Helical reconstruction of *Mycobacterium smegmatis* Mycothiol S-conjugate amidase filaments

Jeremy Gareth Burgess

Thesis presented for the degree of
MSc Med
in Medical Biochemistry

Faculty of Health Sciences
UNVERSITY OF CAPE TOWN

Supervisors:

Prof. B. T. Sewell, Dr B. Weber, Dr J. Woodward

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

DECLARATION

I, JEREMY BURGESS hereby declare that the work on which this dissertation/thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: Signed by candidate

signature removed

Date: 27.03.2017

Abstract:

The metabolic pathway of mycothiol (MSH) is a major cellular defence against oxidative stress, and several antibiotics for mycobacteria, including *Mycobacterium tuberculosis.* The central enzyme used in the clearance of electrophilic toxins is Mycothiol S-conjugate amidase (Mca). Mca is similar to a biosynthetic enzyme MshB, which has partial overlapping substrate activity and is the closest homologue to Mca with a known structure. The basis for the substrate specificity differences in Mca and MshB is not well understood. Several regions of low sequence similarity between MshB and Mca are contained within an active site pocket, and these may affect the observed substrate preferences. However, these regions cannot be modelled in Mca with confidence, which makes it essential to obtain a structure of Mca experimentally. Mca is also a potential drug target, and a structure of Mca would enhance the rational design of inhibitors against the enzyme.

A search for crystalline forms of MsMca (*Mycobacterium smegmatis* Mca) led to the discovery of regular filaments, which showed helical order. Helical symmetry was estimated using power spectra from single filaments. The number of potential symmetry solutions was reduced using phase information from Fourier transforms of single filaments. Three possible solutions to the helical symmetry were suggested, two of which converged on the same symmetry parameters using Iterative Helical Real-Space Reconstruction. The first solution had a selection rule of l = *18m + n*, and the second *l = 20m + n.*

Reconstructions made from the predicted helical symmetries were compared in their power spectra and through rigid-body fitting with an atomic model of MsMca. The first reconstruction, with a final symmetry of $\Delta\phi = 20.05^{o}$ and $\Delta z = 10.27$ Å, better matched the predicted helical symmetry than did the second reconstruction. However, rigid-body fitting did not indicate either reconstruction as being superior.

Following this, the second reconstruction was improved using a number of additional techniques to those used in the initial reconstruction. These included the use of the fortuitous 3-fold cyclic symmetry, the removal of double-walled filaments, use of a cut-off filter for images with low correlation to projections of the 3D reconstruction, and use of a layer-line filter to reduce the noise in the images. These were used individually, then in a single reconstruction, to improve the and agreement between the predicted helical symmetry and that obtained from the reconstruction.

Several of the improved reconstructions were used via rigid-body fitting to assess the favoured handedness of the filament through examination of the major interfaces between subunits. These suggest that the 3-start helix is right-handed. Future work would be to determine the handedness of the filament using alternative techniques, such as metal-shadowing.

This work provides a springboard for high resolution cryo-electron microscopy, to determine a high-resolution structure of MsMca, which will enable rational inhibitor design and give the basis for the different substrate specificity in Mca and MshB.

Acknowledgements

I would like to acknowledge the following people for their support and encouragement over the last two years of this project:

Firstly, I would like to thank my supervisor, Prof. Trevor Sewell, for providing me with the opportunity to pursue this research, and for consistently giving me time, even when you were under pressure, with many deadlines to meet.

I would like to thank Dr Brandon Weber, for his support and supervision in the laboratory, especially in the earlier stages of this project. I also thank Brandon for his emotional support during the difficult times of this project.

I would like to thank Dr Jeremy Woodward, for his assistance in understanding helical symmetry, and the basics of helical reconstruction.

I thank Mohammed Jaffer, for teaching me to use the Electron Microscope, and for sitting with me for many hours as we tried to optimize the many samples for data collection.

I thank Gaynor Yorath, for her consistent support during the trying times of this project. Thank you Gaynor!

I thank my parents and sister, for their radical support for me during the difficult and confusing times of this work. Throughout all of the difficulties in mental health and emotional ups and downs.

I thank Stella Umuhoza, for her support and encouragement during the time of her stay with our family.

I thank my lovely fiancée Lou, for your determination to stick with me through thick and thin. I am so delighted to have you alongside me in this time.

Lastly, I want to thank my Lord, Jesus, for your immense and unending love for me. I also thank you for the incredible way in which you have made our bodies, and other living things. In studying life at the molecular level, we see your hand in the finest of details.

Table of Contents

List of Abbreviations

BOG	β-octylglucoside
CV	Column Volume
DOTS	Directly observed treatment short-course
EM	Electron Microscopy
ETB	ethambutol
ETH	ethionamide
GSH	Glutathione
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIV	Human Immunodeficiency Virus
IHRSR	Iterative helical real-space reconstruction
INH	Isoniazid
IPTG	Isopropyl β-D-1-thiogalactopyranoside
KPi	Potassium phosphate buffer system
LB	Lauria-Bertani broth
mBBr	monobromobimane
MBC	Minimum bactericidal concentration
MBP	Maltose binding protein
Mca	Mycothiol S-conjugate amidase
MDFF	Molecular Dynamics Flexible Fitting
MDR-TB	Multi-drug resistant TB
MES	2-(N-morpholino)-ethanesulfonic acid
MIC	Minimum Inhibitory Concentration
Ms	*Mycobacterium smegmatis*
MSH	Mycothiol
MshB	MSH deacetylase
MSmB	Mycothiol S-bimane
MSR	Mycothiol S-conjugate
MSSM	Mycothiol disulfide
MSSNaph	2-S-(mycothiolyl)-6-hydroxynaphthyldisulfide

MST	Mycothiol S-transferase
Mtr	Mycothiol disulfide reductase
Mt	*Mycobacterium tuberculosis*
NMR	Nuclear Magnetic Resonance
PDKAs	phenyl-diketo acids
PYZ	pyrazinamide
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEM	Scanning Electron Microscopy
STEM	Scanning Transmission Electron Microscopy
TB	Tuberculosis
TCA	trichloroacetic acid
TEM	Transmission Electron Microscopy
TEV	Tobacco Etch Virus
XDR-TB	Extensively drug resistant TB

1. Introduction

1.1. Tuberculosis

Tuberculosis (TB) is a contagious disease caused by *Mycobacterium tuberculosis*. TB accounts for the second highest number of deaths from a single infectious agent, after the human immunodeficiency virus (HIV).

In 2013, 9 million people developed active TB, and 1.5 million people died from TB. HIV-coinfected individuals accounted for 1.1 million of all TB cases, while TB-caused deaths amongst HIV positive individuals were 360 000. [1]

TB remains largely curable, as current short term treatment (directly observed treatment short-course, DOTS) has a global success rate of 86% among new TB cases [1].

Current treatment involves an initial phase of rifampicin (RIF), isoniazid (INH), pyrazinamide (PYZ), and ethambutol (ETB) daily for 2 months. The second phase involves a continuation of RIF and INH for a further 4 months, administered daily or 3 times per week [2].

1.2. Resistance to current therapies

Multi-drug resistant TB (MDR-TB), characterized by resistance to RIF and INH, occurs in approximately 3.5% of all new TB cases. Treatment for MDR-TB currently lasts for 24 months, although shorter regiments have been attempted, with improved success [1]. The global treatment success rate for MDR-TB is only 52%.

Extensively drug resistant tuberculosis (XDR-TB), is MDR-TB, with additional resistance to a fluoroquinolone, and at least one of the core second line agents. Treatment success is estimated at 28% globally, and XDR-TB was estimated to occur in 9.5% of MDR-TB cases [1].

The treatment of MDR-TB and XDR-TB is costly, requiring less effective second line drugs, and longer treatment [2]. In the case of XDR-TB, drugs are often required for which there is little clinical data, and treatment currently lasts for 24 months [3].

The rise of drug-resistant Tuberculosis presents an urgent need for novel anti-tubercular drugs, acting on novel targets within the bacterium.

1.3. TB drug discovery

TB drug discovery typically follows two routes – (a) the drug-to-target approach, or (b) the target-to-drug approach [4]. The drug-to-target approach begins with chemical screens against whole cells. Lead compounds are identified and characterized by their in vivo activity, including their Minimum Inhibitory/Bactericidal Concentrations (MIC/MBC) and cytotoxicity. The target(s) of the lead compounds are then identified and validated. Finally, the lead compounds are used in preclinical development, i.e. the lead compounds are optimized, adapted through medicinal chemistry and *in vivo* studies are undertaken [4].

The target-to-drug approach starts with the identification and validation of a potential drug target. The target enzyme is expressed and purified, before use in enzyme inhibition assays, to identify active compounds. The 3D structure of the enzyme is determined, and fragment-based *in silico* screening is used to complement the results of the *in vitro* enzyme inhibition assays. Lead compounds are tested for their whole cell activity, and where applicable, their MIC/MBCs are determined. Finally, lead compounds are subjected to preclinical development, as in the case of the drug-to-target approach [4].

The target-to-drug approach is more difficult, since hit compounds with activity against a purified enzyme target are often ineffective against whole cells [5]. Various resistance mechanisms exist in mycobacteria to prevent the activity of hit compounds, and these can range from the low permeability of the cell wall, to active efflux mechanisms. In addition, the physio-chemical properties of current drug libraries, from which hit compounds are selected, are optimized to treat human diseases such as cancer and may not be suited to the environment of the prokaryotic pathogen [5].

The target-to-drug approach may still be useful, particularly in the optimization of lead compounds against a target identified in the drug-to-target approach.

1.4. Rational drug design

Rational drug design is useful for either approach to drug discovery. In the drug-to-target approach, after the identification of a target of the active compound, the lead compound may be optimized with knowledge of the 3D structure of the target. The structure of the target can be used to select the library of compounds for screening experiments in the target-to-drug approach [4].

In the following section, we review the use of rational drug design in the development of several novel anti-tubercular agents.

1.4.1. InhA and pyridomycin

InhA is the target of clinical drugs isoniazid (INH) and ethionamide (ETH). INH resistance can occur through mutations in the catalase-peroxidase KatG, required for activation of the isoniazid pro-drug. Using a drug-to-target approach, the natural product pyridomycin (from *Dactylosporangium fulvum*) was shown to inhibit InhA and was active against *katG* mutants [6].

InhA utilizes NADH (a cofactor) and the substrate 2-*trans*-Enoyl-(Acyl-Carrier-Protein). Co-crystallization of pyridomycin and InhA showed pyridomycin to occupy both binding sites, forming interactions with several residues required for binding to NADH [7].

Several analogues of pyridomycin were developed against InhA. The most active analogue was co-crystallized with InhA, and shown to occupy the same position as the parent compound. These structures have improved the development and optimization of pyridomycin and may aid in the search for other direct InhA inhibitors [7].

1.4.2. EthR and ethionamide boosters

Ethionamide (ETH) is a second-line drug that targets the biosynthesis of mycolic acids by inhibition of InhA. ETH is activated by EthA, which is regulated by EthR, a transcriptional repressor of *ethA*. The inhibition of EthR leads to increased levels of EthA and increased potency of ETH [4, 8].

EthR is a homodimer, with two functional domains [9]. Each monomer has 9 α-helices. The first 3 α-helices of each monomer form the DNA-binding domain. The DNA-binding domain is connected to a core-domain, which is necessary for dimerization. In the first crystal structure [9], a hydrophobic ligand was seen in the core domain of each monomer.

From the interactions of the ligand in the binding pocket, a pharmacophore was designed and used to select a small library of compounds for physical screens for inhibition of EthR.

The physical screens revealed a family of compounds with good inhibitory activity. Members of this family of compounds were co-crystallized with EthR, and were found to occupy the same binding-pocket within the core-domain of EthR. The hit compounds were used in disk diffusion assays alongside ETH to test their capacity to boost the action of ETH on the growth of *M. tuberculosis*.

Analogues of the most active hit compound were evaluated regarding their bioavailability in a number of environments. Two analogues were tested with ETH in mice infected with *M. tuberculosis*. While both compounds had shown good *in vitro* activity, only one had a significant boosting effect on ETH in the *M. tuberculosis*-infected mice. Co-administration of this analogue with

ETH caused a similar reduction of bacterial load to treatment with ETH alone at three-times the dose [9].

1.4.3. Malate synthase and phenyl-diketo acids

Rational drug design was used successfully in the targeting of malate synthase, a crucial component of the glyoxylate shunt in mycobacteria [10].

From the crystal structure of GlcB (malate synthase), a small number of substrate-like small molecules were selected for an assay against GlcB. Those displaying activity were phenyl-diketo acids (PDKAs). Modification of the phenyl-ring by the addition of halogen groups was used to improve the stability of the PDKAs. Several analogues of PDKA were co-crystallized with GlcB. Additional analogues were developed to strengthen active site interactions. These were assayed against GlcB *in vitro*, and were active in whole cell assays against *M. tuberculosis*. Pro-drug esters of the analogues were developed, to increase drug permeability. The esters of the PDKA analogues displayed activity in a mouse model of TB [10].

Multiple targets in *M. tuberculosis* may be exploited by rational drug design. Mycobacterial metabolic pathways may harbour several druggable targets, which is useful in the attempt to target drug-resistant strains of *M. tuberculosis*. Pathways that are unique to the organism and essential for the growth and survival of M. tuberculosis are examined first. In this section of the review, we examine one of the unique pathways of *M. tuberculosis*, and discuss the various targets within the pathway, and give special focus to one of the targets, discussing its potential in structure-assisted drug design.

1.5. Oxidative stress and low weight thiols

M. tuberculosis, like other aerobic organisms, must survive in the presence of oxygen. Oxygen is both essential and toxic to aerobic life, and cells have numerous mechanisms to guard against the toxic products of its use [11].

Thiols found in enzymes, co-enzymes and free cysteine are susceptible to oxidation by atmospheric oxygen [12]. Heavy-metal catalysed thiol oxidation is rapid, and results in the production of reactive oxygen species, which are toxic to the cell. Free cysteine is particularly sensitive to metal-catalysed autoxidation because of its free amino- and carboxyl groups, which assist in the binding of heavy

metals [12]. However, heavy metals are also required by the cell as co-factors of many proteins. For this reason, there are specific cellular systems for heavy metal import, export, transport and delivery to target proteins. These are arranged to ensure that only the correct metal ions are inserted into their respective apoproteins [13], and minimize the intracellular concentrations of many free heavy metals.

The cysteine autoxidation can also be reduced if much of the free cysteine supply is converted to other low weight thiols less sensitive to autoxidation [12].

Glutathione, formed from the condensation of glutamic acid, cysteine and glycine [γ-Glu-Cys-Gly, or GSH] is four times less sensitive to autoxidation than free cysteine [12]. Glutathione is the most common low weight thiol present in aerobic organisms, and is involved in a myriad of oxidoreductive processes. The glutathione system is the most important metabolic network for the modification or removal of reactive oxygen species, other endogenous electrophiles and various xenobiotics [11].

This highly important thiol, although ubiquitous amongst eukaryotes, is not found in many prokaryotes, particularly amongst Gram positive bacteria. Instead, these organisms utilize alternative low weight thiols, and this is the case for most *Actinomycetes*, including *M. tuberculosis* [14].

1.6. Mycothiol

The major low weight thiol of *M. tuberculosis* is *mycothiol* (MSH) [14]. It was detected in *Streptomyces clavuligerus*, then in *Mycobacterium bovis*, and its structure was later determined as 1-D-myo-inositol-2-(N-acetyl-L-cysteinyl) amino-2-deoxy-alpha-D-glucopyranoside or AcCys-GlcN-Ins [15–17] (figure 1.1). MSH is similar in function to glutathione, serving as an intracellular protectant against auto-oxidation, various reactive oxygen and nitrogen species, and a variety of toxins [18, 19]. Importantly, since MSH is not produced in eukaryotes, its metabolism has been investigated as a potential target for novel therapy in the treatment of TB [20].

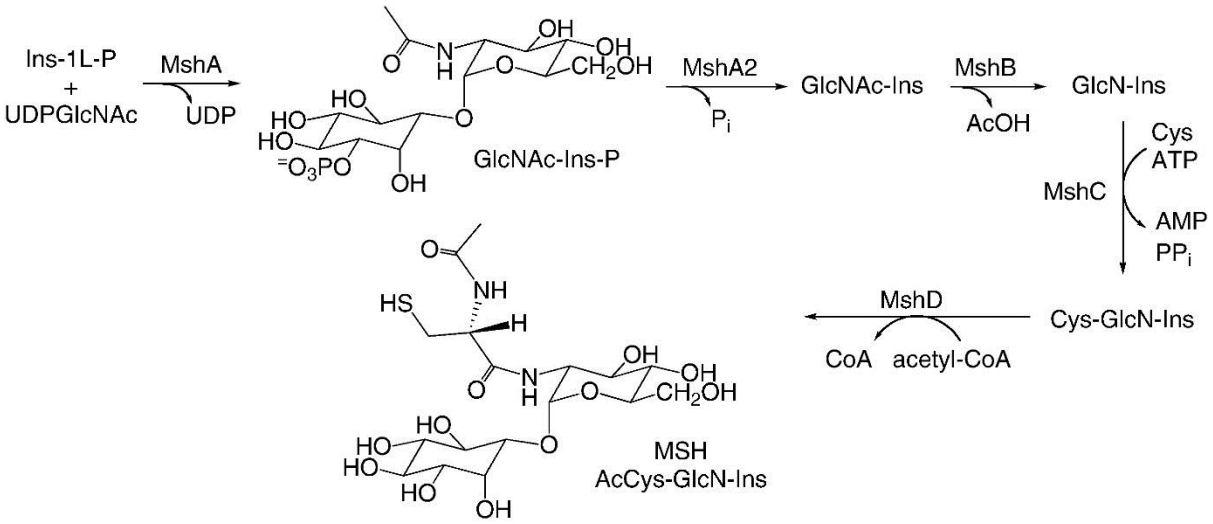


Figure 1.1: Biosynthesis of MSH. Reproduced with permission from R.Fahey, 2013. The names of enzymes and reactions can be found in table 1.1.

1.7. MSH biosynthesis

Following the discovery of MSH, the genes for MSH biosynthesis were rapidly identified using transposon- and chemical mutants [18, 21–23]. A summary of the genes required for MSH biosynthesis can be found in table 1.1.

Gene Enzyme *Rv number*	Catalysed reaction	Essentiality	References
MshA MSH glycosyltransferase *Rv0486*	Transglycosylation of UDP-GlcNAc with Ins-1L-P forming GlcNAc-Ins-P and UDP	Non-essential for growth in mouse models	[18, 24]
MshA2* Inositol monophosphate phosphatase *Unconfirmed*	Hydrolysis of GlcNAc-Ins-P to GlcNAc-Ins	Unknown	[25]
MshB MSH deacetylase *Rv1170*	Deacetylation of GlcNAc-Ins to GlcN-Ins	Non-essential	[26, 27]
MshC MSH ligase *Rv2130c*	ATP dependent ligation of GlcN-Ins and L-cysteine to form Cys-GlcN-Ins	Essential	[22]
MshD MSH synthase *Rv0189*	Transfer of acetyl group from Acetyl-CoA to the free amino group of Cys-GlcN-Ins, forming AcCys-GlcN-Ins (MSH)	Non-essential	[28]

Table 1.1: An overview of the mycothiol biosynthetic enzymes, their catalysed reactions and their requirement for *M. tuberculosis* survival. *- The gene encoding the predicted MshA2 is unknown. Mutagenic studies in putative Ins-P phosphatase genes in *M. tuberculosis* have revealed *impC* (*Rv3137*) as an essential gene, but its Ins-P phosphatase activity has yet to be experimentally verified.

1.8. Mycothiol-dependent enzymes

MSH can act as a substrate for additional enzymes, of which several are mentioned here.

MSH is oxidized to mycothiol disulfide (MSSM) non-enzymatically, or through the action of an MSH-dependent peroxidase (still to be identified) [12]. MSSM is rapidly reduced to MSH in an NADPH-dependent manner by mycothiol disulfide reductase (Mtr) [29]. This helps to maintain the needed intracellular reserves of MSH. Mtr is a homodimeric flavoprotein containing FAD as a cofactor, and is functionally analogous to reductases of other low molecular weight thiols, such as the glutathione and tryptathione reductases. These belong to the flavoprotein disulfide reductase family, which have conserved active site residues and FAD- and NADPH-binding regions [29]. The importance of Mtr to MSH metabolism has stimulated several studies on its potential as a drug target [30].

MSH in reduced form is used in the effective detoxification of multiple endogenous toxins and xenobiotics. This may occur non-enzymatically, through reaction of MSH with highly reactive alkylating agents such as mBBr, or may be catalysed through the action of a Mycothiol S-Transferase (MST) [31]. In both instances an S-conjugate of MSH and the toxin is formed.

The mycothiol S-conjugate (MSR) is rapidly cleaved by a specific amidase, mycothiol S-conjugate amidase (Mca), forming the products GlcN-Ins and AcCysS-R [21] (figure 1.2). AcCysS-R is exported from the cell, through an unknown transporter, and GlcN-Ins may be recycled to MSH biosynthesis.

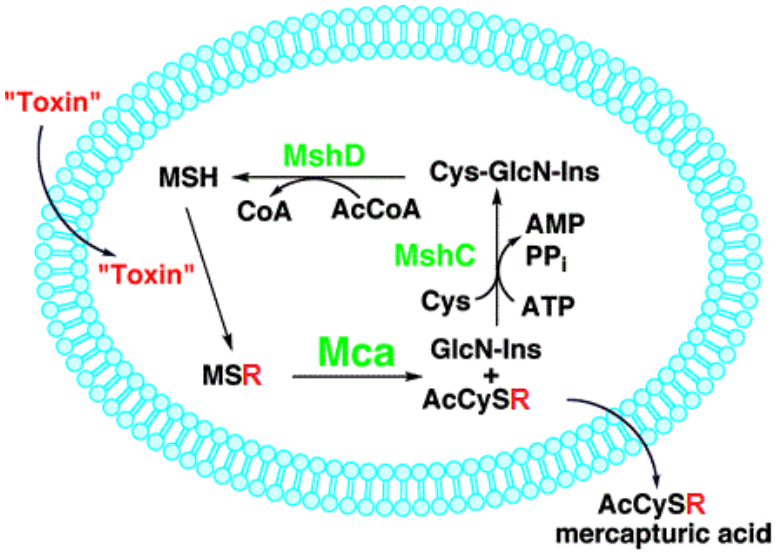


Figure 1.2. MSH-dependent detoxification of drugs and metabolites. Non-enzymatic condensation of MSH with thiol-reactive drugs, or condensation via a mycothiol S-transferase, produces mycothiol S-conjugates (MSRs). These are rapidly cleaved by Mca, to produce GlcN-Ins and AcCysSR. AcCysSR is exported from the cell, while GlcN-Ins may be recycled to MSH biosynthesis. Figure adapted and reproduced with permission from Newton et al. 2008 [32].

1.9. Mca

Mycothiol S-conjugate amidase activity was first detected in *M. smegmatis* cells [21], when cells were treated with monobromobimane (mBBr) to monitor intracellular mycothiol levels. A bimane derivative of N-acetylcysteine, AcCySmB, was detected, as were increased levels of GlcN-Ins, the remaining product of MSmB (the bimane derivative of Mycothiol) cleavage. The enzyme responsible for this was isolated, and showed activity against a large variety of Mycothiol S-conjugates (MSRs). N-terminal sequencing of the enzyme from *M. smegmatis* allowed identification of the corresponding gene Rv1082 in *M. tuberculosis*.

Mca was recombinantly expressed in *E. coli*, and purified successfully [33]. Kinetic analysis of the enzyme showed that the mycothiol moiety must be retained for full activity (removal of the acetyl, Cys or Ins groups of MSmB leads to a 10^3-fold, $3x10^5$-fold, or $8x10^4$-fold loss of activity, respectively.) The preferred S-conjugates substrates of Mca are smaller, hydrophobic adducts, MSmB and the MS-acetophenone adduct. Larger S-conjugates can be accepted with little loss in activity, for example the MS-cerulenin adduct. Mca is even used in the detoxification of rifamycin, though Mca is less active with the MS-rifamycin S adduct than with MSmB (20-40 fold lower activity).

Mca has low activity with MSH itself (10^3-fold less activity than for MsmB), yet at the intracellular levels of MSH (3mM in *M. smegmatis*), the total MSH degradation rate is 20% per hour [33].

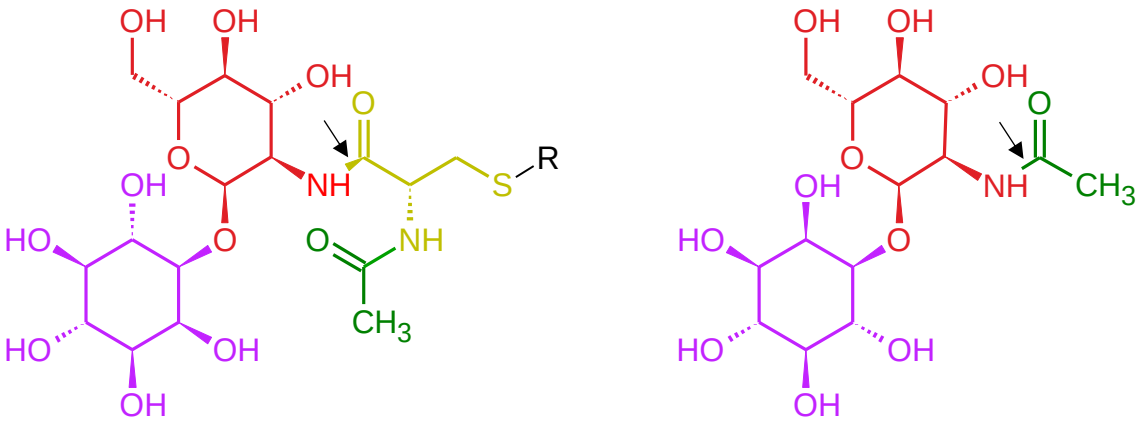
Mca is functionally similar to MshB, the deacetylase involved in MSH biosynthesis. MshB and Mca are both zinc metalloproteases, and are able to cleave the amide bond of a wide variety of MSH-derived substrates [27].

MshB has minor amidase activity with MSRs (for MsmB, MshB has $1.7x10^3$-fold lower activity than does Mca). In contrast, Mca has minor deacetylase activity with GlcNAc-Ins ($4x10^4$-fold lower activity than MshB).

Mca's activity against mycothiol S-conjugates of antibiotics such as cerulenin and rifamycin S [21, 33] have led others to suggest Mca as a potential target for novel inhibitors and it has been the study of several efforts in rational inhibitor design [34–36].

However, the rational inhibitor design against Mca is hindered by the lack of an experimental structure of Mca.

What can be known about the structure of Mca is currently derived from the knowledge of MshB, which is the closest homologue to Mca with a solved crystal structure [37].



Mycothiol, **MSH**, or AcCys-GlcN-Ins GlcNAc-Ins

Figure 1.3. Skeletal structures of Mycothiol (left) and GlcNAc-Ins (right). The natural substrates of Mca and MshB are hydrolysed at the same position, indicated by a black arrow. For Mca, MSH is conjugated to an electrophile through the sulfhydryl group, forming MSR. MSR is cleaved specifically by Mca at the designated position. MshB catalyses the hydrolysis of GlcNAc-Ins (right), at the indicated position.

What follows is a brief description of the structure of MshB, and what has been learned of its catalytic mechanism. Where possible, comparisons with Mca are given.

1.10. MshB

MshB has a mixed α/β fold, with 10 β-strands, and six major α-helices. Seven strands form the major mixed β-sheet (β1-6 and β10), and there is a small three-stranded antiparallel β-sheet (β7-9).

To date, three crystal structures have been determined for MshB. The first contained no ligand in the active site, but revealed a metal-binding triad (H13, D16, H147), co-ordinated by a catalytic zinc ion [37] (figure 1.4). Several residues were suggested to form part of the catalytic mechanism (which was later refined). These included the proposed general base D15, and several others, including H144 and D146. The reaction mechanism proposed then involved a nucleophilic attack by a water molecule on the carbonyl carbon of the acetyl group of GlcNAc-Ins. The tetrahedral intermediate that followed would be stabilized by the Zn^{2+} ion, and the side chain of H144. Proton transfer from D15 to the N of GlcN-Ins would facilitate the collapse of the tetrahedral intermediate, and release of GlcN-Ins, followed by acetic acid [37].

The metal binding residues and those proposed to assist in catalysis in MshB are conserved in Mca. MshB shares moderate sequence identity (32%) with Mca, and 43% over the first 200 residues. A

model of the first 180-residues may be built reliably [37], whereas there is poor agreement for the remaining residues. The modelled region contains the metal-binding residues and the catalytic residues.

The second crystal structure of MshB [38] contained a β-octylglucoside (BOG) ligand in the active site. The bound ligand allowed modelling of GlcNAc-Ins in the active site. The carbonyl oxygen of the amide bond of GlcNAc-Ins is positioned as a ligand of the catalytic Zn^{2+}.

A number of hydrogen bonds were seen which would stabilize the substrate for the purpose of catalysis. The hydrogen bonds formed between the O6 hydroxyl of BOG and H144 and D95, the O4 hydroxyl of BOG and D95 and R68, and the O3 hydroxyl of BOG and R68.

The reaction mechanism proposed from the second structure was similar to that proposed in [34], but the role of the H144 side chain was revised as a stabilizer of the GlcN ring of the substrate and not a direct stabilizer of the proposed tetrahedral intermediate.

The residues H144, D95 and R68 proposed to stabilize the GlcN-moiety of GlcNAc-Ins by MshB are conserved in Mca. This is unsurprising, given that the preferred substrates of MshB and Mca (GlcNAc-Ins and MSR, respectively) share the glucopyranoside ring.

The third crystal structure of MshB [36] contained a molecule of acetic acid and glycerol in the active site (figure 1.5).

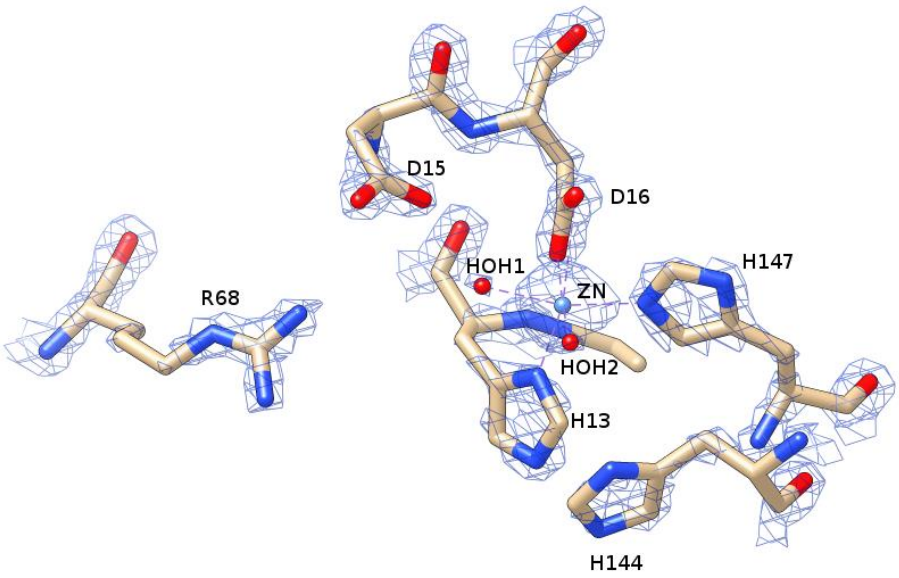
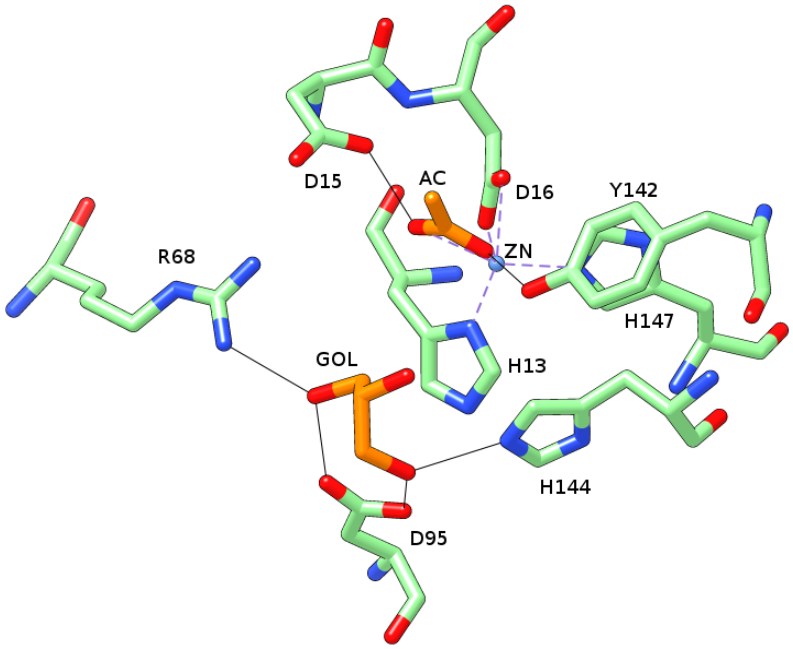


Figure 1.4 Stick representation of MshB active site residues from the crystal structure 1Q74 [37]. Waters are shown as red spheres. Zn as a blue sphere. Carbon atoms – beige, Nitrogen – blue, Oxygen – red. The catalytic Zinc is co-ordinated by the residues H13, D16 and H147, and two water ligands. HOH1 is likely the catalytic water molecule, and HOH2 is likely displaced by the carbonyl oxygen of the substrate [37].

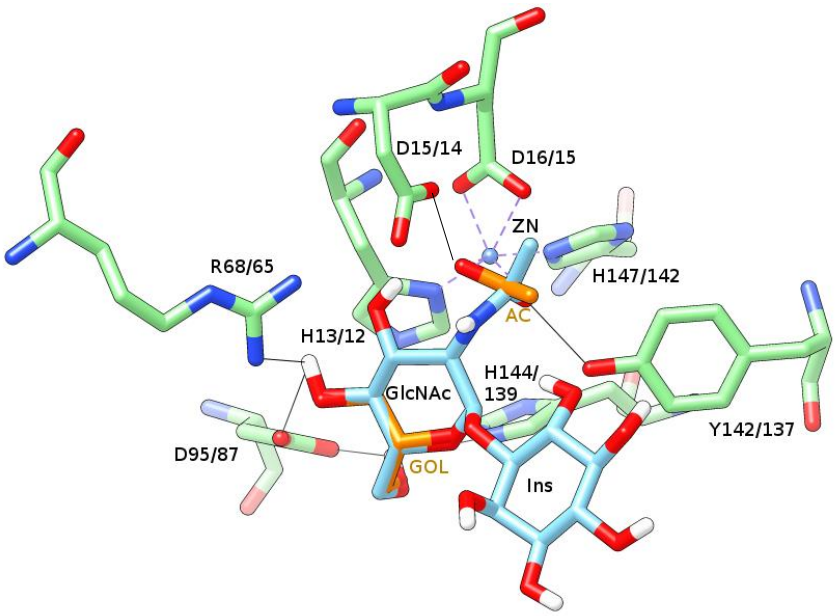
The molecule of acetic acid was positioned with its carboxyl oxygen atoms as ligands of the Zn^{2+} ion. The Y142 side chain was in a different conformation to those seen in the structures from [37] and [35]. The conformation of Y142 in this structure allowed H-bonding between the Y142 hydroxyl and the first carboxyl oxygen of the acetic acid. In addition, H-bonding was possible between the first carboxyl oxygen of D15 and the second carboxyl oxygen of the acetic acid. A glycerol molecule was hydrogen-bonded by residues R68, D95 and H144, in a similar network to that seen for BOG, as described earlier. GlcNAc-Ins was modelled in the active site by superimposition of the glucopyranoside ring onto the corresponding atoms of the glycerol molecule, and the N-acetyl group onto the corresponding atoms of the acetate molecule (figure 1.5).

The position of the amide group is such that D15 is positioned to mediate a nucleophilic attack on the amide carbonyl carbon by a water molecule. The general reaction mechanism is shown in figure 1.6. The main difference between that proposed by Maynes et al. [34] and [36], is that Y142 and not H144 stabilizes the carbonyl oxygen of the reaction intermediate. A conformational change in Y142 enables the release of the second product, acetate, following the exit of GlcN-Ins.

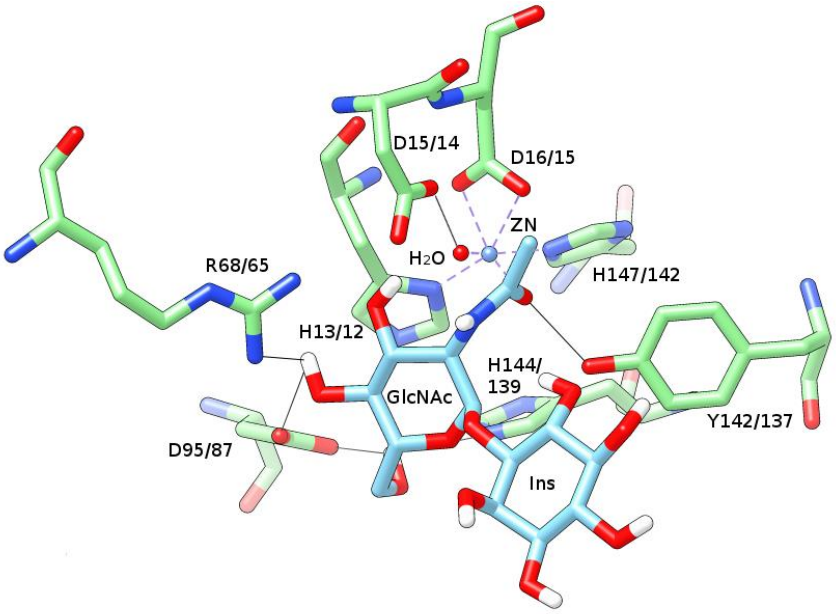
Molecular dynamics simulations were used to model the interactions of the inositol ring in the active site. These revealed hydrophobic interactions between M98 and the inositol ring, hydrogen bonds between E47 and the 3'- and 4'-hydroxyls, and water-mediated hydrogen bonding between N261 and the 5'- and 6'-hydroxyls (figure 1.7). The corresponding residues in Mca implicated in the binding of the inositol moiety are E46 (E47 in MshB), L90 (equivalent to M98 in MshB) and Q257 (N261 in MshB).



(A)



(B)



(C)

Figure 1.5: Stick representation of MshB active site residues (green) from the crystal structure 4EWL [39]. (A) Acetate (AC) and Glycerol (GOL) are bound in the active site. The catalytic Zinc (blue sphere) is co-ordinated by the residues H13, D16, H147 and the two carboxyl oxygens of acetate. Glycerol is stabilized by hydrogen bonds with R68, D95 and H144, while D15 and Y142 form hydrogen bonds (black lines) with acetate. (B) GlcNAc-Ins (cyan) modelled in place by alignment with acetate and glycerol (orange) from the structure of 4EWL. (C) Removal of the acetate and glycerol molecules allows the modelling of the catalytic water, thought to occupy a Zn-ligand site, and is positioned to make a nucleophilic attack on the acetyl carbonyl carbon of GlcNAc-Ins. Carbon atoms – green. Nitrogen – blue. Oxygen – red. Hydrogen bonds are shown in black. Residues numbers are given for MshB/Mca.

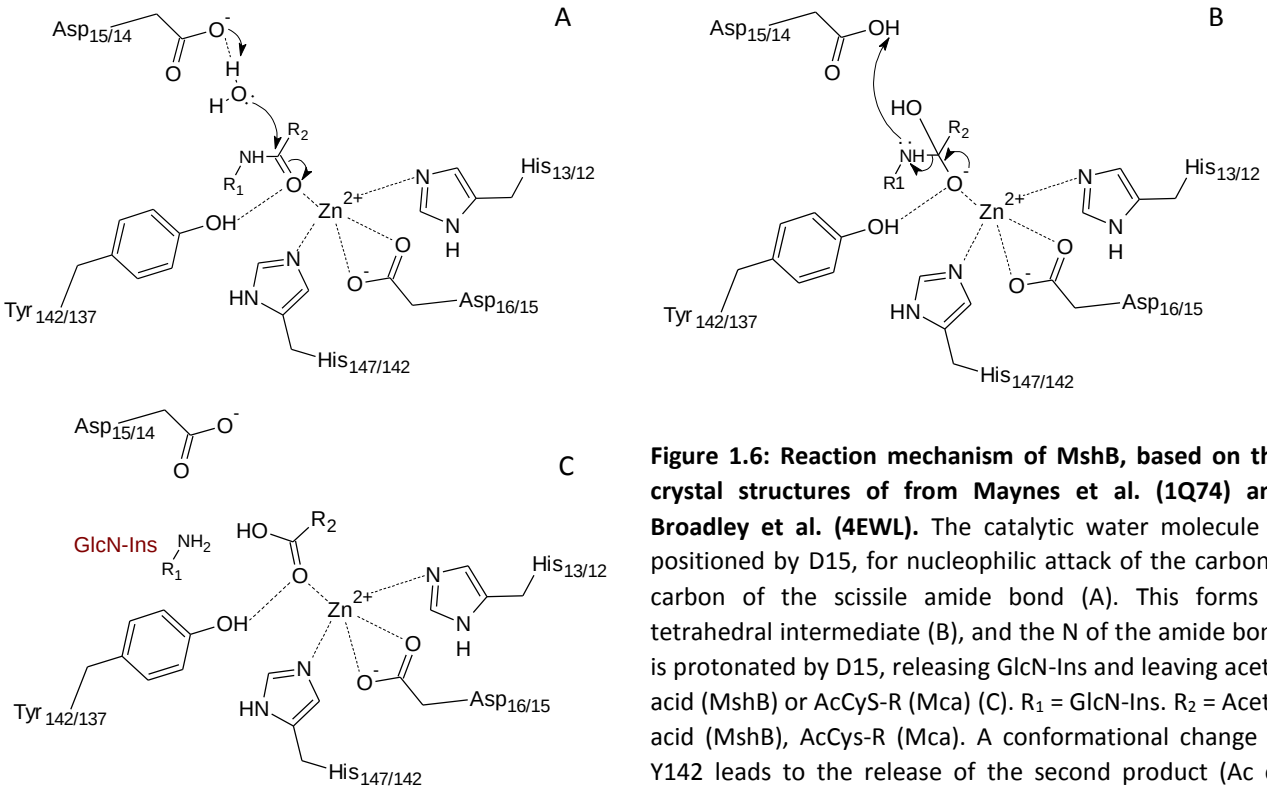


Figure 1.6: Reaction mechanism of MshB, based on the crystal structures of from Maynes et al. (1Q74) and Broadley et al. (4EWL). The catalytic water molecule is positioned by D15, for nucleophilic attack of the carbonyl carbon of the scissile amide bond (A). This forms a tetrahedral intermediate (B), and the N of the amide bond is protonated by D15, releasing GlcN-Ins and leaving acetic acid (MshB) or AcCyS-R (Mca) (C). R_1 = GlcN-Ins. R_2 = Acetic acid (MshB), AcCys-R (Mca). A conformational change in Y142 leads to the release of the second product (Ac or AcCys-R) [36]. Residue numbering is shown for MshB/Mca.

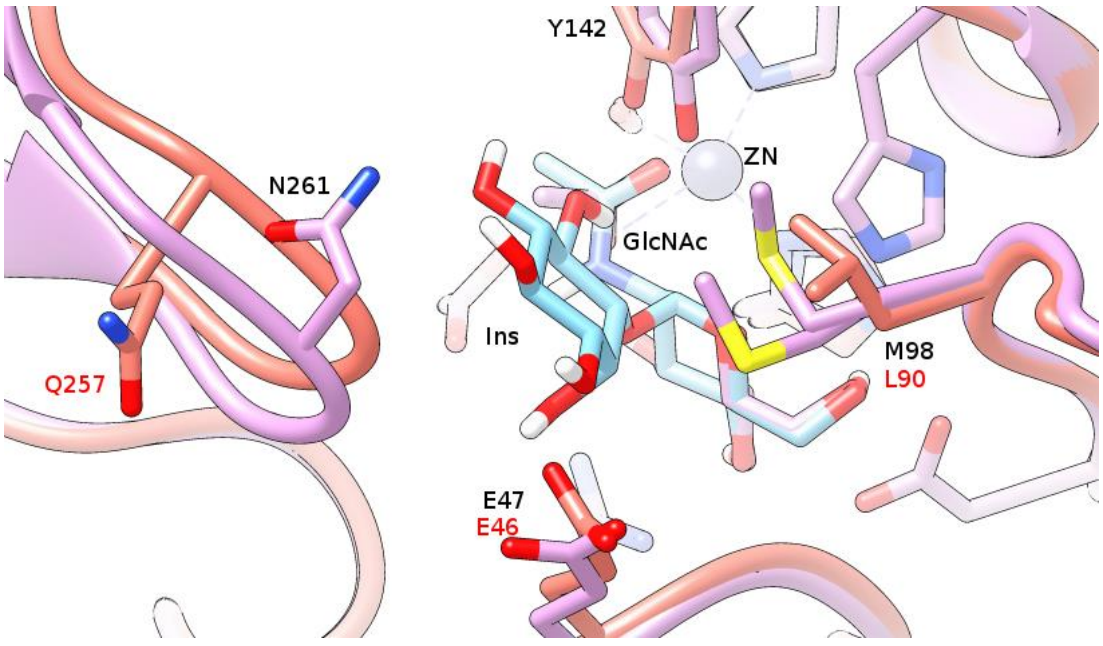


Figure 1.7: Ribbon diagram of the active site of MshB (magenta), superimposed with a homology model of Mca (red). GlcNAc-Ins (cyan) is modelled in the active site. Residues involved in the binding of inositol are labelled in black. Equivalent Mca residues are labelled in red. The residues involved in hydrogen bonding with Inositol are E47 and N261 (indirectly), and in Mca, E46 and Q257 are modelled in similar positions. The hydrophobic interaction of M98 with the inositol ring may be achieved by L90 in Mca.

1.11. Major differences between Mca and MshB

Despite the high degree of structural similarity between Mca and MshB in the active site and in other regions, the structural basis for different substrate specificity between Mca and MshB is not well understood. Several regions of low sequence conservation are highlighted for their possible roles in activity and specificity in Mca and MshB.

The first of these is in the active site itself, and consists of four residues, 19-22, LSNG in MshB and 18-21, SKGA in Mca. These lie near the start of the first α-helix, and L19 forms part of the entrance to a large hydrophobic pocket (figure 1.8), noted by McCarthy et al. 2004, when it was occupied by the octyl chain of a β-octylglucoside molecule. If transferred to Mca, this is likely the binding pocket of the MSR S-conjugate group. In Mca, L19 is replaced by S18. This likely differentiates between the substrates GlcNAc-Ins and MSR on the basis of their acyl groups (Ac- and AcCySR-, respectively). Particularly, concerning the AcCySR conjugate of MSR, L19 in MshB may cause steric hindrance with this group, while S18 in Mca may form favourable interactions with this group, and allow the S-conjugate group into the hydrophobic pocket, resulting in stabilization of the substrate.

Other regions of low conservation between Mca and MshB include the α-helix 187-196 (contributing to the roof of the hydrophobic pocket), the loop from 208-224 (of which 214-216 contribute to the front roof of the hydrophobic pocket), and the mixed loop and β-strand from 261-270 (forming the rear and side walls of the hydrophobic pocket) (figure 1.8). The environment in this pocket will dictate the types of S-conjugates accepted by Mca or MshB. The actual arrangement of residues in the pocket in Mca is not possible to model accurately, given the low degree of sequence conservation between Mca and MshB over these regions, without a structure of Mca or a close homologue (from which an accurate homology model could be built).

Minor regions of low conservation between Mca and MshB include α-helix 231-243, loop 249-255, and loop 279-288. These regions do not appear near to the active site or hydrophobic pocket, and their contribution to substrate specificity is likely minimal.

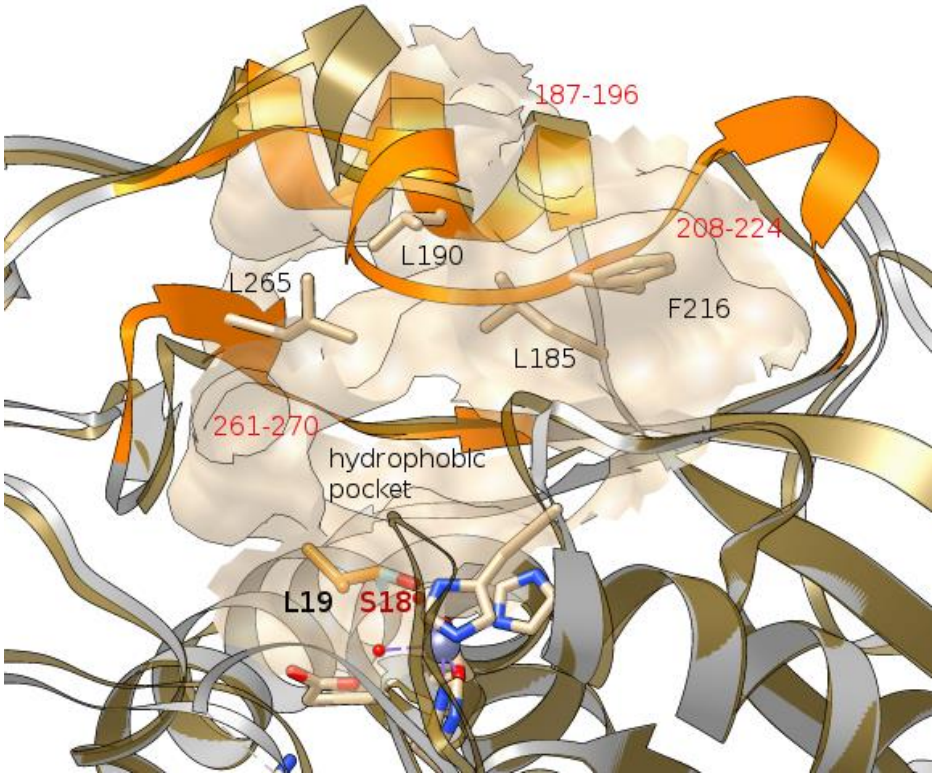


Figure 1.8 Hydrophobic pocket adjacent to the active site in MshB (PDB file 1Q74, grey ribbon). A homology model of Mca (brown ribbon) is superimposed upon MshB. The regions of low conservation between Mca and MshB which contribute to the hydrophobic pocket are coloured orange. Hydrophobic residues lining the hydrophobic pocket are represented as sticks. L19 (MshB) and S18 (Mca) are shown at the entrance to the hydrophobic pocket. L265 contributes to the left wall of the pocket, and L190 and L185 contribute to the roof of the pocket. F216 may also contribute to the roof of the pocket.

To date, MshB is the closest homologue to Mca for which there is a crystal structure available. For this reason, an attempt at determining the structure of Mca itself is quite reasonable. Mca is a small, globular protein [21, 33], which has been expressed recombinantly and purified, but which has not been successfully crystallized.

The aim of this research, therefore, is to determine the structure of Mca, from *M. tuberculosis* or the closely related *Mycobacterium smegmatis* by crystallization or other means. Additional means to obtain the structure of a protein are through nuclear-magnetic-resonance (NMR) spectroscopy and electron microscopy (EM). The use of NMR spectroscopy for protein structure determination is limited to proteins of a size of about 25kDa [40]. The size of Mca (32.6kDa), precludes this method. The use of EM in protein structure determination was limited to large complexes, and was not at the resolution obtained by crystallography or NMR spectroscopy. However, in recent years, the technology used in EM has advanced dramatically, so that near-atomic resolution is now attainable,

even for some small proteins, such as isocitrate dehydrogenase (3.8 Å) [41]. Until the present work, Mca had only been detected as a monomer [33], which is too small for EM-structure determination. Therefore, the only applicable method for structural determination of Mca at the start of this project was by crystallization and X-ray crystallography. However, in the course of the project, large polymers of Mca were found, and these were used in EM-based methods for structural determination. What follows is a description of the methods used to purify the protein (chapter 2), and the EM-based methods used for determination of its low resolution structure (chapter 3).

1.12. Aims of the project

The following aims were made for the project:

1. Recombinantly express and purify *M. tuberculosis* Mca and *M. smegmatis* Mca
2. Crystallize Mca using appropriate methods
3. Determine the crystal structure of Mca if diffracting crystals are grown
4. Use EM to determine the low-resolution structure of MsMca
    a. Prepare a number of negative-stain transmission electron micrographs of the MsMca filament
    b. Process the micrographs before using them in an appropriate algorithm for 3D reconstruction of the filament
    c. Refine the estimations of the helical symmetry of the filament
    d. Assess the quality of the reconstruction using appropriate methods
    e. Determine the handedness of the filament using SEM or suggest the handedness by fitting atomic models of MsMca into the filament reconstruction
    f. Interpret the interfaces between subunits by fitting an atomic model of MsMca into the electron density of the reconstruction

1.13. References

1. *WHO | Global Tuberculosis Report 2016*. World Health Organization; 2016.

2. World Health Organization: *WHO | Guidelines for Treatment of Tuberculosis*. 4th edition. World Health Organization; 2010.

3. WHO WHO: *Multidrug and Extensively Drug-Resistant TB (M/XDR-TB): 2010 Global Report on Surveillance and Response*. 2010.

4. Lechartier B, Rybniker J, Zumla A, Cole ST: **Tuberculosis drug discovery in the post-post-genomic era**. *EMBO Mol Med* 2014, **6**:n/a–n/a.

5. Cole ST, Riccardi G: **New tuberculosis drugs on the horizon**. *Curr Opin Microbiol* 2011, **14**:570–576.

6. Hartkoorn RC, Sala C, Neres J, Pojer F, Magnet S, Mukherjee R, Uplekar S, Boy-Röttger S, Altmann KH, Cole ST: **Towards a new tuberculosis drug: Pyridomycin - nature's isoniazid**. *EMBO Mol Med* 2012, **4**:1032–1042.

7. Hartkoorn RC, Pojer F, Read J a, Gingell H, Neres J, Horlacher OP, Altmann K-H, Cole ST: **Pyridomycin bridges the NADH- and substrate-binding pockets of the enoyl reductase InhA.** *Nat Chem Biol* 2014, **10**:96–8.

8. Flipo M, Desroses M, Lecat-Guillet N, Villemagne B, Blondiaux N, Leroux F, Piveteau C, Mathys V, Flament MP, Siepmann J, Villeret V, Wohlk??nig A, Wintjens R, Soror SH, Christophe T, Jeon HK, Locht C, Brodin P, D??prez B, Baulard AR, Willand N: **Ethionamide boosters. 2. Combining bioisosteric replacement and structure-based drug design to solve pharmacokinetic issues in a series of potent 1,2,4-oxadiazole EthR inhibitors**. *J Med Chem* 2012, **55**:68–83.

9. Willand N, Dirié B, Carette X, Bifani P, Singhal A, Desroses M, Leroux F, Willery E, Mathys V, Déprez-Poulain R, Delcroix G, Frénois F, Aumercier M, Locht C, Villeret V, Déprez B, Baulard AR: **Synthetic EthR inhibitors boost antituberculous activity of ethionamide.** *Nat Med* 2009, **15**:537–544.

10. Krieger I V., Freundlich JS, Gawandi VB, Roberts JP, Gawandi VB, Sun Q, Owen JL, Fraile MT, Huss SI, Lavandera JL, Ioerger TR, Sacchettini JC: **Structure-Guided Discovery of Phenyl-diketo Acids as Potent Inhibitors of M. Tuberculosis Malate Synthase**. *Chem Biol* 2012, **19**:1556–1567.

11. Deponte M: **Glutathione catalysis and the reaction mechanisms of glutathione-dependent enzymes.** *Biochim Biophys Acta* 2013, **1830**:3217–66.

12. Fahey RC: **Glutathione analogs in prokaryotes.** *Biochim Biophys Acta* 2013, **1830**:3182–98.

13. Waldron KJ, Robinson NJ: **How do bacterial cells ensure that metalloproteins get the correct metal?** *Nat Rev Microbiol* 2009, **7**:25–35.

14. Newton GL, Arnold K, Price MS, Sherrill C, Delcardayre SB, Aharonowitz Y, Cohen G, Davies J, Fahey RC, Davis C: **Distribution of thiols in microorganisms: mycothiol is a major thiol in most actinomycetes**. *J Bacteriol* 1996, **178**:1990–5.

15. Newton GL, Bewley CA, Dwyer TJ, Horn R, Aharonowitz Y, Cohen G, Davies J, Faulkner DJ, Fahey RC: **The structure of U17 isolated from Streptomyces clavuligerus and its properties as an antioxidant thiol.** *Eur J Biochem* 1995, **230**:821–5.

16. Sakuda S, Zhou ZY, Yamada Y: **Structure of a novel disulfide of 2-(N-acetylcysteinyl)amido-2-deoxy-alpha-D-glucopyranosyl-myo-inositol produced by Streptomyces sp.** *Biosci Biotechnol*

Biochem 1994, **58**:1347–8.

17. Spies HS, Steenkamp DJ: **Thiols of intracellular pathogens. Identification of ovothiol A in Leishmania donovani and structural analysis of a novel thiol from Mycobacterium bovis.** *Eur J Biochem* 1994, **224**:203–13.

18. Buchmeier NA, Newton GL, Koledin T, Fahey RC: **Association of mycothiol with protection of Mycobacterium tuberculosis from toxic oxidants and antibiotics**. *Mol Microbiol* 2003, **47**:1723–1732.

19. Rawat M, Johnson C, Cadiz V, Av-Gay Y: **Comparative analysis of mutants in the mycothiol biosynthesis pathway in Mycobacterium smegmatis.** *Biochem Biophys Res Commun* 2007, **363**:71–6.

20. Hernick M: **Targeting Mycothiol Biosynthesis and Mycothiol-Dependent Detoxification for the Treatment of Tuberculosis**. *J Anc Dis Prev Remedies* 2013, **1**:e105.

21. Newton GL, Av-Gay Y, Fahey RC: **A Novel Mycothiol-Dependent Detoxification Pathway in Mycobacteria Involving Mycothiol S-Conjugate Amidase**. *Biochemistry* 2000, **39**:10739–10746.

22. Sareen D, Steffek M, Newton GL, Fahey RC: **ATP-dependent L-cysteine: 1D-myo-inosityl 2-amino-2-deoxy-??-D-glucopyranoside ligase, mycothiol biosynthesis enzyme MshC, is related to class I cysteinyl-tRNA synthetases**. *Biochemistry* 2002, **41**:6885–6890.

23. Koledin T, Newton GL, Fahey RC: **Identification of the mycothiol synthase gene (mshD) encoding the acetyltransferase producing mycothiol in actinomycetes.** *Arch Microbiol* 2002, **178**:331–7.

24. Vilchèze C, Av-Gay Y, Attarian R, Liu Z, Hazbón MH, Colangeli R, Chen B, Liu W, Alland D, Sacchettini JC, Jacobs WR: **Mycothiol biosynthesis is essential for ethionamide susceptibility in Mycobacterium tuberculosis.** *Mol Microbiol* 2008, **69**:1316–29.

25. Movahedzadeh F, Wheeler PR, Dinadayala P, Av-Gay Y, Parish T, Daffé M, Stoker NG: **Inositol monophosphate phosphatase genes of Mycobacterium tuberculosis.** *BMC Microbiol* 2010, **10**:50.

26. Rawat M, Kovacevic S, Billman-Jacobe H, Av-Gay Y: **Inactivation of mshB, a key gene in the mycothiol biosynthesis pathway in Mycobacterium smegmatis**. *Microbiology* 2003, **149**:1341–1349.

27. Newton GL, Ko M, Ta P, Av-Gay Y, Fahey RC: **Purification and characterization of Mycobacterium tuberculosis 1D-myo-inosityl-2-acetamido-2-deoxy-alpha-D-glucopyranoside deacetylase, MshB, a mycothiol biosynthetic enzyme.** *Protein Expr Purif* 2006, **47**:542–50.

28. Buchmeier NA, Newton GL, Fahey RC: **A mycothiol synthase mutant of Mycobacterium tuberculosis has an altered thiol-disulfide content and limited tolerance to stress.** *J Bacteriol* 2006, **188**:6245–52.

29. Patel MP, Blanchard JS: **Expression, Purification, and Characterization of Mycobacterium tuberculosis Mycothione Reductase**. *Biochemistry* 1999, **38**:11827–11833.

30. Watermeyer ND: **Design and Synthesis of Potential Inhibitors of Enzymes Involved in the Biosynthesis and Utilisation of Mycothiol**. University of Cape Town; 2012.

31. Newton GL, Leung SS, Wakabayashi JI, Rawat M, Fahey RC: **The DinB superfamily includes novel mycothiol, bacillithiol, and glutathione S-transferases.** *Biochemistry* 2011, **50**:10751–60.

32. Newton GL, Buchmeier N, Fahey RC: **Biosynthesis and Functions of Mycothiol, the Unique Protective Thiol of Actinobacteria**. *Microbiol Mol Biol Rev* 2008, **72**:471–494.

33. Steffek M, Newton GL, Av-Gay Y, Fahey RC: **Characterization of Mycobacterium tuberculosis**

Mycothiol S-Conjugate Amidase†. *Biochemistry* 2003, **42**:12067–12076.

34. Nicholas GM, Newton GL, Fahey RC, Bewley CA: **Novel Bromotyrosine Alkaloids: Inhibitors of Mycothiol S -Conjugate Amidase**. *Org Lett* 2001, **3**:1543–1545.

35. Metaferia BB, Fetterolf BJ, Shazad-ul-Hussan S, Moravec M, Smith JA, Ray S, Gutierrez-Lugo MT, Bewley CA: **Synthesis of natural product-inspired inhibitors of Mycobacterium tuberculosis mycothiol-associated enzymes: The first inhibitors of GlcNAc-Ins deacetylase**. *J Med Chem* 2007, **50**:6326–6336.

36. Riordan SW, Field JJ, Corkran HM, Dasyam N, Stocker BL, Timmer MSM, Harvey JE, Teesdale-Spittle PH: **Synthesis of mycothiol conjugate analogues and evaluation of their antimycobacterial activity.** *Bioorg Med Chem Lett* 2015, **25**:2152–5.

37. Maynes JT, Garen C, Cherney MM, Newton G, Arad D, Av-Gay Y, Fahey RC, James MNG: **The crystal structure of 1-D-myo-inosityl 2-acetamido-2-deoxy-alpha-D-glucopyranoside deacetylase (MshB) from Mycobacterium tuberculosis reveals a zinc hydrolase with a lactate dehydrogenase fold.** *J Biol Chem* 2003, **278**:47166–70.

38. McCarthy AA, Peterson NA, Knijff R, Baker EN: **Crystal Structure of MshB from Mycobacterium tuberculosis, a Deacetylase Involved in Mycothiol Biosynthesis**. *J Mol Biol* 2004, **335**:1131–1141.

39. Broadley SG, Gumbart JC, Weber BW, Marakalala MJ, Steenkamp DJ, Sewell BT: **A new crystal form of MshB from Mycobacterium tuberculosis with glycerol and acetate in the active site suggests the catalytic mechanism**. *Acta Crystallogr D Biol Crystallogr* 2012, **68**:1450–1459.

40. Billeter M, Wagner G, Wüthrich K: **Solution NMR structure determination of proteins revisited.** *J Biomol NMR* 2008, **42**:155–158.

41. Merk A, Bartesaghi A, Banerjee S, Falconieri V, Rao P, Davis MI, Pragani R, Boxer MB, Earl LA, Milne JLS, Subramaniam S: **Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery**. *Cell* 2016, **165**:1698–1707.

2. Expression and Purification of Mca

Mca from *M. tuberculosis* and *M. smegmatis* were produced recombinantly in *E. coli* with an N-terminal His-maltose-binding protein (MBP) tag [1]. The fusion proteins were purified by means of Ni^{2+} affinity chromatography, then incubated with TEV protease to liberate Mca from the MBP fusion tag.

In the present study *M. tuberculosis* (Mt) and *M. smegmatis* (Ms) Mca were purified from *E. coli* BL21 cells, as in [1]. Purification included an additional size-exclusion chromatography step after each step of Ni^{2+} affinity chromatography.

2.1. Purification of *M. smegmatis* Mca and *M. tuberculosis* Mca

2.1.1. Construct information

The genes for *M. smegmatis* Mca and *M. tuberculosis* Mca were given to us in the form of two expression vectors, pVP56K and pVP55A, respectively. These were a gift, from M. Hernick, then at Virginia Tech University.

PVP55A provides the target protein with an N-terminal octahistidine (8xHis) tag, while from pVP56K, the inserted gene is expressed as a fusion protein, with an N-terminally linked maltose binding protein (MBP, from *E. coli*). MBP is produced with an N-terminal 8xHis-tag. The target protein may be cleaved from MBP after purification, via a specific Tobacco Etch Virus (TEV) protease cleavage site, within the linker region.

M. tuberculosis mca was given in the form of pVP55A, while *M. smegmatis mca* was given in the form of pVP56K.

2.1.2. Transformation of *E. coli* BL21 with pVP55A and pVP56K

M. tuberculosis mca (via pVP55A) and *M. smegmatis mca* (in pVP56K) were introduced to *E. coli* BL21 (DE3) by heat shock transformation. Briefly, chemically competent cells were mixed on ice with 1µL of DNA (pVP55A or pVP56K), kept on ice for 30 minutes, then placed in a water bath at 42°C for 30/45sec. Cells were kept on ice for 2 minutes. 500µL Lauria-bertani broth (LB) was added to the cells, which were then incubated at 37°C, with gentle shaking, for 45 minutes. 50µL of this was plated to nutrient agar plates, containing kanamycin (50µg/ml) or carbenicillin (100µg/ml), for cells transformed with pVP56K and pVP55A respectively. Plates were incubated overnight at 37°C.

Single colonies were selected from the plates for pVP56K and pVP55A, and inoculated to 10ml cultures of LB broth, containing kanamycin (50µg/ml) and carbenicillin (100µg/ml) respectively. Cultures were grown at 37°C overnight, with shaking at 120rpm. Cells from each overnight culture were mixed separately with equal volumes of 50% glycerol, flash frozen in liquid nitrogen, then stored at -80°C, until use.

2.1.3. Subcloning of *M. tuberculosis mca* to pVP56K

M. tuberculosis mca had been expressed from pVP55A, but no activity was observed in the soluble fraction of induced cells. We thought to improve the soluble yield of Mca, by expression in pVP56K as a fusion protein with His-MBP. *M. tuberculosis mca* was excised from pVP55A, using 5' *SgfI* and 3' *PmeI* restriction sites, and ligated to pVP56K (digested with *SgfI* and *PmeI, Flexi enzyme blend* (Promega)), using T4 DNA ligase (New England Biolabs). The ligation product was transformed to *E. coli* DH5-α and the successful insertion of *M. tuberculosis mca* into pVP56K was confirmed by DNA sequencing (CAF, Stellenbosch University). pVP56K:*mtmca* was transformed to *E. coli* BL21 (DE3) for protein expression.

2.1.4. Protein detection by SDS-PAGE

SDS - Poly Acrylamide Gel Electrophoresis (SDS-PAGE), was performed using standard methods [BioRad Mini-PROTEAN]. SDS-PAGE was used to detect the presence of His-MBP-Mca and Mca in samples taken during the purification steps. 10-12% Polyacrylamide gels were used. Electrophoresis was performed at 25mA constant current, for 1 hour. Gels were stained using Aquastain.

2.1.5. Cell culture

Glycerol stocks of *E. coli* BL21 transformed with pVP56K:*mtmca* or pVP56K:*msmca* were used to inoculate starter cultures of 10ml nutrient broth (Merck) with Kanamycin (50µg/ml). Cultures were grown overnight with shaking at 37°C. 1ml of saturated starter culture was inoculated to 1L nutrient broth (Merck) containing 50µg/ml Kanamycin, and grown at 37°C until cell density reached 0.5-0.8.

Cells were induced with the addition of IPTG (Isopropyl β-D-1-thiogalactopyranoside) to each culture (final concentration of 1mM), and cultures were incubated at 30°C overnight, with shaking at 120rpm. Cells were harvested by centrifugation at 6000g (12min, at 4°C). Cell pellets were stored at -20°C. Frozen pellets were thawed, and resuspended in lysis buffer (20mM KPi (Potassium phosphate) pH 7.4, 1x cOmplete™ Mini EDTA-free Protease Inhibitor Cocktail (Sigma), 150mM NaCl), using 20ml lysis buffer per 4-5g cell pellet.

Cells were sonicated using a mysonix sonicator, with a ½ inch horn, power setting of 5.0 (48W), for 7 minutes, in 15 second pulses, followed by 15 second intervals. Cells were kept in an ethanol-ice slurry during sonication. The total cell lysate was centrifuged at 20 000g, for 30 min (at 4°C).

The pellet was discarded, and the supernatant was sonicated briefly (½ inch flat horn, at 48W, for 2 minutes, in 15 second pulses, followed by 15 second intervals).

2.1.6. Isolation of His-MBP-Mca by Ni^{2+} Affinity Chromatography

The clarified sonicate of the induced cells was loaded to a Ni^{2+} HiTrap 5ml Column (GE Lifesciences) pre-equilibrated with buffer A (20mM KPi, 0.5M NaCl, 20mM imidazole, pH 7.8). The column was washed with 8 column volumes (CV) of buffer A then buffer B (20mM KPi, 0.5M NaCl, 500mM imidazole, pH 7.4) was applied in a linear gradient of 0-75% against buffer A over 20 CV. The column was washed with 5 CV of 100% buffer B, then re-equilibrated.

2.1.7. Size-Exclusion Chromatography of His-MBP-Mca

Ni^{2+} affinity chromatography fractions containing purified His-MBP-MsMca (determined by SDS-PAGE) were pooled, and buffer exchanged to 50mM Tris-HCl pH 8.0, using a HiPrep 26/10 Desalting Column (GE Lifesciences). Fractions were concentrated using an Amicon 15 ultrafiltration device (Merck) (10KMWCO), then purified by size exclusion chromatography, through a Gilson HPLC system, using a Superdex 200 HiLoad 16/600 Column (GE Lifesciences), equilibrated with 50mM Tris-HCl, pH 8.0, at a flow rate of 1ml/min. The column was pre-calibrated per the manufacturer's instructions, using the BioRad Gel Filtration Standard.

Fractions containing His-MBP-Mca (determined by SDS-PAGE) were pooled, and incubated with His-tagged TEV (His-TEV) protease, at a ratio of 1mg His-TEV protease: 50-100mg His-MBP-Mca, for 2.5 hours, at 15°C, with gentle shaking. His-TEV protease had been purified from *E. coli* BL21 transformed with pRK793, a gift from David Waugh (Addgene plasmid # 8827) [2].

2.1.8. Ni^{2+} Affinity Chromatography for isolation of free Mca and final Size Exclusion Chromatography

Free MsMca was separated from His-MBP by Ni^{2+} affinity chromatography. Sample containing MsMca, His-MBP and His-TEV protease was loaded to the Ni^{2+} HiTrap 5ml Column (GE Lifesciences) pre-equilibrated with buffer A, at a flow rate of 1ml/min. Following sample loading, buffer A was applied for 10 CV, then buffer B was applied in a linear gradient of 0-75%, over 10 CV. Fractions containing purified MsMca (determined by SDS-PAGE) were pooled, concentrated and loaded to a

Superdex 200 HiLoad 16/600 Column (GE Lifesciences), at a flow rate of 1ml/min. 50mM Tris-HCl pH 8.0 was used as the running buffer.

Peak fractions containing free MsMca (determined by SDS-PAGE) were pooled, concentrated, and stored at 4°C, or used in crystallization experiments.

2.1.9. Protein Quantitation

Protein concentration was estimated from UV absorbance at 280nm. Typically, a small sample was quantified by Nanodrop, at 280nm, using extinction coefficients of 1.465 M^{-1} cm^{-1} and 1.364 M^{-1} cm^{-1} for His-MBP-Mca and free Mca respectively [see appendix].

2.1.10. Activity Assays

Mca activity assays were performed to determine the conversion of the hydroxynaphthyldisulfide derivative of mycothiol, 2-S-(mycothiolyl)-6-hydroxynaphthyldisulfide (MSSNaph) by Mca. The assay was adapted from [3]. Specific activities were calculated as the number of nmoles MSSNaph consumed per minute per mg protein (nmol.min^{-1}.mg $protein^{-1}$).

Reactions were prepared in 320µL, containing 50mM HEPES pH 7.5, 125µM MSSNaph, and 5µL of the enzyme sample, added last to initiate the reaction. Reactions were performed at 37°C in a water bath. At 0, 5 and 20 minutes, 80µL was removed from each reaction and mixed with 20µL of 100% acetonitrile and 20µL of 5% trichloroacetic acid (TCA). Samples were heated at 65°C for 5 minutes, and kept on ice until being centrifuged at 16 000g for 4 minutes. 100µL of the supernatant was transferred to a fresh vial and analysed via HPLC, using a Phenomenex C18 column (Luna 5u C18(2) 100A), measuring UV absorbance at 280nm.

	0 min	**5 min**	**20 min**	**25 min**	**30 min**	**35 min**
BUFFER A (%)	75%	75%	55%	10%	10%	75%
BUFFER B (%)	25%	25%	45%	90%	90%	25%

The following method was used:

Buffer A was 0.1% TCA (filtered through a 0.22µm filter), and Buffer B was 100% acetonitrile (HPLC grade). The column was equilibrated for 5 minutes between runs, with 75% Buffer A and 25% Buffer B. The injection volume was 50µL, and the flow rate was 0.8ml/min.

Changes in substrate peak areas between time points were converted to changes in substrate concentration, and expressed as Δ[S] (nmol)/Δt (min). Where possible, activities were normalized against total protein (mg) present in a reaction to give specific activities (nmol.min^{-1}.mg protein^{-1}).

2.2. Results

2.2.1. Expression and purification

MtMca and MsMca were expressed solubly from pVP56K. A large band at ~70kDa (see figure 2.2.1), not present in the uninduced cell fraction, but appearing in the induced cell fraction and remaining in the clarified sonicate, is consistent with the expected position of the fusion protein, His-MBP-Mca.

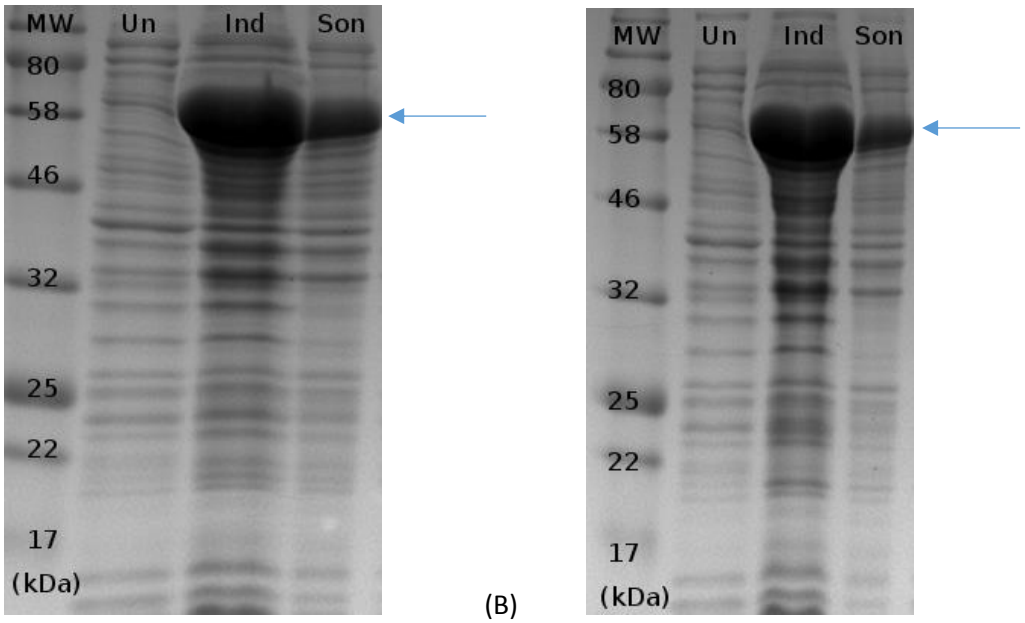
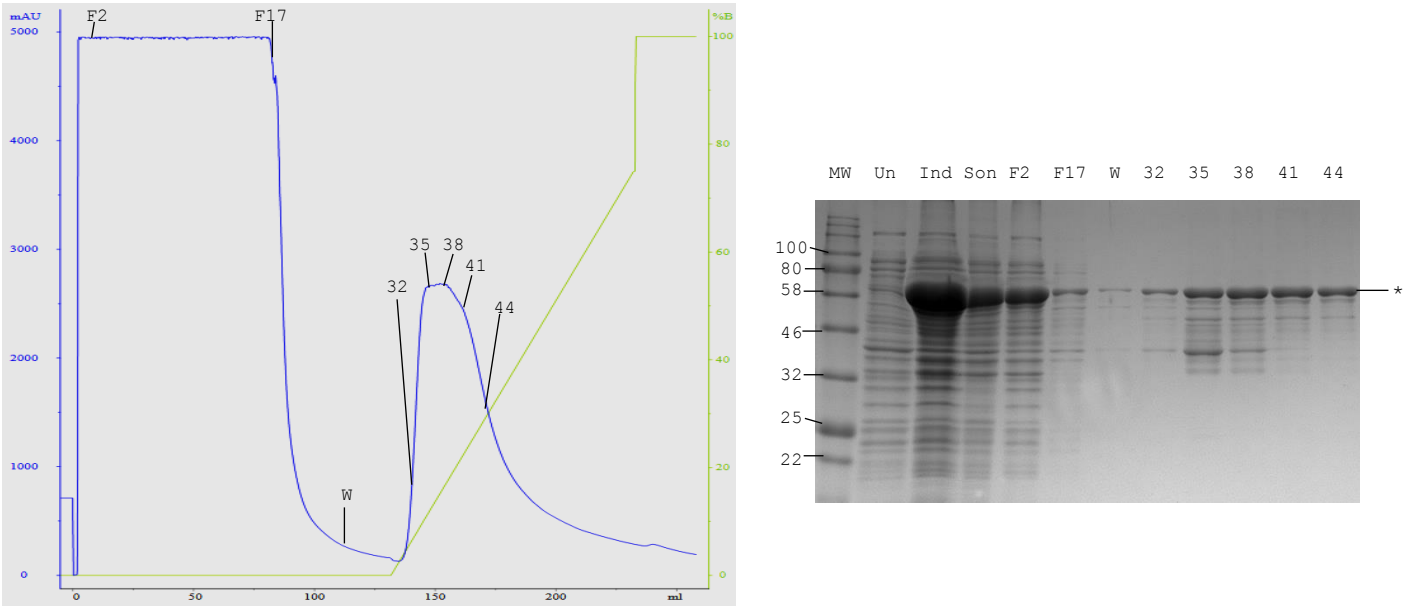


Figure 2.2.1. 12% SDS PAGE showing expression of (A) MtMca or (B) MsMca respectively, from pVP56K in *E. coli* BL21 (DE3). MW – Prestained Protein ladder (New England Biolabs). Un – Uninduced cell fraction. Ind – Induced cell lysate. Son – Clarified sonicate. The broad band at ~70kDa is indicated by an arrow. His-MBP-Mca has an expected weight of 77kDa.

2.2.2. Ni^{2+} Immobilized Metal Affinity Chromatography

2.2.2.1. His-MBP-MtMca



(A) (B)

Figure 2.2.2. Ni^{2+} affinity chromatography purification of His-MBP-MtMca from clarified sonicate. (A) Elution profile showing A_{280} (mAU) versus eluted volume (ml). Buffer B % is in green. Fractions sampled for SDS-PAGE are labelled. (B) 12% SDS-PAGE analysis of samples taken from cell culture and the Ni^{2+} affinity purification. MW – Molecular weight marker (New England Biolabs), Un – Uninduced cells, Ind – Induced cells, Son – Clarified sonicate, Lanes 5-12 – fractions from Ni^{2+} affinity chromatography (A).

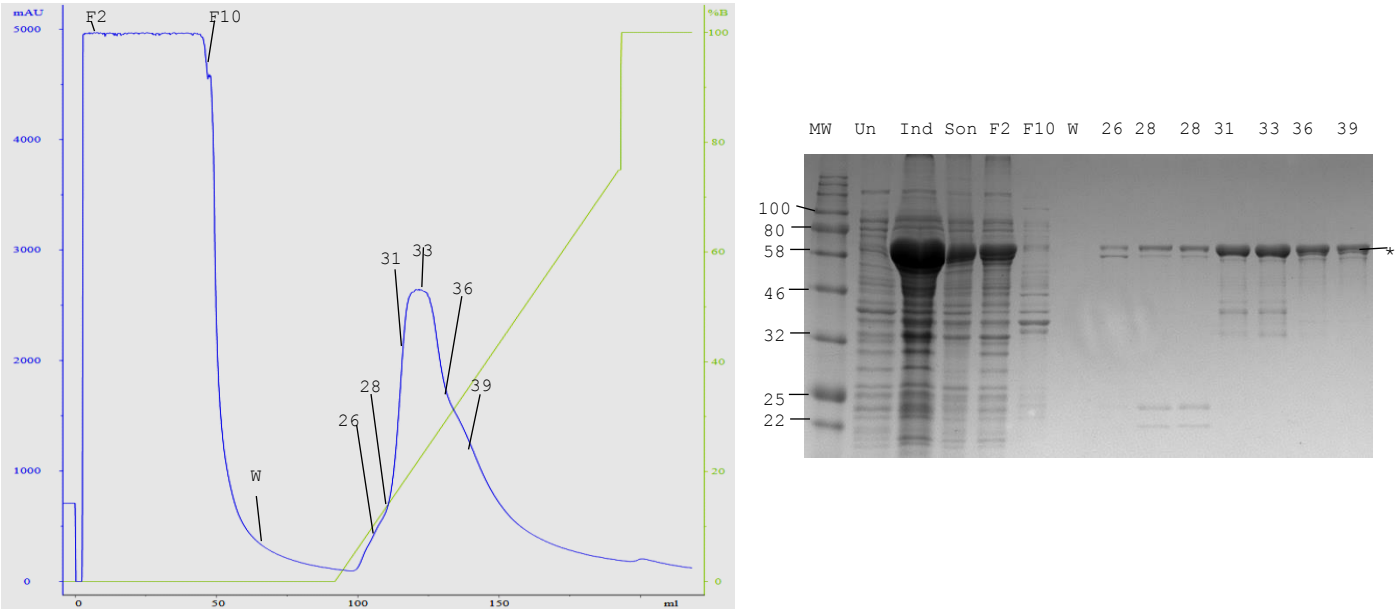
His-MBP-MtMca was successfully purified from the host lysate by Ni^{2+} affinity chromatography (figure 2.2.2). The major band in fractions 32-44 from the eluted protein (figure 2.2.2.B, lanes 8-12) appears only slightly before the expected weight (77kDa) of the fusion protein.

Fractions 38 and 41 were tested for activity. After 5 and 20 minutes, no product formation was observed (see figure 2.2.4). Even after an extended reaction of 60 minutes, negligible product formation was seen (see figure 2.2.4). His-MBP-MtMca therefore, appeared inactive.

The eluent fractions containing His-MBP-MtMca (determined by SDS-PAGE) were pooled, and 5ml was applied directly to the Superdex 200 HiPrep 16/600 column for size-exclusion chromatography. The results are seen in figure 2.2.5. Almost all of the protein in the sample appeared at void volume, indicating aggregation of His-MBP-MtMca. His-MBP-MtMca was not purified further.

The remaining work concerns the purification and characterization of His-MBP-MsMca.

2.2.3. His-MBP-MsMca



(A) (B)

Figure 2.2.3. Ni^{2+} affinity chromatography purification of His-MBP-MsMca from clarified sonicate. (A) Elution profile showing A_{280} (mAU) versus eluted volume (ml). Buffer B % is in green. Fractions sampled for SDS PAGE are labelled. (B) 12% SDS-PAGE analysis of samples taken from cell culture and the Ni^{2+} affinity purification. MW – Molecular weight marker (New England Biolabs), Un – Uninduced cells, Ind – Induced cells, Son – Clarified sonicate, Lanes 5-14 – fractions from Ni^{2+} affinity chromatography (A).

His-MBP-MsMca was also successfully purified from the host lysate by Ni^{2+} affinity chromatography (figure 2.2.3). The major band in fractions 26-39 from the eluted protein (figure 2.2.3.B, lanes 8-14) appears slightly before the expected weight (77kDa) of the fusion protein.

Fractions 33 and 39 both showed strong specific activity for Mca. HPLC analysis of a reaction prepared using fraction 39 and sampled at 5 minutes can be seen in figure 2.2.4, frame 6.

This activity was observed before TEV protease-cleavage of the fusion protein, indicating that MsMca is active with the His-MBP tag.

2.2.4. Activity assays for Mca

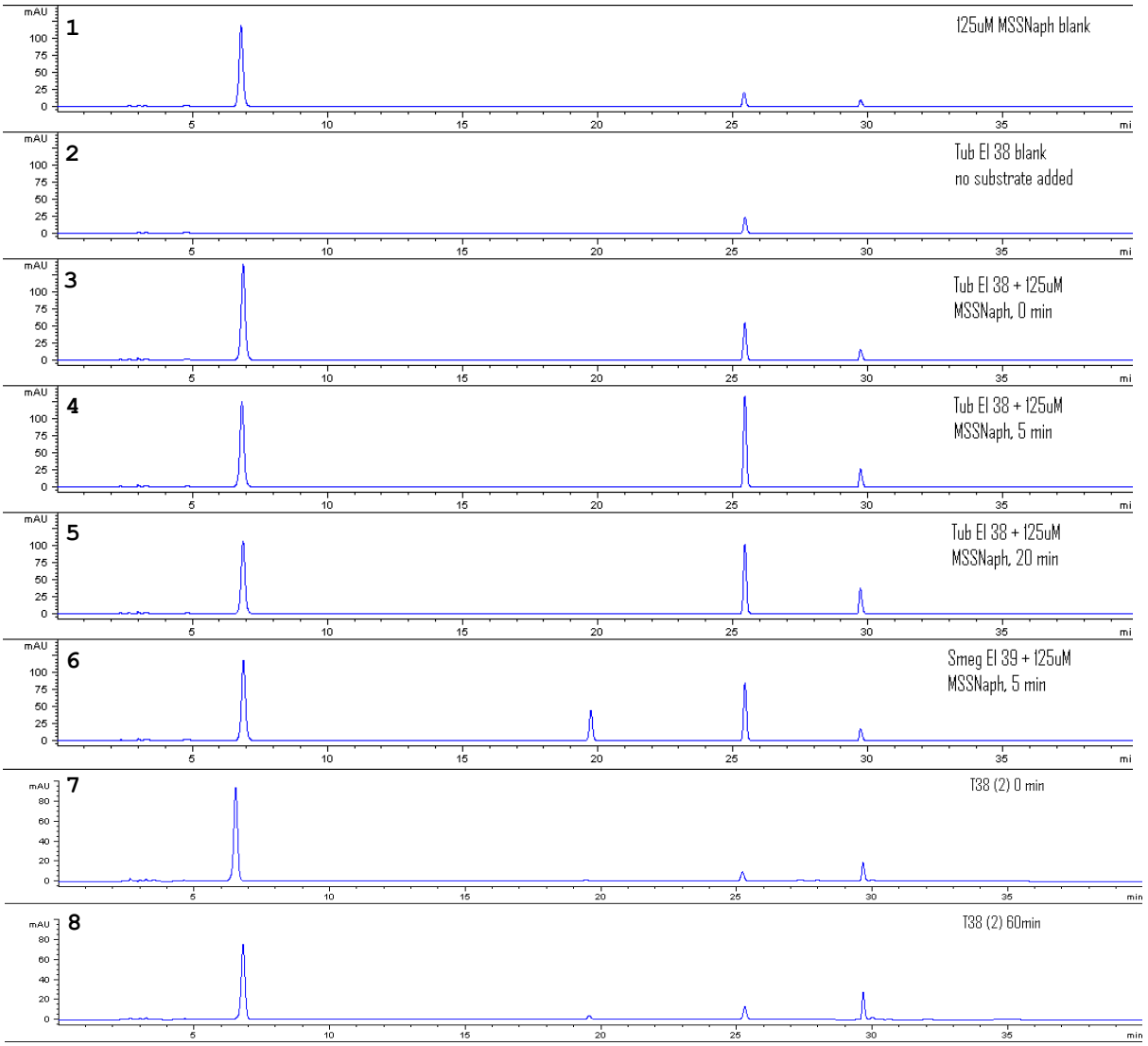


Figure 2.2.4. HPLC elution profiles of samples taken at specified intervals from Mca activity assays with MSSNaph as the substrate. 1 – Substrate blank. 2 – Enzyme blank (fraction 38, or f38, Ni^{2+} affinity purification). 3 – 5 - Reaction of 125µM MSSNaph with His-MBP-MtMca (fraction 38, shown in figure 2.2.2) sampled at 0, 5 and 20 minutes respectively. 6 – Reaction of 125µM MSSNaph with His-MBP-MsMca (fraction 39, shown in figure 2.2.3) sampled at 5 minutes. 7, 8 – Repeated reactions of 125µM MSSNaph with His-MBP-MtMca (f38) sampled at 0 and 60 minutes respectively. MSSNaph and the product AcCySSNaph appear at 6.7 and 19.6 minutes, respectively. No peak is seen at 19.6 minutes in the reaction with His-MBP-MtMca (frames 3 – 5, 7 and 8). Absorbance at 280nm (mAU) was measured versus time of elution (min). HPLC method 1 (see 2.2.9) was used.

2.2.5. Size exclusion chromatography

2.2.5.1. His-MBP-MtMca

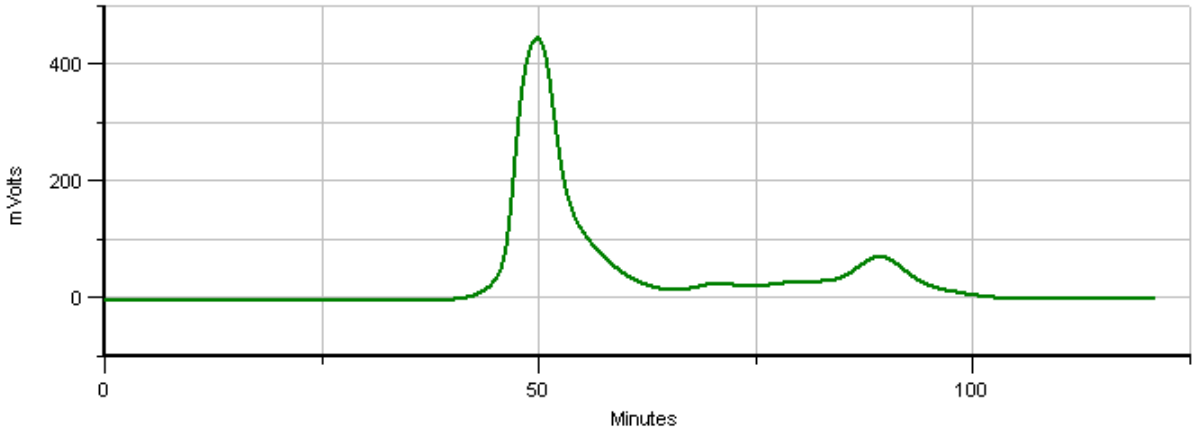


Figure 2.2.5. Size exclusion chromatography (Superdex 200) elution profile of Ni^{2+} affinity-purified His-MBP-MtMca. A280nm (mVolts) is shown versus the time of elution (minutes). The predominant peak is at void volume. Minor peaks at 70.5 and 89.2 minutes correspond to M_r of 116 000 and 29 000 respectively.

2.2.5.2. His-MBP-MsMca

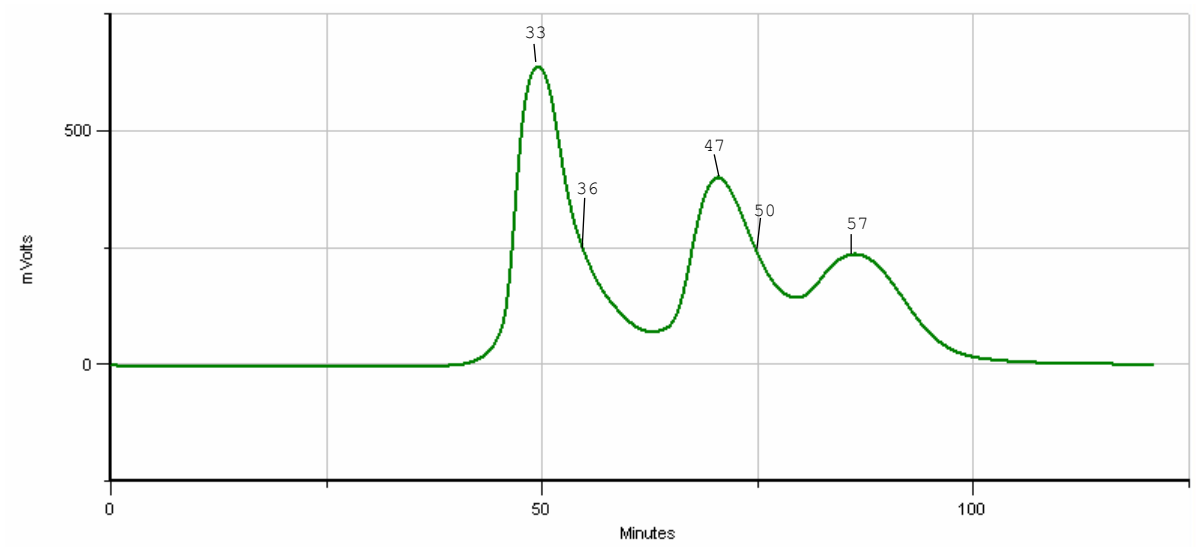


Figure 2.2.6. Size exclusion chromatography (Superdex 200) elution profile of Ni^{2+} affinity-purified His-MBP-MsMca. A280nm (mVolts) is shown versus the time of elution (minutes). The largest peak is at void volume (49.5 minutes). Major peaks are at 70.6 and 86.5 minutes, corresponding to M_r of 115 000 and 35 000 respectively. Fractions analysed by SDS-PAGE are labelled.

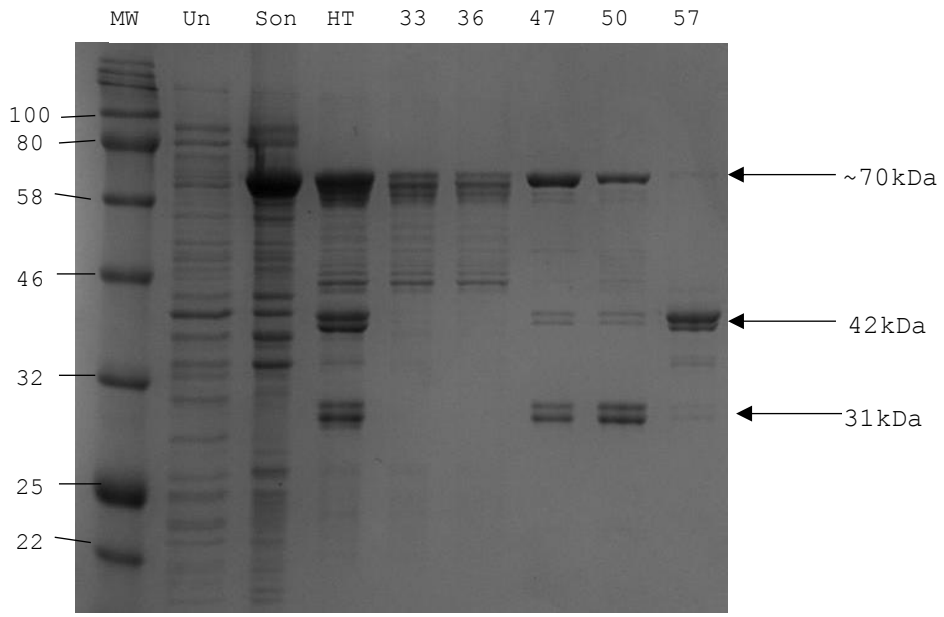


Figure 2.2.7. 12% SDS-PAGE analysis of Superdex 200 purification of His-MBP-MsMca. MW – Molecular weight marker (New England Biolabs), Un – Uninduced cells, Son – Clarified sonicate, HT – pooled fractions 29-45 from Ni^{2+} IMAC. Lanes 5-9 – fractions from S200 purification of the pooled sample (see figure 2.2.6). The positions of major bands described in the text are indicated by arrows.

Purification by Superdex 200 of His-MBP-MsMca obtained from Ni^{2+} affinity chromatography yielded three distinct peaks (figure 2.2.6). The first peak appeared at void volume, indicating aggregation of the His-MBP-MsMca and other contaminants (figure 2.2.7, lane 5). The second peak had an M_r of 115 000, and the two major proteins (of ~70kDa and 31kDa by SDS-PAGE (lane 7, figure 2.2.7)), indicating a complex of His-MBP-MsMca and a smaller protein, possibly MsMca (expected MW of 33kDa). The third peak had an M_r of 35 000, and SDS PAGE shows a major band at ~42kDa (lane 9, figure 2.2.7), consistent with the expected size for His-MBP (MW of 44kDa).

Fractions 33, 47 and 57 were tested for activity. Reactions were in duplicate, and contained 500ng protein per reaction, in a total volume of 350µL. Fraction 33 displayed no activity. Fractions 47 and 57 had specific activities of 7440 ± 219 and 1970 ± 28 $nmol.min^{-1}.mg\ protein^{-1}$ respectively (given with one standard deviation of the mean).

Fractions from the second peak were pooled and incubated at 15°C with His-TEV-protease as described. Figure 2.2.8, lane 5, shows that complete conversion of His-MBP-MsMca was achieved, with bands at 31 and 42kDa indicating MsMca and His-MBP respectively.

The products were separated using a Ni^{2+} affinity column, where MsMca appeared in the flow through (FT), and His-MBP in the eluate fraction (seen in figure 2.2.8, lanes 6-7 and 8, respectively).

FT fractions containing MsMca (determined by SDS-PAGE) were pooled and concentrated, then loaded to a superdex 200 HiPrep 16/600 column for size-exclusion chromatography. Free MsMca eluted as a single, symmetrical peak with an M_r of 83 000 (figure 2.2.9 A), indicating that MsMca is a dimer under these conditions. Previous purification of Mca from *M. smegmatis* cell extracts [4] under different running conditions for SEC showed Mca to be a monomer.

Fractions 48-53 (figure 2.2.9 B) were pooled and concentrated to ~19.5mg/ml.

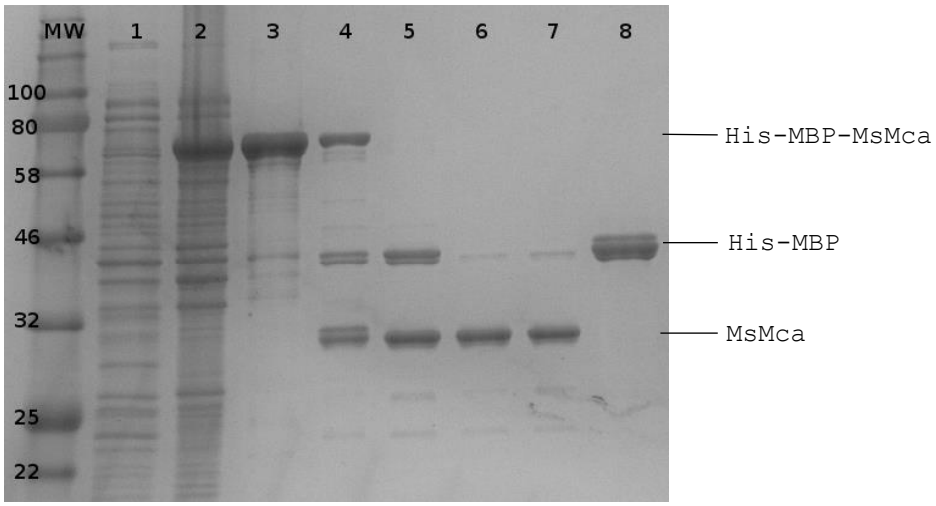


Figure 2.2.8. 12% SDS PAGE summary of His-MBP-MsMca purification, and TEV-protease mediated separation of His-MBP and MsMca. MW – Molecular weight marker (New England Biolabs), 1 – Uninduced cells, 2 – Clarified sonicate, 3 – pooled fractions 29-45 from Ni^{2+} IMAC, 4 – Pooled fractions from Superdex 200 SEC, 5 – (4) after incubation with TEV protease, 6-8 – Ni^{2+} affinity chromatography of (5), 6 and 7 - flow through, 8 – eluate. Molecular weights are given in kDa.

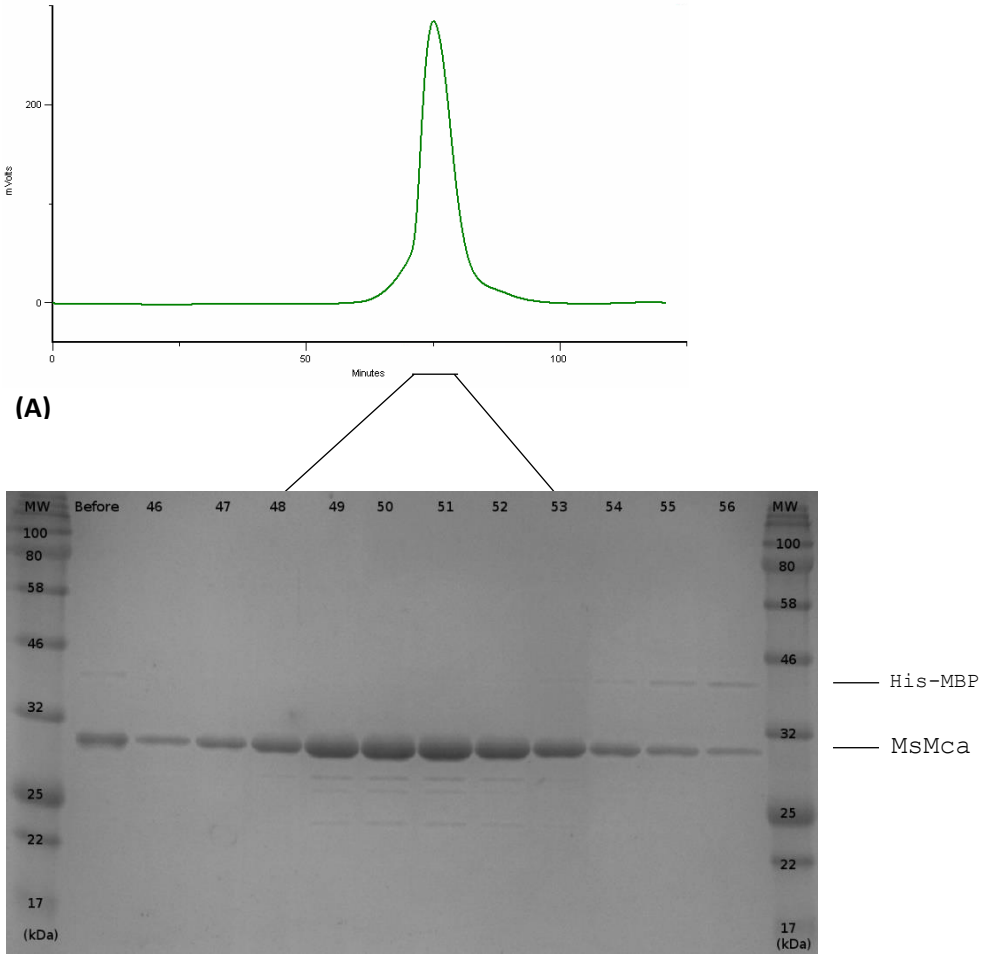


Figure 2.2.9. Size exclusion chromatography of purified MsMca. (A) Elution profile of protein absorbance (mVolts) at 280nm versus time of elution (minutes). The single symmetrical peak has a M_r of 83 000, consistent with an MsMca dimer. The positions of fractions 48-53 are indicated by a solid line. **(B) 12% SDS PAGE of peak fractions from (A).** MW – Molecular weight marker (NEB). Before – pooled flow through fractions from Ni^{2+} affinity chromatography. 46-56 – peak fractions obtained from SEC. Gel was stained for longer than usual, with Aquastain. Minor contaminants remain, of 24, 26 and 27 kDa. Fractions 48-53 were pooled and concentrated to 19.5 mg/ml.

2.3. Discussion

MtMca and MsMca were successfully overexpressed from *E. coli* BL21 (DE3) with His-MBP N-terminal tags. The His-tags were accessible, as seen by the large quantities of Ms and Mt Mca fusion proteins that were purified by Ni^{2+} affinity chromatography. Despite large soluble yields of the fusion proteins, only His-MBP-MsMca was active. His-MBP-MtMca, although soluble, and purified by Ni^{2+} affinity chromatography, was inactive and formed large aggregates, which were detected by size exclusion chromatography.

MtMca had been sub cloned from pVP55A to the pVP56K vector, using the flanking 5' and 3' restriction enzyme sites *SgfI* and *PmeI* respectively. Sequencing over the length of the *M. tuberculosis mca* gene suggested that no mutations were present. However, sequence data was only available for the *mca* gene and part of the linker sequence, whereas the upstream *mbp* gene was not checked. Frame-shift mutations here would cause expression of a completely different polypeptide. To verify this, sequencing over the entire *his-mbp-linker-mtmca* insert should be performed.

MsMca was successfully purified, to >99% purity. The cleavage of the fusion protein was achieved successfully by TEV-protease, indicating that the cleavage site was accessible. Conversely, mutants generated from the MtMca fusion protein were not cleaved by TEV-protease (results not shown). The presence of MBP-MtMca in the void volume following Ni^{2+} affinity chromatography, then size-exclusion chromatography, indicates possible misfolding of MBP-MtMca. This misfolding may cause the TEV-protease cleavage site to be made inaccessible, preventing cleavage of the fusion protein.

The purification of MsMca was performed successfully, and the final size-exclusion chromatography step of the free MsMca, revealed MsMca to be a dimer. Previously, MsMca was detected as a monomer, but the running conditions of the size-exclusion chromatography were different to those in the present study, which may be the cause of the dimer formation.

Crystallization experiments require protein solutions of 95% or greater purity, so MsMca is at sufficient purity for use in crystallization experiments. The high specific activity of MsMca is also an indicator of correct protein folding, which is required for meaningful crystallization and structure determination.

The purification of MsMca was rapid, and gave high yields of active protein. The purification steps followed may be used in further experiments for structural or other characterization of MsMca.

2.4. References

1. Kocabas E, Liu H, Hernick M: **Identity of cofactor bound to mycothiol conjugate amidase (Mca) influenced by expression and purification conditions**. *BioMetals* 2015, **28**:755–763.

2. Kapust RB, Tozser J, Fox JD, Anderson DE, Cherry S, Copeland TD, Waugh DS: **Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency**. *Protein Eng Des Sel* 2001, **14**:993–1000.

3. Marakalala MJ: **Inhibition of a Mycothiol Biosynthetic Enzyme and a Detoxification Enzyme as Anti-tubercular Drug Targets**. University of Cape Town; 2008.

4. Newton GL, Av-Gay Y, Fahey RC: **A Novel Mycothiol-Dependent Detoxification Pathway in Mycobacteria Involving Mycothiol S-Conjugate Amidase**. *Biochemistry* 2000, **39**:10739–10746.

3. Helical filaments and their 3D reconstruction

The following chapters give an account of the helical reconstruction of hollow filaments formed by MsMca under conditions of high salt concentrations. Chapter 3 begins with the theory of helical structures, their symmetry and the effect of their symmetry on helical three-dimensional reconstructions. Chapter 4 describes the refinement of the helical symmetry of MsMca filaments and their reconstruction.

3.1. The helix

A helix may be described in Cartesian co-ordinates (x, y, z) by three equations:

$$x = r\cos\varphi$$

$$y = r\sin\varphi$$

$$z = z$$

A more convenient description can be obtained using cylindrical co-ordinates, (r, φ, z), where *r* is the radius of the helix, *z* is the translation along the helical axis, and $\varphi = 2\pi\frac{z}{P}$, and P is the pitch of the helix (see figure 3.1.1 A).

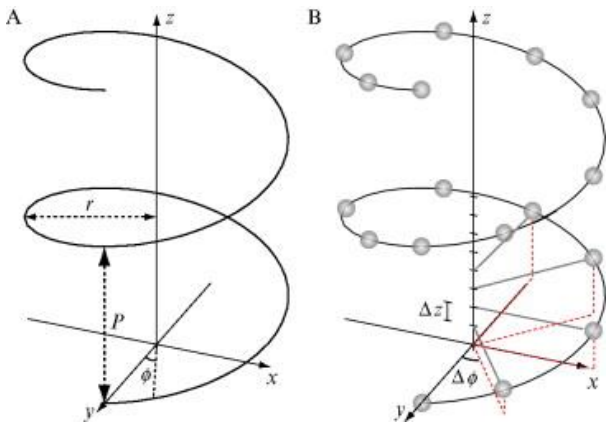


Figure 3.1.1. Helix geometry. (A) A continuous helix, described by the pitch (P) and radius (r) adopted by the spiral. A Cartesian coordinate system (x, y, z) or cylindrical coordinates (r, ϕ, z) can be used. (B) A discontinuous helix, composed of identical subunits arranged along the path of a continuous helix. The incremental translation and rotation per subunit are given by Δz and Δϕ. The above helix contains eight subunits per turn, i.e. Δϕ = 45°. Figure reproduced from [1].

A biological helix is rarely continuous, but is built from a discrete array of subunits (see figure 3.1.1. B), positioned at repeated intervals along the helical path. There are several ways to describe a biological helix. The first, is through the angular interval ($\Delta\varphi$) and axial interval (Δz) between subunits.

A helix may be represented by a helical net which is a 2D radial projection of the helix (see figure 3.1.2). The helical net reveals the basic arrangement of subunits on the helix. A number of helical families can also be represented on the helical net. These families of helices are visible in the helical lattice, as a result of the discontinuous helix, not the continuous helix, which has a single helical family.

In figure 3.1.2, three families are shown, (blue, red and green). The number of times a family of helices crosses the equatorial vector (dashed horizontal line) represents the *startedness*, or *order* (n) of the helical family.

3.2. The Fourier Transform of a helix

The helix is represented by a Fourier transform that has a number of discrete horizontal lines, called 'layer lines'. These layer lines are similar to the discrete reflections seen in the Fourier transform of a two-dimensional lattice.

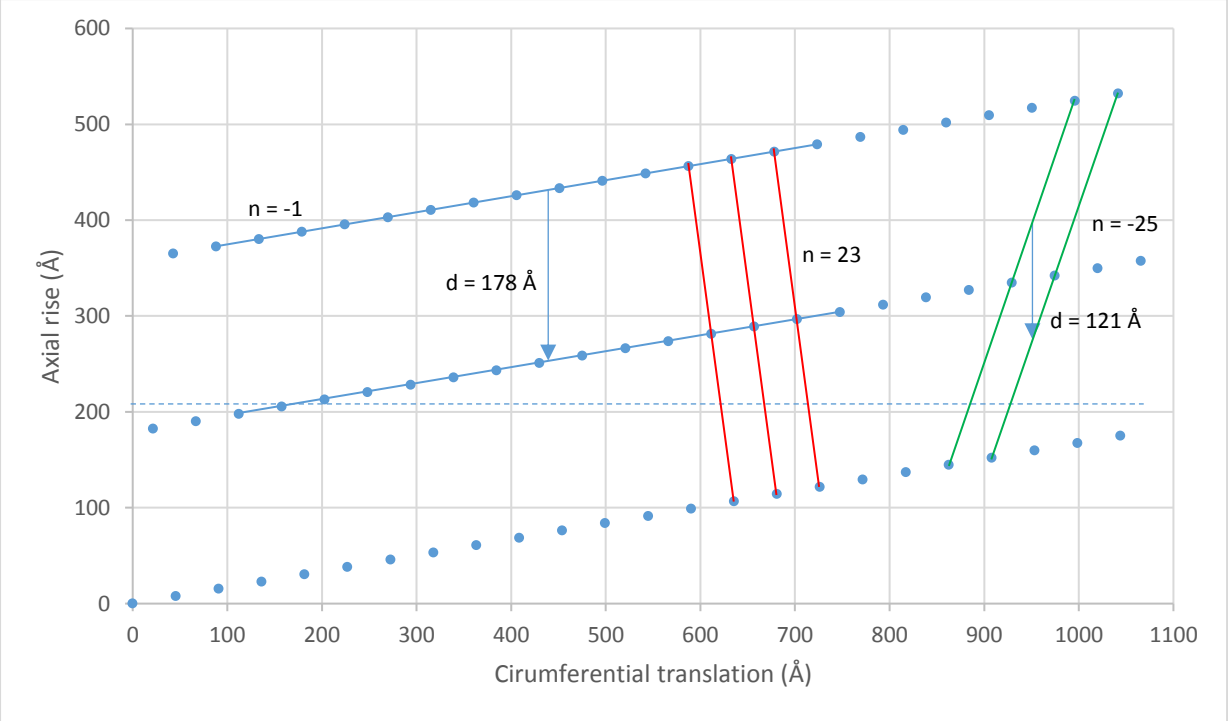
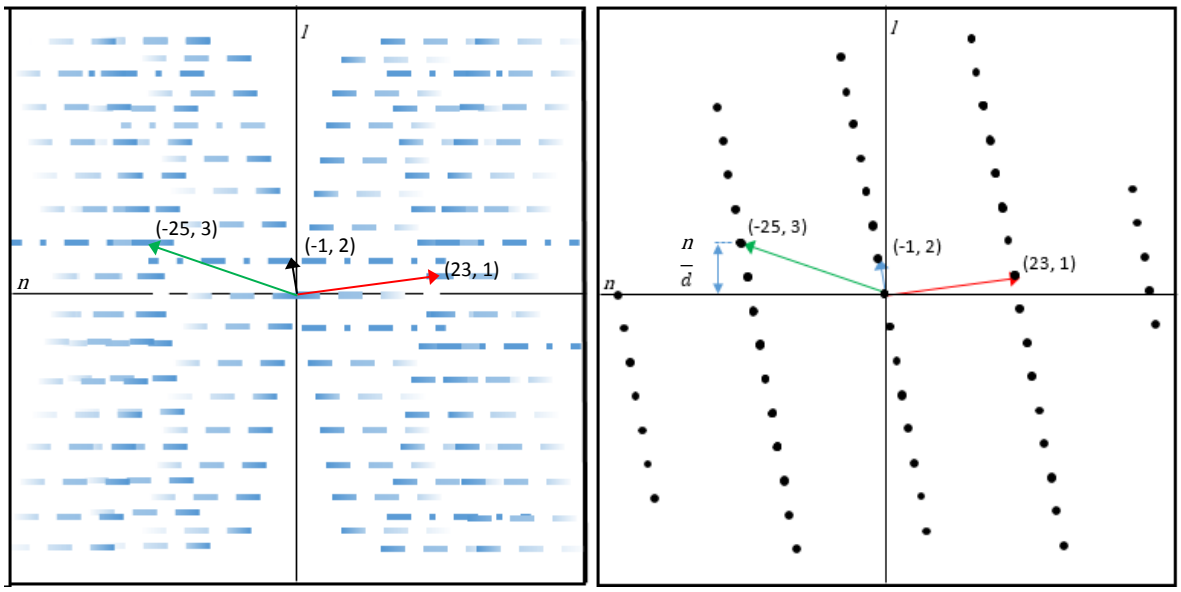


Figure 3.1.2. Helical net of a helical assembly. Helical parameters Δz 7.6 Å and $\Delta\phi$ 15.3°. Axial rise (Å) is shown versus the circumferential translation (Å). The pitch (d) order (n) of several helical families are shown.



(A) (B)

Figure 3.1.3: (A) Fourier transform of the helix described in 3.1.2. Each reflection is convoluted by an oscillating function, the Bessel function, with order n, to form a layer line. Each layer line corresponds to a helical family described in 3.1.2, and has a vertical position n/d, where d is the pitch of the helical family, and n is the order of the helical family. (B) Reciprocal lattice of the helical net shown in 3.1.2.

The differences arise from the fact that a projection of a helix shows both front and rear helical lattices, and the cylindrical nature of the helix affects the Fourier transform, such that reciprocal lattice reflections are extended horizontally to form layer lines.

The Fourier transform of a helix is non-zero only on layer lines, on the Z-axis where $Z = l/c$, where l (the layer line number) is an integer. c is the distance along the z-axis required for the helical object to repeat. For a helical object, the number of subunits in a helical repeat (u) and the number of turns in the same repeat (t) are related to the layer line number by the *selection rule* [2]:

$$l = u.m + t.n$$

n is the order of the helix, and m is an integer.

In 3.1.2, the helix n = 1, there are 47 subunits in 2 turns. The repeat distance is twice the pitch of a single turn. The selection rule is then:

$$l = 47.m + 2.n$$

The selection rule may then be used to determine the positions of other layer lines, and their respective orders.

3.3. 3D reconstruction of the helix

Once the helical symmetry of the object is known, this can aid in the reconstruction of the object. A popular method was derived from single-reconstruction techniques, and is called the Iterative Helical Real-Space Reconstruction (IHRSR) method [3]. The method generates models of the helical structure, which are refined over several cycles, as are the helical symmetry parameters, Δz and Δϕ.

This method requires that the helical objects seen in electron micrographs be divided into segments, for use in the algorithm. The basic algorithm for IHRSR is displayed in figure 3.1.6.

The final volume obtained by the IRHSR approach can be assessed by comparison of the power spectrum of a 2D-projection of the volume with the averaged power spectrum of aligned raw segments and the final volume should have good agreement between the subunits of the map and an atomic model of the protein in question. However, at low resolution, there may not be enough detail to determine the correct symmetry from several possible solutions.

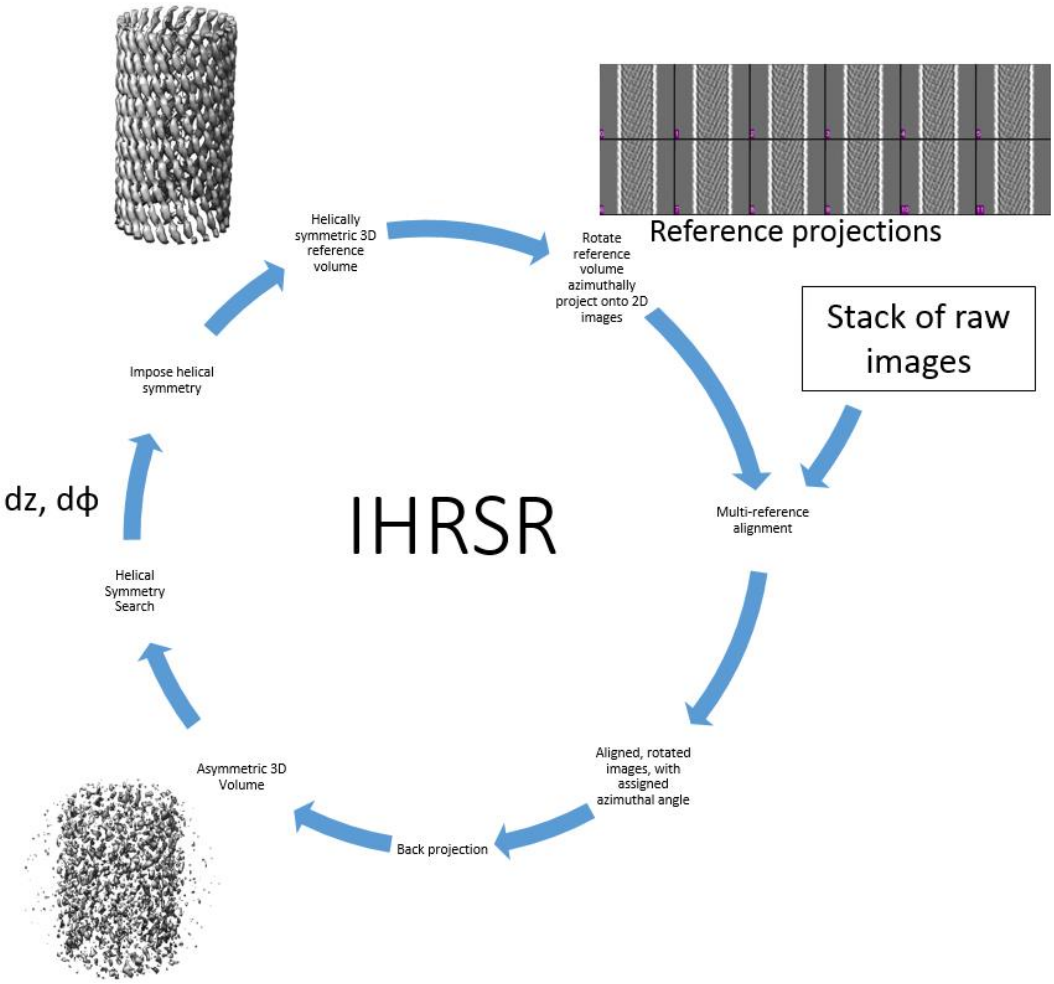


Figure 3.1.6. Overview of the Iterative Helical Real-Space Reconstruction cycle. A helically symmetric reference volume is rotated azimuthally, to generate a set of 2D reference projections. The set of reference projections is used in a multi-reference alignment with a set of raw image segments. For each segment, the azimuthal angle is determined from the best matching reference projection, as are the x- and y-shifts and in-plane rotation of the segment required to bring the segment into register with the reference projection. The segments are aligned, rotated, and back projected from their determined azimuthal angles to generate an asymmetric 3D volume.

The asymmetric volume is used in a helical symmetry search, where a least squares approach is used to determine the values of Δz and Δϕ. Helical symmetry is imposed on the asymmetric volume using these estimates of Δz and Δϕ, to generate a helically symmetric volume. The helically symmetric volume is used as the reference volume for the new cycle. Additional cycles are performed until there is little change in the values of Δz and Δϕ.

3.4. References

1. Diaz R, Rice WJ, Stokes DL: **Fourier-Bessel reconstruction of helical assemblies.** *Methods Enzymol* 2010, **482**:131–65.

2. Cochran W, Crick FH, Vand V: **The structure of synthetic polypeptides. I. The transform of atoms on a helix**. *Acta Crystallogr* 1952, **5**:581–586.

3. Egelman EH: **A robust algorithm for the reconstruction of helical filaments using single-particle methods**. *Ultramicroscopy* 2000, **85**:225–234.

4. Electron Microscopy and Helical Reconstruction

4.1. Methods

4.1.1. Crystallization

Crystallization plates were prepared using the sitting drop vapour diffusion method, by means of a Mosquito® auto pipetting robot (TTP Labtech). Purified free MsMca (19.6mg/ml, in 50mM Tris-HCl pH 8.0) was mixed with precipitant solution at ratios of 1:1, 3:1 and 1:3. Precipitant solutions were prepared using 0.8-2.8M ammonium sulfate, with 0.1M MES pH 7.0. Precipitant solutions were filtered through a .22µm filter before use. Final drop sizes were 200nL (3:1 and 1:3) and 400nL (1:1). Plates were stored at 20^{o}C. Images of drops were recorded using a Leica DFC 320 camera.

Later crystallization plates were prepared using MsMca at (9.8mg/ml, in 50mM Tris-HCl pH 8.0) and precipitant solution 1.4M ammonium sulfate, 0.1M HEPES pH 7.0, at a ratio of 1:1 and 3:1 protein:precipitant solution.

4.1.2. Sample Preparation

Initially, drops were used 1-2 days after the formation of crystals. Drops were diluted with 1µL precipitant solution, mixed briefly by aspiration, then covered for 5-10minutes. 1µL from the enlarged drop was transferred to a glow-discharged copper grid, and after 60 seconds, the grid was washed several times using 2% uranyl acetate, blotted and air-dried.

Later, samples were prepared using a wash step after application of the diluted crystallization drop to the grid. After 60 seconds, the 2µL of 0.3M ammonium sulfate, 0.1M HEPES pH 7.0, was applied to the grid. After 60 seconds, the grid was washed several times using 2% uranyl acetate, blotted, and air-dried.

4.1.3. Electron Microscopy

Electron Microscopy was performed using a Tecnai F20 microscope, operating at 200kV. Images were recorded using minimum dose (<20e-/Å^2) at a magnification of 29,000x on a Gatan 4K CCD camera, with a sampling of 3.8Å/pixel.

4.1.4. Image Processing

Filaments were selected from a number of micrographs using *e2helixboxer* (EMAN2 [1]. Filaments were divided into segments of size 256x256 pixels. The overlap between segments was 95%. Segments were processed using SPIDER [2]. The segments were aligned rotationally and translationally by a reference-free alignment protocol (AP SR). The segments were then masked, using a circular mask with Gaussian cut-off, and were padded to 1024x1024 pixels, to generate an average power spectrum for the purpose of indexing the helical symmetry of the filament.

4.1.5. Iterative Helical Real-Space Reconstruction (IHRSR)

IHRSR runs using a dataset taken from micrographs of negatively-stained samples were performed using SPIDER scripts modified from the script generated by the *generator* tool from IHRSR v1.4. Segments were aligned rotationally and translationally before use in IHRSR. IHRSR was performed over 50 cycles. In each cycle, 90 reference projections were generated, with an azimuthal rotation of 4 degrees between reference projections. Segments were 256x256 pixels, and the sampling was kept at 3.8 Å/pixel. A translational search limit of 5 pixels was allowed, as was a maximum search ring of radius 110 pixels. No rotational symmetry was applied for the initial reconstruction, though in later reconstructions, cyclic symmetry of n=3 was applied, and only 30 reference projections were used.

The first starting models used were featureless cylinders. In the additional reconstructions using selection rule 2, two featureless helices were used as the starting models. At a particular step for improvement of the reconstruction, two separate reconstructions were made from the same dataset, starting from different symmetries, and different starting models.

4.2. Results

4.2.1. Crystallization

After the plates were set up, needle-like crystals were seen after 1 day in drops containing 1.2M ammonium sulfate, 0.1M MES pH 7.0 and 19.6mg/ml MsMca (in 50mM Tris-HCl-pH 8.0), in a ratio of 1:1 and 3:1 protein:precipitant solution.

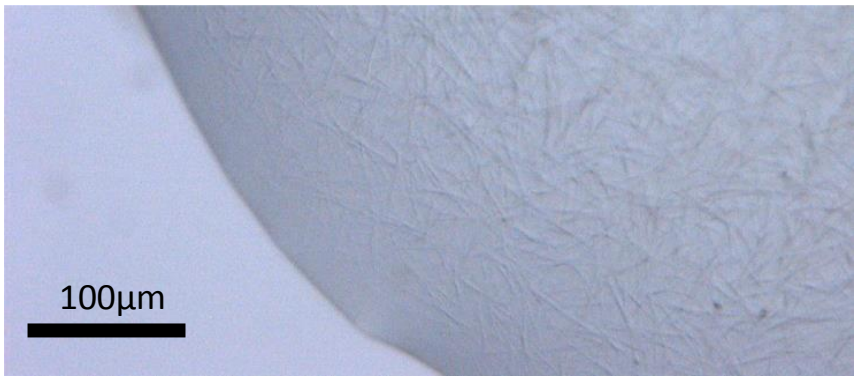


Figure 4.2.1 Light micrograph of a drop containing needles. The drop was prepared using 1.2 M ammonium sulfate, 0.1 M MES pH 7.0 and 19.6mg/ml MsMca (in 50mM Tris-HCl-pH 8.0), in a ratio of 3:1 of protein:precipitant solution. The image was taken after 2 days. Image contrast is enhanced to increase the visibility of the needles.

Crystallization conditions were modified to attempt to improve the morphology of the tubular crystals.

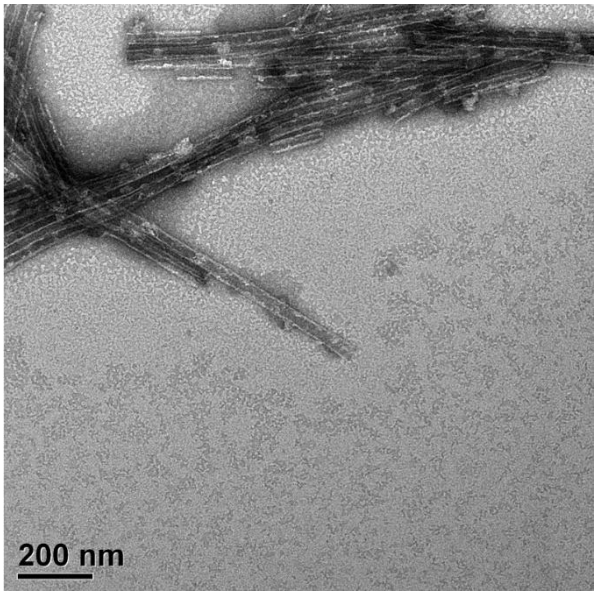
Varying concentrations of ammonium sulfate were tested (0.8, 1.0, 1.2, 1.4 and 1.6 M ammonium sulfate, 0.1M MES pH 7.0). pH was varied from 6.8-7.2 (in 1.2 M ammonium sulfate, 0.1 M MES), and several stabilizing additives were tested, in the presence of 1.2 M ammonium sulfate, 0.1 M MES pH 7.0. The additives included 5-10% glycerol, 10-20mM N-glucosamine (GlcN, used instead of GlcNAc owing to availability), 10-20mM myo-Inositol (Ins), 10mM GlcN + 10mM Ins, 1.2-4.8mM $MnCl_2$ and 2-5mM $CaCl_2$. $MnCl_2$ and $CaCl_2$ had been identified through thermofluor assays with MsMca to improve MsMca thermostability (data not shown).

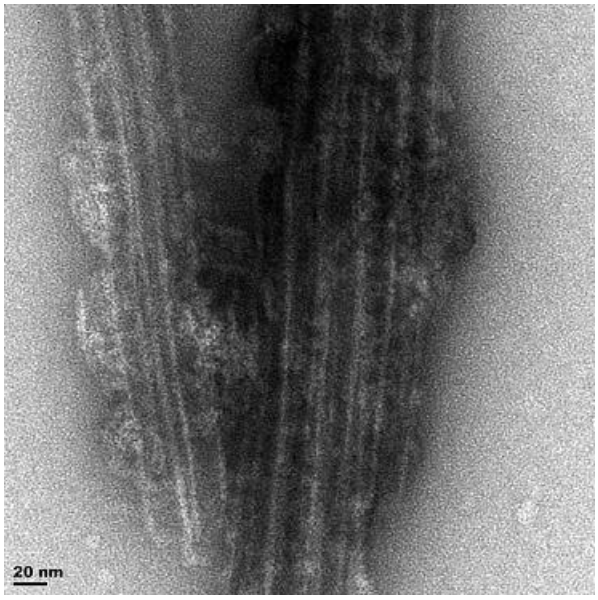
From this, needles only formed in the range of 1.2-1.4 M ammonium sulfate, and similar growth was seen over the pH range 6.8-7.2. While the presence of glycerol slowed needle growth, needles were of similar size and appearance to those grown in the absence of glycerol. The remaining additives did not prevent the growth of needles, and there was no change to needle morphology.

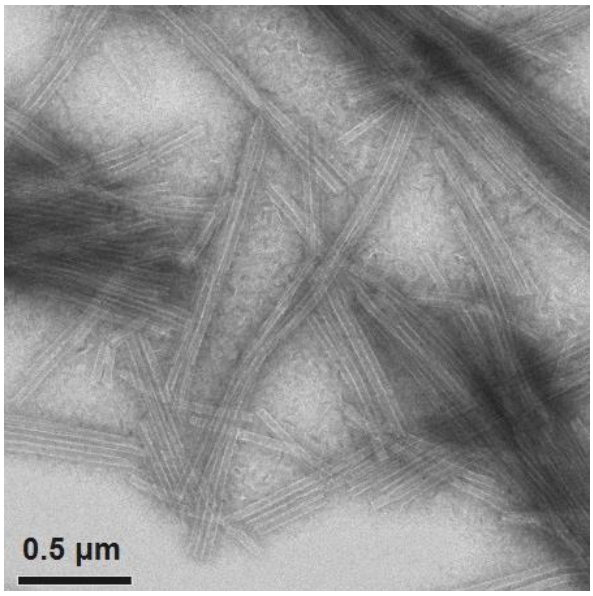
Little change was observed after the modification of crystallization conditions, so attempts were made to view the needles under negative stain electron microscopy, to determine the presence of crystalline order, using the power spectra generated from micrographs of the needles.

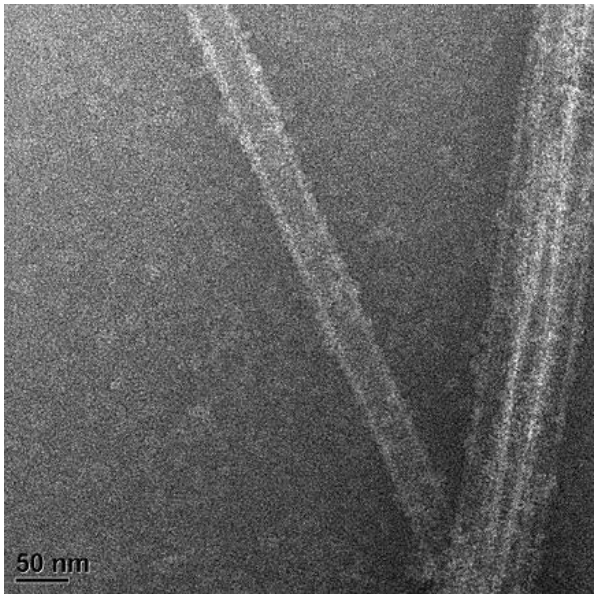
4.2.2. Electron Microscopy

Negatively-stained grids prepared using samples containing needles, showed large clusters of hollow filaments. The clusters (figure 4.2.2) appeared disordered, and were not suitable for use in 3D helical reconstruction.









F**igure 4.2.2. Selected micrographs of hollow filaments obtained from crystallization of MsMca.** (A-C) Show the presence of clusters of MsMca filaments, which could not be used for reconstruction of the filaments. (D) Shows an isolated filament, of outer diameter of approximately 30nm. Electron micrographs (A, B) were taken using a T20 G^2, operating at 120kV. Micrographs (C, D) were taken using a Tecnai F20 TEM, operating at 200kV.

Thereafter, sample preparation was modified, to obtain single filaments, which could be used for reconstruction. It was determined that drops containing lower protein concentrations produced needle-like crystals at a slower rate. From then on, crystallization conditions were modified to use 10mg/ml MsMca, in a 3:1 and 1:1 ratio with 1.4 M AS, 0.1 M HEPES, pH 7.0. Drops contained very thin spindle-like objects. These were mixed with different amounts of crystallization buffer, to increase the drop volume, and to enable loading of the drop to the carbon-coated copper grid. The micrographs obtained from such negatively-stained grids showed the presence of low numbers of aggregated filaments, but a large number of single, isolated filaments. The average diameter for the filaments is on the order of 35nm.

Filaments were selected from a number of micrographs using *e2helixboxer* (EMAN2 [1]). These were divided into segments of size 256x256 pixels (figure 4.2.3). The overlap between segments was 95%. The segments were aligned rotationally and translationally in SPIDER [2] by a reference-free alignment protocol (AP SR).

The segments were masked, using a circular mask with Gaussian cut-off, and were padded to 1024x1024 pixels, and their power spectra were averaged (figure 4.3.1).

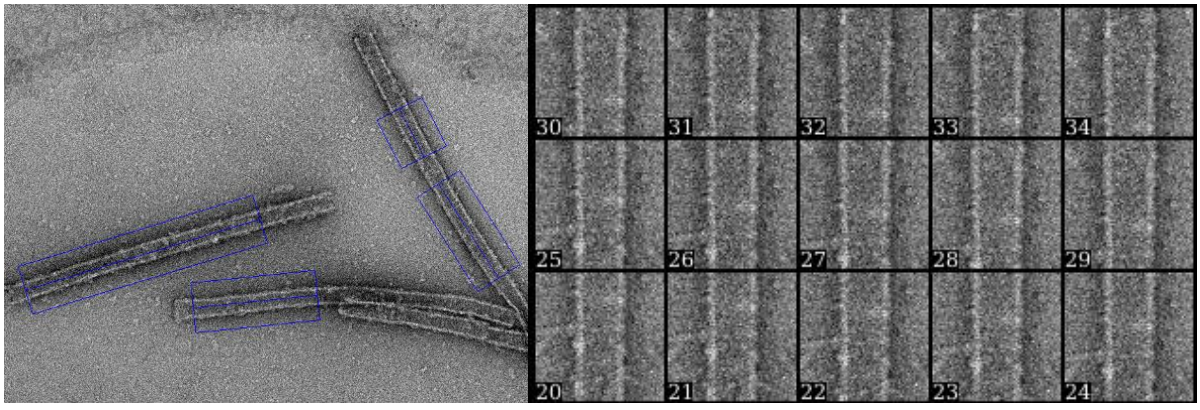


Figure 4.2.3. Boxed filaments from a micrograph of a negatively stained sample (left). Segments extracted from the boxed filaments are shown on the right.

4.3. Indexing

The positions of the major reflections in the diffraction pattern, or power spectrum (figure 4.3.1), of the filament were indexed, which provided an estimation of the helical symmetry of the MsMca filament. The helical symmetry is refined in the iterative helical real-space reconstruction (IHRSR) process, until it better matches the true helical symmetry of the filament.

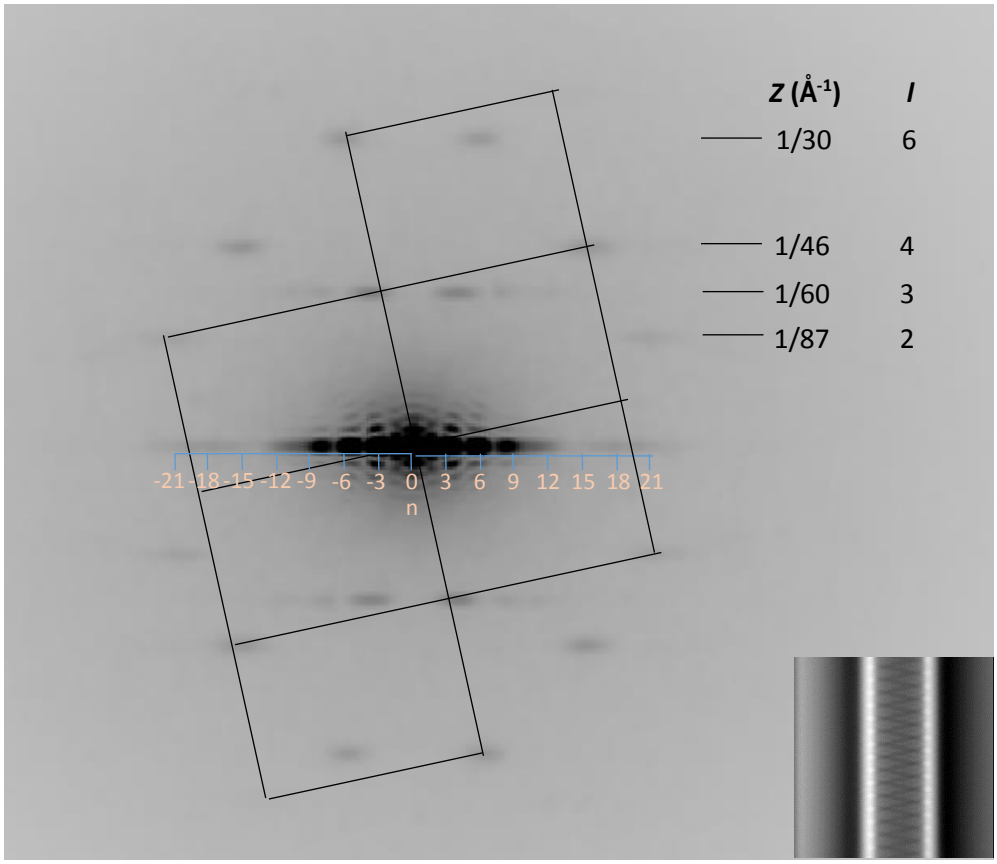
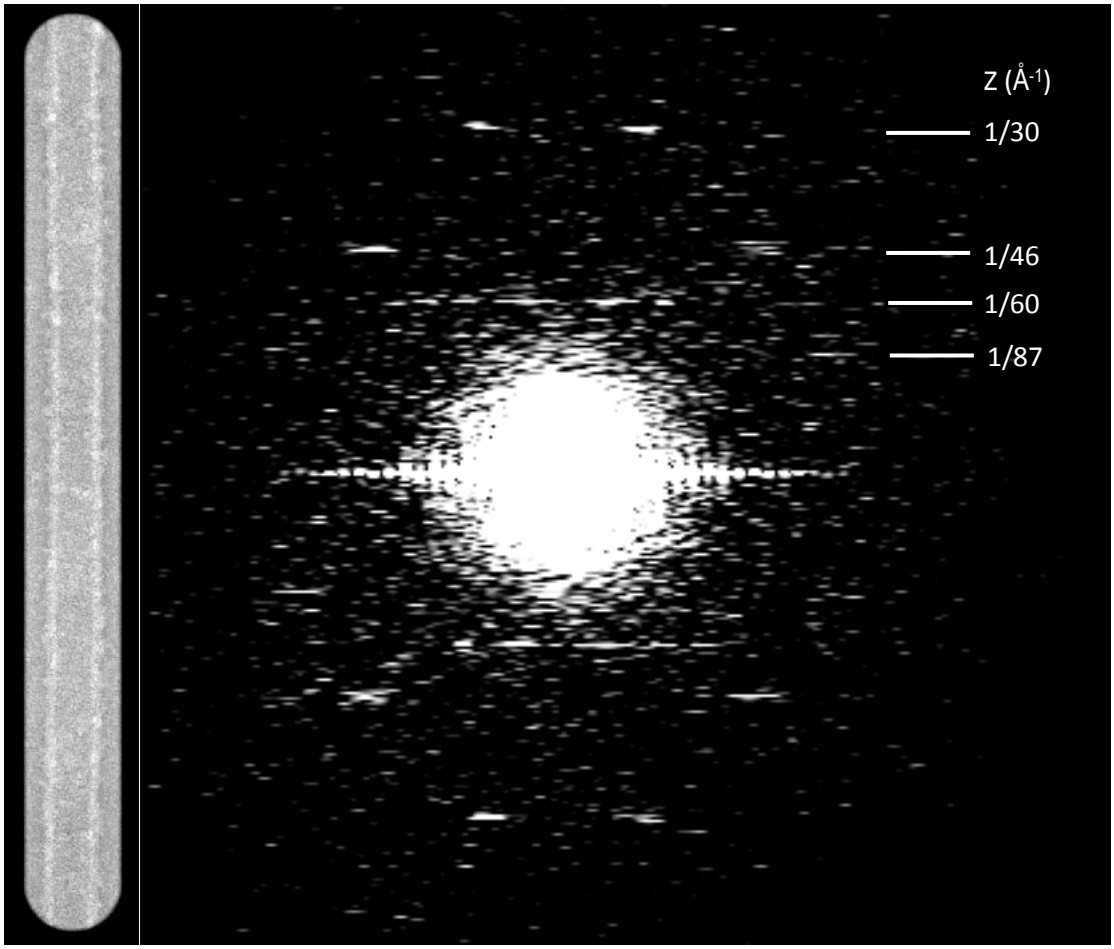
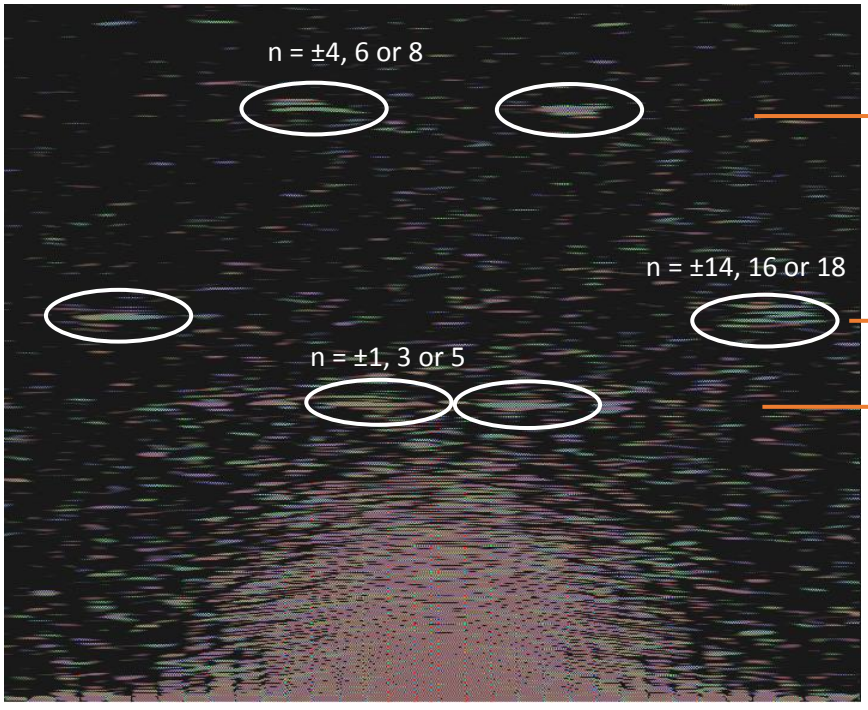


Figure 4.3.1. Average power spectrum of centred, aligned segments after padding and masking. Estimated Bessel orders (n) are shown on the horizontal axis. Layer lines (*l*) and their Z-positions (Å) are shown on the vertical axis. An estimated lattice is overlaid on the power spectrum. (Inset) Average of centred, aligned segments. Layer line positions are shown.

The positions of several layer lines were identified from the average power spectrum of rotationally aligned, centred filament segments (figure 4.3.1). The Bessel orders of the first major reflections on each layer line were estimated using the equation $2\pi Rr \sim 1.1n + 0.9$. The radius, r, of the filament was estimated at 170 Å, from the average (inset, figure 4.3.1). The distance, R, was taken as the distance along the horizontal axis from the meridional to the chosen reflection. The reflections are spread along the layer lines, thus there are a range of possible values of n for each layer line.



(A) (B)



1/30 Å

1/46 Å

1/60 Å

(C)

-180° 0 +180°

Figure 4.3.2: Fourier transform of a single MsMca filament. (A) A single MsMca filament, masked by a rounded rectangular mask. (B) The amplitudes of the Fourier transform of the masked filament in (A). Layer lines are indicated, and their Z positions are shown. (C) The Fourier transform of (A), only the upper half is shown. The Fourier Transform was generated using the FFT function of *e2display.py* from the EMAN2 suite [1]. Phases are represented by colours. The estimated orders of visible layer lines are shown. The reflections on layer line 3 are out of phase, while those on layer line 4 are in phase. The reflections on layer line 6 are also in phase. Therefore, the order of layer line 3 is odd, the order of layer line 4 is even, and the order of layer line 6 is even.

To reduce the number of possible orders for each layer line, one can examine the phases of the first reflections of each layer line. This is done by calculation of the Fourier transform of a single filament (figure 4.3.2). When the two major reflections on a layer line have a phase difference of ~180^{o}, the order is odd. A phase difference between the two reflections of ~0^{o} indicates that the order is even.

The layer lines and their predicted orders are shown in table 4.2.1.

There still exists a number of possible values of n for each layer-line examined. This can be reduced if the layer line *l* = 6 contains the second order reflections of those in the layer line *l* = 3. In this case, the order of the layer line *l* = 3 must be ±3, and the order of the layer line at *l* = 6 must be ±6.

The order of the layer line *l* = 4 is either ±14, 16 or 18.

	Initial estimates		With phase information		Predicted orders		
l	n_{min}	n_{max}	Odd/Even	n	Sel 1	Sel 2	Sel 3
1	20.4	22.9	n.d.	20-23	-17	-19	-21
2	20.4	27.6	n.d.	20-28	20	22	24
3	0.9	6.2	odd	1,3,5	3	3	3
4	14.2	19.9	even	14,16,18,20	-14	-16	-18
5	-*	-	-	-	-	-	-
6	3.2	8.7	even	4,6,8	6	6	6
7	11.6	17.4	n.d.	11-18	-11	-13	-15

Table 4.2.1 Layer lines and the associated Bessel orders for the MsMca filament. Initial estimates for the order (n) are shown in columns 2 and 3. After calculation of phase differences, the revised estimates for n are given in column 5. The predictions of the orders for each layer line from the three selection rules (Sel 1, Sel 2 and Sel 3) are shown in columns 6-8. *- no reflection on layer line 5 was seen in the power spectrum of the MsMca filament, or in the average power spectrum of rotated, aligned filament segments. n.d. - Not determined. Phase information was only available for layer lines 3, 4 and 6.

4.3.2. Selection rule

Using the rule $Z = l/c$ for the positions of the layer lines, we can calculate the position of the first layer line. Substituting the Z-positions of $l = 2$, $l = 3$, $l = 4$ and $l = 5$, the average value of $c = 179.5$ Å. The number of subunits per turn can be estimated from the selection rule:

$$l = um + tn$$

For layer line $l = 4$, if the order is -14, the selection rule is satisfied with

$$l = 18m + n \qquad (1)$$

where m = 1. In this case, there are 18 subunits per helical repeat, which is one turn. The pitch of the helix is then equal to the basic helical repeat, $P = c = 179.5$ Å. The axial translation of each subunit along the helical axis (Δz) is equal to the pitch of the helix, divided by the number of subunits per turn. Therefore, Δz = 9.97 or 10 Å. Δϕ is equal to the number of degrees in a single turn, divided by the number of subunits per turn. Therefore, Δϕ = 20°.

If the order of layer line $l = 4$ is -16, the selection rule is satisfied by

$$l = 20m + n \qquad (2)$$

where m = 1. Here, Δz = 8.975, or 9 Å, and Δϕ = 18°.

Where for layer line $l = 4$, n = -18, the selection rule is satisfied by

$$l = 22m + n \qquad (3)$$

where m = 1. Δz = 8.16 Å, and Δϕ = 16.4°.

Layer lines $l = 3$ and $l = 6$ are satisfied by each of the three selection rules, where m = 0 (See table 4.2.1). Layer line $l = 2$ was not used in the phase calculation, as the intensity of the major reflection was too low to permit unambiguous measurement of the phase.

4.3.3. Assessing the selection rules through IHRSR

IHRSR iteratively refines the helical symmetry parameters and simultaneously generates a 3D reconstruction of the helical object.

To eliminate potentially incorrect symmetries and establish the orders of layer lines 4 and 2, IHRSR was performed using each of the three estimated symmetries.

4.4. IHRSR

In this process, a single starting model was used for each of the IHRSR runs. Different starting symmetries were applied, and the same raw image stack was used in each reconstruction. The procedure was performed over 50 cycles.

The reconstructions converged on two different models. The first model was produced from a starting symmetry of Δz = 10 Å, and $\Delta\phi = 20^{o}$, corresponding to selection rule 1. The symmetry converged on Δz = 10.27 Å, and $\Delta\phi = 20.05^{o}$ (figure 4.4.1), indicating 18 subunits per turn, and a pitch of 184.9 Å.

The second model was produced from starting symmetries of Δz = 9 Å and $\Delta\phi = 18^{o}$, and Δz = 8.16 Å and $\Delta\phi = 16.4^{o}$, from selection rules 2 and 3, respectively. The symmetry in each run converged on Δz = 9.12 Å, and $\Delta\phi = 17.98^{o}$, indicating 20 subunits per turn, and a pitch of 182.4 Å.

Power spectra were generated from projections of the symmetrized volume. The power spectrum from reconstruction 1 shows strong reflections on layer lines l = 3, l = 4 and l = 6. The power spectrum from reconstruction 2 shows strong reflections only on layer lines l = 3 and l = 6.

The orders of layer lines 3 and 6 are 3 and 6 respectively. The reconstruction has a strong 3-start helix. On each subunit, there are two extensions, which are spaced approximately 30 Å apart. These extensions, when observed over successive subunits in the three-start helix, produce a strong 6-start helix. A third, 15-start helix is visible, of the opposite hand to the 3- and 6-start helices. This is in contrast to the expected 14-start helix, according to prediction from the selection rule.

The second reconstruction has a strong 3- and 6-start helix. The subunits have a different appearance to those in the first reconstruction, however, the subunits maintain the feature of two raised extensions, which produce the 6-start helix. The asymmetric unit contains three subunits, aligned almost vertically, on the three-start helix. In both reconstructions, there is minor heterogeneity between subunits in the asymmetric unit.

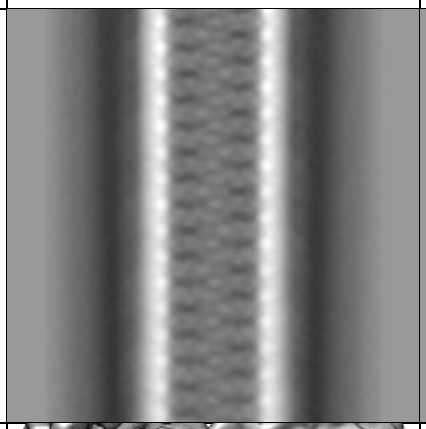
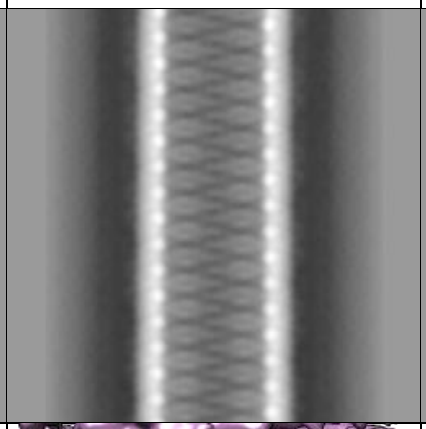
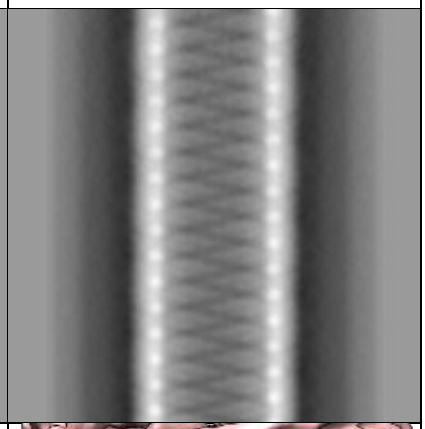
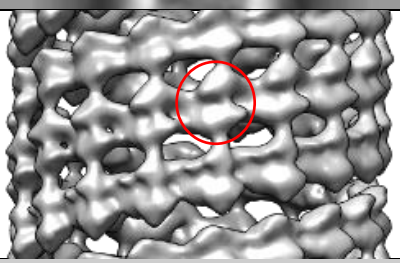
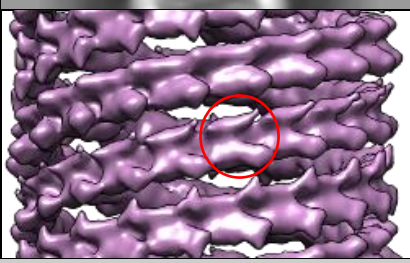
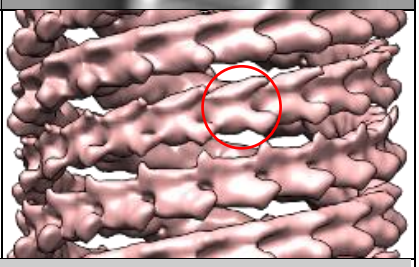
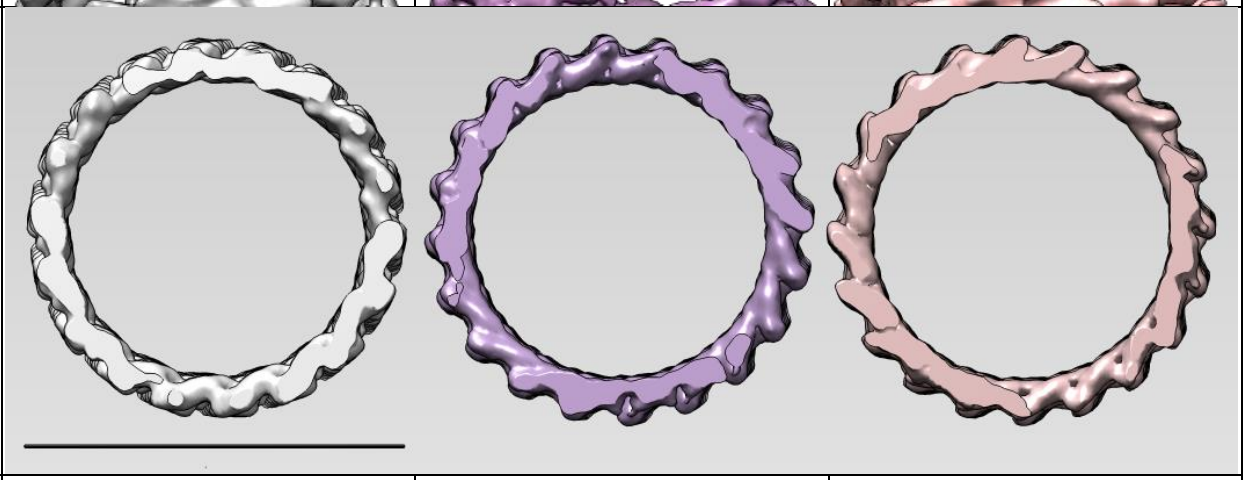
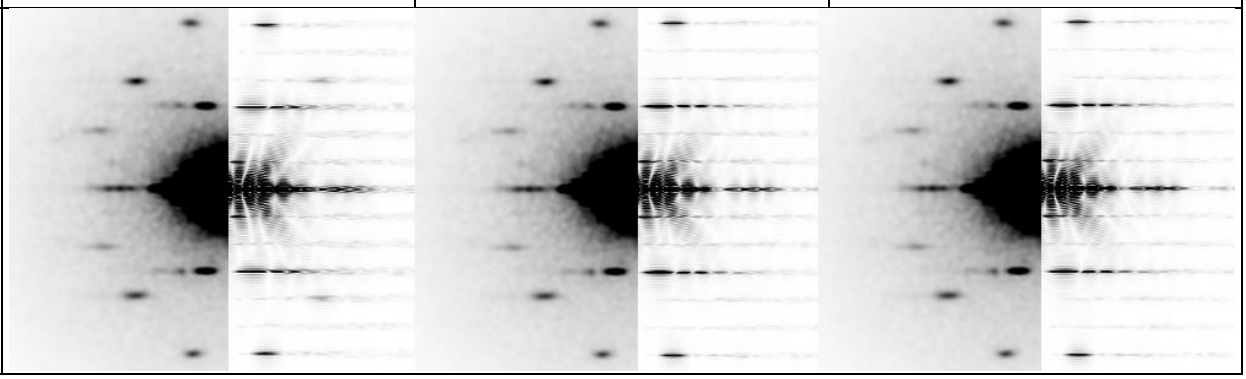
NAME	SELECTION RULE 1	SELECTION RULE 2	SELECTION RULE 3
2D PROJECTION OF SYMMETRIZED FINAL VOLUME			
3D RECONSTRUCTION			
CROSS-SECTION ALONG THE Z-AXIS			
IMAGES	5525	5525	5525
STARTING SYMMETRY	20.0° 10.0 Å	18.0° 9.0 Å	16.4° 8.16 Å
FINAL SYMMETRY	20.05° 10.27 Å	17.98° 9.12 Å	17.98° 9.12 Å
INNER AND OUTER RADIUS	270 Å 345 Å	265 Å 355 Å	265 Å 355 Å
POWER SPECTRA OF (A) ALIGNED FILAMENT SEGMENTS AND (B) A PROJECTION OF THE RECONSTRUCTION			

Table 4.4.1: Comparison of the IHRSR reconstructions from the selection rules 1, 2 and 3. Row 1 – 2D projections of the symmetrized volume. Row 2 – Surface representations of the three-dimensional symmetrized volume. A single subunit is circled in red. Row 3 – Cross-section along the z-axis of the symmetrized reconstructions (from left: reconstruction 1, reconstruction 2, reconstruction 3). The scale bar represents 345 Å. Rows 4-7 – Details of the reconstruction: Number of images used, starting and final symmetry, and inner and outer radii of the reconstructed volume. Row 8 – Power spectra of elongated projections of the symmetrized volumes.

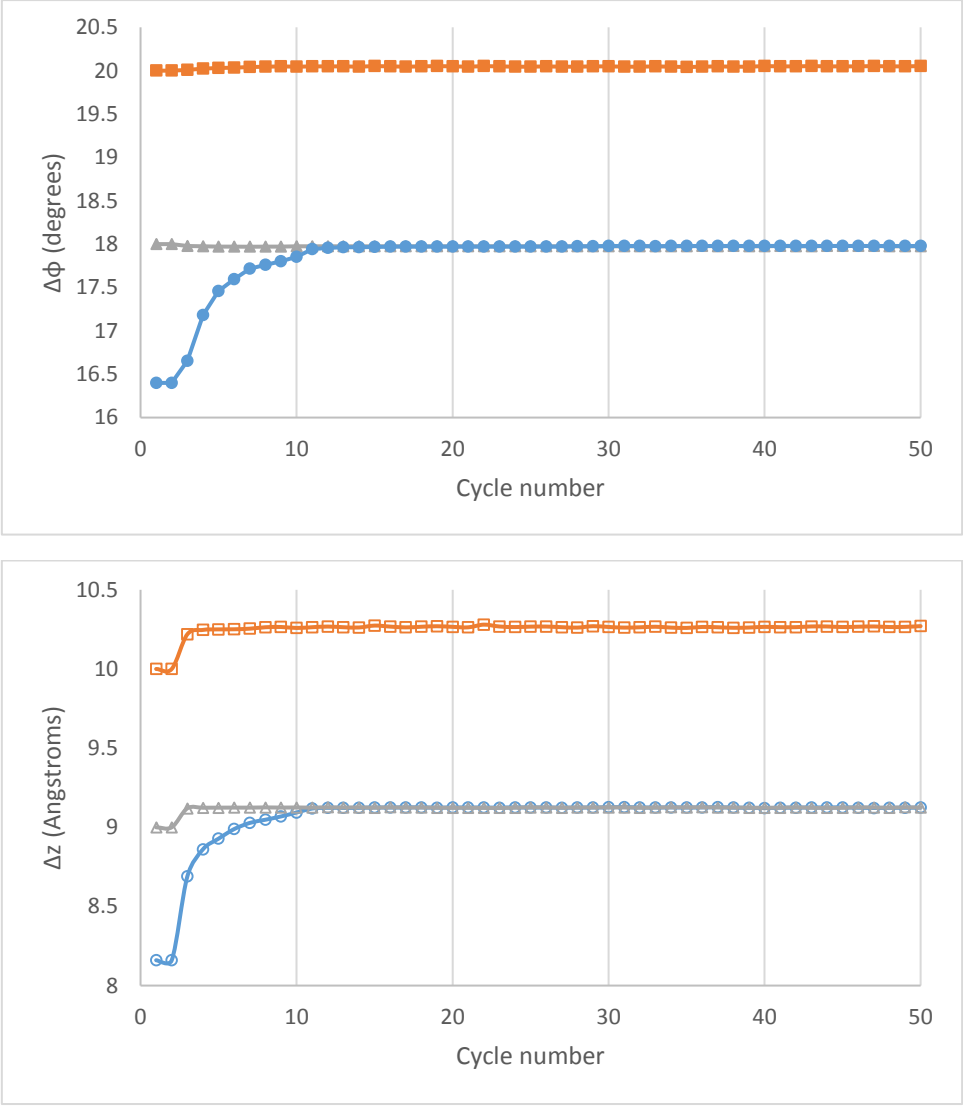


Figure 4.4.1: Refinement of helical symmetry parameters by IHRSR. The first reconstruction started from Δz = 10.0 Å (open squares), and Δϕ = 20.0° (filled squares). The second reconstruction started from Δz = 9.0 Å (open triangles), and Δϕ = 18.0° (filled triangles). The third reconstruction started from Δz = 8.16 Å (open circles), and Δϕ = 16.4° (filled circles). Reconstruction 1 converged on Δz = 10.26 Å and Δϕ = 20.05°, after approximately 4 cycles. Reconstructions 2 and 3 converged on Δz = 9.12 Å and Δϕ = 17.98°, after 11 cycles.

4.4.1. Fitting a homology model of MsMca into the reconstructions

To assess the reconstructions generated from selection rules 1 and 2, studies were performed using several homology models of MsMca [appendix], which were fit into the reconstructions (figures 4.4.2 and 4.4.3). The UCSF Chimera *Fit-in-map* tool [3], was used to semi-automatically fit the monomers into the reconstruction maps. For each subunit, a map was simulated from the atomic model, at a resolution of 20 Å, and the fit was achieved by optimizing the mean cross-correlation between the simulated map and the reconstruction map. The threshold of the map was set to enclose a volume of $4x10^4$ $Å^3$ per subunit, from the mass of MsMca (33 000 Da), and a protein density of 0.73 cm^3/g [4].

In both cases, the atomic models fit into the reconstruction maps with a good cross-correlation score, typically between 0.64 and 0.71, due to minor variation between subunits on the three-start helix. When the subunits of the reconstruction are compared with the simulated map from the atomic models, there is not complete agreement between the simulated map and that of the reconstruction. Although the central density of the simulated map of the atomic model resembles that of the subunits in the reconstructions, the upper and lower extensions seen on the subunits of the two reconstructions are not seen in the simulated maps of the atomic models.

Therefore, both reconstructions remain as possible solutions to the structure of the actual MsMca filament.

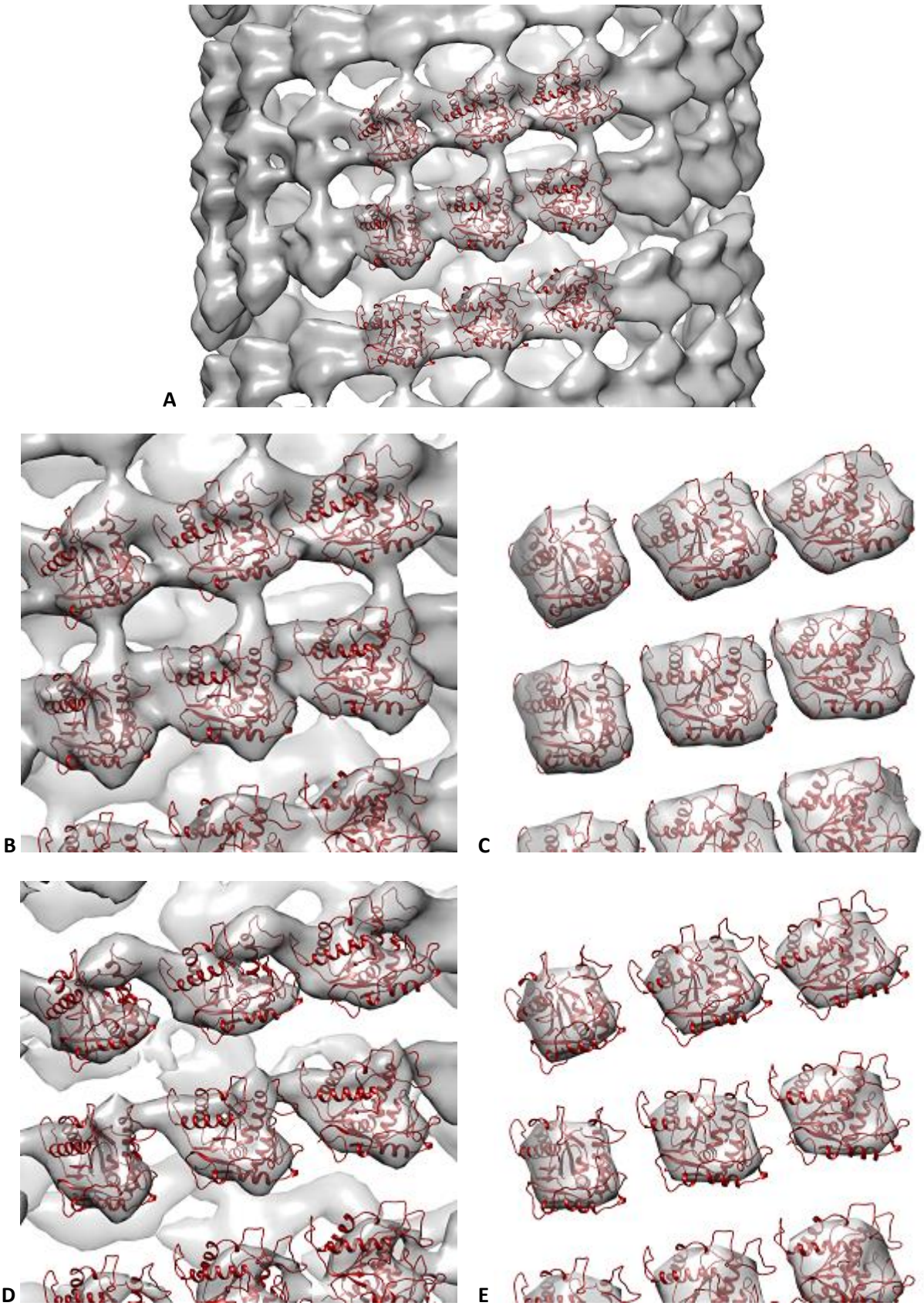


Figure 4.4.2: Fitting of atomic models of MsMca into the first reconstruction. (A) 9 atomic models of MsMca docked into the entire reconstruction. Atomic models of MsMca are docked into the reconstruction (B), and simulated maps from the atomic models (C), at the threshold expected for the volume of MsMca. The active site of the MsMca model is pointed in the direction of the basic helix. (D) and (E), atomic models of MsMca docked into the reconstruction, or simulated maps from the atomic models, respectively, at a threshold for half of the expected volume for MsMca.

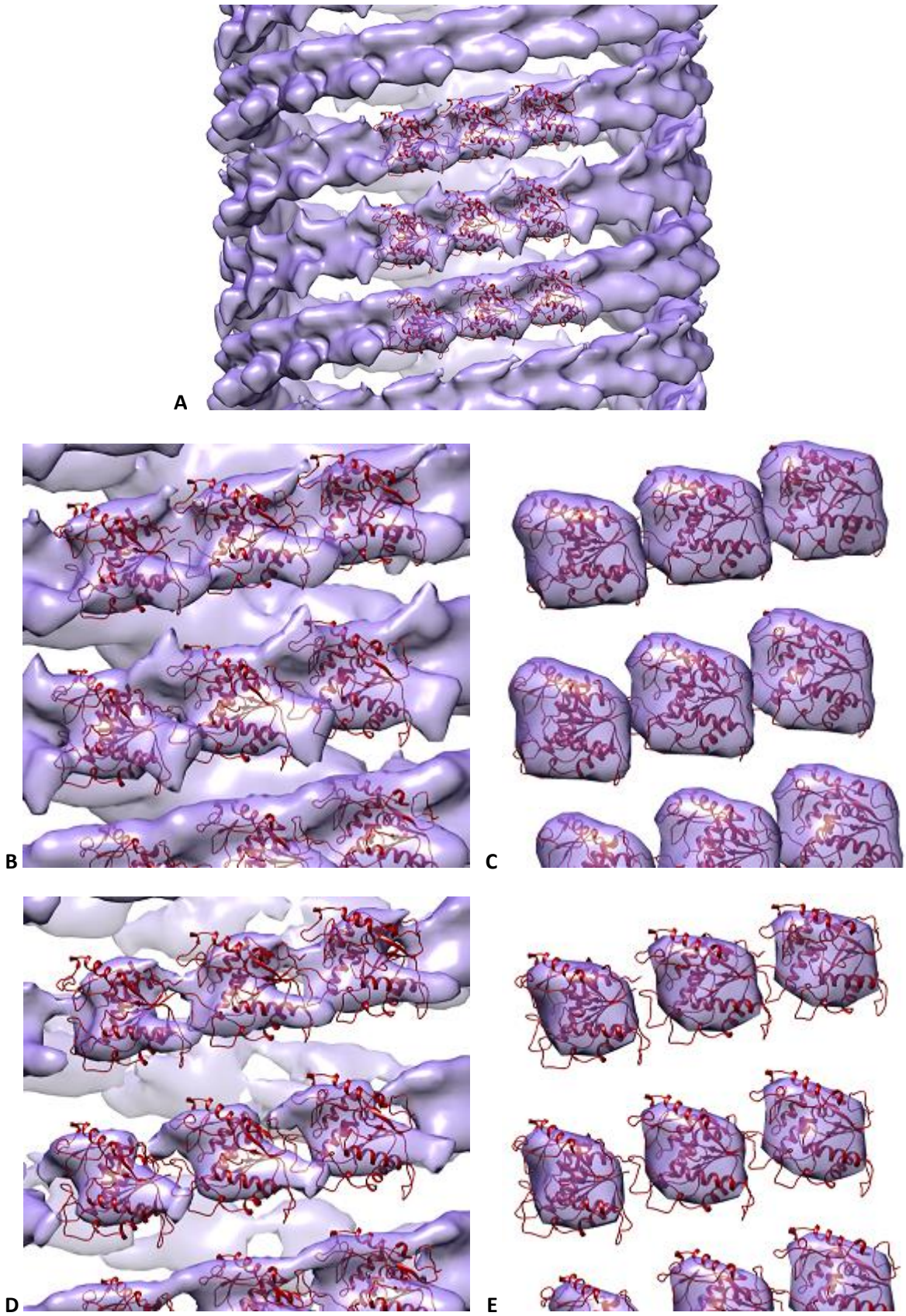


Figure 4.4.3: Fitting of atomic models of MsMca into the second reconstruction. (A) 9 atomic models of MsMca docked into the entire reconstruction. Atomic models of MsMca are docked into the reconstruction (B), and simulated maps from the atomic models (C), at the threshold expected for the volume of MsMca. The active site of the MsMca model is pointed in the opposite direction to the direction of the basic helix. (D) and (E), atomic models of MsMca docked into the reconstruction, or simulated maps from the atomic models, respectively, at a threshold for half of the expected volume for MsMca.

4.5. Improvements to the second reconstruction

Both the first and second reconstructions led to possible solutions for the symmetry of the MsMca filament.

In this work, the second reconstruction was refined using several methods not used in the initial reconstruction process. These steps are described in the following section, and their contributions to model improvement are discussed. These steps included: (A) - removal of segments of alternative diameters, (B) - use of layer-line filtering, (C) - (A) with imposition of 3-fold cyclic symmetry and (D) - combined use of (B) and (C) with an additional filter to remove images with low cross-correlation scores.

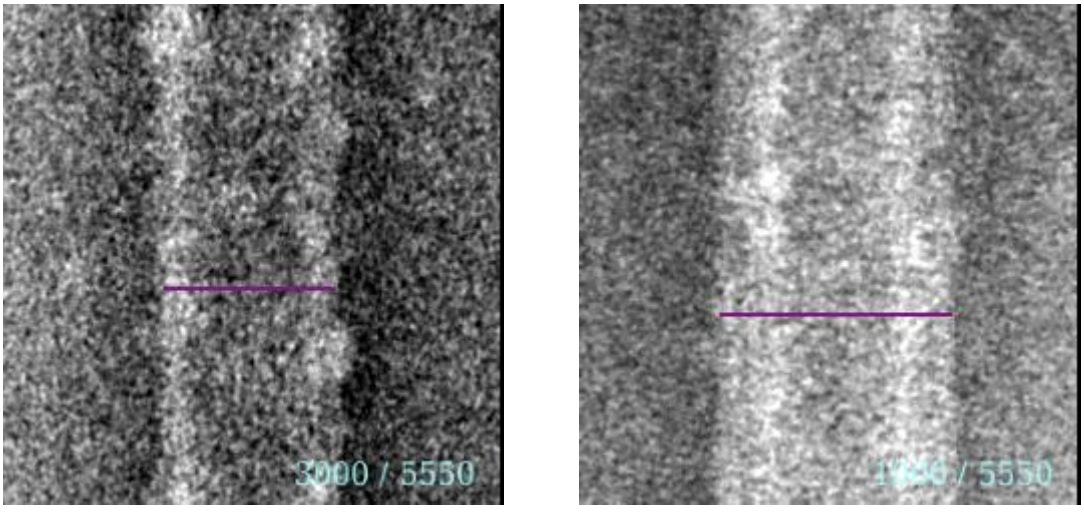
In the future, these methods will be applied to the first reconstruction, to better distinguish between the two suggested symmetries of the MsMca filament.

4.5.1. Filtered image stack reconstruction

In the original dataset, there was a mixed population of segments, with respect to filament diameter. A number of segments (~10%) in the dataset were taken from filaments with a twin-layered wall (figure 4.5.1), whereas the remainder (~90%) contained only a single layer. The twin-walled segments were included in the IHRSR process, and contributed to the original reconstruction. In this step, these twin-walled segments were removed from the dataset, in an attempt to improve the symmetry estimations from the first reconstruction.

Two separate reconstructions from the filtered dataset were made, and converged on symmetry parameters of Δz 9.13 and 9.14 Å, and Δϕ 17.97 and 17.99 degrees, respectively (figure 4.5.2). There is a major groove along each of the three-start helices, which passes through the centre of each subunit. This was seen in the original reconstruction.

The power spectrum of an elongated 2D projection of the reconstruction is compared with that of the first reconstruction (Table 4.5.1). The layer line Z positions are conserved, and the major reflections on layer lines 3 and 6 are conserved in their distance (R) from the meridional. The intensities of the major reflections on layer lines 1, 2 and 4 are weak, as is the case for the power spectrum of the original reconstruction.



(A) (B)

Figure 4.5.1: Comparison of filament segments of single-walled and double-walled filaments. (A) A single-walled filament segment. The scale bar represents ~330 Å. (B) A double-walled filament segment. The scale bar represents ~440 Å.

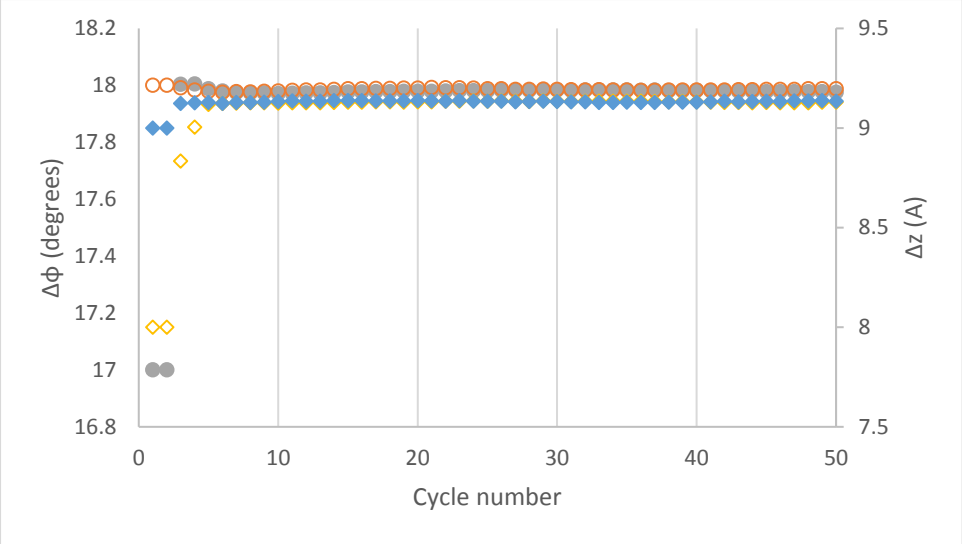


Figure 4.5.2: Refinement of helical symmetry parameters by IHRSR. The first reconstruction started from Δz = 8.0 Å (open diamonds), and Δϕ = 17.0° (filled circles). The second reconstruction started from Δz = 9.0 Å (filled diamonds), and Δϕ = 18.0° (open circles). Both reconstructions converged on Δz = 9.13 Å and Δϕ = 18.0°, by cycle 10.

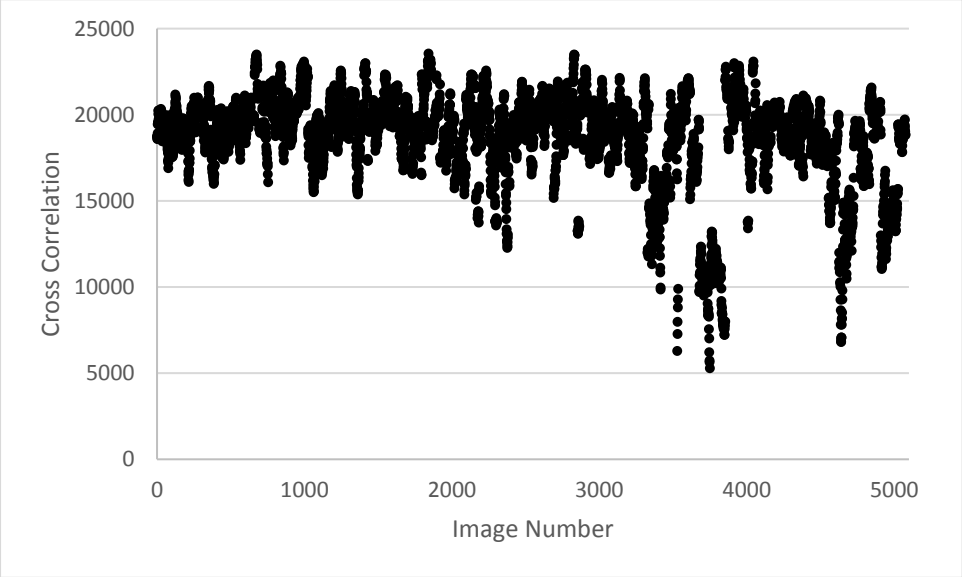


Figure 4.5.3: The cross-correlation between each image in the filtered dataset and its best matching reference projection at cycle 50.

The cross-correlation for each image with its best matching reference projection is shown in figure 4.5.3. The cross-correlation mean for the entire dataset was 18617.54±75.15 (95% confidence interval). The standard deviation of the dataset was 2731.8, with 90% of the images having a cross-correlation of 15 000 or greater with their best matching reference projection. The images having a cross-correlation of less than 15 000 were predominantly 3330-3440, 3530-3535, 3679-3853, 4565-4721, and 4917-5019.

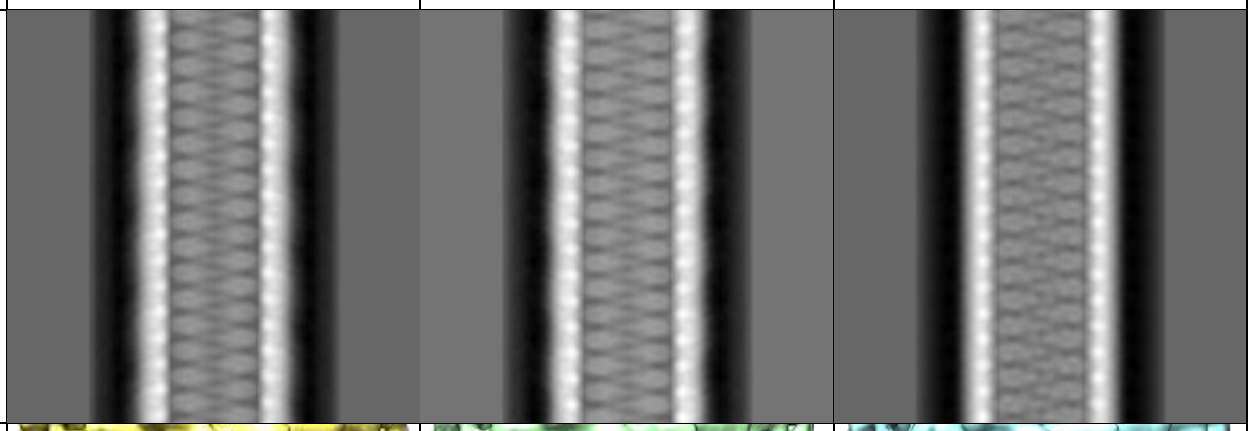
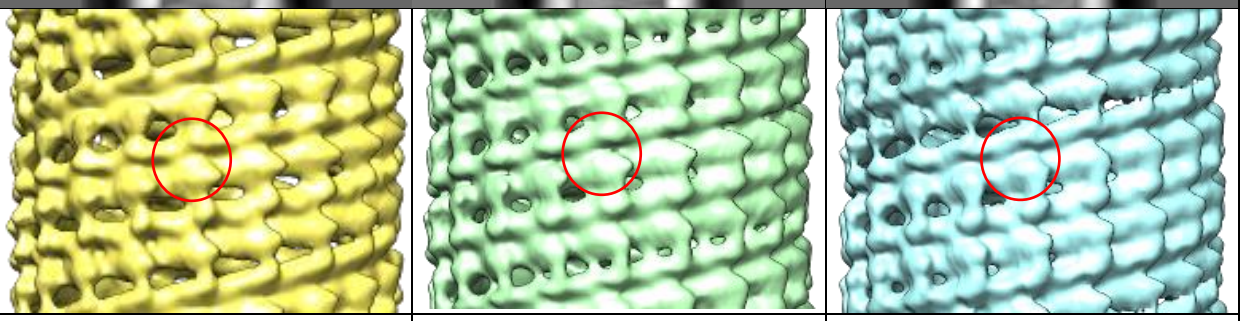
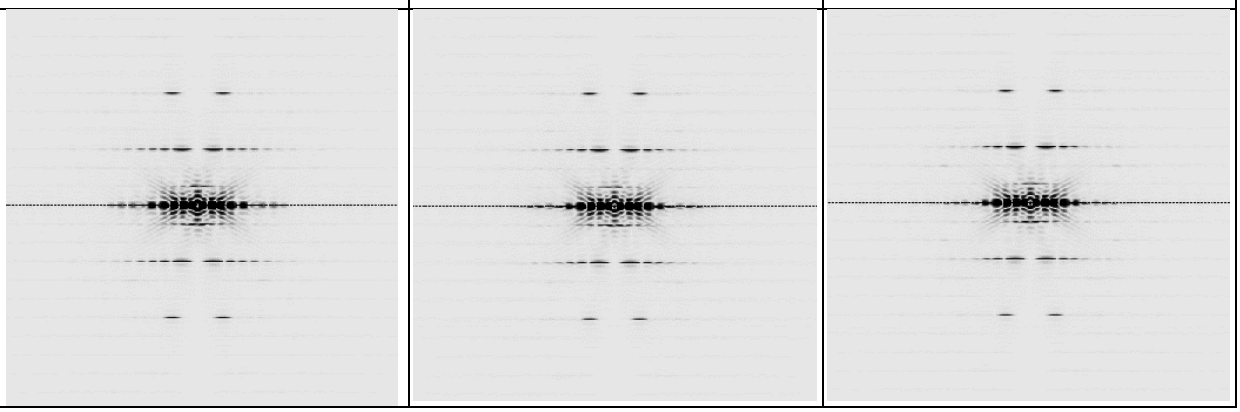
NAME	ORIGINAL RECONSTRUCTION FROM SINGLE- AND TWIN – WALLED FILAMENT SEGMENTS	RECONSTRUCTION FROM SINGLE-WALLED FILAMENT SEGMENTS ONLY	
		Reconstruction A	Reconstruction B
2D PROJECTION OF SYMMETRIZED FINAL VOLUME			
3D RECONSTRUCTION			
NUMBER OF IMAGES	5525	5078	5078
STARTING SYMMETRY	14.0° 7.0 Å	17.0° 8.0 Å	18.0° 9.0 Å
FINAL SYMMETRY	18.0° 9.13 Å	17.97° 9.13 Å	17.99° 9.14 Å
POWER SPECTRUM OF ELONGATED MODEL PROJECTION			

Table 4.5.1: Comparison of the original reconstruction with the reconstructions made from the dataset containing only single-walled filament segments. Row 1 – 2D projections of the final volumes after 50 cycles of IHRSR. Row 2 – Surface representations of the final volumes. A single subunit is circled in red. The contour level of the reconstructions was set to show continuous density along the z-axis. Row 3-5 – Summary statistics for (3) the number of segments used in the cycles of each reconstruction, (4) and (5) the initial and final symmetry parameters respectively of rotation (degrees) and rise (A) per subunit after 50 cycles of the IHRSR process. Row 6 – Power spectra calculated from elongated 2D projections of the symmetrized final reconstruction. Layer line positions are shown.

4.5.2. Reconstruction from layer-line filtered images

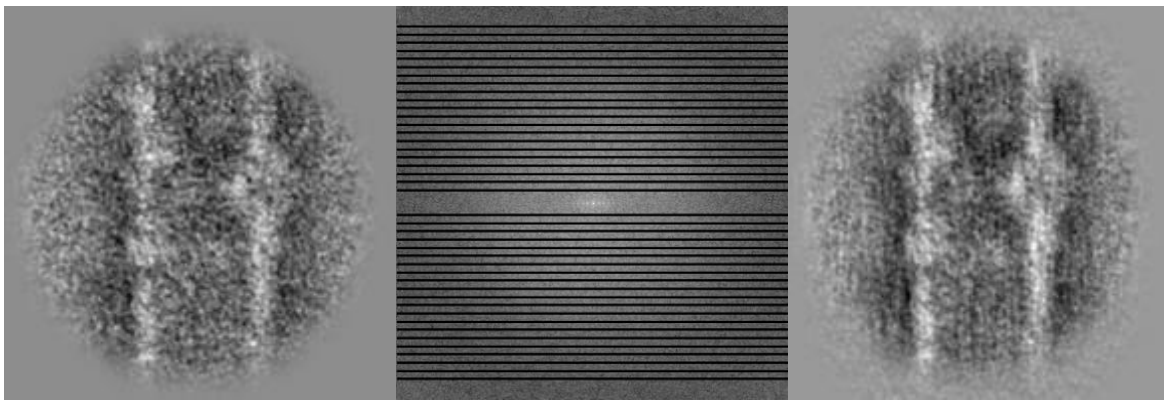


Figure 4.5.4: The effect of layer line filtering on the segments of Mca filaments. (Left) The original segment, masked with a circular mask, filtered with a Gaussian low-pass filter, and padded to 512x512 pixels. (Centre) The Fourier transform of the segment in (A), filtered to omit data between layer lines. (Right) The segment in (A), after layer-line filtering. Subunits are visible, and noise is reduced.

Layer line filtering was used in the reconstruction of ATP-ases from large crystalline tubes to reduce noise, and to enable better estimation of initial symmetry parameters by IHRSR [5]. In this work, layer-line filters were applied at varying degrees to omit data between layer lines. The layer line filters were designed in ImageJ [6], after the calculation of the Fourier transform of masked, padded segments (figure 4.5.4). The masked Fourier transform was used to generate the filtered masked, padded segment. In the example in figure 4.5.4, layer-line filtering reduces noise considerably, and potential subunits are defined, which was not the case for the unfiltered segment, where the potential subunits are poorly defined. The 'subunits' may also be an artefact of the filtering process.

The width of the layer line mask was tested in reconstructions using a small dataset of 1000 images. In this, layer line masks omitting 1, 3 or 5 rows (in pixels) (masks A, B and C, respectively) between the layer lines were tested against unfiltered images. IHRSR runs were performed using identical starting models and starting symmetry parameters. The results after 50 cycles are shown in table 4.5.2.

The 2D projections of the symmetrized final volumes do not show many large differences between images filtered with the layer line filter, and unfiltered images. The same basic three-start helix with pitch of ~182 Å is seen for all reconstructions.

2D power spectra generated form the different reconstructions show consistent reflections on layer lines 3, 4 and 6. The strongest reflections are seen on these layer lines for reconstructions using

layer-line masks A and B. The basic helical symmetry of each reconstruction was not affected by the use of the layer line filter.

The 3D reconstructions show small differences in subunit appearances between unfiltered and the filtered images. The reconstruction with greatest similarity to that from the unfiltered image set was produced using layer-line mask B.

From this, a reconstruction was made using the original dataset of 5525 segments, filtered using the layer-line mask B (see Table 4.5.3). The reconstruction converged on symmetry parameters of Δz 9.14 Å and Δϕ 18.0 (see figure 4.5.5). Slight differences are visible in the subunits from the layer-line filtered reconstruction compared to those of the original reconstruction. No large differences were observed, and the final symmetry parameters are almost identical.

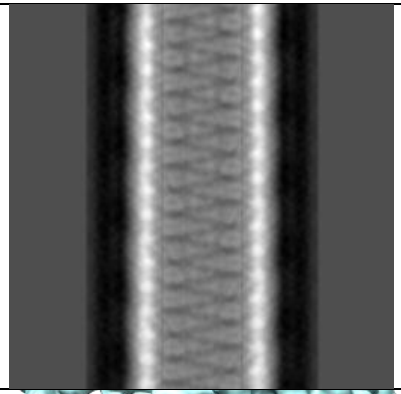
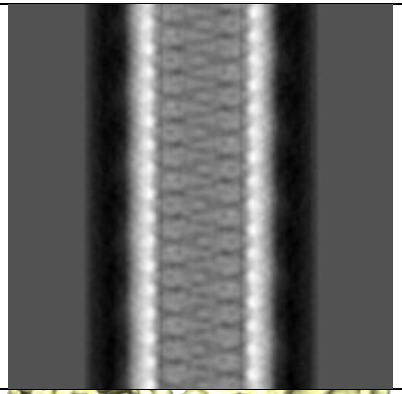
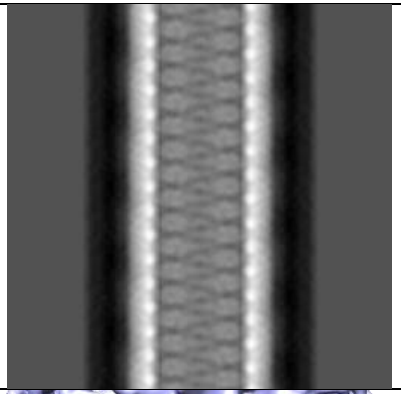
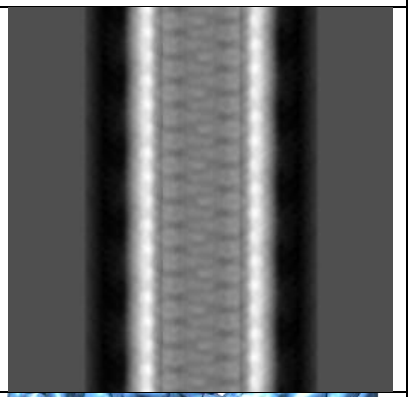
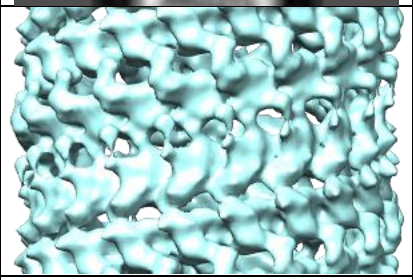
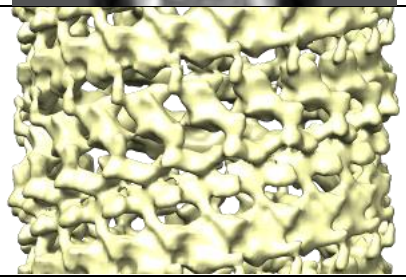
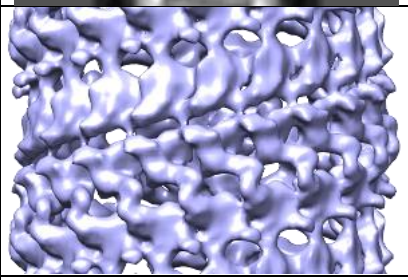
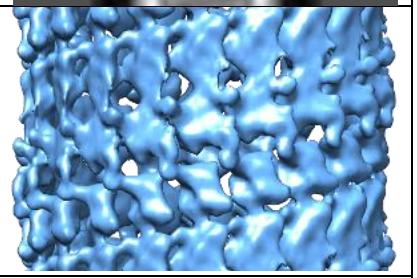
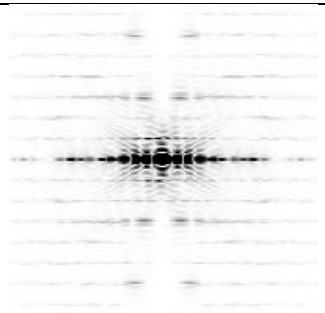
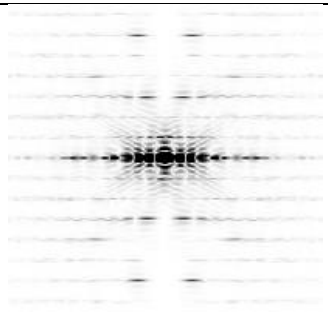
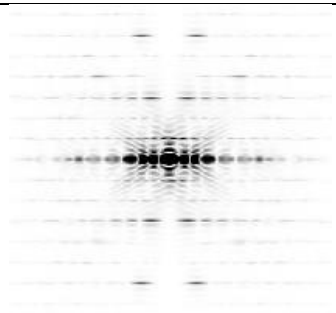
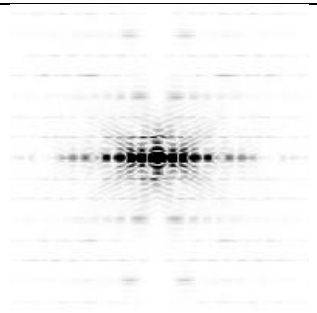
NAME	NO LAYER LINE MASK	LAYER LINE MASK A	LAYER LINE MASK B	LAYER LINE MASK C
2D PROJECTION OF SYMMETRIZED FINAL VOLUME				
3D RECONSTRUCTION				
NUMBER OF IMAGES	1000	1000	1000	1000
STARTING SYMMETRY	17.0° 8.0 Å	17.0° 8.0 Å	17.0° 8.0 Å	17.0° 8.0 Å
FINAL SYMMETRY	17.97° 9.12 Å	17.97° 9.14 Å	17.99° 9.15 Å	18.00° 9.12 Å
POWER SPECTRUM OF ELONGATED MODEL PROJECTION				

Table 4.5.2: Results of four separate reconstructions to test the effect of varying layer-line filters on images used in the reconstruction. 2D projections of the final reconstructions are shown in row 1. Images of the final 3D reconstructions are shown in row 2. The 6th row shows the power spectra generated from projections of each reconstruction.

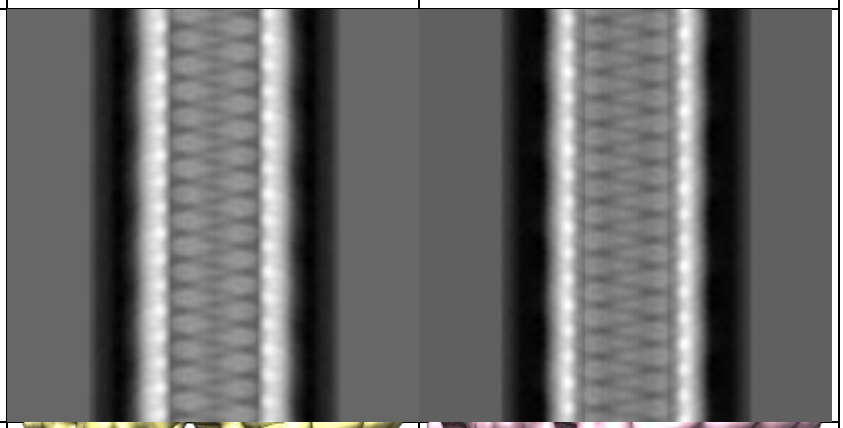
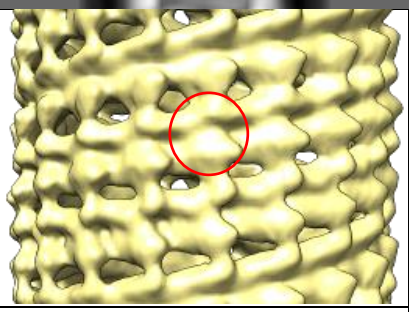
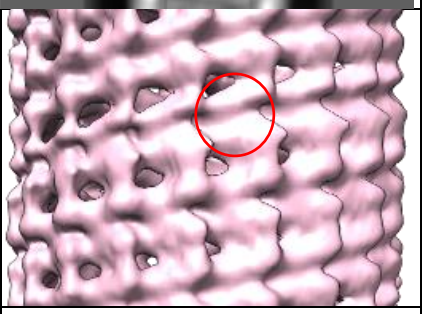
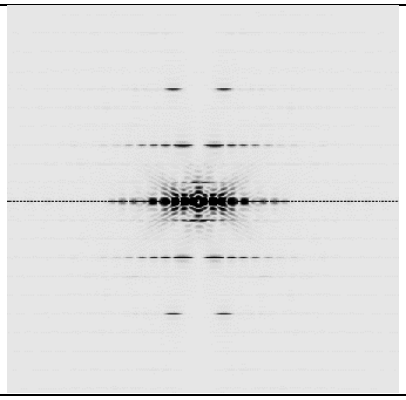
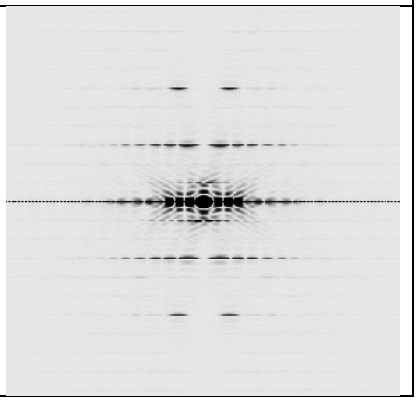
NAME	UNFILTERED IMAGES	LAYER LINE MASK B
2D PROJECTION OF SYMMETRIZED FINAL VOLUME		
3D RECONSTRUCTION		
NUMBER OF IMAGES	5525	5525
STARTING SYMMETRY	14.0° 7.0 Å	18.0° 9.0 Å
FINAL SYMMETRY	18.0° 9.13 Å	18.0° 9.14 Å
POWER SPECTRUM OF ELONGATED MODEL PROJECTION		

Table 4.5.3: Comparison of the reconstruction from unfiltered images and the reconstruction from images filtered using layer-line mask B. Row 1 – 2D projections of the final volumes after 50 cycles of IHRSR. Row 2 – Surface representations of the final volumes. A single subunit is circled in red. The contour level of the reconstructions was set to show continuous density along the z-axis. Row 3-5 – Summary statistics for (3) the number of segments used in the cycles of each reconstruction, (4) and (5) the initial and final symmetry parameters respectively of rotation (degrees) and rise (A) per subunit after 50 cycles of the IHRSR process. Row 6 – Power spectra calculated from elongated 2D projections of the symmetrized final reconstruction. Layer line positions are shown.

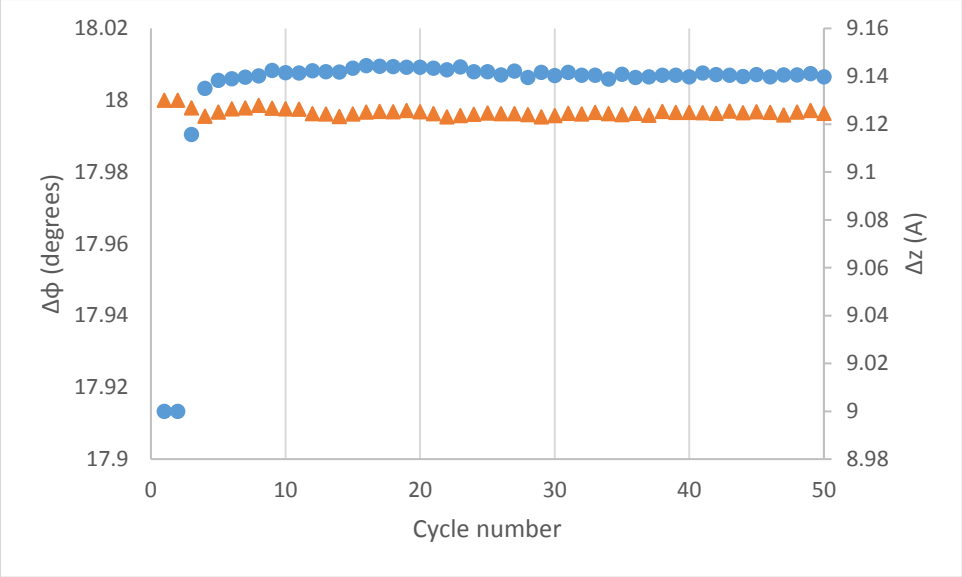


Figure 4.5.5: Refinement of helical symmetry parameters by IHRSR. The reconstruction started from Δz = 9.0 Å (filled circles), and Δϕ = 18.0° (filled triangles). The reconstruction converged on Δz = 9.14 Å and Δϕ = 18.0°, within 10 cycles.

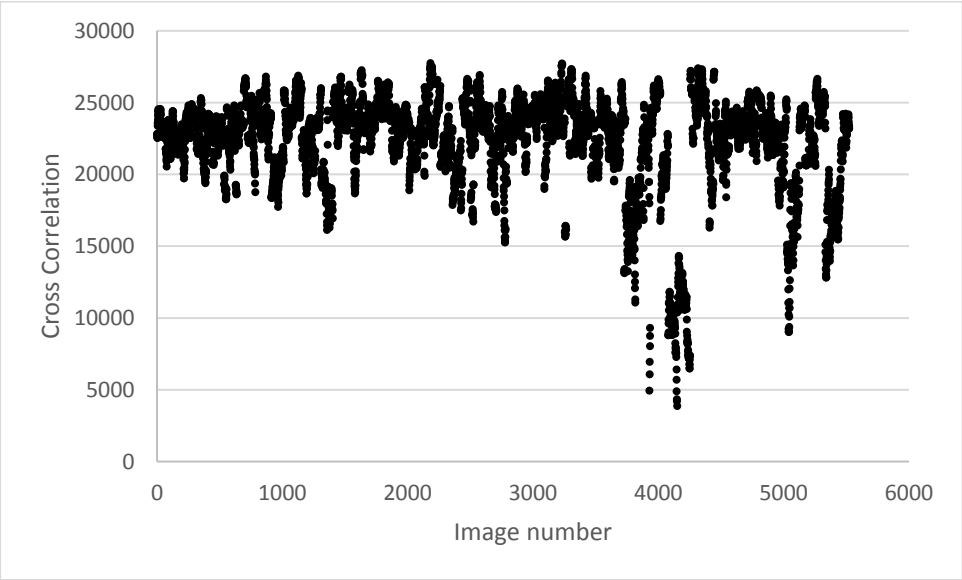


Figure 4.5.6: Cross correlation scores for each image with its best matching reference projection in the final cycle of IHRSR.

The cross-correlation of each image with its best matching reference projection is shown in figure 4.5.6.

The cross-correlation mean for the layer-line filtered dataset is 22355.9 ± 93.8 (95% confidence interval). The standard deviation was 3410.8. 95% of the images had a cross-correlation score of

15 000 or more with the best matching reference projection. The images having a cross-correlation of less than 15 000 were predominantly 3729-3735, 3754-3761, 3793-3797, 3811-3818, 4080-4255, 5029-5052, 5077-5080, 5336-5346 and 5367-5372. The distribution of cross-correlation scores is similar to that seen for previous reconstructions, but the mean cross-correlation score is much higher.

4.5.3. Reconstruction using 3-fold cyclic symmetry

Cyclic symmetry was suspected from the initial reconstruction, which had an asymmetric unit of three subunits. To test this, three-fold cyclic symmetry was imposed during reconstructions from the dataset containing only single-walled filaments.

Reconstructions starting from symmetries of $\Delta\phi$ 17.0 and 18.0 degrees, and Δz 8.0 and 9.0 Å respectively, converged to 17.94 degrees and 9.12 Å (figure 4.5.9).

A significant change is seen in the subunit appearance after imposing 3-fold cyclic symmetry (C_3) (see figure 4.5.7). A central groove runs through the subunits, as in the original reconstruction. However, the extensions of each subunit flanking the central groove are shifted with respect to one another in the C_3 reconstruction. As a result, several other sets of parallel helices are visible.

The power spectrum of the projection of the final volume mirrors this (figure 4.5.8). Strong reflections are seen along layer lines at positions 1/185.3 Å and 1/90.5 Å, which were not seen on these layer lines in the power spectra from previous reconstructions.

The helices seen on the 3D reconstruction are, in addition to the strong 3- and 6-start right handed helices, a 21-start right handed helix, with a pitch of 185 Å, corresponding to the major reflection on the first layer line from the equator. There is an 18-start left-handed helix, with a pitch of ~90 Å, corresponding to the major reflection on the second layer line from the equator. These helices are not seen in the reconstructions generated without 3-fold cyclic symmetry.

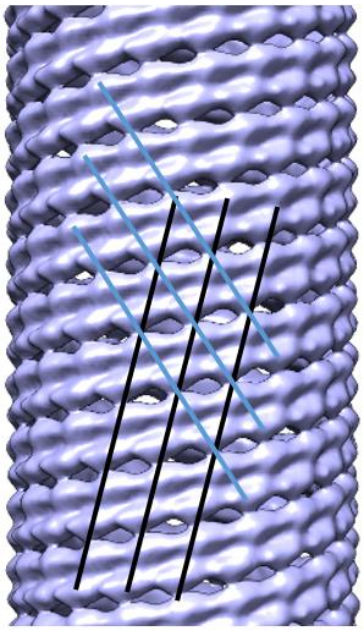


Figure 4.5.7: Final volume after 50 cycles of the reconstruction made using 3-fold cyclic symmetry. In addition to the prominent 3- and 6-start right-handed helices, an 18-start left-handed helix (blue lines) of z-spacing ~90 Å, and a 21-start right-handed helix (black lines) are visible.

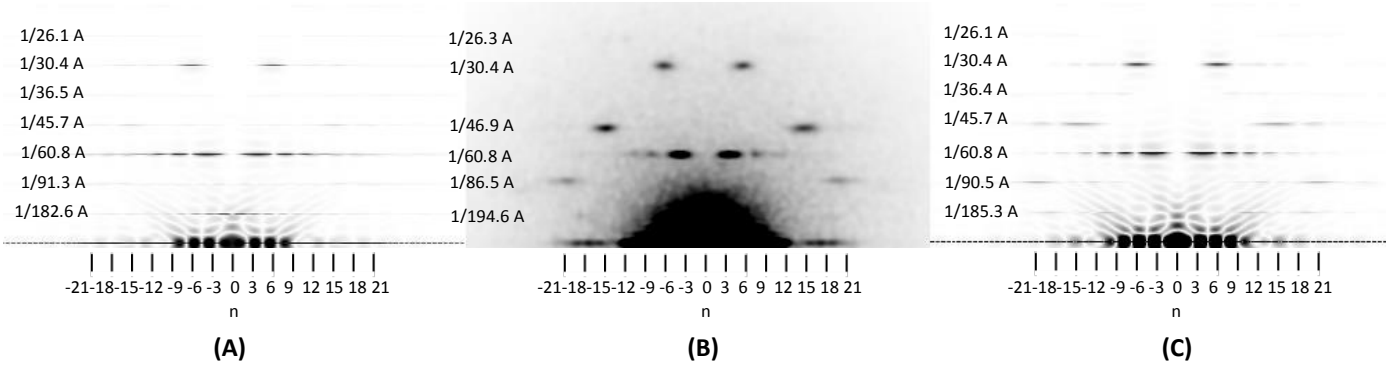


Figure 4.5.8: Comparison of 2D power spectra generated from (A) a projection of the original reconstruction, (B) centred, aligned segments, and (C) a projection of the reconstruction generated using 3-fold cyclic symmetry. Layer lines are seen at 1/185.3, 1/90.5, 1/60.8, 1/45.7, 1/36.4, 1/30.4 and 1/26.1 Å in (C), and in (A) and (B) there are minor variations in the heights of layer lines 1, 2 and 4. The power spectrum in (C) better matches the average power spectrum of the image segments (B) than does (A).

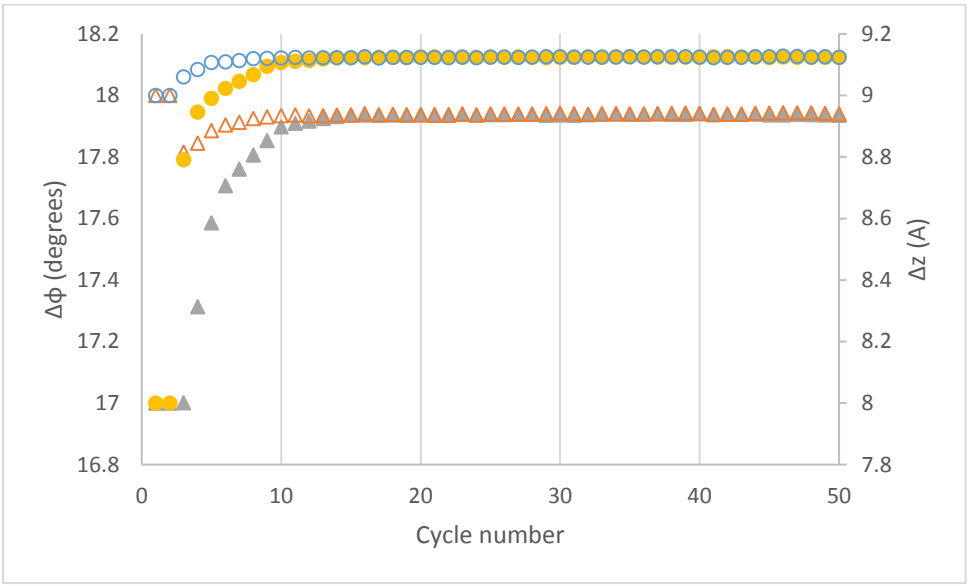


Figure 4.5.9: Refinement of helical symmetry parameters by IHRSR. The first reconstruction started from Δz = 8.0 Å (filled circles), and Δϕ = 17.0° (filled triangles). The second reconstruction started from Δz = 9.0 Å (open circles), and Δϕ = 18.0° (open triangles). Both reconstructions converged on Δz = 9.12 Å and Δϕ = 17.93°, by cycle 15.

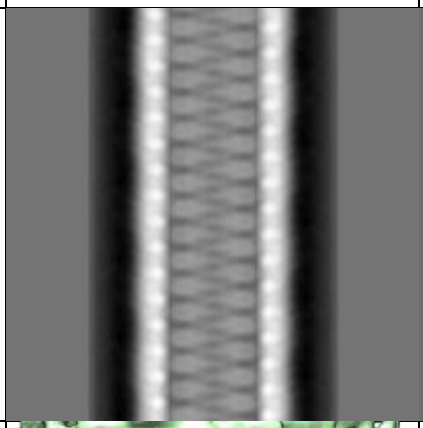
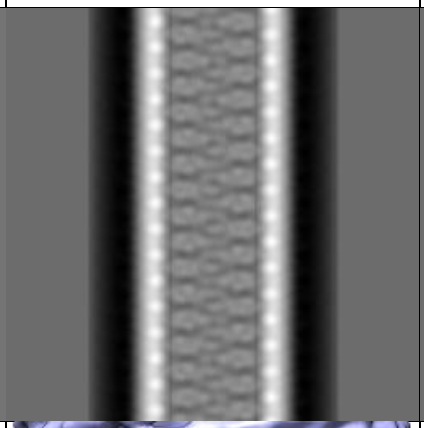
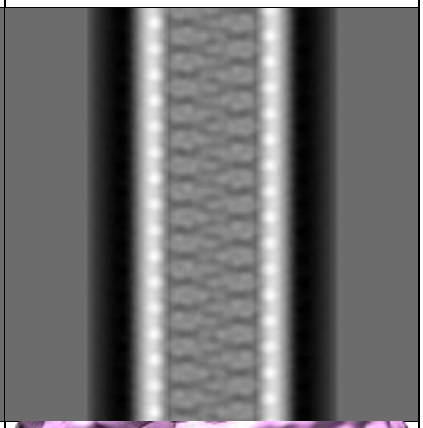
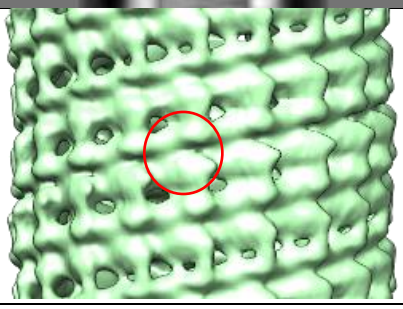
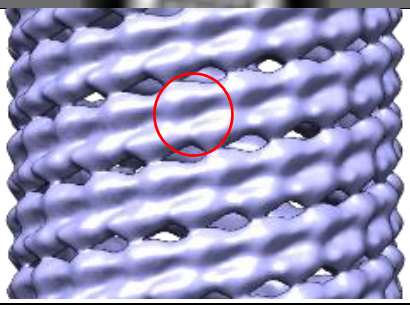
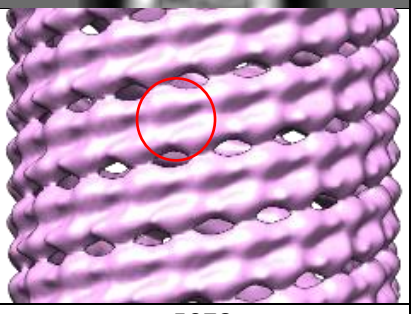
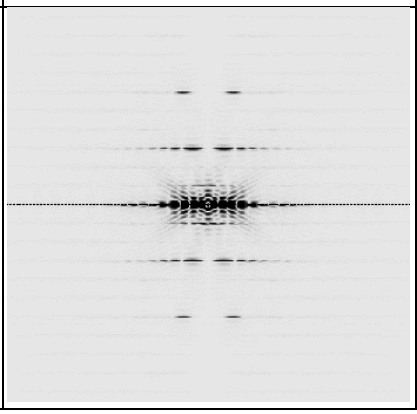
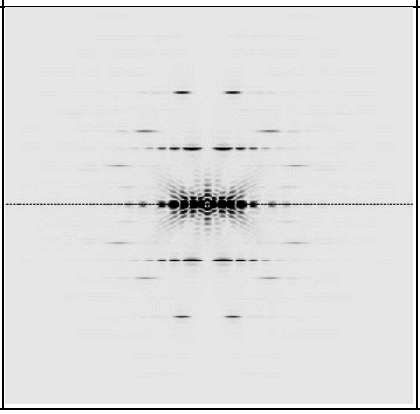
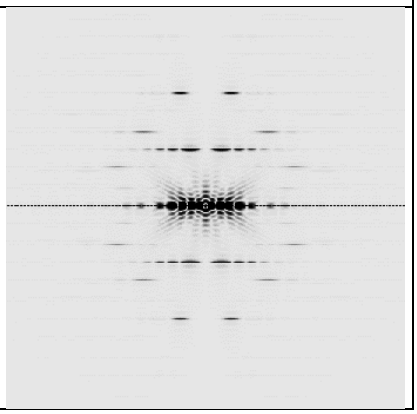
NAME	SINGLE-WALLED TUBES RECONSTRUCTION	3-FOLD CYCLIC SYMMETRY (A)	3-FOLD CYCLIC SYMMETRY (B)
2D PROJECTION OF SYMMETRIZED FINAL VOLUME			
3D RECONSTRUCTION			
NUMBER OF IMAGES	5078	5078	5078
CYCLIC SYMMETRY (FOLD)	1	3	3
STARTING SYMMETRY	17.0° 8.0 Å	17.0° 8.0 Å	18.0° 9.0 Å
FINAL SYMMETRY	17.97° 9.13 Å	17.94° 9.12 Å	17.94° 9.12 Å
POWER SPECTRUM OF ELONGATED MODEL PROJECTION			

Table 4.5.4: Comparisons of the reconstruction made from the single-walled filament segments with the reconstruction made from the same segments, with 3-fold cyclic symmetry imposed. Row 1 – 2D projections of the final volumes after 50 cycles of IHRSR. Row 2 – Surface representations of the final volumes. A single subunit is circled in red. The contour level of the reconstructions was set to show continuous density along the z-axis. Row 3-6 – Summary statistics for (3) the number of segments used in the cycles of each reconstruction, (4) the order of cyclic symmetry imposed, (5) and (6) the initial and final symmetry parameters respectively of rotation (degrees) and rise (A) per subunit after 50 cycles of the IHRSR process. Row 7 – Power spectra calculated from elongated 2D projections of the symmetrized final reconstruction. Layer line positions are shown.

Quality assessment of the reconstruction

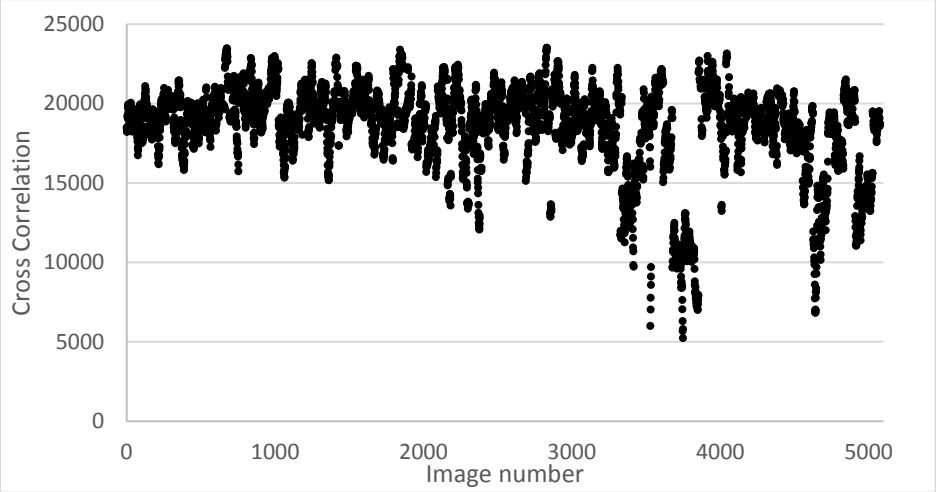


Figure 4.5.10: The cross-correlation between each image in the filtered dataset and its best matching reference projection in the final cycle of IHRSR.

Figure 4.5.10 shows the distribution of cross-correlation scores of each image with its best matching reference projection. The mean cross-correlation for the dataset was 18513.86 ± 75.63 (95% confidence interval). The standard deviation was 2749. 89% of the images had a cross-correlation score of 15 000 or more with the best matching reference projection. The distribution of cross-correlation scores for each image is similar to that seen in previous reconstructions, where 3-fold cyclic symmetry was not imposed. The images having a cross-correlation of less than 15 000 were predominantly 3328-3440, 3530-3535, 3679-3853, 4560-4723, and 4911-5022.

4.5.4. Final Reconstruction using a combined approach

Following the results of the above steps to improve the initial reconstruction, a final reconstruction was made using all the above approaches simultaneously. Images included segments with only single walls, and were filtered using the layer-line mask. 3-fold cyclic symmetry was imposed during the reconstruction. Furthermore, in each cycle, images with a low cross correlation with the reference projections (more than two standard deviations below the mean cross correlation) were excluded from the reconstruction.

The reconstructions A and B converged on Δz 9.14 and 9.15 Å, and Δϕ 17.98 and 17.98 degrees respectively, after 27 cycles (figure 4.5.14).

The reconstruction is very similar to the previous reconstruction generated using 3-fold cyclic symmetry, and the single-walled filament stack (see Table 4.5.5). The subunits have a central groove, and two flanking extensions, which are slightly offset from one another. The three- and six-start right handed helices are clearly visible, as are the 18-start left-handed helix and the 21-start right-handed helix. These correspond to the strong reflections on layer lines 3, 6, 2 and 1 respectively. Additionally, a left-handed 15-start helix is present, with a pitch of ~45 Å. It corresponds to the strong reflection shown on layer line 4, at 1/45.7 Å (see figure 4.5.13).

The power spectrum of a projection of the reconstruction is very similar to that of the previous reconstruction where 3-fold cyclic symmetry was imposed (see figure 4.5.12). Layer lines are present at positions 1/185.3, 1/90.5, 1/60.8, 1/45.7, 1/30.4 and 1/26.1 Å. The intensities of the major reflections on these layer lines are very similar for both the previous reconstruction, and the final reconstruction.

Few changes to the reconstruction or power spectrum were seen after the addition of the layer-line mask and the filter for removal of images with low cross-correlation with reference projections, when compared with the previous reconstruction generated from the single-walled filament dataset, using 3-fold cyclic symmetry.

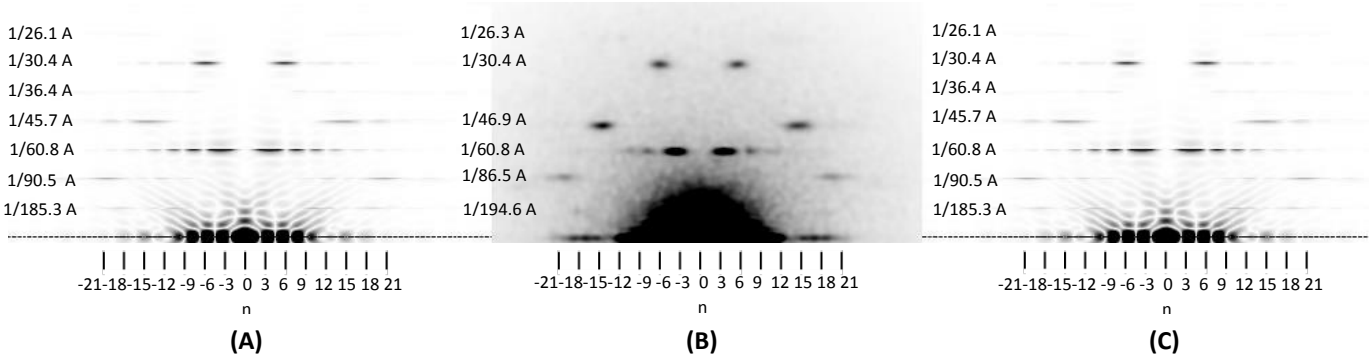


Figure 4.5.12: Comparison of 2D power spectra generated from (A) a projection of the final reconstruction, (B) centred, aligned filament segments and (C), a projection of the previous reconstruction, generated using 3-fold cyclic symmetry, from the single-walled filament dataset. Layer line positions are conserved between (A) and (C), and the intensities of the major reflections on layer lines 1, 2, 3, 4 and 6 are very similar.

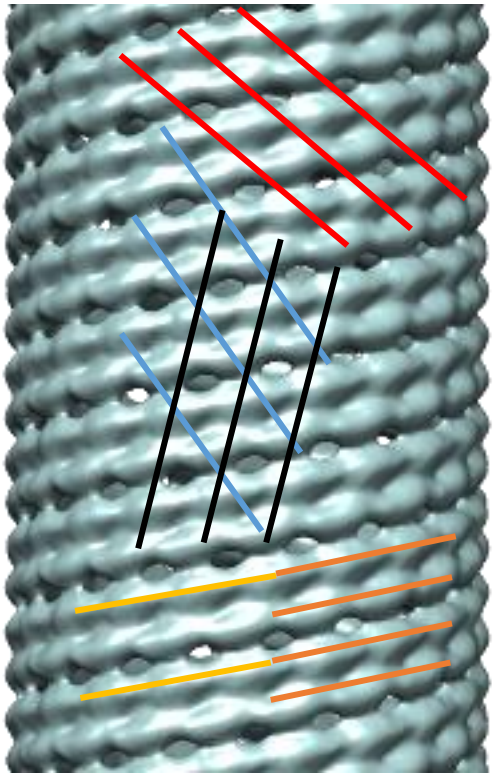


Figure 4.5.13: Final volume of the final reconstruction. The reconstruction was made from single-walled filament segments, filtered using layer-line mask B, and 3-fold cyclic symmetry was imposed. In addition, a filter was applied to remove images with cross-correlation scores of more than two standard deviations below the mean. Prominent 3- and 6- start right-handed helices are shown by yellow and orange lines respectively. A 15-start left-handed helix is shown in red, and has z-spacing of 45.7 Å between parallel helices. The 18-start left-handed helix is shown in blue. A right-handed 21 start helix is shown in black.

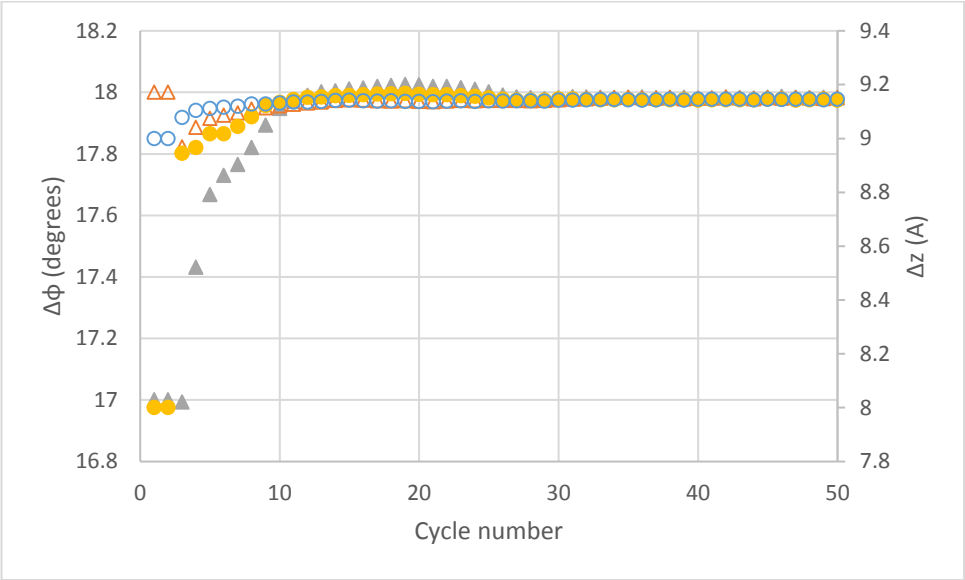


Figure 4.5.14: Refinement of helical symmetry parameters of the final reconstruction by IHRSR. The first reconstruction started from Δz = 8.0 Å (filled circles), and Δϕ = 17.0° (filled triangles). The second reconstruction started from Δz = 9.0 Å (open circles), and Δϕ = 18.0° (open triangles). Reconstructions A and B converged on Δz 9.14 and 9.15 Å, and Δϕ 17.98 degrees respectively, after 27 cycles.

Quality assessment

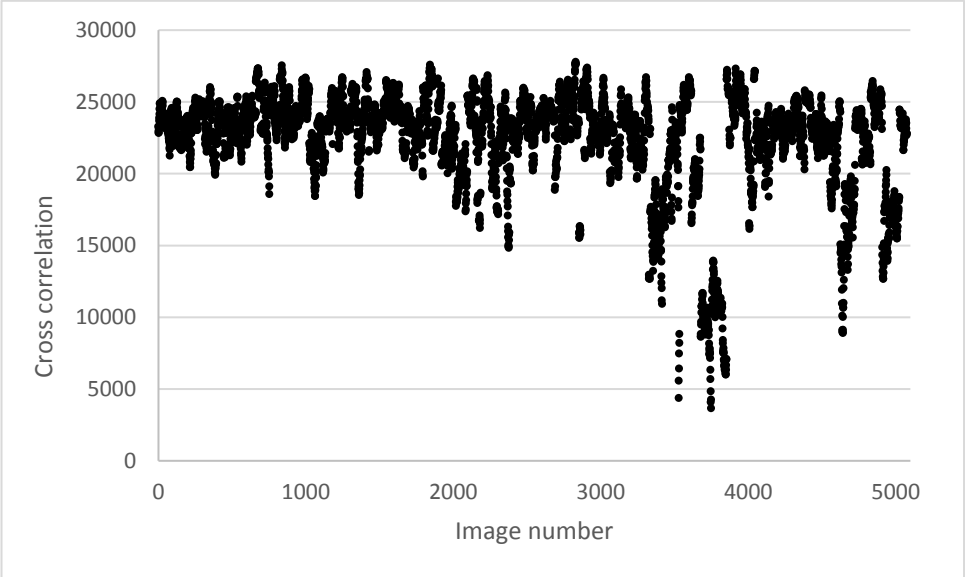


Figure 4.5.15: The cross-correlation between each image in the final dataset and its best matching reference projection in the final cycle of IHRSR

The cross-correlation between each image and its best matching reference projection in the final cycle is shown in figure 4.5.15. The mean cross-correlation score for the dataset was 22345.52 ± 98.64 (95% confidence interval). The standard deviation was 3585.6. 95% of the images had a cross-correlation score of 15 000 or more with the best matching reference projection. The images having a cross-correlation of less than 15 000 were predominantly 3328-3353, 3389-3417, 3530-3535, 4628-4651, 4675-4683 and 4911-4921 and 4943-4946. The distribution of cross-correlation scores is similar to that seen for all previous reconstructions.

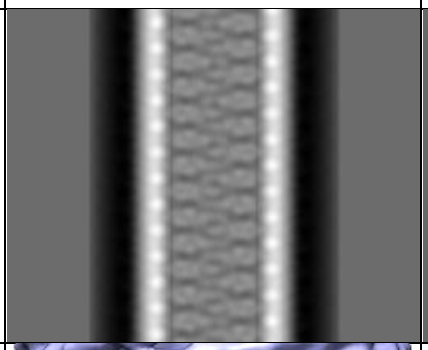
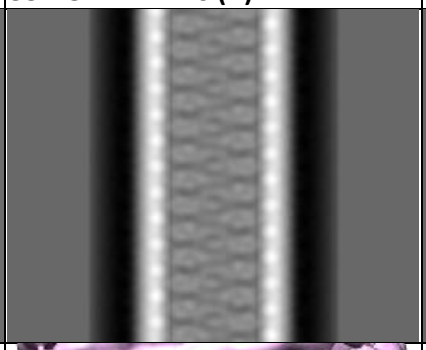
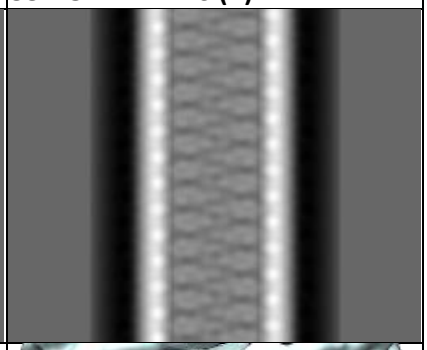
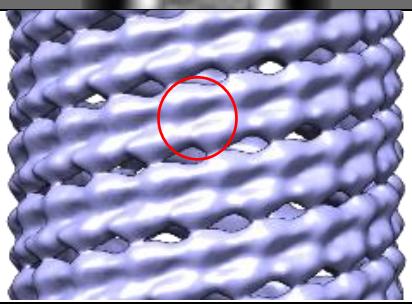
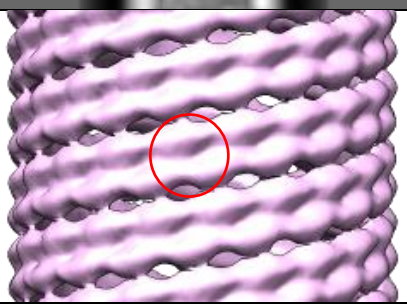
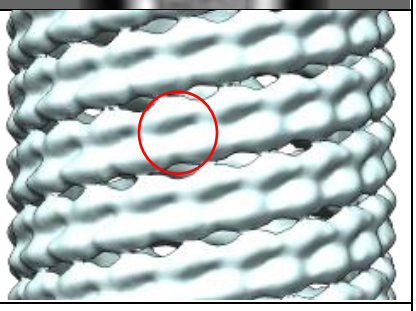
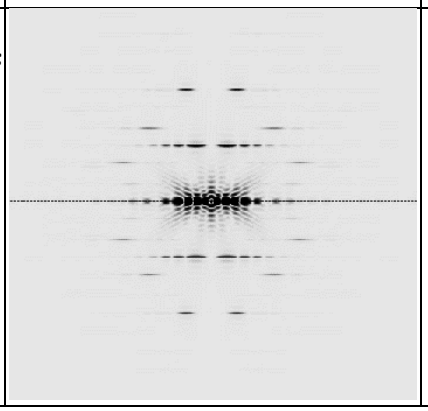
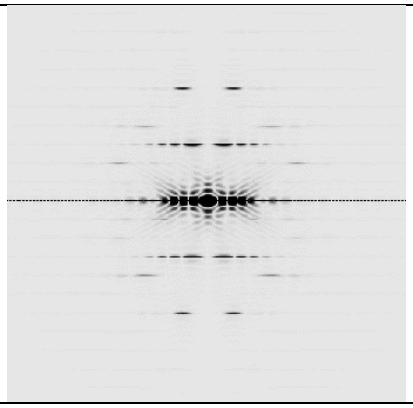
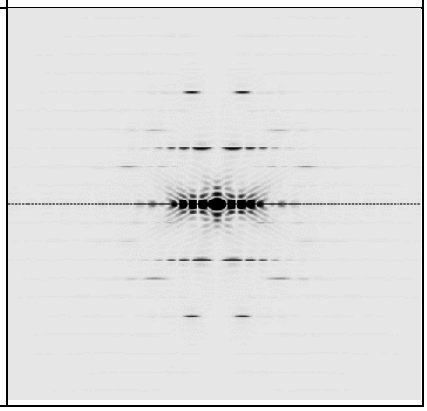
NAME	3-FOLD CYCLIC SYMMETRY (A)	LAYER-LINE AND CORRELATION CUT-OFF FILTERS (A)	LAYER-LINE AND CORRELATION CUT-OFF FILTERS (B)
2D PROJECTION OF SYMMETRIZED FINAL VOLUME			
3D RECONSTRUCTION			
NUMBER OF IMAGES	5078	5078	5078
CORRELATION CUT-OFF	None	Two standard deviations below the mean	Two standard deviations below the mean
STARTING SYMMETRY	17.0° 8.0 Å	17.0° 8.0 Å	18.0° 9.0 Å
FINAL SYMMETRY	17.94° 9.12 Å	17.98° 9.14 Å	17.98° 9.15 Å
LAYER LINE FILTER	No	Layer line mask B	Layer line mask B
POWER SPECTRUM OF ELONGATED MODEL PROJECTION			

Table 4.5.5: Comparison of the reconstruction made using 3-fold cyclic symmetry with the two final reconstructions incorporating the previous steps to improve the reconstruction, as well as a cut-off filter to remove images with a cross-correlation with best matching reference projections of more than 2 standard deviations below the mean. Row 1 – 2D projections of the final volumes after 50 cycles of IHRSR. Row 2 – Surface representations of the final volumes. A single subunit is circled in red. The contour level of the reconstructions was set to show continuous density along the z-axis. Row 3-7 – Summary statistics for (3) the number of segments used in the cycles of each reconstruction, (4) the presence of a low-correlation cut-off filter, (5) and (6) the initial and final symmetry parameters respectively of rotation (degrees) and rise (A) per subunit after 50 cycles of the IHRSR process, and (7) - the use of a layer-line filter. Row 8 – Power spectra calculated from elongated 2D projections of the symmetrized final reconstruction. Layer line positions are shown.

4.6. Rigid-body fitting for handedness estimation

To assess the possible handedness of the improved reconstructions made using selection rule 2, rigid-body fitting was performed using several homology models of MsMca [appendix], which were fit into two reconstructions from section 4.5. Each reconstruction was also mirrored in the Z-axis, to provide reconstructions of the opposite handedness. The UCSF Chimera *Fit-in-map* tool was used to semi-automatically fit the monomers into the reconstruction maps. For each subunit, a map was simulated from the atomic model, at a resolution of 20 Å, and the fit was achieved by optimizing the mean cross-correlation between the simulated map and the reconstruction map. The threshold of the map was set to enclose a volume of $4x10^4$ $Å^3$ per subunit, from the mass of MsMca (33 000 Da), and a protein density of 0.73 cm^3/g [4].

The most favourable handedness was estimated by comparison of the interfaces generated by fitting homology models of MsMca into the reconstructions of original and reversed handedness.

4.6.1. Final reconstruction

Through fitting several copies of the MsMca homology model into the map of the final reconstruction, the principal A interface along the 3-start helix was confirmed, and a possible B interface was modelled between monomers on adjacent parallel helices (figure 4.6.1 and figure 4.6.2).

The A interface contained many steric clashes, as the interface regions in chain A and B were brought close together. The clashing regions are shown in figure 4.6.3.

Possible favourable interactions include a potential hydrogen bond between H58 (chain A) and E167 (chain B) (figure 4.6.3), and possible hydrophobic interactions between W169 (chain B) and M52 (chain A), L54 (chain A), P55 (chain A), and V57 (chain A), respectively. F201 (chain A) may also participate in a hydrophobic interaction with P274 (chain B) or L277 (chain B), respectively.

The B interface contained two possible hydrophobic interactions (figure 4.6.4), between L161 (chain B) and W208 (chain C), and P163 (chain B) and W204 (chain C), respectively. A potential hydrogen bond was modelled between Y162 (chain B) and H207 (chain C). However, the potential interacting residues at the B-interface were too far apart for strong interactions (≥ 14Å).

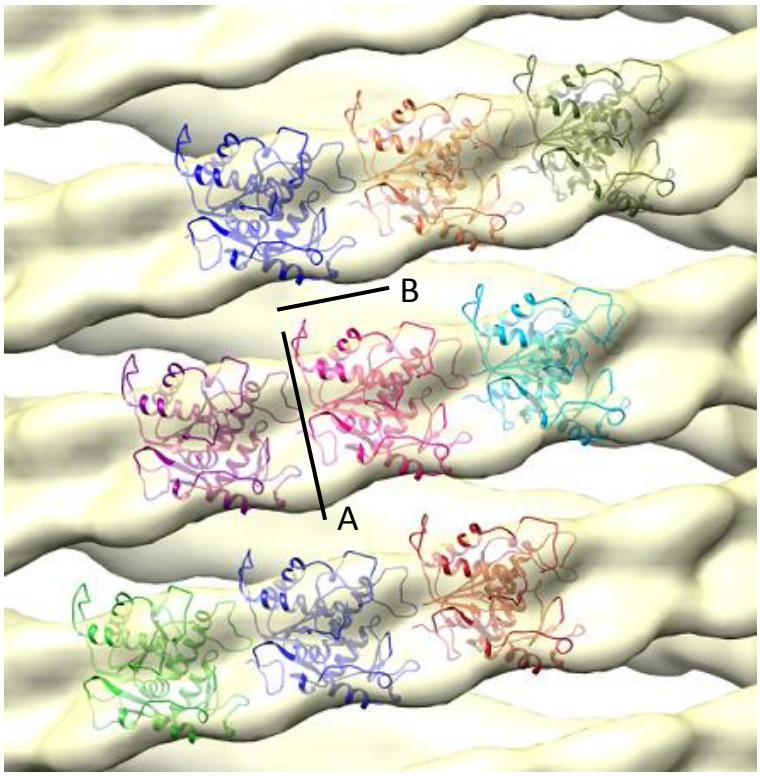


Figure 4.6.1: MsMca atomic homology models fit into the final reconstruction, which has a right-handed 3-start helix. The principle interfaces are indicated by black lines. The A interface occurs along the 3-start helix, while the B interface occurs between parallel helices.

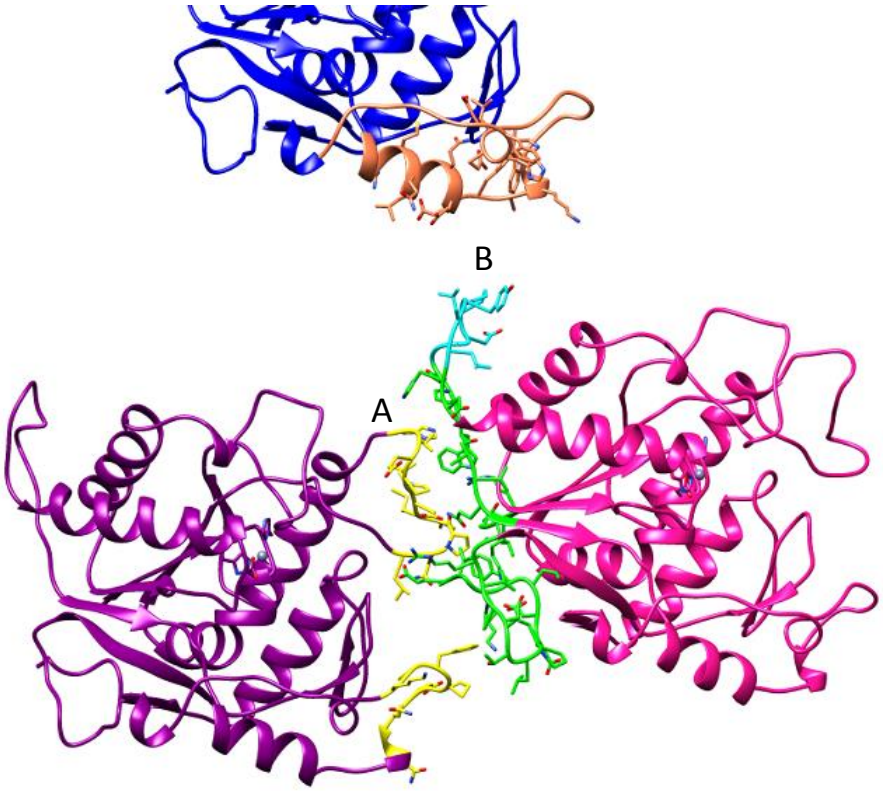


Figure 4.6.2: The A and B interfaces between adjacent MsMca homology models in the right-handed reconstruction. Chains A, B and C are shown in purple, pink and blue, respectively. The A Interface regions are shown in gold (chain A) and green (chain B). The A interface residues in chain A are from 48-58 and 195-203. The interface residues in chain B are 3-5, 158-173 and 271-287. The B interface residues are shown in cyan (chain B) and coral (chain C). The B interface residues in chain B are from 160-166. The B interface residues in chain C are from 182-215.

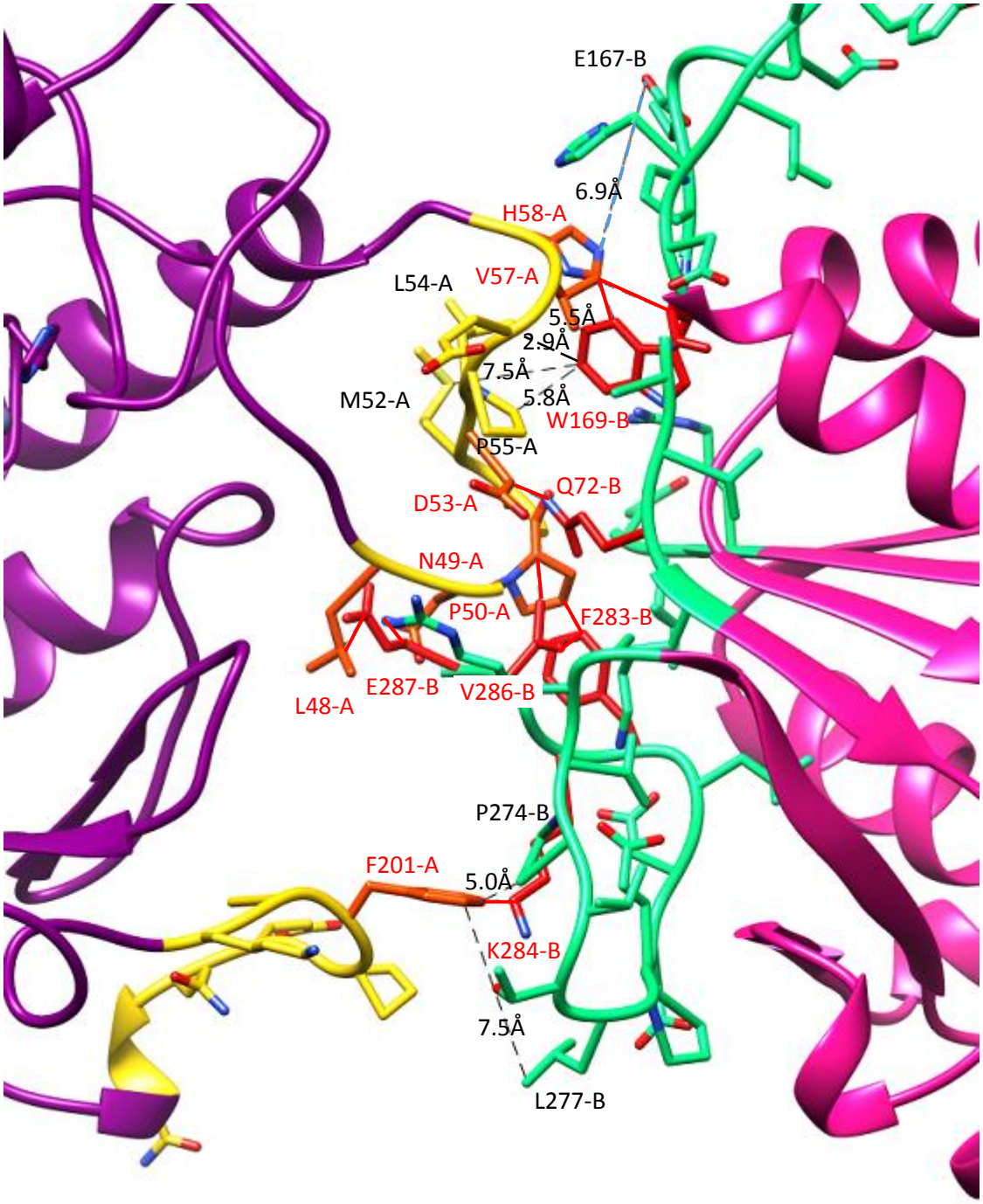


Figure 4.6.3: Potential interactions along the A interface, between chain A (residues 48-58 and 195-203, shown in gold) and chain B (residues 3-5, 158-173 and 271-287, shown in green). Steric clashes are shown in red, and the residues involved in steric clashes are shown in orange (chain A) and red (chain B). A possible hydrogen-bonding interaction is shown as a blue dashed line with the distance shown in Å. Possible hydrophobic interactions are shown as black dashed lines, with the distances in Å.

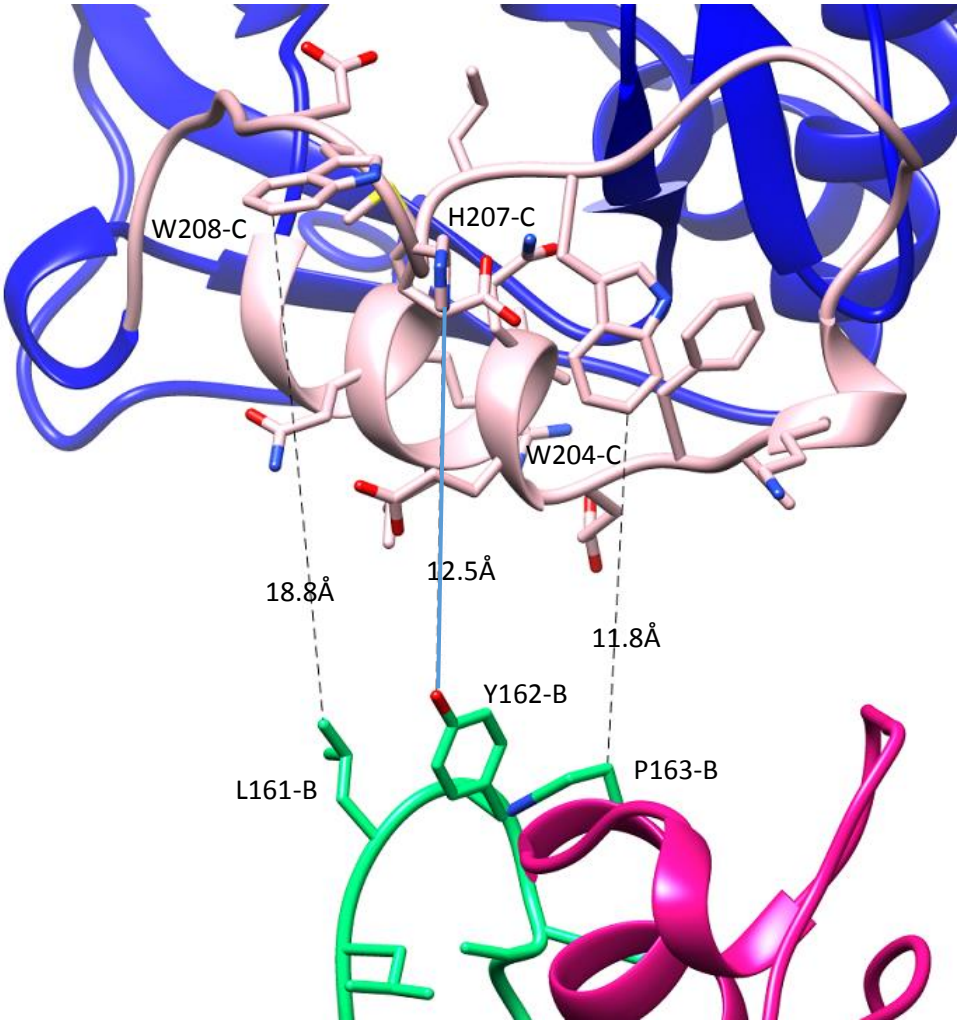


Figure 4.6.4: Potential interactions at the B interface, between chain B (dark pink) and chain C (blue). Interface regions are shown in green (chain B, residues 160-166) and light pink (chain C, residues 182-215). Possible hydrogen bonds are shown by blue lines. Potential hydrophobic interactions are shown by black dashed lines. Distances are given in Å.

4.6.2. Reconstruction of reversed handedness

Fitting MsMca homology models into the reconstruction of reversed handedness confirmed the primary A interface along the three-start helix and a possible B interface was modelled between monomers on adjacent parallel helices. These are shown in figures 4.6.5 and 4.6.6.

The A interface included many steric clashes, shown in figure 4.6.7. These are shown between residues E287 (chain B) and L48 (chain A) and P253 (chain A) respectively, P50 (chain A) and E3 (chain B), L4 (chain B) and H126 (chain B) respectively, W169 (chain B) and D53 (chain A), P55 (chain A) and V57 (chain A) respectively, and L54 (chain A) and R124 (chain B).

Several possible hydrogen bonds at the A interface are shown in figure 4.6.8. These include hydrogen bonds between P50 (chain A) and R5 (chain B), D53 (chain A) and R124 (chain B), H58 (chain A) and E167 (chain B), and D53 (chain A) and Q172 (chain B). Possible hydrophobic interactions are seen between W169 (chain B) and P55 (chain A) and V57 (chain A), respectively, and F283 (chain B) and P50 (chain A) (figure 4.6.9).

The potential interacting residues at the B-interface were too far apart for strong interactions (>8Å, see figure 4.6.8).

Both right- and left-handed reconstructions showed an A interface with a large number of steric clashes. The possible favourable interactions at the A interface in both left- and right-handed reconstructions were similar in number. Therefore, the preferred handedness of the final reconstruction, as determined from the rigid-body fitting of MsMca homology models into the reconstruction, is not known.

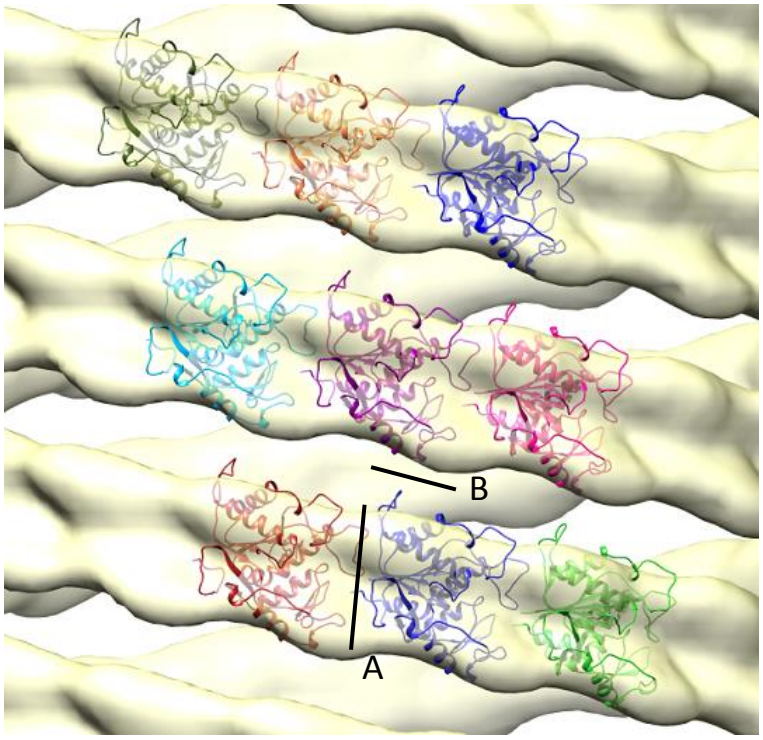


Figure 4.6.5: MsMca atomic homology models fit into the final reconstruction, which has a left-handed 3-start helix. The principle interfaces are indicated by black lines. The A interface occurs along the 3-start helix, while the B interface occurs between parallel helices.

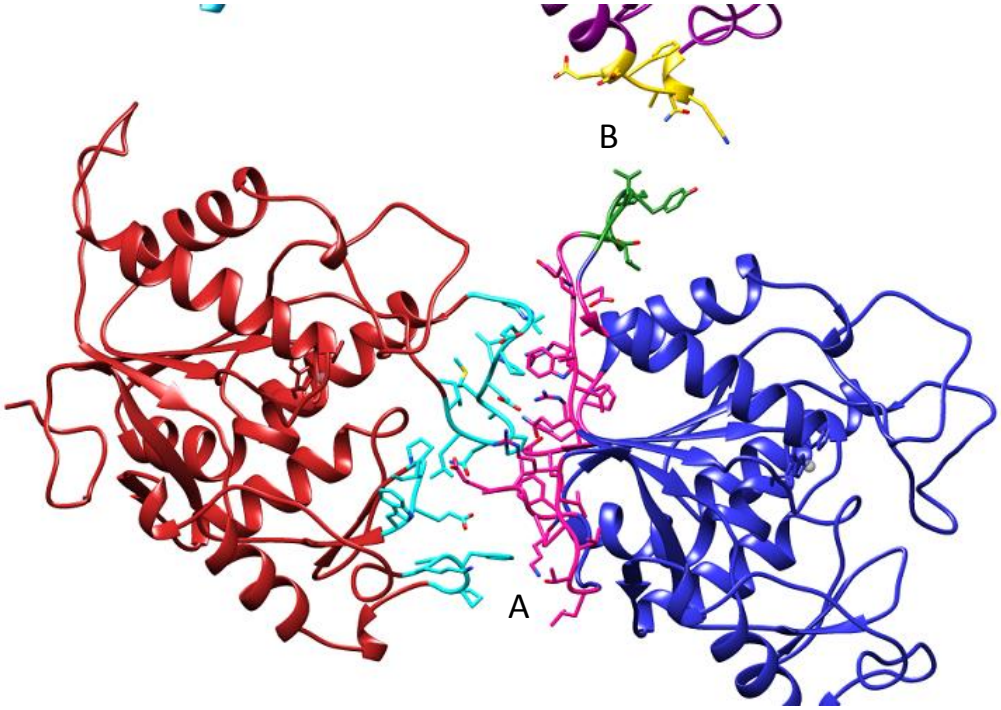


Figure 4.6.6: The A and B interfaces between adjacent MsMca homology models in the left-handed reconstruction. The A interface regions are shown in cyan (chain A) and pink (chain B). The B interface regions are shown in green (chain B) and gold (chain C). The A interface residues in chain A are from 47-59, 199-203 and 253-256. The A interface residues in chain B are 3-5, 124-126, 157-158, 166-172, 272-277 and 283-287. The B interface residues in chain B are 160-165, and the B-interface residues in chain C are 190-196.

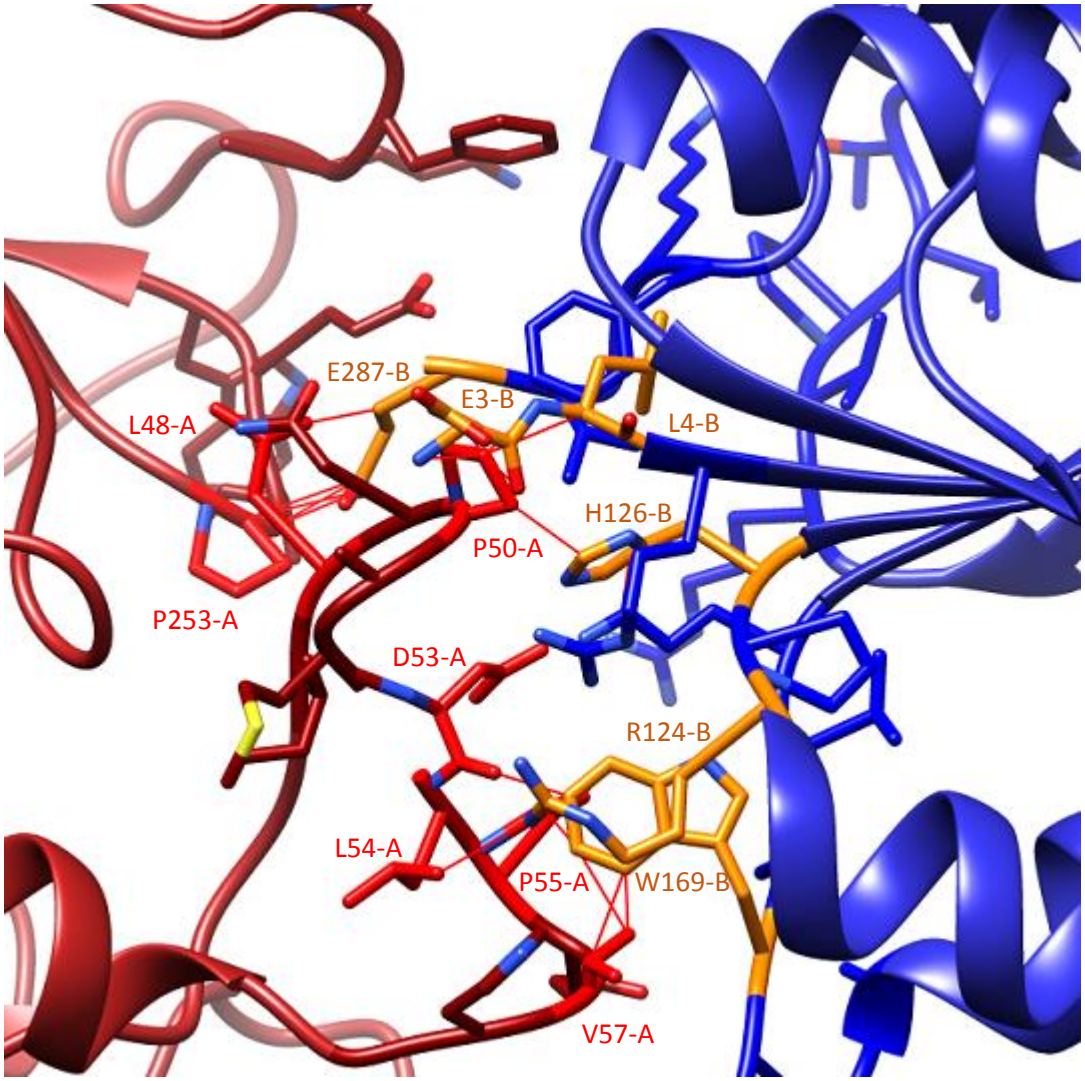


Figure 4.6.7: Steric clashes (red lines) between residues at the A interface between adjacent MsMca homology models in the left-handed reconstruction. Chains A and B are shown in dark red and blue respectively. Residues from chain A involved in steric clashes are shown in red. Residues from chain B involved in steric clashes are shown in orange.

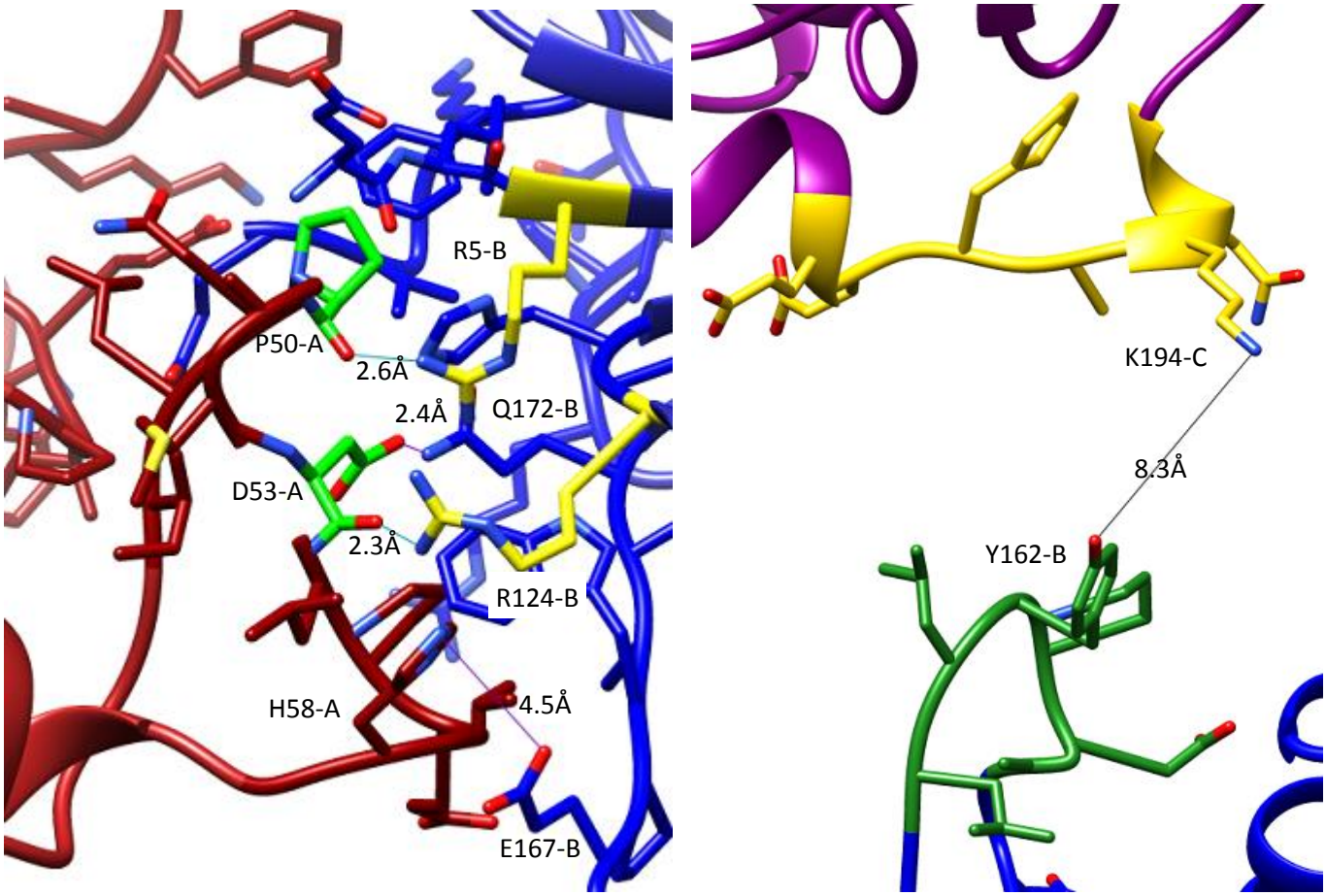


Figure 4.6.8: (A) Possible hydrogen bonds at the A interface between adjacent MsMca homology models. (B) The B-interface between adjacent MsMca homology models. Chains A and B are shown in dark red and blue respectively. Hydrogen bonds found using UCSF Chimera's *FindHBond* [3] tool are displayed as light blue lines. Residues from chain A involved in these hydrogen bonds are shown in green, and residues from chain B involved in these hydrogen bonds are shown in yellow. Other possible hydrogen bonds are shown as purple lines, and the residues involved are coloured in dark red (chain A) or in blue (chain B). The closest pair of residues at the potential B-interface are shown with the inter-residue distance. Potential interface residues are shown in green (chain B) and yellow (chain C).

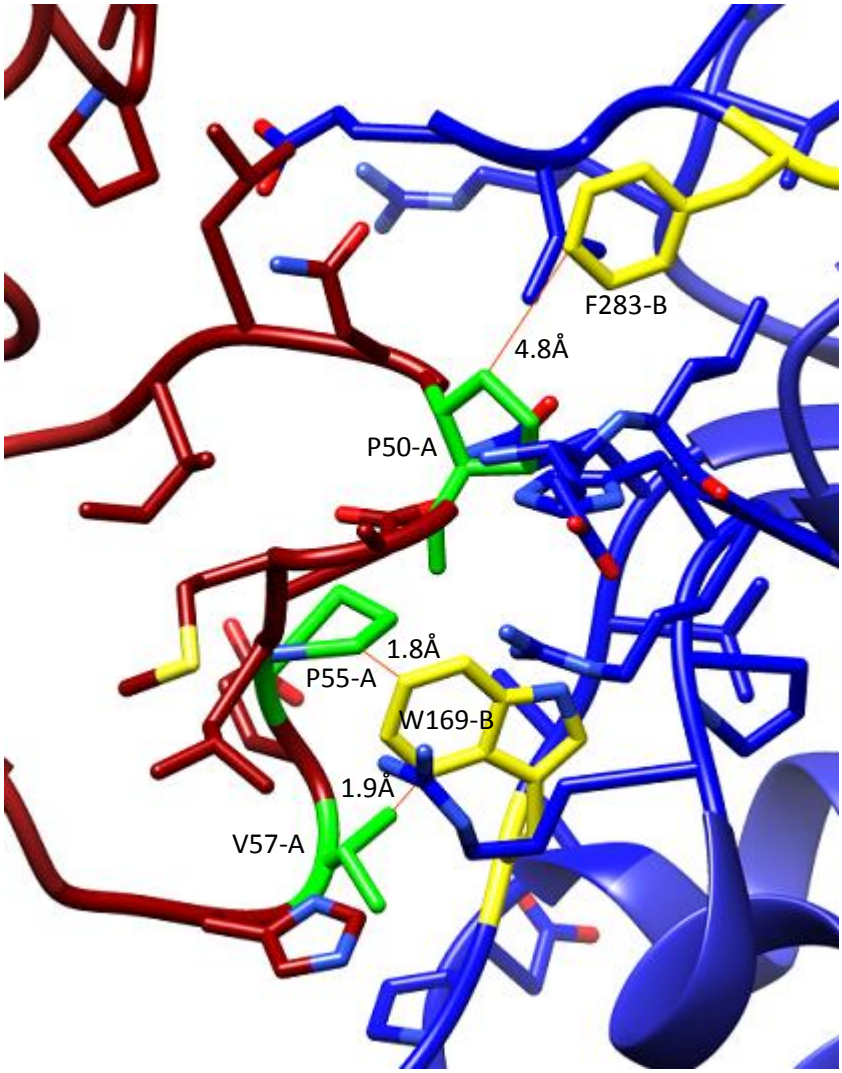


Figure 4.6.9: Possible hydrophobic interactions at the A interface between adjacent MsMca homology models. Chains A and B are shown in dark red and blue respectively. Residues involved in possible hydrophobic interactions are shown in green (chain A) or yellow (chain B). Possible hydrophobic interactions are shown as orange lines. The inter-residue distances are given in Å.

4.6.3. Single-walled filament reconstruction

A reconstruction from single-walled filament segments [see section 4.5.1] was also used in rigid-body fitting exercises. The single-walled filament reconstruction was used in the same approach as used for the final reconstruction [sections 4.6.1, 4.6.2].

The fitting exercise showed a primary A interface, along the 3-start helix, and a secondary B interface, across parallel helices (see figures 4.6.10 and 4.6.11).

Along the A interface, there are a small number of steric clashes (figure 4.6.12). These are between residues E3 (chain A) and D53 (chain B) and P55 (chain B) respectively, and E56 (chain B) and R124 (chain A) and W169 (chain A) respectively.

Hydrogen bonds are shown between the carbonyl O of P55 (chain B) and the NH of R5 (chain A) and R124 (chain A) respectively, the amido N of E3 (chain A) and a carboxyl O of D53 (chain B), a carboxyl O of E56 (chain B) and H126 (chain A) and R124 (chain A) respectively, and the NH of Q172 (chain A) and a carboxyl O of E56 (chain B) (figure 4.6.13).

Possible hydrophobic interactions are seen between W169 (chain A) and P95 (chain B), and between P168 (chain A) and P97 (chain B), though the inter-residue distances are too large (>8Å) for strong interactions (figure 4.6.14).

Residues across the B interface are too far apart (>7Å) for strong interactions (figure 4.6.15).

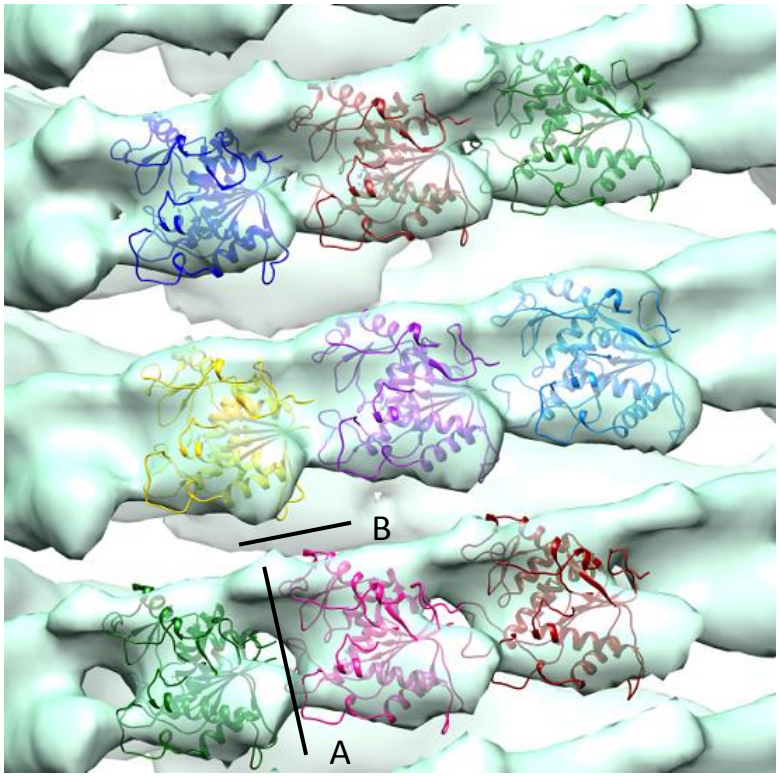


Figure 4.6.10: MsMca atomic homology models fit into the reconstruction from single-walled filaments, without the imposition of 3-fold cyclic symmetry. The 3-start helix is right-handed. The principle interfaces are indicated by black lines. The A interface occurs along the 3-start helix, while the B interface occurs between parallel helices.

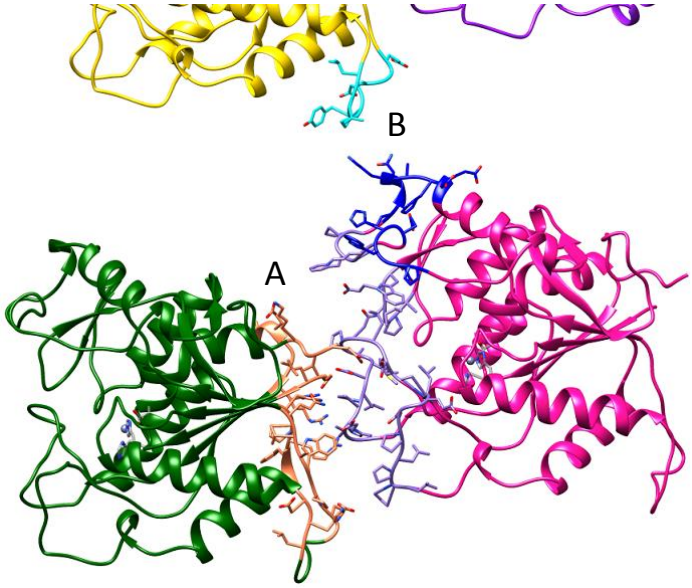


Figure 4.6.11: The A and B interfaces between adjacent MsMca homology models in the reconstruction from single-walled filaments. Chain A is shown in green, chain B in pink and chain C in yellow. The A interface regions are shown in coral (chain A) and purple (chain B). The B interface regions are shown in blue (chain B) and cyan (chain C). The A interface residues in chain A are from 3-5, 29-33, 122-126, 158-160, 166-172 and 284-287. The A interface residues in chain B are 46-58, 90-99, 199-203 and 252-256. The B interface residues in chain B are 190-197 and 206-210, and the B-interface residues in chain C are 160-167.

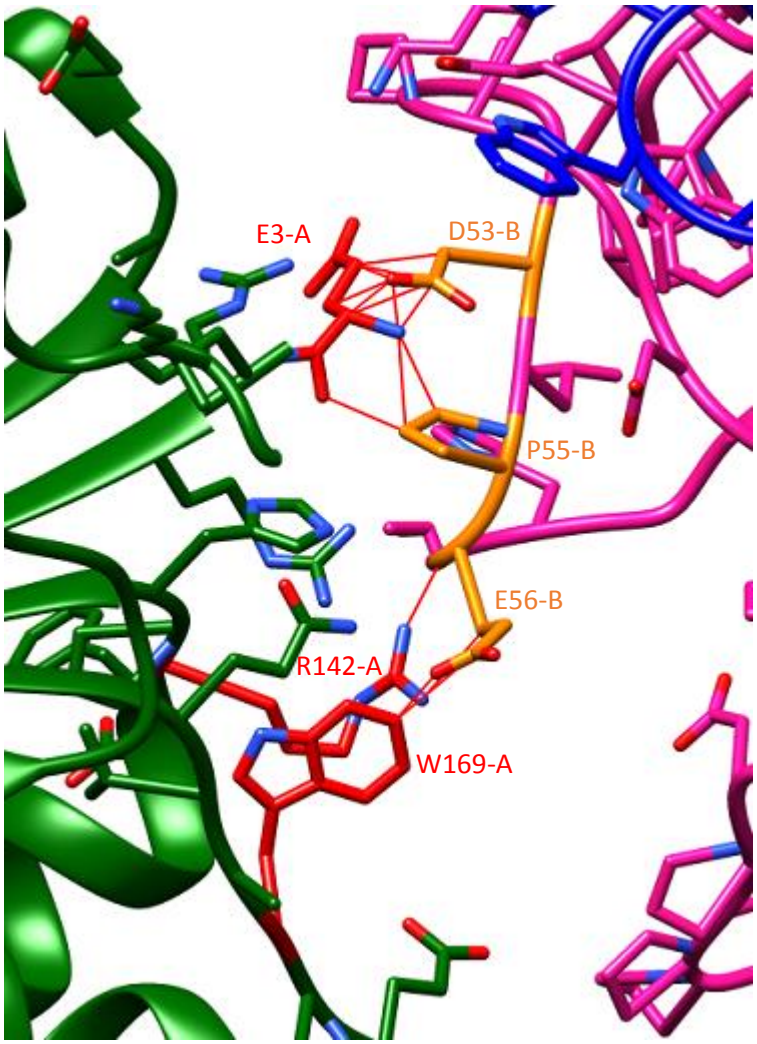


Figure 4.6.12: Steric clashes (red lines) between residues at the A interface between adjacent MsMca homology models in the reconstruction from single-walled filaments. Chains A and B are shown in green and pink respectively. Residues from chain A involved in steric clashes are shown in red. Residues from chain B involved in steric clashes are shown in orange. Residues 286-287 (chain A) are omitted for clarity.

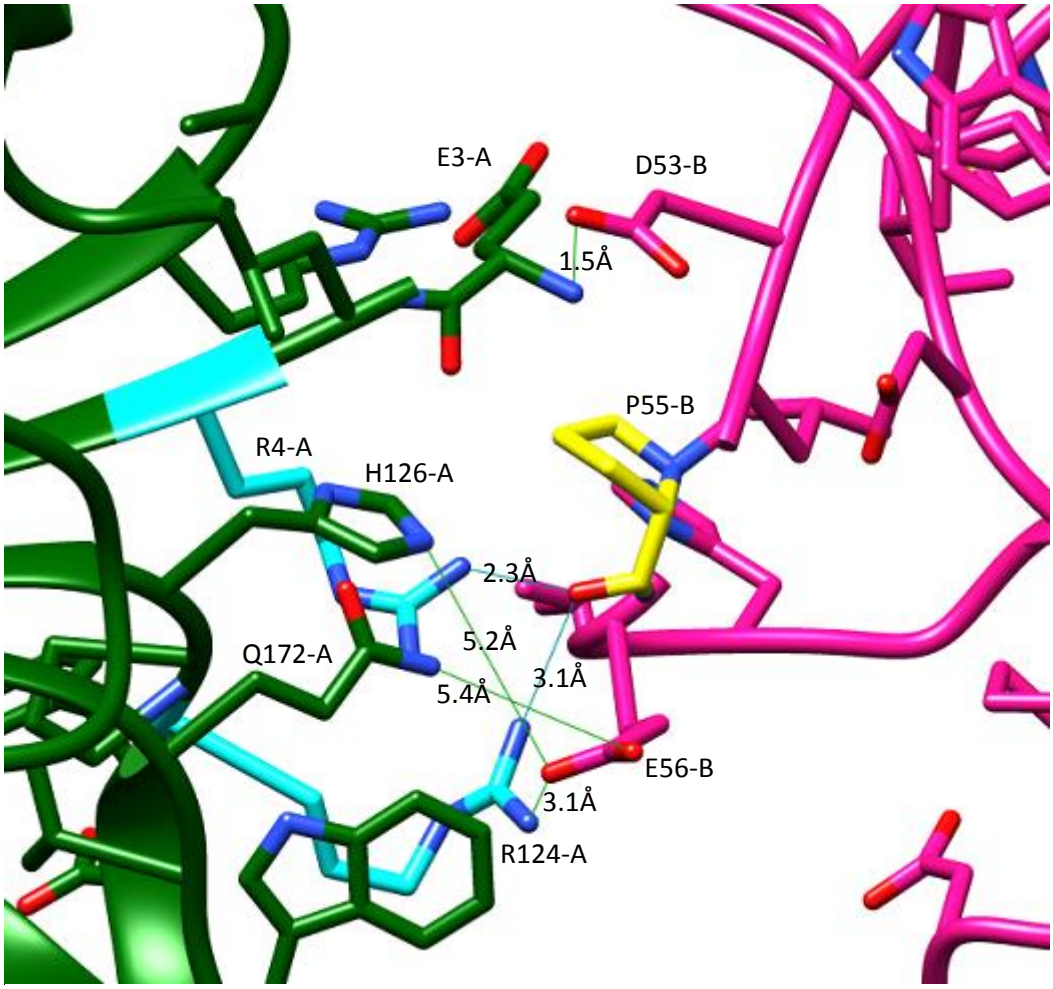


Figure 4.6.13: (A) Possible hydrogen bonds at the A interface between adjacent MsMca homology models in the reconstruction from single-walled filaments. Chains A and B are shown in dark green and pink respectively. Hydrogen bonds found using UCSF Chimera's *FindHBond* tool are displayed as light blue lines. Residues involved in these hydrogen bonds are shown in cyan (chain A), or yellow (chain B). Other possible hydrogen bonds are shown as green lines, and the residues involved are coloured in dark green (chain A) or in pink (chain B).

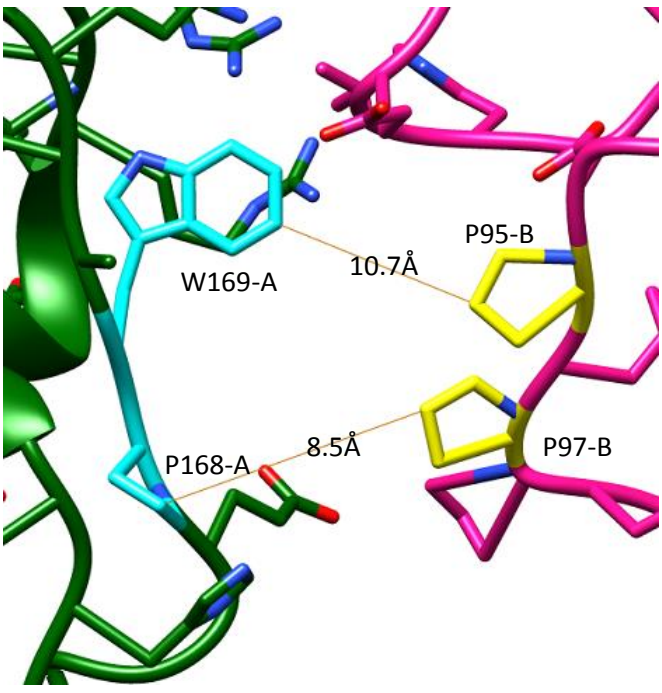


Figure 4.6.14: Possible hydrophobic interactions at the A interface between adjacent MsMca homology models in the reconstruction from single-walled filaments. Chains A and B are shown in dark green and pink respectively. Residues involved in possible hydrophobic interactions are shown in cyan (chain A) or yellow (chain B). Possible hydrophobic interactions are shown as orange lines. The inter-residue distances are given in Å.

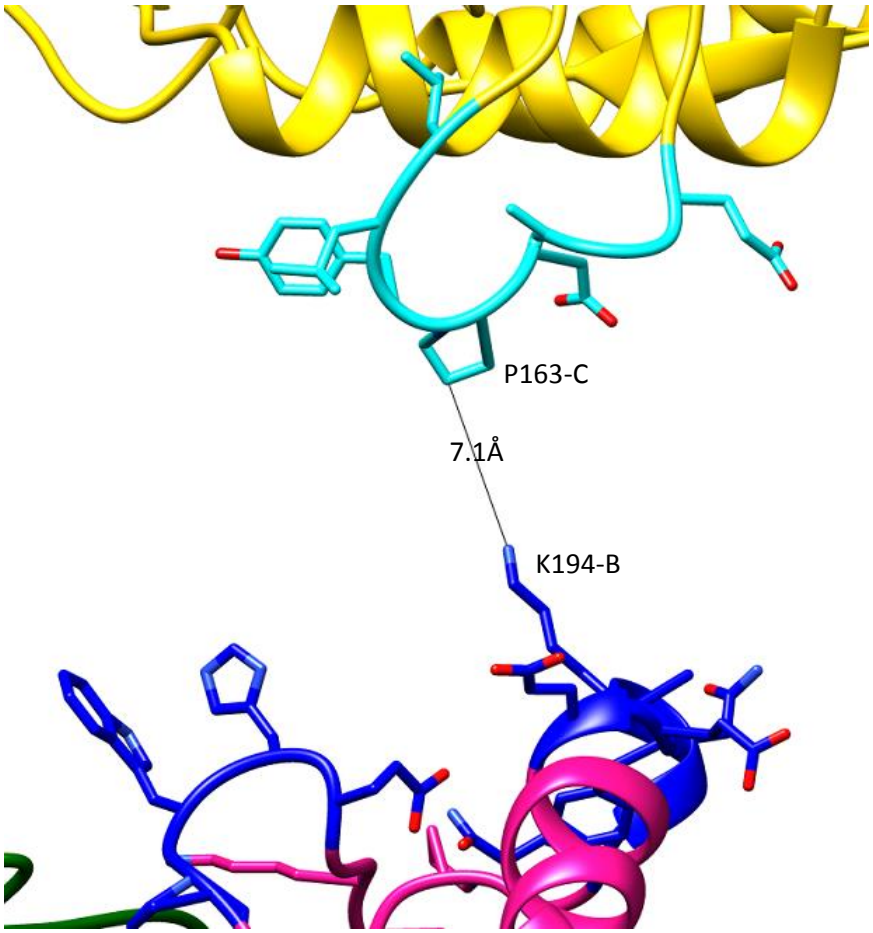


Figure 4.6.15: The B-interface between adjacent MsMca homology models in the reconstruction from single-walled filaments. Chains B and C are shown in pink and yellow respectively. The closest pair of residues at the potential B-interface are shown with the inter-residue distance. Potential interface residues are shown in blue (chain B) and cyan (chain C).

4.6.4. Reconstruction of reversed handedness

Fitting MsMca homology models into the reconstruction from single-walled filaments of reversed handedness confirmed the primary A interface along the left-handed three-start helix and a possible B interface was modelled between monomers on adjacent parallel helices (figure 4.6.16 and 4.6.17).

At the A interface, many steric clashes are present (figure 4.6.18). These are shown between residues T276 (chain A) and P200 (chain B), R272 (chain A) and L48 (chain B) and N49 (chain B) respectively, A170 (chain A) and A51 (chain B), and H58 (chain B) and D158 (chain A) and H159 (chain A) respectively.

Several possible hydrogen bonds are seen, between residues N49 (chain A) and R272 (chain B), H58 (chain A) and H159 (chain B) (figure 4.6.19). Additional hydrogen bonds are possible between N49 (chain A) and Q172 (chain B), E198 (chain A) and K284 (chain B) and H159 (chain B) and the carbonyl O of H58 (chain A), though these will require movement of the surrounding loops to bring the residues into H-bonding distance.

A number of potential hydrophobic interactions are seen, between residues W169 (chain A) and M52 (chain B), P274 (chain A) and P200 (chain B), and V275 (chain A) and F201 (chain B) (figure 4.6.20).

Residues on opposite chains at the B-interface are ≥ 9.7 Å apart, thus no strong interactions can be modelled (figure 4.6.20).

There are fewer steric clashes at the A interface in the original reconstruction with a right handed 3-start helix, than in the reconstruction with reversed handedness. The number of potentially favourable interactions across the interfaces is comparable between the two reconstructions.

Therefore, the preferred handedness based on the analysis of interfaces formed from rigid-body fitting of MsMca homology models into the reconstruction from single-walled filaments, is that the filament has a right-handed 3-start helix – i.e. the handedness of the original reconstruction is correct.

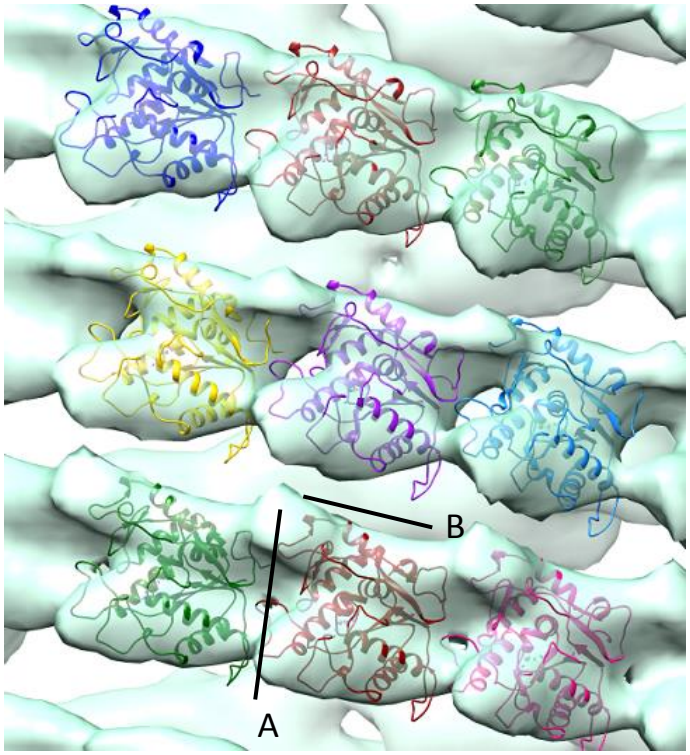
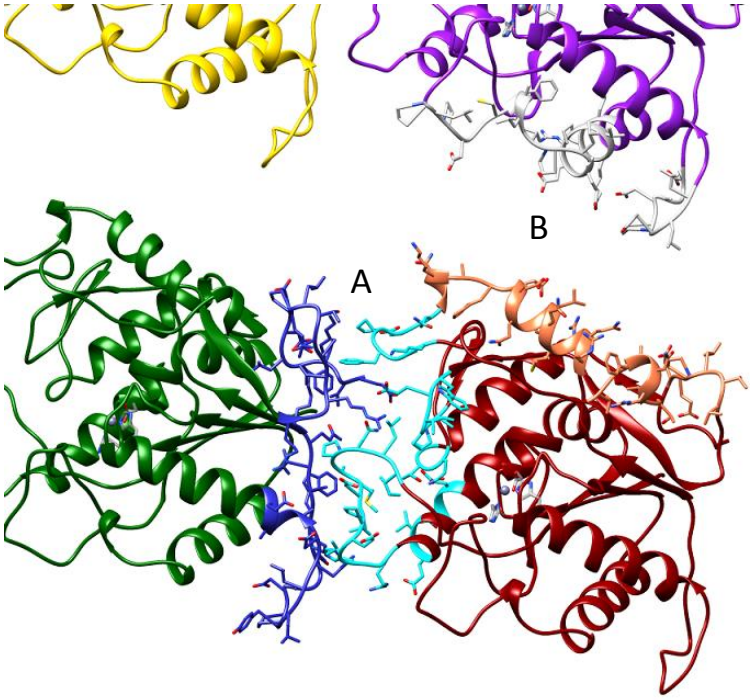


Figure 4.6.16: MsMca atomic homology models fit into the reconstruction from single-walled filaments, mirrored in the Z-axis. The major 3-start helix is left-handed. The principle interfaces are indicated by black lines. The A interface occurs along the 3-start helix, while the B interface occurs between parallel helices.

Figure 4.6.17: (right) The A and B interfaces between adjacent MsMca homology models in the reconstruction from single-walled filaments, mirrored in the Z-axis. Chain A is shown in green, chain B in dark red and chain C in purple. The A interface regions are shown in blue (chain A) and cyan (chain B). The B interface regions are shown in coral (chain B) and grey (chain C). The A interface residues in chain A are from 154-173 and 270-287. The A interface residues in chain B are 47-60, 63-64, 67-68, 197-204, 240-242 and 251-256. The B interface residues in chain B are 182-196, 214-220 and 275-280. The B-interface residues in chain C are 98-116 and 160-167.



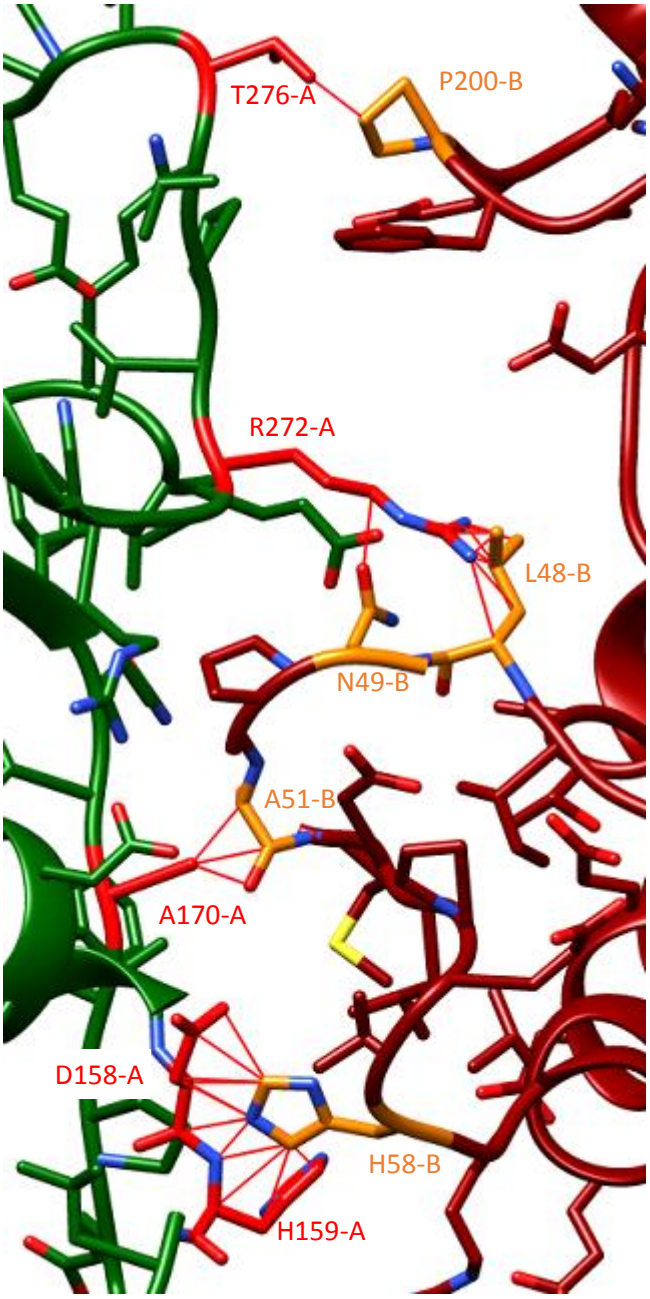


Figure 4.6.18: (left) Steric clashes (red lines) between residues at the A interface between adjacent MsMca homology models in the reconstruction from single-walled filaments, mirrored in the Z-axis. Chains A and B are shown in green and dark red respectively. Residues from chain A involved in steric clashes are shown in red. Residues from chain B involved in steric clashes are shown in orange.

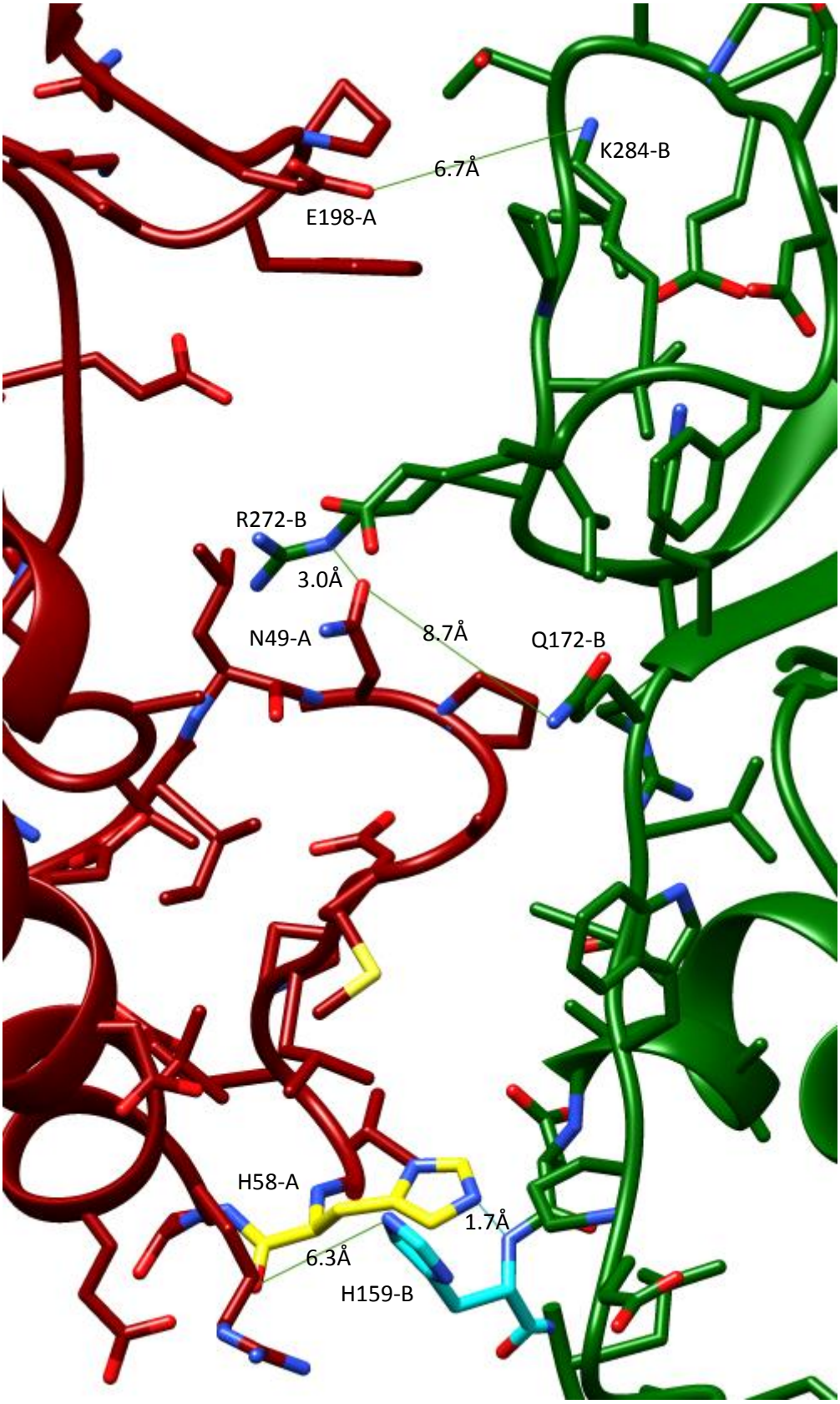
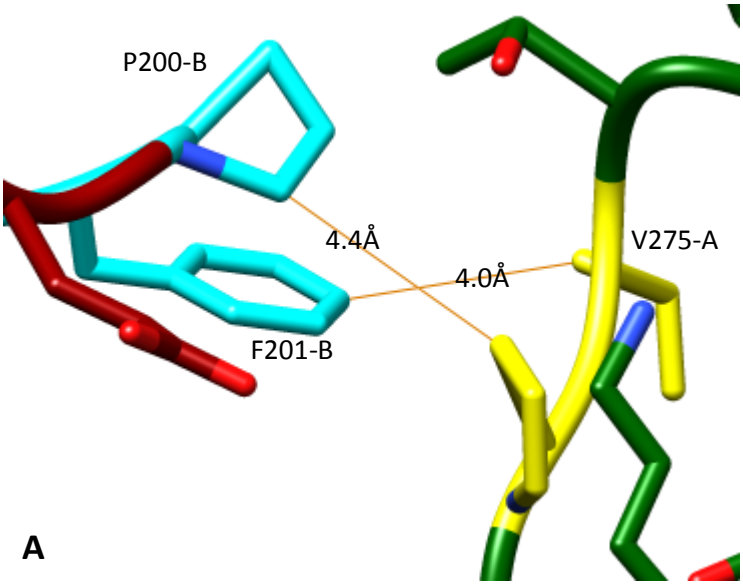
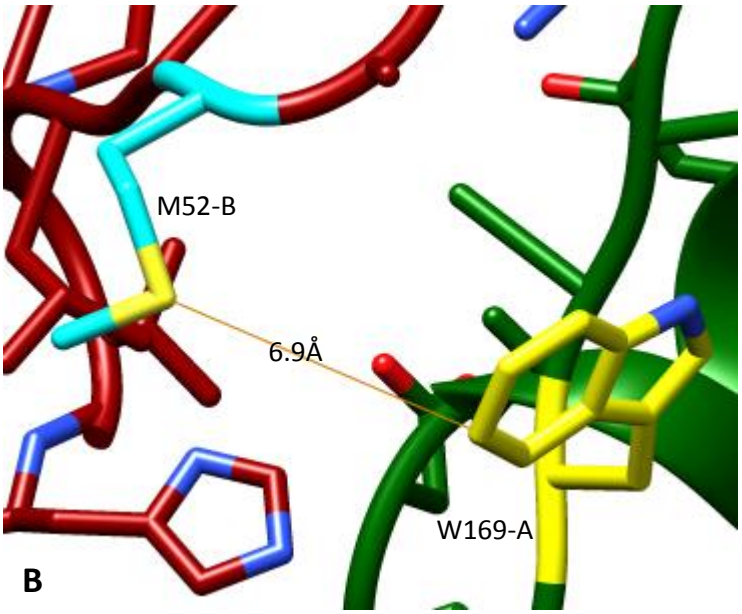


Figure 4.6.19: Possible hydrogen bonds at the A interface between adjacent MsMca homology models in the reconstruction from single-walled filaments, mirrored in the Z-axis. Chains A and B are shown in dark green and dark red respectively. Hydrogen bonds found using UCSF Chimera's *FindHBond* tool are displayed as light blue lines. Residues involved in these hydrogen bonds are shown in cyan (chain A), or yellow (chain B). Other possible hydrogen bonds are shown as green lines, and the residues involved are coloured in dark green (chain A) or in dark red (chain B).





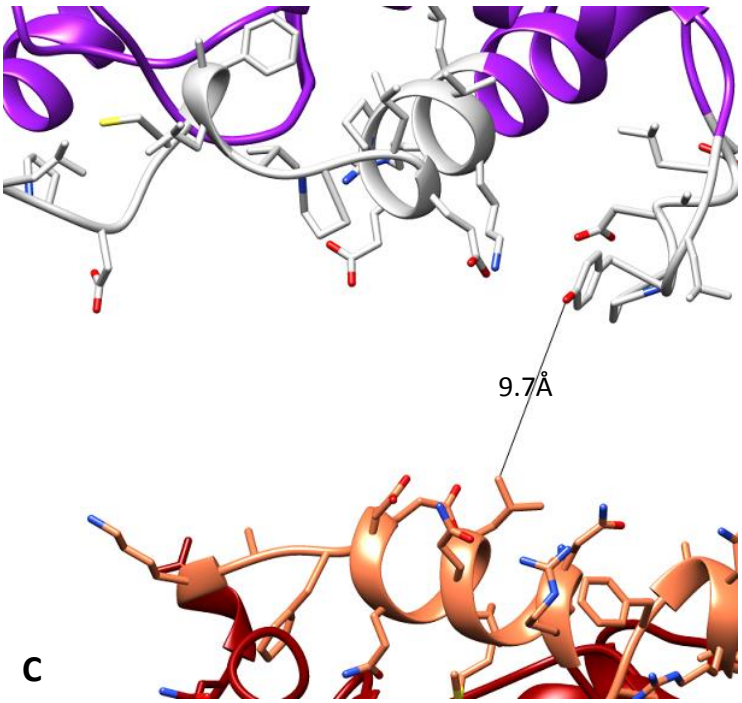


Figure 4.6.20: (A) and (B) - Possible hydrophobic interactions at the A interface between adjacent MsMca homology models in the reconstruction from single-walled filaments, mirrored in the Z-axis. Chains A and B are shown in dark green and dark red respectively. Residues involved in hydrophobic interactions are shown in cyan (chain B) and yellow (chain A). Inter-residue distances are given in Å.

C) The B interface between adjacent MsMca homology models in the reconstruction from single-walled filaments, mirrored in the Z-axis. Chains B and C are shown in dark red and purple respectively. Interface regions are shown in coral (chain B) and grey (chain C). The inter-residue distance is given in Å.

4.7. Discussion

4.7.1. Crystallization

Crystallization of MsMca was achieved using a screen containing ammonium sulfate as the precipitant. The crystallization conditions were varied, but no change to the basic needle-like morphology was observed. The needle-like crystals also grew in the presence of substrate fragments (GlcN, and Ins), which could be seen in the active site in the case of a high-resolution reconstruction.

The thin and fragile nature of these needle-like objects meant that they could not be used easily in X-ray diffraction experiments. However, when they were examined by Electron Microscopy, they were determined as disordered bundles of hollow filaments. The helical nature of these filaments meant that they could potentially be reconstructed from EM-based techniques, after the presence of helical order was determined from analysis of the power spectra generated from isolated filaments.

4.7.2. Indexing and symmetry determination

The reconstruction of helical assemblies through EM-based techniques has been possible since the first reconstruction of the Tobacco Mosaic Virus. The advances in electron detectors and sources, and the advent of Cryo-electron microscopy, have increased the resolution possible for EM-based reconstructions [7]. In addition, development of an algorithm derived from single-particle reconstructions [8] and modified for helical assemblies, has enabled the reconstruction of assemblies which were difficult to reconstruct using earlier methods [8,9], and has enabled increased resolution of structures previously determined [5]. The algorithm has been improved since its release, and several recent reconstructions of helical assemblies have been made to near-atomic resolution [10].

In this work, selection rules were used in the description of the helical symmetry of the MsMca filament. However, the selection rule is a simplistic approximation of helical symmetry. In our case, the helical repeat was given as either 18, 20, or 22 subunits per turn. As noted in [11], helical symmetry need not be a ratio of small integers, given by the selection rule. By a small change in the axial rotation per subunit, a large difference in the helical repeat may be produced. However, the limitations of the selection rule as a description of helical symmetry may be overcome using IHRSR, which refines the helical symmetry to give an accurate representation of the helical structure. In our case, the use of IHRSR led to small changes in the helical symmetry from the predictions made using the selection rules. The refined symmetry may better reflect the actual symmetry of the filament, and should be reported in the final description of the filament. Therefore, the selection rule may be a useful preliminary step to obtaining the helical symmetry of the filament, but IHRSR should be used to refine the helical symmetry.

In the present work, the helical symmetry of the MsMca filament was unknown, and this is an important preliminary step before the use of IHRSR [12]. There were a range of possible orders for each layer line. These were reduced by examination of the phases of the Fourier transform of a single filament. The given phase differences were the average result from 10 filaments, since there was some variability between filaments. This could have been through filaments that were not perfectly centred, or in the case of variable symmetry along a filament. Helical filament flattening would have caused an increase in particle diameter, with a change in helical symmetry. In the case of the MsMca filament, the degree of flattening could be measured by comparison of the filament diameter with a CryoEM reconstruction of the filament.

Ultimately, three possible solutions to the helical symmetry for the MsMca filament were proposed.

4.7.3. IHRSR

The three possible symmetries were tested using IHRSR. Reconstructions from selection rules 2 and 3 converged on the same solution. The reconstruction from the first selection rule, however, converged on a different solution. The reconstruction from the first selection rule appeared to have symmetry that better matched that of the actual filament, seen through comparison of the simulated power spectra from the reconstructions and that of the aligned raw images. The reflection on layer line 4 seen from the power spectrum of a filament was not seen in the power spectra from reconstructions made using selection rules 2 and 3, but was seen in the reconstruction made using selection rule 1.

Further comparisons between the reconstructions were made through fitting an atomic model of MsMca into the reconstructions.

4.7.4. Rigid-body fitting exercises

To assess the reconstructions generated from selection rules 1 and 2, rigid-body fitting was performed using several homology models of MsMca [appendix], which were fit into the reconstructions. The UCSF Chimera *Fit-in-map* tool was used to semi-automatically fit the monomers into the reconstruction maps.

In both cases, the atomic models fit into the reconstruction maps with a good cross-correlation score, typically between 0.64 and 0.72, due to minor variation between subunits on the three-start helix. When the subunits of the reconstruction were compared with the simulated map from the atomic models, there was not complete agreement between the simulated map and that of the reconstruction. The upper and lower extensions seen on the subunits of the two reconstructions are not seen in the simulated maps of the atomic models. The reasons for this are unclear, but may stem from slight inaccuracies of the homology model, or in the reconstructions themselves. At this resolution, it is difficult to determine the correct symmetry solution from the two reconstructions.

One of the major limitations of the fitting exercise in this case is that a rigid body fitting procedure was used, which prevented the rearrangement of flexible loops and other regions into the available density of the reconstruction map, which would have improved the fit of the atomic models into the reconstruction and would have enabled a better comparison of the two reconstructions.

Furthermore, the accuracy of the homology model is low for the region 180-288, due to low sequence conservation with the template structure [appendix]. This would be improved by a crystal structure of MsMca, or a superior template for the atomic homology model of MsMca.

At low resolution, it is not always possible to distinguish between reconstructions with the correct or incorrect symmetry. At high resolutions, these can be distinguished by examination of the secondary structure elements, or in cases of near atomic resolution, side chains [13]. Alternatively, the different symmetries make different predictions about the mass per unit length of the filament, and could be distinguished using scanning transmission electron microscopy (STEM) [12].

4.7.5. Model improvement

Both the first and second reconstructions led to possible solutions for the symmetry of the MsMca filament.

In this work, the second reconstruction was chosen to be refined using several methods not used in the initial reconstruction process. In future work, improvement to the first reconstruction could be made, and this may enable a better comparison between the two symmetries, yet the actual solution may still not be determined at this low resolution.

Several approaches were taken to attempt to improve the second reconstruction. Initially, the segment stack was inspected, and there were two populations of filament segments. Most filament segments were from single-walled hollow filaments, while a small percentage of segments were from filaments with twin-walls. These were thought to lower the resolution of the reconstruction, due to their increased diameters, and presumed different helical symmetry. Upon reconstruction from the segment stack after removal of twin-walled filament segments, the general subunit appearance was unchanged. Therefore, it is possible that their internal walls have a similar helical symmetry to that of the single-walled tubes. To investigate this, a reconstruction could be made from only twin-walled tubes. The *hsearch* and *himpose* functions would need to be used separately for the two walls, since a search over both walls would find symmetry parameters that do not represent the individual symmetries of either wall.

Layer-line filters have been used in the processing of helical particles for some time [5]. In our example, after the reciprocal lattice had been estimated, and the positions of each layer line determined, a layer-line mask could be applied to the Fourier transform of filament segments, and the inverse Fourier transform would give segments reduced in noise. The mean cross-correlation scores for filtered segments versus unfiltered segments against the best matching reference projections of the model was much higher in the case of layer-line filtered segments. The

reduced noise of the images after layer-line filtering may have been the reason for the increased cross-correlation scores. However, little improvement in the subunit appearance was seen in the reconstructions after layer-line filtering.

One potential problem of layer line filtering is that minor layer lines which are poorly visible, may be excluded, thus eliminating some of the data which could be used in the reconstruction [5]. In this work, generous masks were provided over each layer-line, to account for instances of minor differences in helical symmetry between segments, and to include some of the possible minor layer lines, which were not visible in the power spectrum from projections of earlier reconstructions, or the average power spectrum of rotated, aligned segments.

The second reconstruction appeared to have three-fold cyclic symmetry, which was not imposed during the original reconstruction process. This was examined by imposing three-fold cyclic symmetry during the reconstruction. Despite the small changes in helical symmetry, the reconstruction made using three-fold cyclic symmetry has several important advantages over the previous reconstructions. The altered appearance of the subunits led to the formation of several alternative helical families, in addition to the already seen 3- and 6-start helices. The orders of the helical families were slightly different from those predicted. The selection rule $l = 20m + n$ is not compatible with a 3-fold cyclic symmetry with the layer lines here, as n must be a multiple of 3 [14]. An alternative selection rule of $l = 19m + n$ does satisfy the requirements of 3-fold cyclic symmetry. This is in conflict with the results of the phase differences in 4.2.3, which require an even number of subunits per turn. The phase differences were calculated from a number of filaments, and there was some variation between phase differences amongst the different filaments. It is possible that different symmetries exist between filaments, or within a single filament. Therefore, the selection rule of $l = 19m + n$ may yet be valid. A reconstruction made with 3-fold cyclic symmetry, and starting symmetry $\Delta\phi = 18.95^{o}$ and $\Delta z = 9.6$ Å, converged on $\Delta\phi = 19.05^{o}$ and $\Delta z = 9.66$ Å after 14 cycles (results not shown). Therefore, the alternative symmetry of $\Delta\phi = 18.95^{o}$ and $\Delta z = 9.6$ Å should be investigated in future work. Furthermore, the power spectrum of a projection of the C_3 reconstruction was the best match with the average power spectrum of aligned filament segments, when compared with the power spectra of projections of any of the preceding reconstructions.

The final reconstruction incorporated all the preceding steps, while also including a cut-off filter to remove images with very low correlation to the reference projections. For each preceding reconstruction, assessment of the cross-correlation between images and their best matching reference projections showed a subset of images with very low cross-correlation scores. These images appeared to be single-walled filaments, but with larger diameter than the other filaments, and therefore with different helical symmetry. These were removed from cycles in the final reconstruction by removing images with cross-correlation scores of more than two-standard deviations below the mean cross-correlation score.

The final reconstruction differed little from the previous reconstruction where 3-fold cyclic symmetry was used. Given the use of a large number of segments, and the imposition of 3-fold cyclic symmetry, this alone may have been sufficient to reach a resolution close to the maximum resolution for negative-stain samples. The additional

filters included in the final reconstruction would not have significantly improved the reconstruction, but may have been more effective had a smaller dataset been used.

4.7.6. The asymmetric unit

MsMca was previously determined as a monomer in solution, but in the purification followed here, it existed as a dimer. In the purification of MsMca, a buffer of 50mM Tris-HCl, pH 8.0 was used. However, in the crystallization conditions, 1.4M ammonium sulfate was used, with 0.1 M HEPES, pH 7.0. It is not clear whether dimers would have formed under these conditions. Therefore, the basic asymmetric unit of the filament is unknown. The IHRSR method was performed with the assumption of a monomeric asymmetric unit, hence subunits related to one another by the helical symmetry are averaged. However, if the IHRSR script were modified to allow a dimer as the asymmetric unit, then differences in the appearance of the subunits may be visible. This would result in a more accurate reconstruction of the filament, if the asymmetric unit is a dimer.

4.7.7. Handedness of the filament

The handedness of the filament is unknown. Attempts were made to determine the handedness of negatively stained filaments through Scanning Electron Microscopy (SEM). The imaging conditions that were used were insufficient to obtain fine structural details of the filament, and after processing the segments, the positions of reflections in the averaged power spectrum of SEM-derived segments were not consistent with those obtained in the power spectrum of TEM-derived MsMca filament segments. Therefore, the handedness of the MsMca filament remained unknown.

It is possible to dock an atomic model into the subunits of reconstructions of either handedness, and to determine the handedness from the reconstruction allowing the best fit [15]. Therefore, in section 4.6, reconstructions of both handedness were used in the search for the major interface between MsMca subunits. The most favourable handedness was determined from the fit of the atomic models, and comparison of interactions across the interface, using reconstructions of either handedness.

4.7.7.1. Final reconstruction

Based on analysis of the interfaces formed between MsMca atomic models after fitting them into the final reconstruction and the final reconstruction mirrored about the Z-axis, the handedness of the reconstruction could not be definitively determined. There were many steric clashes at the A interface in reconstructions of either handedness. This is a limitation of rigid-body fitting, which does not allow changes to the structure of the atomic model during fitting of the model into the reconstruction map. Flexible fitting should be used in the future to avoid the steric clashes at the interfaces generated through rigid-body fitting.

4.7.7.2. Single-walled filament reconstruction

Based on analysis of the interfaces formed between MsMca atomic models after fitting them into the reconstruction maps of either handedness, the single-walled reconstruction is estimated to have a right-handed 3-start helix, thus the original reconstruction is deemed to be correct.

However, due to the limitations of the rigid-body fitting approach, the notation of handedness of the filament is suggestive only and alternative methods should be used to confirm the estimation of handedness given in this work. These methods are discussed in greater detail in the following sections.

4.7.7.3. Conclusions for the handedness estimation

The accurate placement of residues at the interfaces, particularly over the region from 181 and above, is limited in the current homology model, due to low sequence conservation between MsMca and MshB over the region 181-288 (see appendix). The differences in the homology model and the actual structure of MsMca would influence the interfaces modelled in the reconstruction, obtained using either rigid-body or flexible fitting methods, which would affect the estimation of handedness for the filament.

To address this, a better homology model of MsMca is required, but this awaits the determination of a structure with a higher sequence conservation to MsMca than is between MsMca and MshB. At present, MshB is the enzyme with the most similar sequence to MsMca for which there is an available structure.

The limitations of the fitting procedure can be seen in the fact that the rigid body fitting did not allow for the rearrangement of residues or regions involved in the interfaces. Alternative fitting methods, such as molecular dynamics flexible fitting (MDFF) [16] would improve the fit of the atomic models into the reconstruction map and give more favourable interfaces, which could be used to suggest the handedness of the reconstruction.

The handedness could also be confirmed using additional experimental methods. The first, which may be most useful in the case of the MsMca filament, is the use of metal-shadowing to determine the handedness from the surface of the filament [17]. Additionally, one may use tilt-experiments, where one tilts the sample in the EM [18].

Alternatively, at sufficiently high resolution, e.g. through use of newer detectors and CryoEM techniques, the fitting of atomic models into subunits of the reconstructed map can be used to distinguish the correct- and incorrect handedness [15]. This is also the best means to determine the actual symmetry of the filament.

CryoEM techniques, will not only assist in the determination of the symmetry of the filament, but can be used to determine the high-resolution structure of MsMca, which may yield near-atomic details and provide the means for structure-based drug design against MsMca.

4.8. Conclusion

The enzyme Mycothiol S-conjugate amidase is an important member of the mycothiol metabolic pathway. Its characterization by structural determination would assist the development of inhibitors against the enzyme, and increase the understanding of differences in MshB and Mca substrate specificities and reaction mechanisms.

Mca had been characterized previously, but its structure awaits determination. Considering the aims of this research, the following list of accomplishments were made:

A search for crystalline forms of MsMca led to the discovery of filaments, which were regular, and showed helical order. Analysis of power spectra of single filaments showed layer lines at 1/87, 1/60, 1/46 and 1/30 $Å^{-1}$. Three possible symmetry solutions were predicted, which were refined to two solutions through IHRSR.

The two solutions were compared through their power spectra and correlation between simulated maps of atomic models of MsMca with the subunits of the reconstructions. Based on their power spectra, the first reconstruction better matched the predicted symmetry than did the second reconstruction. However, the fitting experiments did not indicate either reconstruction as being superior.

The second reconstruction, of helical symmetry $\Delta\phi = 17.98^{o}$ and $\Delta z = 9.12$ Å, was improved using several additional techniques to those used in the first reconstruction process. These included the use of 3-fold cyclic symmetry, the removal of double-walled filaments, use of a cut-off filter for images with low correlation to projections of the 3D reconstruction, and use of a layer-line filter to reduce the noise in the images. These were used individually, then in a single reconstruction, to improve the agreement between the predicted helical symmetry and that obtained from the reconstruction.

Several reconstructions were used in rigid-body fitting exercises to assess the favoured handedness of the filament through examination of the major interfaces between subunits. These suggest that the major 3-start helix is right-handed. Future work would be to determine the handedness of the filament using alternative techniques, such as metal-shadowing.

This work provides a protocol by which a high-resolution structure may be obtained using CryoEM-based methods. An atomic or near-atomic structure of MsMca will give the basis for the different substrate specificity in Mca and MshB, and will aid in the development of inhibitors against the enzyme and in the targeting of the mycothiol pathway in the treatment of drug resistant Tuberculosis.

This work also shows the use of IHRSR in the production of meaningful low-resolution structures of negatively stained filaments. However, at the low resolution obtained, the correct helical symmetry could not be determined definitively from the suggested solutions.

4.9. References

1. Tang G, Peng L, Baldwin PR, Mann DS, Jiang W, Rees I, Ludtke SJ: **EMAN2: An extensible image processing suite for electron microscopy**. *J Struct Biol* 2007, **157**:38–46.

2. Frank J, Radermacher M, Penczek P, Zhu J, Li Y, Ladjadj M, Leith A: **SPIDER and WEB: Processing and Visualization of Images in 3D Electron Microscopy and Related Fields**. *J Struct Biol* 1996, **116**:190–199.

3. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE: **UCSF Chimera--a visualization system for exploratory research and analysis.** *J Comput Chem* 2004, **25**:1605–12.

4. Erickson HP: **Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy**. *Biol Proced Online* 2009:32–51.

5. Pomfret AJ, Rice WJ, Stokes DL: **Application of the iterative helical real-space reconstruction method to large membranous tubular crystals of P-type ATPases**. *J Struct Biol* 2007, **157**:106–116.

6. Schneider CA, Rasband WS, Eliceiri KW: **NIH Image to ImageJ: 25 years of image analysis.** *Nat Methods* 2012, **9**:671–5.

7. Merk A, Bartesaghi A, Banerjee S, Falconieri V, Rao P, Davis MI, Pragani R, Boxer MB, Earl LA, Milne JLS, Subramaniam S: **Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery**. *Cell* 2016, **165**:1698–1707.

8. Egelman EH: **A robust algorithm for the reconstruction of helical filaments using single-particle methods**. *Ultramicroscopy* 2000, **85**:225–234.

9. Trachtenberg S, Galkin VE, Egelman EH: **Refining the Structure of the Halobacterium salinarum Flagellar Filament Using the Iterative Helical Real Space Reconstruction Method: Insights into Polymorphism**. *J Mol Biol* 2005, **346**:665–676.

10. Ge P, Scholl D, Leiman PG, Yu X, Miller JF, Zhou ZH: **Atomic structures of a bactericidal contractile nanotube in its pre- and postcontraction states**. *Nat Struct Mol Biol* 2015, **22**:377–382.

11. Egelman EH: **The iterative helical real space reconstruction method: surmounting the problems posed by real polymers.** *J Struct Biol* 2007, **157**:83-94.

12. Egelman EH: **Reconstruction of helical filaments and tubes**. *Methods Enzymol* 2010, **482**(C):167–183.

13. Egelman EH: **Ambiguities in helical reconstruction**. *Elife* 2014, **3**.

14. Klug A, Crick FHC, Wyckoff HW: **Diffraction by helical structures**. *Acta Crystallogr* 1958, **11**:199–213.

15. Kim J, Wu S, Tomasiak TM, Mergel C, Winter MB, Stiller SB, Robles-Colmanares Y, Stroud RM, Tampé R, Craik CS, Cheng Y: **Subnanometre-resolution electron cryomicroscopy structure of a heterodimeric ABC exporter.** *Nature* 2015, **517**:396–400.

16. Trabuco LG, Villa E, Mitra K, Frank J, Schulten K. **Flexible Fitting of Atomic Structures into Electron**

Microscopy Maps Using Molecular Dynamics. *Structure* 2008, **16**:673–83.

17. Woodward JD, Weber BW, Scheffer MP, Benedik MJ, Hoenger A, Sewell BT: **Helical structure of unidirectionally shadowed metal replicas of cyanide hydratase from Gloeocercospora sorghi**. *J Struct Biol* 2008, **161**:111–119.

18. LINCK_ RW, AMOS LA: **The Hands of Helical Lattices in Flagellar Doublet Microtubules**. *J Cell Sci* 1974, **14**.

Appendix

A.1 Sequence information for MsMca and the His-MBP-MsMca fusion protein

(A)

```
         10          20          30          40          50          60
MGHHHHHHHH ASKIEEGKLV IWINGDKGYN GLAEVGKKFE KDTGIKVTVE HPDKLEEKFP
         70          80          90         100         110         120
QVAATGDGPD IIFWAHDRFG GYAQSGLLAE ITPDKAFQDK LYPFTWDAVR YNGKLIAYPI
        130         140         150         160         170         180
AVEALSLIYN KDLLPNPPKT WEEIPALDKE LKAKGKSALM FNLQEPYFTW PLIAADGGYA
        190         200         210         220         230         240
FKYENGKYDI KDVGVDNAGA KAGLTFLVDL IKNKHMNADT DYSIAEAAFN KGETAMTING
        250         260         270         280         290         300
PWAWSNIDTS KVNYGVTVLP TFKGQPSKPF VGVLSAGINA ASPNKELAKE FLENYLLTDE
        310         320         330         340         350         360
GLEAVNKDKP LGAVALKSYE EELAKDPRIA ATMENAQKGE IMPNIPQMSA FWYAVRTAVI
        370         380         390         400         410         420
NAASGRQTVD EALKDAQTLI NGDGAGLEVL FQGPENLYFQ AIAMSELRLM AVHAHPDDES
        430         440         450         460         470         480
SKGAATTARY AAEGARVMVV TLTGGERGDI LNPAMDLPEV HGRIAEVRRD EMAKAAEILG
        490         500         510         520         530         540
VEHHWLGFVD SGLPEGDPLP PLPDGCFALV PLEEPVKRLV RVIREFRPHV MTTYDENGGY
        550         560         570         580         590         600
PHPDHIRCHQ VSVAAYEAAA DHLLYPDAGE PWAVQKLYYN HGFLRQRMQL LQEEFAKNGQ
        610         620         630         640         650         660
EGPFAKWLEH WDPDNDVFAN RVTTRVHCAE YFHQRDDALR AHATQIDPKG DFFHAPIEWQ
        670         680         690
QRLWPTEEFE LARARVPVTL PEDDLFKGVE PV
```

Full fusion protein (His-MBP-MsMca)
Molecular weight: 77039.07
Theoretical pI: 5.29
Abs 0.1% (=1 g/l) 1.465

(B)

```
         10          20          30          40          50          60
AIAMSELRLM AVHAHPDDES SKGAATTARY AAEGARVMVV TLTGGERGDI LNPAMDLPEV
         70          80          90         100         110         120
HGRIAEVRRD EMAKAAEILG VEHHWLGFVD SGLPEGDPLP PLPDGCFALV PLEEPVKRLV
        130         140         150         160         170         180
RVIREFRPHV MTTYDENGGY PHPDHIRCHQ VSVAAYEAAA DHLLYPDAGE PWAVQKLYYN
        190         200         210         220         230         240
HGFLRQRMQL LQEEFAKNGQ EGPFAKWLEH WDPDNDVFAN RVTTRVHCAE YFHQRDDALR
        250         260         270         280         290
AHATQIDPKG DFFHAPIEWQ QRLWPTEEFE LARARVPVTL PEDDLFKGVE PV
```

Free cleaved MsMca
Molecular weight: 33028.2
Theoretical pI: 5.21
Abs 0.1% (=1 g/l) 1.364

Figure A.1: Amino acid sequences of (A) the fusion protein His-MBP-MsMca and (B) free MsMca after separation from His-MBP by TEV-protease cleavage. Expasy Protparam [1] results are given with each sequence.

A.2 Sequence comparison of MsMca and MtMca

```
CLUSTAL O(1.2.4) multiple sequence alignment

SP|A0R2W9|MCA_MYCS2 MSELRLMAVHAHPDDESSKGAATTARYAAEGARVMVVTLTGGERGDILNPAMDLPEVHGR 60
SP|P9WJN1|MCA_MYCTU MSELRLMAVHAHPDDESSKGAATLARYADEGHRVLVVTLTGGERGEILNPAMDLPDVHGR 60
                    ********************** **** ** **:**********:*********:****

SP|A0R2W9|MCA_MYCS2 IAEVRRDEMAKAAEILGVEHHWLGFVDSGLPEGDPLPPLPDGCFALVPLEEPVKRLVRVI 120
SP|P9WJN1|MCA_MYCTU IAEIRRDEMTKAAEILGVEHTWLGFVDSGLPKGDLPPPLPDDCFARVPLEVSTEALVRVV 120
                    ***:*****:********** **********:**  *****.*** ****  .: ****:

SP|A0R2W9|MCA_MYCS2 REFRPHVMTTYDENGGYPHPDHIRCHQVSVAAYEAAADHLLYPDAGEPWAVQKLYYNHGF 180
SP|P9WJN1|MCA_MYCTU REFRPHVMTTYDENGGYPHPDHIRCHQVSVAAYEAAGDFCRFPDAGEPWTVSKLYYVHGF 180
                    ************************************.*.  :*******:*.**** ***

SP|A0R2W9|MCA_MYCS2 LRQRMQLLQEEFAKNGQEGPFAKWLEHWDPDNDVFANRVTTRVHCAEYFHQRDDALRAHA 240
SP|P9WJN1|MCA_MYCTU LRERMQMLQDEFARHGQRGPFEQWLAYWDPDHDFLTSRVTTRVECSKYFSQRDDALRAHA 240
                    **:***:**:***::**.*** :** :****:*.::.******.*::** **********

SP|A0R2W9|MCA_MYCS2 TQIDPKGDFFHAPIEWQQRLWPTEEFELARARVPVTLPEDDLFKGVEP 288
SP|P9WJN1|MCA_MYCTU TQIDPNAEFFAAPLAWQERLWPTEEFELARSRIPARPPETELFAGIEP 288
                    *****:.:** **: **:************:*:*.  ** :** *:**
```

Figure A.2: Pairwise alignment of amino acid sequences in *M. smegmatis* Mca (MCA_MYCS2) and *M. tuberculosis* Mca (MCA_MYCTU), generated using CLUSTAL Omega (1.2.4). There is 78.125% sequence identity (225 out of 288 residues) and 14.2% similar residues (41 out of 288 residues).

A.3 Structural comparison of MsMca and MtMca

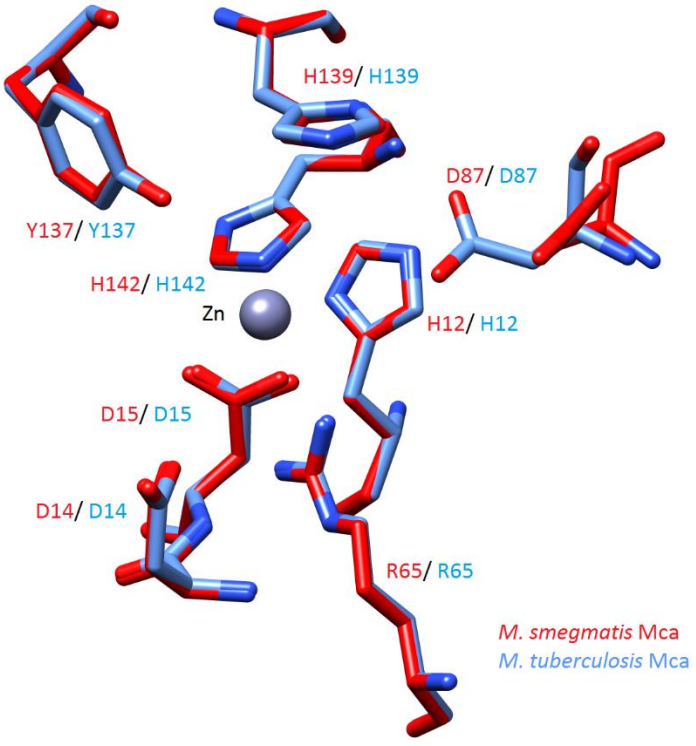


Figure A.3: Active site residues in MtMca (blue) and MsMca (red) homology models, after superimposition through the Match->Align tool in UCSF Chimera [2].

A.4 MsMca homology model made using *M. tuberculosis* MshB, PDB file 4EWL

The MsMca homology model was generated using *M. tuberculosis* MshB from the PDB entry 4EWL, as the template structure. This was performed using the SWISS-MODEL web workspace [3, 4, 5].

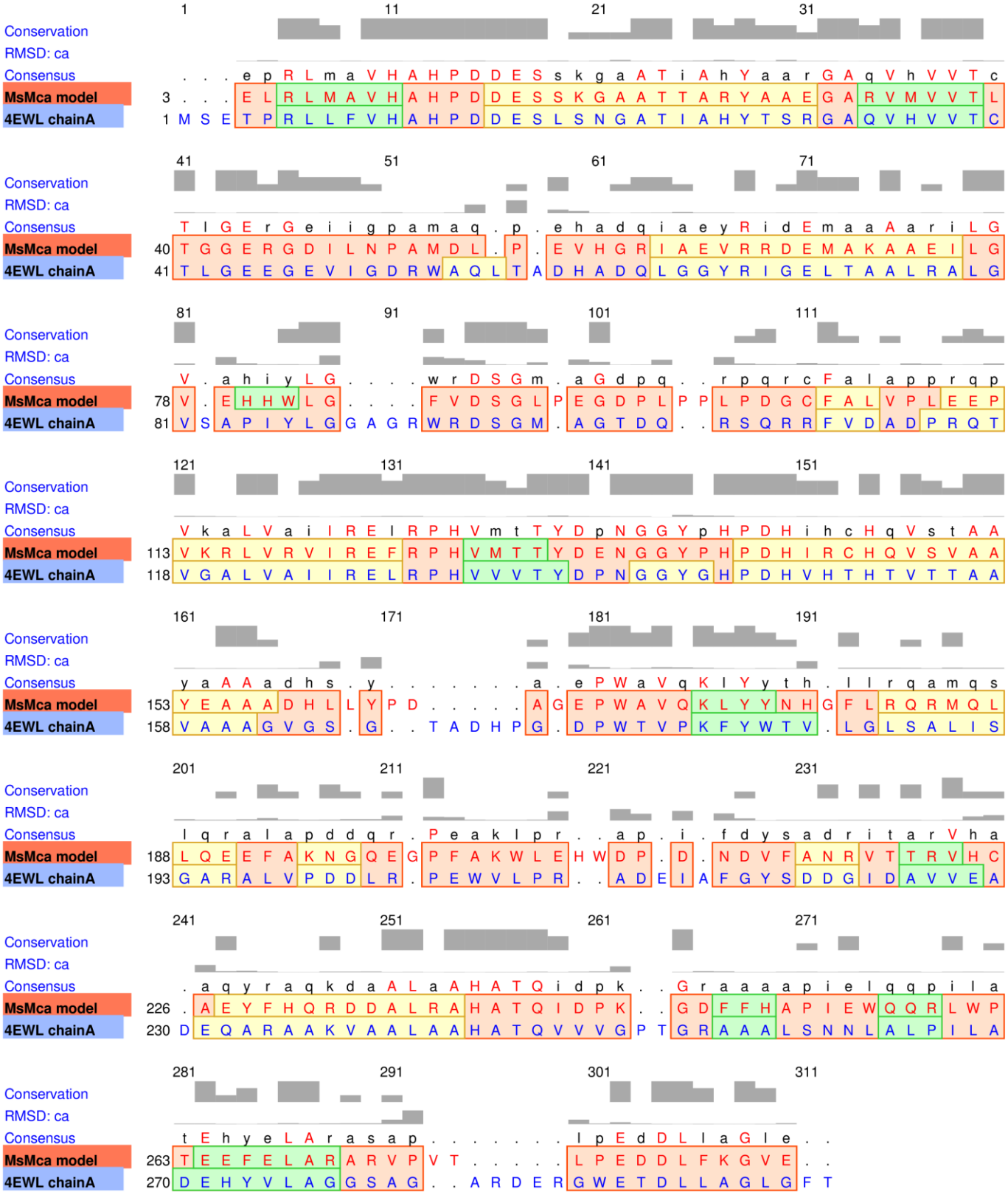


Figure A.4: Sequence alignment of the MsMca homology model and MshB (4EWL, chain A) used as a template. The secondary structure elements are shown (α-helices = red, β-sheets = green, loops = yellow). The consensus sequence, conservation scores and RMSD (Cα) are shown above the aligned residues.

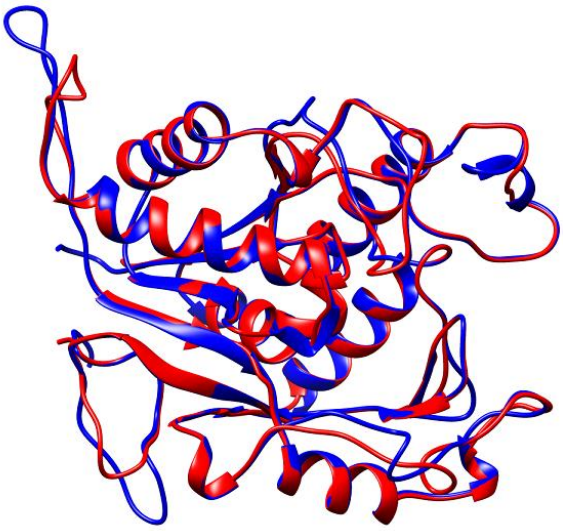


Figure A.5: Superimposition of the MsMca homology model (red) with the template structure, MshB (blue) from the PDB file 4ewl (chain A).

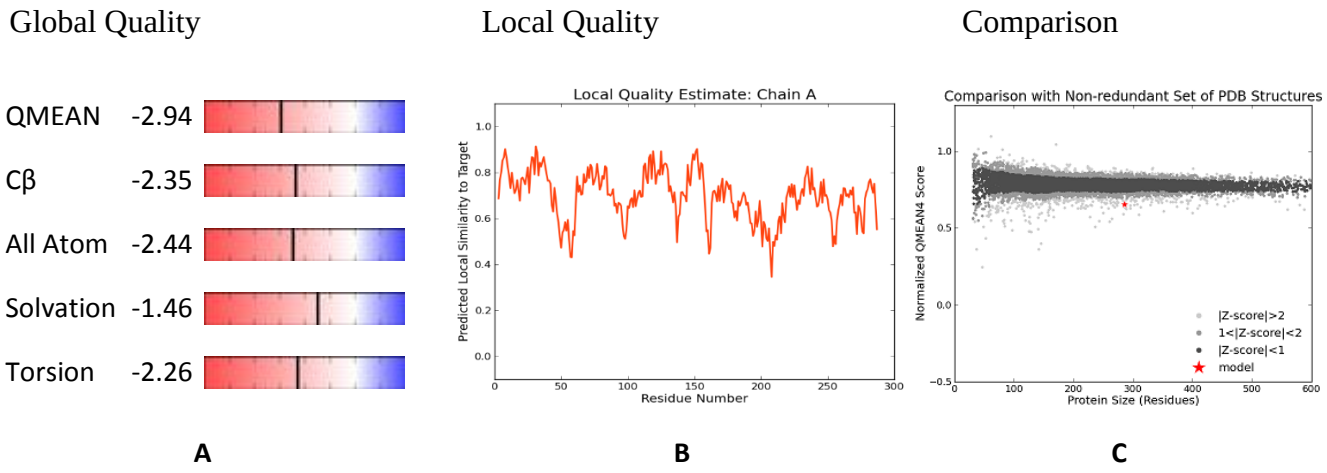


Figure A.6: Summary of the SWISS-MODEL based homology model generation of the MsMca model from the template structure 4EWL [5]. (A) Global quality scores. (B) Local quality scores. (C) Comparison between QMEAN scores of the MsMca homology model and a set of non-redundant PDB structures.

A.5 References:

1. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E, Grosdidier A, Hernandez C, Ioannidis V, Kuznetsov D, Liechti R, Moretti S, Mostaguir K, Redaschi N, Rossier G, Xenarios I, Stockinger H: **ExPASy: SIB bioinformatics resource portal.** *Nucleic Acids Res* 2012, **40**(Web Server issue):W597–603.

2. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE: **UCSF Chimera--a visualization system for exploratory research and analysis.** *J Comput Chem* 2004, **25**:1605–12.

3. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. SWISS-MODEL: **Modelling protein tertiary and quaternary structure using evolutionary information.** *Nucleic Acids Res*. 2014;42.

4. Arnold K, Bordoli L, Kopp J, Schwede T. **The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling.** *Bioinformatics*. 2006;22:195–201.

5. Benkert P, Biasini M, Schwede T. **Toward the estimation of the absolute quality of individual protein structure models.** *Bioinformatics*. 2011;27:343–50.