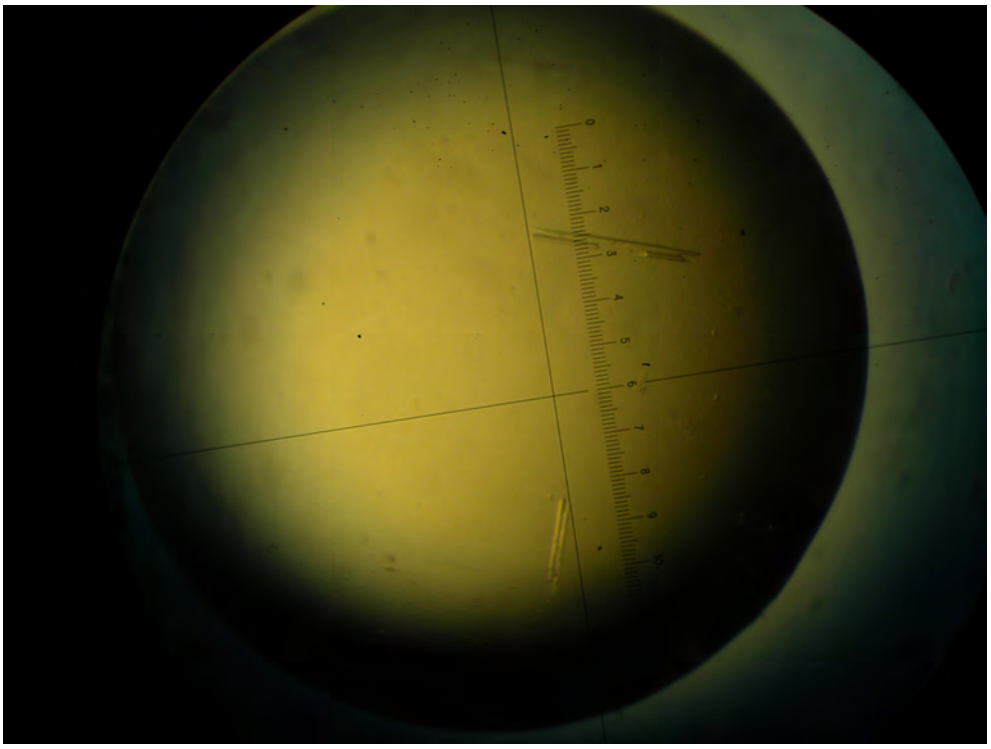
UPPSALA
UNIVERSITET

Structural and functional studies of the Mycobacterial protease MycP3, and RNA polymerase



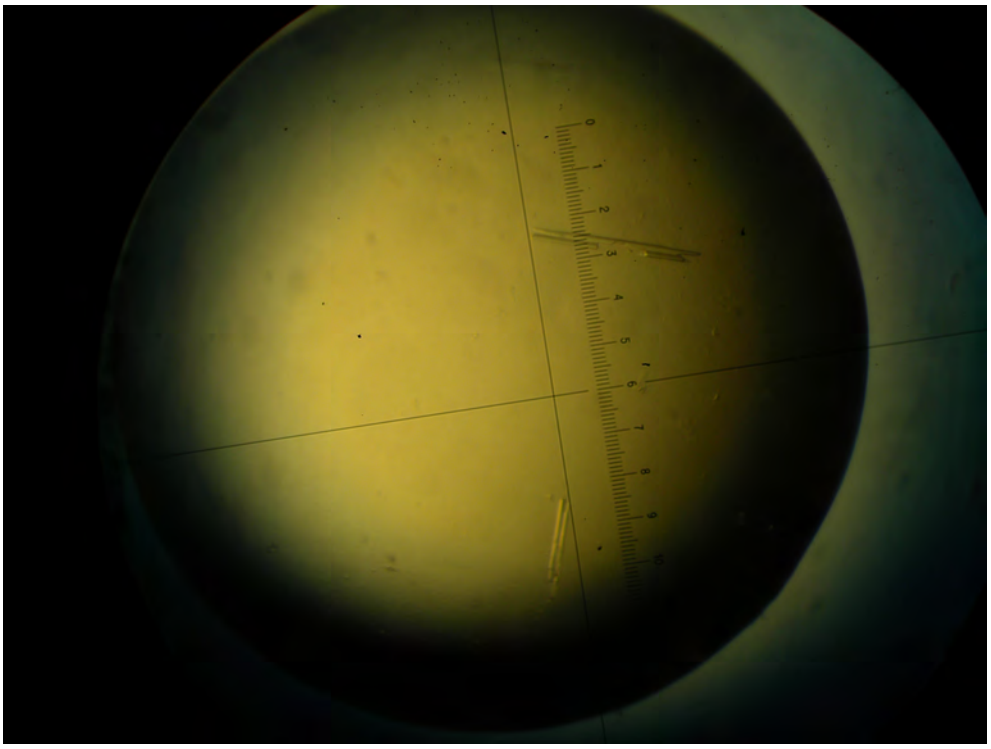
Husam Abdelsalam Babikir

Degree project in biology, Master of science (2 years), 2010
Examensarbete i biologi 30 hp till masterexamen, 2010
Biology Education Centre and Department of Cell and Molecular Biology, Structural Biology, Uppsala University
Supervisors: Professor Torsten Unge and Annette Roos

Table of contents

UPPSALA
UNIVERSITET

Structural and functional studies of the Mycobacterial protease MycP3, and RNA polymerase



Husam Abdelsalam Babikir

Degree project in biology, Master of science (2 years), 2010
Examensarbete i biologi 30 hp till masterexamen, 2010
Biology Education Centre and Department of Cell and Molecular Biology, Structural Biology, Uppsala University
Supervisors: Professor Torsten Unge and Annette Roos

Abbreviations

AMC	aminomethyl coumarin
bp	base pair
BME	β-mercaptoethanol
dNTP	Deoxyribonucleotide triphosphate
E.coli	*Escherichia coli*
ESRF	European Synchrotron Radiation Facility
IMAC	Immobilized metal affinity chromatography
IPTG	Isopropyl-β-D-thiogalactoside
LB	Luria Bertaini broth
Mt	*Mycobacterium tuberculosis*
Msm	*Mycobacterium smegmatis*
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
Pfu	*Pyrococcus furiosis*
RAPID	Rational Approach to Pathogen Inhibitor Discovery
SDS	Sodium dodecyl sulphate
Taq	*Thermus aquaticus*
TB	Tuberculosis

Dedication

This thesis is dedicated to my father, mother, brothers, and sisters. I know I would have made you proud.

Abstract

The mycosin family consists of five subtilisin-like serine proteases found within the periplasmic space of the *Mycobacterium tuberculosis* cell wall. They do not exist in humans, and the most similar human homolog, a transmembrane bound human peptidase, have obvious differences in the active site to make design of species-specific inhibitors possible.

Mycosin3 is one member of the mycosin family and is considerd an attractive target for drug development. The Mycosin 3 gene was successfully cloned and different constructs were designed. Attempts to produce soluble Mycosin 3 protein resulted in formation of inclusion bodies. However, these could be denatured and successfully resolubilized using the dialysis and rapid dilution methods. Crystallization was performed using sitting and hanging drop methods, which resulted in crystals of different quality being obtained. The activity of Mycosin 3 was tested using β-casein and aminomethyl coumarin substrates, but unfortunately it was found to be inactive, a result which may be due to protein misfolding.

Another enzyme, which is considered as a promising tuberculosis drug target is RNA polymerase. The important role of this enzyme as a drug target is a result of its relation with Rifampicin, one of the first line anti-tuberculosis drugs**.** The inhibition of RNA polymerase of *Mt* occurs as a result of Rifampicin binding to the beta subunit of RNA polymerase thus halting the transcription machinery. Using an expression vector containing the *Msm* RNA polymerase, the protein was successfully expressed. Crystallization experiments were performed and promising needles of crystalline material were produced. They were tested at a synchrotron but unfortunately no diffraction was observed possibly due to insufficient cryo protection.

1. Introduction

Mycobacterium tuberculosis (*Mt*) is a gram positive, rod shaped bacteria that is known for its slow growth, its complicated cell envelope and its ability to remain inside the host for a long time after infection (1). It is considered one of the leading causes of death related to bacterial infection according to the World Health organization's (WHO) estimate in 2004. In 2008, WHO reported that more than 9 million people were infected with *Mt*, most of them live in underdeveloped areas in Africa and Asia (Fig.1).The most recent report from WHO in 2009 declares that 2 billion people in the world (approximately one third of the world population) are carriers of *Mt* (2).

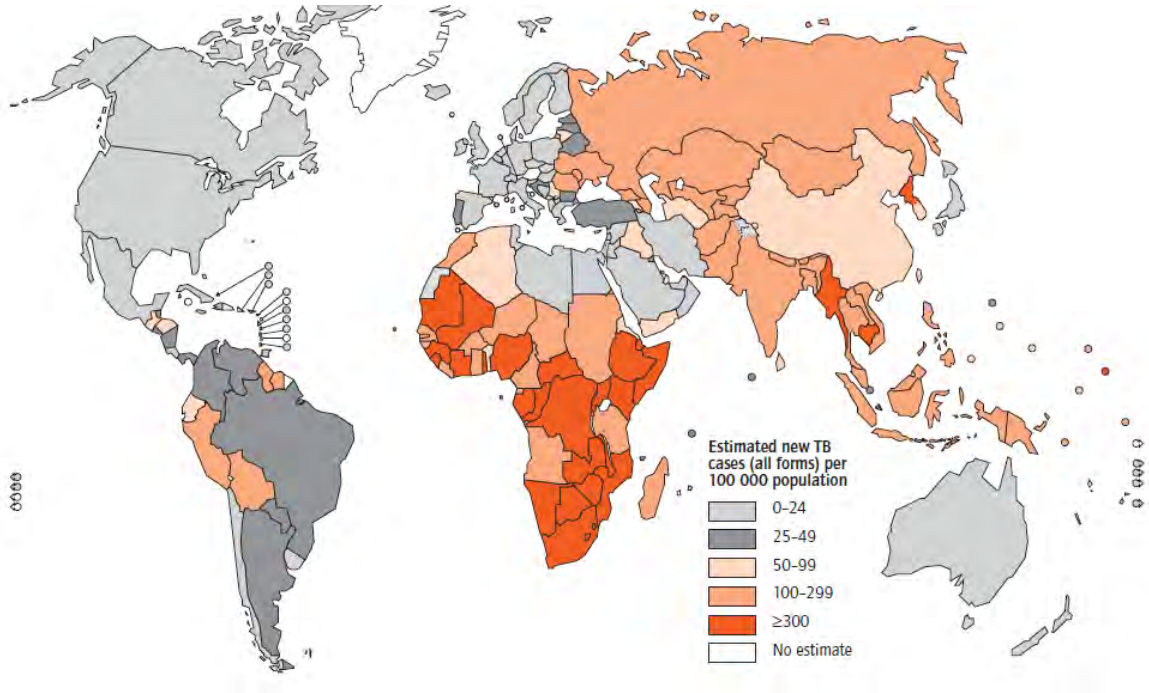


Fig. 1 Tuberculosis incidence rate in 2008, WHO report (2)

The strong relation between tuberculosis and HIV infection is noticeable. Tuberculosis and HIV have been closely linked since the emergence of AIDS the last century. It is believed that TB is one of the main reasons behind the death of people who are HIV positive. Moreover, some studies estimate that HIV patients are 50 times more likely to get a TB infection compared to those who are HIV negative (16).

During the past few years, some strains of *Mt* have evolved resistance to the most available first-line TB drugs isoniazid and rifampicin. These *Mt* strains are classified as multi-drug resistant (MDR-TB). The strains of *Mt* that are resistant to the second-line TB drugs kanamycin , capreomycin and amikacin are described as extensively drug resistant (XDR-TB)(2). China,

India and the Russian Federation are the countries with the highest increase of MDR-TB cases. Taking these facts into consideration, extensive work should be done to design a unique drug that works against drug targets for tuberculosis. Thus, RAPID centre (Rational Approach to Pathogen Inhibitor Discovery) was established at Uppsala University to design and develop novel drugs for treatment of serious diseases like tuberculosis and malaria. In this center, the potential drug target proteins are characterized and their structures are solved using X- ray crystallography. The RAPID centre joins together different scientific disciplines such as medicinal and combinatorial chemistry, computational chemistry and structural biology (3). This study focused mainly on two enzymes, MycP3 from *Mt* and *Msm* and RNA polymerase from *Msm.*

1.1 Mycosin-3 (MycP3)

The vital role of proteases in the pathogenesis of many bacterial species is well known (4). They have a negative impact on the host tissue and inactivate its defense mechanism. However, the role of proteases as virulence factors in *Mt* is poorly known and there is scarcity of information about their involvement and contribution in the pathogenesis of *Mt* (5). There are many and different genes in the genome of *Mt* H37Rv encoding for more than 30 proteases (6) including a family of mycosins. The proteins that are encoded by these genes are Mycp1-5 which are subtilisin-like proteases within the periplasmic space of the *Mt* bacterial cells. A previous study suggests that mycosins have a crucial role in the Snm secretion pathway in mycobacteria, in particular MycP1 (8), which is believed to help in the secretion of two predominant *Mt* antigens, ESAT-6 and AFP-10. In addition, due to the presence of MycP1 in the cell periplasm, it is assumed that MycP1 works as a regulator for the Snm pathway either by blocking an Snm pathway inhibitor or by acting as an activator for a protein in the Snm pathway (8).

Each protease of the mycosin family contains a catalytic triad (Asp, His, Ser) within a highly conserved sequence (5). These three amino acids are the main players in the cleaving activity of proteases from the subtilisin family. Furthermore, all proteases of the mycosin family share the same feature of having a C-terminal transmembrane region and an N-terminal signal sequence (Fig.2). Because of the high similarity between the five mycosin proteases, it is hypothesized that if an inhibitor molecule binds to one of them, then the possibility to bind to other mycosins is also high. Moreover, mycosins do not exist in humans, and the most similar human homolog, a transmembrane bound human peptidase, has obvious differences in the active site, which gives a chance for design of species specific inhibitors.

MycP3 has been chosen among the members of mycosin protease family. It is conserved within the most of virulent mycobacterial species (7) and is also present in *Msm* making it a good starting target to know more about this protein family. The lack of information about its role in the biology of *Mt* and the possibility of its responsibility for survival in the host, make MycP3 an attractive target for drug development and motivate research and investigation about it.

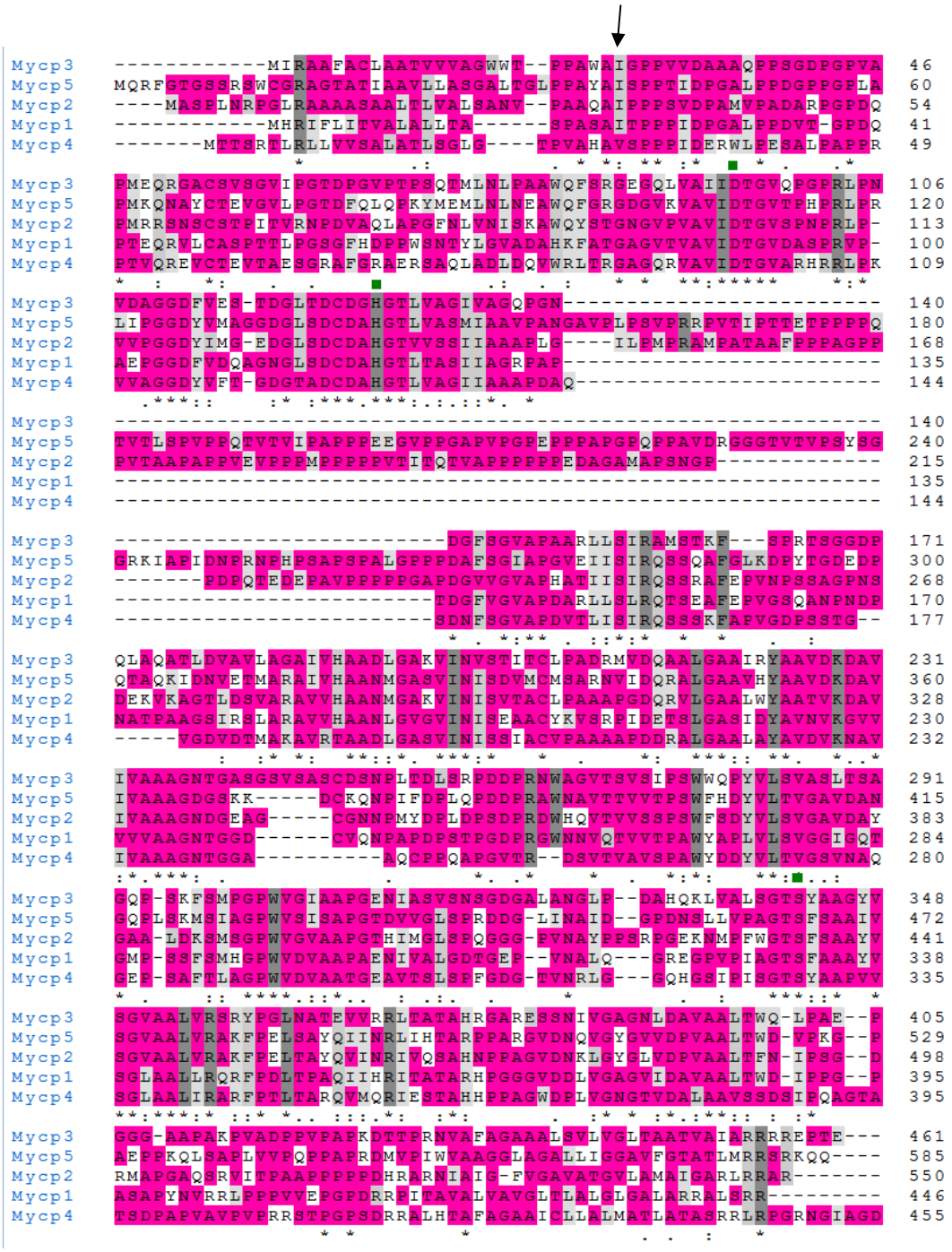


Fig. 2. *ClustalW alignment for protein sequences of five proteases of the mycosin family illustrating the highly conserved catalytic triad (Asp, His, Ser), which is indicated by green squares. An arrow shows the signal peptide.*

1.2 RNA polymerase

The transcription machinery of *Mt* is considered one of the best drug targets for treatment of tuberculosis. Its activity could be prevented by inhibition of RNA polymerase. Many compounds have shown some inhibitory activity against *Mt* RNA polymerase, but the only inhibitor that was approved for clinical use is Rifampicin (9). It is a natural product derived from *Amycolatopsis rifamycinica* and considered one of the first-line TB drugs. Its mechanism of action depends on Rifampicin binding to the β subunit of RNA polymerase and thus halting the transcription machinery (10). One of Rifampicin's advantages as a TB drug is that it targets mainly the RNA and DNA from the *Mt* cell and thus there is no effect on the host transcription mechanism (10). On the other hand, the effect of Rifampicin is weak during the stationary phase of *Mt* lifecycle or on the persisters (bacterial cells that neither grow nor die when antibiotics are used) (9). This is attributed to the long period of chemotherapy treatment for tuberculosis or some mutation that may lead to structural change in RNA polymerase. A previous study dealing with resistance of *Mt* cells to Rifampicin revealed slight levels of RNA polymerase activity after using this drug (11). This was justified by existence of other forms of RNA polymerase holoenzymes that reduce the affinity of RNA polymerase to Rifampicin (9).

The similarity between the genomes of the *Mt* and *Msm* is obvious, and for this reason the RNA polymerase subunit beta protein from *Msm* and the same protein in *Mt* are 91% identical in sequence. It is known that proteins from *(Mt)* are studied for exploring new drugs against the disease tuberculosis. However, past experience has shown that the *Msm* form of the protein is often easier to handle *in vitro.* Testing the inhibitors against RNAP from *M)* takes much time and the purification process is more difficult than RNAP from *Msm*. In addition, an earlier study reported that RNAP activity from *Msm* is higher than that of RNAP from *Mt in vitro* (10).

2. Aim of thesis

The aim of this thesis is to explore the possibility of cloning, constructing, and expressing MycP3 to produce more soluble material. Refolding can also be performed to achieve this goal. Crystallization and data collection can be performed to solve the structure of MycP3 protein.

The second protein in this study is *Msm* RNA polymerase. An expression vector containing the gene encoding for this enzyme had been constructed by collaborators. The aim of this study is to investigate if the culturing of the cells and expression can be performed for *Msm* RNA polymerase. Crystallization of purified protein and data collection at synchrotron can be performed to solve the structure of this protein.

3. Material and methods

3.1 Mycosin3

3.1.1 Sequence and structure predictions

Information about MycP3 and RNA polymerase from *Mt* was obtained from different databases. Sequences and structures of genes and proteins were collected from tuberculist database at http://tuberculist.epfl.ch/index.html and from Uniprot database at http://www.uniprot.org/. Further information about the protein such as extinction coefficient, isoelectric point, and amino acids types was found with the ProtParam tool at http://www.expasy.ch/tools/protparam.html. Sequence alignment was performed using the ClustalW program at http://www.ebi.ac.uk/Tools/clustalw2/index.html and the Uniprot alignment tool at http://www.uniprot.org/. Primers were designed and analyzed using IDT tool at http://eu.idtdna.com/Home/Home.aspx. Prediction of transmembrane helices was done using Hidden Markov Models for Rv0291 at http://tuberculist.epfl.ch/index.html

3.1.2 Primer design for MycP3

There are some considerations that should be taken when primers are designed. These include length, the content of G and C bases and melting temperature. As a general rule, primers should be between 25 and 45 bases in length, the melting temperature should be in the range of 55 to 65 °C, and the contents of primers should be at least 40% G and C bases.

In regard to the MycP3 gene, several constructs were made. Forward and reverse primers were designed for each construct and ordered from Invitrogen.

M2-1 construct primers are:

Forward ATGGCTCATCATCATCATCATCATGGCACCGATCCAGGC

Reverse GCTACGGTGTGGTGTCTTTGGG

M2-3 construct primers are designed to contain the his-tag in the reverse primer and thus the translated protein would contain the his-tag fused to the C-terminus. The primer sequences are:

Forward ATGGGCACCGATCCAGGCGTAC

Reverse GTTAATGATGATGATGATGATGCCAGGTCAGGGCCGCCAC

MycP3 full- length primers are:

Forward ATGGCTCATCATCATCATCATCATATCCGTGCCGC

Reverse TCATTCGGTGGGCTCCCTTC

M3 Construct primers are designed to contain the his-tag in the reverse primer and thus the translated protein would contain the his-tag fused to the C-terminus. The primer sequences are:

Forward ATGATCCGTGCCGCATTTG

Reverse TCAATGATGATGATGATGATGAGCGGCGAATGCGACGTTG

3.1.3 Isolation and amplification of a gene

In order to produce soluble protein, purify it and crystallize it, many constructs were designed using different primers. The best construct should have the ability to produce globular, compact protein without flexible and disordered ends.

Polymerase Chain reaction PCR was used to isolate and amplify the target gene from total TB genome H37RV. Techgene thermocycler was used and the PCR program was as shown (Table.1).

Table.1. PCR programm used for isolation and amplification of the target gene from total TB genome H37RV.

PCR condition	Number of cycles	PCR temperature	Time
Activation of enzyme & denaturing	1	95°C	2 min
Denaturing	25	95°C	1 min
Annealing of primers	25	57°C	1 min
Extension	25	72°C	1 min/kb (1 min- 30 sec)
Extension	1	72°C	10 min

Two types of amplifications were performed according to the type of DNA polymerase, the first reaction was done using PfuUltra polymerase (Invitrogen) which has a proofreading reading activity that confers accuracy for the amplified PCR product. The other reaction was done using Taq polymerase (Fermentas). Taq polymerase has the same role of PfuUltra polymerase which amplifying DNA fragment. The difference between them that Taq polymerase has less replication fidelity and higher error rate making it cheaper than PfuUltra polymerase. The master mix was prepared containing dNTP (Invitrogen) PCR buffer (Invitrogen), DNA polymerase (Invitrogen), forward and reverse primers, dH_2O and template. Agarose gel electrophoresis was used to verify the actual size of the PCR product and for purifying it. The constructs with the right size were cut out and transferred to DNA filtration columns placed in eppendorf tubes without lids which were centrifuged at 13000 rpm for 5 minutes, the DNA was collected and stored at -20 °C.

3.1.4 Ligation and transformation

The addition of adenine nucleotides is crucial for ligation of amplified PCR product to the vector. Adenylation was done by adding 17 µl DNA (purified with agarose gel electrophoresis)

to 2 µl Taq buffer, 0.5 µl dNTP and 0.2 µl Taq polymerase. The mixture was incubated at 72 °C for 10 minutes. Ligation reaction was achieved by adding 1 µl from previous mixture to 0.5 µl pEXP5 vector (Invitrogen), 1 µl water and 0.5 µl salt solution (from the pEXP5-TOPO kit), this mixture was incubated at room temperature for 20 minutes.

Transformation of the vector to Top 10 competent cells (Invitrogen) was performed by thawing 50 µl of the Top 10 cells on ice for 3 minutes and then adding 1 µl of the ligation solution and keeping the tube on ice for 5 minutes. Following this, the cells were heat shocked at 42 °C for 40 seconds and directly transferred to ice for additional 2 minutes. Finally, the solution was spread on LB plates containing 50 µg/ml ampicillin (Sigma) and the plate was incubated at 37 °C over night.

Of the grown colonies, 6 individual ones were picked and used to inoculate 6 times 500 µl LB with ampicillin in a 24 well culture plate. After 3 hours incubation at 37 °C, 10 µl were taken from each well and used as a template for a colony PCR reaction, which was performed according to above PCR program (Table.1). An additional 2 ml of LB media and ampicillin was added to each well. The plates were incubated at 37 °C over night. After growth, the culture was spun down and the pellet was used for plasmid preparation according to the protocol of the QIAprep Spin Miniprep kit (QIAGEN).

To confirm the correct orientation of the construct in relation to vector, an analytical PCR was performed using T7 universal reverse primer, which binds to the vector (Invitrogen) and the designed forward primer of the construct. The PCR program was performed according to above PCR program (Table1) and agarose gel electrophoresis was used to verify the accuracy of the PCR product depending on the band and its quality.

3.1.5 Transformation to BL21-AI competent cells

BL21-AI cells were used as the protein expression strain. These cells are known for their tight regulation and high expression yield of protein, especially proteins which are toxic to *E. coli*. The araBAD promoter is used and the expression is induced by L-arabinose. The transformation of the plasmid was done in the same way as described for transformation to Top 10 cells. The plates were incubated at 37 °C over night.

3.1.6 Expression Test

To test the expression of the target protein, the grown colonies were transferred to eppendorf tubes containing 2 ml LB and ampicillin. Then they were incubated at 37 °C for 4 hours till the OD_{600} reached 0.5-1.0. Prior to induction, 400 µl were taken as an uninduced sample for sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) analysis. The expression was induced by addition of 2 mg/ml arabinose at 37 °C for 3 hours. The induced culture was spun down at 11000 rpm for 5 minutes, the supernatant was discarded and the pellet was kept for SDS-PAGE analysis.

Large-scale expression was performed according to promising result of expression test which was analysed by SDS-PAGE and Phastgel technique(GE Health care). 25 ml of LB media and

ampicillin was inoculated with 5 cultivated colonies from BL21-AI plate and incubated at 37 °C for 4 hours till the OD_{600} reached the value 0.5-1.0.

The above 25 ml culture was used as a seeder for 1 L of growing media. For large-scale expression, a mixture of LB and minimal media was used as growth media; 1 liter was prepared consisting of 12.8g $Na_2H\ PO_4.7H_2O$, 3g KH_2PO_4, 1.5g NaCl, 1g NH_4Cl, 2g Triton X-100 and 1g yeast extract. 2 ml of $MgSO_4$ (1 M), 100 µl $CaCl_2$ (1M), 20 ml of 20% glycerol and 1 ml ampicillin (50 mg/ml). The starter culture was used to inoculate the above liter which was incubated at 37 °C for 4 hours till OD_{600} reached 0.5-1.0; at this point the expression was induced using L-arabinose to a final concentration of 0.2%. After 3 hours incubation at 37 °C, the cells were harvested by centrifugation at 4500 rpm for 20 minutes. The supernatant was discarded and the pellet was dissolved in 30 ml 1xSSP buffer containing 150 mM NaCl, 10 mM NaH_2PO_4 pH 7.5 and 1 mM EDTA. The solution was spun down at 4000 rpm for 12 minutes and the pellet was kept at –20 °C.

3.1.7 Protein Purification using IMAC (Native condition)

The pellet from the large-scale expression was dissolved in 15 ml of lysis buffer containing 50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 10 mM imidazole, 0.5% TritonX-100, 10% glycerol, 0.01 mg/ml RNase, 0.02 mg/ml DNase and 0.02 mg/ml lysozyme. The disruption of cells was performed using a Constant Cell disruptor system (Constant Systems Limited). 75 µl of phenylmethanesulfonylfluoride fluoride PMSF (protease inhibitor) was added to the lysate, which was centrifuged at 18000xg for 20 minutes. The pellet was stored at 4 °C and the supernatant was taken for further purification steps. To purify the expressed protein with his-tag, immobilized metal ion affinity chromatography (IMAC) was applied under native conditions using nickel–nitrilotriacetic acid (Ni-NTA) agarose slurry. The equilibration of the column was done using the lysis buffer (without Triton). Then the cell lysate was applied to the column and incubated for 45 minutes. The flow-through was collected for SDS-PAGE analysis. Afterwards, the column was washed using a buffer containing 50 mM $NaH_2\ PO_4$ pH 8.0, 300 mM NaCl, 20 mM imidazole, and 10% glycerol. A sample of wash fraction was saved for SDS-PAGE analysis. The elution of the target protein was done by applying an elution buffer containing 50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 250 mM imidazol, 10% glycerol. The eluted protein was collected and stored in 4 °C.

3.1.8 Refolding of inclusion bodies

The starter culture was prepared as above (section 2.1.6). LB media was used as growing media (instead of LB with minimal media that used the previous time). One liter from LB media was inoculated with the above starter culture. Incubation, induction and centrifugation of culture were performed as the previous time. The obtained pellet was collected and stored at –20 °C for further use. Two methods were used for refolding the inclusion bodies, the first one was the rapid dilution method and the second one the dialysis method. For each method, two liters of refolding buffer were prepared containing 50 mM sodium phosphate (38 ml NaH_2PO_4 0.5 M + 162 ml Na_2HPO_4 0.5 M) then 15 mM beta-mercaptoethanol (BME) was added and the pH was adjusted to 7.4 . The two liters were divided into two bottles, 1 M urea was added to the first liter (buffer A), the second liter without urea was called buffer B.

The refolding of inclusion bodies by the dialysis method was performed by first washing the pellet obtained from the large-scale culture. The washing was repeated four times using a buffer containing 50 mM sodium phosphate, 300 mM NaCl, 5 mM BME and 0.5% Triton X-100, pH 8.0. Following the washing, the pellet was resuspended in 20 ml of buffer B mixed with 8 M urea and centrifuged at 12000 rpm for 10 minutes. The supernatant was transferred to a dialysis tube, which was kept in buffer A for 4 hours in the cold room. Then the buffer was changed and the dialysis tube was kept in buffer B overnight in the cold room. The second day, the contents of the dialysis tube was transferred to an SS34 tube which was centrifuged at 18000 rpm for 15 minutes to spin down any aggregated or unfolded protein left in the tube.

The refolding of inclusion bodies using the rapid dilution method was performed by washing of the pellet obtained from large-scale culture in the same way described above. After centrifugation, the supernatant was slowly added drop-wise into one liter of buffer B while being stirred and then kept in the cold room overnight. An IMAC column was used to isolate the target his-tagged protein from the liter of buffer. The same steps of purification were performed and the protein was eluted as described in section 3.1.7.

SDS-PAGE analysis was performed for all samples obtained from the two methods to check the solubility and expression of the harvested proteins.

3.1.9 PD-10 column and size exclusion chromatography

In order to change the buffer and to reduce the salt concentration a PD-10 column (Amersham Pharmacia Biotech) was used. For this purpose a buffer was prepared consisting of 50 mM BisTris propane, 75 mM NaCl, 5% glycerol and 20 mM BME, the pH was adjusted to 7.5. The PD10 column was equilibrated with 25 ml of this buffer. The four eluted protein fractions from the Ni purification were pooled. The volume of 2.5 ml protein was loaded onto the column, the flow through was discarded and the protein was eluted using 3.5 ml of the above buffer.

A HiLoad 16/60 Superdex 75 (Pharmacia Biotech) column was used for high degree of protein purification by size. It was connected to an AKTA prime purification system (Pharmacia Biotech). The equilibration of the column was performed using the same buffer used above for PD-10 column. The eluted protein fraction was loaded on the column. The absorbance at OD_{280} was monitored and the fractions containing the target protein were collected.

3.1.10 Crystallization set up

To concentrate the purified protein, diafiltration was utilized with a Vivaspin column (Sartorious Stedium Biotech) and the molecular weight cut-off of 10 kDa. The protein solution was centrifuged at 8000 rpm and the protein concentration was checked each five minutes by measuring the OD_{280}.

The target proteins in this experiment were crystallized by vapor diffusion, using the sitting drop and hanging drop methods. The common approach of these two methods is to reduce the solubility of the protein solution. This occurs when the solvent molecules diffuse from the protein-precipitate droplet to the reservoir till equilibrium is reached. Gradually, the protein concentration increases and saturation of the solution occurs. Thereby, the proteins may organize into crystals if the conditions are optimal.

In the sitting drop method, crystal quick 96-well plate (MRC-2) was used, and the reservoirs were filled with 75 µl of precipitant from a crystal screen. In this experiment, JCSG crystal screen (Molecular Dimensions Limited) was used containing pre-made solutions for each of the 96 wells. A liquid dispensing robot (Douglas Instrument-Oryx) was used in order to dispense the crystallization trial drops accurately; the volume of each drop is 0.8 µl. In case of crystal formation by one of the conditions in the 96 well plates, the hanging drop method was used to optimize the crystallization conditions and to produce more crystals. In this case, a 24-well plate (QIAGEN) with screw-on lids was used. The reservoirs were filled with 500 µl of solution that was prepared according to the contents of the condition that produced the crystal. However, each reservoir contains different concentration in order to obtain low supersaturation of the solution where the crystal grows best. Inside the screw-on lid, 2 µl of the protein and 2 µl of condition buffer (obtained from the crystal screen) were placed. The plate was incubated and checked for crystal formation. When there was no growth, the droplets in the screw-on lids were seeded by touching the crystals obtained from the sitting drop method gently with a horsetail hair, which was then drawn through the droplets that didn't show crystal growth.

3.1.11 Crystal prepration for data collection

Obtained crystals were fished out of their drops using nylon loops before being transferred to a cryo solution, which was prepared containing 10% glycerol, 10% ethylenglycol and 80% of the condition buffer(0.1 M Bicine as buffer at pH 9.0, no salt, and 10% w/v PEG 6K as precipitant). Cryo solution works as a protective agent and prevents ice formation in and around the crystal. Ice will give strong diffraction rings and will disturb the measurement of crystal diffraction. Following cryo solution, the crystals were kept in liquid nitrogen to protect them from damage by the high intensity X-ray radiation used during data collection.

3.1.12 General Activity assay for testing MycP3 using B-casein as substrate

Different constructs were designed for MycP3; one of them was M2-1. Its protein was tested with B-casein substrate to check its activity. β-casein was dissolved to a concentration of 0.65% in 50 mM KPO_4 buffer pH 7.5 at 37 °C. The solution was heated at 80 °C for 10 minutes to denature the substrate protein. M2-1 protein was prepared as two 200 µl different stock solutions, one with a concentration of 3 mg/ml and the other with 6 mg/ml. 5 mM $CaCl_2$ and 5 mM BME were added to the above solutions and 50 mM of KPO_4 was used for diluting the protein. The reaction between substrate and M2-1 was performed by mixing 50 µl substrate and 10 ul protein solutions. The mixture was kept at 37 °C for 2h. Three solutions of 1 mg/ml trypsin, 1 mg/ml chymotrypsin and 1 mg/ml thermolysin were prepared in 50 mM KPO_4, 5 mM $CaCl_2$ and 5 mM BME. These solutions were used as positive controls. Then SDS PAGE was run to analyze the activity of the protease.

3.2 RNA polymerase

3.2.1 Msm RNA polymerase cloning & preparation of 7H9 media

An expression vector containing the *Msm* RNA polymerase was constructed at Astra Zeneca India. *Msm* cells were kindly obtained from there containing expression vector with histagged beta subunit of RNA polymerase. To express the RNAP in *Msm* cells, 7H9 was used as growth

medium. Firstly, 7H9 Broth base was prepared containing ammonium sulfate 4.2 mM, disodium phosphate 19.7 mM, monopotassium phosphate 8.1 mM, sodium chloride 57 mM, magnesium sulfate 0.46 mM, calcium chloride 0.004 mM, zinc sulfate 0.006 mM, copper sulphate 0.006 mM, ferric ammonium sulphate 0.092 mM, L-glutamic acid 3.7 mM, pyridoxine 0.006 mM, biotin 0.002 mM, Tween 80 0.424 mM and 2 ml glycerol. The volume was made up to 900 ml with water. The supplement for 7H9 broth was prepared containing Bovine albumin (fraction V) 754 mM, glucose 111 mM, catalase 0.116 mM. The volume was made up to 250 ml with water. It was kept at 4 °C for further use. After autoclaving, 100 ml of the supplement was added to 900 ml of the 7H9 broth to be used as growth media for *Msm.* 7H9 plates were prepared by warming up the refrigerated supplement to 45-50°C. Aseptically, 50 ml of the supplement was added to 450 ml of 7H9 broth media and hygromycine 50 μg/ml was used as antibiotic marker (CalBiochem). The mixture was poured into sterile plates, after getting solid the plates were kept in cold room.

3.2.2 Growing of *Msm* in 7H9 media

The *Msm* cells (strain MC^2155) obtained from India containing his-tagged beta subunit of RNAP were scratched and streaked on a 7H9 plate that was incubated at 37 C for three days. The grown cells were used to inoculate 10 ml starter culture which consisted of 10 ml 7H9 broth media, 0.5% Glucose, 0.05% Tween 80, 50 μg/ml of hygromycin. The culture was incubated at 37 °C for 24 hours. When the OD_{600} reached 0.9, the starter culture was used to inoculate one liter large scale autoinduction media which consists of 700 ml 7H9 broth media, 200 ml LB (as nutrient) 100 ml ADC enrichment media (without glucose because it inhibits the expression), 0.05 % glucose, 0.5 % glycerol, 0.05 Tween 80, 0.2% alpha lactose, 50 μg/ml of hygromycin. The culture was incubated at 37 °C for three days, then the cells were harvested by centrifugation at 4500 rpm for 20 minutes. The supernatant was discarded and the pellet was dissolved in 30 ml 1xSSP buffer containing 150 mM NaCl, 10 mM NaH_2PO_4 pH 7.5 and 1 mM EDTA. The solution was spun down at 4000 rpm for 12 minutes, the supernatant was discarded and the pellet was kept at – 20 °C (12).

In addition to the 7H9 large-scale autoinduction media described above, the 7H9 media was used for the second time but the induction was done using IPTG. The media was prepared in the same way as described above, 0.3 mM IPTG was added to the culture and the flask was kept at 37 °C for three days. Moreover, LB media was utilized as large-scale autoinduction media, so it was used without adding other ingredients except hygromycin 50 μg/ml.

3.2.3 Protein purification using IMAC

The RNAP protein was purified using IMAC following the same protocol as for MycP3; the only difference was adding 5 mM BME to the equilibration buffer.

3.2.4 Analysis of crystal quality

The assay was performed by preparing 10 mM rifampicin (SIGMA) as inhibitor (0.8 mg rifampicin was added to 100 μl H_20). 4 μl from this solution was mixed with 40 μl from RNAP protein. A crystallization trial was set up, robot and JSCG structure screen were used as described previously.

4. Results and Discussion

4.1 Mycosin 3

4.1.1 Protein primary and secondary structure analysis

The Uniprot and TubercuList databases were used to find the sequence of MycP3 gene and its translated protein. The gene length is 1386 bp translated to 461 amino acids with 46.12 kDa molecular weight. The protein parameters were computed using Expasy Proteomics server. According to this analysis, the MycP3 protein contains 5 cysteine amino acids in the whole protein sequence. This number of cysteines may affect negatively on the refolding process of MycP3protein because the wrong cysteine pairs could form disulfide bonds and thus lead to misfolding.

4.1.2 Protein truncation and primer design

The MycP3 protein contains an N-terminal signal sequence and a C-terminal transmembrane region (5): thus these regions are very rich in hydrophobic residues (Fig.3). As a result of this hydrophobicity the over expression of MycP3 in *E. coli* cells may produce insoluble protein. To increase the chances of getting MycP3 in soluble form, these hydrophobic regions could be removed by truncation of the protein at specific amino acids.

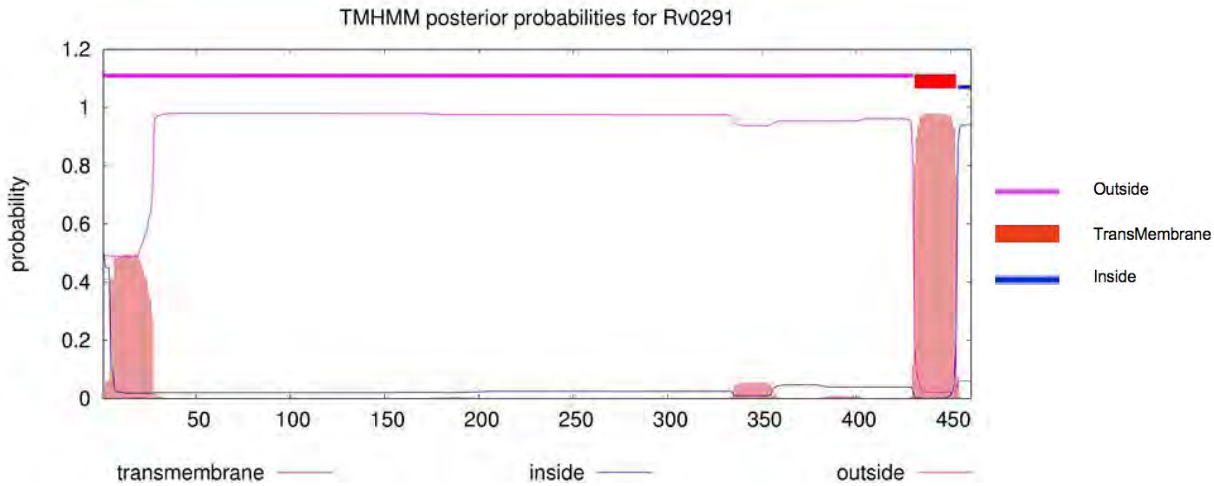


Fig. 3. *Transmembrane prediction for MycP3 protein using TMHMM Server v. 2.0. Predicted TM segments in N-terminal region are signal peptides from amino acid 9 to 19. The C- terminal is predicted to contain a transmembrane helix from amino acid 431 to 453.*

Several constructs of the MycP3 protein were made, all of them containing the active site residues Asp, His and Ser. To make a proper truncation, several amino acids were left before the active site residues. Therefore, M2-1, M2-3, M3 and full length constructs of MycP3 were designed (Table 1). After truncation M2-1 is 375 residues, M2-3 is 345 residues and M3 construct is 441 residues. A 6-histidine tag (his-tag) was attached to each construct to facilitate

the purification process using IMAC. The function of the his-tag is to separate the target protein from undesired proteins originating from the bacterial host. Following the cell disruption, the lysate is mixed with affinity media containing nickel, which binds to the his-tagged proteins. The removal of unbound protein is performed by washing the column using phosphate buffer. To elute the target protein, imidazole, which also binds to the nickel, is applied to the column and will replace the his-tag because of its high concentration (13). In this experiment, the primers were designed to contain a his-tag. Thus, eighteen bases encoding for 6 histidines were attached after the start codon. After the PCR reaction and expression of the target protein, the his-tag will be present at the N- or C-terminal of the construct (Table 2).

Table. 2 *Different constructs of MycP3 with their sequences. Attached His-tag is labeled with red color. The primers of each construct are also present.*

MycP3 Constructs	Sequence	Primer
M2-1	MAHHHHHHGTDPGVPTPSQTMLNLPAAWQFSRGEGQLVAIIDTG VQPGPRLPNVDAGGDFVESTDGLTDCDGHGTLVAGIVAGQPGND GFSGVAPAARLLSIRAMSTKFSPRTSGGDPQLAQATLDVAVLAGA IVHAADLGAKVINVSTITCLPADRMVDQAALGAAIRYAAVDKDA VIVAAAGNTGASGSVSASCDSNPLTDLSRPDDPRNWAGVTSVSIP SWWQPYVLSVASLTSAGQPSKFSMPGPWVGIAAPGENIASVSNSG DGALANGLPDAHQKLVALSGTSYAAGYVSGVAALVRSRYPGLN ATEVVRRLTATAHRGARESSNIVGAGNLDAVAALTWQLPAEPGG GAAPAKPVADPPVPAPKDTTP	Forward ATGGCTCATCATCATCATCATCATGGC ACCGATCCAGGC Reverse GCTACGGTGTGGTGTCTTTGGG
M2-3	MGTDPGVPTPSQTMLNLPAAWQFSRGEGQLVAIIDTGVQPGPRLP NVDAGGDFVESTDGLTDCDGHGTLVAGIVAGQPGNDGFSGVAPA ARLLSIRAMSTKFSPRTSGGDPQLAQATLDVAVLAGAIVHAADLG AKVINVSTITCLPADRMVDQAALGAAIRYAAVDKDAVIVAAAGN TGASGSVSASCDSNPLTDLSRPDDPRNWAGVTSVSIPSWWQPYVL SVASLTSAGQPSKFSMPGPWVGIAAPGENIASVSNSGDGALANGL PDAHQKLVALSGTSYAAGYVSGVAALVRSRYPGLNATEVVRRLT ATAHRGARESSNIVGAGNLDAVAALTWHHHHHH	Forward ATGGGCACCGATCCAGGCGTAC Reverse GTTAATGATGATGATGATGATGCCAG GTCAGGGCCGCCAC
M3	MIRAAFACLAATVVVAGWWTPPAWAIGPPVVDAAAQPPSGDPGP VAPMEQRGACSVSGVIPGTDPGVPTPSQTMLNLPAAWQFSRGEG QLVAIIDTGVQPGPRLPNVDAGGDFVESTDGLTDCDGHGTLVAGI VAGQPGNDGFSGVAPAARLLSIRAMSTKFSPRTSGGDPQLAQATL DVAVLAGAIVHAADLGAKVINVSTITCLPADRMVDQAALGAAIR YAAVDKDAVIVAAAGNTGASGSVSASCDSNPLTDLSRPDDPRNW AGVTSVSIPSWWQPYVLSVASLTSAGQPSKFSMPGPWVGIAAPGE NIASVSNSGDGALANGLPDAHQKLVALSGTSYAAGYVSGVAALV RSRYPGLNATEVVRRLTATAHRGARESSNIVGAGNLDAVAALTW QLPAEPGGGAAPAKPVADPPVPAPKDTTPRNVAFAAHHHHHH	Forward ATGATCCGTGCCGCATTTG Reverse TCAATGATGATGATGATGATGAGCGG CGAATGCGACGTTG
Full Length	MHHHHHHIRAAFACLAATVVVAGWWTPPAWAIGPPVVDAAAQP PSGDPGPVAPMEQRGACSVSGVIPGTDPGVPTPSQTMLNLPAAWQ FSRGEGQLVAIIDTGVQPGPRLPNVDAGGDFVESTDGLTDCDGHG TLVAGIVAGQPGNDGFSGVAPAARLLSIRAMSTKFSPRTSGGDPQ LAQATLDVAVLAGAIVHAADLGAKVINVSTITCLPADRMVDQAA LGAAIRYAAVDKDAVIVAAAGNTGASGSVSASCDSNPLTDLSRPD DPRNWAGVTSVSIPSWWQPYVLSVASLTSAGQPSKFSMPGPWVG IAAPGENIASVSNSGDGALANGLPDAHQKLVALSGTSYAAGYVSG VAALVRSRYPGLNATEVVRRLTATAHRGARESSNIVGAGNLDAV AALTWQLPAEPGGGAAPAKPVADPPVPAPKDTTPRNVAFAGAAA LSVLVGLTAATVAIARRRREPTE	Forward ATGGCTCATCATCATCATCATCATATC CGTGCCGC Reverse TCATTCGGTGGGCTCCCTTC

4.1.3 Cloning, ligation and transformation

The *Mt* genomic DNA (H37Rv) was kindly provided by Annette Roos and the MycP3 gene was successfully cloned. The primers containing his-tag were used for DNA amplification by PCR. The agarose gel was used to assess and confirm the size of truncation PCR product (Fig.4). According to the gel picture, the M3 clone was successfully amplified from *Mt* genomic DNA (H37Rv) and its band is corresponding to the size of 1300 bp. The same clone was used as a template and histag was attached to the reverse primer.

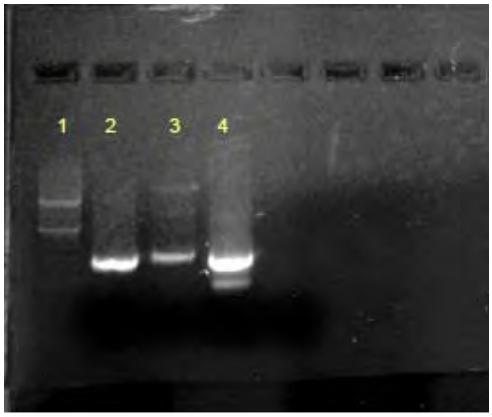


Fig. 4. *PCR using the amplified M3 from previous PCR as template, the his-tag was attached to the reverse primer, thereby introducing a C-terminal his-tag to M3. Lane1 marker, Lane 2 M3 without histag, Lane 4 M3 with his-tag , lane 3 MycP3 full length.*

The amplified PCR product with his tag was extracted from the gel and ligated to PEXP5/CT vector using taq polymerase. Following the ligation, the vector was transformed to Top10 competent cells. Selection of positive transformed cells was achieved depending on the ampicillin-resistant gene that was constructed in the vector. For the M2-1 construct, around 40 colonies were obtained after transformation and the same number of colonies was obtained after transformation of the M2-3 construct.

Analytical PCR was utilized to verify the integrity of the PCR product into the vector and to confirm the orientation of ligation. The reaction was done using a forward primer from the construct and reverse primer from vector. In regard to M2-1 and M2-3 constructs, analytical PCR was conducted and the positive clones were confirmed by agarose gel. For M2-1 construct, 6 colonies were picked for analytical PCR and 3 of them were positive. Regarding M2-3, two colonies were picked and both of them are positive (Fig. 5).

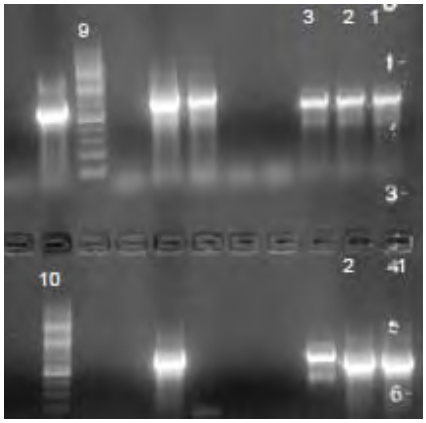


Fig. 5. *Analytical PCR for M2-1 and M2-3 constructs. Lanes 1,2, and 3 from the upper row explains the positive clones of M2-1 while lanes 1 and 2 of the bottom half illustrates the positive clones of M2-3.*

Analytical PCR was done for full length MycP3: only one clone was positive as it appears on lane 5 of the right gel picture (Fig. 6) while other faint bands in the gel indicate the low concentration of the plasmid.

In regards to the M3 construct, analytical PCR was performed and agarose gel electrophoresis was used to confirm the positive clones. The obtained bands were faint, even for the positive control where T7 reverse and forward primers were used. There is one faint positive clone as it appears on lane 2 of the bottom left gel. The strong band in lane 4 of this experiment was obtained after using M3 as template and the forward and reverse primers belonging to the construct (Fig. 6). The faint bands from the positive control and strong band where construct primers were used could be attributed to some contamination with the vector primers. The lanes with no bands indicate no DNA amplification, which means that the insert is oriented in the wrong direction.

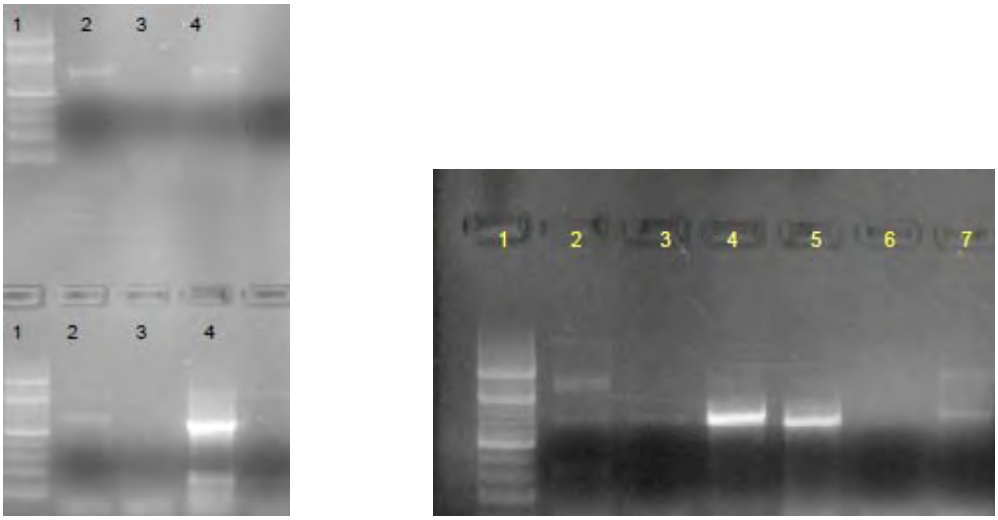


Fig. 6. *Analytical PCR for full length and M3 construct. The left gel picture shows the analytical PCR for full length, lanes 2 and 4 of top half illustrate faint bands belong to M3 positive control. In the bottom half, lane 2 explain faint positive clone of M3 using T7 forward primer and reverse primer from M3 construct. Lane 4 shows a strong band for M3 using forward and reverse primers belong to the M3 construct. The right gel picture shows analytical PCR for full length. Lane 4 explains positive control for full length while lane 5 shows positive clone for full length.*

Following analytical PCR, positive clones were confirmed by DNA sequence analysis (Uppsala Genome Centre, Rudbeck Lab). The sequence results were analyzed using 4peaks software, the M2-1, M2-3 and M3 sequences were correct. The result of full length was incorrect

The result of the expression test for M2-1 and M2-3 construct was interesting (Fig. 7). Some clones from M2-1 (lanes 5 and 7) expressed the target protein with the right size which is 37.7 KDa. In lane 9, only one clone from M2-3 expressed the right protein, which in this case is 34.9 kDa (there was an emergence of the SlyD protein in lane 9, it is considered as a common contaminant in purification of his-tagged proteins). The expression test for M3 and full length constructs was unsuccessful; the target protein with the actual size was not expressed after induction. This failure could be explained by one of the following reasons; the low concentration of plasmid that transformed to BL21-AI, the toxicity of the target protein, or mutations leading to frame shifts that could be detected through sequence analysis. The existence of these mutations results in emergence of many stop codons in the sequence which prevents the transcription of the gene and thus no translation and expression of the target protein.

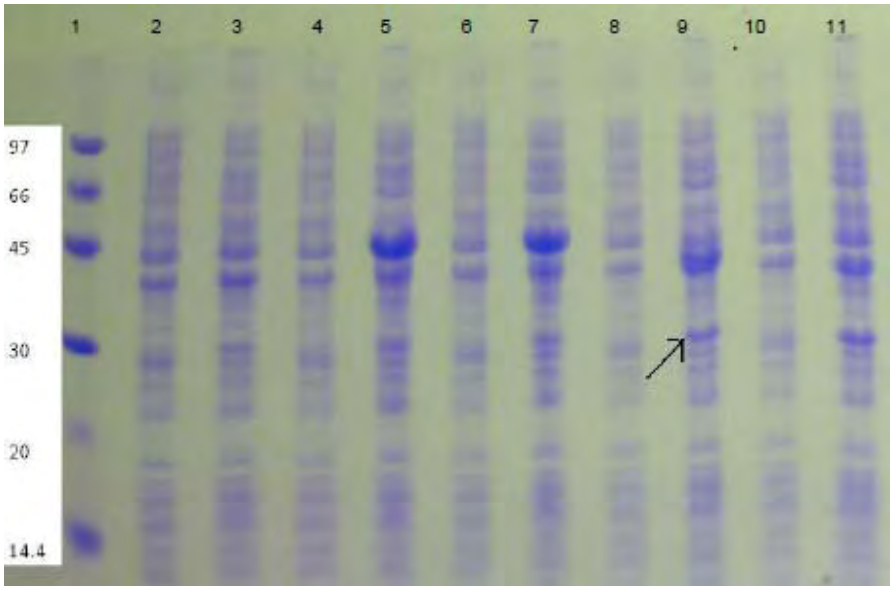


Fig. 7 *20% homogenous SDS gels of test expression. Lane 1 ladder, Lane 2 pre M2-1, Lane 3 induced M2-1, Lane 4 pre M2-1, Lane 5 induced M2-1, lane 6 pre M2-1, lane 7 induced M2-1, lane 8 pre M2-3, lane 9 induced M2-3, Lane 10 pre M2-3, Lane 11 induced M2-3. The arrow in lane 9 points to SlyD protein.*

4.1.4 Large scale expression, purification using IMAC and refolding

According to the positive results from the expression test, large scale expression was carried out to produce a larger amount of the target protein. At OD600=0.5-1.0, the cells were induced with L-arabinose, and after 3 hours incubation at 37 C, the obtained pellet was disrupted and the lysate was purified using IMAC.

The above steps were applied on the M2-3 construct (Fig.8), samples before and after induction were taken and analyzed by SDS PAGE. The band of after induction sample matches the size of protein of interest which is 34.9 kDa. After disruption, the cells were spun down; supernatant was analyzed as soluble fraction while the pellet was analyzed as insoluble fraction. The lane of supernatant contains several bands reflecting different cell proteins with different sizes. In the lane containing the pellet, a strong band appears revealing the formation of inclusion bodies due to over-expression of the protein of interest, which results in incorrect folding and thus precipitation into large clusters. In other words, the emergence of inclusion bodies is due to heterologous expression of foreign genes in *E.coli* (14). After applying the lysate to the column, the washing step was done and the flow through was collected and analyzed by SDS-PAGE, the lane that contains the flow through samples displays a smear of bands representing unbound protein (without his-tag) and cellular DNA. The lanes representing the elution contain no bands. Elution of the target protein was done using imidazole and there was no harvested protein due to formation of inclusion bodies.

To refold the protein, the dialysis method was used. The concentrated denatured protein was dialysed against refolding buffers overnight. SDS PAGE analysis reveals the presence of the

target protein in the pellet while there is no soluble protein in the supernatant (Fig. 8), which indicates that the refolding process has failed with the M2-3 construct.

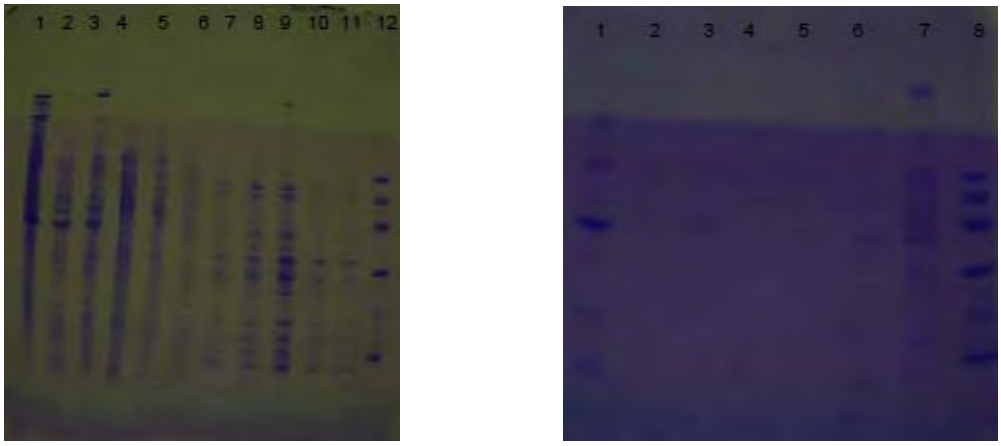


Fig. 8. *Left SDS gel picture explains the purification of M2-3 protein using IMAC. Lane 1 is the insoluble fraction, showing inclusion bodies. Lanes 7-11 are the eluted fractions, the target protein is not purified. The right SDS picture explains the refolding process of M2-3 inclusion bodies using dialysis. Lane 1 reveals the presence of the target protein in the pellet while there is no soluble protein in the supernatant in lane 2.*

The same steps and conditions as for M2-3 were used for expression and purification of the M2-1 construct. The result was again formation of inclusion bodies in insoluble fractions and no soluble protein in the eluted fractions (Fig. 9). This time, two methods of refolding were applied: the rapid dilution and the dialysis methods. Both methods were successful in removing and refolding the inclusion bodies. It is noticeable that the concentration of the cultivated protein looks higher with the dialysis method compared to the rapid dilution (Fig. 10).

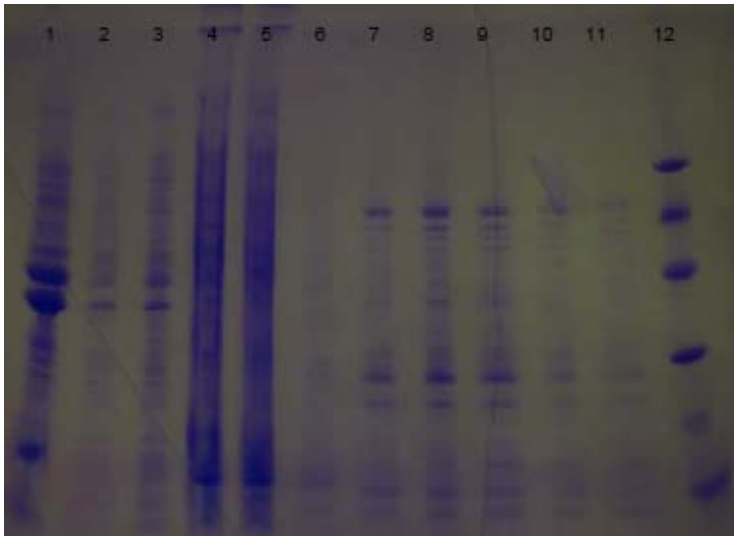


Fig. 9. *SDS gel picture explain the purification of M2-1 protein using IMAC. Lane 1 is the insoluble fraction showing the formation of inclusion bodies. Lanes 7-11 are the eluted fractions, the target protein was not purified.*

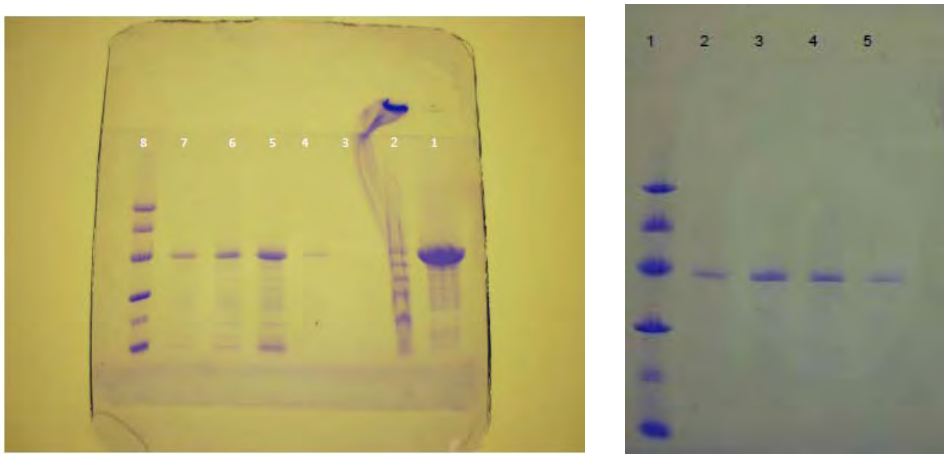


Fig. 10. *Left SDS picture explain the success of dialysis method in refolding of inclusion bodies. Lane 1 supernatant after dialysis, lane 2 pellet after dialysis, lane 3 wash fraction, Lanes 4-7 eluted fractions of M2-1 protein, lane 8 ladder. The right SDS showing the removal of inclusion bodies using rapid dilution method, lanes 2-5 illustrates bands matching the size of M2-1protein.*

To confirm the above result and to produce good amount of the soluble protein, the above experiment was done again but this time LB media was used as growing media instead of LB with minimal media. The reason is that LB is richer in nutrition. In addition, the main purpose of using LB with minimal media was to produce soluble protein but once this was found to not work, then LB was used to produce more inclusion bodies which can be refolded by one of the above methods. Another point behind using LB with minimal media was to reduce the expression of protein by slowing down the metabolism but in case of LB media, the main target

was to maximize the protein expression to get good amount of inclusion bodies, which can be refolded, for reaching the goal of protein solubility.

The same conditions and parameters were applied for purification and refolding of inclusion bodies. To get the soluble protein highly purified, an additional purification step was taken using size exclusion chromatography. A PD-10 column and a HiLoad 16/60 Superdex 75 column were used. Fractions corresponding to the Superdex peak were analyzed using SDS PAGE; according to the marker they match the right size of the target protein, which is 37.7 KD (Fig. 11). The protein was concentrated using Vivaspin concentrator to 2.5 mg/ml.

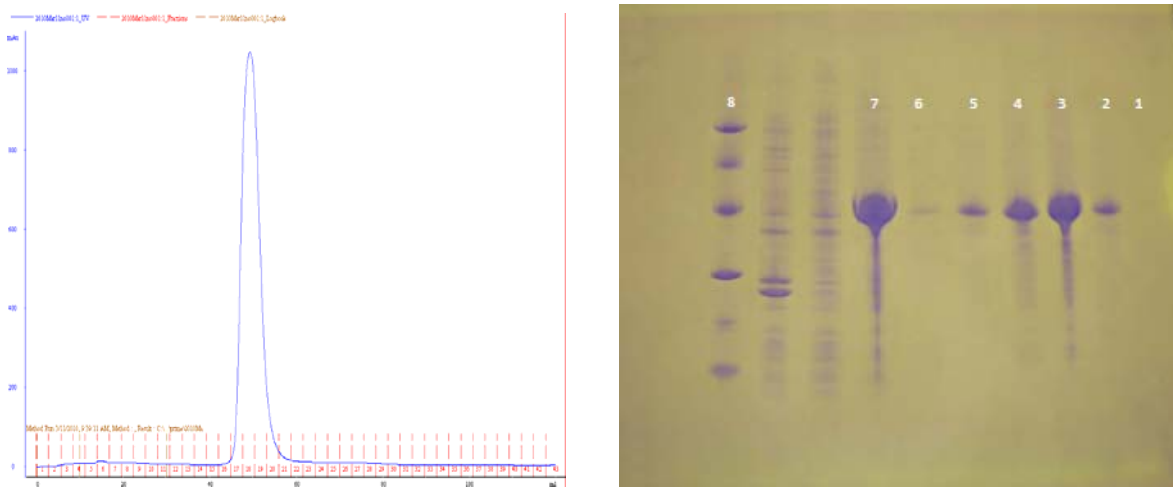


Fig. 11. *Left picture illustrates the final purification of M2-1 protein using Superdex 75 column. The peak explains the purified fractions of M2-1 protein that was analysed by SDS(right picture). Lanes 2-6 explaining the tubes with high fractions where the protein was eluted. Lane 7 sample after using PD10 column and before using Superdex column.*

4.1.5 Crystallization

A crystallization trial was set up using the sitting drop method and JCSG crystal screen. After 2-3 days of incubation, many crystals appeared, having the shape of needle-like or small rod-like structures. The quality of most of them was not satisfying and many were formed from salt source. Some crystals were interesting and promising; they looked like protein crystals according to the polarization test. The promising crystallization condition was 0.1 M Bicine as buffer at pH 9.0, no salt, and 10% w/v PEG 6K as precipitant (Fig. 12).

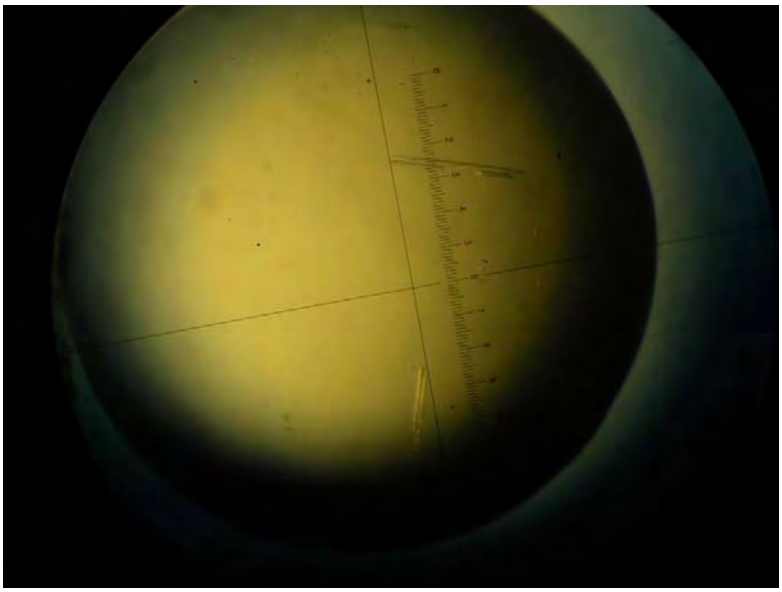


Fig. 12. *Promising crystals for M2-1 protein from sitting drop method using JCSG crystal screen.*

To optimize the crystallization conditions, hanging drop method was utilized and a 24 wells plate with screw-on lids was used. After many days of incubation, there was no crystal formation. The droplets in the screw-on lids were seeded by touching the crystals obtained from sitting drop method gently with a horsetail hair, which was drawn through the droplets in the screw-on lids. The plate was checked again after three days and there was still no growth, this was attributed to the difference between the kit solution used for sitting drop and the home made solution that was used for the hanging drop plate. Also the kinetics of the sitting drop plate is different from the hanging drop method.

4.1.6 General assay for testing MycP3 activity

M2-1 protein was tested with β-casein substrate to check its activity. $CaCl_2$ was added to the enzyme substrate mixture to optimize the reaction. Thermolysin and trypsin were used as positive controls. According to SDS PAGE analysis, the M2-1 protein is not active because it did not digest the substrate. On the other hand, SDS analysis shows the activity of thermolysin as protease when it digests the β-casein substrate. The activity assay was performed again but this time the $CaCl_2$ was not added to the reaction. The enzyme was found to be inactive as it did not digest the substrate, which means that there is no effect of $CaCl_2$ on enzyme activity (Fig. 13).

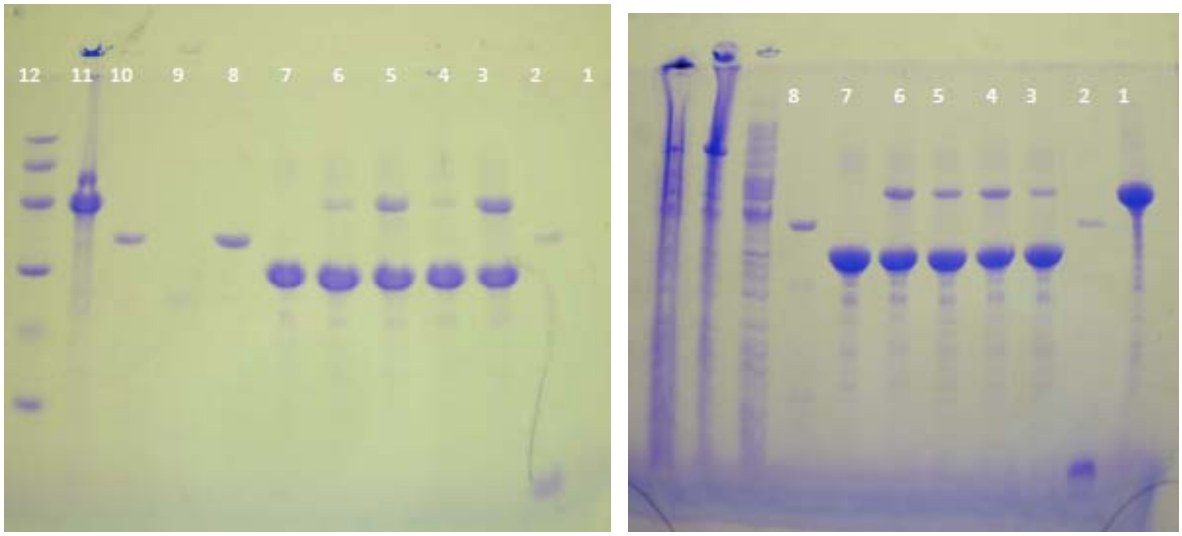


Fig. 13. *The left SDS picture shows M2-1 enzyme activity as protease with* β*-casein substrate. Lane 2 thermolysin with substrate as positive control, Lanes 3-6 reveal no digestion of B-casein by M2-1enzyme. Lane 7* β*-casein (substrate) without enzyme, lane 8 thermolysin without substrate, lane 11 pure M2-1 protein, lane 12 ladder. The right SDS picture explains the same reaction but this time* $CaCl_2$ *was not used. Lane1 pure M2-1 protein, lane 2 thermolysin with substrate as positive control, lanes 3-6 reveal no digestion of* β*-casein by M2-1enzyme, lane 7 B-casein substrate, lane 8 thermolysin without substrate*

According to the above result, M2-1 protein activity was tested by our collaborators using a panel of AMC (aminomethyl coumarin) substrates (15). The experiment was performed and it revealed that the M2-1 protein is inactive against these substrates.

The above result of M2-1 protein inactivity against different substrates despite its solubility could be attributed to incorrect refolding of this protein. Many factors may be responsible for protein misfolding. One of them is the formation of incorrect disulfide bonds. The M2-1 protein contains 5 cysteines, which means there is the potential of misfolding to occur during the refolding process because the wrong cysteine pairs could form. For this reason, BME was added to keep the cysteine reduced and prevent the formation of disulfide. To check the existence of possible disulfide bonds in the M2-1 protein, BME was used and SDS PAGE was run to compare the protein samples with BME and without BME (Fig. 14). According to this experiment, there is no disulfide bond in the M2-1 protein that may lead to incorrect folding. Therefore, the lack of correct disulfide maybe the reason behind the inactivity of M2-1 as the protein may need a disulfide to fold correctly.

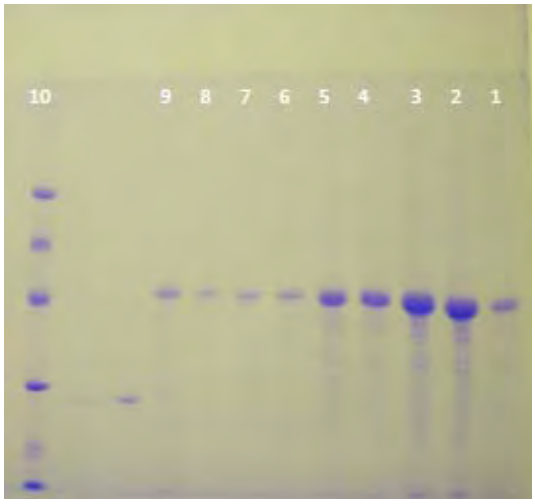


Fig. 14. *Reaction performed to check the existence of disulfide bonds in M2-1 protein. Lanes 3 and 4 were run without BME while lanes 2 and 5 contain samples run with BME. Apparently there is no difference between bands in these lanes and thus mean no existence of disulfide in M2-1 protein.*

To make a comparison between M2-1 construct and M3 construct, an experiment was made and the same steps and reaction parameters were applied to each construct. Following the culturing, the inclusion bodies were refolded by the dialysis method. An SDS gel was used to analyze the proteins solubility. It is obvious that the M2-1 protein was successfully harvested as a soluble fraction after dialysis while there was no soluble protein obtained from M3 after refolding (Fig. 15). According to this result, it is clear that the M2-1 is the most promising of the MycP3 constructs.

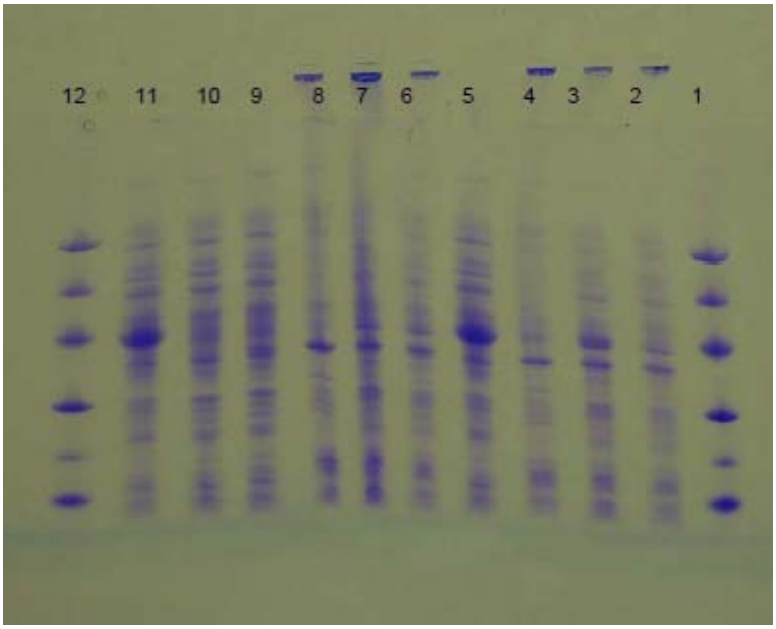


Fig. 15. *Comparison between M2-1 & M3 constructs. Soluble protein was obtained from M2-1 while there is no protein obtained from M3. Lane 1 ladder, lane 2 M2-1 uninduced, lane 3 M2-1 induced, lane 4 M2-1 pellet after dialysis, lane 5 M2-1 supernatant after dialysis, lane 6 M3 uninduced, lane 7 M3 induced, lane 8 M3 pellet after dialysis, lane 9 M3 supernatant after dialysis.*

4.2 RNA Polymerase

4.2.1 Bacterial strains and culture conditions

Construction of hexa-histidine tagged RpoC gene was done at Astra Zeneca India. pNCO26 was the final suicide vector. It was constructed to carry a 1042 bp fragment from 3' end of the RpoC gene, which contains DNA coding for his-tag. This fragment was cloned upstream of the hygromycin cassette. Further, 1039 bp from downstream RpoC gene was cloned to the other side of the hygromycin cassette. The above pNCO26 vector was treated with alkali and transformed to *Msm* cells mc^2155 by electroporation. Relying on the homologous recombination strategy, selection of double recombinants was done in the presence of hygromycin. SM07 clone was chosen as one of the double recombinants and used as a genomic mutant of RpoC for further reactions (10).

Msm cells mc^2155 containing the above vector were grown in 1 l of autoinduction media. SDS PAGE was used to check the expression and solubility of RNAP protein. Thereby, no protein was obtained in this experiment and this was attributed to some defect with the preparation of autoinduction media or as a result of improper purification.

To avoid the mistake that happened in the previous experiment, *Msm* cells mc^2155 were grown in 1 l media consisting 7H9 broth supplemented with LB and other ingredients. The difference was that this time that ADC enrichment was not used (only traces of catalase was added to media). Furthermore, the culture was induced using IPTG. The purification with IMAC was

monitored with SDS-PAGE, which revealed the expression of B-subunit of RNAP from SM07 (Fig. 16).

To confirm the role of IPTG in the induction of RNAP expression, the above experiment was done and the same steps and parameters were applied. The IPTG was not added, and the autoinduction media was used. SDS PAGE analysis demonstrated the expression of B-subunite of RNAP (Fig. 16). This result explains no great effect of IPTG on the induction of the target protein and that autoinduction media is useful for its expression.

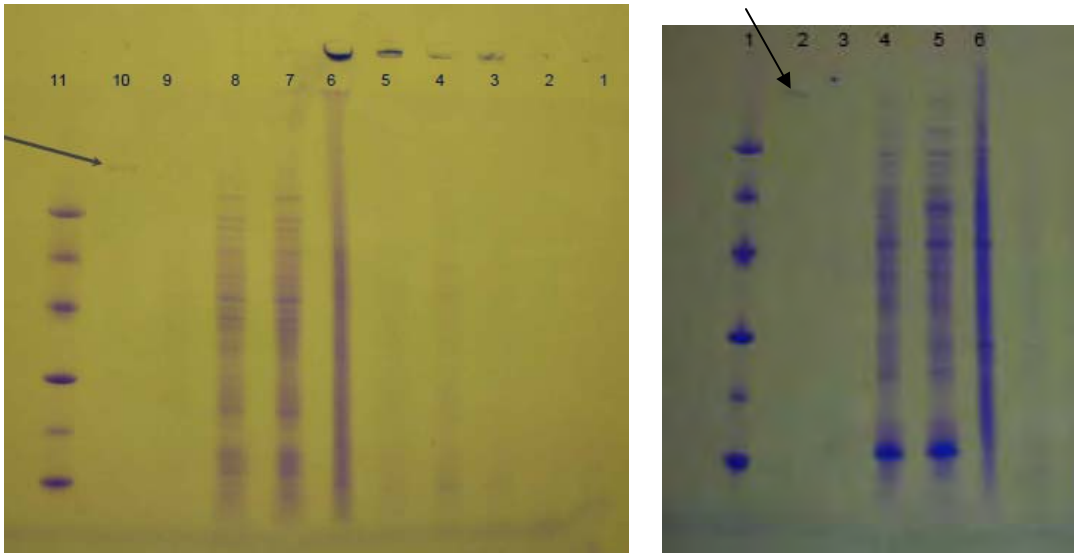


Fig. 16. *The left SDS illustrates the purification of beta subunit of RNAP using IPTG for induction. Lane 10 explains the eluted fraction and the arrow point to the band that corresponding to the size of RNAP B-subunit. Lane 9 wash fraction, lane 8 flow through, lane 7 soluble fraction, lane 6 insoluble fraction. The right SDS picture reveals the same fact but this time the expression of RNAP was performed using autoinduction media. Lane1 ladder, lane 2 eluted protein, lane 3 wash fraction, lane 4 flow through.*

Moreover, to check the possibility of obtaining a soluble RNAP protein from different culture, LB was used as growing autoinduction media. This time 7H9 was not used and the other minerals were not added also. The culture was incubated for five days at 37 °C then purified using IMAC. Then SDS PAGE was run and it was interesting that the polymerase β-subunit RNA protein was obtained also when LB was used (Fig. 17). Its concentration is low compared to protein yield from other growing media (IPTG induced media and autoinduction media). In addition, the obtained results from the above confirm the molecular weight of RNAP protein which is around 146 KDa. The same result was obtained in previous studies but from different media and methods (9,10).

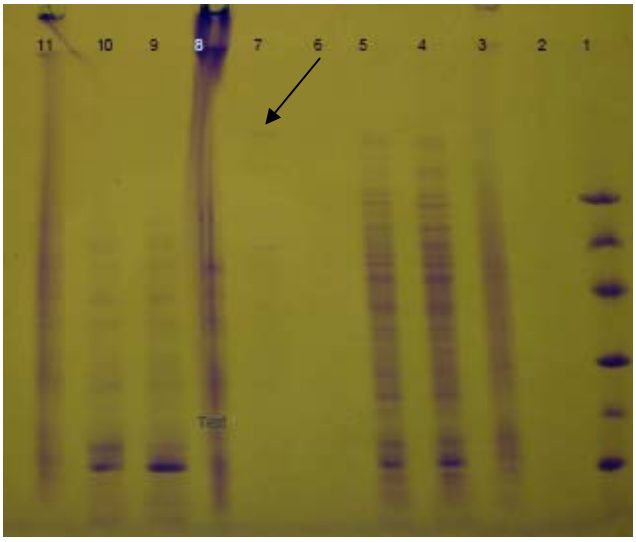


Fig. 17. *Purification of B- subunit RNAP using LB as autoinduction and growing media. the arrow points to band corresponding the size of the target protein.Lane3 insoluble fraction, lane 4 soluble fraction, lane 5 flow through, lane 6 wash fraction, lane 7 eluted protein.*

4.2.2 Crystallization of *M. smegmatis* RNA polymerase

The RNAP protein that was harvested from IPTG induced media and the other fraction obtained from autoinduction media were pooled together and concentrated using Vivaspin concentrator to 7.0 mg/ml. A crystallization robot and JCSG structure screen were used for crystallization trial following the sitting drop method. The plate was incubated for three days, and then there was formation of many crystals in different conditions. Most of them were salt crystals according to polarization test. There were some promising crystals with a good quality in one condition. They were needle shaped and looked like protein crystals. This successful condition was 0.1 M CAPS as buffer (*N*-cyclohexyl-3-aminopropanesulfonic acid), no salt, pH 10.5 and 40% v/v MPD (2-methyl-2,4-pentandiol) as precipitant.

Then hanging drop method was used for optimization of crystallization conditions. 24 wells plate with screw-on lid was incubated for many days but there was no formation of crystals. The reason behind this was explained by the main differences between hanging and sitting drop methods, like the volume of protein drop, the different kinetics and the difference between the home made solution used in hanging drop and the kit solution used in sitting drop.

4.2.3 Analysis of crystal quality

This experiment was performed to see if adding the inhibitor could improve the quality of the crystals. Often an inhibitor will make the protein molecules pack in a more ordered fashion and the crystals will form better.

To analyze the crystal quality of RNAP, rifampicin was used as inhibitor. It was added to the isolated polymerase β-subunit before crystallization. Crystallization trials were performed using sitting drop technique. 4 µl of 10 mM rifampicin was added to 40 µl of RNAP and robot was

used to dispense the above mixture over JCSG plate. After three days of incubation, there was formation of needle-shaped crystals like those obtained when the crystallization trial was performed without rifampicin. Also, these crystals were formed in the same condition of previous experiment.

Before data collection, the crystals from sitting drop method were transferred to cryo solution which was prepared according to contents of the condition that produced the crystals (500 µl was prepared containing 200 µl MPD (40% v/v) + 100 µl CAPS + 200 µl water). The MPD in this solution has the same protective role as glycerol in the standard cryo solution. The crystals were immediately taken to liquid nitrogen. Data were collected at European Synchrotron Radiation Facility (Grenoble, France). ESRF. Unfortunately, there was no diffraction obtained from RNAP protein crystal due to ice formation around the crystals that might result from insufficient cryo protection.

5. Conclusion

For MycP3, different constructs were designed and cloned to vector. The attempt to produce soluble protein from full-length construct was unsuccessful because of the hydrophobic residues in the signal peptide and transmembrane region. Also, no protein expression obtained from M3 construct and this was attributed to law concentration of the plasmid. M2-3 construct yielded inclusion bodies and the refolding of them was unsuccessful. The most promising construct was M2-1, the inclusion bodies were successfully removed and soluble protein was harvested. Following crystallization, promising crystals were produced but no diffraction was obtained. The activity assay was performed for M2-1 protein and it was inactive against different substrates. The future of MycP3 rely on designing new constructs for it and changing the refolding protocol and buffers for M2-3 to overcome the problem of inclusion bodies. To obtain a good diffraction of MycP3 crystals, optimization of crystallization conditions should be achieved and proper cryo protection must be found. The inactivity of M2-1 protein could be solved by changing the buffers and making some modification in the refolding protocol.

In regard to RNA polymerase, it was successfully expressed using 7H9 and LB as autoinduction media. Also its expression was performed using IPTG as inducer. Needle shape Crystals were produced from this protein but no diffraction was observed. In the case of RNA polymerase, optimizing the crystallization conditions as well as finding a proper cryo solution to protect the crystals during data collection are needed.

Acknowledgments

I would like to show my sincere gratitude to my supervisors Torsten Unge and Annette Roos for their inspiration, knowledge, patience, and for giving me a chance to work with interesting enzymes.

Also, this is great opportunity to express my respect to Nina for her help and technical support in the lap. Moreover, am pleased to thank Lu Lu for the good discussion about proteins and crystallography. Xiaohu, for being a good office mate. Anna J and Agata for taking my crystals to ERSF. Avinash, Cha San, and Ana L for answering a lot of my questions and for nice time we had together. Taha, for helping me with my computer problems and for the endless talks about everything between earth and sky. It is my pleasure to thank Dr. Salwa and Dr. Yahia from Central Veterinary Research Lab Khartoum, Sudan without them this thesis would not be possible.

Finally, I am heartily thankful to Meriam, who is no more, for her encouragement & support and for making my life in Uppsala memorable.

References

1. Brennan, P. J. & Nikaido, H. (1995). The envelope of mycobacteria. *Annu Rev Biochem* **64**, 29-63.
2. World Health Organization. (2009).
3. http://xray.bmc.uu.se/alwyn/DotMac/RAPID/RAPID.html (2010), corresponding address : alwyn@xray.bmc.uu.se , website visited July 2010.
4. Maeda, H. (1996). Role of microbial proteases in pathogenesis. *Microbiol Immunol* **40**, 685-99.
5. Brown, G. D., Dave, J. A., Gey van Pittius, N. C., Stevens, L., Ehlers, M. R. & Beyers, A. D. (2000). The mycosins of Mycobacterium tuberculosis H37Rv: a family of subtilisin-like serine proteases. *Gene* **254**, 147-55.
6. Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E., 3rd, Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M. A., Rajandream, M. A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J. E., Taylor, K., Whitehead, S. & Barrell, B. G. (1998). Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* **393**, 537-44.
7. Ribeiro-Guimaraes, M. L. & Pessolani, M. C. (2007). Comparative genomics of mycobacterial proteases. *Microb Pathog* **43**, 173-8.
8. Converse, S. E. & Cox, J. S. (2005). A protein secretion pathway critical for Mycobacterium tuberculosis virulence is conserved and functional in Mycobacterium smegmatis. *J Bacteriol* **187**, 1238-45.
9. Dey, A., Verma, A. K. & Chatterji, D. Role of an RNA polymerase interacting protein, MsRbpA, from Mycobacterium smegmatis in phenotypic tolerance to rifampicin. *Microbiology* **156**, 873-83.
10. Mukherjee, R. & Chatterji, D. (2008). Stationary phase induced alterations in mycobacterial RNA polymerase assembly: A cue to its phenotypic resistance towards rifampicin. *Biochem Biophys Res Commun* **369**, 899-904.
11. Hu, Y., Mangan, J. A., Dhillon, J., Sole, K. M., Mitchison, D. A., Butcher, P. D. & Coates, A. R. (2000). Detection of mRNA transcripts and active transcription in persistent Mycobacterium tuberculosis induced by exposure to rifampin or pyrazinamide. *J Bacteriol* **182**, 6358-65.
12. Goldstone, R. M., Moreland, N. J., Bashiri, G., Baker, E. N. & Shaun Lott, J. (2008). A new Gateway vector and expression protocol for fast and efficient recombinant protein expression in Mycobacterium smegmatis. *Protein Expr Purif* **57**, 81-7.
13. Hengen, P. (1995). Purification of His-Tag fusion proteins from Escherichia coli. *Trends Biochem Sci* **20**, 285-6.
14. Tsumoto, K., Ejima, D., Kumagai, I. & Arakawa, T. (2003). Practical considerations in refolding proteins from inclusion bodies. *Protein Expr Purif* **28**, 1-8.

15. Ohol, Y. M., Goetz, D. H., Chan, K., Shiloh, M. U., Craik, C. S. & Cox, J. S. Mycobacterium tuberculosis MycP1 protease plays a dual role in regulation of ESX-1 secretion and virulence. *Cell Host Microbe* **7**, 210-20.
16. Reinhardt, E.(2005) UN Chronicle 42(2) 17-19.