs caused.

The canonical treatment of TB requires a combination of drugs to be administered to the patient, which lends its own difficulties and complications when trying to encourage patients to finish their course of medication. Table 1-1 contains a list of the most common drugs and associated side effects that are used to cure TB infected patients.

The time needed for patients to complete a course of medication that leads to clearing of the current infection usually takes about 6 to 9 months. The time factor combined with the side effects increases the number of people who stop taking their medication prematurely. As mentioned previously, this then increases the likelihood of resistant TB infections emerging.

TB is often found in patients who are concurrently infected with HIV, due to their compromised immune system. The problem with this is that some TB drugs interact with anti-retroviral drugs in a negative way (Aaron *et al.* 2004). For example, when rifampicin is administered together with anti-retrovirals, this causes sub-therapeutic blood levels of anti-retroviral drugs and toxic levels of rifampicin (Small, Fujiwara 2001).

Commercial drug companies perceive a lack of profit opportunity because the highest number of new cases of TB are in under-developed countries, with the African region having the highest incidence per 100 000 population (WHO 2011), while the number of incidences in the developed world is decreasing. Combined with this is the inherent cost of developing new drugs, which is estimated to be anywhere between US$ 161 million and US$ 1.8 billion (Morgan *et al.* 2011). There is also a perceived problem facing clinical trials; few products have been newly registered for the treatment of TB in the past three decades and thus previous guidelines are seen as outdated and current guidelines are often poorly defined

(O'Brien, Nunn 2001). Lead discovery is currently a bottleneck in the pipeline for much needed novel drugs for diseases such as malaria and tuberculosis (Nwaka and Hudson 2006).

To ensure compliance and enhance positive treatment, new drugs are needed to either increase the sensitivity of TB to current drugs, or kill the bacteria independently. It would be best if these new drugs also target pathways or enzymes that are not present in the host and therefore will not have as many adverse drug interactions as the current therapies. This would shorten the regimen needed for successful treatment. New drugs may decrease the cost of treating TB and decrease the duration of treatment with drugs which cause the patient to feel ill and therefore increase the compliance and treatment rate.

One of the pathways that are of interest with respect to investigating drug targets is the mycothiol biosynthetic pathway of *M. tb*.

Table 1-1: Adverse side effects of current first and second line drugs used to cure TB

Drug	Adverse Side effects§	Comments
First Line Medication		
Isoniazid	Elevated hepatic enzymes, peripheral neuropathy, hepatitis, central nervous system (CNS) effects.	Overdose may be fatal.
Rifampicin	Hepatitis, fever, reduces blood levels of many drugs including warfarin, birth control pills, protease inhibitors and non-nucleoside reverse-transcriptase inhibitors.	Causes orange discolouration of secretions, interactions with many drugs lead to decreased levels of rifampicin. May complicate management of diabetes.
Pyrazinamide	Gastrointestinal (GI) disturbance, hepatic effects, rash.	May complicate management of diabetes.
Ethambutol	Decreased red-green discrimination, decreased visual acuity, rash.	Toxicity in the eyes may be unilateral therefore each eye must be checked separately.
Streptomycin	Auditory, vestibular and renal effects, electrolyte effects.	
Second Line Medication		
Capreomycin	Auditory, vestibular and renal effects, electrolyte effects.	
Ciprofloxacin¶	GI disturbance, insomnia, headache, photosensitivity, interactions with warfarin.	Variable absorption, increased caffeine effects.
Clofazimine¶	GI symptoms that can mimic appendicitis, visual disturbances in rare cases.	Efficacy unproved, causes orange-brown discolouration of skin.
Cycloserine	Psychosis, seizures, headache, depression, rash, other CNS effects.	
Ethionamide	GI disturbance, bloating, hepatoxic effects, hypothyroidism.	
Kanamycin/amikacin¶	Auditory and renal toxic effects, vestibular effects, electrolyte effects.	
Levofloxacin¶	Similar to effects of ciprofloxacin.	Variable absorption, increased caffeine effects.
Ofloxacin¶	Probably similar to effects of ciprofloxacin.	Variable absorption, increased caffeine effects.
Aminosalicylic acid	GI disturbance, hypersensitivity, hepatoxic effects, hypothyroidism.	
Rifabutin¶	Rash, hepatitis, fever, reduces levels of many drugs including protease inhibitors, non-nucleoside reverse-transcriptase inhibitors and birth control pills.	Causes orange discolouration of body fluids
Rifapentine	Nausea, vomiting, dizziness, rash, elevated values on liver function tests.	Cross-resistance with rifampicin, induces cytochrome P-450 system, causes orange discolouration of body fluids.

Notes:

§ Not all toxic effects are listed.

¶ The drug has not been approved by the USA Food and Drug Administration for the treatment of tuberculosis.

Adapted from "Management of tuberculosis in the United States" Small and Fujiwara, 2001 (Small, Fujiwara 2001).

1.2 Mycothiol

Eukaryotes and gram negative bacteria make use of a low molecular weight thiol called glutathione (GSH) in their defence against toxins and oxidative stress (Forman, Zhang & Rinna 2009). The major low molecular weight thiol in mycobacteria, however, is 1D-*myo*-inosityl-2-(*N*-acetyl-L-cysteinyl)amido-2-deoxy-α-D-glucopyranoside (MSH, Figure 1-2).

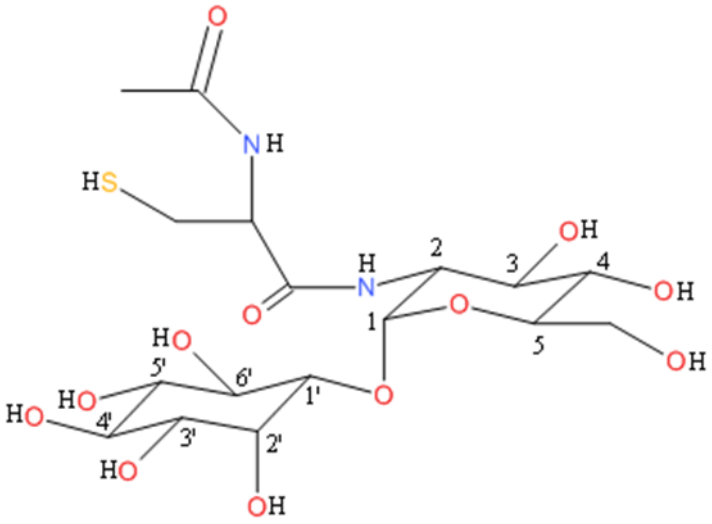


Figure 1-2: Structure of mycothiol

1D-*myo*-inosityl-2-(*N*-acetyl-L-cysteinyl)amido-2-deoxy-α-D-glucopyranoside (mycothiol or MSH) (Figure adapted from Newton and Fahey (2002)). MSH is a conjugate of *N*-acetylcysteine (AcCys) with 1-D-*myo*-inosityl-2-amino-2-deoxy-α-D-glucopyranoside (GlcN-Ins) (Newton *et al.* 1996).

The common name mycothiol was proposed by Spies and Steenkamp (1994). MSH is a conjugate of *N*-acetylcysteine (AcCys) with 1-D-*myo*-inosityl-2-amino-2-deoxy-α-D-glucopyranoside (GlcN-Ins) (Newton *et al.* 1996). MSH is involved in a variety of different cellular processes, including detoxification of electrophilic compounds, inactivation of reactive oxygen and nitrogen species, reductions and implicated to be involved in the

isomerisation of maleylpyruvate to fumarylpyruvate in some actinomycetes including *Corynebacterium* spp. (Newton, Buchmeier & Fahey 2008).

By analogy to GSH, MSH is proposed be very important in the protection of cells from a variety of toxins and reactive oxygen species. It has been shown that MSH deficient strains of *Mycobacterium spp.* are sensitive to stress (Rawat *et al.* 2002, Newton, Av-Gay & Fahey 2000, Buchmeier *et al.* 2003, Hayward, Wiid & van Helden 2004). Preventing bacteria from protecting themselves from toxic levels of reactive oxygen species that are abundant in an aerobic environment will also help in crippling dormant *M. tb*. In addition, some genes in the mycothiol biosynthesis pathway are essential for the growth of *M. tb*, namely *mshA* and *mshC* (Buchmeier, Fahey 2006, Sareen *et al.* 2003).

MSH is important to mycobacteria and since there is no MSH pathway in humans, it is possible that MSH biosynthesis can be inhibited in the bacterium without affecting the human host. This makes the MSH biosynthesis pathway a potential target for novel TB drugs (Duncan, 2004).

1.3 MSH biosynthetic pathway

The MSH pathway starts with the conjugation of 1L-*myo*-inositol-1-phosphate (Ins-P) to uridine-diphosphate-*N*-acetyl-glucosamine (UDP-GlcNAc), catalysed by MshA. Uridine-diphosphate (UDP) is removed during the reaction and the product formed is 1-*O*-(2-acetamido-2-deoxy-α-D-glucopyranosyl)-D-*myo*-inositol-3-phosphate (GlcNAc-Ins-P) (Newton *et al.* 2003, Newton *et al.* 2006). The second step is catalyzed by an enzyme that has yet to be identified (but has been referred to as MshA2), and involves the dephosphorylation of GlcNAc-Ins-P to form 1-*O*-(2-acetamido-2-deoxy-α-D-glucopyranosyl)-D-*myo*-inositol (GlcNAc-Ins) (Newton, Buchmeier & Fahey 2008). The third step involves the deacetylation of GlcNAc-Ins to 1-*O*-(2-amino-2-deoxy-α-D-glucopyranosyl)-D-*myo*-inositol (GlcN-Ins) by MshB (Newton, Av-Gay & Fahey 2000). The fourth step is catalysed by MshC and entails the ATP-dependent ligation of cysteine with GlcN-Ins to form 1-*O*-[2-[[(2*R*)-2-amino-3-mercapto-1-oxopropyl]amino]-2-deoxy-α-D-glucopyranosyl]-D-*myo*-inositol (Cys-GlcN-Ins) (Sareen *et al.* 2002). The fifth and final step in MSH biosynthesis results in the acetylation of the amino group of cysteine in Cys-GlcN-Ins by acetyl-CoA. This step is catalyzed by MshD and yields the final product 1D-*myo*-inosityl-2-(*N*-acetyl-L-cysteinyl)amido-2-deoxy-α-D-glucopyranoside (MSH) (Koledin, Newton & Fahey 2002). Figure 1-3 provides a graphical overview of the biosynthetic pathway and shows the substrates, products and known protein structures.

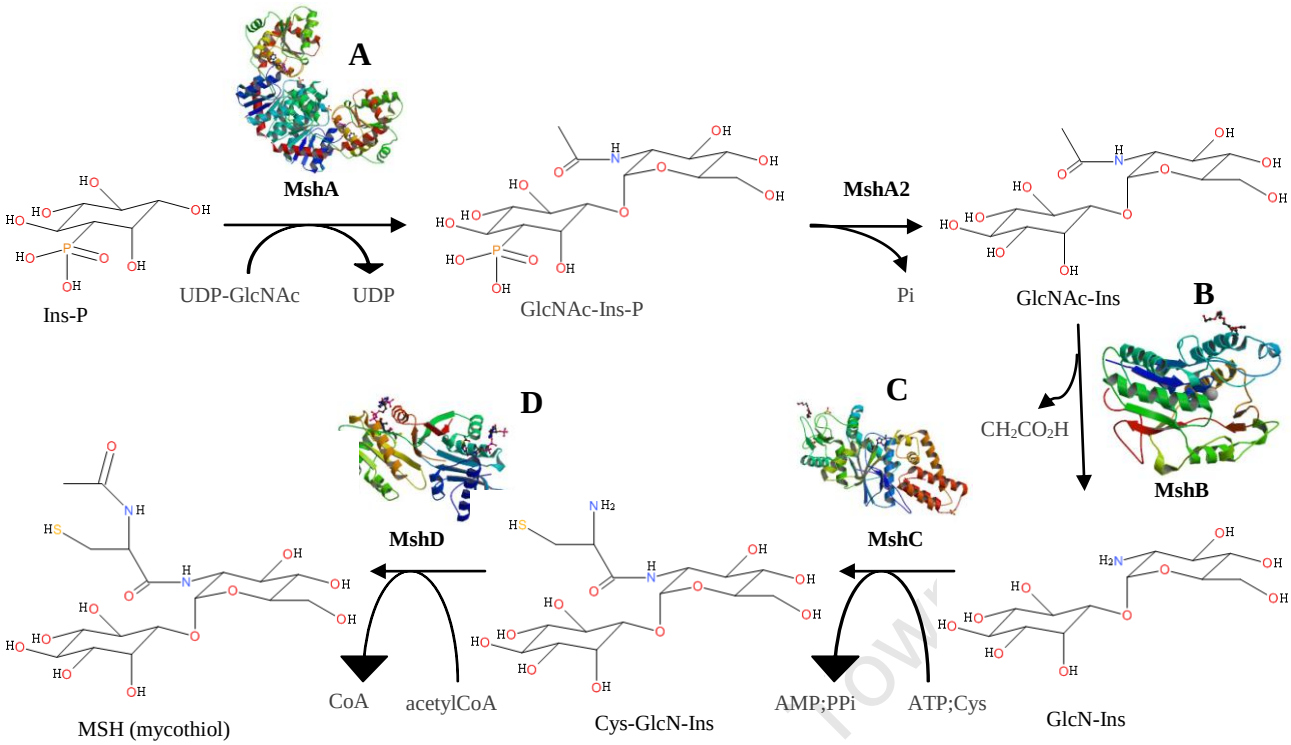


Figure 1-3: MSH biosynthetic pathway.

The five step biosynthesis of MSH is shown, adapted from Jothivasan and Hamilton (2008). The pathway begins with the conjugation of Ins-P to GlcNAc by MshA, followed by the removal of phosphate by a hypothetical protein MshA2. The acetyl group is then removed by MshB and a cysteine is added by MshC. Finally MshD catalyses the addition of another acetyl group.

(A) MshA: The structure of MshA from *Corynebacterium glutamicum* is shown (PDB ID: 3C4Q). (B) MshB: The structure of MshB from *M. tb* is shown (PDB ID: 1Q74). (C) MshC: The structure of MshC from *Mycobacterium smegmatis* is shown (PDB ID: 3C8Z). (D) MshD: The structure of MshD from *M. tb* is shown (PDB ID: 1OZP).

1.4 MSH-dependent enzymes

In one well-studied pathway (Figure 1-4), MSH forms conjugates with various toxins and antibiotics via the reactive sulphur and forms a compound referred to as MSR (Buchmeier *et al.* 2003, Steffek *et al.* 2003). MSH *S*-conjugate amidase (Mca) cleaves these conjugates producing GlcN-Ins and mercapturic acid. GlcN-Ins is then recycled to make more MSH by re-entering the MSH biosynthetic pathway via MshC. The mercapturic acid is a conjugate of the toxin and the acetyl cysteine and is thought to be excreted from the bacteria in a similar fashion to the GSH detoxification pathway (Newton, Fahey 2002). No Mca mutants have been reported in *M. tb* so far, but Mca mutants in *Mycobacterium smegmatis (M. smegmatis)* have been shown to have increased sensitivity to a variety of toxins and antibiotics (Rawat *et al.* 2004). This pathway is thought to be a route of antibiotic resistance and would therefore also be a good drug target.

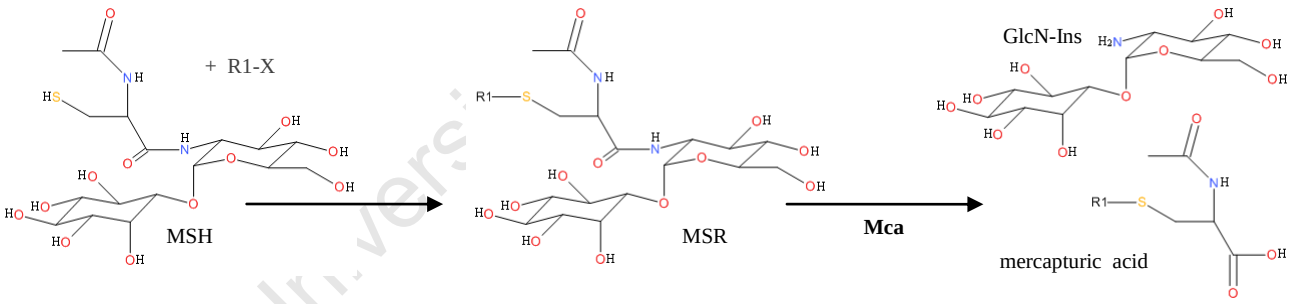


Figure 1-4: The role of Mca in recycling GlcNAc-Ins.

MSH forms conjugates (termed MSR) with toxins and antibiotics (R1-X). MSR is a substrate of Mca. The products of the reaction where MSR is hydrolysed by Mca are a mercapturic acid, which is a conjugate of the toxin (R) and the acetyl cysteine (AcCys), and GlcN-Ins, which is then recycled back into the biosynthetic pathway.

1.5 MSH-related drug targets

As MshA and MshC are considered essential to *M. tb*, these proteins would be good drug targets. The first obstacle with these targets is that there are no structures currently reported of these *M. tb* enzymes, which makes structure-based drug design difficult. There are however structures of these targets reported in other related organisms.

MshA has been solved from *Corynebacterium glutamicum* (*C. glut*) to a resolution of 2.8 Å (PDB ID 3C4Q). MshA is related to the GT-B structural family whose members have two domains with both domains showing a Rossmann-type fold. The Rossmann fold is a common three layer fold consisting of a β sheet layer sandwiched between two layers of α helices. According to a multiple alignment using the online Basic Local Alignment Search Tool (BLAST) (Altschul *et al.* 1997, Altschul *et al.* 2005), the *C. glut* MshA and the *M. tb* MshA share 51 % identity, and 74 % of the amino acids that are within 5 Å of the active site bound ligands in chain A of 3C4Q are the same. The similarities are in favour for potentially successful modeling attempts. If the measure of success of a model is a situation when the position of the C-α backbone atoms have an r.m.s.d. less than 2 Å when compared to the template, then protein structure templates for homology modeling are generally considered successful templates if they share more than 50 % sequence identity with the target protein to be modelled (Schwede *et al.* 2000, Eswar *et al.* 2006). Therefore a potentially successful model could be built, however a great limitation to evaluating drugs that target MshA, is the difficulty in obtaining active purified MshA from *M. tb* (Newton, Buchmeier & Fahey 2008).

MshC from *M. smegmatis* was solved to a resolution of 1.6 Å (PDB ID 3C8Z). The MshC tertiary structure resembles that of cysteinyl-tRNA synthetase, with a Rossman fold catalytic domain. Statistics from a BLAST search indicate that MshC from *M. smegmatis* shares 79 % identity to the MshC from *M. tb*, and 95 % of the amino acids that are within 5 Å of the

active site bound ligands in chain A of 3C8Z are the same. Due to the high sequence identity between the published structure and the protein of interest in *M. tb,* MshC in *M. tb* could be modelled fairly accurately using MshC from *M. smegmatis* as a template.

In one study a high-throughput screen identified lead molecules that could inhibit MshC (Newton *et al.* 2006). One hit, NTF1386, was found to interact well with MshC (Fahey *et al.* 2008). *M. smegmatis* treated with the MshC inhibitor NTF1836 produced less MSH but also showed reduced growth of the bacteria. This is important to note as MSH is not essential for growth in *M. smegmatis* (Newton, Buchmeier & Fahey 2008). Therefore current inhibitors of MshC are also likely to affect other essential pathways, which is advantageous in one sense as it would be unlikely for the bacteria to develop resistance to a drug that affects two essential pathways. However inhibitors of MshC are also likely to inhibit cysteinyl-tRNA transferase, which could be the cause of the reduced growth in the aforementioned study, but which is also present in humans (Newton, Buchmeier & Fahey 2008) and thus would more than likely cause considerable harm to the human host (a particularly undesirable trait in a drug lead) as well as the pathogen.

The next likely target is MshD. Its structure has been solved in *M. tb* and the tertiary structure consists of two tandem-repeated domains, each resembling the Gcn5-related *N*-acetyltransferase (GNAT) fold. MshD mutants are among the most sensitive mutants surveyed for survival in mouse macrophages (Rengarajan, Bloom & Rubin 2005), however MshD is not essential for *M. tb* growth (Buchmeier, Newton & Fahey 2006), and knockout strains still produce small amounts of MSH. No drugs have been reported that are active against MshD.

MshB is the last enzyme of the MSH biosynthetic pathway that will be discussed here. The structure of MshB closely resembles the Rossman fold of lactate dehydrogenase, although it

is interesting to note that MshB does not have a related fold among the known zinc-dependent enzymes (Maynes *et al.* 2003).

Some authors suggest that MshB is not a suitable drug target, reasons for this include the finding that MshB is not essential for the growth of *M. tb* and MshB mutants still produce MSH (most likely due to activity of an homologous enzyme, Mca) and show resistance to isoniazid and ethionamide (Xu *et al.* 2011). Isoniazid and ethionamide are pro-drugs that require activation by a catalytic peroxidise (katG) or by a flavoprotein monoxygenase; the mechanism by which the MSH pathway participates in the activation of these two drugs is unclear (Newton, Buchmeier & Fahey 2008).

Results have shown however that MshB mutants are sensitive to a range of other toxins including rifampicin and that MshB acts as a rate limiting step controlling the amount of MSH produced by way of a negative feedback loop (Buchmeier *et al.* 2003). MshB seems to be the most likely target in the mycothiol biosynthetic pathway, it has also previously been studied at the University of Cape Town and therefore will be the object of this research. MshB is also a very close homologue to Mca, and inhibitors that act against both of these enzymes simultaneously could be highly detrimental to the viability of *M. tb* and therefore lead to potential drug development.

1.6 MshB and Mca

MshB in *M. tb* was the first enzyme to be discovered in the MSH biosynthetic pathway and was found as a homologue to the *M. smegmatis* Mca (Newton, Av-Gay & Fahey 2000). Mca and MshB are similar in a number of ways. According to a multiple alignment using the online Basic Local Alignment Search Tool (BLAST), the *M. tb* Mca (taxonomy ID 1773) and MshB share 43% amino acid homology over the first 200 residues. Secondly, they are both Zn^{2+} metalloproteins. Thirdly there is evidence for a slight overlap in substrate specificity (Steffek *et al.* 2003, Newton *et al.* 2006).

There is currently no structure reported for Mca. Due to the similarities between Mca and MshB, drug leads that are produced using the MshB structure could potentially be used to stabilise the Mca enzyme enough for it to be crystallised, that is of course if the problem surrounding crystallisation of Mca is due to the fact that there are flexible loops around the active site that prevent it from crystallising readily. An Mca structure would add a great deal of insight into the binding mechanism of substrates and contribute significantly to the knowledge surrounding MSR metabolism.

Newton, Buchmeier and Fahey (2008) inferred that Mca was not essential to *M. tb* by referencing a study by Sassetti, Boyd & Rubin (2003), however the Sassetti study assumed that genes were only to be considered essential, if they were essential to both *Mycobacterium bovis* and *M. tb*. However, it is known that homologous MSH biosynthetic enzymes that are not essential in *M. smegmatis*, are essential in *M. tb*. For example, MshC mutants can be found in *M. smegmatis*, but not in *M. tb* (Sareen *et al.* 2003 and Rawat *et al.* 2002). Therefore citing the results of the Sassetti study as a reason to ignore Mca as a potential drug target would not be entirely correct.

In vitro experiments have shown that Mca is able to cleave the natural substrate of MshB, and it is thought to be an alternative route to MSH in *M. tb* mutants deficient in MshB (Buchmeier *et al.* 2003). However it must be noted that recombinant Mca that is over-expressed in a mutant lacking MshB in *M. smegmatis* was unable to recover wild type levels of MSH (Rawat *et al.* 2003).

A recent study has shown that *M. smegmatis* mutants with a double deletion of *mshB* and *mca* produced no detectable MSH (Xu *et al.* 2011). Inhibiting these enzymes could mean that no MSH would be produced, and therefore the bacteria would not be able to excrete toxins and agents that cause oxidative stress via this route and thus would be more sensitive to pro-oxidant drugs.

This has led to a focus on identifying and producing chemicals that inhibit both MshB and Mca as possible leads for drugs in the treatment of tuberculosis (Newton, Av-Gay & Fahey 2000, Steffek *et al.* 2003, Newton *et al.* 2006, Fetterolf, Bewley 2004, Gammon *et al.* 2003, Knapp *et al.* 2002, Nicholas *et al.* 2003, Nicholas *et al.* 2002, Nicholas *et al.* 2001, Pick *et al.* 2006, Metaferia *et al.* 2007b, Metaferia *et al.* 2007a, Gammon *et al.* 2010).

1.7 Structure-based drug design

In order to design drugs rationally, ideally an enzyme structure is needed with a transition state substrate bound in the active site. This provides information regarding important pharmacophores that are available in the active site and implicates the mechanism by which the substrates are processed. Based on that information one can perform *in silico* drug screens and formulate a few drug leads that should effectively inhibit the enzyme and decrease the cost of empirically screening hundreds of random small molecules.

Unfortunately, up until now, no-one has been able to visualise either substrate or inhibitors in the active site of MshB, and no-one has reported a crystal structure of Mca. Attempts to model the active site of Mca have revealed very little with regards to the changes that could be causing the differences in substrate specificity (Maynes *et al.* 2003). Therefore, there has been little investigation into structure-based drug design for inhibitors of both MshB and Mca. Structure based drug design depends on knowledge of inhibitor binding and the mechanism of action, therefore an analysis will be made of the present state of knowledge of MshB and Mca.

1.8 Structure and mechanism of MshB

In August of 2003 two groups published the structure of MshB from *M. tb* derived from X-ray crystallography. Maynes *et al.* (2003) deposited the first structure in the Protein Data Bank (PDB) at a resolution of 1.70 Å (PDB code 1Q74) in space group P1. Five days later, the second MshB structure was deposited by McCarthy *et al.* (2004) with a resolution of 1.90 Å (PDB code 1Q7T) in space group $P2_12_12_1$. These structures showed that MshB is a Zn^{2+}-dependent metallohydrolase, the N-terminal domain of which resembles the lactate dehydrogenase fold while the C-terminal domain has an α/β fold (Figure 1-5).

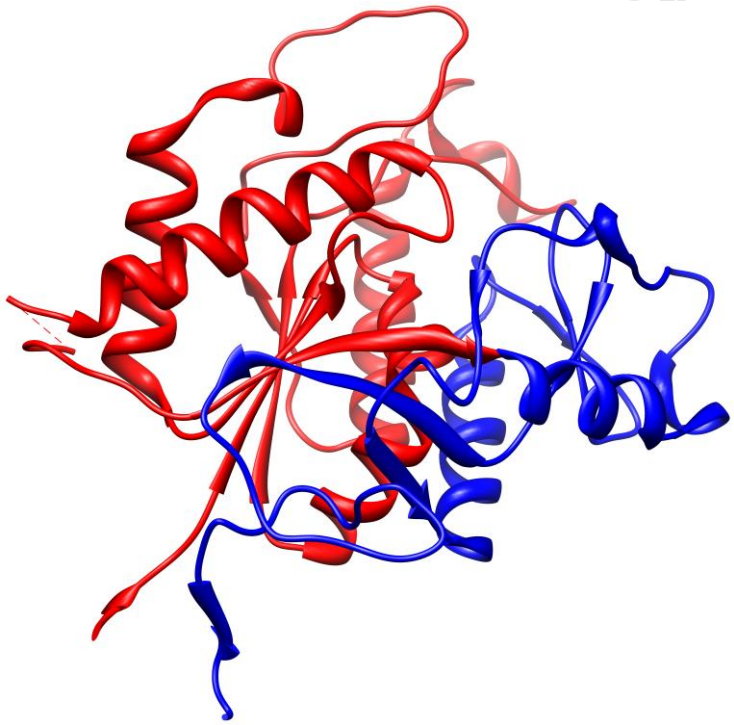


Figure 1-5: Cartoon representation of MshB (PDB code 1Q7T, chain A).

The N-terminal domain (red) resembles the lactate dehydrogenase fold while the C-terminal domain (blue) has an α/β fold.

The zinc binding site (Maynes *et al.* 2003) and the mercury binding site (McCarthy *et al.* 2004) in both structures is coordinated by His-13, Asp-16 and His-147. The MshB structure

with PDB code 1Q74 shows two water molecules coordinating with the catalytic zinc ion (Figure 1-6) and the MshB structure with PDB code 1Q7T shows the fortuitous binding of β-octyl glucoside (BOG) which is hypothesised to indicate the putative active site due to BOG resembling the glucose moiety of the substrate GlcNAc-Ins (Figure 1-7) (McCarthy *et al.* 2004). The binding position of BOG in chain A is hypothesised to more accurately predict the binding orientation of the glucose moiety of GlcNAc-Ins.

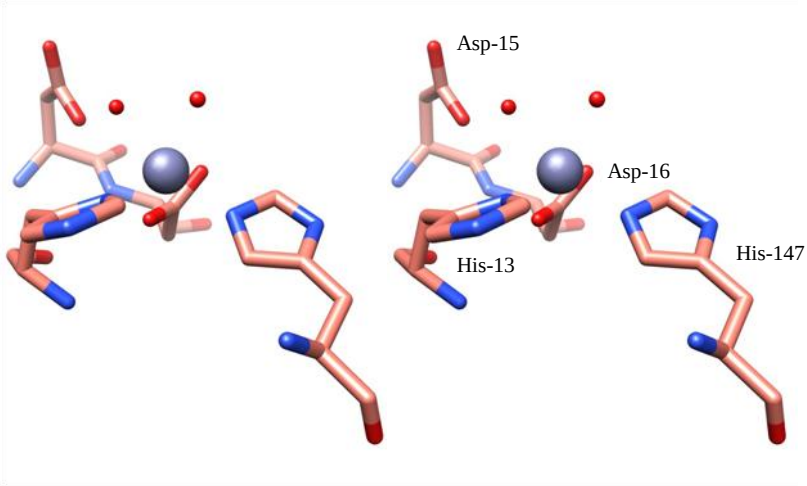


Figure 1-6: Stereo view of MshB (PDB code 1Q74) showing the zinc ion (grey sphere) and water molecules (red spheres) in the putative active site.

The zinc atom binding site is shown comprising His-13, His-147 and Asp-16. The carboxyl of Asp-15 is ideally positioned to interact with the proposed catalytic water, while the water closest to His-147 is hypothesised to be displaced by the substrate upon binding.

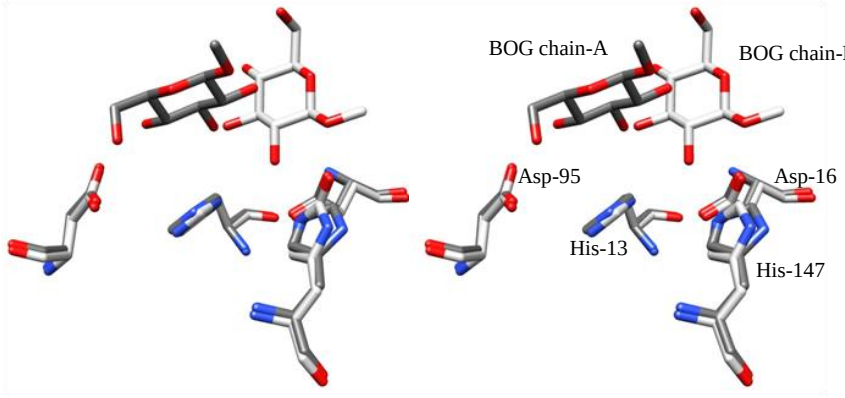


Figure 1-7: Stereo view of the superimposition of the active sites of MshB (PDB code 1Q7T) showing β-octyl glucoside.

The position of β-octyl glucoside **(**BOG) indicates the putative active site. Chain A in dark grey and Chain B in white. Asp-95 is on the far left and the zinc atom binding site comprised of His-13, His-147 and Asp-16 are on the right. The BOG binding in chain A is thought to more accurately predict the binding position of the glucose moiety of the substrate GlcNAc-Ins.

The catalytic mechanism proposed by Maynes *et al*. (2003) involves the substrate binding to MshB so that the carbonyl oxygen of the acetyl group displaces the water molecule closest to His-147 (Figure 1-6) and coordinates with the zinc ion.

The Asp-15 residue serves as a general base catalyst, assisting the nucleophilic attack of water on the amide carbonyl which is polarized by the zinc ion. The resulting tetrahedral transition state would be stabilised by the positively charged zinc ion and the side chain of His-144, which is hydrogen bonded to Asp-146. Finally, the carboxyl group of Asp-15 acts as a general acid for the proton transfer to the substrate nitrogen and facilitates the collapse of the tetrahedral intermediate with cleavage of the amide bond to leave GlcN-Ins and acetate (Figure 1-8).

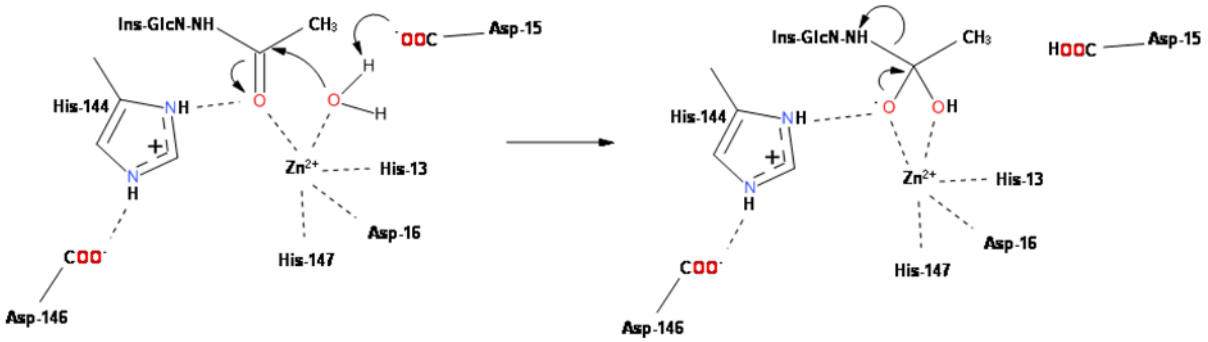


Figure 1-8: Schematic of the catalytic mechanism for MshB proposed by Maynes *et al.* (2003).

Asp-15 acts as a general base activating a water molecule which then performs a nucleophilic attack on the amide carbonyl. The tetrahedral intermediate is stabilised by the zinc ion and His-144 and collapses, via proton transfer from Asp-15 behaving as a general acid, to form GlcN-Ins and acetate.

This mechanism has found wide acceptance (Maynes *et al.* 2003, Jothivasan and Hamilton 2008, Newton, Buchmeier & Fahey 2008, Fan *et al.* 2009).

However McCarthy *et al.* (2004) proposed His-144 to be oriented differently, that the $N^{\varepsilon 2}$ of His-144 would coordinate with the *N*-acetylglucosamine moeity of the natural substrate (Figure 1-9). The discrepancies in the McCarthy *et al.* (2004) paper are discussed further on.

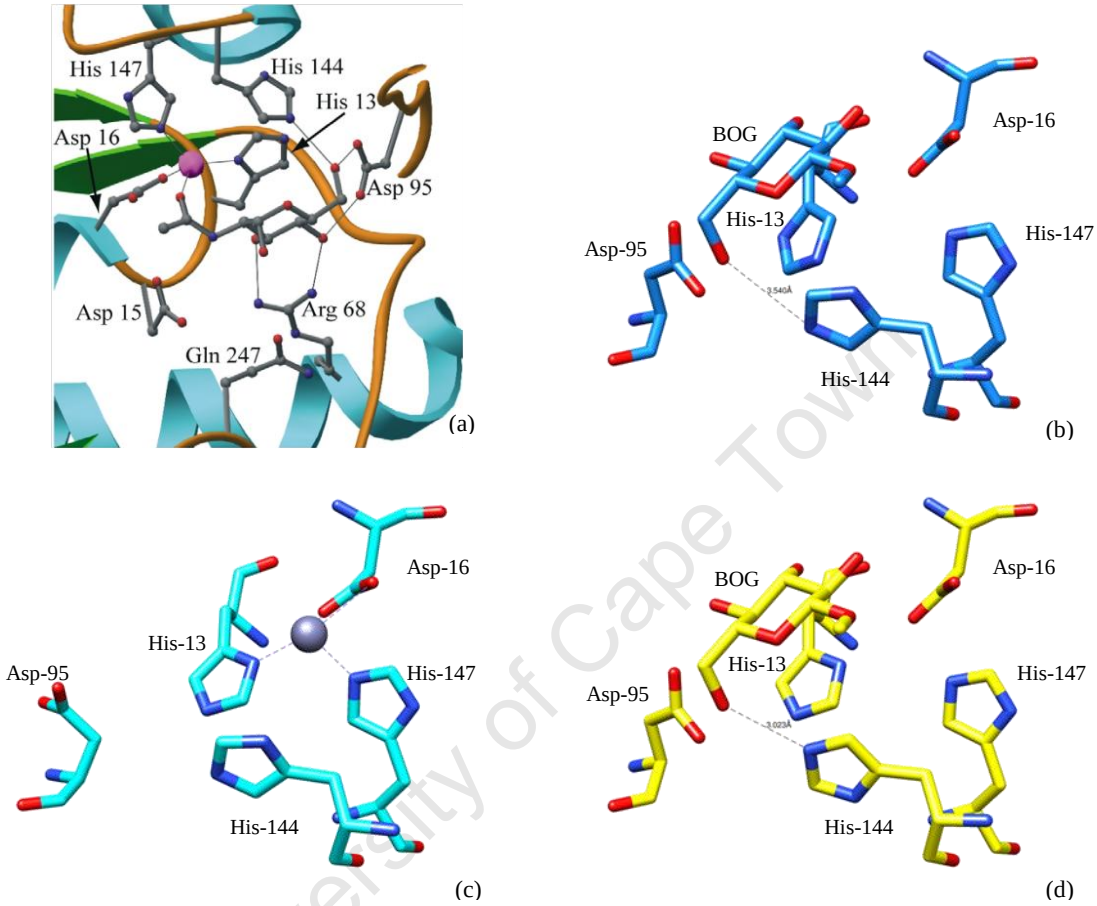


Figure 1-9: Comparison between the orientation of His-144 from McCarthy *et al.* (2004) (a,b,d) and Maynes *et al* (2003) (c).

The figure in (a) was taken from McCarthy *et al.* (2004) and shows the modelled orientation of HIS-144 upon substrate binding in MshB. Note the position of the $N^{\varepsilon 2}$ facing away from the catalytic center and towards O6 on the BOG moiety found in the active site, in addition note the orientation of His-13 with the $N^{\delta 1}$ facing towards the modeled zinc. BOG was used to predict the binding of the glucosyl ring of the natural substrate. The figure in (b) shows the deposited structure of MshB (with PDB code 1Q7T, Chain A). Note the His-144 $N^{\varepsilon 2}$ facing towards the catalytic center, and the His-13 $N^{\delta 1}$ facing away from the proposed zinc binding site. The figure in (c) shows the deposited structure of MshB (with PDB code 1Q74, Chain A) in which His-13 was modeled with $N^{\delta 1}$ orientated towards the zinc and with His-144 $N^{\varepsilon 2}$ facing towards the catalytic center. The histidines in figure in (d) have been re-orientated to match figure (a) for ease of reference with the other figures.

The *N*-acetylglucosamine moiety was modelled on the binding mode observed for β-octylglucoside (BOG) in the active site of chain A of 1Q7T, which they believed predicted the binding of the glucose moiety of the natural substrate, and therefore His-144 could not participate in the catalytic mechanism as the nitrogen was not facing the proposed position of the amide bond.

There is a discrepancy in the orientation of His-144 between the description of the crystal structure published by McCarthy *et al.* (2004) with PDB code 1Q7T, and the description of the orientation of His-144 when part of the natural substrate (the GlcNAc moiety) was modelled into the active site. In the paper (McCarthy *et al.* (2004)), His-144 $N^{\varepsilon 2}$ (Chain A) is depicted as being hydrogen bonded to the 6-OH of the glucosyl ring of BOG (with a distance of 3.54 Å), in the same orientation as shown in the coordinate file, with $N^{\delta 1}$ directed towards to active site. However, in the description of the modeling of the GlcNAc moiety into the active site, His-144 is oriented differently with $N^{\delta 1}$ directed away from the active site and $N^{\varepsilon 2}$ still hydrogen bonded to the 6-OH of the GlcNAc moiety. The modelled distance is not reported, but if the orientation of His-144 in their deposited structure is changed to match the description of binding to GlcNAc, then the distance is approximately 3 Å. This distance would be more favourable energetically. In addition, His-13 in Chain A of MshB with PDB code 1Q7T was modelled with the $N^{\delta 1}$ facing away from the proposed zinc binding site and should be facing towards in order to coordinate the zinc ion.

Considerable insight into the mechanism was derived from the kinetic experiments of Huang and Hernick (2012). Huang and Hernick (2012) published a paper where mutations in the active site of MshB were characterised (Figure 1-10). The loss of the upper and lower pKa's were noticed when Tyr-142 and Asp-15 were mutated, indicating their importance in the

catalytic mechanism. Based on the structures deposited at the time it was suggested that His-144 acts as a general acid catalyst and that the $N^{\varepsilon 2}$ of His-144 was pointed in the direction of, and could aid catalysis of, the amide bond of the natural substrate when bound in the active site of MshB. However when His-144 was mutated, the activity was reduced but the pKa's remained largely the same. Their work also showed that Tyr-142 plays an important role in the catalytic mechanism and included Tyr-142 in their schematic (Figure 1-11).

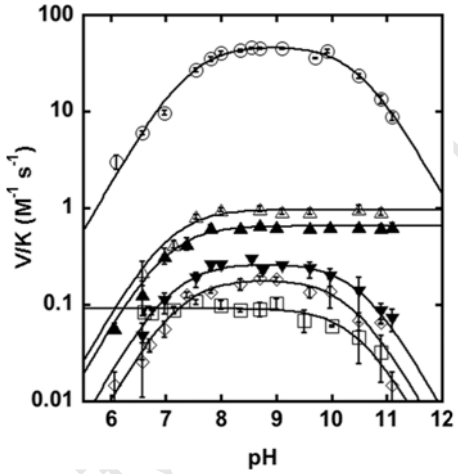


Figure 1-10: pH rate profiles figure take from Huang and Hernick (2012).

The pH dependence of the MshB-catalysed reaction is shown for WT (O), Asp-15 to Ala (□), Tyr-142 to Ala (△), Tyr-142 to Phe (▲), His-144 to Ala (▼), and Asp-146 to Ala (◇) as described in Huang and Hernick (2012).

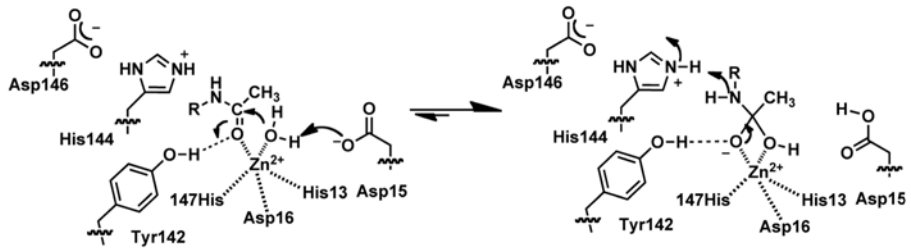


Figure 1-11: Schematic of the catalytic mechanism proposed by Huang and Hernick (2012).

On binding, Tyr-142 and the catalytic metal polarize the carbonyl group of the substrate. Asp-15 acts as a general base to activate a water for attack on the carbonyl substrate. The tetrahedral intermediate is stabilised by Tyr-142 and the metal ion. Finally His-144 acts as a general acid to facilitate the breakdown of the tetrahedral intermediate.

1.9 Homology modeling of Mca

Homology modeling is a technique used to build a structure of a previously unknown molecule based on closely related molecules. As previously stated, the degree of homology between MshB and Mca suggests they could be structurally similar to each other. Therefore, a model of Mca tertiary structure based on the protein sequence of Mca and the structural scaffold of MshB could produce a structural model of Mca that might be used to account for the differences in substrate specificity between Mca and MshB.

An attempt to model the first 180 residues of Mca using MshB as a template has been published by Maynes *et al*. (2003), it was reported that the only major difference between the two models was that Lys-19 in Mca was found in the location corresponding to Ser-20 in MshB. Maynes *et al.* (2003) suggested that the glucosyl moiety of GlcNAc-Ins (substrate of MshB) could bind in this location and that this difference could cause a steric clash in Mca. However MSR (substrate of Mca) has the same glucosyl moiety as GlcNAc-Ins, therefore Maynes *et al.* (2003) believed that either MSR binds in a different orientation or there are differences in other parts of the active site that might compensate for the unfavourable binding. For example, residues that bind to the AcCys residue of MSH were hypothesised to be Gln-247 and Ser-260, and increased binding affinities here could overcome the steric clash introduced by Lys-19. They were unable to model that region of Mca and therefore were unable to comment further. As mentioned in the previous section however, the structure with PDB code 1Q7T was published 5 days later with BOG bound nearer to Asp-95 and His-13. This evidence suggested that the glucosyl moiety bound in this region rather than near Lys-19. There is a need to build a complete homology model of Mca on the basis of current knowledge.

1.10 Co-crystallisation of MshB with an inhibitor

The purpose of co-crystallizing the enzyme with inhibitors is to permit the visualization of the interactions between the protein and the inhibitor. The insights obtained from this are used to design molecules that satisfy all the criteria necessary for them to be used as drugs and still maintain the critical interactions with the enzyme in question.

Gammon *et al.* (2010) synthesised 21 substrate analogue inhibitors of the hydrolysis of GLcNAc-Ins by MshB and Mca. Because they were substrate analogues they were proposed to act by competitive inhibition. The most effective inhibitors (Figure 1-12) were found to be the thiophenylglycosides (with the thiophenyl group in the same position as the inositol moiety, and the glycoside group and amide bond in the same position of that in the natural substrate GlcNAc-Ins) linked to plumbagin, and were referred to as compounds VU2-5.

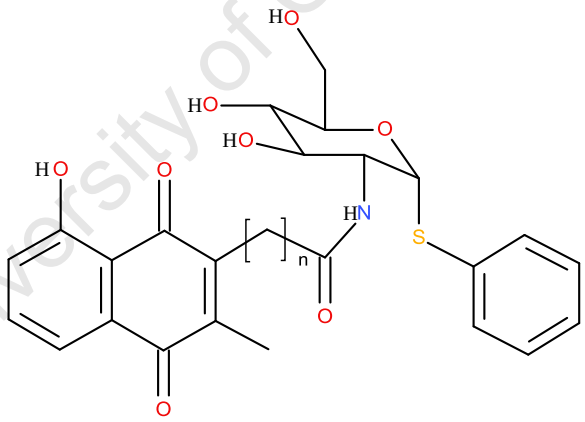


Figure 1-12: General structure of VU inhibitors.

VU inhibitors (Gammon *et al.* 2010, Marakalala 2008) are plumbagin linked via 2 to 5 methylene carbons and an amide linkage to phenyl-2-deoxy-2-amino-1-thio-α-D-glucopyranoside. These inhibitors are named VU*n*, with *n* being the number of carbons in the tether, e.g. for VU5 *n*=5.

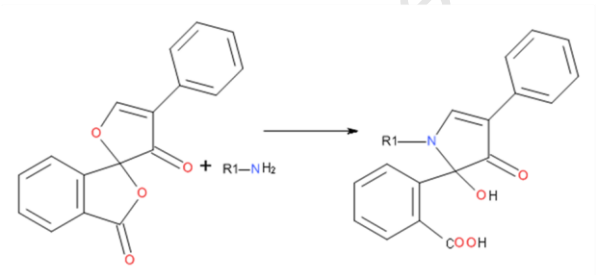
Interestingly the percentage of inhibition was found to be directly related to the length of the link, where an increase in the number of carbons in the link increased the amount of

inhibition of MshB and Mca activity (Gammon *et al.* 2010, Marakalala 2008). VU5, which has a five carbon linker, was found to be the most potent inhibitor of MshB and Mca activity. For this reason, VU5 was used as the inhibitor for co-crystallisation in this study. It must be noted that no chemical characterisation was reported for VU5 (solubility etc).

1.11 Derivitisation of the product of the MshB activity assay

Gammon *et al.* (2010) derivatised GlcN-Ins, the product of the MshB catalysed deacetylation of GlcNAc-Ins, with 6-aminoquinolyl-*N*-hydroxysuccinamidal carbamate (AccQ-Fluor, Waters, USA) using a method published by Anderberg, Newton & Fahey (1998) which included visualising the derivatised product using HPLC.

An alternative method of derivitising the GlcN-Ins product was developed more recently using the fluorescent substrate fluorescamine (FSA, Figure 1-13) by Huang and Hernick (2011).



Fluorescamine Fluorophore

Figure 1-13: Chemical structure of fluorescamine and conversion to fluorescent product once the fluorescamine molecule had reacted with a primary amine.

FSA only becomes fluorescently active once it has reacted with a primary amine. The amount of fluorescence is thus directly proportional to the amount of free primary amine in the solution and can be measured accurately by fluorescent spectrophotometry in a 96 well plate. This method of derivitisation is simple, quick and less costly than other assays using AccQ-Fluor. In short, instead of being based on multiple time-consuming runs on an HPLC

system, it relies on fluorescent labelling of the primary amine in 96 well plates which are easily read on a fluorescent plate reader.

Another advantage of the Huang and Hernick assay is that the cheaper and commercially available substrate analogue GlcNAc was used instead of the natural substrate GlcNAc-Ins which is not commercially available and is relatively difficult and expensive to produce.

Some notable differences between the Gammon *et al.* (2010) and the Huang and Hernick (2011) assays is that Gammon *et al*. (2010) used half the concentration of Hepes and NaCl, and did not include any reducing agent, whereas Huang and Hernick (2011) used 50 mM Hepes and 50 mM NaCl, and included 1 mM TCEP to maintain a reducing environment. The assay is described in detail in section 2.5.

1.12 Aims and objectives

The aim of this work, as initially conceived, was the direct visualization of VU5 in the active site of MshB by X-ray crystallography. This would then lead to an analysis of the binding of this molecule which would be returned to the synthetic chemists for the development of further compounds.

This would be accomplished by:

- Expression and purification of MshB for use in crystallisation,
- Co-crystallising the protein with the VU5 inhibitor,
- Solving the resulting structure by molecular replacement using data obtained from X-ray diffraction.

It will be seen that the route to the understanding of VU5 binding was not nearly as straightforward as initially conceived and the work outlined above led instead to a new understanding of the mechanism of MshB which enables a considerably more insightful model of VU5 binding to be built than was previously possible.

2 Materials and Methods

2.1 Expression of recombinant MshB protein in E. coli

MshB was expressed in *E. coli* BL21 (DE3) pLysS using a pET17b plasmid. The *E. coli* carrying the plasmid was a gift from Daniel Steenkamp, Division of Chemical Pathology, University of Cape Town. Colonies were picked, from stock cultures freshly plated on Luria Agar (1 % w/v tryptone, 1 % w/v NaCl, 0.5 % yeast extract and 1.5 % agar containing 100 µg/ml ampicillin and 34 µg/ml chloramphenicol), and inoculated into starter cultures of 10 ml sterile Luria Broth (LB) (1 % w/v tryptone, 1 % w/v NaCl and 0.5 % yeast extract containing 100 µg/ml ampicillin and 34 µg/ml chloramphenicol). The starter cultures were incubated overnight at 37 °C in a shaking incubator. The 10 ml culture was then added to 1 L sterile LB (containing 100 µg/ml ampicillin and 34 µg/ml chloramphenicol) and incubated at 37 °C in a shaking incubator until the optical density (Absorbance of light with a wavelength of 600 nm, OD_{600nm}) reached 0.4-0.6. Once the cells had reached the appropriate density (OD_{600nm} = 0.4-0.6), over-expression of the recombinant protein was induced by the addition of isopropyl-1-thio-β-D-galactopyranoside (IPTG) to a final concentration of 0.4 mM and incubated overnight at 20 °C in a shaking incubator (100 r.p.m.).

2.2 Purification of MshB

2.2.1 Lysis of MshB-expressing *E. coli* BL21 cells

Induced cells were collected by centrifugation at 5 000 x *g* for 15 min at 4 °C. The cells were resuspended in a total volume of 10 ml lysis buffer (100 mM NaCl, 50 mM Hepes, pH 7.5) containing cOmplete™ EDTA-free protease inhibitors (Roche, USA).

Disruption of the cell membranes was ensured by freezing the cell pellet overnight at -20 °C, prior to sonication on ice using a Misonix Sonicator 3 000. The total sonication process time was 4 min at an initial output level of 5 for a sample volume of 10 ml. Processing was done with cycles of 15 s with the pulse on and 15 s with the pulse off. The insoluble cell debris was separated from the soluble cell content by centrifugation at 27 000 x *g* for 30 min at 4 °C.

2.2.2 Ammonium sulphate precipitation

The supernatant from cell lysate was incubated with 35% ammonium sulphate for one hour at 4 °C, while being stirred to ensure even distribution and maximum solubility of ammonium sulphate in the sample. The solution was centrifuged at 10 000 x *g* for 15 min at 4 °C. The supernatant was then removed into a clean tube and the ammonium sulphate concentration increased to 55%. The sample was then incubated for another hour at 4 °C, while being stirred. The 55% solution was centrifuged at 10 000 x *g* for 15 min at 4 °C.

The supernatant was removed and the protein pellet from the 55% ammonium sulphate sample was resuspended in 50 mM Hepes buffer containing 100 mM NaCl, pH 7.5. Excess ammonium sulphate was removed by dialysis using 10 kDa Snakeskin® molecular weight cut-off dialysis tubing (Thermo Scientific, USA), overnight at 4 °C in 2 L 50 mM Hepes buffer containing 100 mM NaCl, pH 7.5 while stirring.

2.2.3 Anion exchange chromatography

A Q-Sepharose High Performance (Q-HP) column (GE Healthcare, USA) (column volume of 30 ml) was used for anion exchange chromatography. The Q-HP column was prepared and washed in the reverse direction of flow with 2 column volumes (CV) 2 M NaCl, followed by 2 CV distilled water, 2 CV 1 M NaOH and again with 2 CV distilled water. The column was then equilibrated in the normal direction of flow with 5 CV of 50 mM Hepes binding buffer containing 100 mM NaCl, pH 7.5.

The pre-column sample was passed through a 0.45 μm filter and loaded onto the Q-HP column at a flow rate of 2.5 ml/min, using a 50 ml superloop injection loop (GE Healthcare, USA) in conjunction with a Gilson 321 pump. The column was then washed with the Hepes binding buffer until the optical density stabilised at a baseline value.

Protein elution was monitored using a Waters 484 tunable absorbance detector (Waters, USA; set at 280 nm) and fractions collected using a Gilson FC 204 fraction collector (Gilson, USA).

Bound proteins were eluted by increasing the salt concentration over 10 column volumes (CV) using a linear gradient from 100 mM NaCl to 1 M NaCl. Fractions of 5 ml were collected in 60 tubes over 110 min. The column was finally washed with 5 CV of the Hepes elution buffer to ensure that no more protein was bound, and then re-equilibrated with 5 CV of the Hepes binding buffer before washing with distilled water and storing in 20 % ethanol.

The data were transferred to a computer using a Gilson 506C system interface and was interpreted using Gilson Unipoint software, version 5.11.

2.2.4 Removal of contaminating proteins by size exclusion chromatography

The size exclusion column was packed in-house with Sephacryl™ S300 high resolution beads (GE Healthcare, USA) to a column volume of 110 ml with a bed height of 63 cm and radius 0.8 cm. The column was used in conjunction with a Gilson 305 pump (at a flow rate of 0.5 ml/min), Gilson 806 manometric module, Gilson UV/Vis-151 absorbance detector (set at 280 nm) and a Gilson FC 203B fraction collector (Gilson, USA). The data were transferred to a computer using a Gilson 506C system interface and was interpreted using Gilson Unipoint software, version 3.3.

The column was washed and equilibrated with 2 CV 50 mM Hepes buffer containing 150 mM NaCl, pH 7.5. Protein-containing fractions from the Q-HP column were analysed by SDS-PAGE; those fractions containing a band of an apparent molecular weight (MW) similar to that of the calculated molecular weight of MshB were pooled and concentrated to 5 ml. The concentrated sample was passed through a 0.2 μm filter to remove particulate matter and loaded onto the column. The proteins were eluted with 2 CV buffer and fractions of 0.5 ml were collected in 80 tubes. Fractions containing putative MshB were pooled and concentrated to 4-5 ml which averaged 1 mg/ml protein concentration.

Protein-containing samples were either stored at 4 °C for use immediately or they were diluted with glycerol to a final concentration of 20 %, flash frozen in liquid nitrogen and stored at -80 °C.

2.2.5 Molecular weight estimation by size exclusion chromatography

A calibration curve was created by eluting seven standard molecules (Table 2-1) from a Sephacryl^{TM} S300 size exclusion column under the same conditions as the experiment.

Gel filtration molecular weight standards were bought from Bio-rad (USA) and used as per manufacturer's instructions.

Table 2-1: Table of gel filtration standards

Standard	MW
Tobacco Mosaic Virus	50000 kDa
Thyroglobulin	670 kDa
Gammaglobulin	158 kDa
Ovalbumin	44 kDa
Myoglobulin	17.5 kDa
Vitamin B-12	1.35 kDa
Acetone	0.58 kDa

The partition coefficient (K_{av}) is used to compare different sized molecules eluted off a size exclusion column by their elution volumes. It is calculated by subtracting the elution volume (V_e) from the void volume (V_o) and then dividing by the internal volume (V_{in}) (Equation 2.2-1).

Equation 2.2-1: Equation to calculate K_{av} for comparing elution volumes of different proteins. V_o is the void volume, V_e is the elution volume and V_{in} is the internal volume.

$$K_{av} = V_o - V_e / V_{in}$$

When K_{av} equals zero, that molecule elutes in the void volume ($V_e = V_o$). When K_{av} equals one, that molecule elutes in the internal volume plus the void volume ($V_e = V_{in} + V_o$). For very large molecules, the time spent in the internal volume is limited as they cannot fit into the pores of the beads and K_{av} approaches zero. For very small molecules, the time spent in the internal volume is maximised as they can fit into all the pores of the beads and K_{av} approaches one.

K_{av} was calculated for each molecule and was plotted against the log of each molecule's molecular weight (logMW), the equation of the resulting straight line was estimated using the trendline generated by Excel (Microsoft, USA).

The molecular weight of MshB was estimated by calculating its K_{av} value for the Sephacryl^{TM} S300 column, and substituting it into the equation from the standard curve (Equation 2.2-2).

Equation 2.2-2: Standard curve equation from S300 molecular weight standards.

$$LogMW = -4.4K_{av} + 6.5$$

2.2.6 Concentration of protein samples

Protein samples were concentrated using a stirred cell filtration unit (Amicon, USA) with a 10 kDa molecular weight cut-off ultra filtration membrane (Amicon, USA) at 4 °C.

2.2.7 Protein concentration determination

Protein concentration was determined using a Bradford protein assay (Bio-Rad, USA). The assay was performed as instructed in the manufacturer's manual. Bovine serum albumin was used as the standard in 0.1 mg/ml increments from 0.1 mg/ml to 0.5 mg/ml (Figure 2-1).

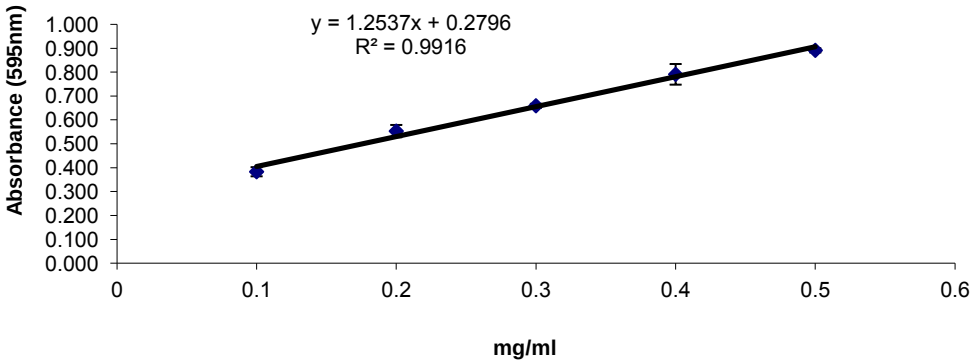


Figure 2-1: BSA standard curve.

The Bradford standard curve was made using BSA concentrations ranging from 0.1 to 0.5 mg/ml in 0.1 mg/ml increments. The resulting equation used for converting absorbance units to concentration (mg/ml) is presented on the graph. Measurements were performed in triplicate and the y-axis error bars are shown on the graph.

2.2.8 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

Sample purity and protein size was estimated using 16 % or 20 % SDS-PAGE stained with Coomassie Brilliant Blue R-250 as per the standard protocol (Laemmli 1970). Pageruler ™ prestained protein ladder and prestained protein ladder plus (Fermentas, Canada) were used as the molecular markers with 5 ul being loaded on the gel. Samples were prepared by mixing 5 ul of 4x sample application buffer (SAB) (200 mM Tris-HCl pH 6.8, 8 % SDS, 40 % glycerol, 4 % β-mercaptoethanol, 50 mM EDTA and 0.08 % bromophenol blue) with 15 ul sample and boiled for five minutes to ensure that the proteins were sufficiently denatured. Depending on the protein concentration of the sample, 5 – 20 ul were loaded onto the gel. The gels were run for 90 min at 10 mA in running buffer (0.025 M Tris-HCl, 0.192 M glycine, 0.1 % SDS, pH 8.3).

2.2.9 Protein Purity Estimation

Protein purity was estimated by densitometry using the computer program ImageLab (Bio-Rad) (Appendix A). SDS-PAGE analysis was carried out on samples from S300 size exclusion chromatography experiments and subsequently analysed with ImageLab. Lanes were created manually, while the bands were first detected automatically using the high sensitivity setting and then manually adjusted as necessary.

2.3 X-ray Crystallography

2.3.1 Crystallisation

Protein crystallisation refers to the technique where molecules in solution are allowed to precipitate out of solution slowly in a controlled manner creating a crystalline solid that produces an X-ray diffraction pattern. The most common way of achieving this is by forcing the effective concentration of the enzyme to increase in a slow, controlled manner. This can be done by regulating the evaporation of the solvent which in turn increases the concentration of the enzyme and other molecules in the solution that act as precipitants. The protein effectively reaches a supersaturated stage and once nuclei form, crystal growth can be achieved. Nuclei can take the form of anything present in the solution that initiates crystal formation. Crystals have to be grown until they are at least big enough to be hit with the X-ray beam. Bigger crystals with good long-range order will give better signal-to-noise ratios when compared to smaller crystals or crystals with poor long range order; it is therefore easier to obtain higher quality data from large, well-ordered crystals.

There are various crystallisation techniques; including batch crystallisation, liquid-liquid diffusion, hanging drop vapour diffusion, sitting drop vapour diffusion and dialysis; that crystallographers use to control precipitation conditions and induce crystal growth. In this study, the hanging drop vapour diffusion method was used to grow crystals (Jan Drenth 1999). Figure 2-2 is a simplified phase diagram used to explain the process of protein crystallisation with respect to protein and precipitant concentration.

Hanging drop vapour diffusion involves suspending a one-to-one solution of protein to precipitant solution on a cover slip over a well of concentrated precipitant. The well is sealed and the protein/precipitant solution is allowed to equilibrate with the precipitant in the well by evaporation (Figure 2-3).

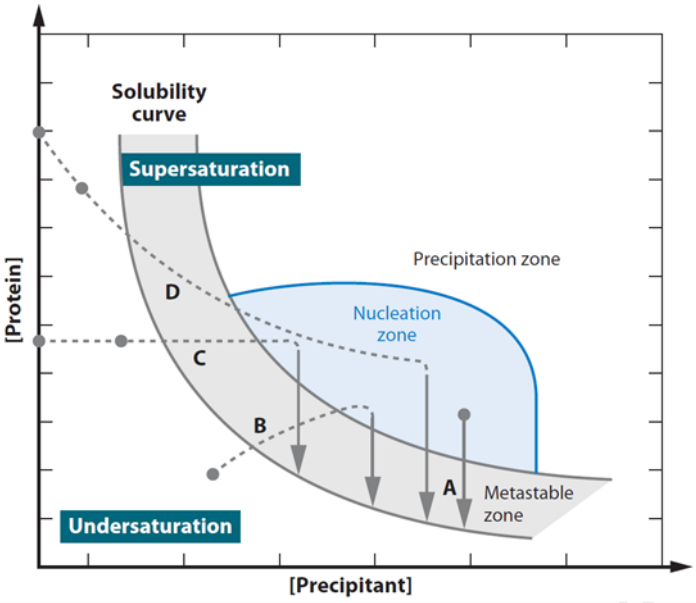


Figure 2-2: Simplified phase diagram for protein crystallisation.

Paths for batch (A), hanging/sitting drop vapour diffusion (B), dialysis (C), and liquid-liquid diffusion (D) are shown. Crystals only form in the zone of supersaturation (right of the solubility curve), where the concentration of the protein in the precipitant solution is above its solubility. In the zone of undersaturation (left of the solubility curve), crystals dissolve and cannot form. The metastable zone indicates conditions where crystals are stable and will not dissolve. In the zone of precipitation protein will precipitate in an unordered fashion. The zone of nucleation indicates the conditions where nucleation can occur thus initiating crystal growth (Figure and caption from Li and Ismagilov (Li, Ismagilov 2010) International Union of Crystallography©.

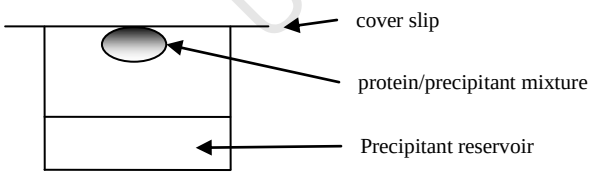


Figure 2-3: Hanging drop vapour diffusion

Hanging drop vapour diffusion involves suspending a solution of protein to precipitant (where the concentration of precipitant does not immediately cause precipitation) on a cover slip over a reservoir of concentrated precipitant. The well is sealed and the protein/precipitant solution is allowed to equilibrate by diffusion of water from the droplet to the reservoir. This results in the concentration of the precipitant causing crystal formation.

Protein crystallisation is a process involving many variables. In order to reduce these variables to a manageable number of conditions, it is common to use "sparse matrix" screening, i.e. a process of trial and error covering a wide range of different conditions. The phase diagram (Figure 2-2) shows how the protein and precipitant concentration affect protein solubility and illustrates how, over time, the concentration of protein changes during a crystallisation experiment. To grow crystals, conditions need to pass through the supersaturated zone and into the nucleation zone, without forming a precipitate. Once nucleation is achieved, the protein crystals should start growing and cause the concentration of protein to drop (Figure 2-2). A large number of diverse conditions need to be screened before the search can be narrowed to the optimum conditions. Once hits are obtained, conditions are optimised to improve the rate at which crystal growth occurs and improve the size of the crystals. These include changes to the concentration of the precipitant, the pH, the temperature and the starting concentration of the protein.

Alterations to the concentration of precipitant and protein have obvious consequences, as increasing the concentration of either will increase the rate of precipitation. Changing the pH of the solution alters the charges on the molecules that are present in solution, and thus changes the repulsion or attraction between molecules. Changing the temperature may either increase or decrease the solubility of the protein.

Prior to crystallisation attempts in this study, the protein solution was dialysed against 50 mM Hepes buffer, pH 7.5, overnight at 4 °C to reduce the salt concentration. MshB was incubated with a 5 times greater molar concentration of the VU5 inhibitor at 4 °C for 30 min before setting up the crystal trials. The stock concentration of MshB was approximately 19 mg/ml and was diluted as necessary during the optimisation procedure.

2.3.2 Identifying crystallising conditions for MshB

Initial crystallisation hits were identified using the Hampton crystal screen and PEG/ion screen kits (Hampton Research, USA). Equal volumes of protein and precipitant solution from the screening kits (2 µl each) were mixed on a cover slip and inverted and sealed with vacuum grease over 1 ml reservoir solution as per the hanging drop vapour diffusion method (Drenth, 1999). Protein samples were centrifuged at 5 000 x *g* for 1 min and filtered through a 0.2 µm filter prior to all crystallisation attempts. All crystallisation attempts were incubated at 22 °C and monitored daily. Hits were identified as those conditions that at best gave small crystals, or at worst formed a precipitate over the period of a week.

2.3.3 Optimisation of crystallisation conditions for MshB

Once conditions are found that show signs of precipitation, nucleation or crystal formation, the conditions around these hits are explored by altering the pH, precipitant concentration, protein concentration and temperature. The phase diagram shows how conditions that form precipitates can be considered as hits. If the precipitating factors in these conditions are decreased, the experiment can enter the nucleation zone before forming precipitate and can potentially form crystals instead. These optimisations are usually done using hanging or sitting drop methods as they can accommodate higher volumes and therefore bigger crystals. Once the optimum conditions have been found, crystallisation can be optimised further by methods such as streak seeding and macro seeding. Seeding is used to try control the number of nucleation sites. Streak seeding involves crushing or stroking existing crystals with a seeding tool so that shards of crystals get stuck to the seeding tool. Subsequently the seeding tool is rinsed to get rid of excess crystal shards so that only a few nucleation sites are introduced in an attempt to obtain a few large crystals rather than many small crystals. The seeding tool is then dragged through a drop of protein/precipitant mixture that has been

allowed to equilibrate for an hour, after which the drops are returned to their wells and allowed to equilibrate further. The few crystal shards that fall off into the protein/precipitant solution then act as nucleation sites in the equilibrated supersaturated drops and hopefully form a few large, good-quality crystals (Figure 2-4).

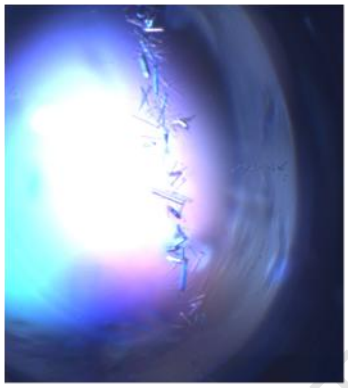


Figure 2-4: Crystals forming along a streak line after streak seeding.

Micrograph of *Haemophilus influenzae* hypothetical protein HI1161 crystals produced by streak seeding homologous protein through equilibrated drops prior to using the hanging drop vapour diffusion method (Georgiev *et al.* 2006).

Hits were optimised in this study by varying the precipitant concentration, the pH and/or the protein concentration. Protein concentration was varied from 0.5 – 7 mg/ml by diluting the stock protein appropriately with 50 mM Hepes buffer, pH 7.5. Protein and precipitant concentration were varied by diluting the hanging drops. This was achieved by the addition of 50 mM Hepes buffer, pH 7.5 in 0.5 µl increments, effectively diluting the precipitant and protein by 11% when 0.5 µl was added, 20% when 1 µl was added, 27% when 1.5 µl was added, and 33% when 2 µl was added.

2.3.4 Optimising crystal growth for MshB crystals

Crystal growth was optimised by micro-seeding existing MshB crystals into hanging drops that had been pre-equilibrated for 1 hour (homologous seeding). A home-made micro-seeding tool (human eyelash) was brushed against an existing crystal, washed in buffer (50 mM Hepes buffer, pH 7.5) to remove excess crystal shards, and subsequently drawn through the pre-equilibrated drops. The micro-seeding tool was washed in distilled water (dH_2O) and dried between successive micro-seeding attempts.

2.3.5 Preparation of MshB crystals for transport and X-ray diffraction

For X-ray data collection, crystals need to be transferred into a loop of nylon that is approximately the same size as the crystal and fixed onto a metal capillary that can be attached to the goniometer head. The goniometer is the device that alters the angle of presentation of the crystal to the X-ray beam in the diffractometer. To try to stabilise the crystal and also limit the damage caused by the X-ray beam, the crystals are soaked briefly in a cryo-protectant and then flash-frozen in liquid nitrogen. Freezing prevents evaporation of the solvent, which is important in stabilising the protein because solvent channels make up about 40-60 % of a crystal's volume; if the solvent were to evaporate, the crystal can collapse and be destroyed. The decrease in temperature also minimizes the vibrational energy of the atoms which can enhance the diffraction pattern. The cryo-protection and flash freezing are done to prevent ice crystals from forming, as ice crystals destroy the protein crystals and produce diffraction rings which can obscure data originating from the protein crystal. It is important to move the crystal through the air above the liquid nitrogen as fast as possible; the layer of cold air/evaporated nitrogen can start the cooling and freezing process which can complicate flash freezing (Rhodes, 2006).

Prior to storage in liquid nitrogen and subsequent transport to the beam-line, MshB crystals were mounted on a nylon loop and soaked in cryo-protectant (precipitant solution diluted to 20 % glycerol) for 10 seconds prior to being flash frozen in liquid nitrogen.

2.3.6 X-ray diffraction, model building and model refinement of MshB

Figure 2-5 provides an overview of the steps required to produce a protein structure from a protein crystal. Diffraction data were collected at beam-line BM14 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. Collecting data on a synchrotron has an advantage over home sources because of the increased beam intensity. BM14 also has an advantage in that data collection is highly automated. Data frames in this study were collected at one degree oscillations over a range of 180 °, under a stream of nitrogen vapour at a temperature of 100 K. The exposure time for each frame was 20 s and each frame was collected on a MAR225 CCD detector.

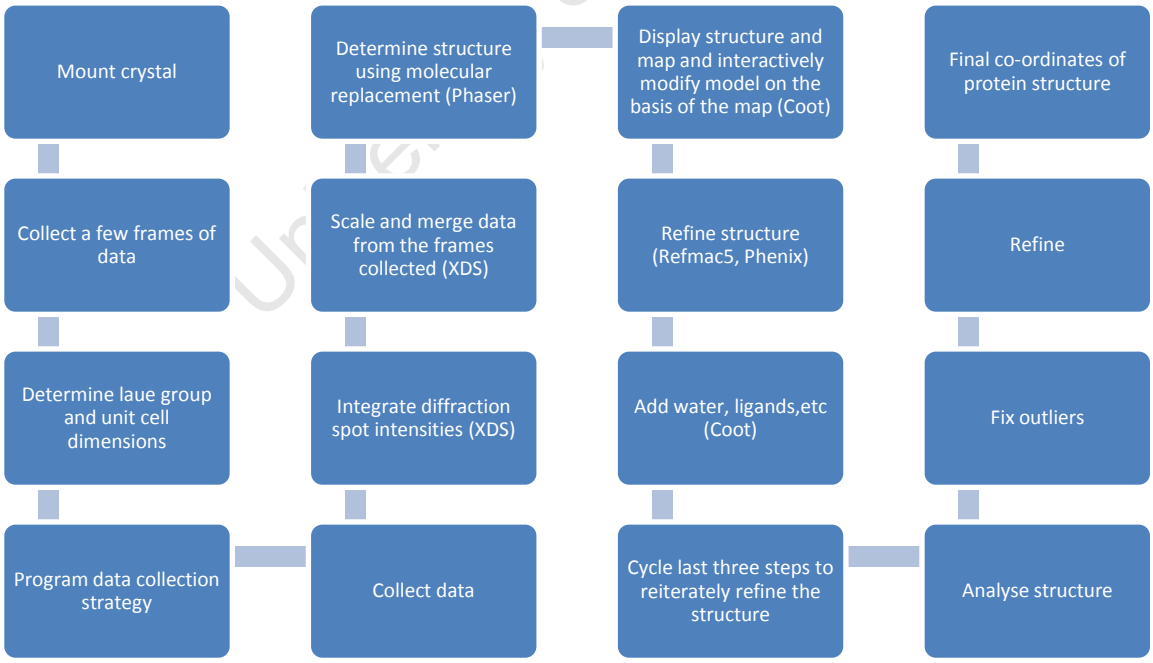


Figure 2-5: This chart describes the steps required to solve a structure from a crystal.

XDS (Kabsch, 2010) is one of many programs that are used for partially automated indexing, integration, scaling and refinement. First the crystal properties must be determined and refined by indexing and integration. The detector and experimental properties are entered in manually after which the program finds (and refines) the reflections (from two image frames 90° apart) and determines the unit cell dimensions and tries to work out the crystal symmetry, orientation and space group. Knowing these crystal parameters, the program can predict where reflections will be on each data frame, and which reflections will span more than one frame. Integration of the data provides three dimensional profiles from predicted reflection positions which XDS (Kabsch, 2010) uses to re-refine the crystal properties by minimizing the differences between the observed and calculated positions of the reflections. The data are then scaled so that symmetry related reflections throughout the dataset are of the same intensity and the reflections are corrected for absorbance by the crystal and surrounding mother liquor.

Phaser (McCoy *et al.* 2007) is one of several available programs for solving X-ray data and forms part of a suite of programs in CCP4 (Collaborative Computational Project number four, 1994). The advantage of Phaser is that it is highly automated and the steps involved in molecular replacement require very little user input. Phaser uses Patterson functions to determine the correct orientation of the model by first investigating the cross-rotation and then the cross-translation functions. The program then calculates the phases derived from the structural model which has been placed in the correct orientation. Patterson functions can be derived directly from amplitude data without phases; this is advantageous because during the collection of X-ray crystallographic data, the phase information is lost.

Like Phaser, Refmac (Murshudov, Vagin & Dodson 1997) is also part of the CCP4 suite. Refmac optimises the correspondence between the measured and calculated diffraction

intensities subject to restraints that the bond angles and bond distances should be correct. PHENIX (Adams *et al* 2010) was also used to refine bond angles and bond distances.

In short, the data were integrated, scaled and averaged using XDS (Kabsch, 2010), the phases were calculated using Phaser_MR (McCoy *et al.* 2007) and refined using Refmac_5 (Murshudov, Vagin & Dodson 1997) and PHENIX (Adams *et al* 2010) in conjunction with manual model-building using Coot V0.6.1 (Emsley *et al.* 2010) while 5 % of the data were kept aside to calculate R-free; model refinement was stopped when R-free stopped decreasing. The initial model used for calculating the phases was chain B of MshB from PDB entry 1Q7T (McCarthy *et al.* 2004). Molecular graphics were visualised using Chimera V1.5 (Pettersen *et al.* 2004).

2.4 Modeling of Mca

Homology modeling of Mca based in the structure of MshB determined in this dissertation was carried out using Modeller 9v8 (Sali, Blundell 1994) and the model with the lowest zDOPE score was chosen. zDOPE stands for a normalised Discrete Optimized Protein Energy (DOPE), which is an atomic distance-dependent statistical score where negative values indicate better models (Shen, Sali 2006). The Mca amino acid sequence was submitted to GenThreader (Lobley, Sadowski & Jones 2009); the highest score was an alignment with chain A of MshB with PDB code 1Q74 (Figure 3-20). The structure of MshB that was solved in this study (Figure 3-13) was used as the reference structure. The Mca secondary structural elements were constrained to those predicted by Psipred (Buchan *et al.* 2010) (Figure 3-21). CastP (Dundas *et al.* 2006), a program that analyses pockets, was used to compare the pockets found in the MshB structure solved here and those found in the Mca structure modelled here. A web based server, ProSA-web (Sippl, 1993, Weiderstein, Sippl, 2007), was used to validate the model generated by Modeller. The ProSA-web server provides an overall z-score which indicates overall model quality, and an energy plot which shows local model quality. The z-score is compared to z-scores of all deposited protein structures in the Protein Data Bank (PDB), therefore an acceptable z-score is one that falls within the range of z-scores found in the PDB. The energy associated with the amino acid sequence is acceptable if it shows negative values, and positive values show parts of the model that could be problematic (Weiderstein, Sippl, 2007).

2.5 Characterisation of MshB Activity

All plots were derived from experiments that were performed in triplicate and the standard deviation from the mean is shown using y-axis error bars. Where the errors bars cannot be seen, they are hidden by the size of the marker.

2.5.1 MshB activity assays

MshB activity was assayed with GlcNAc (*N*-acetyl Glucosamine) using an assay adapted from Huang and Hernick (2011).

Various concentrations of substrate (0, 5, 25, 50, 100, 150 and 200 mM GlcNAc) were made up to 1.5 ml in 50 mM Hepes buffer with 50 mM NaCl and 10 mM dithiothreitol (DTT), pH 7.5, and incubated in a 37 °C water bath prior to starting the reaction by the addition of enzyme at 1 µM final concentration.

Each reaction was performed in triplicate in 2 ml Eppendorf tubes. The Eppendorf tubes were vortexed before each time point was taken (0, 10, 20, 30, 40, 50 and 60 min). The reaction was stopped by removing a 200 µl sample and incubating it with 70 µl 20 % tri-chloro acetic acid (TCA) for 10 min in a 96 well plate. The pH was returned to neutral by removing a 25 µl sample into 75 µl 1 M sodium borate buffer, pH 9.0 in a separate 96 well plate. Once all time points were transferred into borate buffer they were derivatised with 30 µl 10 mM FSA (in acetonitrile) and incubated for 10 min at room temperature. Fluorescence was measured on a Cary Eclipse Fluorescence Spectrophotometer (Varian, Australia) using an excitation wavelength of 390 nm and an emission wavelength of 475 nm (Huang, Hernick 2011). The rate of GlcN produced was plotted and the initial rates were found using the linear portion of the graphs which corresponded to the 0-30 min time points.

2.5.2 Effect of DTT and VU5 on MshB activity assayed with GlcNAc substrate

The possibility of DTT inhibiting MshB activity was investigated by increasing the DTT concentration from 10 mM to 100 mM.

Additionally, the effect of VU5 on the activity of MshB alone and MshB with GlcNAc was investigated using various concentrations of VU5, 0.1, 1 and 5 mM in the presence of 10 mM and 100 mM DTT.

2.5.3 Graphics and structural analysis

Images of small molecules and mechanism pictures were re-drawn in Accelrys Draw (Accelrys 2011). Chimera V1.5 was used for structural analysis including alignment of protein molecules and hydrogen bond prediction (Pettersen *et al.* 2004).

3 Results

3.1 Protein expression

In order to obtain sufficient amounts of protein for crystallisation and activity assays, MshB was recombinantly expressed in *E. coli*.

Cells were grown to an OD_{600} of 0.4-0.6 at which time an uninduced sample was removed and recombinant expression of MshB was induced by the addition of IPTG. The cells were left to express overnight and another sample was removed. The two samples were analysed using SDS-PAGE (Figure 3-1). After comparison of the total protein content in the cells pre- and post- induction, protein with an apparent molecular weight of 33kDa (corresponding to the calculated MW of MshB) was found to be over-expressed after overnight induction. The recombinant method of over-expression yielded sufficient MshB protein to carry on with purification.

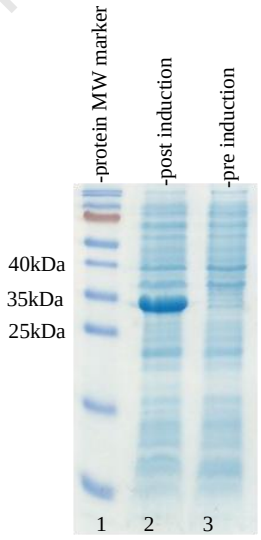


Figure 3-1: 20 % SDS-PAGE showing expression of MshB

Protein with an apparent MW of 33 kDa, corresponding to the calculated MW of MshB was found to be over-expressed in the IPTG induced fraction.

3.2 Protein purification scheme

3.2.1 Purification scheme overview

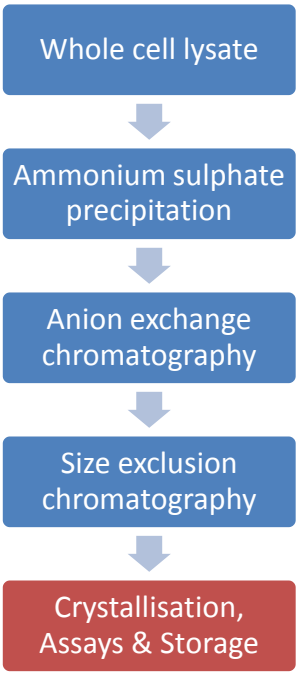


Figure 3-2: Purification scheme overview

Flow diagram giving a basic overview of the steps involved in purification of MshB in this study (Blue blocks). The final block in the scheme shows the different uses of the purified protein samples (Red block). Each time MshB was purified, the procedure was carried out in full (I.E. each step was carried sequentially and once only, per round of purification).

The soluble fraction of whole cell lysate was subjected to ammonium sulphate precipitation. The 35-55 % ammonium sulphate cut subjected to anion exchange chromatography and finally, the fractions containing MshB were passed through a size exclusion column to further improve purity and to examine potential oligomerisation.

The results that follow are representative of different repeated steps of a single purification scheme.

3.2.2 Ammonium sulphate precipitation

After overnight expression the cells were lysed by sonication and the soluble cell content was separated from the insoluble cell content by centrifugation. Ammonium sulphate precipitation was carried out on the soluble cell content as the first purification step in order to remove as many contaminating proteins as possible before performing anion exchange chromatography. This was done to improve the efficiency of the anion exchange step.

Initially a range of ammonium sulphate cuts were performed to investigate the appropriate percentage of ammonium sulphate that gave the best separation of putative MshB from contaminating *E.coli* proteins. Soluble cell content was sequentially incubated with increasing percentages of ammonium sulphate, after each increase the precipitated protein was removed and resuspended in 50 mM Hepes buffer containing 100 mM NaCl, pH 7.5 for analysis by SDS-PAGE. The best separation was found to be given by a 35-55 % ammonium sulphate cut of the soluble cell content. The following result (Figure 3-3) shows representative analysis of subsequent experiments where the soluble cell content was subjected to 35 % ammonium sulphate precipitation, after which the supernatant fraction (SNF) was removed and subjected to 55 % ammonium sulphate precipitation.

SDS-PAGE revealed the presence of a band indicating over-expressed protein with an apparent molecular weight similar to that of the calculated molecular weight of MshB in the soluble cell content and in the resuspended 35-55 % ammonium sulphate cut. The 35-55 % ammonium sulphate precipitation cut successfully separated the putative MshB from many contaminating proteins (lane 8 Figure 3-3) and was dialysed, centrifuged and filtered prior to anion exchange chromatography.

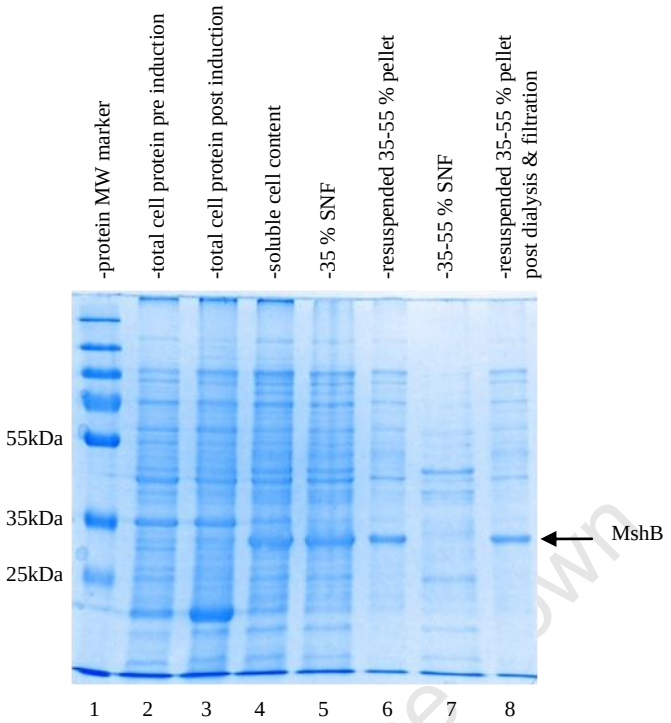


Figure 3-3: 16 % SDS-PAGE of ammonium sulphate precipitation steps

The band corresponding to MshB is indicated. Ammonium sulphate precipitation was effective in removing some contaminating *E.coli* proteins as can be seen by comparing lane 4 with lane 8.

3.2.3 Anion exchange chromatography

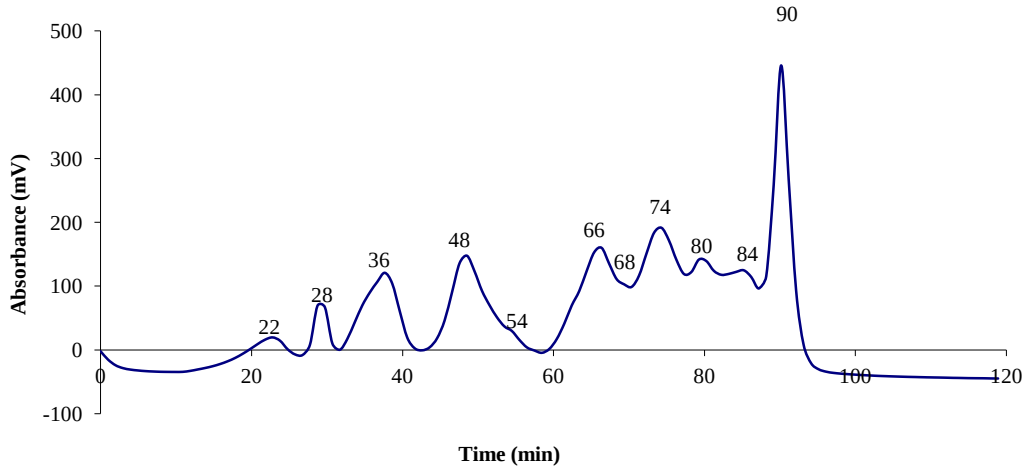
Anion exchange chromatography was used in an attempt to further separate MshB from contaminating *E.coli* proteins. In order to optimise the purification of MshB, multiple attempts with various techniques were employed. In the first round of purification the ammonium sulphate precipitation step was not included, and will be discussed first.

Soluble cell content obtained from *E.coli* that was induced to express putative MshB overnight was filtered and loaded onto a QHP anion exchange column with a column volume (CV) of 30 ml. The column was pre equilibrated with 50 mM Hepes binding buffer containing 100 mM NaCl, pH 7.5. Proteins were eluted in 5 ml fractions using a linear

gradient of NaCl from 100 mM to 1 M over 10 column volumes (CV) at a flow rate of 2.5 ml/min.

Figure 3-4 shows the resulting chromatogram and corresponding SDS-PAGE analyzing column fractions. Many protein peaks eluted from the Q-HP column, with the first three separating well from other peaks while the remaining peaks were overlapping with other peaks.

(a)



(b)

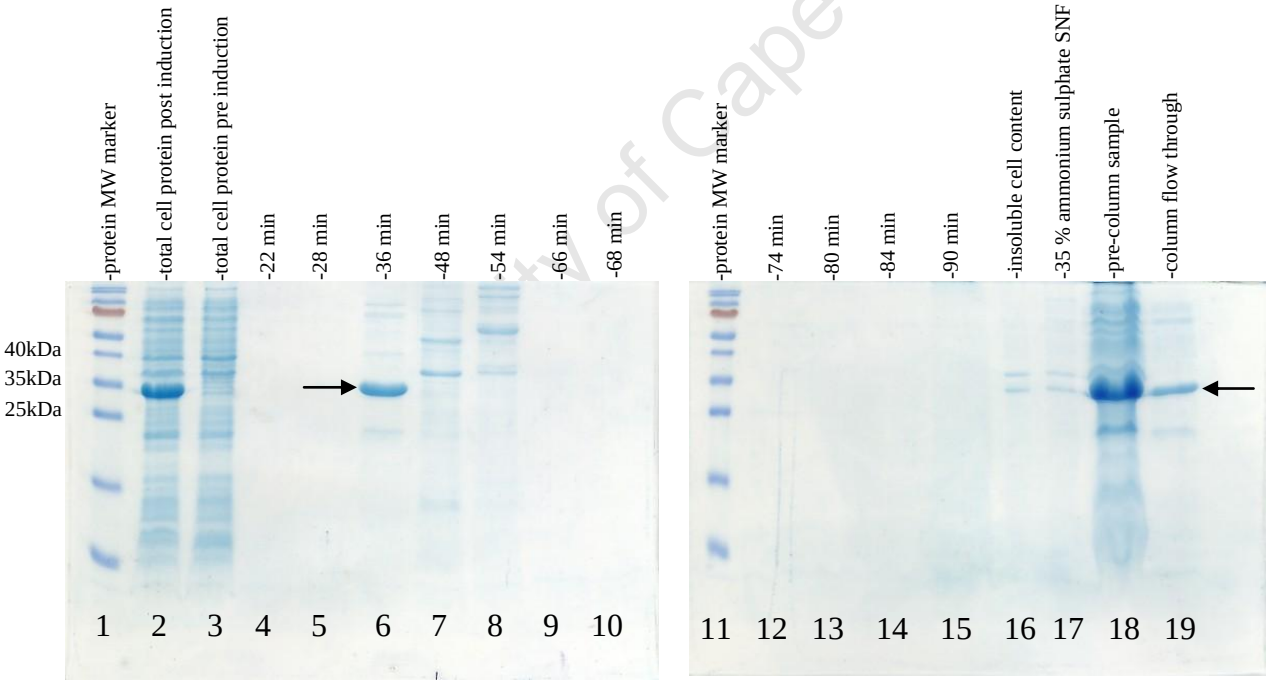


Figure 3-4: Analysis of soluble cell content eluted off an anion exchange column.

(a) Chromatogram of soluble cell content eluted off an anion exchange column. Peaks are labeled according to elution time.

(b) Analysis by SDS-PAGE found that MshB eluted at 36 minutes (lane 6) and that a considerable amount of MshB was seen in the flow through fraction (lane 19). The volume of the flow through was 200 ml. Bands corresponding to MshB are labeled with an arrow.

Analysis of the fractions by SDS-PAGE (Figure 3-4) revealed that the fractions containing MshB eluted off the column in a discrete peak at 36 min. This showed that anion exchange chromatography was effective in separating putative MshB from the majority of *E.coli* contaminating proteins.

It was noticed however that the flow through (FT) contained a significant amount of putative MshB (Figure 3-4 (b), lane 19). To investigate whether this putative MshB was not binding to the column due to it being outcompeted by *E.coli* proteins or if the putative MshB was just not interacting with the column for some reason, the flow through was re-loaded onto the column and the resulting chromatogram is shown in Figure 3-5.

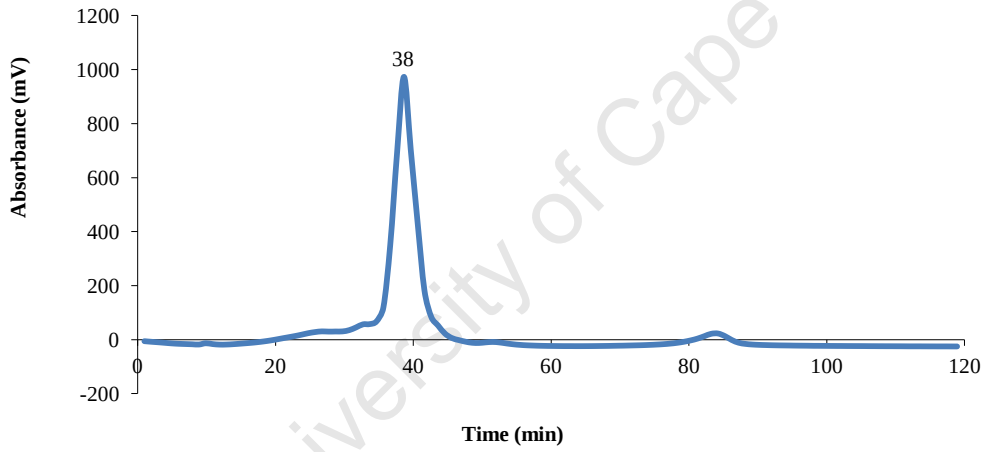


Figure 3-5: Chromatogram of the FT from the previous run, eluted from Q-HP.

The peak corresponding to MshB eluted off the column at 38 min.

SDS-PAGE analysis of column fractions (Figure 3-5) showed that MshB eluted at 38 min. This result suggested that removal of contaminating proteins might improve the performance of anion exchange chromatography. For this reason ammonium sulphate precipitation was included as a purification step prior to anion exchange chromatography in subsequent purification attempts.

Figure 3-6 is a representative chromatogram of anion exchange chromatography of the soluble cell content after ammonium sulphate precipitation. The chromatogram shows fewer peaks when compared to Figure 3-4 (a) and the MshB containing peak at 36 min (corresponding to fraction 19 & 20) was well resolved after the inclusion of the ammonium sulphate precipitation step prior to anion exchange chromatography.

Figure 3-7 is a representative SDS-PAGE analysis of subsequent purifications where an ammonium sulphate precipitation step (Figure 3-3) was included prior to elution from the Q-HP column.

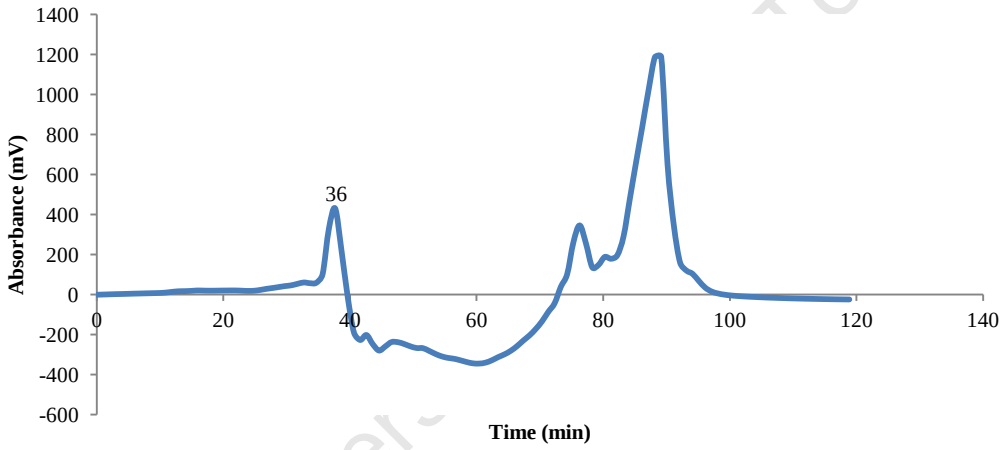


Figure 3-6: Representative chromatogram of anion exchange chromatography of the soluble cell content after ammonium sulphate precipitation.

MshB eluted off the column at 36 min and was well separated from the other peaks present on the chromatogram. The apex of the peak was split over two fractions, 19 and 20.

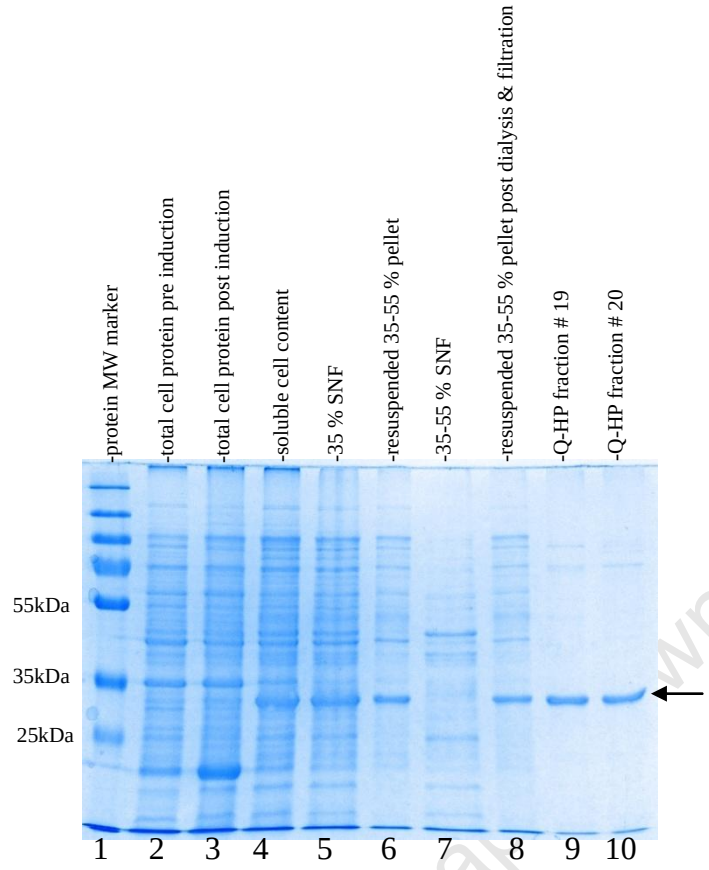


Figure 3-7: SDS-PAGE analysis of MshB purification by ammonium sulphate precipitation followed by anion exchange chromatography.

16 % SDS-PAGE analysis of representative samples from ammonium sulphate precipitation and subsequent Q-HP anion exchange chromatography steps. The bands corresponding to MshB are labeled with an arrow. It is important to note that there are still some high MW contaminants present. The QHP fractions 19 and 20 constituted the protein-containing peak at 36 min in Figure 3-6

There were still high MW contaminating proteins in the MshB containing fractions after anion exchange chromatography (Figure 3-7, lane 9 & 10), therefore these fractions were pooled and concentrated and subjected to size exclusion chromatography. .

3.2.4 Sephacryl™ S300 calibration

Size exclusion chromatography (SEC) was used as the final purification step, apart from removing high MW contaminants from the MshB containing sample, SEC allows estimation of the molecular weight and investigation of the possibility of quaternary structure.

Molecular weight standards were eluted from the Sephacryl™ S300 column under identical conditions to those of the MshB elution (Figure 3-8). That is, Sephacryl™ S300 was packed in house to a column volume of 110 ml. 0.5 ml fractions were collected over 2 column volumes at a flow rate of 0.5 ml/min with 50 mM Hepes buffer containing 150 mM NaCl, pH 7.5.

(a)

Standard	MW	Label on chromatogram	Elution time
Tobacco Mosaic Virus	50000 kDa	1	107 min
Thyroglobulin	670 kDa	2	120 min
Gammaglobulin	158 kDa	3	157 min
Ovalbumin	44 kDa	4	174 min
Myoglobulin	17.5 kDa	5	197 min
Vitamin B-12	1.35 kDa	6	239 min
Acetone	0.58 kDa	7	263 min

(b)

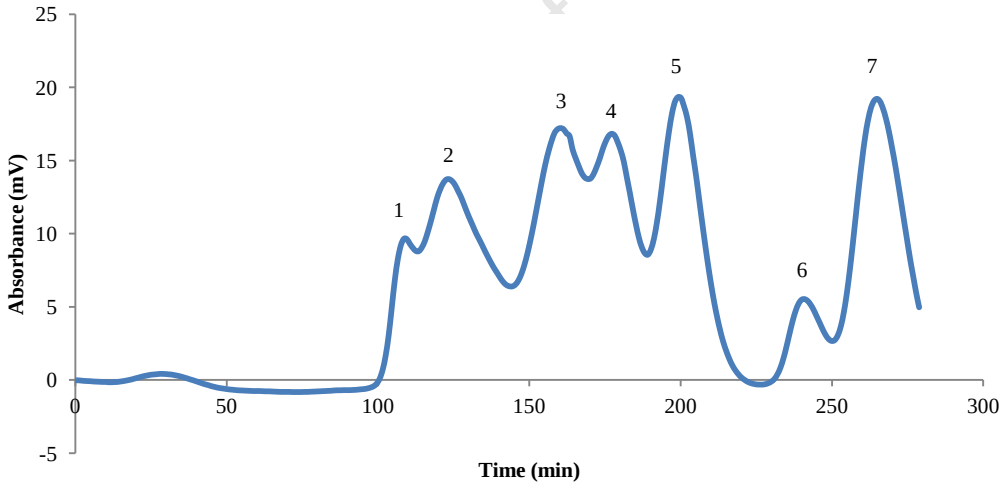


Figure 3-8: Calibration of the Sephacryl™ S300 column

(a) Table of protein standards used in the Sephacryl™ S300 size exclusion column calibration showing their respective molecular weights, elution time and label on chromatogram.

(b) Chromatogram of protein standards passed through the Sephacryl™ S300 size exclusion column.

The apex of the peaks corresponding to the size exclusion chromatography MW standards were separated by the Sephacryl^{TM} S300 column, however many of the sides of the peaks were overlapping and therefore poorly resolved. Despite this, the column should still be able to separate putative MshB (33 kDa) from high MW contaminants visible in Figure 3-7. The calibration curve was calculated by plotting the LogMW of the standards against their partition coefficient (Figure 3-9).

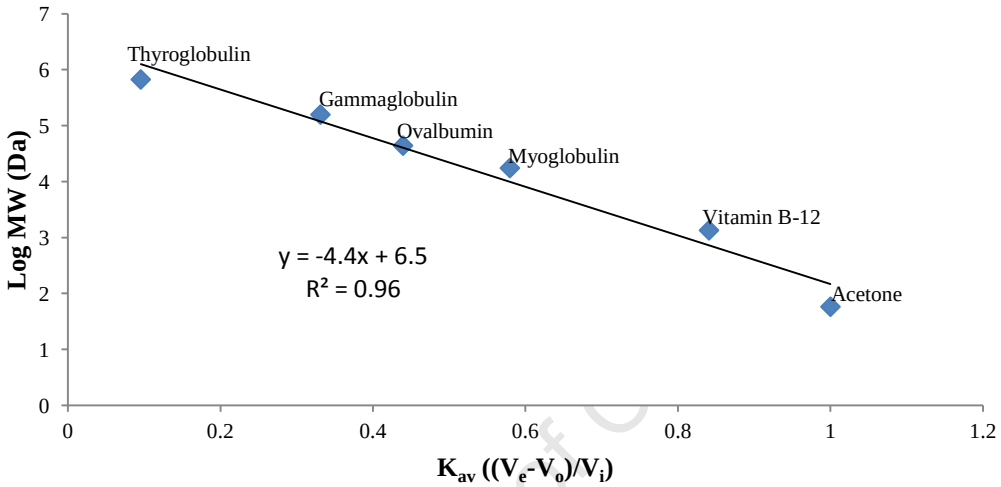


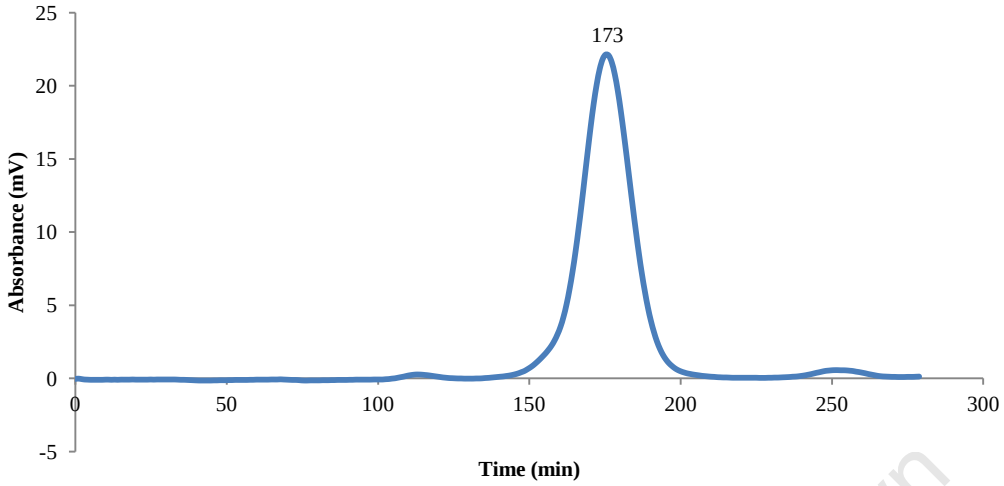
Figure 3-9: Sephacryl^{TM} S300 Calibration curve

This curve was plotted using data obtained in Figure 3-8 The higher MW proteins have a partition coefficient tending towards 0 while the lower MW proteins have a partition coefficient tending towards 1 as expected.

3.2.5 Sephacryl^{TM} S300 Size exclusion chromatography of the MshB containing fractions

Fractions containing MshB from the anion exchange run were pooled, concentrated, filtered and loaded as an individual 5 ml sample onto the Sephacryl^{TM} S300 size exclusion column and were eluted under identical conditions to those of the size exclusion standards (Figure 3-10).

(a)



(b)

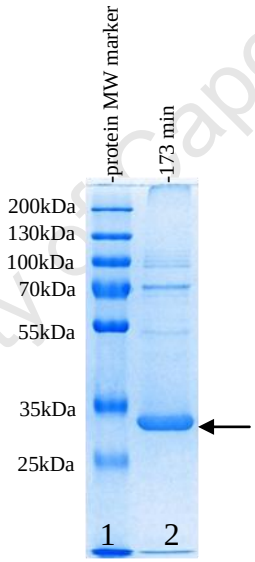


Figure 3-10: Sephacryl™ S300 Size exclusion analysis

(a) The chromatogram shows one major peak at 173 min. (b) SDS-PAGE analysis of this peak shows MshB indicated by an arrow, and persistent high molecular weight contaminants.

Figure 3-10 shows that a major peak eluted at 173min. SDS PAGE analysis of this peak indicated that the majority of the protein corresponded to MshB. The peak seemed well resolved, although small amounts of high MW contaminants were present when the fractions were analysed by SDS-PAGE.

The estimated molecular weight of the putative MshB was calculated to be 46 kDa using the K_{av} value obtained from the size exclusion experiment, and the equation calculated from the calibration curve (Figure 3-9). Even though the calculated MW is higher than the published MW of approximately 33 kDa, this is most likely due to a lack of sensitivity as the peaks on the chromatogram of the SEC standards (Figure 3-8) are not well resolved and the least squares regression analysis of the calibration curve falls below the molecular weight standards used that represent 44 and 17.5 kDa (Figure 3-9). MshB most likely elutes as a monomer as there is no evidence that indicates dimer formation, although in future a thorough investigation should be performed to confirm this.

Size exclusion chromatography did not significantly separate MshB from high MW contaminants.

3.2.6 Purification table

A purification table (Table3-1) was drawn up based on specific activity determined by activity assays using GlcNAc as a substrate and the total protein concentration determined by a Bradford assay calibrated using BSA (Figure 2-1). The purification table shows that a large proportion of MshB activity was retained over the duration of purification, with a purification factor of over 70 fold and a final yield of over 70 %. The pooled and concentrated S300 protein containing fractions were used in all further crystallography trials and enzyme assays as these samples were deemed to be the most pure. Protein purity was estimated by densitometry using the computer program ImageLab (Bio-Rad). An SDS-PAGE analysis was carried out on the protein containing fractions from S300 Size Exclusion Chromatography elution (four fractions centred on the peak of the resulting S300 chromatogram) and subsequently analysed with ImageLab (Appendix A). The average purity of the band corresponding to the estimated MW of MshB in each lane was estimated to be almost 85 %.

Table3-1: MshB purification table

Sample	protein concentration (mg/ml)	Total volume (ml)	Total protein (mg)	Activity units (mM/min)	Total activity (units.ml)	Specific activity (mM/min/mg)	Yield (%)	Purification factor
Soluble cell content	7.29	17.0	123.96	5.50E-03	9.35E-02	4.44E-05	100.00	1.00
55% pellet sample from ammonium sulphate precipitation	1.79	16.0	28.60	5.80E-03	9.28E-02	2.03E-04	99.25	4.57
Fraction #19 from Q-HP	1.11	5.0	5.54	1.24E-02	6.20E-02	2.24E-03	66.31	50.44
Fraction #20 from Q-HP	0.98	5.0	4.88	2.26E-02	1.13E-01	4.63E-03	120.86_a	104.31_a
S300 protein containing fractions, pooled and concentrated	1.04	4.5	4.66	1.50E-02	6.75E-02	3.22E-03	72.19	72.49

The value designated by "$_a$" in Table3-1 seems unlikely at first glance but could be explained by the possibility that partial separation of active and inactive MshB provided an artificially increased specific activity in fraction 20, this produced artificially increased yield and purification factor results.

3.3 X-ray Crystallography of MshB crystals

An attempt was made to co-crystallise purified MshB (MshB containing fractions that were pooled and concentrated after S300 Size Exclusion Chromatography during this study, and MshB containing samples that were stored frozen by Dr Marakalala (2008)) with VU5 in order to investigate the crystal structure with a bound inhibitor.

3.3.1 Identifying and optimising crystallizing conditions

Purified MshB protein was incubated with 5 times molar concentration of VU5 and subsequently used in crystallisation trials using the hanging drop method with the Hampton crystal screen and PEG/ion screen kits. The VU5 and MshB sample was mixed 1:1 with mother liquor on a cover slip and allowed to equilibrate until crystals were noticed in the hanging drop. The conditions that provided crystals were optimised by altering the pH and protein concentration. Crystals that were grown were then used to streak seed the most promising conditions that were identified after optimisation. Figure 3-11 shows crystals that were grown and used to obtain diffraction data.

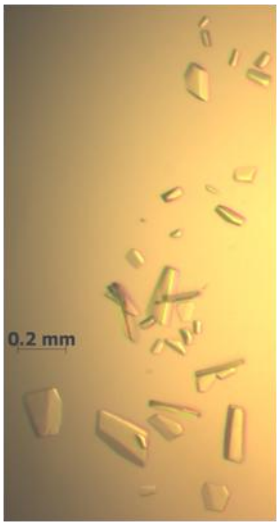


Figure 3-11: MshB crystals grown after streak seeding

A pre-equilibrated hanging drop containing MshB (pre-incubated with VU5) and mother liquor was streak seeded with MshB micro crystals and allowed to equilibrate further at room temperature.

Crystals were found to grow best overnight in mother liquor containing 0.1 M sodium acetate, 0.2M ammonium sulphate, 25% PEG 4000, pH 4.6, that had been diluted 1:1 with 4 mg/ml MshB (purified in this study and pre-incubated for 30 min with 5 times molar concentration of VU5 i.e. approximately 500 µM VU5) in 50 mM Hepes pH 7.5 and streak seeded with micro crystals (initial crystals were made using MshB purified by Dr Marakalala (2008)), using hanging drop vapour diffusion. Approximate crystal dimensions were 0.2 mm x 0.25 mm x 0.1 mm (Figure 3-11).

The protein material used to make crystal seeds was material that had been expressed, purified and stored at -80 °C in 20 % glycerol by Dr Marakalala during his PhD (Marakalala 2008). This material was used (as opposed to material produced during this study) because the material produced in this study failed to form crystals without the help of homologous seeding. Dr Marakalala included a zinc bound Immobilised Metal Affinity Chromatography (IMAC) step in his purification procedure. This purification step presumably removed non-

zinc bound MshB, thus producing a more homogeneous sample of zinc bound MshB which was easier to crystallise.

3.3.2 Crystal quality, space group and unit cell determination

Synchrotron diffraction data were collected from a single crystal at the ESRF beamline BM14. The crystal diffracted to 1.95 Å and was indexed in space group *C*2.

The unit cell dimensions were as follows: a = 147.42 Å, b = 71.90 Å, c = 97.69 Å, with angles α = 90 °, β = 128.29 °, γ = 90 °. The data were processed, scaled and integrated using XDS (Kabsch, 2010). Table 3-2 shows the data collection statistics.

Table 3-2: Data collection statistics

Data collection statistics*	
Data set source	BM14[a]
Space group	*C*2
Unit cell parameters	a = 147.72 (Å)
	b = 71.90 (Å)
	c = 97.69 (Å)
	α= 90.0°
	β= 128.29°
	γ= 90.0°
Molecules in ASU	2
Resolution range (Å)	76.67-1.95 (2.07 - 1.95)
No. of observed reflections	221610 (35224)
No. of unique reflections	58226 (9287)
Completeness (%)	99.3 (98.8)
Redundancy	3.8 (3.8)
Mean (I) /$\sigma(I)$	16.66 (3.68)
R_{merge} (%)	8 (49)

3.3.3 Molecular replacement and structure refinement

The diffraction data resulting from the MshB containing crystals was solved by molecular replacement as there are models deposited in the Protein Data Bank that were previously solved.

The model used for molecular replacement was chain B of MshB from the PDB entry 1Q7T. After an acceptable solution was found by the program Phaser (McCoy *et al* 2007), the model was built into electron density that was unaccounted for using the program Coot (Emsley *et al* 2010). Refinement was performed using the programs PHENIX (Adams *et al* 2010) and

REFMAC5 (Murshudov *et al* 2011). Alternating cycles of interactive model building and automatic structure refinement were performed until all major density was accounted for in the model and until the R-factor stopped decreasing. The coordinates were deposited in the Protein Data Bank with code 4EWL.

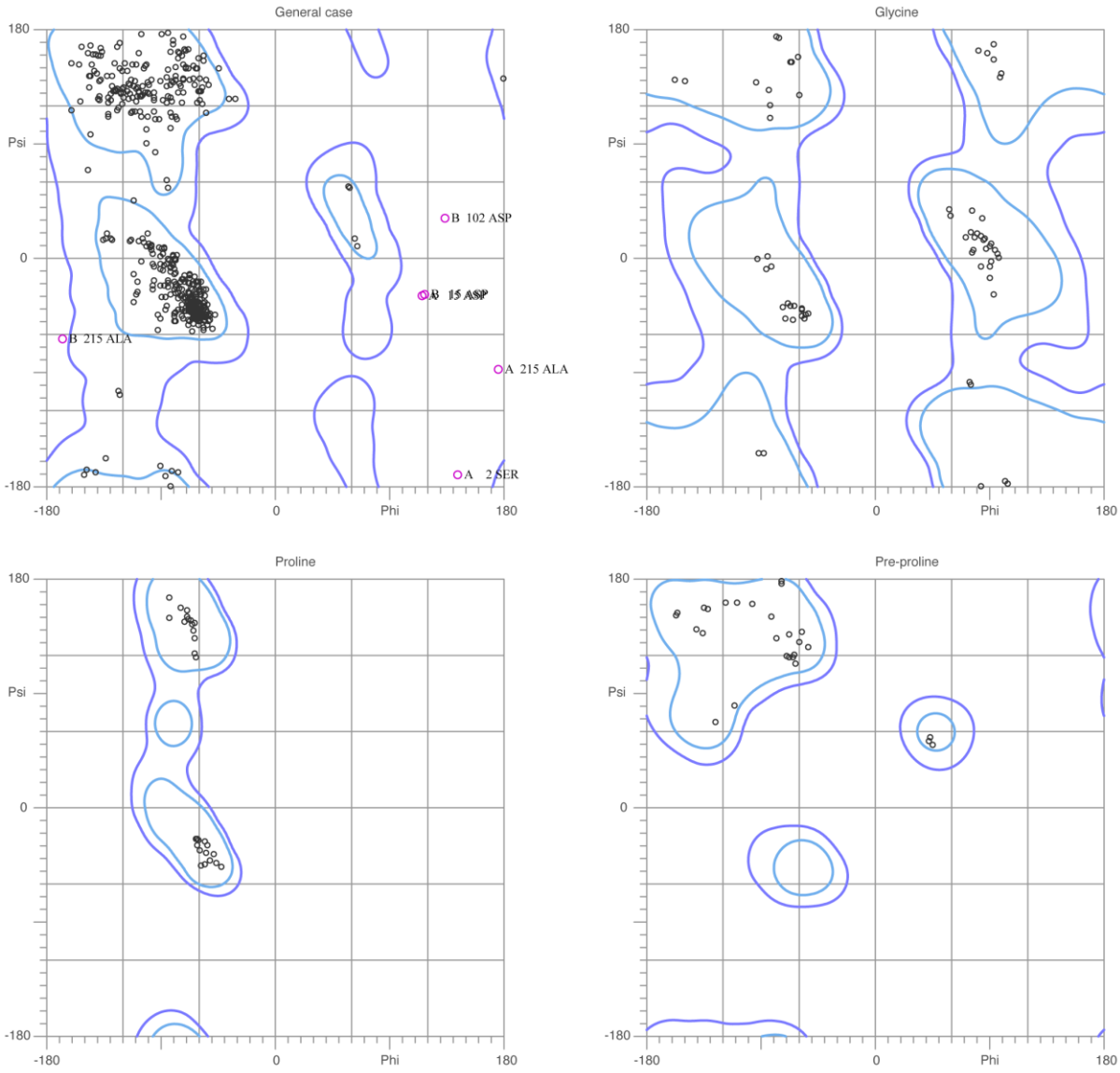
Refinement statistics are shown in Table 3-3, and the resulting Ramachandran plot is shown in Figure 3-12. It will be shown further on that the strained configuration of Asp-15 might be due to its position and likely action as general acid base catalyst in the active site

Table 3-3: Refinement and model validation statistics

Refinement statistics	
Resolution range (Å)	76.67 – 1.95
No. of atoms	
Protein	4382
Ligand/ion	112
Water	570
R-factor (%)	17
R-free (%)	20
Average B value (Å^2)	28.5
Model validation	
R.m.s. deviations from ideality	
Bond lengths (Å)	0.019
Bond angles (°)	1.413
Bad rotamers	22
Ramachandran outliers	9

R-factor = Σ_h $|F_o - F_c|$ / Σ_h F_o where F_o and F_c are observed and calculated structure factor amplitudes of reflection *h* of the working set of reflections, respectively. R-free is equal to R-factor for *h* belonging to the test set of reflections.

* Data in parentheses refer to the highest resolution shell.



96.4% (565/586) of all residues were in favored (98%) regions.
99.0% (580/586) of all residues were in allowed (>99.8%) regions.

There were 6 outliers (phi, psi):
A 2 SER (143.6, -170.2)
A 15 ASP (115.5, -29.1)
A 215 ALA (175.4, -87.4)
B 15 ASP (117.2, -28.6)
B 102 ASP (134.0, 32.8)
B 215 ALA (-168.8, -63.4)

Figure 3-12: Ramachandran analysis output from MolProbity (Lovell *et al.* 2003).

Image made with Kinemage (http://kinemage.biochem.duke.edu). This analysis shows that six of the amino acids have phi and psi angles that are not normally represented. The amino acids were modelled with these angles based on the calculated electron density available, and could not be modelled in any other way. It is important to note that Asp-15 has a strained configuration and that this is likely due to its position and likely action as the general acid base catalyst in the active site.

3.3.4 Structural analysis of MshB

Two MshB molecules were found in the asymmetric unit (chain A and B, Figure 3-13).

Figure 3-13: Cartoon representation of solved crystal structure of MshB

MshB (chain A in green, chain B in gold) showing the PEG (C) molecule at the interface, and the Zn^{2+} ion (grey sphere), acetate and glycerol molecules (oxygen atoms in red, carbon atoms coloured the same as the respective chain) in both active sites (A and B) present in the unit cell. The loop which is visualised in this structure for the first time is designated by D.

An acetate, zinc ion and glycerol molecule were found in the active site of each of the two MshB molecules (Figure 3-14), two molecules of polyethylene glycol 4000 (PEG) were found at the interface between the two units (The PEG could be one molecule, but no density

was found that connect the two). A further seven glycerol molecules were found interacting with the surface of the protein. No density corresponding to VU5 was found.

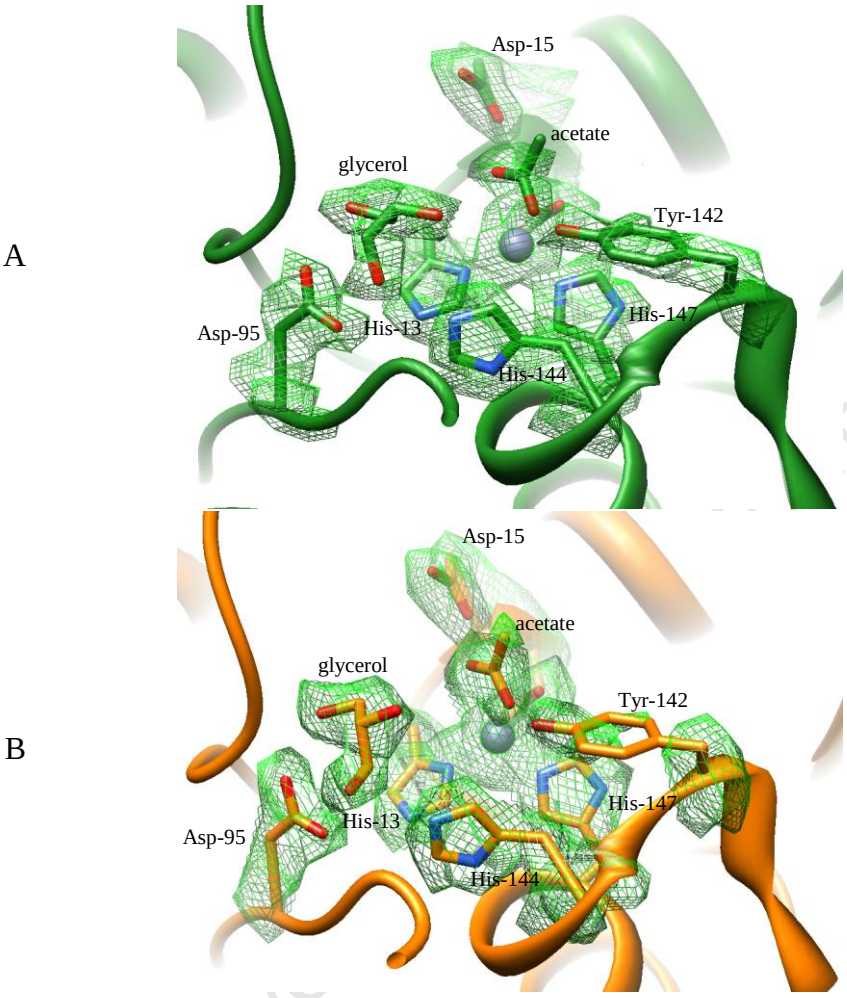


Figure 3-14: Electron density maps of both active sites in the MshB structure solved here.

The electron density maps show the density ($2F_{obs}$-F_{calc}) after the atoms were modelled. The atoms and bonds have been modelled as sticks. The map is contoured at 1 sigma. MshB chain A (A) is shown in green, chain B (B) in gold, zinc atom as a grey sphere.

The density corresponding to chain A enabled every amino acid from1 to 299 to be modeled. This is the first time the loop comprising residues 165 to 170 has been interpreted (Figure 3-15). The visualization was possibly assisted because the loop is stabilised by interactions with Glu-231A and Gln-232A of symmetry related molecules.

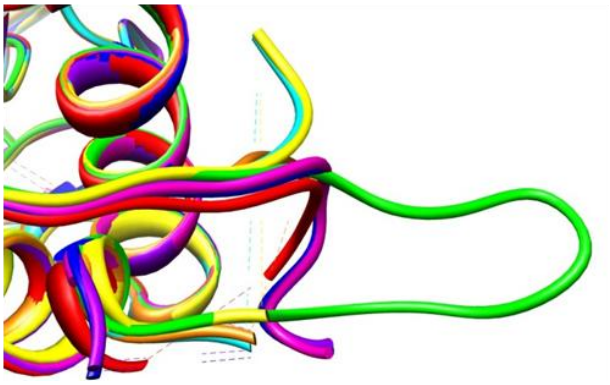


Figure 3-15: The loop in chain A of MshB which is described here for the first time

The loop described here for the first time found in Chain A of the structure solved here (green) compared to all deposited chains of MshB in the PDB. Chain B of the structure solved here is shown in orange. MshB with PDB code 1Q7T chain A is shown in light blue. MshB with PDB code 1Q7T chain B is shown in yellow. MshB with PDB code 1Q74 chain A is shown in red. MshB with PDB code 1Q74 chain B is shown in dark blue. MshB with PDB code 1Q74 chain C is shown in light purple. MshB with PDB code 1Q74 chain D is shown in dark purple.

Chain B is missing density corresponding to the residue range 165 to 172 and was only modelled up to residue 298. The structure revealed a large six-stranded β-sheet sandwiched between a series of α-helices (two on one side and three on the other) with a smaller three-stranded β-sheet creating a "roof" on the active site.

In Figure 3-16 the relatively high B factors corresponding to loops labelled A (residue range 207-217) and B (residue range 163-174) are areas of the protein that are most likely to be quite flexible and are located far from the active site. The high B factors corresponding to the loops labelled D (residue range 97-104) are situated very close to the active site and their flexibility could indicate an induced fit model which is stabilised when the natural substrate (GlcNAc-Ins) binds.

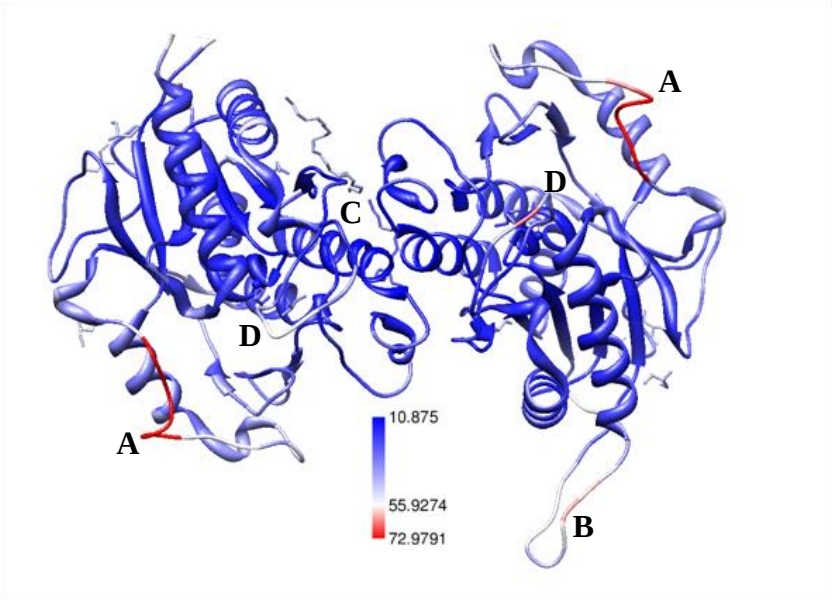


Figure 3-16: MshB showing the B factors of the modelled residues in cartoon representation.

The low B factors in the area of C correspond to the areas surrounding the active site which is likely to be stabilised by the bound zinc ion, glycerol and acetate, but also by the interaction between the molecules mediated by PEG. The B factor scale indicates that low B factors are represented by the colour blue, and high B factors are represented by the colour red.

There is a cavity that is formed by the two β-sheets and a short bend-forming α-helix next to the catalytic zinc ion (Figure 3-17).

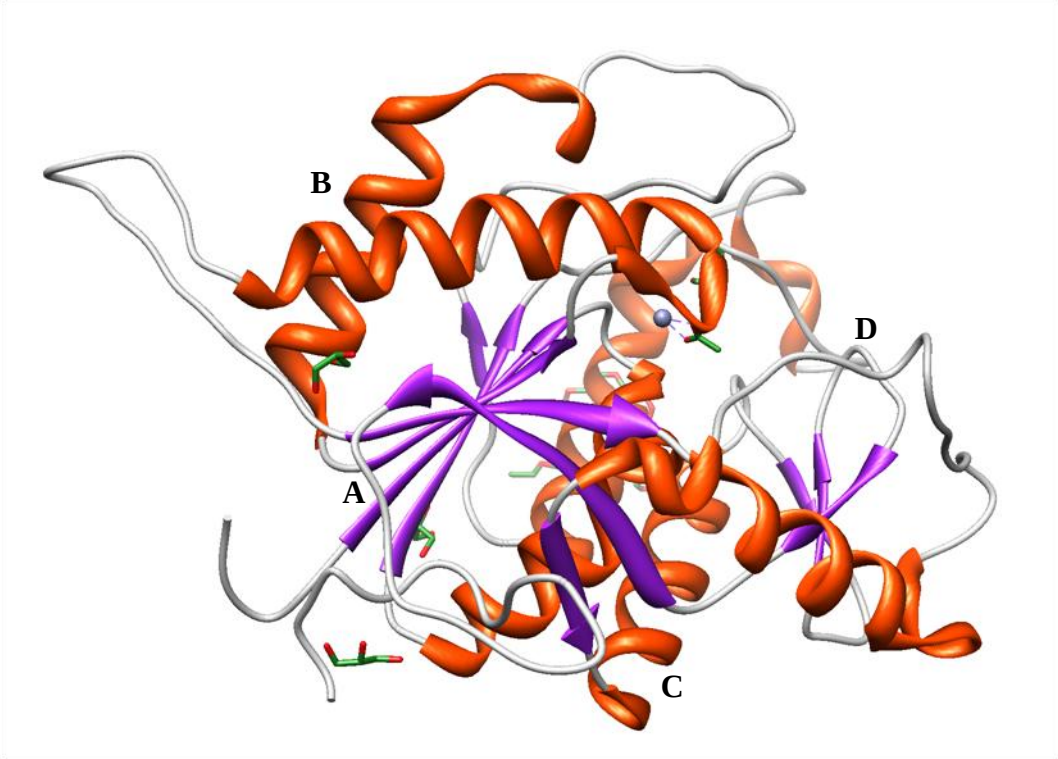


Figure 3-17: Cartoon representation of Chain A of MshB.

The zinc ion is shown as a grey sphere and depicts the location of the active site. (A) is a large β-sheet that is sandwiched between the groups of α-helices indicated by (B) and (C). (D) is the smaller β-sheet that forms part of the "roof" of the active site. The glycerol, acetate and PEG molecules are modeled as sticks, with their carbon atoms in green and their oxygen atoms in red.

In both chains the zinc ion was found interacting with $N^{\varepsilon 2}$ His-147, $N^{\delta 1}$ His-13 and both carboxylate oxygens of Asp-16, which correspond to the previously reported zinc ion binding site (Maynes *et al.* 2003, McCarthy *et al.* 2003, McCarthy *et al.* 2004).

The oxygen atoms of the acetate molecules were found coordinating with the zinc ion in both active sites (These occupy similar positions to the water molecules modeled in the Maynes *et al.* (2003) structure). One acetate oxygen (O) is located in an oxyanion hole formed by the Zn^{2+} and the OH of Tyr-142 (Figure 3-18), while the other acetate oxygen (OXT) is found within hydrogen bonding distance to the $O^{\delta 2}$ of Asp-15.

The glycerol molecule was found interacting with $N^{\varepsilon 2}$ His-144, both carboxylate oxygens of Asp-95 and $N^{\eta 1}$ Arg-68 in both active sites. The glycerol molecule in the active site of each chain was modeled slightly differently according to the observed density, but were found to have similar interactions with the active site residues (Figure 3-14). Distances between relevant atoms are reflected in Table 3-4. Figure 3-18 gives a graphical representation of the information presented in Table 3-4.

Table 3-4: Distances between relevant atoms in each chain of the structure solved here

Ligand	Residue	Distance in chain A (Å)	Distance in chain B (Å)
Glycerol			
1-OH	$N^{\eta 1}$ Arg-68	2.9	2.7
1-OH	$O^{\delta 2}$ Asp-95	2.6	2.6
3-OH	$O^{\delta 1}$ Asp-95	2.8	2.8
3-OH	$N^{\varepsilon 2}$ His-144	2.8	3.0
Acetate			
O	O Tyr-142	2.6	2.6
O	Zn^{2+}	2.0	1.9
OXT	$O^{\delta 2}$ Asp-15	2.7	2.5
OXT	Zn^{2+}	2.6	2.7
Zn^{2+}			
Zn^{2+}	$N^{\delta 1}$ His-13	2.0	2.0
Zn^{2+}	$N^{\varepsilon 2}$ His-147	2.1	2.0
Zn^{2+}	$O^{\delta 2}$ Asp-16	2.0	2.1
Zn^{2+}	$O^{\delta 1}$ Asp-16	2.3	2.3
Additional distances suggesting hydrogen bonds important for active site geometry			
$O^{\delta 1}$ Asp-146	O^{γ} Ser-96	2.6	2.6
$O^{\delta 2}$ Asp-146	N Ser-96	2.7	2.8
$O^{\delta 2}$ Asp-146	$N^{\varepsilon 2}$ His-13	3.0	2.7
N Asp-146	$N^{\delta 1}$ His-144	3.1	3.1

Note: - Atoms are numbered according to the International Union of Pure and Applied Chemistry (IUPAC) and International union of Biochemistry and Molecular Biology (IUBMB) joint Commission on Biochemical Nomenclature (CBN) recommended numbering (Moss, 2009).

All the ligands found resolved in the crystal structure are easily explained as the protein had been stored frozen in glycerol, and the acetate and PEG are constituents of the crystallisation medium.

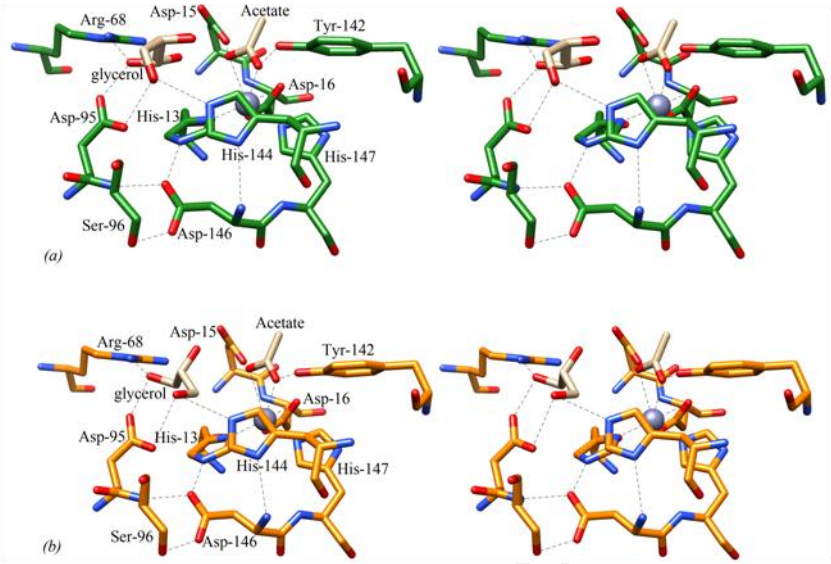


Figure 3-18: Stereo view of glycerol and acetate in the active site.

Lines indicate important distances shown in Table 3-4 (dotted lines are potential hydrogen bonds while purple lines are interactions with the zinc ion). The zinc ion is shown as a grey sphere. Chain A is shown in green and Chain B is shown in gold.

3.3.5 GlcNAc-Ins modelled into the active site of MshB solved here

Glycerol is a subcomponent of the glucopyranoside ring of the natural substrate, GlcNAc-Ins, and acetate is a product of the reaction arising from the hydrolysis of the amide bond. It is possible to superimpose a model of the natural substrate onto the corresponding atoms of glycerol and acetate using Chimera (Figure 3-19). This exercise clearly positions GlcNAc in the active site and enables the identification of specific enzyme substrate interactions.

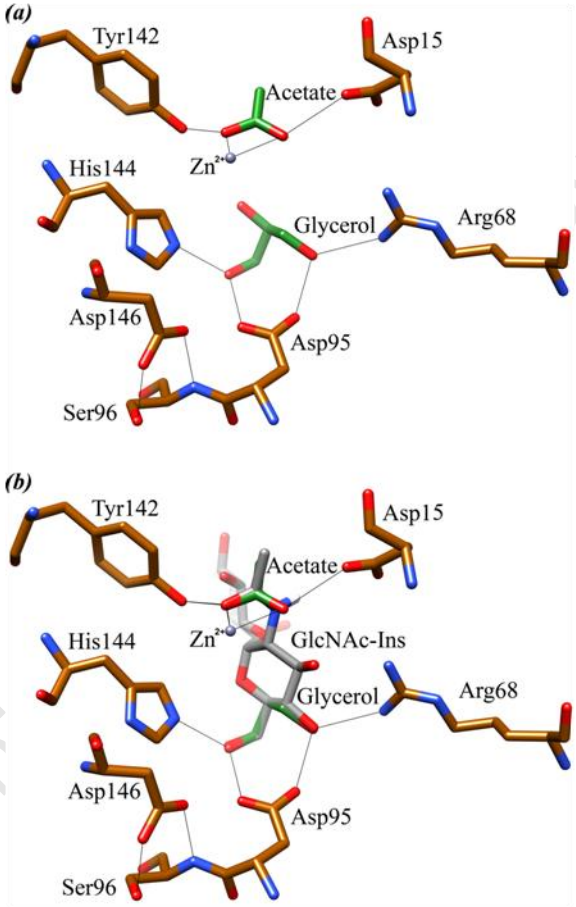


Figure 3-19: GlcNAc-Ins modeled into the active site of MshB solved here

(a) The active site atoms from ligands in chain A are shown in green and active site amino acids are shown in brown, with nitrogen in blue and oxygen in red. (b) GlcNAc-Ins (grey) is superimposed almost perfectly onto the atoms of the ligands present in the active site.

3.4 Modeling of Mca

Mca is of interest because of the overlapping substrate specificity with MshB. Mca and MshB share almost 33 % sequence identity (and 43 % for the first 200 residues) and are predicted to fold in a similar tertiary conformation, as such, a model of Mca could potentially be built on the scaffold of MshB to provide insights into the architecture of the active site of Mca and therefore shed light on possible active site differences between Mca and MshB. This information could prove useful when designing inhibitors of Mca.

The Mca sequence was submitted to Psipred to retrieve the predicted secondary structure and to GenThreader check for the closest predicted structural homologue. Subsequently this information was used in conjunction with the computer program Modeller to model Mca.

The top scoring multiple alignment from GenThreader, using the Mca amino acid sequence as the query, was with a published MshB structure (PDB code 1Q74) (Figure 3-20). Using Modeller, the secondary structure constraints that were predicted using the Psipred server (Buchan *et al.* 2010) (Figure 3-21) were imposed on the Mca amino acid sequence and the MshB structure solved here (PDB code 4EWL) was then used as a template to constrain the tertiary structure.

The best zDOPE score achieved was -0.04 (the best zDOPE score achieved when MshB was modeled to MshB was -1.12). Better zDOPE scores are indicated by lower numbers, a zDOPE score of less than -1 is an indication of a relatively accurate model (Lasker *et al.* 2012). It would be interesting to attempt to model the first 200 residues of Mca only and compare the zDOPE obtained from that study to the zDOPE obtained here (-0.04) to see if the divergence of amino acid identity in the last 100 amino acids is causing the zDOPE to get worse. Unfortunately due to time constraints this was not attempted.

In order to provide an overview of the Mca model produced by Modeller relative to the MshB structure solved here and the secondary structural elements of Mca predicted by Psipred, a multiple structural alignment was constructed using Chimera (by superimposing the MshB structure solved here onto the Mca model made here). Figure 3-22 shows this alignment together with the structural elements predicted by Psipred. The mismatches between predicted secondary structures and those present in the model arise from the difficulty for computer programs to accurately predict protein structures directly from the amino acid sequence and add weight to the need for a crystal structure to be solved before any meaningful inferences be made regarding the specific 3-dimensional spatial arrangement of amino acids in Mca.

The MshB and Mca models were submitted to CastP (Dundas *et al.* 2006) and the active site pockets were compared. MshB returned with a solvent-accessible pocket area of 620.9 Å and the Mca model with 574.3 Å. This is unexpected as Mca is able to accommodate much larger substrates (MSR) than MshB (GlcNAc-Ins) and therefore it is expected to have a larger solvent accessible pocket associated with the active site.

```
- Key ---------------------------------------------------------------
Conf   = Description of confidence level
Score  = Raw score from SVM
p-val  = Probability of false positive
Epair  = Pairwise energy for model
Esolv  = Solvation energy for model
AlnSc  = Sequence alignment score
Alen   = Length of alignment
Dlen   = Length of PDB entry
Tlen   = Length of target sequence
PDB_ID = PDB identifier (+ chain code + domain code in CATH format)
---------------------------------------------------------------------
Confidence levels:
CERT   p-value < 0.0001
```

Conf	Score	p-val	Fpair	Fsolv	AlnSc	Alen	Dlen	Tlen	PDB_ID
CERT	142.831	2e-13	-240.7	-4.1	805.0	277	287	288	1q74A0

```
>>> Alignment with 1q74A0:
                 10        20        30        40        50
         -CCCEEEEEECCCCHHHHHHHHHHHHHHHCCCEEEEECCCCCCCCCCCCCCCCCCCCCC
1q74A0   -ETPRLLFVHAHPDDESLSNGATIAHYTSRGAQVHVVTCTLGEEGEVIGDRWAQLTADHA
             ||  |||||||||    || | |   |  | ||| | || ||      |    |
Query    MSELRLMAVHAHPDDESSKGAATLARYADEGHRVLVVTLTGGERGEILNP--AMDLPDVH
         CCCCEEEEECCCCHHHHCCHHHHHHHHHHCCCCEEEEEECCCCCCCCCCC--CCCCCCHH
                 10        20        30        40        50


                  70        80        90          100       110
         CCHHHHHHHHHHHHHHHCCCCCCEECCCCCCCC---CCCCCCCCCCCCCCCCHHHHHHHHH
1q74A0   DQLGGYRIGELTAALRALGVSAPIYLGGAGRWR---DSGMARSQRRFVDADPRQTVGALV
               |  | | |   |||     ||         |         |         |||
Query    GRIAEIRRDEMTKAAEILGVEH-TWLGFVDSGLPKGDLPPPLPDDCFARVPLEVSTEALV
         HHHHHHHHHHHHHHHHHHCCCE-EEECCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHH
                   70        80         90        100       110


          120       130       140       150         160       170
         HHHHHHCCCEEEEECCCCCCCCHHHHHHHHHHHHHHHHHC-----CCCCCCCCCEEEEE
1q74A0   AIIRELRPHVVVTYDPNGGYGHPDHVHTHTVTTAAVAAAGV-----HPGDPWTVPKFYWT
            || ||||  ||| |||| ||||    | |  ||  |||        | |||| | |
Query    RVVREFRPHVMTTYDENGGYPHPDHIRCHQVSVAAYEAAGDFCRFPDAGEPWTVSKLYYV
         HHHHHHCCCEEEEECCCCCCCCHHHHHHHHHHHHHHHHCCCCCCCCCCCCCCEEEEEEEE
         120       130       140       150       160       170


                180       190       200       210       220
         ECC-HHHHHHHHHCCCCCCCCCCCCC-CCCCCCCCCCCCCCCCEEEECCHHHHHHHHHHH
1q74A0   VLG-LSALISGARALVPDDLRPEWVLP-RADEIAFGYSDDGIDAVVEADEQARAAKVAAL
                         |                          ||           ||
Query    HGFLRERMQMLQDEFARHGQRGPFEQWLAYWDPDHDFLTSRVTTRVECS-KYFSQRDDAL
         CCCHHHHHHHHHHHHHHCCCCCCHHHHHCCCCCCCCCCCCCCEEEEEECC-CHHHHHHHHH
         180       190       200       210       220       230


                 240       250       260       270       280
         HHCCCCCEECCCCCEEECCCCCEEECCCEEEEEEEECCCCCCCCCCCCCCCCCCCCCC
1q74A0   AAHATQVVVGPTGRAAALSNNLALPILADEHYVLAGGSAGARDERGWETDLLAGLGFT
          |||||    |     |            |   ||     |   |  || | ||
Query    RAHATQI--DPNAEFFAAPLAWQERLWPTEEFELARSRIPA---RPPETELFAGIEP-
         HHHHCCC--CCCCCCCCCCHHHHHHCCCCCEEEECCCCCCC---CCCCCCCCCCCCC-
          240         250       260       270          280


Percentage Identity = 32.8.
```

Figure 3-20: Structural alignment reported by GenThreader.

The Mca amino acid sequence was submitted to GenThreader. The alignment with chain A of MshB PDB code 1Q74 returned with the highest alignment score.

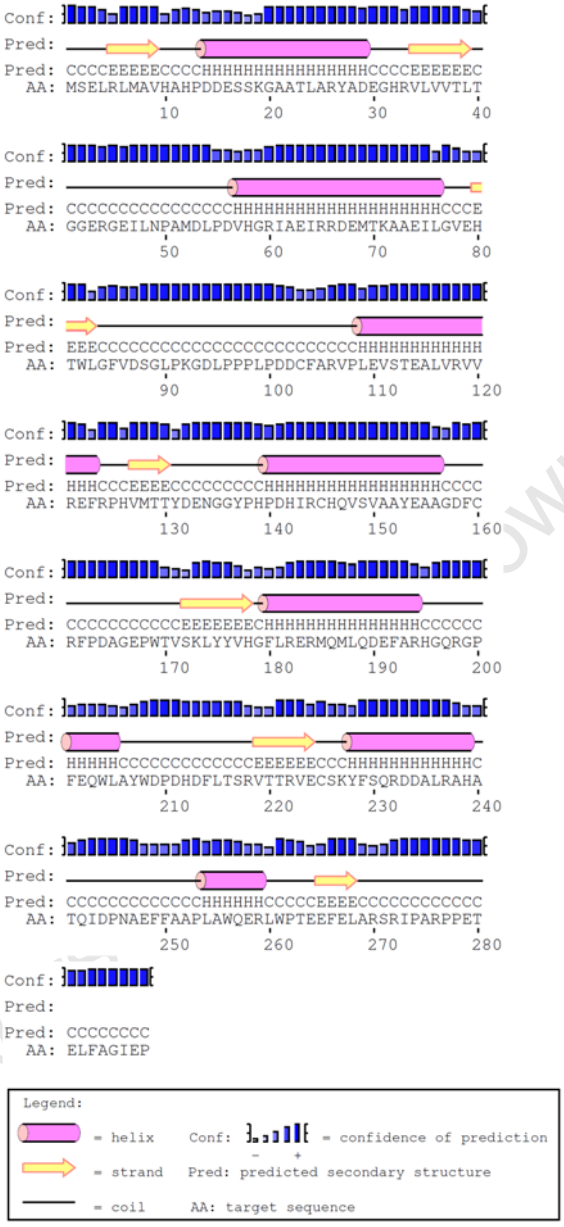


Figure 3-21: Secondary structural elements of Mca predicted by Psipred.

Secondary structural elements were predicted by entering the Mca amino acid sequence into the Psipred server (Buchan *et al.* 2010).

```
chain A MshB 1   MSETPRLLFVHAHPDDESLSNGATIAHYTSRGAQVHVVTCTLGEEGEVIGDRWAQLTADH
Mca PRED     1   MSEL.RLMAVHAHPDDESSKGAATLARYADEGHRVLVVTLTGGERGEILNP..AMDLPDV
Mca ACTL     1   MSEL.RLMAVHAHPDDESSKGAATLARYADEGHRVLVVTLTGGERGEILNP..AMDLPDV

chain A MshB 61  ADQLGGYRIGELTAALRALGVSAPIYLGGAGRWRDSGMAGTD...QRSQRRFVDADPRQT
Mca PRED     58  HGRIAEIRRDEMTKAAEILGVE.HTWLG....FVDSGLPKGDLPPPLPDDCFARVPLEVS
Mca ACTL     58  HGRIAEIRRDEMTKAAEILGVE.HTWLG....FVDSGLPKGDLPPPLPDDCFARVPLEVS

chain A MshB 118 VGALVAIIRELRPHVVVTYDPNGGYGHPDHVHTHTVTTAAVAAAGVGSGTADHPGDPWTV
Mca PRED     113 TEALVRVVREFRPHVMTTYDENGGYPHPDHIRCHQVSVAAYEAAGDFCRFPD.AGEPWTV
Mca ACTL     113 TEALVRVVREFRPHVMTTYDENGGYPHPDHIRCHQVSVAAYEAAGDFCRFPD.AGEPWTV

chain A MshB 178 PKFYWTVLGLSALISGARALVPDDLRPEWVLPRADEIAFGYSDDGIDAVVEADEQARAAK
Mca PRED     172 SKLYY.VHG...FLRERMQMLQDEFARHGQRGPFEQWLAYWDPDHDFLTSRVTTRVECSK
Mca ACTL     172 SKLYY.VHG...FLRERMQMLQDEFARHGQRGPFEQWLAYWDPDHDFLTSRVTTRVECSK

chain A MshB 238 V.....AALAAHATQVVVGPTGR..AAALSNNLALPILADEHYVLAGGSAGARDERGWET
Mca PRED     228 YFSQRDDALRAHATQ..IDPNAEFFAAPLAWQER..LWPTEEFELARSRIPARPP...ET
Mca ACTL     228 YFSQRDDALRAHATQ..IDPNAEFFAAPLAWQER..LWPTEEFELARSRIPARPP...ET

chain A MshB 291 DLLAGLGF
Mca PRED     281 ELFAGIEP
Mca ACTL     281 ELFAGIEP
```

Figure 3-22: Multiple structural alignment of Mca model and MshB structure solved here.

"Mca PRED" indicates the Mca amino acid sequence with structural elements predicted by Psipred. "Mca ACTL" indicates the Mca amino acid sequence with actual structural elements found in the model made here. Amino acids highlighted in yellow indicate α-helices, while those in green indicate β-strands. Red residues underlined in italics indicate putative important active site residues.

The active site of the Mca model presented here (Figure 3-23 and Figure 3-24) shows that there is a Lys-19 in Mca where Ser-20 is in MshB. Where Maynes *et al.* (2003) predicted large differences in the amino acid corresponding to Gln-247 (MshB), it was found that Gln-242 (Mca) was modelled and presents no amino acid difference. At the position of Ser-260 (MshB) though, Ala-255 (Mca) was modeled.

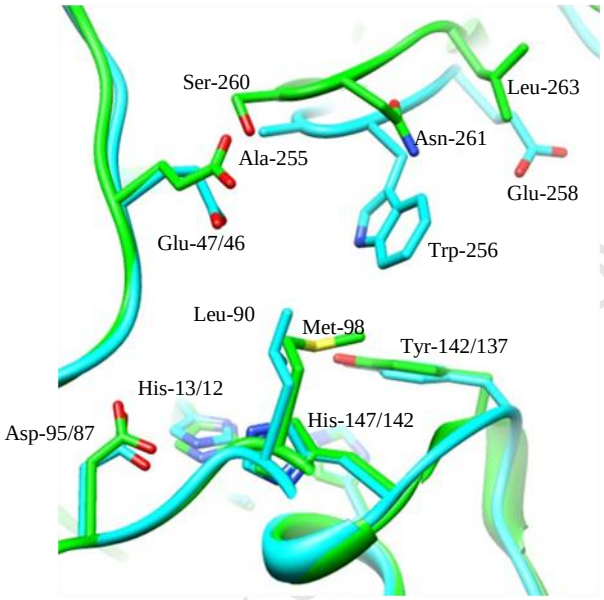


Figure 3-23: Active sites of MshB structure solved here and Mca model made here superimposed.

MshB is shown in green and Mca in blue. The major proposed differences in the immediate vicinity of the active site are shown; namely Ser-260/Ala-255, Asn-261/Trp-256, Leu-263/Glu-258 and Met-98/Leu-90 (for MshB/Mca).

The biggest difference found was in the amino acid that could be in the vicinity of the inositol moiety if MSR binds in the active site the same orientation that GlcNAc-Ins is proposed to bind, the polar Asn-261 (MshB) is replaced by a bulkier non-polar Trp-256 (Mca). In the current orientation of the tryptophan and with the current position of the loop on which the tryptophan is situated, this could cause a steric clash which would be very unfavourable energetically (Figure 3-24) and could cause larger differences in other portions of the Mca

protein upon substrate binding. Other potentially important differences include that Met-98 (MshB) at the mouth of the active site is replaced by Leu-90 (Mca) and deeper into the cavity and away from the active site, the non-polar Leu-263 (MshB) is replaced by an acidic Glu-258 (Mca).

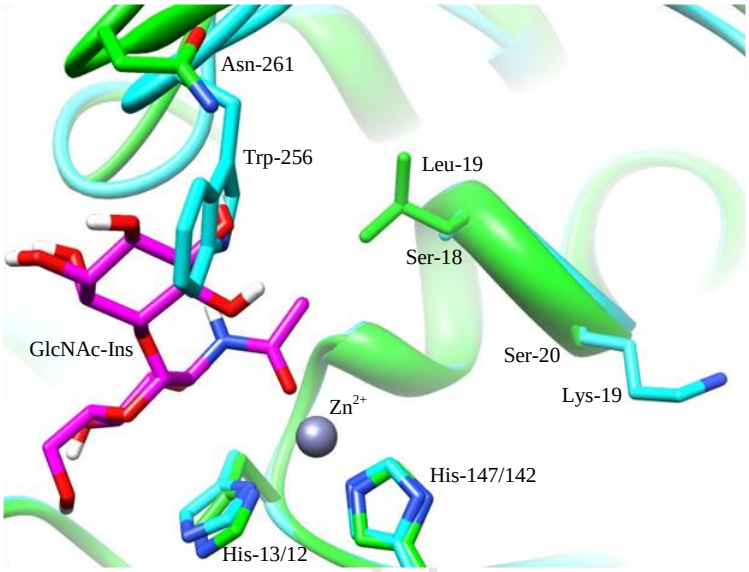


Figure 3-24: GlcNAc-Ins modeled into the active site of MshB (PDB code 1Q7T), superimposed onto the Mca model made here.

This image shows the natural substrate (GlcNAc-Ins) (pink) that was modelled into MshB with PDB code 1Q7T, superimposed into the active sites of MshB solved here (white), and the Mca model (blue) that was solved here. The zinc ion is shown in grey. There is a steric clash between GlcNAc-Ins and Trp-256 in Mca.

To validate the Mca model produced by Modeller, the PDB coordinates were submitted to the online web server, ProSA-web (Wiederstein & Sippl, 2007). The MshB model solved here was also submitted for reference. The Mca model showed better local energy results for the first 200 residues as can be seen by Figure 3-25 (a) but the amino acids of the model in the last 100 residues have energies that are unfavourable and are likely due to the lower percentage of sequence identity in this area of the enzyme compared to MshB, this includes

Trp-256 which is predicted by the model to clash with the proposed binding of the natural substrate and this means that in the real structure (due to the uncertainty of the model) might in fact be in a different position entirely. Figure 3-26 (a) shows how the overall z-score of -5.38 for the Mca structure modelled here is acceptable when compared to z-scores of other structures in the Protein Data Bank. Figure 3-25 (b) and Figure 3-26 (b) show the local energy plot and the position of the z-score (-7.35) respectively for MshB (PDB code 4EWL), both showing acceptable results.

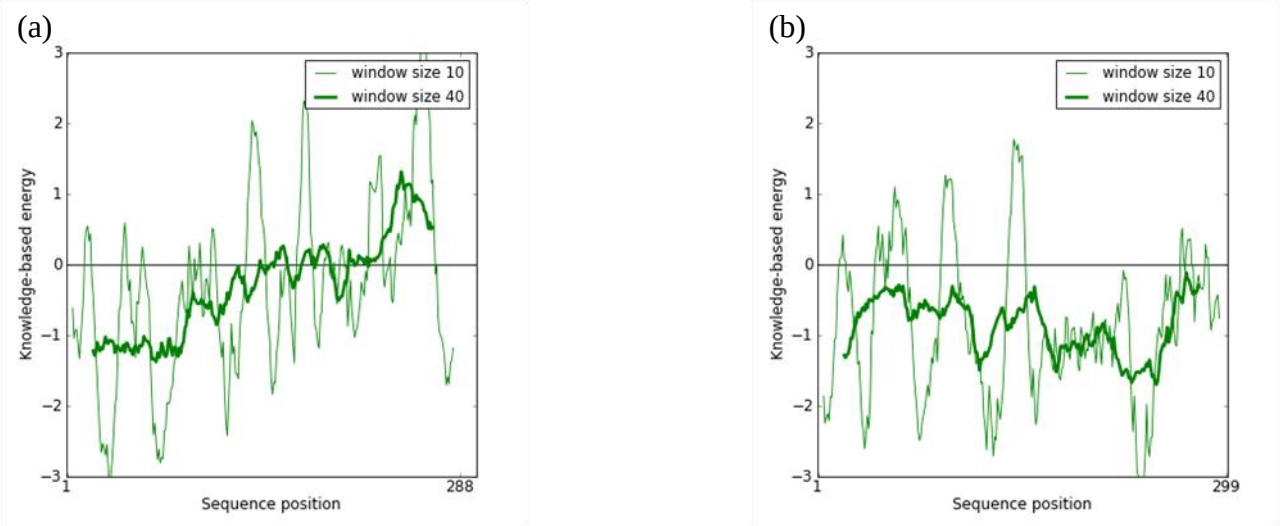


Figure 3-25: Local model quality plots of Mca modeled here and MshB solved here.

The Mca model (a) showed better local energy results for the first 200 residues, but the amino acids of the model in the last 100 residues have energies that are unfavourable and are likely due to the lower percentage of sequence identity in this area of the enzyme compared to MshB. The MshB structure solved here (b) showed acceptable energies across the entire sequence.

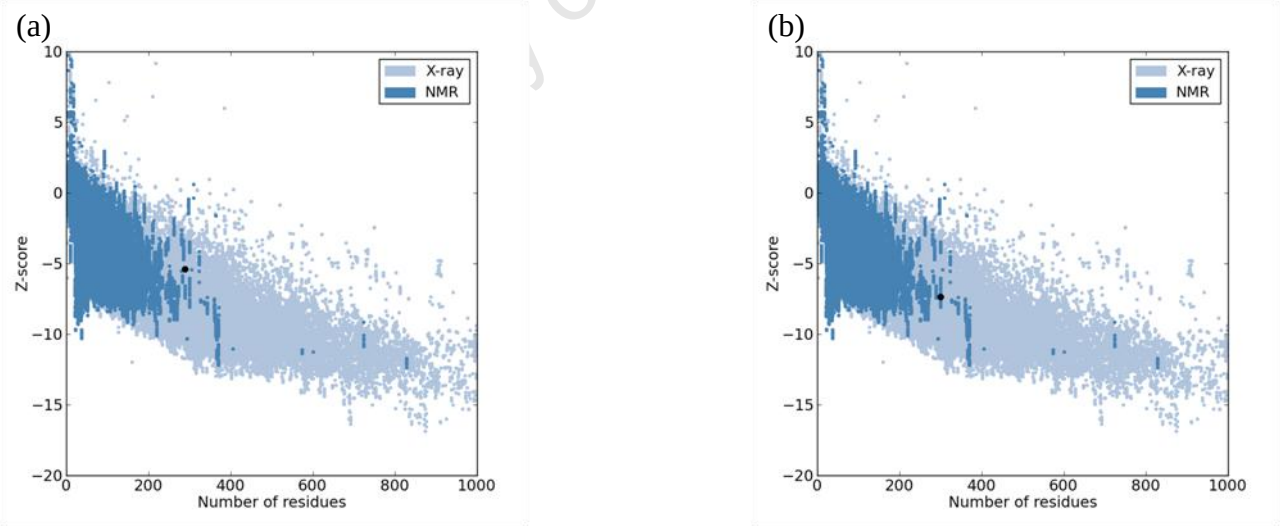


Figure 3-26: Overall model quality plots of Mca modeled here and MshB solved here.

The overall z-score of -5.38 for the Mca structure modelled here (a) and the overall z-score of -7.35 for the MshB structure solved here (b) is acceptable when compared to z-scores of other structures in the Protein Data Bank

3.5 MshB activity

The activity of MshB (MshB containing fractions from the S300 Size Exclusion Chromatography were pooled and concentrated and it was this sample that was used in the activity assays.) with GlcNAc was investigated in order to construct the purification table and provide an indication as to the efficacy of the purification protocol.

3.5.1 MshB assays

MshB activity was assayed using the method of Huang and Hernick (Huang, Hernick 2011), with slight modifications (Chapter 2.5) using protein material purified during this study that was concentrated after size exclusion chromatography.

The assays were performed using substrate concentrations ranging from 5 – 200 mM GlcNAc. The upper limit of GlcNAc solubility is approximately 226 mM in water. The rate of fluorescence was determined using the data that represents a constant rate of increase in fluorescence (straight line portion of the curve) which corresponds to the data in the 0 – 30 min range (Figure 3-27 and Figure 3-28).

The rate of fluorescence seemed to increase with an increase in substrate concentration which is expected. At 0 mM substrate concentration the rate of fluorescence was found to be 0.01 fluorescent units per minute (Fl/min) and at 200 mM GlcNAc the rate had increased to 7.88 Fl/min. The rates obtained in this study were also compared to those published previously by Huang and Hernick (2011), as this would give an indication of how active the protein prepared in this study was (Table 3-5). It is clear that the rates of fluorescence obtained in this study were greatly reduced when compared to those rates previously published.

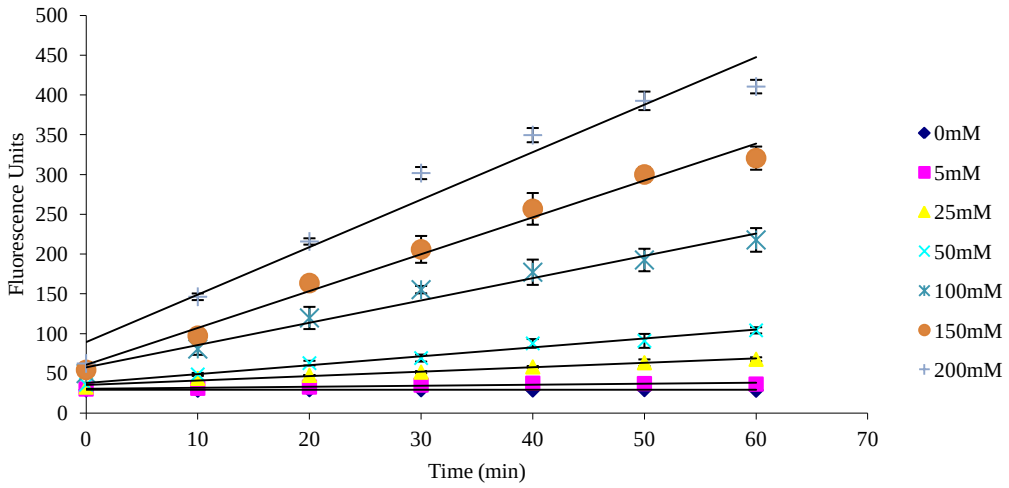


Figure 3-27: MshB assay showing increase of fluorescence as a function of time.

This figure shows a comparison of all the original fluorescent data that was obtained from the range of GlcNAc concentrations assayed with MshB. Each line represents a different concentration of substrate as shown in the legend on the right hand side of the table.

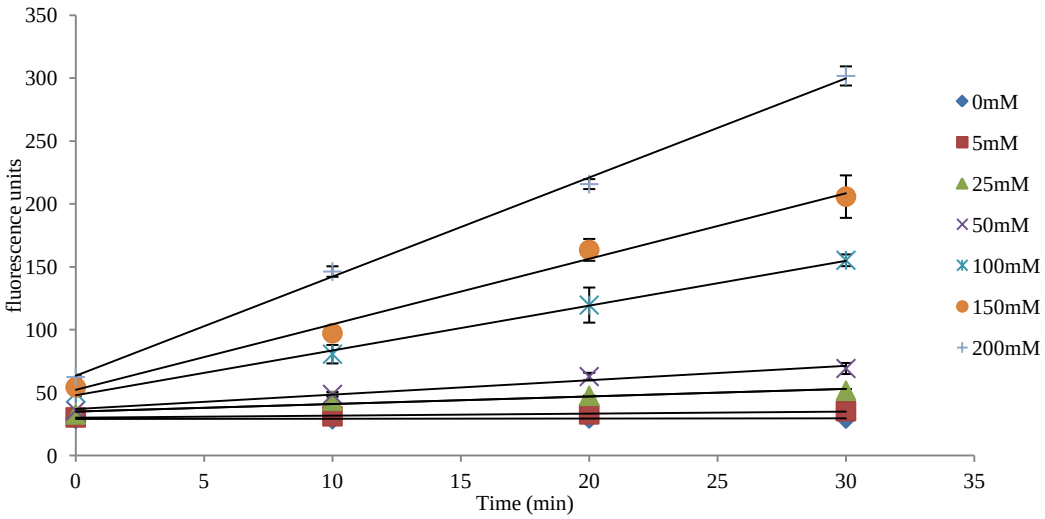


Figure 3-28: Initial velocity of MshB assays showing increase of fluorescence as a function of time.

This figure shows a comparison of the initial (first 30 mins) velocity fluorescent data that was obtained from the range of GlcNAc concentrations assayed with MshB. Each line represents a different concentration of substrate as shown in the legend on the right hand side of the table.

Table 3-5: Equations of each graphical trend line from the initial velocity of the MshB assays and comparison of the resulting rates of fluorescence with previously published data.

	MshB activity 0-200mM GlcNAc Fluorescence Vs Time	Rate of fluorescence increase Fluorescence units per minute	Huang and Hernick (2011) Rate of fluorescence increase Fluorescence units per minute
0mM	y = 0.01x + 29.12 R^2 = 0.33	0.01	0.81
5mM	y = 0.16x + 29.99 R^2 = 0.95	0.16	73.10
25mM	y = 0.60x + 34.92 R^2 = 0.92	0.60	226.40
50mM	y = 1.14x + 37.01 R^2 = 0.98	1.44	298.40
100mM	y = 3.56x + 47.97 R^2 = 0.99	3.56	407.50
150mM	y = 5.21x + 52.15 R^2 = 0.99	5.21	448.00
200mM	y = 7.87x + 63.49 R^2 = 0.99	7.88	No data available

3.5.2 Effect of DTT

Because DTT was substituted for TCEP in the activity assays performed here, a preliminary experiment was conducted to investigate whether DTT could be the cause of the decreased activity witnessed in the activity assays (Table 3-5). Figure 3-29 shows that a 10 fold increase in DTT corresponded to an almost 10 fold decrease in activity.

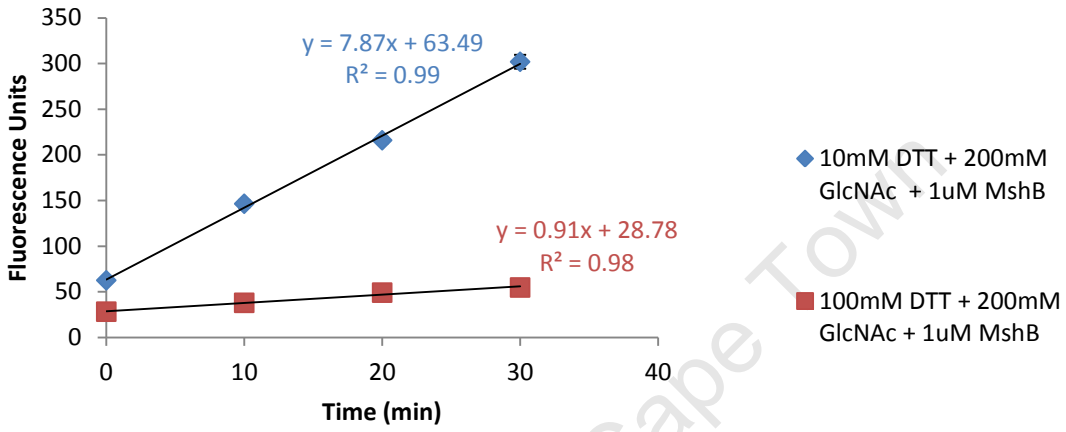


Figure 3-29: Plot showing the effect of DTT on MshB activity, showing the rate of increasing fluorescence. This plot shows that a ten fold increase in DTT concentration decreases the rate of increasing fluorescence by almost ten fold. The equation derived from each plot is coloured the same as the points of the corresponding plot.

3.5.3 Effect of VU5

A preliminary investigation was also made into the effect of the putative inhibitor VU5 on MshB activity with GlcNAc, as previously inhibition effects had only been performed with GlcNAc-Ins (Gammon *et al.* 2010).

VU5 was included in assays with and without GlcNAc, and the resulting rates were compared.

The rate of increasing fluorescence observed when 0.1 mM VU5 was incubated with MshB at 100 mM DTT (note these are extreme inhibitory conditions when MshB is incubated with

GlcNAc) (Figure 3-30 A) was found to be approximately twice more than that observed when 5 mM GlcNAc was incubated with MshB at 10 mM DTT (note these conditions are not as inhibitory) (Table 3-5).

It was found that VU5 was able to increase the rate of fluorescence and potentially recover the activity of MshB when the assay was performed in 100 mM DTT (Figure 3.5-10 A and B). VU5 also increased the rate of increase of fluorescence by almost 40 times that of GlcNAc when incubated at the same molar quantities (Figure 3-30 C).

VU5 seems to be a better substrate than GlcNAc, which is not surprising as it has a cleavable amide bond together with the glucosyl moiety and phenyl ring (in place of the inositol moiety) and therefore the important pharmocophores are represented more in VU5 than in GlcNAc. Despite this, the fact that VU5 seems to be a cleavable substrate needs to be investigated further if any significant proposals are to be made.

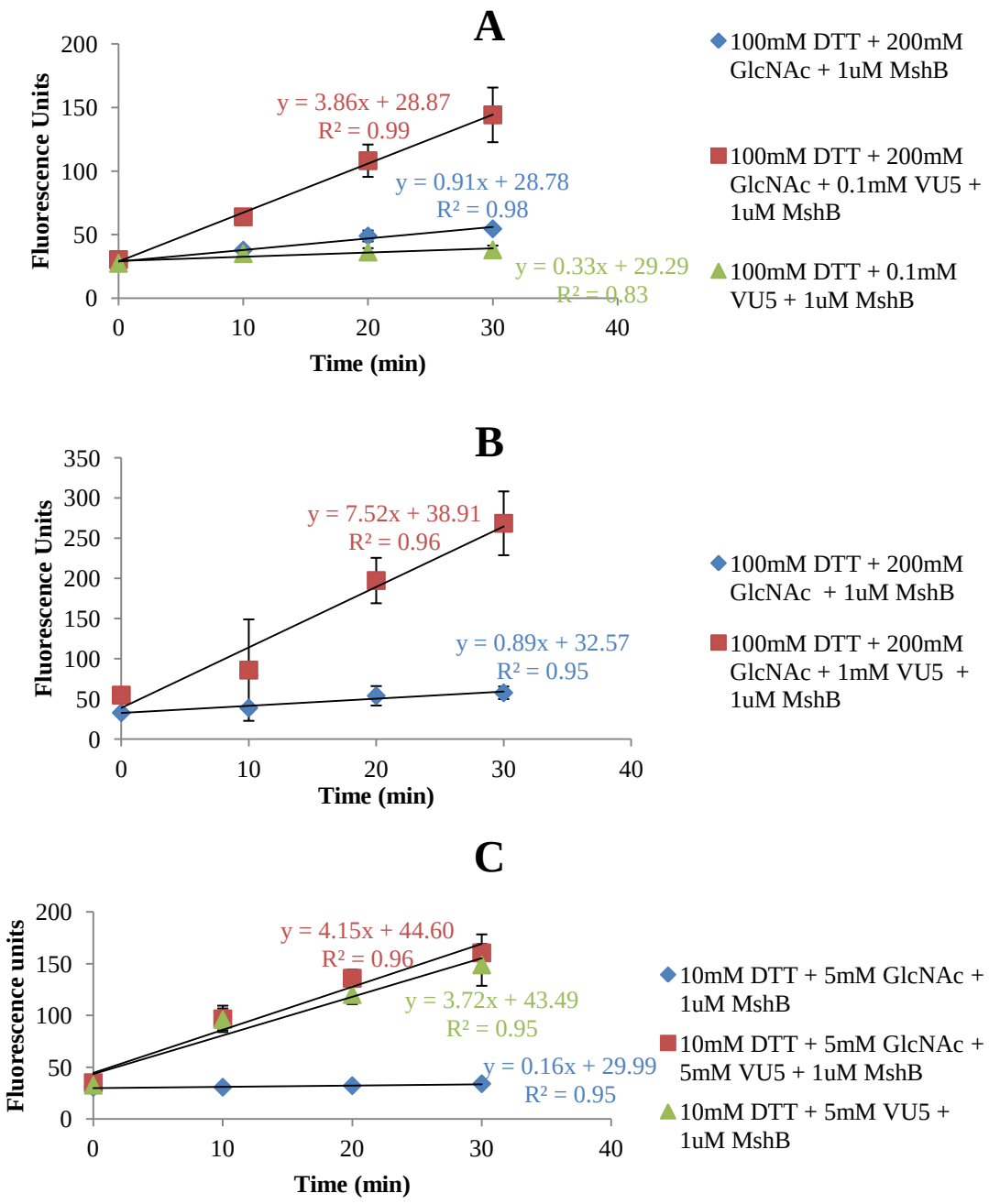


Figure 3-30: Plots showing the effect of VU5 on MshB activity, showing the rate of fluorescence.

The equation for each trend line is shown on the graph in the same colour of the corresponding plot. The slope indicates the initial velocity of the reaction. These graphs compare the effect that different concentrations of VU5 has on the activity of MshB under inhibiting conditions (high DTT) and under conventional conditions. The blue graphs represent conditions without VU5, maroon graphs represent conditions with VU5 and the green graph represents conditions with VU5 and without GlcNAc.

4 Discussion

MshB was successfully purified and crystallised, X-ray diffraction data were collected and the structure solved by molecular replacement to a resolution of 1.95 Å. The coordinates were also successfully deposited into the Protein Data Bank under the code 4EWL. Glycerol and acetate were found in the active site and GlcNAc-Ins was modelled into the active site using glycerol and acetate as a guide. It will be argued below that the mechanism proposals of both Maynes et al (2003) and Huang and Hernick (2012) are inconsistent with the structure determined in this study and a new mechanism which resolves these difficulties but remains consistent with known kinetic information is proposed.

The structure of Mca modelled here could indicate some important amino acid differences between the active sites of MshB and Mca, however it will be shown that the homology modeling done here is not sufficient to explain the differences in substrate specificity.

The activity of MshB under different assay conditions implied that high concentrations of DTT inhibit the enzyme and that VU5 seems to increase the rate of primary amine production.

4.1.1 MshB purification

The assay conditions used in this study were not optimal (due to the potential inhibitory nature of DTT on the reaction) and thus prevent any informative analysis on the kinetic properties of the enzyme from being made in this study. However the resulting purification table can still provide an indication of the efficacy of the purification scheme. The purification table for MshB preparation (Table3-1) shows an overall yield of over 70 % and a purification of over 70 fold. The first three steps, namely the cell free extract, ammonium sulphate precipitation and the first MshB containing fraction (fraction number 19) from anion

exchange chromatography, provided expected trends of yield and purification where the yield decreased after each step, the purification factor increased and the total protein content decreased. The last two steps however show some data that initially comes across as impossible (for example a 104.31 fold purification and a yield of 120.86 %). The Q-HP sample shows what seems to be an unlikely increase in overall activity which corresponds to an irrational yield of over 100 %. This could possibly be attributed to partial separation of inactive and active enzyme.

When the purification procedure performed in this study is compared to that done by Dr Marakalaka (2008) (Table 4-1), it seems that the purification procedure performed in this study managed to obtain over 2-fold increase in purification factor, and almost double the percentage yield after Ion Exchange Chromatography. The increase in apparent activation of MshB in later purification steps also seems to be a trend that is observed by Marakalala (2008) and Newton *et al.* (2000).

The purification procedure in this study was deemed a success because active protein was produced and it was sufficiently pure to grow crystals (with the assistance of homologous streak seeding with crystals grown using Dr Marakalala's stored MshB enzyme).

Purification Step	Protein Concentration (mg/ml)	Total Volume (ml)	Total Protein (mg)	Total Activity (nmols/min)	Specific Activity (nmol/mg/min)	Yield TA/TA1 %	Purification Factor
Crude Extract	9.88	40	395.2	2544	6.44	100	1
DEAE Sepharose	5.55	18	99.9	1321	13.2	51.9	2.05
IMAC	7.27	8.0	58.16	783.2	13.5	30.8	2.1
Sephacryl S-300	1.63	8.8	14.3	2587	181	101	28

Table 4-1: MshB Purification table drawn up by Dr Marakalala (2008)

It seems that the purification procedure performed in this study managed to obtain over 2-fold increase in purification factor, and almost double the percentage yield after Ion Exchange Chromatography when compared to that observed by Dr Marakalala.

4.2 Biological unit of MshB

The estimated MW calculated using the SEC standard curve was found to be 45.8 kDa, which was larger than the published MW. The published MW corresponds with the apparent MW calculated from SDS-PAGE which is approximately 33 kDa (Newton *et al.*, 2006).

The molecules in the asymmetric unit reported in this study share an interface, orientation and interfacial interactions mediated by PEG, with two of the chains reported in the MshB structure with PDB code 1Q74 (Figure 4-1).

Figure 4-1: Chain A (green) and B (gold) of the MshB structure solved here superimposed onto the corresponding chains in MshB PDB code 1Q74 (yellow).

The chains were superimposed using MatchMaker in Chimera. The zinc ion is shown as a grey sphere and PEG, glycerol and acetate are modeled as sticks with their carbon atoms coloured green and gold for chain A and B respectively with PEG from 1Q74 in yellow, while oxygen atoms are in red. This image shows that the interface between these two monomers in their respective crystal structures is almost identical.

Maynes *et al.* (2003) did not report any information regarding the biological unit of MshB, even though their structure (PDB code 1Q74) showed four molecules in the asymmetric unit. The PDB coordinates obtained in this study were submitted to the Protein Interfaces, Surfaces and Assemblies service (PISA) at the European Bioinformatics Institute (http://www.ebi.ac.uk/pdbe/prot_int/pistart.html) and the analysis revealed that the structures represented in the PDB coordinates are unlikely to form a complex in solution (Krissinel, Henrick 2007). As the molecules did not show any kind of clear interfacial interactions, one can infer that either the crystallisation conditions were not conducive to the biological unit forming, or that MshB acts as a monomer. This is consistent with the findings by McCarthy *et al.* (2004) who stated that "The crystal structure shows no evidence of any oligomeric association of the MshB monomers, consistent with size exclusion chromatography, which showed a monomeric species." Even though prior purification and structural methods had shown that MshB is likely to act as a monomer, Newton *et al.* (2006) found that MshB eluted as dimers and tetramers from size exclusion chromatography. At low Hepes concentration, dimers and tetramers were found, and at a higher Hepes concentration, only dimers. Even though the buffer used in the size exclusion chromatography performed by Newton *et al.* (2006) was almost identical to the buffer used in this study, it is believed that in this study MshB eluted from the size exclusion column as a monomer. An in depth study of the buffer and salt conditions leading to the various possible mono, di and tetrameric species of MshB eluted off size exclusion chromatography could provide clarity on possible inter-molecular interactions of MshB monomers. Even though the formation of higher order structures is most likely due to protein aggregation under low ionic conditions, future work could include cross-linking experiments under different buffer conditions together with activity assays to investigate the biological unit more thoroughly.

4.3 X-ray Crystallography

4.3.1 Co-crystallisation of inhibitor

VU5 was not visualised in the crystal structure. One possible reason is that the poor solubility of VU5 would hamper its ability to co-crystallise. There was not enough VU5 to allow a full characterisation of the small molecule during this study, therefore details such as the exact solubility of VU5 in water are currently unknown.

Although the mode of inhibition by VU5 is currently unproven, it is likely that competitive inhibition is the mode of action of this inhibitor as it is a substrate analogue. Modeling has offered preliminary views of how VU5 might bind into the active site. If the thiophenylglycoside moiety binds in a similar position to the predicted binding site of the *myo*-inosityl-glucopyranoside moiety of the natural substrate, and if their amide bonds coordinate with the zinc ion in a similar way, then the tethered plumbagin could be accommodated in a cavity within the enzyme. If this is the case then it would be easy to understand why molecules with a longer tether could inhibit the enzyme better (Gammon *et al.* 2010). The longer tethers allow the plumbagin moiety to be positioned further into a pocket near the active site and therefore be stabilised by optimising hydrophobic interactions (Figure 4-2).

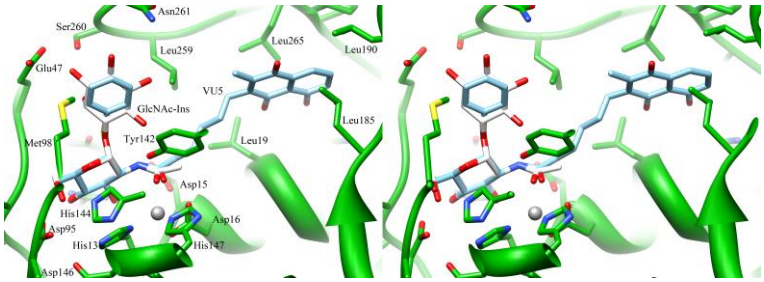


Figure 4-2: Stereo view comparison of VU5 and GlcNAc-Ins modelled into the active site of chain A of the MshB structure solved in this study.

As VU5 is a substrate analogue, it can be modelled into the active site by superimposition over the corresponding atoms of the natural substrate. MshB chain A is shown in green, VU5 in blue, GlcNAc-Ins in white and the Zinc ion as a grey sphere.

In the study by Gammon *et al.* (2010) the pocket which interacts with the plumbagin moiety is stated as being made by Val-184 and Leu-185. In the model presented here however, Val-184 is too far away to interact and instead the pocket is made by Leu-185, Leu-190 and Leu-265. An additional interaction might be made by Lys-237, but that can only be substantiated by solving a crystal structure of MshB with VU5 bound because in the current structure it is too far (more than 4Å) away to make any meaningful interaction. Also if VU5 does bind in this orientation in the active site, the amide bond is well positioned for catalysis by MshB and this could be another reason why VU5 was not visualised in the active site, as the products of the reaction could have been displaced by the relatively high concentrations of acetate and glycerol. The fact that MshB seems active against VU5 adds more evidence that supports this mode of binding.

4.3.2 MshB Structure

The structure of MshB was solved here in space group *C*2. There are two other reported crystal structures in the Protein Data Bank (PDB), MshB with PDB code 1Q74 has the space group P1 (Maynes *et al.* 2003) in spite of very similar molecular packing to that determined here. MshB with PDB code 1Q7T had the space group$P2_12_12_1$ and had a substantially different molecular packing (McCarthy *et al.* 2003).

A high percentage of the data (> 98 % complete) were observed with a redundancy of 3.8. A high completeness together with the observed redundancy means that the data collected represents almost all the data available and that the single axis rotation data collection strategy resulted in each spot being recorded multiple times. This improved the signal to noise ratio, resulting in better quality data for refinement. This in turn can be seen by the ratio of the merged intensity to the error assigned to the merged intensity of 16.66. The R_{merge} obtained in this study (8 %) is good too as it is an indication of the agreement between multiple measurements of the same reflections, lower values indicating a better agreement. The mosaicity observed is 0.21 ° which is an acceptable value (mosaicity angles below 1 ° are considered acceptable). Mosaicity is an indication of the degree of orientational divergence of the mosaic blocks that make up the crystal. It could be improved by exploring the cryo conditions under which the crystal was frozen. Often the mosaicity can be increased when the freezing conditions are not optimal (Jeffrey 2006).

Approximately 5 % of the data were retained for model validation using the R_{free} statistic. The correlation between the resolution and R_{free} was found to be below average for structures deposited before the year 2000. This is an indication that the model is fairly accurate (Kleywegt, Jones 2002). The difference between R-free and R-factor is 0.03; if the difference is less than 0.05 then the possibility of model bias is low enough to be acceptable

(Kleywegt, Jones 2002). The Ramachandran plot (Ramachandran et al. 1963) from MolProbity (Lovell *et al.* 2003) showed over 96 % of all residues falling in the favoured and allowed regions, an indication of correctness in the backbone geometry of all structures (Figure 3-12). This analysis showed that six of the amino acids have phi and psi angles that are not normally represented. The amino acids were modelled with these angles based on the calculated electron density available, and could not be modelled in any other way. In particular the strained conformation of Asp-15 could be explained by its involvement in the active site. Interaction with acetate bound in the active site, in addition to the active site environment, could cause a strained conformation. The model had reasonable bond lengths and bond angles, as judged from the root mean square deviation (r.m.s.d.) values that are within the acceptable range for structures of similar resolution.

The structure solved here, together with those previously solved, allow us to observe a conformational change in the flexible loop located near the active site of MshB (Figure 4-3). The changes are indicative of an induced fit model where the active site geometry is fairly unstable or flexible prior to substrate binding, but assumes a stable conformation after the substrate binds to the active site, this causes significant changes in the areas surrounding the active site which allow the substrate to be positioned for catalysis. Specifically the binding of Asp-95 to the substrate could trigger Asp-146 binding to Ser-96 and a closing of the loop with Met-98 aligning with the inositol moiety of the natural substrate.

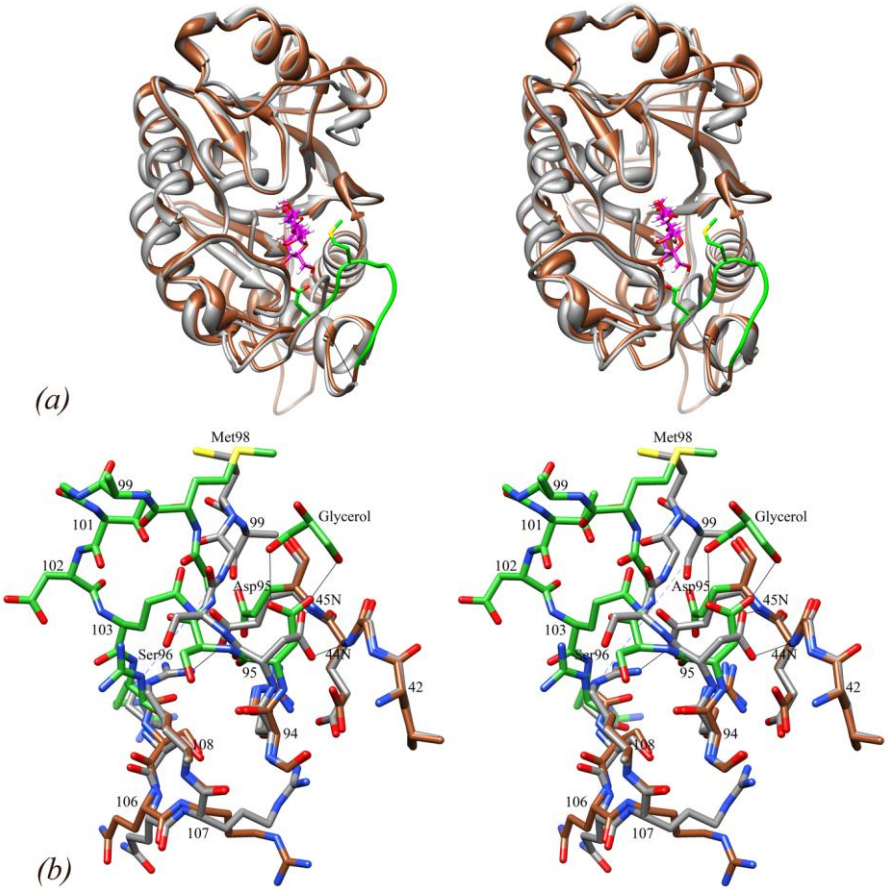


Figure 4-3: Conformational changes in MshB upon binding of substrate.

(a) A ribbon diagram of chain A of MshB from Maynes *et al.*, (2003) (grey) is superimposed on the glycerol bound form solved here (brown and green). The modelled substrate is shown with pink carbon atoms. The comparison highlights the functionally important conformational change in the residues that form the substrate binding pocket. The rigidity of the loop shown in green (residues 95 to 107) is increased as a result of the interaction of the substrate with Asp-95. These changes bring about the binding pocket for the D-*myo*-inositol ring. In particular, according to our model, Met-98 moves so as to interact with the proposed position of the inositol ring of the natural substrate GlcNAc-Ins. (b) In the form of MshB in which no ligands are bound in the active site pocket, Asp-95 forms two hydrogen bonds, one to each of the carboxylate oxygens, from the main chain NH groups of Glu-44 and Glu-45. In the glycerol bound structure, two new hydrogen bonds to the 1-OH and 3-OH of the glycerol respectively, are formed. This causes a number of changes that result in the movement of parts of the molecule which are coloured green. 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. (Figure and caption adapted from Broadley *et al.* (2012)).

The movement of Tyr-142, when the structure reported here is compared to all known chains of MshB in the PDB, shows another important structural change in the active site, and provides structural evidence for the proposal by Huang and Hernick (2012) that Tyr-142 plays an important role in stabilizing the tetrahedral intermediate and could play an important part in product removal. Tyr-142 has rotated about its Cα-Cβ bond such that in our structure it is found interacting with the acetate, and in the other structures it is found pointing out and away from the active site (Figure 4-4).

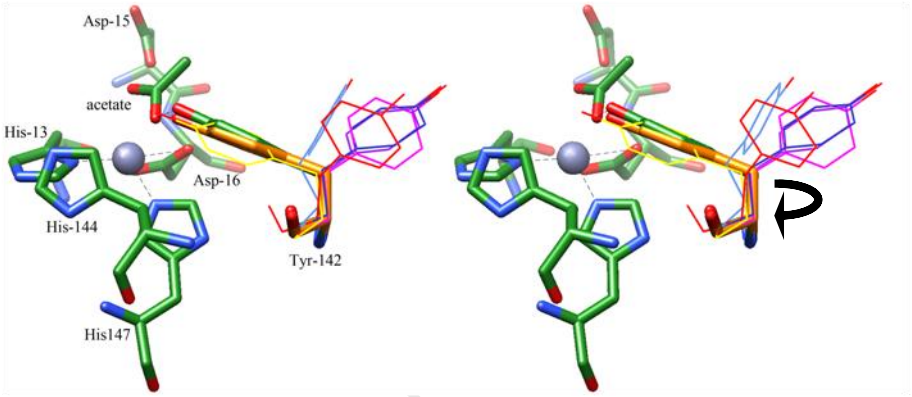


Figure 4-4: Stereo view showing the movement observed for Tyr-142 in all known crystal structures of MshB, and the formation of the oxyanion hole.

The rotation about the Cα-Cβ bond of Tyr-142 is visible as the position it occupies is different in every instance of MshB structure solved thus far. It is important to note that in the presence of acetate the observed position is almost identical in both chains demonstrating the role this residue plays in the oxyanion hole. This position also resembles the hypothetical position suggested by Huang and Hernick (2012). Chain A of the structure solved here is shown in green, Chain B in gold. The other structures currently deposited in the PDB are shown in thin wires, yellow representing Tyr 142 from chain B in the structure with PDB code 1Q7T, red representing A from 1Q74, light purple representing C from 1Q74, dark blue representing B from 1Q74, dark purple representing D from 1Q74 and light blue representing A from 1Q7T.

McCarthy *et al.* (2004) modeled GlcNAc into the active site of MshB based on the β-octyl glucoside (BOG) molecule that was found in the A chain of their crystal structure. The atoms

of the BOG that were found binding to Asp-95 (MshB with PDB code 1Q7T, chain A) correspond very closely to the atoms of the bound glycerol in this structure, and the BOG found in chain B binds in a similar position to acetate (Figure 4-5).

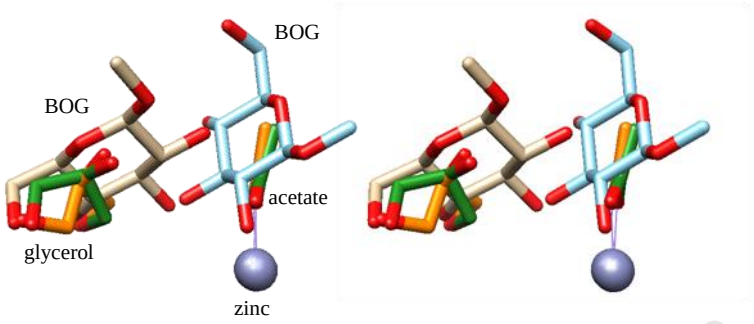


Figure 4-5: Stereo view of ligands found in the active sites of MshB structures with PDB codes 1Q7T and 4EWL, superimposed upon one another.

The β-octyl glucosides crystallised in the MshB structure with PDB code 1Q7T (chain A in cream and chain B in light blue), were superimposed onto the glycerol and acetate molecules found in the active site of the solved structure (chain A in green and chain B in gold).

The glycerol molecule was found to bind to each chain in a slightly different orientation, but they were both found to be interacting with Asp-95 and His-144. The two distal oxygen's of the glycerol molecule seem to bind in a similar position to the β-octyl-glucoside (BOG) from the crystal solution of MshB by McCarthy *et al.* (2004). This indicates the importance of the amino acids that they interact with, namely Asp-95 and His-144, in substrate recognition.

There was an acetate molecule that was found bound to the zinc ion in each active site. The acetate together with the glycerol could indicate a possible lead towards new drug development. The acetate molecule also resembled the geminal diol that was formed when the ester containing kAW inhibitor was co-crystallised with another zinc ion dependent peptidase, angiotensin converting enzyme (ACE) (PDB code 3BKL) (Watermeyer *et al.*

2008) that has been extensively studied at the University of Cape Town. Despite the finding by Gammon *et al.* (2010), that a non cleavable substrate analogue was not a very successful inhibitor of either Mca or MshB, this avenue has perhaps not been exhaustively investigated and my still prove fruitful. Gammon *et al.* (2010) only produced a compound (with an ester in place of the amide) with a cyclohexyl moiety in place of the inositol moiety. This part of the substrate has been shown to be very important in substrate recognition by MshB and by simply replacing the inositol moiety with a cyclohexyl moiety (keeping everything else in the natural substrate the same) the activity is reduced by more than 17 times (Gammon *et al.* 2010). For example, if a substrate analogue inhibitor with an ester was thoroughly investigated rather than an amide, one might get inhibition in the same manner that kAW inhibits ACE.

4.3.3 Catalytic mechanism

Previously researchers have hypothesized that during catalysis, the tetrahedral transition state of the amide carbonyl would be stabilised by the zinc ion and the side chain of His-144 (Newton, Buchmeier & Fahey 2008, Jothivasan, Hamilton 2008, Maynes *et al.* 2003, Fan *et al.* 2009). It is felt that the orientation of histidine 13 and 144 in the active site of MshB Chain A deposited under PDB code 1Q7T were modelled incorrectly. This discrepancy may have given rise to confusion in later work (for example in the study by Huang and Hernick, 2012).

In the MshB structure that was solved here, His-144 was shown to be interacting with the bound glycerol molecule and Tyr-142 was found interacting with the acetate coordinated to the zinc ion. It is proposed that, as opposed to previous authors (Maynes *et al.* 2003, Jothivasan and Hamilton 2008, Newton, Buchmeier & Fahey 2008 and Fan *et al.* 2009) who believe that His-144 acts as the general acid catalyst, His-144 instead acts to stabilise substrate binding, while the transition state and resulting acetate is stabilised by the oxyanion hole formed by the zinc ion and Tyr-142, and that Asp-15 acts as both the general base and general acid catalyst (Figure 4-6).

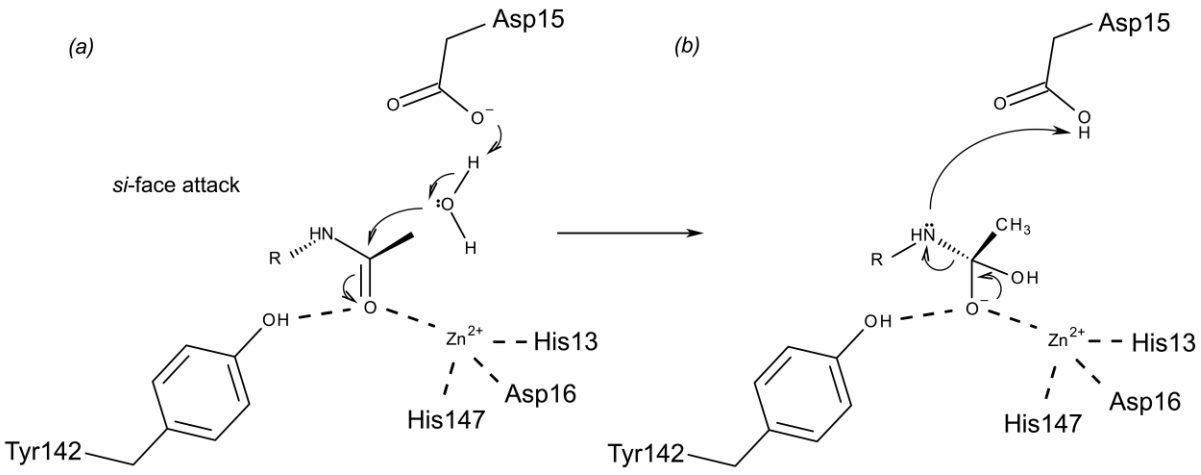


Figure 4-6: Schematic of the catalytic mechanism for MshB that is proposed here.

Proposed catalytic mechanism of MshB showing how Tyr-142 is implicated in stabilising the tetrahedral intermediate state instead of His-144 and that Asp-15 acts first as the general base catalyst and then as a general acid catalyst.

Figure 4-7 shows the catalytic mechanism using deposited structures and models of ligands. McCarthy *et al.* (2004) support the hypothesis that His-144 binds to the substrate and is not involved in stabilizing the tetrahedral transition state (McCarthy *et al.* 2004). The biochemical study published by Huang and Hernick (2012) explored the effect of different active site mutations on the activity of the enzyme (Figure 4-8). It was shown that Tyr-142 is involved in the catalytic mechanism and believed to occupy multiple conformations during the catalytic cycle. The structure reported here provides structural confirmation that Tyr-142 indeed plays an important role in the catalysis of the substrate. Huang and Hernick (2012) show that mutation of Asp-146 to alanine also decreases activity, and increases K_M by a factor of more than 10. They state that this is due to the interaction of Asp-146 with His-13 and/or His-144. Our structure indicates that an interaction with His-13 is possible but that an interaction with His-144 is unlikely as the closest approach of an Asp-146 oxygen to His-144 is 3.3 Å and this is to a carbon atom. Instead we propose that Asp-146 instead interacts with

Ser-96 and His-13 to stabilise the region surrounding the active site when the substrate binds. The biochemical data (Huang and Hernick, 2012) supports our model in that the pK_a values of the functional groups in enzymes in which either His-144 or Asp-146 were mutated were the same as those of the wild-type protein. This suggests that neither of these residues are involved in ionizations that affect the catalytic events. However when Tyr-142 and Asp-15 were mutated to alanine, the upper and lower pK_a values were lost which indicate that these residues play an integral role in the catalytic mechanism and are responsible for the observed pKa values.

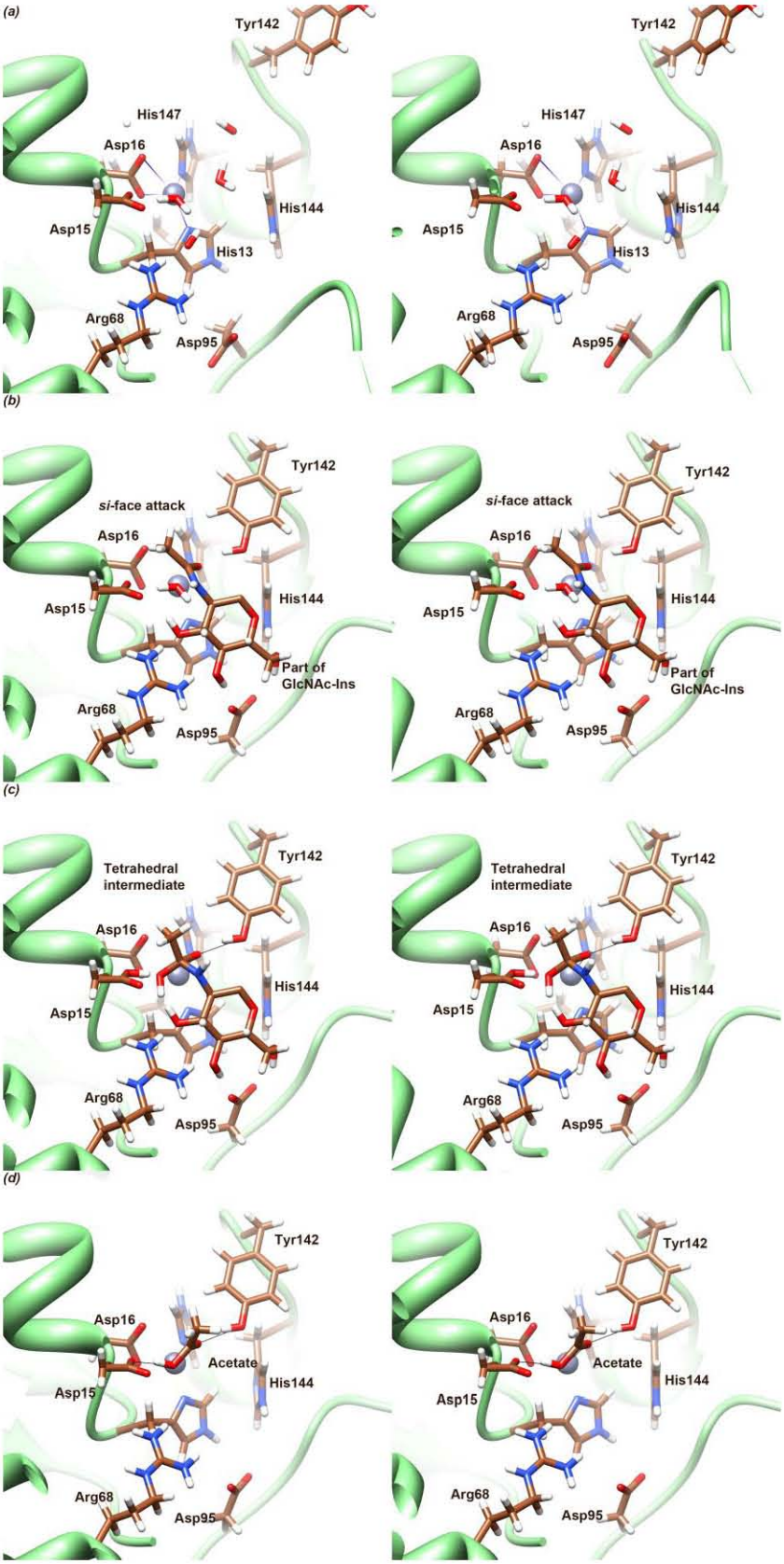


Figure 4-7: Models supporting the catalytic mechanism proposed here.

See caption on next page.

Figure 4-7 caption:

(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 Tyr-142 and Asp-95 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 Tyr-142 and Asp-95 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 Asp-15. (c) The modelled tetrahedral intermediate stabilized by the oxyanion hole formed by Zn^{2+} and by Tyr-142. The carboxylate of Asp-15 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 Asp-15 $O^{\delta 2}$ and 3-OH of GlcNAc. (d) The product acetate with one O atom hydrogen-bonded to Asp-15 and the other located in the oxyanion hole based on the structure described here.

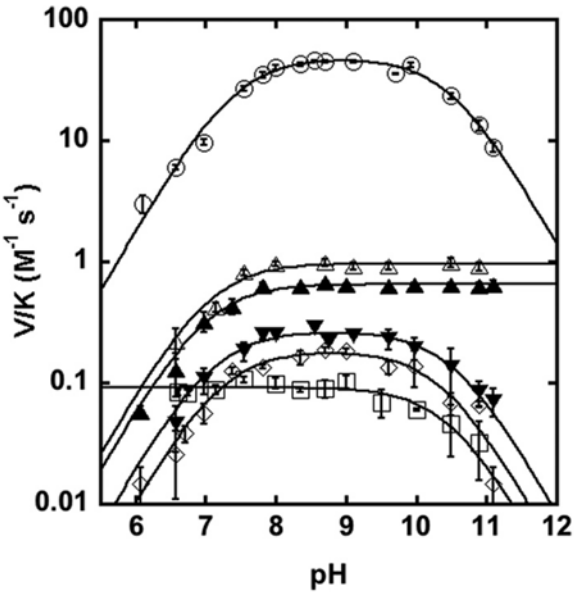


Figure 4-8: pH rate profiles figure take from Huang and Hernick (2012).

The pH dependence of the MshB-catalysed reaction is shown for WT (O), Asp-15 to Ala (□), Tyr-142 to Ala (△), Tyr-142 to Phe (▲), His-144 to Ala (▼), and Asp-146 to Ala (◇) as described in Huang and Hernick (2012).

Huang and Hernick (2012) dismiss the idea of a single amino acid performing as a general acid – base catalyst due to results obtained in two experiments.

Firstly because the decreases in activity for the Asp-15 (530 fold) and His-144 (180 fold) mutations to alanine were less than that expected for a single amino acid functioning as a bifunctional general acid – base catalyst. Huang and Hernick (2012) state that for a bifunctional general acid – base catalyst, one would expect a decrease in activity of $10^3 – 10^5$ fold when that amino acid is mutated. However, the references that they cite to support this expectation involve mutations of glutamic acid and histidine, and not an aspartic acid (Jackman, Raetz and Fierke 1999, Cha and Auld 1997, Wetterholm *et al* 1992), therefore their expectations might not necessarily be appropriate in this circumstance.

Jackman *et al* 1999 investigate an enzyme (UDP-3-*O*-(*R*-3-hydroxymyristoyl)-*N*-acetylglucosamine deacetylase or LpxC) that in a previous paper had been isolated with a mutation from His-19 to a tyrosine (Hyland, Eveland and Anderson 1997). Hyland, Eveland and Anderson (1997) speculated that this mutation disrupted a metal ligand, and that this was the reason for the decrease in activity.

Cha and Auld (1997) describe a mutation in an enzyme (matrilysin) where Glu-198 was mutated to aspartic acid, cysteine, glutamine and alanine. They found a 4.1, 160, 590 and 1900 fold decrease in activity respectively. According to the structure solved by Browner, Smith and Castelhano (1995), Glu-198 is not located in the active site and it serves to help coordinate one of two structural calcium ions visible in the enzyme crystal structure. Disruption of this amino acid must lead to structural effects in the active site and cause a loss of activity.

Wetterholm *et al* (1992) investigated mutations of Glu-296 from human leukotriene A_4 hydrolase. A mutation from glutamic acid to glutamine actually increased the enzyme's

epoxide hydrolase activity by 1.5 fold but decreased the enzyme's peptidase activity by 466 fold. A mutation to alanine decreased the epoxide hydrolase activity by 8 fold and decreased the peptidase activity by 700 fold. The active site glutamic acid in this family of metalloproteases is proposed to act as the general base catalyst in the hydrolysis of peptides (Helgstrand *et al* 2011).

The second experiment that Huang and Hernick (2012) used to infer whether the catalytic mechanism used a single amino acid as the general acid – base catalyst or two amino acids, one being the general acid and the other as the general base catalyst, was to see if they observed any solvent isotope effects when activity of MshB mutants was measured in 95 % D_2O and compared to measurements taken in H_2O. Solvent isotope effects of 0.8 and 5.4 were observed when Asp-15 and His-144 were mutated respectively. Huang and Hernick (2012) believe that the observed effects are consistent with a mechanism that uses a general acid – base pair. It has been shown here that His-144 is involved in substrate binding and is very important in positioning the substrate, so there are many subtle changes that could be taking place with regards to the way that the substrate binds when His-144 is removed. To propose a mechanism based on the solvent isotope effect that is observed could be premature, as there are other possible reasons for observing a solvent isotope effect. For example, the substrate could be interacting with the active site slightly differently and the D_2O could be causing multiple isotope effects that we are not aware of, including secondary effects.

Huang and Hernick (2012) were not in a position to consider key information that is detailed here for the first time, in particular, the proximity of $N^{\varepsilon 2}$ His-144 to 3-OH of the bound glycerol is clearly indicative of hydrogen bonding. This being the case, the atom directed towards the scissile bond is a carbon which is located ~5.5A away making it impossible for His-144 to act as an acid catalyst. In addition previous structures have misinterpreted the role of Asp-146 which is uniquely well defined in the structure described here. This is not

hydrogen bonded to His-144 as depicted by Maynes *et al.* (2003) but to Ser-96, stabilising the highly mobile loop that brings Met-98 to align with the proposed binding position of the inositol moiety of the natural substrate, and to His-13, stabilising the coordination of His-13 to the zinc ion. There is potentially a hydrogen bond from the backbone N of Asp-146 to $N^{\delta 1}$ His-144 (depending on the protonation state of the Histidine) which would aid in the stabilisation of His-144 and thus the ability to aid alignment of the natural substrate when bound in the active site.

4.4 Modeling of Mca

Secondary structure constraints that were predicted using the Mca amino acid sequence submitted to Psipred (McGuffin, Bryson & Jones 2000) were entered into the Modeller script and the MshB structure solved here was then used as a template to constrain the tertiary structure. The best zDOPE score achieved was -0.04 (the best zDOPE score achieved when MshB was modeled to MshB was -1.12).

Maynes *et al.* (2003) modelled Mca and reported that the only major difference between the two models (Mca vs MshB) is that there is a Lys-19 in Mca where there is a Ser-20 in MshB. They suggest that this is where the glucosyl moiety could bind and this difference could cause a steric clash in Mca with the glucosyl moiety on the MSH. The substrate of Mca (MSH conjugated to a toxin, MSR) also contains the glucosyl moiety and therefore either binds in a different orientation or there are differences that might compensate for the unfavourable binding, by residues that bind to the AcCys residue of MSH, Gln-247 and Ser-260, but they were unable to model this region of Mca.

The Mca model generated in this study also shows that Lys-19 in Mca is located where Ser-20 is in MshB, but taking the glycerol binding information together with the BOG binding information (McCarthy *et al.* 2004) it strongly suggests that the glucosyl moeity actually binds to Asp-95. With regards to the question of substrate specificity, in the MshB structure there is the possibility of a steric clash between Asp-16 and the Acetyl group on MSH and could explain why MSH is a poor substrate for MshB, in the Mca model there is slightly more space in this region as the Asp-16 is replaced by Ser-18.

Where Maynes *et al.* (2003) predicted large differences in the position of Gln-247 (MshB) it was found in our model that Gln-242 (Mca) occupied the same position and thus no difference was observed. At the position of Ser-260 (MshB), Ala-255 (Mca) was found. The

biggest differences that were found, were in the amino acids that are believed to bind to the inositol moiety of MSH/MSR. The polar Asn-261 (MshB) is replaced by a bulkier non-polar Trp-256 (Mca). Another big difference at the mouth of the active site is that the Met-98 (MshB) is replaced by Leu-90 (Mca). As one moves deeper into the cavity and away from the active site, the non-polar Leu-263 (MshB) is replaced by acidic Glu-258 (Mca). This change could influence the recognition of R-groups that are conjugated to MSH.

The predicted structural elements do not match up perfectly with the elements present in the model, this is an indication as to the unreliable nature of predictions made using a query with a single sequence, however the differences are discussed below and still provide a platform from which to work from with regards to designing mutagenesis experiments.

A β-strand predicted (but not present in our model) from residue 80 to 83 is close (almost 10 amino acids away) to the active site and would form part of the large β-sheet that runs through the center of the protein. Forming part of the β-sheet could influence the geometry of the active site slightly. Another potentially important prediction that is not accounted for in the model is a helix predicted from residue 254 to 259. This predicted helix holds Trp-256 and forms part of the "roof" of the active site which is implicated in recognition of the inositol moiety of the natural substrate MSR. These alterations might have a consequence that would account for the differences in substrate specificity. Further work could include mutating these residues in Mca such that a β-sheet is not predicted from residue 80 to 83 and that the Trp-256 is changed to Asn. These mutations could alter the activity and substrate specificity of Mca and could provide more insight into the substrate specificity of these two enzymes.

Pockets observed in the Mca model solved here were analyzed using CastP (Dundas *et al.* 2006), viewed in Chimera (Pettersen *et al.* 2004) and compared to those of the MshB

structure solved here. The active site cavity in the modeled Mca structure was smaller than in the MshB structure. This does not account for Mca's ability to hold larger structures in its active site, if the substrate binds into the active site in a similar fashion to MshB, as proposed.

The substrates of Mca (MSH conjugated to an electrophile) are inherently bigger than the MshB substrate GlcNAc-Ins. This information together with the replacement of Asn-261 (MshB) with the bulkier Trp-256 (Mca) suggest that there must be greater folding differences between MshB and Mca that account for the differences in substrate specificity between these two homologous enzymes, and that these differences cannot be fully investigated by modeling alone.

The active site seems to be able to be quite variable as it is surrounded by large loops that could accommodate various sizes of substrate (MSR). It would be difficult to interpret the amount of movement that these loops can accommodate, without having transition state analogues bound in the crystal structures.

Further work is needed to conclusively investigate the substrate specificity of these two enzymes. It would be very interesting to perform mutagenesis experiments in an attempt to convert each enzyme's activity to act on the other enzyme's substrates.

4.5 MshB Activity

4.5.1 MshB assays

MshB activity was assayed with VU5 and GlcNAc (*N*-acetyl Glucosamine) using an assay adapted from Huang and Hernick (2011). GlcNAc is used as it is far cheaper and easier to obtain than the natural substrate, GlcNAc-Ins, and it has been used before in published articles to measure the enzyme kinetics of MshB (Huang and Hernick 2011) and measure the pH dependence of MshB and mutants thereof (Huang and Hernick 2012).

It must be noted that there are differences between the assay published by Hernick and Huang (2011) and that published by Gammon *et al.* (2010). Gammon *et al.* (2010) used half the concentration of Hepes and NaCl, and did not include any reducing agent, whereas Huang and Hernick (2011) used 50 mM Hepes and 50 mM NaCl, and included 1 mM TCEP to maintain a reducing environment.

The differences between the Huang and Hernick (2011) published method and the method that was used in this study will be discussed below. The samples in this study were not centrifuged after precipitating the enzyme with TCA. This was done to increase the speed at which the assay can be completed and to decrease the cost involved in assaying hundreds of reactions. As long as the concentration of enzyme was kept constant, this would not affect the measurement of the rate of the reaction. It would only affect the baseline value as the protein would obviously have primary amines that would react with the fluorescamine reagent. The reaction volume in this study was also increased from 200 µl to 1.5 ml. This was done to reduce the chance of pipetting error when pipetting very small volumes.

The assay as implemented in this study has DTT substituted for TCEP. DTT was freely available in the lab and is cheaper to purchase than TCEP. As TCEP is a much stronger

reducing agent than DTT (Han, Han 1994), together with guidance from colleagues in the lab, 10 mM DTT was added to compensate. This change in the assay was not useful for detailed kinetic analysis because, in hindsight, DTT is likely to be an inhibitor. It did however enable the relative activity of the MshB preparations to be assessed. The assay was used in a preliminary manner to ascertain whether or not active protein was being produced and to give an indication as to the effect of VU5 on the activity of MshB for future studies.

The differences between the derivitisation method published by Huang and Hernick (2011) and the method used by Gammon *et al.* (2010) (adapted from Anderberg, Newton & Fahey 1998) will be discussed below.

The Anderberg, Newton and Fahey (1998) method utilises derivitisation of primary amines by 6-aminoquinolyl-*N*-hydroxysuccinamidal carbamate (AccQ-Fluor) whereas the Huang and Hernick (2011) method utilises derivitisation of primary amines by fluorescamine (FSA). With AccQ-Fluor derivitisation, the AccQ-Fluor must be incubated with the protein sample for 1 min at room temperature, followed by 10 min at 55 °C, after which the sample has to be eluted off a reverse phase column using High Pressure Liquid Chromatography (HPLC) which takes approximately 30 min (not taking into account the time needed to prepare the column for use and the time needed to prepare the column for storage again). Additionally only one sample can be eluted at a time. With FSA derivitisation however, the FSA must be incubated for approximately 10 min in a 96 well plate before the fluorescence can be determined by a plate reader with fluorescent capabilities. The time needed to assay 96 samples is approximately 48 hours when using AccQ-Fluor, and only 10 min when using FSA. The benefits for using a system involving FSA and 96 well plates are increased further when taking into account the amount of reagent used in each method.

4.5.2 Effect of Glycerol on MshB activity

Huang and Hernick (2012) demonstrate a loss of activity in the presence of high concentrations of glycerol and sucrose which they attribute to an increase in the viscosity of the solution. The structure solved here shows glycerol located in the active sites of both chains A and B. These interactions are similar to the hydrogen bonding interactions with the glucopyranoside ring of BOG (McCarthy *et al*., 2004) and suggest that similar interactions position the natural substrate in the active site of MshB. Structural evidence therefore suggests that the observed loss in activity could be attributed to competitive inhibition in addition to possible viscosity effects (Brouwer & Kirsch, 1982). This could also be another reason why VU5 was not seen in the crystal structure as the enzyme had been stored with 20 % glycerol which is approximately 2.2 M glycerol. This high concentration of glycerol would have been diluted down to approximately 0.23 M when the protein was diluted to the appropriate concentration for crystallisation, but could have still out-competed the even lower (approximately 500 µM) concentration of VU5 present in the crystallisation conditions.

4.5.3 Effect of DTT and VU5 on MshB activity

The rates of fluorescence obtained during activity assay measurements made in this study were compared to those obtained in the study by Hernick and Huang (2011) in order to crudely investigate the activity of the MshB produced in this study. Initially it seemed that the MshB that was produced in this study was less active than previously published by Huang and Hernick (2011). However, after noticing that an increase in DTT concentration further inhibits the activity of MshB with GlcNAc, we propose that the difference in activity was due in fact to DTT acting as an inhibitor. Perhaps by performing activity assays over a range of DTT concentrations alongside experiments with a similar range of TCEP concentrations, the apparent inhibition by DTT could be further investigated. Unfortunately due to time and cost constraints the necessary experiments could not be performed with respect to investigating the effects of DTT on MshB activity.

Under conditions containing 100 mM DTT, the amount of GlcN produced by the hydrolysis of 200 mM GlcNAc by MshB was reduced by almost 89 % compared to data obtained with 10 mM DTT. If 1 mM VU5 was added to the above reaction, then it seems that the activity was restored. If 0.1 mM VU5 was added, then the activity was only partially restored. The mechanism of action by VU5 that prevents DTT from inhibiting MshB activity is unknown and requires further investigation.

It must be noted that there is no noticeable auto-fluorescence that takes place by GlcNAc, VU5 or even when they are incubated together without MshB.

Under conditions containing 100 mM DTT, when 200mM GlcNAc and 0.1mM VU5 is incubated with MshB, the rate of fluorescence is more than 4 times that of 200mM GlcNAc and MshB and more than 11 times than that of 0.1 mM VU5 and MshB. It is suggested that the combination of VU5 and GlcNAc somehow activates the production of primary amines

by MshB. Whether the amines produced come from VU5 or GlcNAc, or a mixture of both sources, remains unknown.

Thus while attempting to assay the published inhibitory action of VU5 using GlcNAc as the substrate, it was noticed that VU5 could be a better substrate than GlcNAc.

When equivalent amounts of GlcNAc and VU5 were incubated with MshB separately, the rate of increasing fluorescence in the VU5 containing reaction was almost 23 times faster than the rate of the GlcNAc containing experiment. Simultaneous incubation of VU5 and GlcNAc led to a rate of fluorescence almost 1.12 times that of VU5 alone.

The increased activity of VU5 compared to GlcNAc can possibly be attributed to the VU5 having a phenol ring in place of the inositol when compared to the natural substrate, GlcNAc-Ins, whereas GlcNAc does not have the corresponding ring. The inositol group has been shown to increase substrate specificity (Newton *et al.* 2006, Gammon *et al.* 2010) and because the VU5 has a cleavable amide bond, it is not particularly surprising that it could have increased activity when compared to GlcNAc.

There is also the possibility that the interaction of GlcN and VU5 or the interaction of VU5 and fluorescamine, causes an increase in fluorescence that is un-related to the amount of free primary amine reacting with the fluorescamine, however this also needs to be investigated further.

5 Concluding remarks

In this study it was shown that glycerol and acetate interact with the MshB active site and bind well enough to be visualised by X-ray crystallography. This supports previous proposals with regards to the orientation of binding of the natural substrate, and reiterates the importance of specific residues in substrate/inhibitor recognition. Studies have so far shown that *mshB* and *mca* mutants are more resistant to isoniazid and ethionamide, however perhaps if these enzymes are chemically inhibited, and are not made inactive by mutation or deletion, the isoniazid and ethionamide sensitivity could be retained while MSH levels are kept as low as possible, thus causing the bacteria to become more vulnerable to reactive oxygen species and other toxins and antibiotics.

An alternate catalytic mechanism is proposed where instead of His-144, Tyr-142 stabilises the tetrahedral intermediate while His-144 helps to stabilise substrate binding in the active site. Thus it is believed that His-144 neither forms part of the oxyanion hole as suggested by Maynes *et al.* (2003) nor does it act as a general acid catalyst as suggested by Huang and Hernick (2012). Rather it is important in substrate recognition (assisting in the changes necessary for the suggested induced fit of the substrate into the active site) and in binding the GlcNAc moiety of the natural substrate via a hydrogen bond between $N^{\varepsilon 2}$ His-144 and the 6-OH of the GlcNAc moiety.

Future work needs to include a full characteristic and kinetic study of the effect of VU5 and DTT on MshB activity with GlcNAc and GlcNAc-Ins, this will clarify the specific conditions under which MshB is inhibited or activated.

Non hydrolysable substrate analogues that co-crystallise with the MshB would be a great step forward in fully elucidating the active site geometry. This would also potentially be of use in Mca co-crystallisation attempts and could create the opportunity to make inhibitors that result

in a geminal diol for example, like the kAW inhibitor of the angiotensin converting enzyme (ACE) (Watermeyer *et al.* 2008). ACE is also a zinc ion dependent peptidase and the geminal diol that forms in the active site when co-crystallised with kAW resembles the proposed transition state between substrate and product, however the inhibitor is non-hydrolysable and therefore binds very tightly to the target active site. This could be a goal for the inhibition of MshB and Mca. If a substrate analogue inhibitor could be produced (that included pharmocaphores important in substrate recognition) that is non-hydrolysable (for example containing and ester instead of an amide) and could form a geminal diol, then perhaps it could experience a similar success to the kAW inhibition of ACE.

6 References

Aaron, L., Saadoun, D., Calatroni, I., Launay, O., Memain, N., Vincent, V., Marchal, G., Dupont, B., Bouchaud, O. & Valeyre, D. 2004, "Tuberculosis in HIV-infected patients: a comprehensive review", *Clinical Microbiology and Infection*, vol. 10, no. 5, pp. 388-398.

Accelrys, I. 2011, *Accelrys Draw*, 4.0th edn, Accelrys, Inc.

Afonine, P.,*Crystallographic Structure Validation*. Available: http://www.phenix-online.org/presentations/latest/pavel_validation.pdf

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. 1997, "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic acids research*, vol. 25, no. 17, pp. 3389-3402.

Altschul, S.F., Wootton, J.C., Gertz, E.M., Agarwala, R., Morgulis, A., Schäffer, A.A. & Yu, Y.K. 2005, "Protein database searches using compositionally adjusted substitution matrices", *Federation of European Biochemical Societies Journal*, vol. 272, no. 20, pp. 5101-5109.

Anderberg, S.J., Newton, G.L. & Fahey, R.C. 1998, "Mycothiol biosynthesis and metabolism", *Journal of Biological Chemistry*, vol. 273, no. 46, pp. 30391-30397.

Broadley, S.G., Gumbart, J.C., Weber, B.W., Marakalala, J.M., Steenkamp, D.J. & Sewell, B.T. 2012, "A new crystal form of MshB from *Mycobacterium tuberculosis* with glycerol and acetate in the active site suggests the catalytic mechanism", *Acta Crystallographica section D: Biological Crystallography*, vol. 68, pp. 1450-1459.

Brouwer, A. C. & Kirsch, J. F. 1982, "Investigation of diffusion-limited rates of chymotrypsin reactions by viscosity variation", *Biochemistry*, vol. 21, no. 6, pp. 1302-1307.

Browner, M.F., Smith, W.W & Castelhano, A.L. 1995, "Matrilysin-inhibitor complexes: common themes among metalloproteases", *Biochemistry*, vol. 34, no. 20, pp. 6602-6610.

Buchan, D., Ward, S., Lobley, A., Nugent, T., Bryson, K. & Jones, D. 2010, "Protein annotation and modelling servers at University College London", *Nucleic Acids Research*, vol. 38, suppl. 2, pp. W563-W568.

Buchmeier, N.A., Newton, G.L. & Fahey, R.C. 2006, "A mycothiol synthase mutant of Mycobacterium tuberculosis has an altered thiol-disulfide content and limited tolerance to stress", *Journal of Bacteriology*, vol. 188, no. 17, pp. 6245-6252.

Buchmeier, N. & Fahey, R.C. 2006, "The mshA gene encoding the glycosyltransferase of mycothiol biosynthesis is essential in Mycobacterium tuberculosis Erdman", *Federation of European Microbiological Societies Microbiology Letters*, vol. 264, no. 1, pp. 74-79.

Buchmeier, N.A., Newton, G.L., Koledin, T. & Fahey, R.C. 2003, "Association of mycothiol with protection of Mycobacterium tuberculosis from toxic oxidants and antibiotics", *Molecular Microbiology*, vol. 47, no. 6, pp. 1723-1732.

Cha, J. & Auld, D.S. 1997, "Site-directed mutagenesis of the active site glutamate in human matrilysin: investigation of its role in catalysis", *Biochemistry*, vol. 36, no. 50, pp. 16019-16024.

Collaborative Computational Project Number 4 1994, "The CCP4 suite: programs for protein crystallography", *Acta Crystallographica Section D: Biological Crystallography*, vol. 50, pp. 760-763.

Drenth, J. 1999, *Principles of protein X-ray crystallography*, Springer Verlag.

Duncan, K. 2004, "Identification and validation of novel drug targets in Tuberculosis", *Current Pharmaceutical Design*, vol. 10, pp. 3185-3194.

Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y. & Liang, J. 2006, "CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues", *Nucleic Acids Research*, vol. 34, suppl. 2, pp. W116-W118.

Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. 2010, "Features and development of Coot", *Acta Crystallographica Section D: Biological Crystallography*, vol. 66, no. 4, pp. 486-501.

Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M., Pieper, U. & Sali, A. 2006, "Comparative protein structure modeling using Modeller", *Current Protocols in Bioinformatics*, vol. 5, no. 5.6, pp. 1-30.

Fahey, R.C., Newton, G.L., Ta, P. & Buchmeier, N. 2008, "Dibenzothiazepinones and oxides thereof as inhibitors of bacterial MshC biosynthesis enzymes, homologs thereof, and inhibitor identification methods", U.S. patent WO 2008036139.

Fan, F., Vetting, M.W., Frantom, P.A. & Blanchard, J.S. 2009, "Structures and mechanisms of the mycothiol biosynthetic enzymes", *Current Opinion in Chemical Biology*, vol. 13, no. 4, pp. 451-459.

Fetterolf, B. & Bewley, C.A. 2004, "Synthesis of a bromotyrosine-derived natural product inhibitor of mycothiol-S-conjugate amidase", *Bioorganic & Medicinal Chemistry Letters*, vol. 14, no. 14, pp. 3785-3788.

Forman, H.J., Zhang, H. & Rinna, A. 2009, "Glutathione: overview of its protective roles, measurement, and biosynthesis", *Molecular Aspects of Medicine*, vol. 30, no. 1-2, pp. 1-12.

Gammon, D.W., Hunter, R., Steenkamp, D.J. & Mudzunga, T.T. 2003, "Synthesis of 2-deoxy-2-*C*-alkylglucosides of *myo*-inositol as possible inhibitors of a *N*-deacetylase enzyme in the biosynthesis of mycothiol", *Bioorganic & Medicinal Chemistry Letters*, vol. 13, no. 12, pp. 2045-2049.

Gammon, D.W., Steenkamp, D.J., Mavumengwana, V., Marakalala, M.J., Mudzunga, T.T., Hunter, R. & Munyololo, M. 2010, "Conjugates of plumbagin and phenyl-2-amino-1-thioglucoside inhibit MshB, a deacetylase involved in the biosynthesis of mycothiol", *Bioorganic & Medicinal Chemistry*, vol. 18, no. 7, pp. 2501-2514.

Georgiev, A., Vorobiev, S., Edstrom, W., Song, T., Laine, A., Hunt, J. & Allen, P. 2006, "Streak Seeding Automation Using Silicon Tools", Columbia University, New York.

Han, J.C., Han, G.Y. 1994, "A procedure for quantitative determination of tris(2-carboxyethyl)phosphine, an odorless reducing agent more stable and effective than dithiothreitol", *Analytical Biochemistry*, vol. 220, no. 1, pp. 5-10.

Hayward, D., Wiid, I. & van Helden, P. 2004, "Differential expression of mycothiol pathway genes: are they affected by antituberculosis drugs?", *International Union of Biochemistry and Molecular Biology: life*, vol. 56, no. 3, pp. 131-138.

Helgstrand, C., Hasan, M., Uysal, H., Haeggstrom, J.Z. & Thunnissen, M.M.G.M. 2011, "A leukotriene A_4 hydrolase-related aminopeptidase from yeast undergoes induced fit upon inhibitor binding", *Journal of Molecular Biology*, vol. 406, no. 1, pp. 120-134.

Huang, X. & Hernick, M. 2011, "A fluorescence-based assay for measuring *N*-acetyl-1-D-*myo*-inosityl-2-amino-2-deoxy-α-D-glucopyranoside deacetylase (MshB) activity", *Analytical Biochemistry*, vol. 414, pp. 278-281.

Huang, X. & Hernick, M. 2012, "Examination of the mechanism of *N*-acetyl-1-D-*myo*-inositol-2-amino-2-deoxy-α-D-glucopyranoside deacetylase (MshB) reveals an unexpected role for a dynamic tyrosine", *The Journal of Biological Chemistry*, vol. 287, no. 13, pp. 10424-10434.

Hyland, S.A., Eveland, S.S. & Anderson, M.S. 1997, "Cloning, expression, and purification of UDP-3-*O*-acyl-glcnac deacetylase from *Psudomonas aeruginosa*: a metalloprotease of the lipid A biosynthesis pathway", *Journal of Bacteriology*, vol. 179, no. 6, pp. 2029-2037.

Jackman, J.E., Raetz, C.R.H. & Fierke, C.A. 1999, "UDP-3-*O*-(*R*-3-hydroxymyristoyl)-*N*-acetylglucosamine deacetylase of *Escherichia coli* is a zinc metalloenzyme", *Biochemistry*, vol. 38, no. 6, pp. 1902-1911.

Jeffrey, P. 2006, *X-ray Data Collection Course*. Available: http://xray0.princeton.edu/~phil/facility/guides/XrayDataCollection.html

Jothivasan, V.K. & Hamilton, C.J. 2008, "Mycothiol: synthesis, biosynthesis and biological functions of the major low molecular weight thiol in actinomycetes", *Natural Product Reports*, vol. 25, no. 6, pp. 1091-1117.

Kleywegt, G.J. & Jones, T.A. 2002, "Homo Crystallographicus--Quo Vadis?", *Structure*, vol. 10, no. 4, pp. 465-472.

Knapp, S., Gonzalez, S., Myers, D.S., Eckman, L.L. & Bewley, C.A. 2002, "Shortcut to mycothiol analogues", *Organic Letters*, vol. 4, no. 24, pp. 4337-4339.

Koledin, T., Newton, G.L. & Fahey, R.C. 2002, "Identification of the mycothiol synthase gene (mshD) encoding the acetyltransferase producing mycothiol in actinomycetes", *Archives of Microbiology*, vol. 178, no. 5, pp. 331-337.

Krissinel, E. & Henrick, K. 2007, "Inference of molecular assemblies from crystalline state", *Journal of Molecular Biology*, vol. 372, pp. 774-797.

Laemmli, U.K. 1970, "Cleavage of structural proteins during the assembly of the head of bacteriophage T4", *Nature*, vol. 227, no. 5259, pp. 680-685.

Lamprecht, D.A., Muneri, N.O., Eastwood, H., Naidoo, K.J., Strauss, E., Jardine, A. 2012, "An enzyme initiated Smiles rearrangement enables the development of an assay of MshB, the GlcNAc-Ins deacetylase of mycothiol biosynthesis", *Organic & Biomolecular Chemistry*, vol. 10, pp. 5278-5288.

Lasker, K., Velazquez-Muriel, J.A., Webb, B.M., Yang, Z., Ferrin, T.E. & Sali, A. 2012, "Macromolecular assembly structures by comparative modeling and electron microscopy", *Methods in Molecular Biology*, vol. 857, pp. 331-350.

Li, L. & Ismagilov, R.F. 2010, "Protein crystallisation using microfluidic technologies based on valves, droplets, and SlipChip", *Annual Review of Biophysics*, vol. 39, pp. 139-158.

Lobley, A., Sadowski, M.I. & Jones, D.T. 2009, "pGenTHREADER and pDomTHREADER: new methods for improved protein fold recognition and superfamily discrimination", *Bioinformatics*, vol. 25, no. 14, pp. 1761-1767.

Lovell, S.C., Davis, I.W., Arendall III, W.B., de Bakker, P.I.W., Word, J.M., Prisant, M.G., Richardson, J.S. & Richardson, D.C. 2003, "Structure validation by Cα geometry: ϕ, ψ and Cβ deviation", *Proteins: Structure, Function, and Bioinformatics*, vol. 50, no. 3, pp. 437-450.

Marakalala, M.J. 2008, "Inhibition of a Mycothiol Biosynthetic Enzyme and a Detoxification Enzyme as Anti-Tubercular Drug Targets", University of Cape Town, PhD thesis.

Maynes, J.T., Garen, C., Cherney, M.M., Newton, G., Arad, D., Av-Gay, Y., Fahey, R.C. & James, M.N. 2003, "The crystal structure of 1-D-*myo*-inosityl 2-acetamido-2-deoxy-α -D-glucopyranoside deacetylase (MshB) from *Mycobacterium tuberculosis* reveals a zinc hydrolase with a lactate dehydrogenase fold", *The Journal of biological chemistry*, vol. 278, no. 47, pp. 47166-47170.

McCarthy, A.A., Knijff, R., Peterson, N.A. & Baker, E.N. 2003, "Crystallisation and preliminary X-ray analysis of *N*-acetyl-1-D-*myo*-inosityl-2-deoxy-α-D-glucopyranoside deacetylase (MshB) from Mycobacterium tuberculosis", *Acta crystallographica Section D: Biological Crystallography*, vol. 59, no. 12, pp. 2316-2318.

McCarthy, A.A., Peterson, N.A., Knijff, R. & Baker, E.N. 2004, "Crystal structure of MshB from *Mycobacterium tuberculosis*, a deacetylase involved in mycothiol biosynthesis", *Journal of Molecular Biology*, vol. 335, no. 4, pp. 1131-1141.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. & Read, R.J. 2007, "Phaser crystallographic software", *Journal of Applied Crystallography,* vol. 40, no. 4, pp. 658-674.

McGuffin, L.J., Bryson, K. & Jones, D.T. 2000, "The PSIPRED protein structure prediction server", *Bioinformatics,* vol. 16, no. 4, pp. 404-405.

Metaferia, B.B., Ray, S., Smith, J.A. & Bewley, C.A. 2007a, "Design and synthesis of substrate-mimic inhibitors of mycothiol-S-conjugate amidase from *Mycobacterium tuberculosis*", *Bioorganic & Medicinal Chemistry Letters,* vol. 17, no. 2, pp. 444-447.

Metaferia, B.B., Fetterolf, B.J., Shazad-Ul-Hussan, S., Moravec, M., Smith, J.A., Ray, S., Gutierrez-Lugo, M.T. & Bewley, C.A. 2007b, "Synthesis of natural product-inspired inhibitors of Mycobacterium tuberculosis mycothiol-associated enzymes: the first inhibitors of GlcNAc-Ins deacetylase", *Journal of Medicinal Chemistry,* vol. 50, no. 25, pp. 6326-6336.

Mitchell, E., Kuhn, P. & Garman, E. 1999, "Demystifying the synchrotron trip: a first time user's guide", *Structure,* vol. 7, pp. R111-R121.

Morgan, S., Grootendorst, P., Lexchin, J., Cunningham, C. & Greyson, D. 2011, "The cost of drug development: A systematic review", *Health Policy,* vol. 100, no. 1, pp. 4-17.

Moss, G.P. 2009, "Biochemical nomenclature committees", Available: http://www.chem.qmul.ac.uk/iupac/jcbn/ http://www.chem.qmul.ac.uk/iupac/misc/ppep4.html

Murshudov, G.N., Vagin, A.A. & Dodson, E.J. 1997, "Refinement of macromolecular structures by the maximum-likelihood method", *Acta Crystallographica Section D: Biological Crystallography,* vol. 53, no. 3, pp. 240-255.

Newton, G.L., Arnold, K., Price, M.S., Sherrill, C., Delcardayre, S.B., Aharonowitz, Y., Cohen, G., Davies, J., Fahey, R.C. & Davis, C. 1996, "Distribution of thiols in microorganisms: mycothiol is a major thiol in most actinomycetes", *Journal of Bacteriology,* vol. 178, no. 7, pp. 1990.

Newton, G.L., Av-Gay, Y. & Fahey, R.C. 2000, "A novel mycothiol-dependent detoxification pathway in mycobacteria involving mycothiol S-conjugate amidase", *Biochemistry,* vol. 39, no. 35, pp. 10739-10746.

Newton, G.L., Av-Gay, Y. & Fahey, R.C. 2000, "*N*-Acetyl-1-D-myo-inosityl-2-amino-2-deoxy-alpha-D-glucopyranoside deacetylase (MshB) is a key enzyme in mycothiol biosynthesis", *Journal of Bacteriology,* vol. 182, no. 24, pp. 6958-6963.

Newton, G.L. & Fahey, R.C. 2002, "Mycothiol biochemistry", *Archives of Microbiology,* vol. 178, no. 6, pp. 388-394.

Newton, G.L., Koledin, T., Gorovitz, B., Rawat, M., Fahey, R.C. & Av-Gay, Y. 2003, "The glycosyltransferase gene encoding the enzyme catalyzing the first step of mycothiol biosynthesis (mshA)", *Journal of Bacteriology,* vol. 185, no. 11, pp. 3476-3479.

Newton, G.L., Ta, P., Sareen, D. & Fahey, R.C. 2006, "A coupled spectrophotometric assay for 1-cysteine:1-D-*myo*-inositol 2-amino-2-deoxy-α-D-glucopyranoside ligase and its application for inhibitor screening.", *Analytical Biochemistry,* vol. 353, no. 2, pp. 167-173.

Newton, G.L., Ta, P., Bzymek, K.P. & Fahey, R.C. 2006, "Biochemistry of the Initial Steps of Mycothiol Biosynthesis", *Journal of Biological Chemistry,* vol. 281, no. 45, pp. 33910-33920.

Newton, G.L., Ko, M., Ta, P., Av-Gay, Y. & Fahey, R.C. 2006, "Purification and characterization of *Mycobacterium tuberculosis* 1-D-myo-inosityl-2-acetamido-2-deoxy-α-D-glucopyranoside deacetylase, MshB, a mycothiol biosynthetic enzyme", *Protein Expression and Purification,* vol. 47, no. 2, pp. 542-550.

Newton, G.L., Buchmeier, N. & Fahey, R.C. 2008, "Biosynthesis and functions of mycothiol, the unique protective thiol of Actinobacteria", *Microbiology and Molecular Biology Reviews,* vol. 72, no. 3, pp. 471-494.

Nicholas, G.M., Eckman, L.L., Newton, G.L., Fahey, R.C., Ray, S. & Bewley, C.A. 2003, "Inhibition and kinetics of *Mycobacterium tuberculosis* and *Mycobacterium smegmatis* mycothiol-S-conjugate amidase by natural product inhibitors", *Bioorganic & Medicinal Chemistry,* vol. 11, no. 4, pp. 601-608.

Nicholas, G.M., Eckman, L.L., Ray, S., Hughes, R.O., Pfefferkorn, J.A., Barluenga, S., Nicolaou, K.C. & Bewley, C.A. 2002, "Bromotyrosine-derived natural and synthetic products as inhibitors of mycothiol-S-conjugate amidase", *Bioorganic & Medicinal Chemistry Letters,* vol. 12, no. 17, pp. 2487-2490.

Nicholas, G.M., Newton, G.L., Fahey, R.C. & Bewley, C.A. 2001, "Novel bromotyrosine alkaloids: inhibitors of mycothiol S-conjugate amidase", *Organic Letters,* vol. 3, no. 10, pp. 1543-1545.

Nwaka, S. & Hudson, A. 2006, "Innovative lead discovery strategies for tropical diseases", *Nature Reviews Drug Discovery,* vol. 5, no. 11, pp. 941-955.

O'Brien, R.J. & Nunn, P.P. 2001, "The need for new drugs against tuberculosis. Obstacles, opportunities, and next steps", *American journal of respiratory and critical care medicine,* vol. 163, no. 5, pp. 1055-1058.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. & Ferrin, T.E. 2004, "UCSF Chimera - a visualization system for exploratory research and analysis", *Journal of Computational Chemistry,* vol. 25, no. 13, pp. 1605-1612.

Pflugrath, J.W. 1999, "The finer things in X-ray diffraction data collection", *Acta Crystallographica Section D: Biological Crystallography,* vol. 55, no. 10, pp. 1718-1725.

Pick, N., Rawat, M., Arad, D., Lan, J., Fan, J., Kende, A.S. & Av-Gay, Y. 2006, "In vitro properties of antimicrobial bromotyrosine alkaloids", *Journal of Medical Microbiology,* vol. 55, no. 4, pp. 407-415.

Ramachandran, G.N., Ramakrishan, C. & Sasisekharan, V. 1963, "Stereochemistry of polypeptide chain configurations", *Journal of Molecular Biology*, vol. 7, pp. 95-99.

Rawat, M., Newton, G.L., Ko, M., Martinez, G.J., Fahey, R.C. & Av-Gay, Y. 2002, "Mycothiol-deficient *Mycobacterium smegmatis* mutants are hypersensitive to alkylating agents, free radicals, and antibiotics", *Antimicrobial Agents and Chemotherapy*, vol. 46, no. 11, pp. 3348-3355.

Rawat, M., Uppal, M., Newton, G., Steffek, M., Fahey, R.C. & Av-Gay, Y. 2004, "Targeted mutagenesis of the *Mycobacterium smegmatis* mca gene, encoding a mycothiol-dependent detoxification protein", *Journal of Bacteriology*, vol. 186, no. 18, pp. 6050-6058.

Rawat, M., Kovacevic, S., Billman-Jacobe, H. & Av-Gay, Y. 2003, "Inactivation of mshB, a key gene in the mycothiol biosynthesis pathway in Mycobacterium smegmatis", *Microbiology*, vol. 149, no. Pt 5, pp. 1341-1349.

Rengarajan, J., Bloom, B.R. & Rubin, E.J. 2005, "Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 23, pp. 8327-8332.

Rogers, I. 2011, "Rationalising the Inhibition of M. tuberculosis MshB by a Series of Inhibitors Constructed From Plumbagin Conjugated Via a Variable Alkyl Linker to a Phenyl Thioglycoside", University of Cape Town, Masters dissertation.

Rhodes, G. 2006, "Crystallography made crystal clear: a guide for users of macromolecular models", *Academic Press*.

Sali, A. & Blundell, T.L. 1994, "Comparative protein modelling by satisfaction of spatial restraints", *Protein Structure by Distance Analysis*, , pp. 64–86.

Sareen, D., Newton, G.L., Fahey, R.C. & Buchmeier, N.A. 2003, "Mycothiol is essential for growth of *Mycobacterium tuberculosis* Erdman", *Journal of Bacteriology*, vol. 185, no. 22, pp. 6736-6740.

Sareen, D., Steffek, M., Newton, G.L. & Fahey, R.C. 2002, "ATP-dependent l-cysteine: 1-D-*myo*-inosityl 2-amino-2-deoxy-α-D-glucopyranoside ligase, mycothiol biosynthesis enzyme MshC, is related to class I cysteinyl-tRNA synthetases", *Biochemistry*, vol. 41, no. 22, pp. 6885-6890.

Sassetti, C.M., Boyd, D.H. & Rubin, E.J. 2003, "Genes required for mycobacterial growth defined by high density mutagenesis", *Molecular Microbiology*, vol. 48, no. 1, pp. 77-84.

Schwede, T., Diemand, A., Guex, N. & Peitsch, M.C. 2000, "Protein structure computing in the genomic area", *Research in Microbiology*, vol. 151, no. 2, pp. 107-112.

Shen, M. & Sali, A. 2006, "Statistical potential for assessment and prediction of protein structures", *Protein Science*, vol. 15, no. 11, pp. 2507-2524.

Sippl, M.J. 1993, "Recognition of errors in three-dimensional structures of proteins", *Proteins: Structure, Function, and Genetics,* vol. 17, pp. 355-362.

Small, P.M. & Fujiwara, P.I. 2001, "Management of tuberculosis in the United States", *New England Journal of Medicine,* vol. 345, no. 3, pp. 189-200.

Smith, A. 2008, "Design and synthesis of carbohydrate based derivatives as antimicrobial compounds", Dublin Institute of Technology, PhD Thesis.

Spies, H.S. & Steenkamp, D.J. 1994, "Thiols of intracellular pathogens. Identification of ovothiol A in *Leishmania donovani* and structural analysis of a novel thiol from *Mycobacterium bovis*". *European Journal of Biochemistry*, vol. 224, no. 1, pp. 203-213.

Steffek, M., Newton, G.L., Av-Gay, Y. & Fahey, R.C. 2003, "Characterization of Mycobacterium tuberculosis Mycothiol S-Conjugate Amidase", *Biochemistry,* vol. 42, no. 41, pp. 12067-12076.

Ung, K.S.E. & Av-Gay, Y. 2006, "Mycothiol-dependent mycobacterial response to oxidative stress", *Federation of European Biochemical Societies Letters,* vol. 580, no. 11, pp. 2712-2716.

Watermeyer, J.M., Kroger, W.L., O'Neill, H.G., Sewell, B.T. & Sturrock, E.D. 2008, "Probing the basis of domain-dependent inhibition using novel ketone inhibitors of angiotensin-converting enzyme", *Biochemistry,* vol. 47, no. 22, pp. 5942-5950.

Wetterholm, A., Medina, J.F., Radmark, O., Shapiro, R., Haeggstrom, J.Z., Vallee, B.L. & Samuelsson, B. 1992, "Leukotriene A_4 hydrolase: abrogation of the peptidase activity by mutation of glutamic acid-296", *Proceedings of the National Academy of Sciences,* vol. 89, pp. 9141-9145.

WHO 2011, *World Health Organisation*. Available: http://www.who.int/en/.

Wiederstein, M., Sippl, M.J. 2007, "ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins", *Nucleic Acids Research,* vol. 35, pp.w407-w410.

Xu, X., Vilcheze, C., Av-Gay, Y., Gomez-Velasco, A. & Jacobs Jr, W.R. 2011, "Precise Null Deletion Mutations of the Mycothiol Synthesis Genes Reveal Their Role in Isoniazid and Ethionamide Resistance in *Mycobacterium smegmatis*", *Antimicrobial Agents and Chemotherapy,* vol. 55, no. 7, pp. 3133.

7 Appendix A: Densitometry analysis

Protein purity was estimated by densitometry using the computer program ImageLab (Bio-Rad) (Appendix A). SDS-PAGE analysis was carried out on samples from S300 size exclusion chromatography experiments and subsequently analysed with ImageLab. Lanes were created manually, while the bands were first detected automatically using the high sensitivity setting and then manually adjusted as necessary. The average purity was calculated by averaging the individual percentage purity of the hypothetical MshB containing bands from lane 2 to 5. The calculation is shown below:

74.6/100 + 84.5/100 + 85.5/100 + 93.4/100 = 338/400

(338/400) / 4 = 84.5% average purity.

Figure 7-1 on the following page shows the image that was analysed. Lane 1 is the protein molecular weight marker, lane 2 represents sample that eluted at 166 min, lane 3 represents sample that eluted at 173 min, lane 4 represents sample that elute at 176.5 min and lane 5 represents sample that eluted at 183.5 min. These samples cover the apex of the peak from the chromatogram in Figure 3-10 (a).

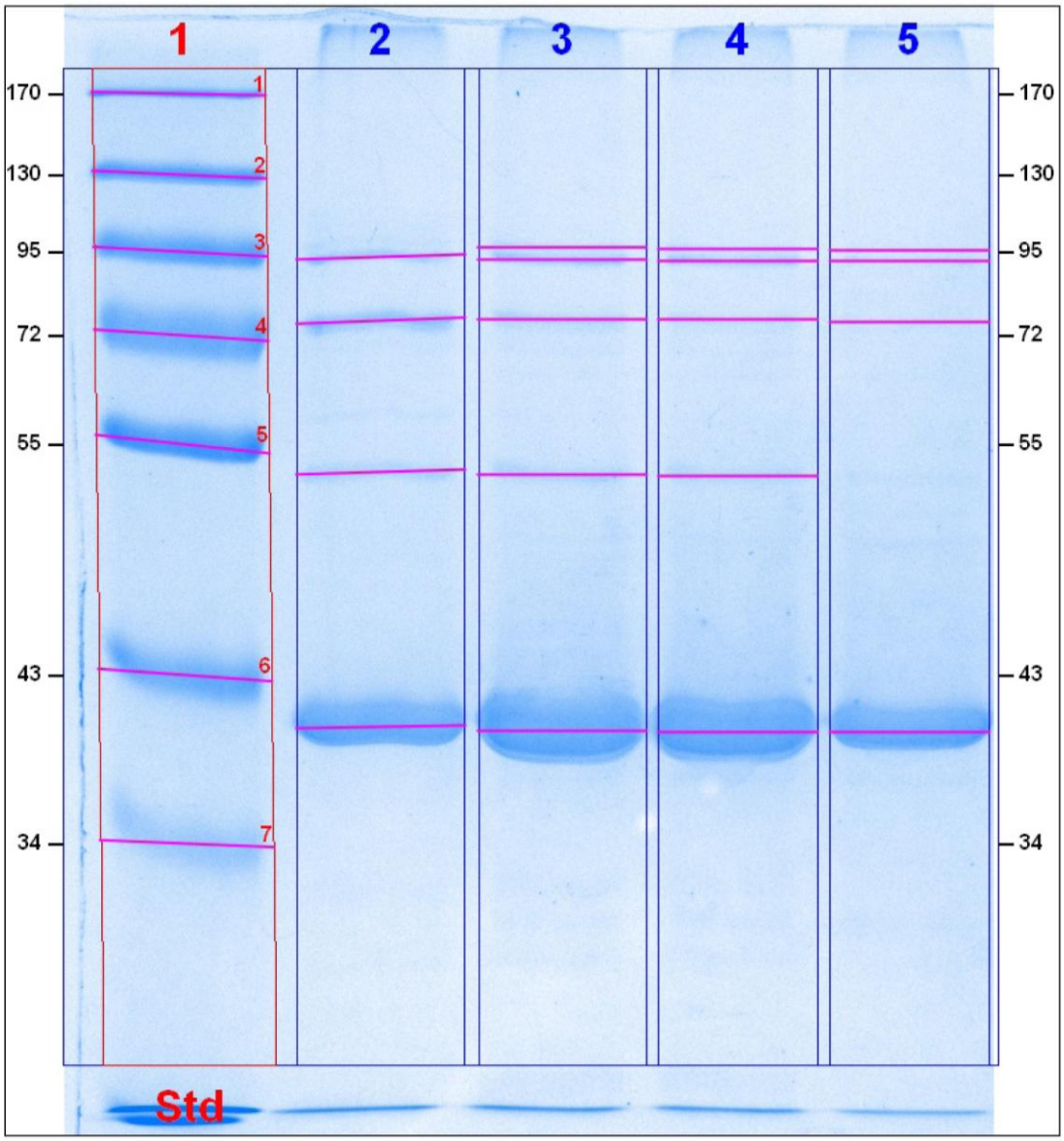


Figure 7-1: Densitometric analysis of SDS-PAGE from samples taken from S300 chromatogrphy

Acquisition Information

Imager	ChemiDoc XRS+
Exposure Time (sec)	0.029 (Auto - Intense Bands)
Flat Field	Applied (White)
Serial Number	721BR01814
Software Version	2.0.1
Application	Coomassie Blue
Excitation Source	White Trans illumination
Emission Filter	Standard Filter

Image Information

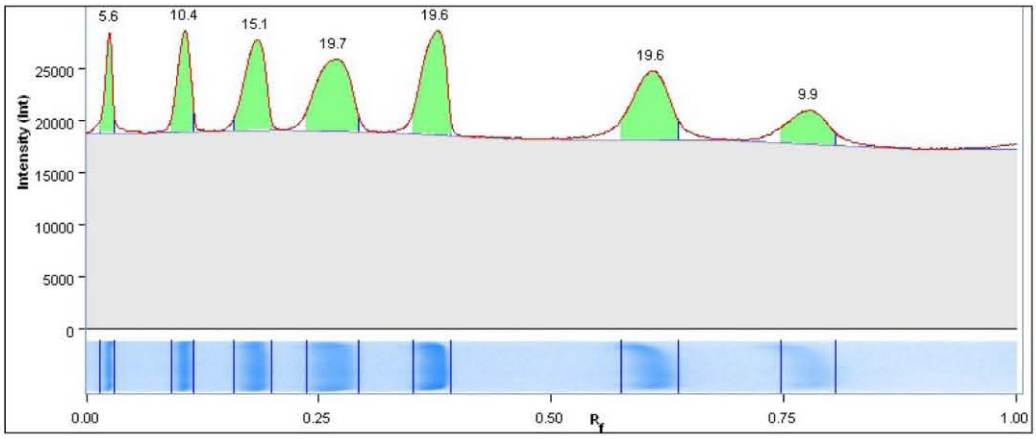
Acquisition Date	2011/12/08 04:06:53 PM
User Name	ChemiDoc XRS
Image Area (mm)	X: 47.8 Y: 58.1
Image Pixels	X: 700 Y: 852
Pixel Size (um)	X: 68.2 Y: 68.2
Data Range (Int)	13416 - 55913

Analysis Settings

Detection	Lane detection: Manually created lanes Band detection: Automatically detected bands with sensitivity: High Manually adjusted bands Lane Background Subtraction: Lane background subtracted with disk size: 10
Mol. Weight Analysis	Standard: Fermentas PageRuler Prestained Protein Ladder Standard lanes: first Regression method: Linear (semi-log)

Lane And Band Analysis

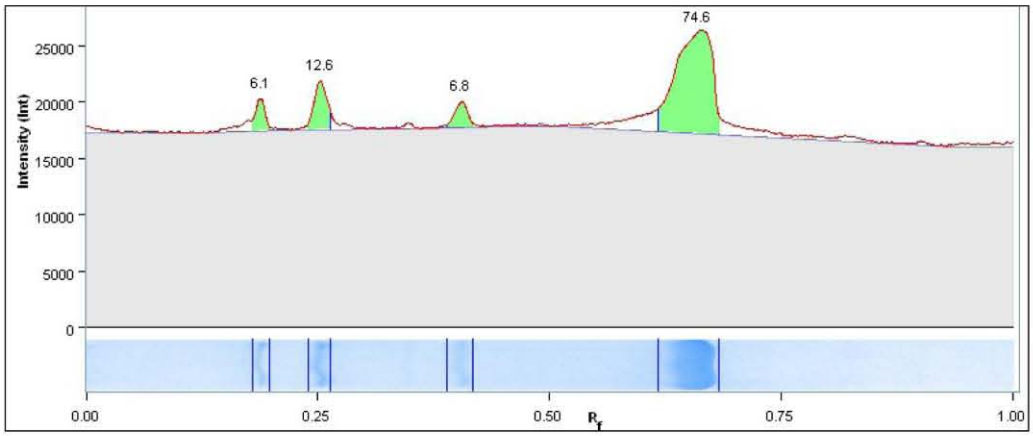
Lane 1 - Fermentas PageRuler Prestained Protein Ladder



Band No.	Mol. Wt. (KDa)	Band %
1	170.0	5.6
2	130.0	10.4
3	95.0	15.1
4	72.0	19.7
5	55.0	19.6
6	43.0	19.6
7	34.0	9.9

Band Detection	Automatically detected bands with sensitivity: High
Lane Background	Lane background subtracted with disk size: 10
Regression Equation	y = -0.899 * x + 2.17 R-squared value: R^2 = 0.931907

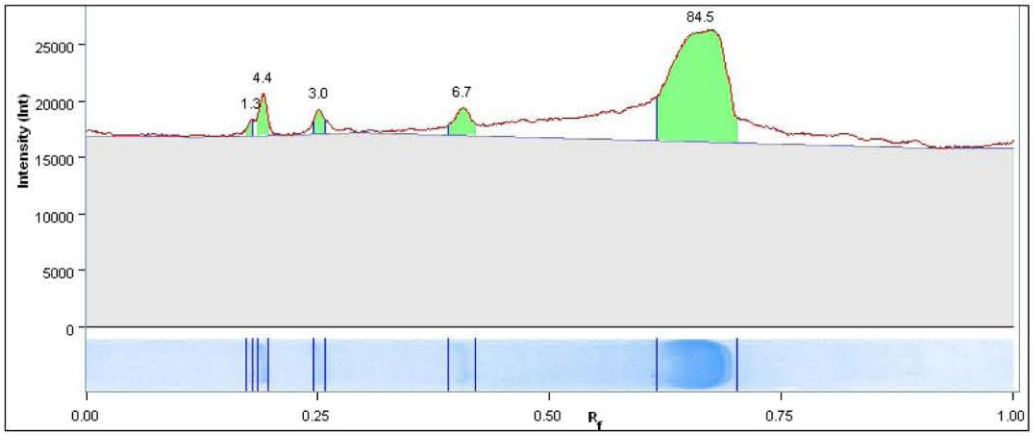
Lane 2



Band No.	Mol. Wt. (KDa)	Band %
1	100.1	6.1
2	87.7	12.6
3	64.1	6.8
4	37.7	74.6

Band Detection	Automatically detected bands with sensitivity: High
Lane Background	Lane background subtracted with disk size: 10
Regression Equation	y = -0.899 * x + 2.17 R-squared value: R^2 = 0.931907

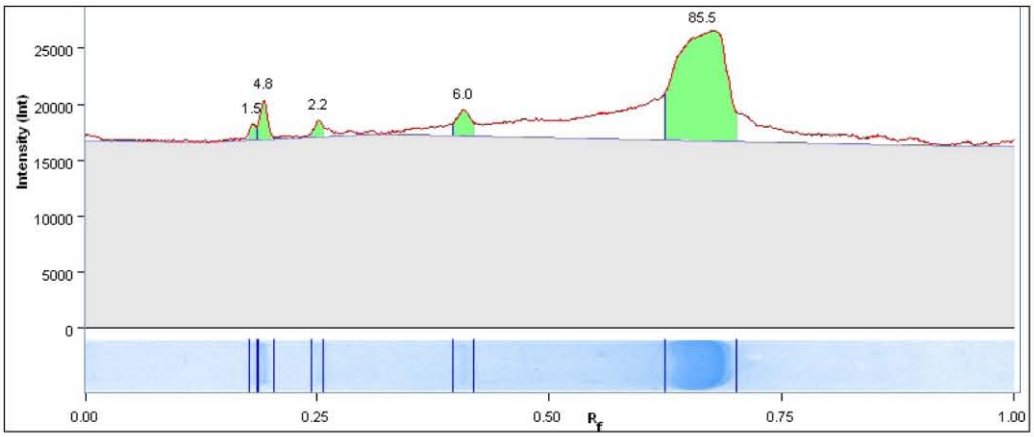
Lane 3



Band No.	Mol. Wt. (KDa)	Band %
1	102.1	1.3
2	99.6	4.4
3	88.0	3.0
4	63.7	6.7
5	37.4	84.5

Band Detection	Automatically detected bands with sensitivity: High
Lane Background	Lane background subtracted with disk size: 10
Regression Equation	y = -0.899 * x + 2.17 R-squared value: R^2 = 0.931907

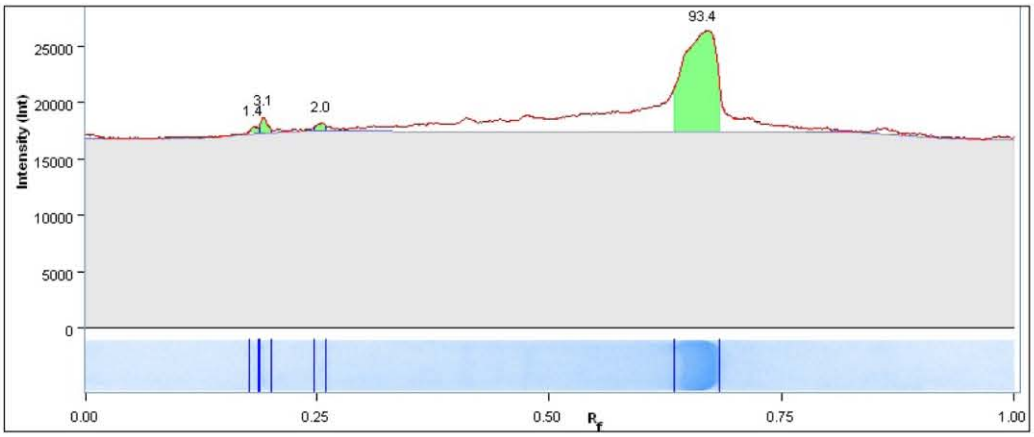
Lane 4



Band No.	Mol. Wt. (KDa)	Band %
1	101.8	1.5
2	99.3	4.8
3	88.0	2.2
4	63.5	6.0
5	37.3	85.5

Band Detection	Automatically detected bands with sensitivity: High
Lane Background	Lane background subtracted with disk size: 10
Regression Equation	y = -0.899 * x + 2.17 R-squared value: R^2 = 0.931907

Lane 5



Band No.	Mol. Wt. (KDa)	Band %
1	101.5	1.4
2	99.3	3.1
3	87.5	2.0
4	37.3	93.4

Band Detection	Automatically detected bands with sensitivity: High
Lane Background	Lane background subtracted with disk size: 10
Regression Equation	y = -0.899 * x + 2.17 R-squared value: R^2 = 0.931907