University of South Bohemia in České Budějovice
Faculty of Science

Expression, purification, and crystallization of the putative transcriptional regulator MSMEG_6227 from *Mycobacteria*

Bachelor thesis

Simon Halas

Supervisor: RNDr Roman Tůma, Ph.D
Co-Supervisor: RNDr Pavel Grinkevich, Ph.D

České Budějovice 2021

Halas, S., 2021: Expression, purification, and crystallization of the putative transcriptional regulator MSMEG_6227 from *Mycobacteria*. BSc. Thesis, in English. – 51 p., Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic.

Annotation

MSMEG_6227, a putative transcriptional regulator from *Mycobacterium smegmatis*, was cloned and expressed in *E. coli* BL21 cells. The expressed protein was purified by chromatographic methods and an optimized protocol for the expression and purification of MSMEG_6227 was established. With the purified protein initial crystallization trials were set up, which need to be further optimized and scaled up in the future to resolve the X-ray structure of MSMEG_6227.

Declaration

I hereby declare that I have worked on my bachelor's thesis independently and used only the sources listed in the bibliography. I hereby declare that, in accordance with Article 47b of Act No. 111/1998 in the valid wording, I agree with the publication of my bachelor thesis, in full to be kept in the Faculty of Science archive, in electronic form in publicly accessible part of the STAG database operated by the University of South Bohemia in České Budějovice accessible through its web pages. Further, I agree to the electronic publication of the comments of my supervisor and thesis opponents and the record of the proceedings and results of the thesis defense in accordance with aforementioned Act No. 111/1998. I also agree to the comparison of the text of my thesis with the Theses.cz thesis database operated by the National Registry of University Theses and a plagiarism detection system.

Taufkirchen an der Pram, 17.08.2021

…….

Simon Halas

Acknowledgments

Firstly, I want to thank RNDr Roman Tůma, Ph.D for letting me write my thesis in his laboratory, for supervising my work, and giving valuable input for my experimental work.

Especially, I want to thank Pavel Grinkevich, Ph.D for guiding me through the experimental work and showing me everything in the laboratory. Moreover, I am very grateful that he explained everything to me, was always open to my questions, and gave me valuable tips for writing and presenting.

Finally, I want to thank everybody in the laboratory for the very welcoming and including environment. It was very enjoyable to work in the lab and to be able to write my thesis here.

Table of Contents

1 Introduction

1.1 Mycobacteria

The family of *Mycobacteria,* which belongs to the phylum of Actinobacteria, comprises over 100 known species with very similar features. Generally, species from the *Mycobacterium* genus are aerobic and belong to the Gram-positive bacteria, although they stain poorly in Gram staining. The general appearance of mycobacterial strains is a rod-shaped form, which is typically 0.3 - 0.5 µm wide and 1.5 to 4 µm long, depending on the state of the culture. The genome size of different mycobacterial strains can vary between approximately 3,000,000 and 7,000,000 base pairs (bp), depending on the strain. [1] [2]

1.1.1 Mycobacterial cell wall

A unique feature to Mycobacteria is their cell wall. Although designated as Gram positive bacteria, Mycobacteria have a very exclusive cell wall structure not found in other Gram-positive prokaryotes. The mycobacterial cell wall is with 8 nm thicker than in other prokaryotic species. [2]

The peptidoglycan layer of the mycobacterial cell wall is modified in contrast to the cell walls of other Gram-positive bacteria. In the interchanging N-acetylglucosamine – N-acetylmuramic acid residues some N-acetyl functional groups of the muramic acid residues are changed to N-glycolyl. It is believed that this feature enhances the durability and strength of the peptidoglycan layer in mycobacterial species. [3] [4]

Linked with the peptidoglycan in the cell wall, arabinogalactan can be found in Mycobacteria. Arabinogalactan is linked covalently to the cell walls peptidoglycan layer and it is a highly branched molecule. Arabinogalactan provides additional stability for the cell wall and a point of attachment for mycolic acid residues. [3] [5]

Mycolic acids are attached to the non-reducing arabinogalactan ends. These 70–90-C chains are essential for the antibiotic resistance and the durability of the prokaryotes. However, this is only achieved due to the tight packing of the mycolic acid residues. [5]

All these features common to Mycobacteria make the mycobacterial cell wall especially resistant and impermeable to toxic compounds found in the environment. Their unique cell

wall provides the mycobacterial species with a natural resistance against antibiotics especially those that target the cell wall, such as penicillin. [2]

Due to the unique cell wall architecture, acid fastness is common to prokaryotic species from the Mycobacterium family. This means that acid-stable complexes with dyes based on phenols, such as fuchsin, crystal violet and auramine O, are formed during the Ziehl-Neelsen staining procedure. These complexes, however, resist destaining with acid alcohols, hence acid fastness. [6] [7] [8]

1.2 *Mycobacterium tuberculosis* and tuberculosis

1.2.1 Epidemiology of tuberculosis

Tuberculosis (TB) is an airborne disease, primarily of the lungs but it can affect other organs as well. The pathogen causing TB comes from the *Mycobacterium* genus, namely *Mycobacterium tuberculosis*. Estimates suggest that approximately 1.4 million people die from a TB infection annually, and that approximately 10 million people develop an active TB infection every year. Since TB especially breaks out when the hosts immune system is weakened by other infections, like an HIV-infection, or in states of poor health, like malnutrition and other chronic diseases, poorer, less developed countries are affected more than wealthier countries. According to data from the World Health Organizations (WHO) annually published TB report African and Asian countries have a higher TB prevalence. In the last years, however, a downwards trend was seen in the TB incidence. [9] [10]

Figure 1 shows the new infections with TB worldwide. A majority of the cases can be accounted to less developed, poorer African and Asian countries, like South Africa, India, and Indonesia. [10]

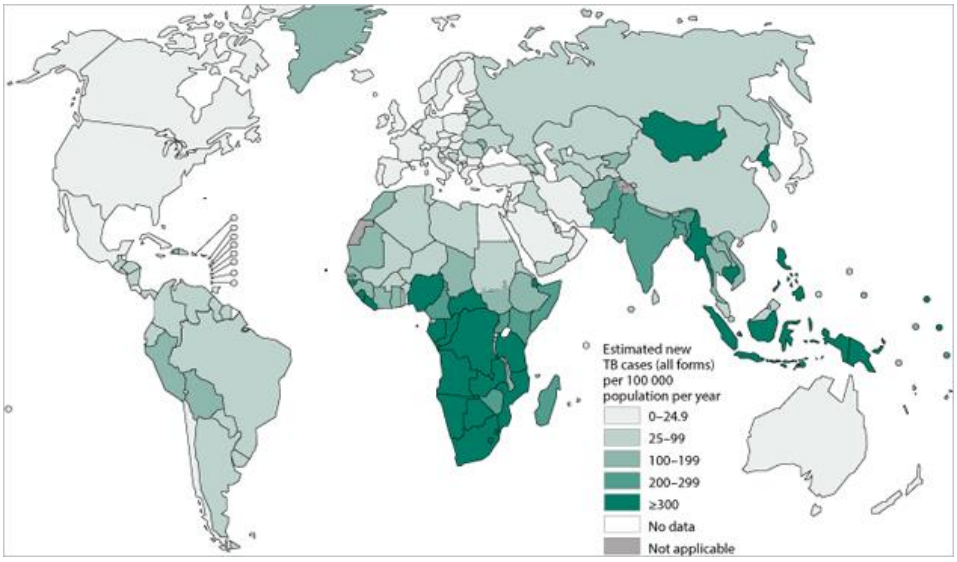


Figure 1: global TB cases (estimated and all TB forms) per 100,000 persons in 2015. Taken from the global TB report 2016 of the World Health Organization (WHO)

Source: https://www.who.int/publications/10-year-review/tb/en/index1.html on 14.11.2020

1.2.2 Pathophysiology

A TB infection starts with the mycobacterial pathogen entering the body of the host through droplets suspended in the air. The infectious air droplets are released by an infected TB patient with the active form of TB by coughing. Upon entering the body of the host, *M. tuberculosis* penetrates the alveolar macrophages in the lower regions of the lung. The bacterium triggers an immune response of the host but can resist and stop the defence mechanisms of the alveolar macrophages. The immune response, however, attracts more alveolar macrophages, neutrophil granulocytes, and monocytes to the affected area, which accumulate into so-called granuloma. When a granuloma has formed, the pathogen is brought under control by the hosts immune system. At this time *M. tuberculosis* can undergo transformation into a dormant state in which it cannot be detected by the host's immune system anymore; a latent form of TB establishes. However, not all cells of *M. tuberculosis* undergo transformation into the dormant state, some cells persist in the active state and act as sensors for reactivation. [11] [12]

Cells in present in the dormant state can undergo transformation back to the active state. This back-transformation happens with the help of the sensor mycobacterial cells and upon special triggers of the host's immune system, which signal *M. tuberculosis* that the host's immune system is weakened by other infections. The active form of TB is re-established. *M.*

tuberculosis begins to replicate again, and the granulomas which formed upon the entering of the pathogen begin to change. Lesions of the lung start to form, and first symptoms can occur. Typical symptoms include coughing, coughing with blood but also chest pain etc. Untreated TB can lead to death. [11] [13]

M. tuberculosis is a difficult target for antibiotics due to the unique cell wall structure, which is further exacerbated when it is present in the dormant state. Usually, long treatment regimens with a combination of antibiotics, like isoniazid and rifampin, are required. Due to this fact, many *M. tuberculosis* species develop drug resistances. [11]

It is, however, not a prerequisite of an active TB infection to establish through a latent phase of the disease but the active form can also establish right after the infection.

1.2.1 *M. Smegmatis* as model organism for *M. tuberculosis*

Because of the pathogenicity of *M. tuberculosis* and the serious disease it can cause, other very similar species of the *Mycobacterium* genus are used as model organisms for *M. tuberculosis.* The most popular model organism is *Mycobacterium smegmatis* (*M. smegmatis*) because of its fast growth rate and non-pathogenicity. The commonly used strain *M. smegmatis mc^2 155* has a genome size of 6,983,267 bp, a proteome of 6,602 proteins and was first isolated in 1884 by Lustgarten. Because of the high transformation efficiency of the *mc^2 155* strain it is a frequently used strain for the study of gene expression, gene function and replication in *Mycobacterium* species. [1] [14]

1.3 Dormancy in *Mycobacteria*

Mycobacteria possess the ability to undergo transformation into a dormant, non-culturable state. In this state the *Mycobacteria* slow down replication processes several folds, are hidden from the host's immune system, and can hardly be targeted with drugs. The cause for the transformation from the active to the dormant state is still not fully understood but there is strong evidence that unfavourable environmental conditions, like hypoxia and low pH-values, can induce this state in *Mycobacteria*. [15] [16]

1.3.1 Dormancy in *M. smegmatis*

M. smegmatis undergoes transition from the active to the dormant, nonculturable state as well as *M. tuberculosis*. An in vitro transformation of *M.* smegmatis can be achieved by acidification of the environment, hypoxic conditions, or nutrient depletion. [15] [16] [17]

When the dormant state in *M. smegmatis* is induced chemically with slow acidification of the environment morphological changes occur, and the bacteria appear as spherical, cocci-like structures instead of rod-shaped cells. Furthermore, the cytoplasm gets less dense, and the cell wall increases in thickness compared to the cells in the active state. The number of ribosomes, and the rate of protein synthesis decreases significantly in the dormant state compared with the active state. [15]

The non-culturable state is accompanied by a low metabolic activity, and slower metabolic pathways, however, a membrane potential is still maintained in the dormant, non-culturable state of *M. smegmatis*. The low metabolic activity, unlike in other bacteria that can be found in this dormant, nonculturable state, is not achieved by the inhibition of respiratory chain enzymes but by the obstruction of enzymes found in the inside of the cell, e.g., enzymes for the citrate cycle. Changes in the overall cell composition in the non-culturable state, that is a significant decrease in lipid content and an increase in protein content of the cell, are known and have been reported. [15] [16]

1.4 Protein expression during dormancy

Not only is the rate of the protein synthesis decreased in the dormant state of *Mycobacteria* but furthermore there are differences in the proteins which are expressed in the dormant state compared to the active state.

1.4.1 Proteome of dormant *M. smegmatis* cells

In dormant *M. smegmatis* cells nearly 50% of the proteome differs from that of the active form. Not only is the proteome different, but also proteins which are expressed in both the active and the dormant states are expressed to significantly different degrees. [16]

It was also shown that the protein composition differed in both, the cytosolic and the membrane fraction of dormant *M. smegmatis* when compared to the active form. In the non-culturable, dormant state fewer proteins responsible for cell wall related processes such as respiration are expressed in cytosolic and membrane-bound fractions. On the other hand, proteins responsible for cell signalling and "information processes" are expressed to a greater extent in the proteome of dormant *M. smegmatis* cells. [16]

Proteins playing a key role in defence mechanisms for *M. smegmatis* were overexpressed in the dormant state to a great extent providing protection against oxidative stress, antibiotics, and oxygen depletion. [16]

Several proteins involved in biosynthetic processes of pyrimidines, purines, histidine, pyridoxine, thymidylate and pantothenate are absent in dormant *M. smegmatis*. Proteins responsible for peptidoglycan synthesis, however, show higher expression levels in dormant cells. Furthermore, the accumulation of energy-rich compounds like glycogen, trehalose, and polyphosphates is more pronounced in the dormant state. [16]

In addition, proteins responsible for degradation processes, such as the hydrolysis of lipids and proteolytic reactions, are found to be more highly expressed in the dormant state. A significant decrease in expression level, however, can be seen in the proteins that account for the transport across membranes. Only ~21% of these proteins are found in the dormant *M. smegmatis* cells when compared to the active form. Some transcriptional factors are only found in the dormant state, while others are only present in the active state. [16]

1.5 Transcriptional regulation

1.5.1 Transcriptional regulation - overview

Transcription in general describes the process of mRNA generation from a DNA template in a prokaryotic cell. The generated RNA template is then further used for translation into an amino acid sequence and ultimately protein production is achieved. [18]

In a prokaryotic organism not every gene or operon of the genome needs to be transcribed all the time. The tryptophan operon which comprises the genes of the tryptophan synthesis pathway in *E. coli* cells, for example, is not transcribed when tryptophan is present in a sufficient concentration. This is achieved by a transcriptional regulator which can bind to free tryptophan, which enables it to repress the transcription of the operon. However, if tryptophan concentrations drop, the transcriptional regulator has no bound tryptophan anymore and cannot interact with DNA anymore; the operon is activated and the genes in this operon are transcribed, producing proteins and enzymes for tryptophan synthesis; the lack of tryptophan can be compensated. Different operons with different functions are known in prokaryotic cells and they all can be regulated in similar ways. [19]

The key components of these systems are transcriptional regulators that take part in either positive or negative feedback loops. Transcriptional regulators are proteins which can specifically and reversibly bind to a DNA motif called operator upstream of the gene coding sequence(s) they control. Upon binding of a transcriptional regulator to this specific DNA motif the binding site for the RNA polymerase (promoter) is changed. When a positive transcriptional regulator is involved the RNA polymerase binding site increases its affinity

to the RNA polymerase. On the contrary if a negative transcriptional regulator is involved the RNA polymerase binding is hindered. By this process expression of genes and operons can be switched on and off selectively and a rapid response to changing environmental conditions is achieved. [18]

1.5.2 Transcriptional regulators - structure and DNA binding

Generally transcriptional regulators contain motifs and structural features that can bind specifically to features on the outside of the DNA double helix. The binding is achieved through hydrogen bonds, hydrophobic interaction as well as ionic bonds. Moreover, many transcriptional regulators can be found as dimers due to the increased affinity of dimers to the DNA sequence. This comes from the fact that a dimer has more available binding sites to the DNA sequence than a monomer. [18] [20]

Over the past decades many transcriptional regulators have been identified but there are some general structural features to explain DNA binding. These structural features give rise to different families of transcriptional regulators. Different transcriptional regulators belong to different classes, depending on their structure:

- Zinc fingers: The zinc finger proteins contain a zinc atom in their structure. The zinc atom coordinates the secondary protein structures, mostly an α helix and a β sheet structure. [18] [21]
- HTH: The basic structural motifs of helix-turn-helix proteins (HTH) contain three right-handed α helices which are connected to each other by a chain of amino acids making up the turn. One of the two helices acts as recognition helix on the DNA; the two helices are held at a protein specific angle. Several subclasses to the HTH proteins are known, like the winged HTH protein (wHTH) which consists of three α helices and three to four β sheets. The β sheets make up the so-called winged part. [18] [19] [22]
- HLH: Helix-loop-helix proteins (HLH) consist of a shorter and a longer α helix connected by an amino acid chain. This amino acid chain makes up the loop. [18] [23]
- Leucine zippers: Two α helices are bound together in a X-shaped form. This is called the leucine zipper because the helices are often held together by the interaction of leucines. [18] [24]

1.6 MSMEG_6227 – a PadR negative transcriptional regulator

1.6.1 PadR transcriptional regulators

PadR stands for Phenolic acid decarboxylases Regulator, and transcriptional regulators from this family act as environmental sensors for microorganisms. At normal environmental conditions with low amounts of phenolic acid derivatives present, the negative transcriptional regulators from the PadR-family bind to a palindromic DNA motif in the operator region. The binding is achieved through two tyrosine residues and a positively charged residue in the N-terminal end of the protein. This tyrosine and positively charged motifs are typical for transcriptional regulators from the PadR-family. Upon binding to the DNA, the transcription of the bound gene, which usually codes for decarboxylases, is repressed. If the amount of phenolic acid derivatives in the environment starts to increase, phenolic acid derivatives can bind to the transcriptional regulator, and inhibit its binding to the DNA. This in turn initiates the transcription of the repressed gene, and decarboxylases which are important for detoxification are produced. The produced decarboxylases convert toxic phenolic acid derivates into vinyl compounds. [25] [26]

Although the PadR-family transcriptional regulators use the same motif for binding DNA they can be divided into two subfamilies, namely subfamily-1 and subfamily-2. PadR subfamily-1 transcriptional regulators dimerize using the C-terminal domain, and the N-terminal domain which is physically separated from the C-terminal domain contains the winged helix-turn-helix (wHTH) motif responsible for DNA-binding. In contrast, Subfamily-2 forms a dimer with the linkage of the C-terminal end and the wHTH motif. [25]

1.6.2 MSMEG_6227

MSMEG_6227 is a putative negative PadR transcriptional regulator from *M. smegmatis mc*^{2} *155* belonging to PadR subfamily 1. However, its structure has not been solved yet, which is thought to be crucial to the understanding how MSMEG_6227 can help *M. smegmatis* in detoxification, and the development of antibiotic resistance in the dormant state. In turn the findings on MSMEG_6227 in *M. smegmatis* could be superimposed to *M. tuberculosis*.

The protein MSMEG_6227 is a transcriptional regulator, which is found in excess in dormant *M. smegmatis* cells and is uniquely expressed to the dormant state of *M. smegmatis*. It is believed that MSMEG_6227 also plays a crucial role in the maintenance of the non-

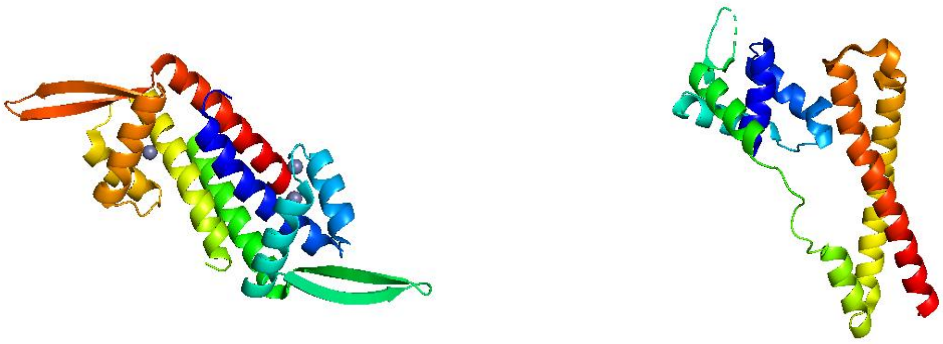
culturable, dormant state, the change of DNA topology, and the natural resistance of dormant cells to antibiotics. [27]

The MSMEG_6227 protein has a molecular weight (MW) of 26.629kDa. The theoretical isoelectric point of the protein was found to be 7.2.

The N-terminal end contains a predicted disordered part, marked with changing Glycine - Proline motifs. This prediction arises from freely rotatable Glycine residues interchanging with the sterically hindered Proline residues. [16] [27] [28]

According to multiple sequence alignments related structures are already known in literature. The transcriptional regulator Rv3488 from *M. tuberculosis* and the Virulence Gene Activator AphA from *Vibrio Cholerae* display similarities of up to 40%. From these similarities it can be assumed that MSMEG_6227 forms a homodimer protein. [29] [30]

Figure 2 shows the solved x-ray structures from the transcriptional factors RV4388 (*M. tuberculosis*) and ApaA (*V. cholerae*) mentioned above.



(a) Rv3488 from M. tuberculosis. (b) AphA from Vibrio cholerae

Figure 2. Solved X-ray structures of RV3488 and AphA.

2 Aims of the thesis

The goal of this bachelor thesis is to develop approaches and procedures to produce, purify and crystallize the protein MSMEG_6227 from *M. smegmatis* the following procedures were performed to achieve this goal:

- Cloning of the MSMEG_6227 gene into an *E. coli* expression system
- Expressing the MSMEG_6227 protein in an *E. coli* expression system
- Purification of the expressed protein
- Crystallization trials with the purified protein
- Structural analysis of the MSMEG_6227 protein

3 Methodology

The DNA of the gene coding for MSMEG_6227 was provided as maxiprep with a concentration of 1559 ng/µl from the Faculty of Science at the University of South Bohemia in České Budějovice. The coding gene was already cloned into a pET19b expression vector with resistance to ampicillin.

3.1 Transformation

To transform a gene, which is inserted a vector, into *E. coli BL21* a transformation procedure using a heat shock is applied.

The *E. coli BL21* cells came from New England Biolabs Inc. The plasmid DNA was diluted to a final concentration of 50 ng/µl. Both cells and vector were kept on ice. For the transformation of the vector into the cells 2.5 µl of the vector and 25 µl of the cells were mixed in a 1.5 ml Eppendorf tube and incubated on ice at -4 °C for 20 minutes. After the incubation time the tube containing the mixture was placed in a water bath at 42°C for 30 seconds. Following the heat shock, the Eppendorf tube was kept on ice for five minutes before 250 µl of SOC growth medium were added to the tube. The SOC growth medium (New England Biolabs, USA) was added to the transformed bacteria. The Eppendorf tube was put onto an Eppendorf Innova S44i shaker at 37 °C and 220 rounds per minute (RPM) for one hour. Afterwards, the transformation product was plated on an agar plate containing with ampicillin (Amp). The plate was the incubated at 37°C overnight (ON) and sealed with parafilm on the next day. All agar plates were stored in the fridge at + 4 °C.

The same procedure of transformation was used to transform the S2-R105 deletion mutant into *E. coli BL21* cells.

3.2 Pilot expression

One colony from the agar plate, containing the transformed bacteria, is put into 5 ml of LB media with50 mg l^{-1} Amp in a 15 ml cultivation tube with a sterile pipette tip. The cultivation tube was then placed in the Eppendorf Innova S44i shaker at 37 °C, 220 RPM, ON. On the next day, the tube was taken out of the shaker and 1 ml of the ON-culture was pipetted into a 50 ml Falcon Tube containing 19 ml of LB media and 50 mg l^{-1} Amp. The tube was then again placed in the shaker and the optical density at a wavelength of 600 nm (OD_{600}) was measured from time to time. When the OD_{600} reached approximately 0.5 the mixture containing the bacterial culture was split into two tubes, each containing 9 ml of culture. One

tube was induced with 1mM isopropyl β-D-thiogalactopyranoside (IPTG) to trigger protein expression in the *E. coli BL21* cells, the other tube was not induced. Before induction 1 ml of the mixture was transferred to a fresh 1.5 ml Eppendorf tube, centrifuged at 11,000 RPM for 60 seconds in a table centrifuge, and the supernatant was discarded. The pellet which appeared after centrifugation was then stored in the freezer at -20 °C and labelled 0-hour time point. 2, 4 and 6 hours after induction 1 ml of culture was transferred to a new 1.5 ml Eppendorf tube, which was again centrifuged in a table centrifuge at 11,000 RPM for 60 seconds; the supernatant was discarded. The pellet from each time point was stored in the freezer at -20 °C. This was done for the induced culture and for the uninduced culture, respectively.

The pellets were stored in the freezer until further experiments were conducted.

3.3 SDS-PAGE

The cell pellet from the pilot expression was dissolved in 70 µl of lysis buffer (50 mM KH_2PO_4, 400 mM NaCl, 100 mM KCl, 10% glycerol, 0.5% Triton X-100, 10 mM imidazole, pH = 7.8). The dissolved pellet was frozen in liquid N_2 and thawed at 37 °C in a heating block 3 times; the treated sample was then centrifuged for 15 minutes at 4 °C with 15,000 RPM. After centrifugation 60 µl of the supernatant were taken and transferred to a new, sterile 1.5 ml Eppendorf tube, the rest of the supernatant was discarded. 20 and 50 µl 4x Laemmle sample buffer (4% SDS, 20% glycerol, 10% β-mercaptoethanol, 0.004% bromophenol blue and 0.125 M Tris HCl, pH approx. 6.8). were added to the supernatant and the pellet, respectively the pellet was redissolved in the sample buffer and all samples were heated at 95°C for 10 minutes in a heating block. The samples were loaded onto a 12.5% isocratic SDS-PAGE gel and run for one hour and 30 minutes at 1000 mA and 100 V. The voltage was changed to 150 V as soon as the samples entered the running gel. Page Ruler Prestained Protein Ladder (Sigma-Aldrich, USA) was used for size determination of the bands.

The SDS-PAGE gels were prepared and polymerized in advanced, wrapped in wetted kitchen towels and stored in the fridge at 4 °C.

The composition of the isocratic SDS-PAGE gel is shown in Table 1.

Table 1. SDS-PAGE gel composition.

Running gel	
30% acrylamide	2.08 ml
18 MΩ H_2O	1.57 ml
1.5 M Tris (pH=8.8)	1.25 ml
10% SDS	50 µl
10% Ammonium Persulfate	50µl
Tetramethylethylenediamine	5 µl
Stacking gel	
30% acrylamide	340 µl
18 MΩ H_2O	1.36 ml
1 M Tris buffer (pH=6.8)	250 µl
10% SDS	20 µl
10% Ammonium persulfate	20 µl
Tetramethylethylenediamine	2 µl

The running gel was polymerized for one hour, the stacking gel for 30 minutes.
When the gel ran to completion, it was stained in Coomassie Brilliant Blue (CBB) for 20 minutes and destained 3-4 times in destaining solution (60% distilled H_2O, 30% methanol and 10% acetic acid).
The destained gel was imaged using the GENESYS G: Box imaging system (Syngene, UK).

3.4 Western Blot

A Polyvinylidene Fluoride (PVDF) membrane was incubated in 100% MeOH for 15 minutes. Then the membrane and an unstained SDS-PAGE gel were incubated in 1x SDS Transfer Buffer for 15 minutes. Afterwards, the membrane and the gel were placed between two filter papers, which were soaked with 1x SDS transfer buffer (filter paper - SDS-PAGE gel - membrane - filter paper). A vertical transfer from the SDS-PAGE gel onto the PVDF membrane was accomplished using the Bio-Rad Trans Blot Turbo midi gel programme with 25 V for 30 minutes. When the transfer was finished the membrane was washed three times

in 1x Tris-buffered saline buffer containing 0.1 Vol% tween (TBS-T buffer) for 10 minutes. Then the membrane was incubated for one hour 15 minutes in a blocking buffer solution, followed by two washes with TBS-T buffer for 10 minutes. 2 µl of penta-His antibody solution was added to 4 ml of blocking buffer, the membrane was placed in a plastic bag, the antibody solution was poured onto the membrane and the plastic bag was incubated for one hour at room temperature. All air bubbles in the plastic bag were removed prior to incubation. Another two washes with TBS-T and one wash with Tris-buffered saline (TBS) were carried out. For the imaging of the bound antibodies 5 ml of Western peroxide solution and 5 ml of Western Luminol Reagent (both from Bio-Rad, USA) were mixed and the membrane was incubated in this mixture for 5 minutes. The bound antibodies were imagined using the GENESYS G: Box imaging system (Syngene, UK).

3.5 Large scale expression

ON cultures of *MSMEG_6227* cloned into the pET19b vector were prepared by transferring one colony of the *E. coli* BL21 cells containing the inserted vector into a 100 ml glass flask with 40 ml of LB media and 50 mg l^{-1} Amp. The suspension was put to the Eppendorf Innova S44i shaker at 37°C, 220 RPM, ON. On the next day, the ON culture was added to 800 ml LB media with 50 mg l^{-1} Amp in a 2l plastic flask. The flask was placed in the shaker at 37°C, 180 RPM until the OD_{600} reached approximately 0.5. When the OD_{600} reached the desired value of 0.5 the culture was put back to the shaker for 4 hours without induction. After 4 hours the flask was taken out of the shaker and the culture was centrifuged at 4 °C and 4000 RPM for 30 minutes. The supernatant was discarded, and the pellet was dissolved in 20 ml of 20 mM Tris-buffer. Afterwards the dissolved pellet was transferred to a 50 ml Falcon Tube, which was then centrifuged at 4 °C and 4000 RPM for 30 minutes. Again, the supernatant was discarded, and the pellet was frozen at -20 °C in the freezer.

3.6 Affinity Chromatography

3.6.1 Sample preparation

A pellet from the big scale expression was prepared for an immobilized metal affinity chromatography. Therefore, 35 ml of lysis buffer (50 mM Tris, 0.5 M NaCl, 5 mM $MgCl_2$, 1 g l^{-1} lysozyme, pH = 8) and were added to the pellet in a 50 ml Falcon Tube, which was kept on ice for 30 minutes and occasionally vortexed to achieve full dissolution of the pellet. 10 mg l^{-1} DNase and 50 mg l^{-1} RNase were added to the sample. Afterwards, the sample was

sonicated using the Hielscher UP100H ultrasonic processor 3 times for 5 minutes with 80% amplitude and 0.5 cycles. Between the sonication cycles the sample was rested on ice for 5 minutes to prevent overheating and after sonication the sample was kept on ice for 30 minutes. Then the sample was centrifuged for 1 hour at 40,000 RPM and 4 °C. a sample for the SDS-PAGE analysis was prepared from the supernatant and the pellet according to section 3.3.

3.6.2 Metal affinity chromatography

The supernatant from the ultracentrifugation was loaded into the equilibrated ÄKTA pure system with a His Trap HP column 5 ml (both from GE Healthcare, USA). The equilibration was done with binding buffer (50 mM Tris, 0.5 M NaCl, 1 mM β-mercaptoethanol) and the flowthrough from the loading procedure was collected. To make sure the sample was fully loaded onto the column the sample pump was washed with binding buffer after the loading was done. Then the column was washed with binding buffer and the flowthrough was collected again. After this washing step a stepwise elution gradient was chosen for the elution. The elution buffer (0.5 M imidazole, 20 mM Tris, 0.5 M imidazole) was set to a concentration of 0.2 M imidazole for 15 minutes before the concentration was set to 0.5 M imidazole. The flow rate during the chromatography was constantly kept at 1 ml min^{-1}.
The collected fractions were analysed by an SDS-PAGE analysis and the samples were prepared according to section 3.3.

3.7 Size exclusion chromatography

After equilibration of the system, 100 µl of the elution fraction collected from the metal affinity chromatography were used for the loading onto the ÄKTA pure system (GE healthcare, USA) with a TSK6000 column (Sigma Aldrich, USA). The size exclusion chromatography (SEC) was run at a constant flow rate of 1 ml/ min to not exceed the maximum pressure. Fractions from the SEC were collected using the fractionator and stored in the fridge at +4°C after the end of the SEC. The progress of the SEC was tracked by a UV-vis measurement at 280 nm. The Dawn 8+ laser system (Wyatt technologies, USA) kept track of the molecule size which came out of the column.

3.8 Transmission electron microscopy

Negative staining with uranium acetate was applied to the sample from the SEC. After the sample was immobilized on a carbon grid, which was made hydrophilic beforehand, the TEM was performed. The sample preparation and microscopy procedure were carried out by the scientific staff at the Electron Microscopy centre at the University of South Bohemia in Budweis.

3.9 MSMEG_6227 S2 - R105 deletion mutant

Because of the results from the experiments with the full-length MSMEG_6227 gene another approach for the experiments was chosen. There was aggregation of the proteins, which is thought to be due to the disordered region on the N-terminal end of the protein. That is why we decided to clone another version of the gene into the pET19b expression vector, where the disordered part was removed by the means of cloning.

3.10 Fragment PCR

The disordered part was deleted, and the deletion mutant with a Serine 2 - Arginine 105 deletion was achieved. The sequences of the primers used in the fragment PCR are shown in Table 2.

Table 2. Fragment PCR primer sequences.

Fragment	Primer name	primer sequence
1	Forward primer T7 promoter	**TAATACGACTCATATAGGG**
1	Reverse primer MSMEG_6227-S2_R105d-R	**GCGCACGTCACC**CATATGCTTGTCGTC GTCGTC
2	Forward primer MSMEG_6227-S2_R105d-F	GACGACAAGCATATG**GGTGACGTGCG CGCC**
2	Reverse primer T7 terminator	**CCGCTGAGCAATAACTAGC**

With the primers above two PCR reactions depicted in Table 3 were mixed in two 0.2 ml PCR tubes. One reaction was set up for fragment 1, the other reaction was set up for fragment 2.

Table 3. Fragment PCR reactions.

Reactant	Volume / µl	Final concentration
5x Q5 reaction buffer	10	1x
10 mM dNTPs	1	200 µM
100 µM Forward primer	0.5	1 µM
100 µM Reverse primer	0.5	1 µM
Q5 HF polymerase (New England Biolabs, USA)	1	0.02 U μl^{-1}
MSMEG_6227 template (diluted 1:30)	0.5	~ 10 ng / reaction
PCR-grade H_2O	37	-
total volume	50	

All other constituents but the primers were the same for the two fragment PCR reactions. The thermocycler was set up with the parameters in Table 4.

Table 4. Fragment PCR - thermocycler setup

Method	Conditions
denaturation	94 °C for 1 minute
primer annealing	58 °C for 20 seconds
elongation	72 °C for 20 seconds
final elongation	72°C
number of cycles	30

The PCR products were taken out from the thermocycler and run on a 1.5% agarose gel to verify their correct length. A 1kB DNA ladder (New England Biolabs, USA) was used for the approximate determination of the fragment length.

For the extraction of the DNA from the agarose gel a NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel, Germany) was used according to the manufacturer's instructions.

3.11 Overlap PCR

To combine both fragments from the fragment PCR reaction an overlap PCR reaction was pipetted.

The sequences of the primers used in the overlap PCR reaction are shown in Table 5.

Table 5. Overlap PCR primer sequences.

Primer name	Primer sequence
T7 promoter	**TAATACGACTCATATAGGG**
T7 terminator	**CCGCTGAGCAATAACTAGC**

Table 6 shows the composition of the overlap PCR mixture.

Table 6. Overlap PCR reaction composition.

Reactant	Volume / µl
5x Q5 reaction buffer	10
10 mM dNTPs	1
100 µM forward primer (T7 promoter)	0.25
100 µM reverse primer (T7 terminator)	0.25
Q5 HF DNA polymerase (New England Biolabs, USA)	1
both fragments from the fragment PCR	1 of each fragment
PCR-grade H_2O	35.5
total volume	50

Table 7 depicts the PCR-program used for the overlap PCR and the primers were added during the PCR reaction to sustain a combination of the two fragments.

Table 7. Conditions for the overlap PCR reaction.

Step	Conditions
Denaturation	94 °C
Annealing	58 °C for 20 seconds
Elongation	72 °C for 20 seconds
Final elongation	72 °C
Number of cycles	35
Primer addition	during cycle 11

The reaction was analysed on a 1.5% agarose gel with a 1 kB DNA ladder (New England Biolabs, USA).

3.12 Restriction

To prepare the vector and the *MSMEG_6227 S2-R105* deletion mutant for ligation a restriction reaction was set up for the vector containing the full-length gene and for the modified gene. Both were cut at an appropriate position with the same restriction enzymes. The restriction reaction depicted in Table 8 was used.

Table 8. Restriction reaction for the S2-R105 deletion mutant.

Reactant	pET19b + MSMEG_6227 original conc. 500 ng/µl	MSMEG_6227 *S2-R105* deletion gene original conc. 86 ng/µl
DNA	2 µl	11.5 µl
CutSmart buffer	2 µl	2 µl
Restriction enzymes	1 µl XbaI 1 µl BamHI HF	1 µl XbaI 1 µl BamHI HF
Water	14 µl	4 µl
total volume	20 µl	20 µl

The restriction reaction was pipetted into 0.2 ml PCR tubes and put to the thermocycler at 37 °C for 1 hour. The expected size of the fragments was 860 bp + 5,600 bp for the pET19b + MSMEG_6227 digestion reaction, and 550 bp for the MSMEG_6227 S2-R105 deletion gene digestion reaction. The digestion was checked on an 1.5% agarose gel with a 1 kB DNA ladder (New England Biolabs, USA).

3.13 Ligation

To combine the vector and the MSMEG_6227 S2 - R105 deletion gene a ligation reaction was carried out. According to the concentration from the restriction reaction the amount of cut vector and insert were calculated to achieve a 15:1 insert:vector molar ratio for the ligation reaction. Due to the low concentrations this ratio could not reached but the reaction was set up in a 0.2 ml PCR tube according to Table 9.

Table 9. Ligation reaction setup.

Reactant	Volume / µl
MSMEG_6227 S2 - R105 deletion gene (cut)	1.4 (= 35 ng/µl)
vector (cut)	0.8 (= 42 ng/µl)
T4 ligase	0.5
T4 buffer	1
PCR-grade H_2O	6.3
total volume	10

The reaction mixture was put into the thermocycler ON at 37 °C and checked on the next day on an 1.5% agarose gel with a 1 kB marker (New England Biolabs, USA).

3.14 Transformation of the S2 - R105 deletion mutant

The transformation of the MSMEG_6227 S2 - R105 deletion into *E. coli* BL21 was using the same heat shock procedure described in section 3.1.

3.15 Pilot expression of the S2 - R105 deletion mutant

Pilot expression experiments with the S2 - R105 deletion mutant of MSMEG_6227 were performed as stated in section 3.2.

3.16 Large scale expression of the S2 - R105 deletion mutant

The large-scale expression experiments with the S2 - R105 deletion mutant were also performed as described in section 3.3 but the culture was induced when the OD_{600} reached approximately 0.5. This induction was performed with 1 mM IPTG.

The preparation of the pellet after the incubation in the shaker did not differ from the procedure reported in section 3.3.

3.17 Affinity chromatography with the S2 - R105 deletion mutant

3.17.1 Metal affinity chromatography

The sample for the metal affinity chromatography was prepared according to section 3.6.1. The His Trap HP column 1 ml was fixed in a stand and equilibrated with a buffer containing no imidazole (0.5 M NaCl, 20 mM Tris, 1 mM β-mercaptoethanol, pH = 8). The supernatant from the centrifugation was taken up by a syringe and loaded directly onto the column through a syringe filter. The column was then washed with at least 10 column volumes (CV) 20 mM imidazole buffer (0.5 M NaCl, 20 mM Tris, 1 mM β-mercaptoethanol, 20 mM imidazole, pH = 8) before it was washed with at least 10 CV of 50 mM imidazole buffer (0.5 M NaCl, 20 mM Tris, 1 mM β-mercaptoethanol, 50 mM imidazole, pH = 8). The fractions of the washes were collected in 15 ml Falcon Tubes. Then the sample was eluted with 500 mM imidazole buffer (0.5 M NaCl, 20 mM Tris, 1 mM β-mercaptoethanol, 500 mM imidazole, pH = 8), and 5 fractions, each 2 ml, were collected in 15 ml Falcon Tubes.
Samples of each fraction collected were prepared for an SDS-PAGE analysis, according to the procedure in section 3.3.

3.18 Size exclusion with the S2 - R105 deletion mutant

A SEC was performed with the elution fractions from the manual metal affinity chromatography on the ÄKTA pure system (GE healthcare, USA). The injection volume was 100 µl, and the column used was a Superdex 200 Increase 10/300 GL gel filtration column with an agarose/dextrose matrix. The flow rate was kept constant at 1 ml/ minute, and the outcoming sample was collected in a 96 well collection plate. The outcoming molecules were analysed using the MALS system and the refractive index detector connected to the ÄKTA pure system.

3.19 Crystallization

Crystallization screenings were conducted for the S2 - R105 deletion mutant with multiple screens. The sitting drop vapor diffusion method was used for the crystallization process.

The screens used for an initial screening were Morpheus I, Morpheus II and Structure Screen. After 6 days and 4 months the plates were checked for crystals.

4 Results

4.1 Expression of the full-length MSMEG_6227 protein in *E. coli* BL 21

The full-length protein inserted in the pET19b vector was cloned into *E. coli* BL 21 cells by heat shock transformation. Colonies were observed after one night of incubation, the agar plates were sealed with parafilm and stored in the fridge at 4 °C. The physical properties of the full length MSMEG_6227 protein are summarized in Table 10.

Table 10. Physical properties of the full length MSMEG_6227 protein including the expression Tag.

molecular weight / kDa	29.4
Amino acids	269
GC content	70.37 %
isoelectric point	6.52
extinction coefficient / M^{-1} cm^{-1}	16960

Pilot expression experiments were performed with the transformed *E. coli* BL 21 cells. Overnight cultures were grown, and the pilot expression was done with 1 mM IPTG induction at an optical density of 0.5 for 4 hours. Another culture was grown in which no induction happened upon reaching an OD of 0.5. However, after reaching an OD of 0.5 the culture was left in the shaker for 4 hours. The experiment was performed at 37 °C for 4 hours, and 5 time points (0 hours, 1 hour, 2 hours, 3 hours, and 4 hours) were taken after induction or after reaching the desired OD of 0.5.
After the expression was finished a 12.5% isocratic SDS-PAGE was run. The results from this trial were that the protein is expressed in *E. coli* BL 21 cells in the soluble part of the cell pellet and the amount of expression increases up to the 4-hour time point. Moreover, the band of the protein was also found in the uninduced cell culture in a comparable amount.

From the results of this first pilot expression, a second pilot expression experiment was made, and the 4-hour time point was analysed on a 12.5% isocratic SDS-PAGE gel. The sample from the time point were split up into soluble – insoluble and induced – uninduced. Furthermore, from this SDS-PAGE two Western Blots were made to confirm that the band corresponds to the desired protein. For this purpose, antibodies detecting the His-tag of the

protein were used in western blotting. The results from these experiments are shown in Figure 3 and Figure 4.

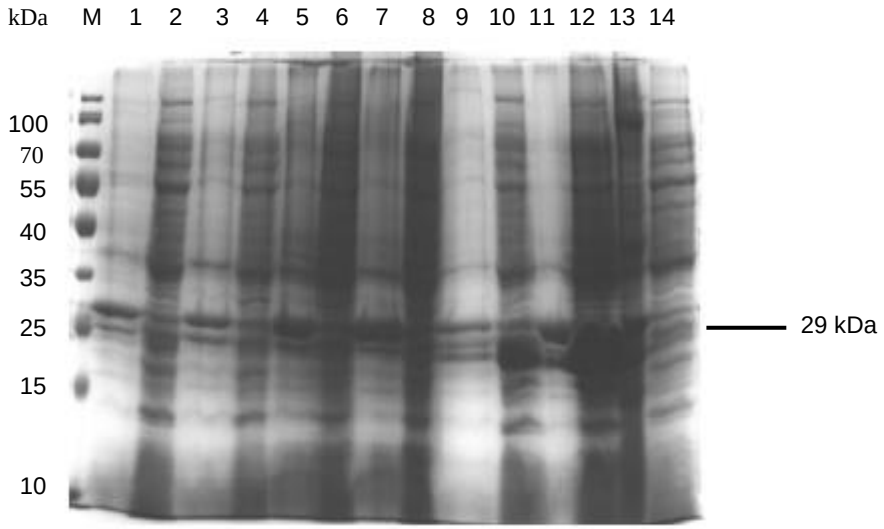


Figure 3.SDS-PAGE analysis of the pilot expression experiments with MSMEG_6227 full length – soluble and insoluble fraction at 4h, induced and uninduced cell cultures.

M-PageRuler prestained protein ladder, 1 – 8- other experiments, 9-4h induced insoluble, 10-4h induced soluble, 11-4h uninduced insoluble, 12-4h uninduced soluble, 13 – 14-other experiment

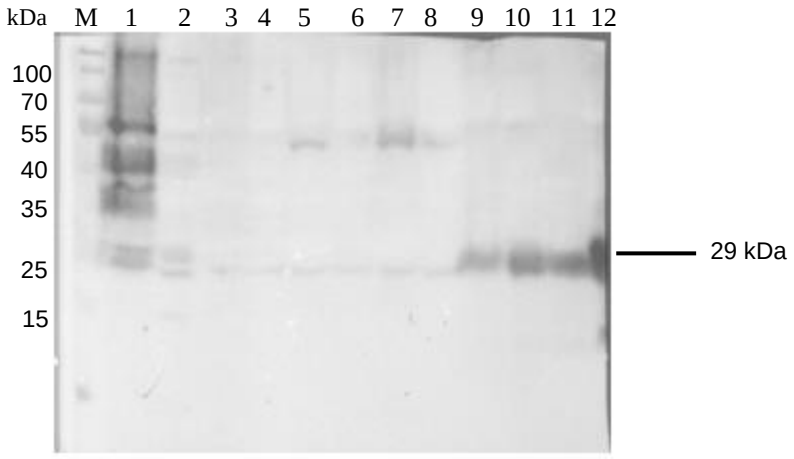


Figure 4. Western Blot of the pilot expression experiments with MSMEG_6227 full length – soluble and insoluble fraction at 4h, induced and uninduced cell cultures.

M- Page Ruler prestained protein ladder, 1 – 8- other experiments, 9-4h induced insoluble, 10-4h induced soluble, 11-4h uninduced insoluble, 12-4h uninduced soluble

The SDS-PAGE gel depicts a band at approximately 29 kDa in the soluble fractions in both, the induced and uninduced, cultures. This band was also detected in the western blot analysis of the same SDS-PAGE gel. In lane number 12 only half of the band is visible. Nevertheless,

the signal can be seen, and it coincides with the signal on the SDS-PAGE. The bands in the Western Blot are at approximately 29 kDa as well.

After the results from the pilot expression, expression experiments with a volume of 800 ml LB-medium were conducted at 37 °C for 4 hours. The 4-hour time point was chosen because in the pilot expression experiments it was the time point with the most pronounced protein band. Two separate expressions were performed, one flask was induced with IPTG and one was not. From the results of the pilot expression experiments a leaking expression without any induction was expected since the protein was observed in the uninduced cultures as well as in the induced ones. The uninduced fraction was left in the shaker for 4 hours after the OD reached 0.5 (this gives a total expression time of 5 hours).
The isocratic 12.5% SDS-PAGE gels showed bands at approximately 29 kDa in the induced and uninduced soluble fractions. The band intensities of the uninduced fractions were comparable with the intensities of the induced fractions. This finding further confirmed the assumption of leaking expression. In the next purification steps whole cell pellets from induced and uninduced cultures were used.

4.2 Purification of the full length MSMEG_6227 protein

4.2.1 Metal affinity chromatography

Metal affinity chromatography was performed to purify the expression pellet. At first both, induced and uninduced, expression pellets were used in a first preliminary purification. The pellets were prepared for chromatography according to section 3.6.1, and the metal affinity chromatography was performed according to section 3.6.2. The purification was done with a stepwise imidazole elution gradient; no chromatogram of this purification is available, as it was not saved.

12.5% isocratic SDS-PAGE gels of the collected metal affinity chromatography fractions showed that the protein can be purified in the induced and in the uninduced pellet. Moreover, a Western Blot confirmed the correctness of the SDS-PAGE gels. Again, the induced as well as the uninduced fraction yielded the desired protein product in the elution fractions. (Figure 5 and Figure 6)

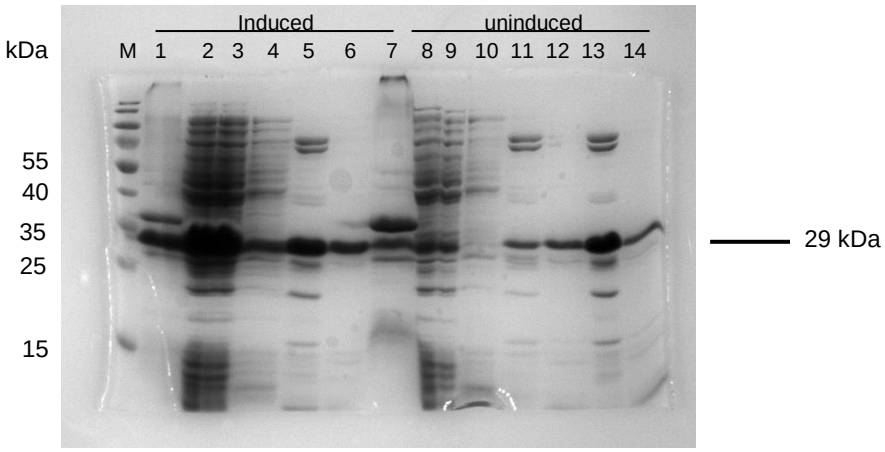


Figure 5. Metal affinity chromatography - MSMEG_6227 full length protein – induced and uninduced pellet. SDS-PAGE analysis

M-PageRuler prestained protein ladder, 1-pellet before chromatography, 2-supernatant before chromatography, 3-flowthrough, 4-wash, 5-elution 1, 6-elution 2, 7-elution 3, 8-pellet before chromatography, 9- supernatant before chromatography, 10-flowthrough, 11-wash, 12-elution 1, 13-elution 2, 14 elution 3

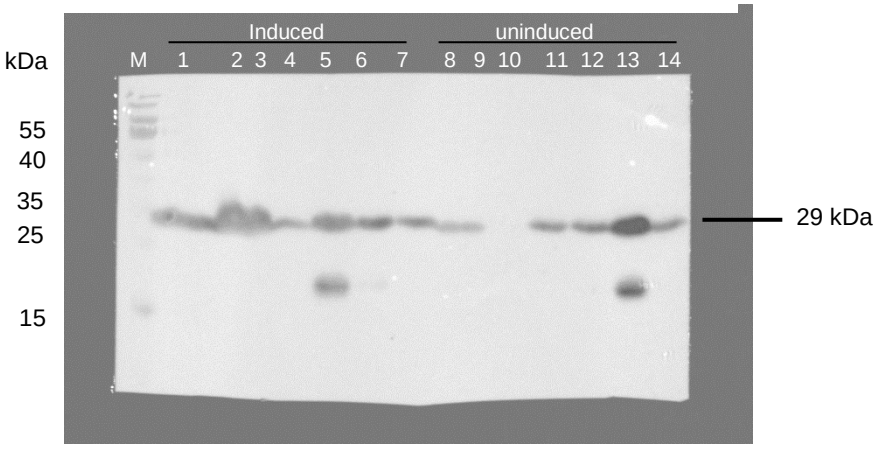


Figure 6. Metal affinity chromatography - MSMEG_6227 full length protein – induced and uninduced pellet. Western Blot.

M-PageRuler prestained protein ladder, 1-pellet before chromatography, 2-supernatant before chromatography, 3-flowthrough, 4-wash, 5-elution 1, 6-elution 2, 7-elution 3, 8- pellet before chromatography, 9- supernatant before chromatography, 10-flowthrough, 11-wash, 12-elution 1, 13-elution 2, 14 elution 3

To further confirm the results from the purification, the bands in lanes 6 and 12 were cut from the SDS-PAGE gel. With the cut bands a MALDI-TOF mass spectrometry analysis was performed at the Faculty of Science by the responsible person. The mass spectrometric analysis confirmed the amino acid sequence of the expressed protein.

After this confirmation by mass spectrometry another metal affinity chromatography with only the uninduced fraction was carried out. The uninduced culture was chosen because it yielded, according to the earlier results, protein amounts comparable to the induced culture.

A stepwise imidazole elution with two steps was chosen. The sample was loaded onto the column and prior to elution the column was washed with 20mM and with 50mM imidazole. Elution was then performed with 200mM imidazole and finally, with 500mM imidazole. Figure 5 shows the chromatogram of the purification. The detection was done by UV-vis measurement at 280 nm. The chromatogram can be seen in Figure 7.

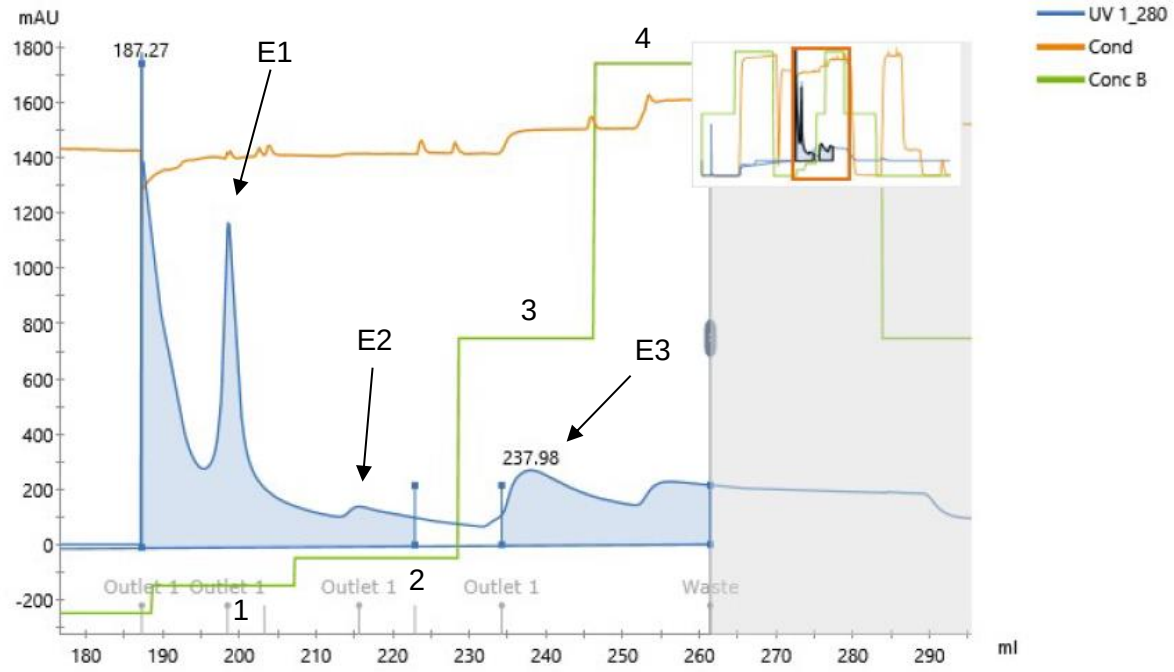


Figure 7. Chromatogram of the metal affinity chromatography with MSMEG_6227 full length – uninduced pellet

A stepwise elution gradient was applied. 1-washing with 20mM imidazole, 2-washing with 50mM imidazole, 3-200mM imidazole elution, 4-500mM imidazole elution.

E1 - elution peak 1, E2 - elution peak 2, E3 - elution peak 3

From the chromatogram there are two elution peaks during the washing steps, one peak during the 20mM imidazole wash and a smaller peak during the 50mM imidazole wash. However, during the 200mM elution a third elution peak appeared.

All fractions were collected and run on a 12.5% isocratic SDS-PAGE gels to check the results. Figure 8 shows the SDS-PAGE analysis of the metal affinity chromatography.

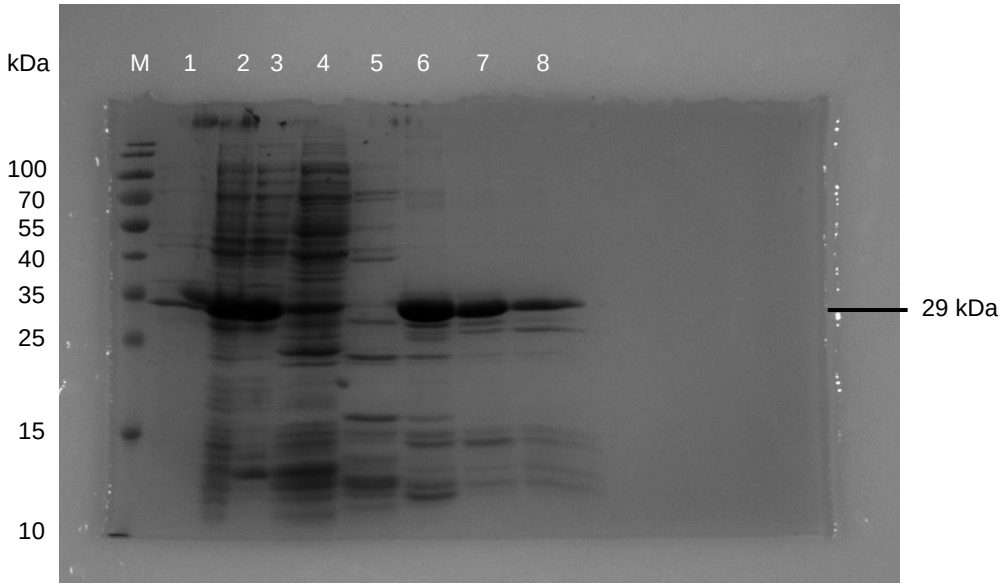


Figure 8. SDS-PAGE analysis of the fractions from the metal affinity chromatography with the uninduced pellet

M-PageRuler prestained protein ladder, 1-pellet before chromatography, 2-supernatant before chromatography, 3-flowthrough, 4-wash 20mM imidazole, 5-wash 50mM imidazole, 6-elution 1, 7-elution 2, 8-elution 3

4.2.2 Size exclusion chromatography

To further determine the size and the state of the non-reduced sample from metal affinity chromatography a subsequent size exclusion chromatography was conducted.

The size exclusion was performed with the elution I fraction from the purification in Figure 7, an injection volume of 100 µl and a TSK gel 600 column were used.

Prior to size exclusion chromatography two equivalents of sample were taken, one was ultracentrifuged, whereas the other was centrifuged with a normal centrifuge. This step was done to remove any big aggregates in the sample and to avoid clogging of the instruments tubing. After equilibration of the instrument the two samples was injected using a syringe. The chromatogram from the SEC is shown in Figure 9. The detection was done by UV-vis measurement at 280 nm.

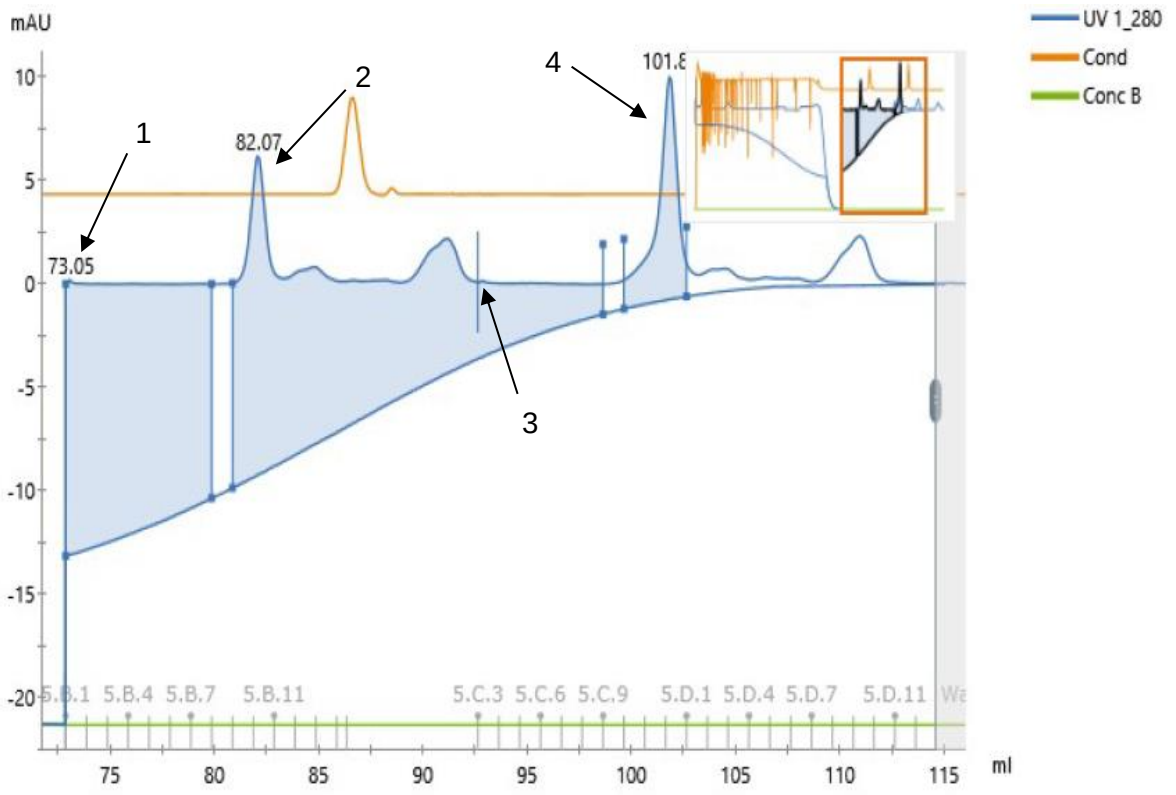


Figure 9. Size exclusion chromatography with two uninduced, purified MSMEG_6227 full length protein samples – first injection with sample after ultracentrifugation, second injection with sample after normal centrifugation

1-injection after ultracentrifugation, 2-elution peak (fractions B9 – C1), 3-injection after normal centrifugation, 4-elution peak (fractions C11 – D3)

Two well defined peaks were observed in the chromatogram from the SEC, one for each sample injection. Additionally, at the end of the column a dynamic light scattering measurement was carried out. The retention time of these peaks on TSK600 gel column, in combination with the dynamic light scattering results suggested an aggregate of approximately 1000kDa with an estimated radius of 20 – 25 nm. During the chromatography, the samples were collected using a fractionator and a 96-well plate. The collection and labelling of the fractions are indicated on the abscissa of the chromatogram.

A 12.5% isocratic SDS-PAGE gel was run with fractions B9 - C1 (ultracentrifuged sample) and C11 - D3 (normally centrifuged sample) from this SEC. The fractions were chosen according to the peaks in the chromatogram. The other fractions were not run. Figure 10 shows the SDS-PAGE analysis of the SEC.

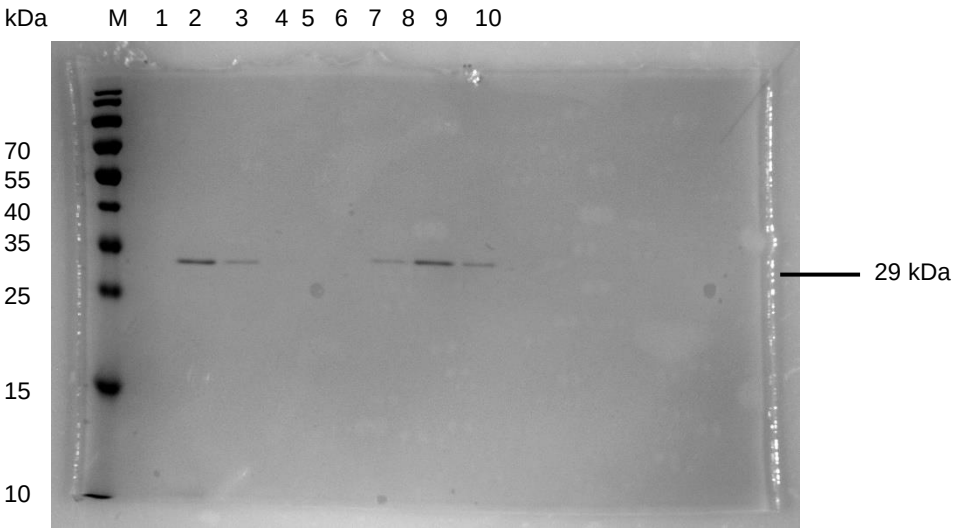


Figure 10. SDS-PAGE analysis of the size exclusion chromatography fractions- one ultracentrifuged sample, one normally centrifuged sample

M-PageRuler prestained protein ladder, 1-B9, 2-B10, 3-B11, 4-B12, 5-C1, 6-C11, 7-C12, 8-D1, 9-D2, 10-D3

To confirm the results from the SEC the two fractions with the most pronounced band in the SDS-PAGE, namely B10 and C12 were further concentrated. 500 µl of each fraction were put into a 10 kDa Amicon filter tube. The fractions were spun down with 5000 RPM at 4 °C to a final volume of approximately 300 µl. The final concentrations of the two fractions were measured with a spectrophotometer at 280 nm and extinction coefficient ε of 16960.

Table 11. Concentration of the SEC fractions.

fraction (each ~ 300 µl)	concentration in mg/ml
B10	0.035
C12	0.069

4.2.3 Transmission Electron Microscopy

With these fractions a negatively stained transmission electron microscopy (TEM) pictures were taken. The sample was fixed on a hydrophobic carbon grid and stained with Uranium acetate. The staining and capturing were done at the Faculty of Science, University of South Bohemia by the responsible person. The pictures of the ultracentrifuged and the normally centrifuged sample can be seen in Figure 11 and Figure 12, respectively.

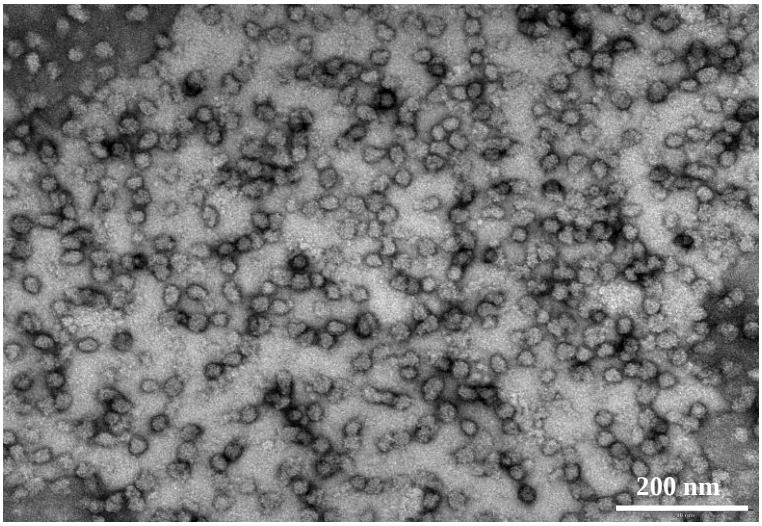


Figure 11: TEM picture of the ultracentrifuged size exclusion chromatography sample after concentration to ~ 300µl (fraction B10)

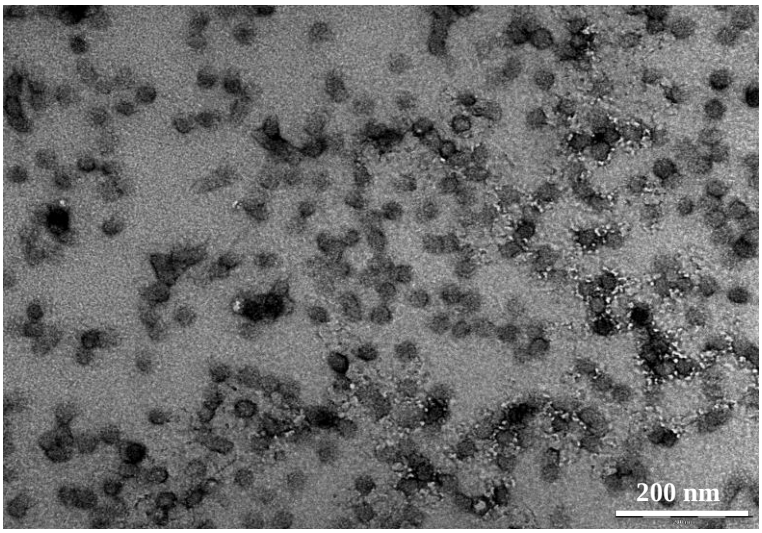


Figure 12: TEM picture of the normally centrifuged size exclusion chromatography sample after concentration to ~ 300µl (fraction C12)

Both pictures show nearly spherical aggregates of sample with an estimated radius of 25 nm. The estimated radius of the aggregates corresponds to the dynamic light scattering data from the SEC experiments.

Figure 11 shows the sample, which was ultracentrifuged, Figure 12 shows the sample after normal centrifugation. The aggregates did not look like they had any specific structure, although many of them did have nearly spherical shapes in the TEM pictures.

The SEC and TEM data confirmed that the proteins expressed in *E. coli* BL 21 cells did aggregate into nearly spherical aggregates. It was assumed that this behavior comes from an amino acid motif in the N-terminal region of the protein. This motif contained many residues

of glycine and proline. Due to the freely rotatable glycine and the sterically hindered proline this part of the sequence should not display any defined structure and was designated a disordered part of the protein.

4.3 Cloning and expression of the S2 - R105 MSMEG_6227 deletion mutant

The approach to prevent this accumulation was to delete the part with the high Glycine and Proline amount by the means of cloning. The part which was decided to be removed went from Serine on position 2 to Arginine on position 105. This S2 - R105 deletion left the predicted PadR domain intact and removed most of the N-terminal domain. The C-terminal part was not modified.

The red part in Figure 13 was deleted from the original protein and the physical properties of the modified version is summarized in Table 12.

Table 12. Physical properties of the modified version of MSMEG_6277.

molecular weight / kDa	18.3
amino acids	165
GC content	65.26 %
isoelectric point	5.17
extinction coefficient / M^{-1} cm^{-1}	16960

The green part shown in Figure 13 is the His-tag with which the protein was tagged for the metal affinity chromatographic retention on the Ni^{2+} column.

The shortened protein version was expressed by *E. coli* BL 21 cells, further purified by metal affinity chromatography and SEC, and eventually crystallization trials were set up.

The removed part was not thought to be involved in DNA binding of the transcriptional regulator protein. The removal of the S2 - R105 part was done by the means of PCR reactions with the sample obtained from the Faculty of Science, University of South Bohemia.

MGHHHHHHHHHHSSGHIDDDDKHMSTPFAFPTGDFGFGPAGRRAMLHHRR
AARREFRDQMRAHLHEAREQALDPRCEMGGPGGPGPMGGPGRGFGGFGFG
FDPGAGFGFGPGGPRGRRGHGRGRRGRRGDVRAAILKLLAERPMHGYEMIQEIGER
TDNLWRPSPGSVYPTLQLLVDEGLISGTEAEGSKKLFELTEAGRAAAEAIETPPWEQIA
EDVDPAAVNLRGAIAQLMGAVAQSAYTATEDQQQRILDVVNNARREIYQILGEE

Figure 13: Modified sequence of MSMEG_6227

The green part specifies the His-tag of the protein and the red part was deleted by PCR reactions; the black part is the shortened version of the protein.

4.3.1 Fragment PCR

Two fragments of the expression vector were produced by two PCR reactions which contained the part before upstream (His-tag) and downstream (PadR-domain and C-terminus) of the part which was to be deleted. The first fragment, containing the His-tag and the start codon, had a length of 159 base pairs and the second fragment, containing the PadR-binding domain and the C-terminus was 506 base pairs long.

The reaction was performed as stated in Section 3.10.

The reaction products from the two reactions were analyzed by gel electrophoresis on an 1% agarose gel with a 1 kb marker (New England Biolabs, USA) and a 100 bp marker (New England Biolabs, USA). The agarose gel is shown in Figure 14.

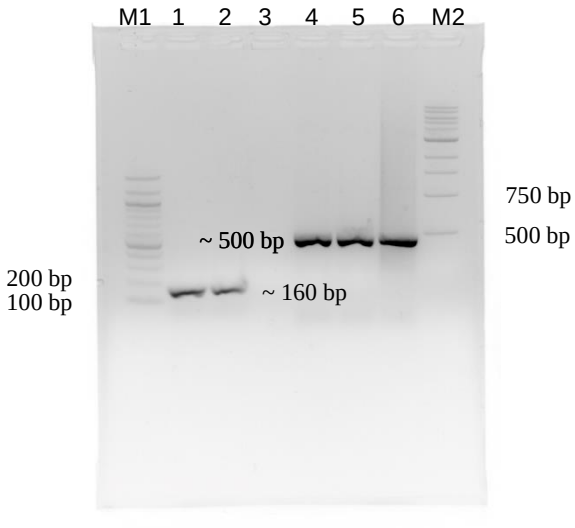


Figure 14: Fragment PCR reaction for the modified MSMEG_6227 version

M1-100bp DNA ladder, 1 – 2-fragment 1, 4 – 6-fragment 2, M2-1kb marker

The observed bands did correspond to the expected size of the fragments. A control for the correctness of the reactions by sequencing was only done at the end of cloning.
The bands in lanes 2, 3, 5, and 6 were cut from the gel. The bands from 2 and 3 were combined in one tube as fragment 1, the bands from 5 and 6 were combined in one as fragment 2. The cut bands were cleaned up from the gel by the commercially available NucleoSpin Gel and PCR clean up kit (Marcherey-Nagel, Germany) and the two fragments were used in the following overlap PCR reaction.

4.3.2 Overlap PCR

The overlap PCR combined the two fragments from the fragment PCR into one sequence with 665 base pairs in length. The reaction was performed as stated in section 3.11.
The result was again checked by gel electrophoresis on an 1% agarose gel with a 100 bp marker (New England Biolabs, USA). The result of the overlap PCR is shown in Figure 15.

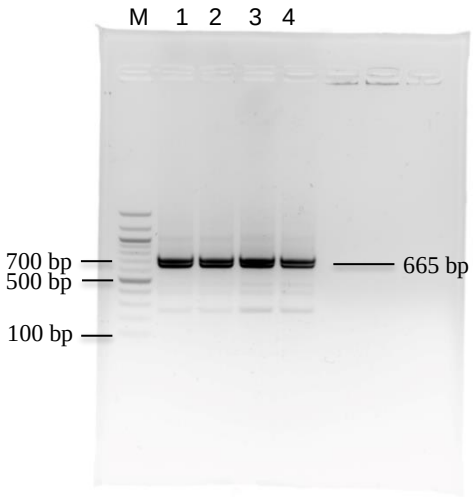


Figure 15: 4 Overlap PCR reactions with fragments 1 and 2

M-100bp marker, 1 – 4 overlap PCR reactions

According to the marker the band was found to be between 600 bp and 700 bp, which confirmed the expected result. The band was again cut from the gel and cleaned up with the NucleoSpin Gel and PCR clean up kit (Marcherey-Nagel, Germany).

4.3.3 Restriction and Ligation

To combine the product from the overlap PCR reaction with the pET19b vector, both were restricted with XBal and BamH-HF restriction enzymes. The restriction reaction was checked by electrophoresis on an 1% agarose gel. The digested fragments are shown in Figure 16 and Figure 17.

The products from the restriction reactions were ligated using T4 ligase with the procedure from section 3.13. The ligation reaction was left overnight at 37 °C, the product was transformed into *E. coli* Neb5α cells by heat shock and plated on an LB - ampicillin plate on the following day according to the procedure in 3.1 and 3.13.

The sequence of the modified protein version was confirmed by sequencing, but in the C-terminal part a Serine codon was changed to a Proline codon. However, the sample was used for further analysis as the codon exchanged was not in the DNA binding region and was not thought to be functionally relevant.

After a Mini prep (Marcherey-Nagel, Germany) from the *E. coli* Neb5α cells the altered version of the protein was transformed into *E. coli* BL 21 cells and *E. coli* BL 21 codon plus cells by heat shock and plated on an LB/ Ampicillin plate.

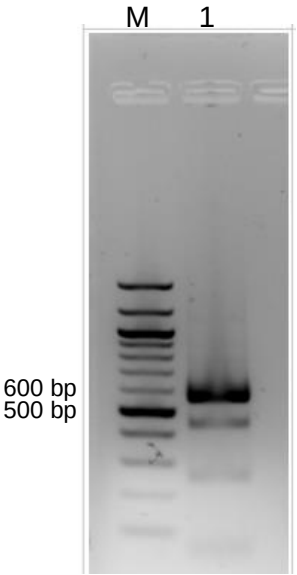


Figure 16: Fragment digestion

M-100bp marker, 1-digested fragment from overlap PCR

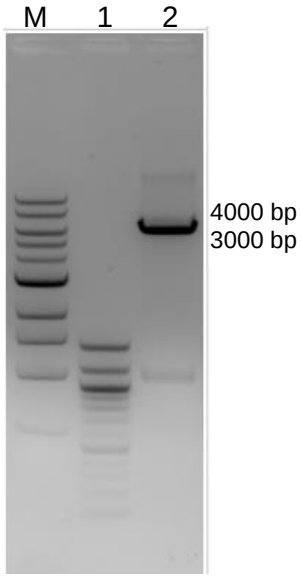


Figure17: Vector restriction

M-1kb marker, 1-other digestion, 2-vector digestion

4.3.4 Pilot expression and big scale expression

With one of the grown colonies from each cell line pilot expression experiments were conducted. The different cell lines were used for the initial trials only.

From the previous expression experiments with the full-length protein only 37 °C was chosen as expression temperature; induction with 1 mM IPTG was not omitted.

At the time points 0 hours, 1 hour, 2 hours, 3 hours, and 4 hours after induction a sample was taken and analyzed on a 12.5% isocratic SDS-PAGE gel. The result of the pilot expression experiments is shown in Figure 18.

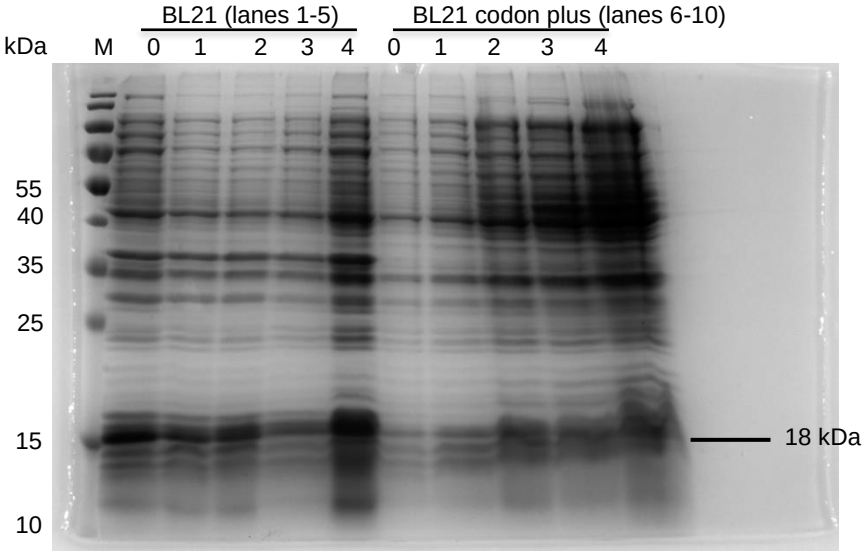


Figure 18. Pilot expression of the modified MSMEG_6227 protein in E. coli BL21 and E. coli T7 cells

M-PageRuler prestained protein ladder, 1 – 5 BL21 whole cell expression products (the number indicates the time point), 6 – 10-BL21 codon plus whole cell expression products (the number indicates the time point)

The desired band at 18.3 kDa can be observed in both cell lines, however, in the *E. coli* BL 21 cells the band is more pronounced.

Repetition of the SDS-PAGE with both cell lines at the 0-hour and 4-hour time point at 37 °C only was conducted with the modified version of MSMEG_6227. Furthermore, the pilot expression SDS-PAGE gel was divided into soluble and insoluble fractions and a Western Blot with penta-His-antibody was performed from this SDS-PAGE gel to confirm the results. The Western Blot was done according to section 3.4. The results from the SDS-PAGEs and the Western Blot are depicted in Figure 19 and Figure 20.

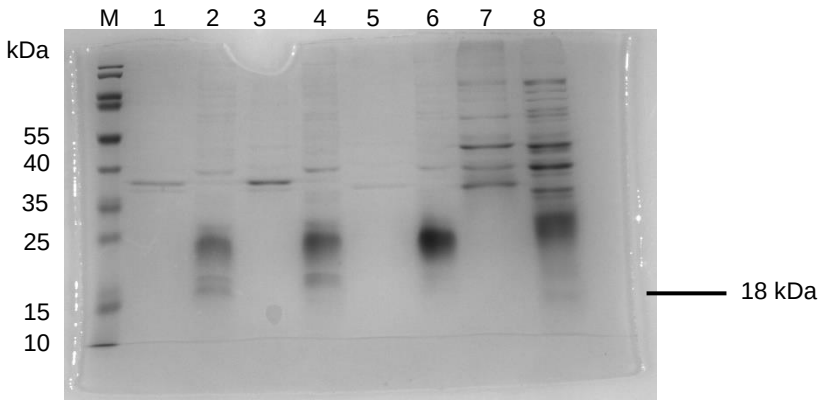


Figure19: SDS-PAGE analysis of the BL21 and BL21 codon plus pilot expression – 0 hour and 4hour time point

M-PageRuler protein ladder, 1-BL21 0 hours insoluble, 2-BL21 0 hours soluble, 3-BL21 4 hours insoluble, 4-BL21 4 hours soluble, 5-BL21 codon plus 0 hours insoluble, 6-BL21 codon plus 0 hours soluble, 7-BL21 codon plus 4 hours insoluble, 8-BL21 codon plus 4 hours insoluble

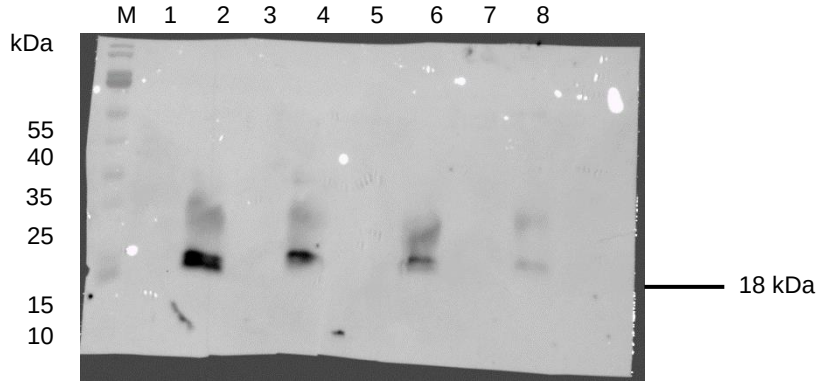


Figure 20: Western Blot analysis of the BL21 and T7 pilot expression – 0 hour and 4hour time point

M-PageRuler protein ladder, 1-BL21 0 hours insoluble, 2-BL21 0 hours soluble, 3-BL21 4 hours insoluble, 4-BL21 4 hours soluble, 5-BL21 codon plus 0 hours insoluble, 6-BL21 codon plus 0 hours soluble, 8-BL21 codon plus 4 hours insoluble, 9-BL21 codon plus 4 hours insoluble

The SDS-PAGE analysis showed expression in the soluble part of *E. coli* BL 21 and in the soluble fraction of the *E. coli* BL 21 codon plus cells. The Western Blot and the SDS-PAGE analysis confirmed the expression in the soluble part of the cells and expression was better in *E. coli* BL 21 cells.

A big scale expression from *E. coli* BL 21 cells was performed as described in section 3.5: 37°C in a shaker at 180 RPM for 4 hours after 1mM IPTG induction at OD_{600} = 0.5. The cells were pelleted and stored in the freezer at - 80°C.

4.4 Purification of the S2 - R105 MSMEG_6227 deletion mutant

4.4.1 Metal affinity chromatography

The pellet from the big scale expression was prepared for a metal affinity chromatography according to section 3.17.1. After a 20 mM washing step a stepwise elution with 500 mM imidazole as the first step and 1M imidazole as the second step was applied. The chromatography was recorded with an UV-vis detection at 280 nm. Figure 21 depicts the chromatogram from the metal affinity chromatography.

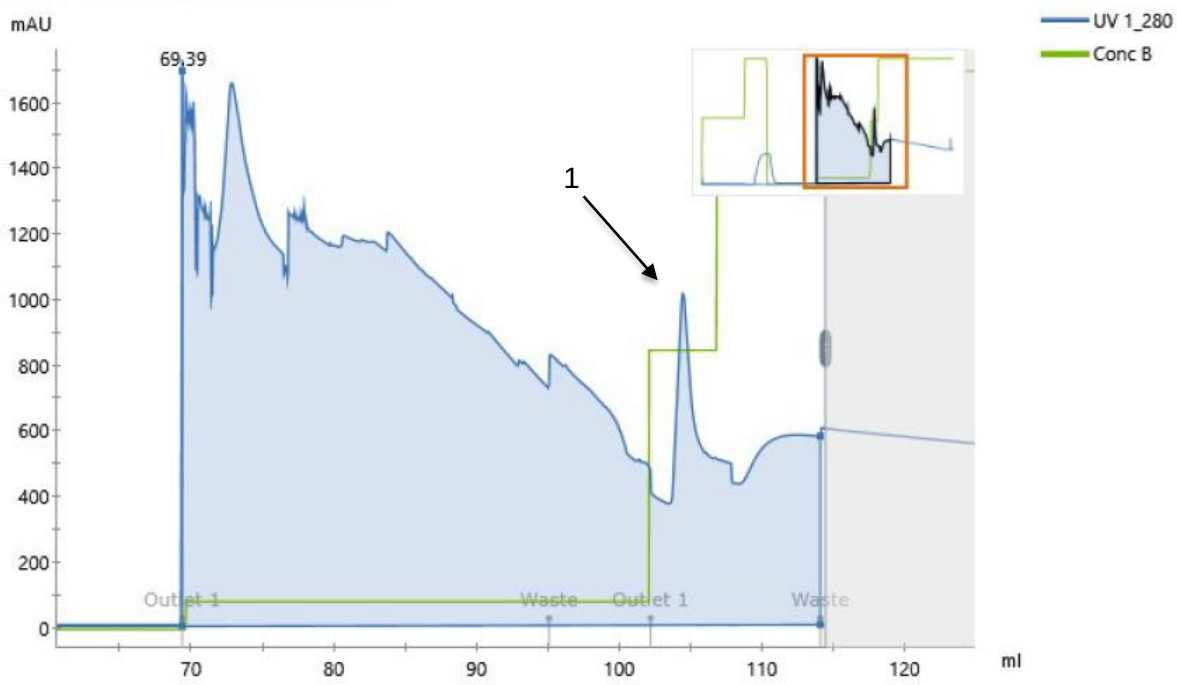


Figure 21: Chromatogram from the metal affinity chromatography with the S2-R105 deletion version.

1 – elution peak at 500 mM imidazole

The chromatogram showed an elution peak (1) at 500 mM imidazole concentration. At 1 M imidazole concentration no further elution peak was observed. The chromatogram before the application of the stepwise elution gradient does not look normal but the abnormal UV-vis data arises from air bubbles which were present in the system from the sample loading

procedure. Fractions from loading, washing and elution were collected and an SDS-PAGE analysis was carried out. The 12.5% isocratic SDS-PAGE gel, which is depicted in Figure 22, showed that the desired band is present in the 500 mM elution fraction from the purification.

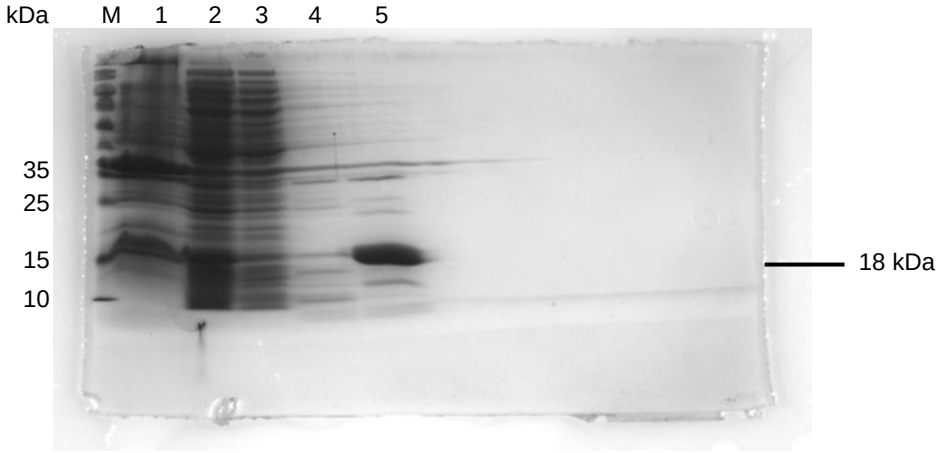


Figure 22: SDS-PAGE analysis from the metal affinity chromatography with the soluble fraction of induced BL21 cells

M-PageRuler prestained protein ladder, 1-pellet before chromatography, 2-supernatant before chromatography, 3-flowthrough, 4-50mM imidazole wash, 5-500mM imidazole elution

4.4.2 Size exclusion chromatography

The 500 mM imidazole elution fraction from the metal affinity chromatography was then used for a SEC to determine the approximate size and the state of the sample after the metal affinity chromatography. The column which was used was a Superdex 200 10/300GL column (Sigma Aldrich, USA). The injection volume was 100 µl. The chromatogram of the SEC is shown in Figure 23.

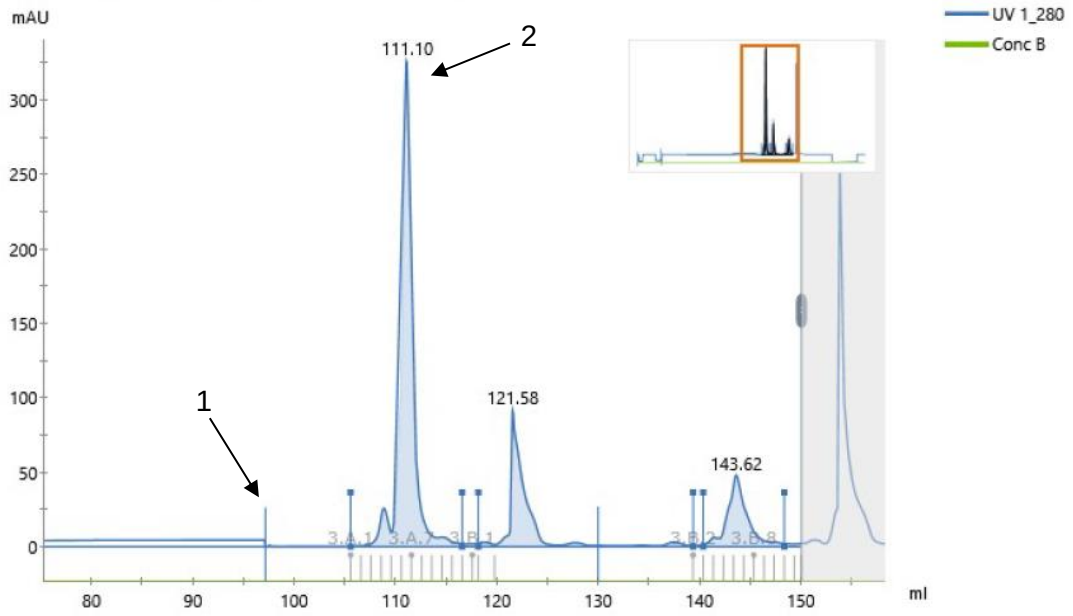


Figure 23: Size exclusion chromatography with the modified version of MSMEG_6227 after metal affinity chromatography

1-sample injection, 2-first elution peak (fractions A5 – A9)

The first elution peak appeared at approximately 14 ml after the sample injection. With this retention on this specific column, and in combination with the dynamic light scattering data a molecular size of approximately 70 kDa was suggested. This could correspond to a tetramer of the modified protein version. The sample was collected using a fractionator and the sample was run on a 12.5% isocratic SDS-PAGE gel. The result from the SDS-PAGE analysis is shown in Figure 24.

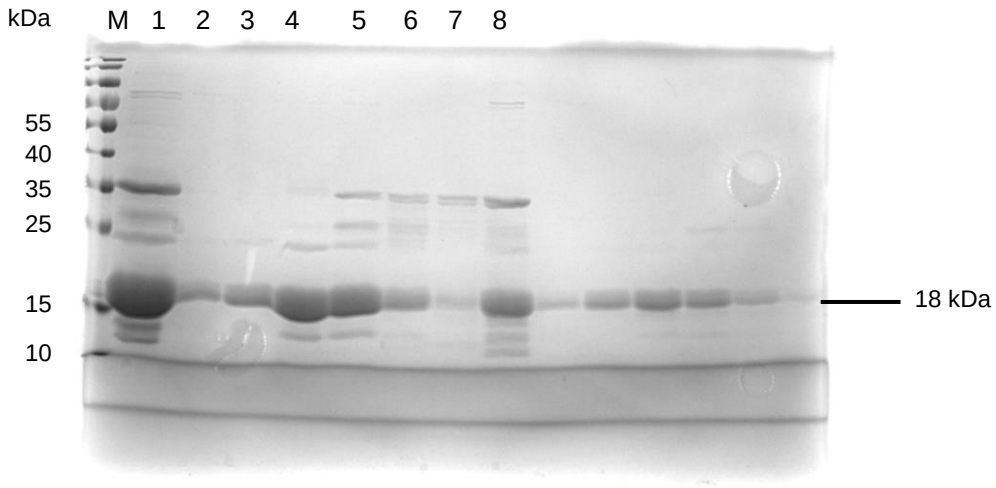


Figure 24: SDS-PAGE analysis of the size exclusion fractions from the elution fraction from metal affinity chromatography with MSMEG_6227 modified version.

M-PageRuler prestained protein ladder, 1-elution fraction, 2-A4, 3-A5, 4-A6, 5-A7, 6-A7, 7-A8, 8-A9

The fractions in lanes 4 and 5 are from the elution peak in Figure 21. The band in the SDS-PAGE corresponds to the expected size of the modified protein version. The fact that the protein had the expected size, that it appeared as a tetramer in the SEC, and that the SDS-PAGE confirmed that it was the desired protein the samples from the elution peak in Figure 23 were used for further experiments.

4.5 Crystallization trials with the S2 - R105 MSMEG_6227 deletion mutant

With the elution fractions from the metal affinity chromatography shown in Figure 21 crystallization trials were set up. The crystallization trials were done using the sitting drop crystallization technique with the following commercially available screens: Structure Screen 1, Structure Screen 2, Morpheus, Morpheus II, PEG/Ion, Shot gun 1, PEG/Rx, MemGold, PGA Screen, and Crystal Screen. All plates were prepared using an automated machine. The pipetted screens were kept in the fridge at 4 °C and checked after 6 days. After 6 days some salt crystals were observed. After 4 months of crystallization protein crystals were found in some conditions. One condition in which a protein crystal was suspected was condition B12 from the Morpheus II screen. Another potential protein crystal was found in well F12 of Crystal Screen.

From the condition B12 of the Morpheus II screen and condition F12 of the Crystal Screen protein crystals were yielded and analyzed at the Synchrotron facility in Berlin, Germany. However, no diffraction was achieved with the yielded crystals.

5 Discussion

Tuberculosis is still one of the widest spread diseases caused by one of the deadliest microorganisms with an annual death toll of approximately 1.4 million people globally. [9] Due to the long treatment regimens with antibiotics to treat the infection with *M. tuberculosis* antibiotic resistance is common among *M. tuberculosis*. Furthermore, the ability of the bacteria to undergo transformation into a dormant state complicates the treatment of tuberculosis. A central research field is the development of effective drugs for the tuberculosis treatment.

A very well-known, non-pathogenic, and fast-growing model organism for the study of *M. tuberculosis* is *M. smegmatis*. Many structural features of *M. smegmatis* can be transposed to *M. tuberculosis*.

The research group of Trutneva and Co. investigated the proteome of dormant *M. smegmatis* cells. One of the proteins they found to be active in the dormant state only was the putative transcriptional regulator MSMEG_6227. From multiple alignments it is believed that this putative transcriptional regulator plays a crucial role in detoxification and antibiotic resistance in *Mycobacteria*. As a potential target for a medication against tuberculosis MSMEG_6227 poses a potential target. To further investigate the function of MSMEG_6227 in dormant *Mycobacteria,* its role in detoxification and antibiotic resistance, and its role as potential drug target its structure needs to be determined.

The goal of this thesis was to produce crystals of the putative transcriptional regulator MSMEG_6227 from *M. smegmatis* and ultimately determine the tertiary protein structure. Extensive structural studies with the putative transcriptional regulator MSMEG_6227 from *M. smegmatis* have not been carried out so far.

The full-length protein was cloned into *E. coli* BL 21 cells and expressed for 4 hours at 37°C without induction. The protein was detected in the soluble fraction of the cell extract and further purified. The purification of the full-length protein was successful. However, it was found that aggregates with the size of approximately 1000kDa are formed. It was not futher investigated why this behavior was observed. A possible explanation may be found in the predicted structure of this protein. The N-terminal end contains a part which is characterized by a high amount of interchanging proline and glycine residues. This may lead to a

disordered domain in the protein structure and the observed aggregation in the experiments. Moreover, it was not determined if the aggregation already happens during expression or if the aggregation is a consequence of the purification steps done on the protein.

To prevent aggregation during expression or purification the sequence of the disorder part, Serine on position 2 until arginine at position 105, was removed by the means of PCR reactions. After ligating the modified DNA sequence with the pET19b plasmid, and transformation into *E. coli* cells, the DNA was sequenced. Sequencing, however, indicated a Serine to Proline mutation in the C-terminal domain.

Nevertheless, because this mutation was not found to be crucial in the protein further experiments with this protein were carried out. The protein which had the domain S2 - R105 removed was expressed and purified. Expression and purification were done at the same conditions as for the full-length protein but with IPTG induction.

However, the shortened version was found to be present as a tetramer after expression and purification. With these results crystallization trials were set up. The plates were checked after 6 days and after 4 months. After 4 months crystals were found in two conditions. Analysis of these crystals was performed at the Synchrotron facility in Berlin, Germany. The crystals were confirmed as protein crystals, but they did not diffract during the X-ray crystallographic measurements. As a result, no structure determination could be performed in this thesis.

Summarizing, a way to express and purify the shortened version of MSMEG_6227 was implemented and crystallization trials were set up.

The next steps to solve the structure of MSMEG_6227 are the following:

- Set up more crystallization trials with the already obtained shortened version of MSMEG_6227.
- Obtain the shortened version of MSMEG_6227 without the S -> P mutation in the C-terminal part.
- Find suitable crystallization conditions and scale up crystallization experiments with the shortened version of MSMEG_6227.
- Determine the structure of the S2 -> R105 MSMEG_6227 deletion mutant by X-ray crystallography.
- Optimization of the expression and purification procedure for the full length MSMEG_6227 version

    - Try different buffer compositions.
    - Try different pH values.
    - Try other purification methods.
- Set up crystallization trials with the full length MSMEG_6227 protein.
- Scale up the crystallization with the full length MSMEG_6227 protein.
- X-ray crystallography and structure determination with the full length MSMEG_6227 protein.

Because of the SARS-CoV-2 pandemic I could not further work on this thesis due to the limited time.

6 References

[1] Cook G.M., Berney M., Gebhard S., Heinemann M., Cox R.A., Danilchanka O., Niederweis M. Physiology of mycobacteria. Adv Microb Physiol. 2009; 55:81-182, 318-9. doi: 10.1016/S0065-2911(09)05502-7. PMID: 19573696; PMCID: PMC3728839.

[2] Brennan P.J., Nikaido H. The envelope of mycobacteria. Annu Rev Biochem. 1995; 64:29-63. doi: 10.1146/annurev.bi.64.070195.000333. PMID: 7574484.

[3] Alderwick L.J., Harrison J., Lloyd G.S., Birch H.L. The Mycobacterial Cell Wall--Peptidoglycan and Arabinogalactan. Cold Spring Harb Perspect Med. 2015;5(8):a021113. Published 2015 Mar 27. doi:10.1101/cshperspect.a021113

[4] Abrahams K.A., Besra G.S. Synthesis and recycling of the mycobacterial cell envelope. Curr Opin Microbiol. 2021 Apr;60:58-65. doi: 10.1016/j.mib.2021.01.012. Epub 2021 Feb 18. PMID: 33610125.

[5] Batt S.M., Burke C.E., Moorey A.R., Besra G.S. Antibiotics and resistance: the two-sided coin of the mycobacterial cell wall. Cell Surf. 2020;6:100044. Published 2020 Sep 2. doi:10.1016/j.tcsw.2020.100044

[6] Bhatt A., Fujiwara N., Bhatt K., Gurcha S.S., Kremer L., Chen B., Chan J., Porcelli S.A., Kobayashi K., Besra G.S., Jacobs W.R. Jr. Deletion of kasB in Mycobacterium tuberculosis causes loss of acid-fastness and subclinical latent tuberculosis in immunocompetent mice. Proc Natl Acad Sci U S A. 2007 Mar 20;104(12):5157-62. doi: 10.1073/pnas.0608654104. Epub 2007 Mar 12. PMID: 17360388; PMCID: PMC1829279.

[7] Yegian D., Vanderlinde R.J. The Nature of Acid-Fastness. J Bacteriol. 1947;54(6):777-783. doi:10.1128/JB.54.6.777-783.1947

[8] Barksdale L., Kwang-Shin K. Mycobacterium. Bacteriological Reviews. 1977;41(1):217-377.

[9] World Health Organization. Global Tuberculosis Report 2020. p. 1-2.

[10] World Health Organization. Global Tuberculosis Report 2016. p. 1-53.

[11] Gengenbacher M., Kaufmann S.H. Mycobacterium tuberculosis: success through dormancy. FEMS Microbiol Rev. 2012 May;36(3):514-32. doi: 10.1111/j.1574-6976.2012.00331.x.

[12] Smith I. Mycobacterium tuberculosis pathogenesis and molecular determinants of virulence. *Clin Microbiol Rev*. 2003;16(3):463-496. doi:10.1128/cmr.16.3.463-496.2003

[13] Loddenkemper R., Lipman M., Zumla A. Clinical Aspects of Adult Tuberculosis. Cold Spring Harb Perspect Med. 2015;6(1):a017848. Published 2015 Feb 6. doi:10.1101/cshperspect.a017848

[14] T JAS., J R., Rajan A., Shankar V. Features of the biochemistry of Mycobacterium smegmatis, as a possible model for Mycobacterium tuberculosis. J Infect Public Health. 2020 Sep;13(9):1255-1264. doi: 10.1016/j.jiph.2020.06.023.

[15] Salina E.G., Zhogina Y.A., Shleeva M.O., Sorokoumova G.M., Selishcheva A.A., Kaprelyants A.S. Biochemical and morphological changes in dormant ("Nonculturable") Mycobacterium smegmatis cells. Biochemistry (Mosc). 2010 Jan;75(1):72-80. doi: 10.1134/s0006297910010098. PMID: 20331426.

[16] Trutneva K., Shleeva M., Nikitushkin V., Demina G., Kaprelyants A. Protein Composition of Mycobacterium smegmatis Differs Significantly Between Active Cells and Dormant Cells with Ovoid Morphology. Front Microbiol. 2018 Sep 4;9:2083. doi: 10.3389/fmicb.2018.02083.

[17] Dick T., Lee B.H., Murugasu-Oei B. Oxygen depletion induced dormancy in Mycobacterium smegmatis. FEMS Microbiol Lett. 1998 Jun 15;163(2):159-64. doi: 10.1111/j.1574-6968.1998.tb13040.x.

[18] Alberts B., Johnson A., Lewis J., Morgan D., Raff M. Molecular Biology of the Cell. 6th ed. 376-377.

[19] Styrer L., Gatto G.J. Jr., Tymoczko J., Berg J. Biochemistry. 9th ed.

[20] Inukai S., Kock K.H., Bulyk M.L. Transcription factor-DNA binding: beyond binding site motifs. Curr Opin Genet Dev. 2017;43:110-119. doi:10.1016/j.gde.2017.02.007

[21] Laity J.H., Lee B.M., Wright P.E. Zinc finger proteins: new insights into structural and functional diversity. Curr Opin Struct Biol. 2001 Feb;11(1):39-46. doi: 10.1016/s0959-440x(00)00167-6.

[22] Aravind L., Anantharaman V., Balaji S., Babu M.M., Iyer L.M. The many faces of the helix-turn-helix domain: transcription regulation and beyond. FEMS Microbiol Rev. 2005 Apr;29(2):231-62. doi: 10.1016/j.femsre.2004.12.008.

[23] Massari M.E., Murre C. Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Mol Cell Biol*. 2000;20(2):429-440. doi:10.1128/mcb.20.2.429-440.2000

[24] Miller M., Shuman J.D., Sebastian T., Dauter Z., Johnson P.F. Structural basis for DNA recognition by the basic region leucine zipper transcription factor CCAAT/enhancer-binding protein alpha. J Biol Chem. 2003 Apr 25;278(17):15178-84. doi: 10.1074/jbc.M300417200.

[25] Park S.C., Kwak Y.M., Song W.S., Hong M., Yoon S.I. Structural basis of effector and operator recognition by the phenolic acid-responsive transcriptional regulator PadR. Nucleic Acids Res. 2017 Dec 15;45(22):13080-13093. doi: 10.1093/nar/gkx1055.

[26] Fibriansah G., Kovács Á.T., Pool T.J., Boonstra M., Kuipers O.P., et al. (2012) Crystal Structures of Two Transcriptional Regulators from Bacillus cereus Define the Conserved Structural Features of a PadR Subfamily. PLOS ONE 7(11): e48015. https://doi.org/10.1371/journal.pone.0048015

[27] Isom C.E., Menon S.K., Thomas L.M., West A.H., Richter-Addo G.B., Karr E.A. Crystal structure and DNA binding activity of a PadR family transcription regulator from hypervirulent Clostridium difficile R20291. BMC Microbiol. 2016 Oct 4;16(1):231. doi: 10.1186/s12866-016-0850-0.

[28] https://www.ncbi.nlm.nih.gov/protein/ABK71118.1 on 14.04.2021 at 21:48

[29] Kumari M., Pal R.K., Mishra A.K., Tripathi S., Biswal B.K., Srivastava K.K., Arora A. Structural and functional characterization of the transcriptional regulator Rv3488 of Mycobacterium tuberculosis H37Rv. Biochem J. 2018 Nov 9;475(21):3393-3416. doi: 10.1042/BCJ20180356.

[30] De Silva R.S., Kovacikova G., Lin W., Taylor R.K., Skorupski K., Kull F.J. Crystal structure of the virulence gene activator AphA from Vibrio cholerae reveals it is a novel member of the winged helix transcription factor superfamily. J Biol Chem. 2005 Apr 8;280(14):13779-83. doi: 10.1074/jbc.M413781200.

7 Appendix

7.1 Nucleotide sequence of the gene coding for MSMEG_6227

ATGAGCACCCCCTTTGCATTTCCCACCGGCGACTTCGGATTCGGCCCCGCCGGC
CGTCGCGCCATGCTCCACCACCGTCGAGCGGCCCGTCGCGAGTTCCGCGACCA
GATGCGGGCCCACCTCCACGAGGCCCGCGAGCAGGCCCTCGACCCGCGCTGCG
AGATGGGCGGGCCCGGCGGTCCCGGCCCCATGGGCGGACCGGGCCGCGGCTTC
GGCGGGTTCGGGTTCGGCTTCGATCCGGGCGCGGGCTTCGGGTTCGGTCCGGG
CGGACCGCGCGGACGCCGCGGCCACGGCCGCGGGCGGCGCGGCCGTCGCGGT
GACGTGCGCGCCGCAATCCTGAAACTGCTGGCCGAGCGCCCCATGCACGGCTA
CGAGATGATCCAGGAGATCGGCGAGCGCACCGACAACCTGTGGCGCCCCAGC
CCCGGTTCGGTGTATCCCACACTGCAACTGCTGGTCGACGAGGGCCTGATCAG
CGGCACCGAGGCCGAGGGCAGCAAGAAGCTGTTCGAACTCACCGAGGCAGGT
CGCGCGGCCGCCGAGGCGATCGAGACCCCGCCGTGGGAGCAGATCGCCGAAG
ACGTGGACCCGGCCGCGGTCAACCTGCGCGGGGCCATCGCCCAGTTGATGGGT
GCGGTCGCCCAGTCGGCGTACACCGCGACCGAGGACCAGCAGCAACGCATCCT
CGACGTGGTCAACAACGCCCGCCGCGAGATCTACCAGATCCTCGGCGAGGAGT
GA

7.2 Amino acid sequence of MSMEG_6227 (full length)

MSTPFAFPTGDFGFGPAGRRAMLHHRRAARREFRDQMRAHLHEAREQALDPRCE
MGGPGGPGPMGGPGRGFGGFGFGFDPGAGFGFGPGGPRGRRGHGRGRRGRRGD
VRAAILKLLAERPMHGYEMIQEIGERTDNLWRPSPGSVYPTLQLLVDEGLISGTEA
EGSKKLFELTEAGRAAAEAIETPPWEQIAEDVDPAAVNLRGAIAQLMGAVAQSAY
TATEDQQQRILDVVNNARREIYQILGEE

7.3 Amino acid sequence of MSMEG_6227 (modified version)

MGDVRAAILKLLAERPMHGYEMIQEIGERTDNLWRPSPGSVYPTLQLLVDEGLISG
TEAEGSKKLFELTEAGRAAAEAIETPPWEQIAEDVDPAAVNLRGAIAQLMGAVAQ
SAYTATEDQQQRILDVVNNARREIYQILGEE

7.4 Vector map of pET19-b

taken from:
https://www.snapgene.com/resources/plasmid-files/?set=pet_and_duet_vectors_(novagen)&plasmid=pET-19b&format=png on 27.04.2021 at 10:30 pm.