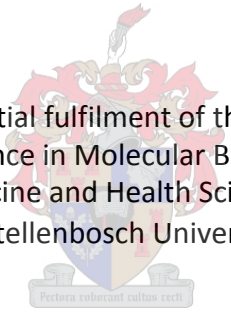
The identification of novel proteins involved in iron-sulphur cluster biogenesis in mycobacteria

by
Jesmine Arries



Thesis presented in partial fulfilment of the requirements for the degree Master of Science in Molecular Biology in the Faculty of Medicine and Health Sciences
at Stellenbosch University

Supervisor: Dr Monique J. Williams
Co-supervisor: Prof Samantha L. Sampson
Faculty of Medicine and Health Sciences

March 2016

Declaration

By submitting this thesis/dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

March 2016

Abstract

Mycobacterium tuberculosis remains one of the world's deadliest pathogens, and although decades of research have been dedicated to the discovery and development of anti-tuberculosis agents and vaccines, tuberculosis (TB) remains an ongoing global public health threat. Given the emergence of drug resistance and the poor treatment outcomes associated with this disease, it is imperative that we improve our understanding of the pathogenesis of *M. tuberculosis* in order to develop new anti-TB drugs and strategies for shortening drug treatment. Iron-sulphur (Fe-S) clusters are ubiquitous cofactors that are required for the maturation of various proteins, many of which are involved in essential biological processes. Multiprotein complexes are required for the *in vivo* assembly of Fe-S clusters, and the SUF system, encoded by the *Rv1460-Rv1461-Rv1462-Rv1463-csd-Rv1465-Rv1466* operon in *M. tuberculosis*, is thought to be the major Fe-S cluster biogenesis machinery in this organism. This process is poorly understood in mycobacteria, and it is currently unclear if proteins outside of this operon are involved in Fe-S cluster biogenesis. In this study, we sought to identify novel proteins involved in Fe-S cluster assembly in mycobacteria using various approaches. Initially we optimised blue-native PAGE for the analysis of mycobacterial cytoplasmic protein complexes, and demonstrated that protein complexes could be identified by immunodetection of BN-PAGE and subsequent gel fraction analysis by mass spectrometry. However, due to the complexity of the cytoplasmic fraction and poor resolution of this technique, we were unable to use it to identify complexes containing the Fe-S cluster biogenesis machinery. In a second approach, we set out to generate an antibody against Csd, an Fe-S cluster assembly protein, to facilitate immunodectection of the Fe-S cluster biogenesis machinery. However, heterologous expression of the *M. tuberculosis* Csd in *Escherichia coli* proved unsuccessful due to insolubility of the recombinant protein. In an attempt to achieve soluble expression of Csd, a 6xHis-tag fusion was expressed in the closely related organism *Mycobacterium smegmatis*. Soluble expression of the fusion protein was optimised and confirmed by immunodetection. We then employed affinity enrichment using nickel-charged magnetic beads and mass spectrometry to identify potential interacting partners for Csd. Strains expressing a 6xHis-tag SufC fusion and a 6xHis-tag mCherry fusion were also included for comparison. Utilising this method, we were able to demonstrate affinity enrichment for Fe-S cluster assembly

proteins, and Fe-S cluster-containing enzymes, as well as identify potential novel interacting partners. This work has established the methodology for identifying novel protein-protein interactions in mycobacteria, and lays the foundation for elucidating the process of Fe-S cluster assembly in mycobacteria.

Opsomming

Mycobacterium tuberculosis is nog steeds een van die mensdom se dodelikste patogene ten spyte van intensiewe navorsing wat fokus op die ontdekking en ontwikkeling van antimikrobiese middels en vaksiene om tuberkulose (TB) te bekamp. Die toename in antimikrobiese weerstandbiedendheid in hierdie patogeen sowel as die onbereidwilligheid van pasiënte om die streng medisynekursus te volg, beklemtoon die noodsaaklikheid om die organisme en sy oorlewingsmeganismes beter te verstaan. Dit kan lei tot die ontwikkeling van nuwe teen-TB middels en die verkorting van middelbehandeling. Yster-swawel (Fe-S) komplekse dien as kofaktore vir verskeie proteïene, waarvan baie noodsaaklike rolle verig in sekere biologiese prosesse. Multiproteïen komplekse is nodig sodat Fe-S komplekse *in vivo* gesintetiseer kan word – 'n proses wat nog nie goed beskryf is vir mikobakterieë nie. Die SUF sisteem, wat in *M. tuberculosis* deur die geenoperon *Rv1460-Rv1461-Rv1462-Rv1463-csd-Rv1465-Rv1466* geënkodeer word, word beskou as die hoofkomponent in die biogenese van Fe-S komplekse. Onsekerheid heers of ander proteïene buite dié van die operon ook 'n rol speel in biogenese. In hierdie studie het ons gepoog om nuwe proteïene te identifiseer wat betrokke kan wees in die biogenese van Fe-S komplekse deur verskeie benaderings te volg. Aanvanklik het ons die tegniek van blou-natuurlike poliakrielamied jelelektroforese (blue-native PAGE) geöptimiseer om mikobakteriële sitoplamiese proteïenkomplekse te analiseer. Ons het daarmee geslaag om proteïenkomplekse te identifiseer met behulp van immuno-opsporing gevolg deur jelfraksie-analise met massa-spektrometrie. Vanweë die kompleksiteit van die sitoplasmiese fraksie en die swak resolusie van die tegniek is geen komplekse wat Fe-S komplekse bevat, geïdentifiseer nie. In die tweede benadering is daar gepoog om 'n teenliggaam te genereer vir Csd, 'n proteïen betrokke in die biogenese van Fe-S komplekse. Dit is ten doel om die immuno-opsporing van die Fe-S komplekse te vergemaklik. Heteroloë uitdrukking van Csd in *E. coli* het gelei tot die produksie van 'n onoplosbare proteïen. Deur die proteïen uit te druk in die nabyverwante *M. smegmatis*, tesame met 'n 6xHis fusie, het gelei tot 'n oplosbare produk wat later bevestig is deur immuno-opsporing. Ons het daarna affiniteit-verryking gebruik met behulp van nikkel-gelaaide magnetiese krale en massa-spektrometrie om proteïene wat moontlike interaksies met Csd het, te identifiseer. Die nodige kontroles (rasse wat 'n 6xHis fusie het met onderskeidelik 'n *sufc* en 'n *mcherry* geen) is ingesluit ter vergelyking. Met die gebruik van

hierdie metode het ons affiniteit-verryking vir proteïene betrokke in die biogenese van Fe-S komplekse gewys asook Fe-S kompleksbevattende ensieme. Ons het terselfdetyd nuwe potensiële proteïene geïdentifiseer wat interaksie toon met Fe-S komplekse. Hierdie studie het ‘n metode daargestel vir die identifikasie van nuwe proteïen-proteïen interaksies in mikobakterieë en lê ‘n fondasie vir die toeligting van die proses van die biogenese van Fe-S komplekse in mikobakterieë.

Acknowledgements

Firstly, I would like to thank my supervisor, Dr Monique Williams, for her generous contribution to the design and conceptualization of this work, for always having an open door, for the hands-on support and for setting an example by never compromising her integrity. Also, thank you for the laughs! ☺

I would like to thank my team, the Fe-S cluster-cluster, for always being willing to help, for their insightful ideas and for always provoking thought and keeping me on my toes.

I would also like to thank my co-supervisor, Prof. Samantha Sampson for her guidance and contribution to the completion of this manuscript.

I would like to extend a special thank you to my colleague and friend, Dr Tiaan Heunis, for his substantial contribution to the success of this project. I couldn't have done it without you.

I would like to thank Niel van Wyk, Jason Limberis and Nastassja Kriel for their invaluable input and support.

Furthermore, I would like to thank all the members of Lab F424 for their support.

I would like to thank the MRC NHSP initiative for their generous financial support.

I would also like to thank the NRF and the Harry Crossley fund for their financial support for this project.

Dedications

I would like to dedicate this work to the memory of the pillars of strength I have lost during my postgraduate studies.

Percy, Luqie, Papa…

The lessons you have taught me cannot be found in a book

I carry them in my heart day by day

I will love and miss you always

Rest in peace

Table of Contents

List of Figures

List of Tables

Nomenclature

WHO	World Health organisation
BCG	Bacillus Calmette–Guérin
DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
HIV	Human immunodeficiency virus
Hyg	Hygromycin
Kan	Kanamycin
Amp	Ampicillin
cDNA	Complementary deoxyribonucleic acid
ATP	Adenosine triphosphate
β	Beta
bp	Base pairs
$MgCl_2$	Magnesium chloride
min	Minutes
ml	Mililitre
mm	Milimetre
mM	Milimolar
mRNA	Messenger RNA
OD	Optical density
PCR	Polymerase chain reaction
qRT-PCR	Quantitative real time PCR
Tween-80	Polyoxyethylene sorbitan monooleate
U	Units
V	Volt
M	Micro
µF	MicroFarad
Ω	Ohm
× *g*	Centrifugal force
PVDF	Polyvinylidene fluoride
CaCl	Calcium chloride
TB	Tuberculosis
SUF	Sulphur mobilization system

BN-PAGE	Blue native polyacrylamide gel electrophoresis
Fe-S	Iron-sulphur
Csd	Cysteine desulphurase
PPI	Protein-protein interaction
Y2H	Yeast two-hybrid
ISC	Iron-sulphur cluster system
NIF	Nitrogen fixation system
PCA	Protein complementation assay
DHFR	Dihydrofolate reductase
BACTH	Bacterial two-hybrid
ATP	Adenosine triphosphate
cAMP	Cyclic adenosine monophosphate
M-PFC	Mycobacterial protein fragment complementation
BiFC	Bimolecular fluorescence complementation
BiLC	Bimolecular luminescence complentation
DmsD	Dimethyl sulfoxide subunit D
Tat	Twin-arginine translocation
BRET	Bioluminescence resonance energy transfer
FRET	Fluorescence resonance energy transfer
YFP	Yellow fluorescent protein
CFP	Cyan fluorescent protein
CoIP	Co-immunoprecipitation
FAS-II	Fatty acid synthase-II
eGFP	Enhanced green fluorescent protein
MS	Mass spectrometry
AP-MS	Affinity purification followed by mass spectrometry
IMEx	International Molecular Exchange
MPIDB	Microbial protein interaction database

PATRIC	Pathosystems Resource Integration Centre
RT	Room temperature
LB	Luria Bertani
LA	Luria Bertani agar
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Atc	Anhydrotetracycline
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
NO	Nitric oxide
DETA/NO	Diethylenetriamine/nitric oxide adduct
H_2O_2	Hydrogen peroxide
LDS	Lithium dodecyl sulphate
MES	2-(N-morpholino)ethanesulfonic acid
HRP	Horse radish peroxidase
LC-MS/MS	Liquid chromatography-mass spectrometry/tandem mass spectrometry

Chapter 1: Literature review

The investigation of protein-protein interactions in mycobacteria

1.1. Introduction

Mycobacterium tuberculosis, the causative agent of tuberculosis (TB), remains one of the leading causes of death worldwide, claiming the lives of approximately 2 million people annually[1]. *M. tuberculosis* was first identified in 1882 by Robert Koch, and has since claimed the lives of thousands[2]. Despite decades of research aimed at the development of anti-TB drugs and vaccines, TB remains a global health threat. Upon initial *M. tuberculosis* infection, most people are able to contain the infection, leading to the development of a latent infection. Latent TB is defined as a persistent immune response to *M. tuberculosis* in the absence of clinical symptoms of disease, and is estimated to affect one third of the world's population[3]. This colossal reservoir of potential active TB patients makes this disease particularly challenging to contain. Reactivation of latent *M. tuberculosis* infection occurs in approximately 10% of infected individuals. An estimated 9 million new infections occur each year, and in sub-Saharan Africa the risk of progression to active disease is greatly increased due to HIV-TB co-infection, where the prevalence is highest. Given the increase in drug resistance and the poor adherence of TB patients to a lengthy and arduous anti-TB drug regimen, it is critical to investigate the virulence pathways of *M. tuberculosis* and find new ways to treat this devastating disease[4].

In recent years, exploring protein-protein interactions (PPIs) has become a popular approach for elucidating bacterial physiology and pathogenesis. Here we review current PPI detection techniques, with a focus on methods applied in mycobacteria for both initial identification of PPIs and the validation thereof. We also discuss high-throughput PPI detection methods and databases used to build and catalogue interaction networks.

1.2. Defining protein-protein interactions

Proteins are large macromolecules consisting of specific amino acid chains necessary for the cellular function of all living organisms, and have diverse roles within the cell. These functions include gene regulation, catalysis of chemical reactions, mediation of the immune response, involvement in signalling cascades, and maintaining the structural stability of the

cell[5]. Studying cellular processes on a protein level is crucial for understanding the physiology and pathogenesis of bacteria[6]. In the pre-genomic era, molecular biologists established the central dogma which described the flow of genetic information from DNA to protein. Genes and proteins were largely studied on an individual level, giving very little insight into how these cellular processes occur spatially and over time. In the post-genomic era, high-throughput techniques have provided an abundance of transcriptomic and proteomic data, giving us a more holistic view of cellular metabolism and allowing us to define the relationship between proteins and genes within a given pathway or complex[7,8]. PPI studies are a recent focus in protein biochemistry. The identification of these interactions and mapping thereof to pathways and complexes have led to the design of PPI networks which all contribute to the construction of the interactome[9].

Proteins rarely function in isolation, but rather rely on interactions with other proteins for their function. This interaction is defined as two proteins coming into contact via a specific docking mechanism to perform a function within the cell that may be distinct from the independent function of these proteins[10]. These interactions are either homo-dimeric, when two copies of the same protein interact, or hetero-dimeric, when two different proteins interact. Proteins can also interact as part of higher order complexes, where more than two proteins interact within a complex or pathway to achieve a singular outcome. Permanent interactions are defined as stable, whereas fleeting interactions are defined as transient[11]. Transient interactions are typically involved in signalling cascades whereas stable interactions are found predominantly in the main cellular machinery such as enzymatic complexes and ribosomal subunits[12]. Although proteins bind with varying affinities, interacting proteins bind at their interface via complementary docking sites with extreme specificity, a direct result of their physicochemical properties. These properties include covalent bonds, hydrogen bonds, hydrophobic bonds, ionic interactions and Van der Waals forces[13]. Binding domains associated with PPIs are highly conserved between species and throughout evolution[14].

Proteins can also be involved in multiple interactions, adding a level of complexity to bacterial physiology that cannot be predicted by genome sequence analysis[15]. PPIs can occur concurrently or consecutively depending on the cellular environment and demand at any point in time. These interactions are tightly regulated within the cell to ensure cellular

survival and maintain homeostasis. The frequency of PPIs are influenced by a number of factors including perturbations in cellular homeostasis caused by environmental stressors and changes in protein concentrations[13]. Previous computational analyses of PPIs have shown that interactions between proteins which are situated adjacent to each other in the genome and are involved in metabolism and regulation occur most frequently[12].

1.3. Methods for studying protein-protein interactions and their application in mycobacteria

Two groups of techniques are generally employed for the detection of PPIs, namely, binary and co-complex techniques. Binary techniques are used to specifically determine whether two proteins interact, whereas co-complex methods are commonly used to identify two or more interacting partners[10].

1.3.1. Binary techniques

1.3.1.1. Yeast two-hybrid

Yeast two-hybrid (Y2H) is a technique used for detecting binary PPIs and was developed in 1989 for application in yeast cells[16]. A bait protein is expressed as a fusion with the DNA-binding domain and an individual protein or libraries of prey proteins are expressed as fusions with the activation domain of a site-specific transcription factor. If the bait and prey proteins interact, the DNA-binding domain and activation domain are brought into close enough proximity for the transcription factor to be reconstituted. This in turn permits the expression of downstream reporter genes and results in growth on selective media or a colour reaction. Once the bait and prey fusion proteins are expressed in the cytosol, they must be able to translocate into the nucleus so that activation of transcription can occur. One of the major limitations is the high frequency of false positives associated with the Y2H system, with an estimated 50% inaccuracy[17]. Y2H is not suitable for detecting interactions involving more than two proteins or those dependent on post-translational modifications[18]. In addition, Y2H cannot be used to study membrane proteins, transcription factors, PPI kinetics or strongly acidic proteins[19]. Y2H may also not be suitable for interactions which only occur under certain conditions, such as oxidative stress. An example of this is the phosphorylation dependent FtsZ-FipA interaction in *M. tuberculosis* which occurs during cell division under oxidative stress conditions, which could not be confirmed

by Y2H[20]. However, due to the limited number of feasible alternatives, Y2H remains the most commonly used technique for detecting PPIs. To overcome this limitation, Y2H is used in conjunction with other methods for detecting PPIs in order to validate Y2H interactions.

The use of Y2H for the study of PPIs was originally proposed by Fields and Song, who later proved this concept by demonstrating the interaction between SNF1 and SNF4[16]. The principle of Y2H was confirmed by Ma and Ptashne using two known interacting proteins, GAL4 and GAL80, which is an inhibitor of GAL4. They showed that fusing GAL80 to an activation domain facilitated the initiation of transcription in the presence of GAL4 fused to the GAL80 DNA binding domain[21]. In the late 1990s, Y2H played a significant role in validating protein interactions involved in bacterial cell division, especially those involved in Z-ring formation. FtsZ, a tubulin homolog, has been shown to interact with FtsA, an actin homolog in *Bacillus subtilis*[22], *Caulobacter crescentus*[23], *Escherichia coli* and *Staphylococcus aureus*[24]. The *E. coli* FtsZ has also been shown to interact with the *B. subtilis* FtsA, indicating evolutionary conservation of the domains responsible for interaction. These domains could provide valuable information about bacterial pathogenicity and could serve as potential antimicrobial drug targets[24].

The application of Y2H in mycobacteria dates back to the year 2000, where it was used to show interaction between the *M. tuberculosis* single-stranded DNA binding protein, *Mtu*SSB, and its *E. coli* homolog, *Eco*SSB. These proteins were shown to be capable of heterooligomerization[25]. The following year, Y2H was employed to investigate the interacting partners of the 4.2 domain of RpoV[26]. The principle sigma factor, RpoV, was previously shown to restore virulence to an attenuated *Mycobacterium bovis* strain containing a point mutation in the 4.2 domain of RpoV[27]. In this study, WhiB3 was shown to interact with the 4.2 domain of RpoV in virulent strains but not with the mutated domain. WhiB3 was proposed to be involved in regulating the transcription of genes that affect the host immune response[26]. WhiB3 was later characterised as a core transcriptional regulator in *M. tuberculosis* which contributes to the maintenance of redox homeostasis and the virulence of *M. tuberculosis*[28,29].

One of the most significant contributions of Y2H to our knowledge of how *M. tuberculosis* evades the host immune response is the discovery of the secretion in

mycobacteria (Snm) pathway required for the secretion of ESAT-6 and CFP-10, two key T-cell antigens. Snm mutants were unable to replicate inside the host macrophage and was unable to inhibit macrophage inflammatory response[30]. Y2H can also be used to screen for antimicrobial agents. Y2H was employed in 2012 to screen for small molecules that inhibit the interaction between mycobacterial L12 and L10, two ribosomal proteins. The study was aimed at identifying anti-TB compounds that could target ribosomal protein interaction[31]. Interaction between Mycobacterium fluoroquinolone resistance protein A (MfpA) and Mycobacterium fluoroquinolone resistance protein B (MfpB) was identified using both Y2H and a protein complementation assay (discussed later). MfpB was previously unknown, and its interaction with MfpA is implicated in protecting DNA gyrase against antibiotics targeting it, such as fluoroquinolone[32].

An example of using Y2H to study the proteins involved in a biosynthetic pathway is its application to studying iron-sulphur (Fe-S) cluster biogenesis in mycobacteria. Y2H was used in 2005 to identify interactions between the sulphur mobilisation (SUF) machinery in *M. tuberculosis*, the only Fe-S cluster biogenesis system identified in this organism[33]. Furthermore, in 2006, Y2H was used to show that SufB, an essential protein in the SUF system[34,35], requires protein splicing for its function, suggesting that the inhibition of protein splicing could be a potential target for anti-TB drugs[36]. Moreover, in 2014, the interaction network of IScSMtb, one of three cysteine desulphurases in *M. tuberculosis*[33,37–41], was identified using both Y2H and pull-down assays, and a novel mechanism of sulphur transfer by direct interaction with Fe-S cluster-dependent enzymes[38].

1.3.1.2. Protein complementation and biomolecular fluorescence

Protein complementation assays (PCAs) are based on the concept of splitting a reporter protein in two and fusing each end to two potential interacting proteins, one fragment fused to the bait protein and the other fused to the prey protein[42]. When the bait and prey proteins come into close enough proximity to each other, the fragmented reporter protein is reconstituted, thereby regaining its activity. The original reporter for PCA was dihydrofolate reductase (DHFR)[43]. Over the years many variations of PCAs have been developed. One such modified PCA is the bacterial adenylate cyclase-based 2-hybrid system (BACTH). Reconstituted adenylate cyclase of *Bordetella pertussis* regains activity thereby converting ATP to cAMP, initiating transcription of downstream reporter proteins. Due to

the nature of this system, the frequency of false results is low[44]. BACTH has been used to confirm the interaction between FtsZ, involved in Z-ring formation at the start of cell division, and ClpX, a potential cell division regulator[45].

Mycobacterial protein fragment complementation (M-PFC) is a variation of PCA involving the heterologous expression of split mammalian DHFR in mycobacterial cells. Mammalian DHFR has a lower affinity than bacterial DHFR to the trimethoprim inhibitor, which allows for selection of bacteria containing positive PPIs in the presence of concentrations of trimethoprim that inhibit the bacterial DHFR but not the mammalian DHFR[46]. This system was used to confirm interactions between secreted antigens ESAT-6 and CFP-10. Similarly, M-PFC was also used to confirm PPI in tryptophan (Trp) auxotrophic *M. smegmatis* and *E. coli* between ESAT-6 and CFP-10, amongst others. Here the reporter protein was Trp1 of *S. cerevisiae*, involved in Trp biosynthesis[47]. Reconstitution of Trp1 upon PPI resulted in growth of bacterial cells on Trp-deficient media[48]. M-PFC has also recently been used on a high-throughput scale to identify small molecule inhibitors of PPIs in mycobacteria[49].

Other variants of PCAs include bimolecular fluorescence complementation (BiFC) and bimolecular luminescence complementation (BiLC) which uses fluorescent proteins and luminescent proteins as reporters, respectively. A positive PPI results in the reconstitution of two non-florescent/luminescent reporter protein components, and generates a measurable fluorescent[50] or luminescent[51] signal. A major advantage of BiFC and BiLC is the ability to detect PPI *in vivo*, in real-time. In addition, the strength of the interaction may also be related to the fluorescent or luminescence intensity. BiFC has recently been employed in *E. coli*, together with BACTH and other *in vitro* techniques, to show interaction between dimethyl sulfoxide subunit D (DmsD), a twin-arginine translocation (Tat) pathway-specific chaperone implicated in the maturation of oxidoreductases, and general chaperones as well as proteins involved in molybdenum cofactor biogenesis. These results point toward DmsD acting as a node while ushering its substrate through a cascade of chaperone-mediated events associated with the maturation of protein folding[52]. BiFC has also been used to investigate PPIs involved in protein folding and export in the Tat pathway of Gram-negative and Gram-positive bacteria[53]. BiFC and BiLC have, to date, not been applied in mycobacteria for the detection of PPIs.

1.3.1.3. Fluorescence and bioluminescence resonance energy transfer

Resonance energy transfer is based on the principle of energy transfer between two molecules. Fluorescent resonance energy transfer (FRET) and Bioluminescence resonance energy transfer (BRET) can be used to study PPIs *in vivo*, in real time.

FRET utilizes a donor and acceptor fluorescent protein with overlapping excitation and emission spectra that are expressed as fusions with two possible interacting partners. Yellow fluorescent protein (YFP) and cyan fluorescent protein (CFP), derivatives of green fluorescent protein, are generally used as reporters during FRET. When interaction does not occur, the emission spectrum is represented solely by that of the donor fluorescent protein. When interaction does occur, the acceptor fluorescent protein is excited, thereby emitting at a distinguishable wavelength. Cells expressing fluorescent proteins can be analysed using a confocal microscope or fluorometry. In 2012, a single-cell FRET-based method was employed to track translocation of *M. tuberculosis* and *Mycobacterium marinum* into the host macrophage cytosol to monitor phagolysosomal rupture. Furthermore, *M. tuberculosis* and *M. marinum* strains lacking the secreted antigen associated with membrane rupture, ESX-1 secreted protein ESAT-6, did not display translocation to the cytosol or phagolysosomal rupture, but was restored upon complementation of the ESX-1 secretion system. It was observed that phagolysosomal rupture was followed by necrotic cell death of host macrophages. This study suggests a novel mechanism of *M. tuberculosis* and *M. marinum* to evade the host innate immune response by cytotoxicity, enabling the spread of virulent mycobacteria to new host cells[54].

BRET is superior to FRET due to the fact that an external source of light is not required for excitation of the donor protein, and background due to auto-fluorescence is significantly reduced. The bait protein is expressed as a fusion with luciferase, a bioluminescent enzyme, from *Renilla reniformis*, an anthozoan coelenterate. Luciferase activity on the coelenterazine substrate results in the emission of blue light at 475 nm which overlaps with the excitation spectrum of the acceptor protein, YFP, fused to the prey protein. In the absence of interaction, blue light from the donor is emitted at 475 nm. When interaction does occur, luciferase transfers energy to YFP, resulting in yellow light emitted at 527 nm. BRET has been used in cyanobacteria to identify interactions between circadian clock proteins[55]. BRET is used predominantly in eukaryotes instead of bacteria, most likely

due to the decreased sensitivity resulting from poor expression of the eukaryotic luciferase. It was however modified in 2014 to use a bacterial luciferase, LuxAB from *Photorhabdus luminescens,* to facilitate energy transfer to eYFP. Multiple interactions were identified using this technique, confirming its usefulness in investigating PPIs in bacteria[56]. To date, this technique has not been applied in mycobacteria.

1.3.2. Co-complex techniques

1.3.2.3. Co-immunoprecipitation and affinity purification

Co-immunoprecipitation (CoIP) and affinity purification (pull-down assays) are co-complex techniques that are used for identifying interactions between two or more proteins, based on the affinity of one protein for another protein in the complex[10]. Affinity purification techniques can be used to confirm interactions identified by binary techniques or to identify interactions that occur natively within the cell. CoIP is used when an antibody generated against the bait protein is available. In the absence of an antibody, an affinity purification can be used which is based on cloning a fusion tag in frame with the bait protein and using the appropriate agarose resin for purification.

CoIP has been used to confirm interactions between Lsr2, a global gene regulator, and *MSMEG_*4334, a *M. smegmatis* hypothetical flavoprotein involved in bacterial resistance to oxidative stress. This interaction was originally identified using bacterial 2-hybrid and pull-down assays[57]. In 2004, CoIP was successful in identifying both homotypic and heterotypic interactions between the known proteins of the fatty acid synthase-II (FAS-II) complex responsible for mycolic acid biogenesis, essential for the survival of *M. tuberculosis*[58]. Later, in 2005, CoIP was again used by the same research team to identify accessory interacting partners of the FAS-II complex. This study suggests that one of the identified interaction, KasB-Pks13, could be targeted as an alternative for peptidomimetic antibiotics[59]. In 2010, CoIP was used to confirm interactions between a MazF protein, Rv1495, believed to play a role in dormancy in *M. tuberculosis*, and MtbTopA, the only DNA topoisomerase I in mycobacteria. This study suggests a novel mechanism of bacterial survival by the regulation of MtbTopA by MazF.

Recently a one-step affinity purification method was developed for the identification of PPIs in mycobacteria. Four epitopes were investigated for the isolation of protein

complexes, namely enhanced green fluorescent protein (eGFP), FLAG, protein A and haemagglutinin. Of the four epitopes tested, eGFP proved to be the most suitable candidate due to its cost-effective purification and relatively low background as determined by high-resolution mass spectrometry and proteomic data analysis. The fluorescent nature of eGFP makes it easy to measure expression and is convenient when determining subcellular localization. In this study, RNA polymerase subunit A was used as bait for the co-purification of other members of the RNA polymerase complex[60].

In 2010, Bolstad and Wood combined trichloroacetic acid crosslinking and affinity purification to characterize PPIs within sulphur trafficking systems of *E. coli*. Using this technique, PPIs that form part of the SUF system were identified. In *E. coli* the SUF system is responsible for synthesizing Fe-S clusters under oxidative stress and iron limitation[61–65]. This method was used to confirm a SufE-SufS and SufE-SufB interaction previously identified in *in vitro* studies, and identified interactions between SufE and other Fe-S cluster proteins that may play an important role in Fe-S cluster assembly[66].

A major limitation of using affinity purification to study PPIs in mycobacteria is the problem of misfolding and aggregation associated with heterologous expression of mycobacterial proteins in *E. coli*, [67]. To overcome this problem, Arbing *et al* investigated the use of a maltose binding protein fusion tag to increase protein solubility[68]. Six Esx protein complexes were expressed and purified using a pull-down assay and the crystal structures were determined. This strategy resulted in a significant increase in the proportion of soluble protein, suggesting that this method may also be useful for affinity purification of mycobacterial complexes[69].

Co-complex techniques are limited by the need for *in vitro* manipulation of protein lysates. Other limitations include a bias towards interactions which occur with the strongest affinity, resulting in the loss of transient interactions. Co-complex methods also have limited sensitivity with a high frequency of non-specific interactions[70].

1.3.3. Building interaction networks

Two main techniques are used for high-throughput analysis of PPIs, namely, modified two-hybrid systems and affinity purification followed by mass-spectrometry[71].

Before the development of these techniques, computational predictions tools were used, and continue to play a valuable role in PPI studies.

1.3.3.1. Computational prediction of PPI

Historically, computational PPI predictions were largely based on subunit interface information obtained from protein structure databases[72]. A major advancement in computational predictions was the availability of whole genome data which allowed predictions to be made based on genome sequence data[73]. Information such as gene fusions[74,75], gene co-occurrence[76] and conserved gene order[77,78] were used to predict PPIs. Computational tools including, but not limited to, data mining and machine learning, homology based approaches, structure, domain and motif based approaches have been used to predict various interaction networks as well as host-pathogen PPIs[79]. Computational tools have been used previously in an attempt to predict and construct the interactome of *M. tuberculosis*[80,81]. This is a challenging task given that approximately one third of the open reading frames encoded by the *M. tuberculosis* genome consist of proteins of unknown function. It is for this reason that current PPI data generated using computational tools and homology protein mapping have only covered a fraction of the *M. tuberculosis* genome[82].

1.3.3.2. High-throughput two-hybrid techniques

PPIs are critical during processes such as protein secretion and signal transduction[83]. However, very little PPI data is available for commonly used model organisms and even fewer with global interaction networks. High-throughput screens have been used in recent years to generate large PPI datasets. Y2H is one of the few PPI techniques that have successfully been used on a large scale. This is typically achieved by expressing the desired bait protein as a DNA-binding domain fusion and creating a cDNA or domain library which is then fused to an activation domain for screening[84]. The majority of PPI data has been generated using a modified Y2H method, and this data has been used largely to generate interaction networks for various species[85], including *E. coli*[86], *Helicobacter pilori*[87], *Saccharomyces cerevisiae*[88–90] and *Drosophila melanogaster*[91], to name a few. Recently, a binary PPI interaction network map has been generated for *E. coli* using only Y2H and covers approximately 70% of the *E. coli* proteome[92]. In 2008, an interaction network map was generated for *S. cerevisiae* by combining data generated using both binary and co-complex techniques as well as curated from literature[93]. High-throughput Y2H was also recently

applied to investigate interactions between secreted *M. tuberculosis* proteins and those of the host macrophage. EsxH, in complex with EsxG, was found to disrupt the function of the host endosomal sorting complex required for transport. This plays a role in facilitating lysosomal degradation by direct interaction with the human hepatocyte growth factor-regulated tyrosine kinase substrate. These findings provide insight into the mechanisms employed by *M. tuberculosis* to evade the host immune response[94].

A significant amount of *M. tuberculosis* H37Rv PPI data was generated in 2010 when the first global interaction network was constructed for this organism using a high-throughput bacterial two-hybrid technique. Over 8000 interactions were identified and validation using affinity purification revealed a success rate of over 60%. Ninety-four conserved complexes were identified using cross-species analysis between *M. tuberculosis* networks and those of other prokaryotes. This global PPI network also enabled the authors to determine an association between sub-networks and protein secretion pathways. An example of these sub-networks is the interaction between two WhiB-like core regulators, namely WhiB3 and WhiB7, proposed to play a role in regulating persistent infection. Protein function predictions of many hypothetical proteins were made using this PPI data and pathway mapping tools[83].

Many high-throughput PPI studies, predominantly based on a modified Y2H method, have been cross-compared to obtain a more comprehensive and high quality binary interaction network[92,95]. These combined datasets often give a clearer picture of interaction networks by giving higher confidence in interactions identified. These interaction networks, however, do not represent all possible interactions that could be identified using Y2H, partly due to the difficulty in using library screening methods to sample all possible combinations of PPIs[84]. The main disadvantage of high-throughput screens is the high frequency of false positives and negatives[96]. This problem can be resolved by using bioinformatics tools to compare interaction data of the organism of interest and selecting those interactions that occur in multiple datasets[97]. High-throughput PPI data is generally of lower quality than PPI data generated using small scale PPI techniques. It is therefore important that high-throughput PPI data be validated using other low-throughput binary methods to surmount the problems of inaccuracy associated with high-throughput techniques and those inherent to Y2H.

1.3.4. Affinity purification – mass spectrometry

This technique is based on CoIP or affinity purification, as described above, to enrich for protein complexes of interest and subsequent identification of interacting partners by mass spectrometry[98]. High-throughput affinity purification followed by mass spectrometry (AP-MS) has been used on a large scale to identify protein complexes in organisms such as *E. coli*[99,100] and *S. cerevisiae*[101–104]. Although AP-MS allows us to identify proteins that form part of the same complex, it gives little information about the direct binary interactions between proteins within the complex. Binary PPI techniques are therefore still required to confirm interactions identified by AP-MS and together, elucidate protein complex topology[92,95].

AP-MS has vastly increased the possibilities for PPI studies. AP-MS can be used to analyse post-translational modifications, stoichiometry, protein abundance and structural organisation. This technique has been enhanced by various cross-linking techniques, aimed at preserving PPIs in their physiologically relevant state[105]. In 2014, an *in vivo* formaldehyde cross-linking system was developed for the identification of mycobacterial protein complexes using AP-MS[106]. Cross-linking followed by MS provides a new avenue for protein modelling by mapping structural details of physiologically relevant protein complexes. Although this is a fairly new and still challenging area of protein biochemistry, it holds great promise for the advancement of structural biology and the understanding of interaction network topology[107].

1.3.5. PPI databases

Storing high-throughput interaction data in publicly accessible PPI databases is an important factor in interactome studies. Several PPI databases exist that are either specific for one organism or many organisms within a certain phylogenetic group. Bacteriome.org is an example of a PPI database designed specifically for *E. coli*, and currently reports 7613 interactions between 2283 proteins when using a combined dataset[108]. Datasets are often graphically represented, where proteins are represented as nodes and interactions between proteins represented by a line which connects them in a “ball-and-stick” fashion. This technique is easily applied to binary PPI datasets, but can be challenging for AP-MS datasets, where direct interactions within complexes are unknown. AP-MS datasets are then represented using a matrix or spokes model[109].

The majority of data found on PPI databases are curated from published literature, and usually mention the PPI detection method used and the reference to the article that published it. Databases are represented in various ways depending on the detection method used and the preferences of the designers. It is for this reason that the International Molecular Exchange (IMEx) consortium was developed; a platform comprised of a number of publicly available interaction databases that have developed a partnership to generate a non-redundant PPI dataset that adheres to the XML-based proteomics standard, named the proteomics standard initiative – molecular interactions (PSI-MI)[110,111]. A limited number of PPI databases are available for bacteria. One of the participating partners in the IMEx consortium is the IntAct molecular interaction database which is currently the most comprehensive database available, reporting over 294 000 interactions for over 100 organisms, including *Homo sapiens* and various bacterial species[112]. The microbial protein interaction database (MPIDB) provides PPI information for various microbial species, and currently boasts 24 295 experimentally determined interactions between 250 bacterial species, including mycobacteria[113]. MPI-LIT is a database comprised entirely from microbial PPI data curated from the literature and links 940 proteins by 746 binary interactions[97]. Another PPI database is STRING, which combines known and predicted interactions representing physical and functional associations. STRING has recently released its tenth version which covers over 2000 organisms[114]. Other databases to consider are: the Database of Interacting Proteins (DIP)[115], the Molecular INTeraction database (MINT)[116], the Biological General Repository for Interaction Datasets (BioGRID)[117], the Biomolecular Interaction Network Database (BIND)[118] and the Pathosystems Resource Integration Center (PATRIC)[119].

1.4. Concluding remarks

In recent years, high-throughput PPI studies have generated a large amount of data available in publicly accessible databases that have contributed significantly to our understanding of dynamic biological interactions and their effect on cellular physiology. PPI databases for specific organisms play an essential role in interactome-based research. PPI databases are not only important for studying intracellular interactions, but also for studying interactions between different species, such as host-pathogen interactions, which contributes to our understanding of bacterial physiology and pathogenesis at the host-pathogen interface[120]. These interspecies PPIs could lead to the identification of clinically

significant interactions with implications for drug treatments. An example of this is the maraviroc anti-HIV drug which prevents HIV infection of CD4 T-cells by inhibiting the interaction between HIV protein gp120 and chemokine receptor 5[121]. PPIs can be used to design target-specific drug treatments or multi-target drug treatments that are aimed at shortening current drug therapies and offer a new avenue to explore given the burden of drug resistance[122].

The use of high-throughput techniques has given us valuable information on thousands of PPIs stored in various databases. However, a large proportion of these interactions are false and therefore validation of high-throughput PPI data is required to provide a reliable and comprehensive interaction network. Few PPI databases exist, only 3 of which are specific for bacteria, namely MPIDB, MPI-LIT and PATRIC. As useful as these databases are, many of the large datasets are incomplete or unreliable, with poor consensus when cross-comparing PPI datasets between databases. To date, there is no complete and reliable PPI database for mycobacteria and the number of interactions confirmed using low-throughput techniques is even lower. Resolving this problem would be a step in the right direction toward a complete and reliable global interaction network for mycobacteria. This will aid in the understanding of virulence pathways and the mechanisms by which *M. tuberculosis* evades the immune response.

Chapter 2: Research rationale

2.1. Background

Iron-sulphur (Fe-S) clusters are chemically versatile cofactors that are found in all living organisms and in *M. tuberculosis* approximately 50 proteins are predicted to require an Fe-S cluster for their function[123]. These proteins are involved in various biological processes including DNA repair, RNA modification, redox signalling and central metabolism[123,124]. Multicomponent systems are required for the *in vivo* assembly of Fe-S clusters and three highly conserved systems have been identified in both prokaryotes and eukaryotes, namely the NIF (nitrogen fixation), ISC (Fe-S cluster) and SUF (sulphur mobilization) systems[123–125].

In *M. tuberculosis*, a single operon (*Rv1460-Rv1461-Rv1462-Rv1463-csd-Rv1465-Rv1466*) displays homology to genes in the SUF system[33]. With the exception of *Rv1460*, all the genes in the *suf* operon were predicted to be essential *in vitro* by forward genetic screens, suggesting an important role of Fe-S cluster assembly in the survival of *M. tuberculosis*[34,126]. Additionally, *Rv1460* was predicted to be essential *in vivo* and shown to be induced in the sputum of TB patients, suggesting a significant contribution of Fe-S cluster assembly to *M. tuberculosis* pathogenicity[35]. Fe-S cluster containing proteins can be inactivated in the presence of reactive oxygen and nitrogen species. Exposure to oxidative stress causes Fe-S clusters to become oxidised, resulting in the release of iron, which in turn causes a further increase in oxidative stress[124]. Due to the fact that *M. tuberculosis* is exposed to oxidative and nitrosative stressors within the host, Fe-S cluster assembly is most likely tightly regulated, as these harsh conditions would be toxic to Fe-S cluster containing proteins[124]. In addition, anti-TB drugs such as clofazamine and isoniazid have previously been shown to generate reactive oxygen species, suggesting a crucial role for Fe-S cluster assembly in *M. tuberculosis* drug susceptibility[127]. The first step in the synthesis of Fe-S clusters requires a cysteine desulphurase, which catalyses the transfer of sulphur from L-cysteine to a scaffold protein. Fe-S clusters are then assembled on the scaffold and transferred either directly or via carrier proteins to the apoprotein[123,124].

Rv1460 encodes a probable transcriptional regulator which displays partial homology to the *SufR* gene found in *Synechocystis* sp. strain PCC 6803[33,82]. SufR contains an Fe-S

cluster and functions as a repressor of the *suf* operon in cyanobacteria[128,129]. Rv1460 is hypothesised to function similarly. *Rv1461* encodes a SufB homolog which is highly conserved between many organisms and is thought to be the sulphur acceptor in the first step of Fe-S cluster assembly as interaction studies have shown that SufB interacts weakly with Csd[33,82]. Rv1462 is a conserved hypothetical which is orthologous to the *E. coli* SufD, while *Rv1463* encodes a probable conserved ATP-binding protein ABC transporter and is orthologous to the *E. coli* SufC. SufC interacts strongly with SufB and SufD, and these three proteins are likely to form the scaffold on which the Fe-S cluster is assembled. *Rv1464* (*csd*) encodes a NifS-like protein and is a cysteine desulphurase orthologous to the *E. coli* SufS. *Rv1465* encodes a NifU-like protein of unknown function which is an ortholog of the *E. coli* IscU (a scaffold protein). *Rv1466* is a conserved hypothetical that does not display homology to any of the previously identified *suf* genes. The *suf* operon of *M. tuberculosis* is atypical since no SufA and SufE homologues have been identified[33,82]. The iron donor remains unknown.

The *isc* operon of *M. tuberculosis* lacks several genes required for the function of the ISC system in other organisms, however IscS (NifS homolog) is conserved[38]. Singh *et al*. showed that IscS (Rv3025c) alone is sufficient for the assembly of the Fe-S cluster in the WhiB3 transcriptional regulator, *in vitro*[130]. Although IscS does not interact with any of the proteins in the SUF system, it has been shown to interact with Fe-S cluster containing proteins. Deletion of *iscS* in *M. tuberculosis* resulted in impaired activity of enzymes containing Fe-S clusters, suggesting that IscS is involved in Fe-S cluster assembly[38].

2.2. Problem statement and research question

Iron-sulphur (Fe-S) cluster biogenesis is a poorly understood area of *M. tuberculosis* physiology. Sequence analysis of the genome of *M. tuberculosis* revealed a single operon (*Rv1460-Rv1461-Rv1462-Rv1463-csd-Rv1465-Rv1466*) which displays homology to Fe-S cluster assembly systems[33]. It is currently unclear if proteins outside of this operon are involved in Fe-S cluster assembly.

This study aims to use a proteomics-based mass spectrometry approach to identify proteins which interact with the Fe-S cluster biogenesis machinery in mycobacteria. This will

contribute to our understanding of this process in mycobacteria, and has the potential to identify novel targets for the development of new anti-TB drugs.

2.3. Approach

A proteomic-based mass spectrometry approach was used together with BN-PAGE and affinity purification to identify novel proteins involved in Fe-S cluster assembly.

Mycobacterium bovis BCG is a non-pathogenic, slow-growing mycobacterium which is closely related to *M. tuberculosis*. BCG is a suitable model organism as more than 99% amino acid sequence identity is observed for the proteins encoded by the *suf* operon. This organism was utilised for the optimisation of the BN-PAGE and 2D BN/SDS PAGE methodology since BL3 conditions are not required. However, due to the fact that this organism is so closely related to *M. tuberculosis*, the methodology developed during this study would be applicable to the pathogenic organisms, should the appropriate facilities be available.

Mycobacterium smegmatis is a non-pathogenic fast-growing mycobacterium which is closely related to *M. tuberculosis*, with an amino acid sequence identity of 83% for proteins encoded by the *suf* operon. *M. smegmatis* was used for all protein expression optimisation and affinity purification experiments due to the time constraints of this study.

2.4. Research hypothesis, aims and objectives

Hypothesis: Proteins outside of the *suf* operon are involved in Fe-S cluster assembly.

Aim: To identify novel proteins which interact with the *M. tuberculosis* Fe-S cluster biogenesis machinery

Objectives:

1. To develop the methodology to identify cytoplasmic protein complexes in mycobacteria using blue native PAGE (BN-PAGE) and mass spectrometry (MS)

2. To identify novel proteins which interact with the Fe-S cluster biogenesis machinery in *M. tuberculosis*

3. To validate the novel interactions identified using other protein-protein interaction techniques

Chapter 3: Methods and materials

3.1. Standard culture conditions

Escherichia coli strains were grown on Luria Bertani agar solid media overnight at 37 ˚C and then cultured to mid-logarithmic phase in Luria Bertani liquid media overnight at 37˚C with shaking at 200 rpm. Strains were stored as freezer stocks prepared by mixing 1 ml culture with glycerol to a final concentration of 12%. *Mycobacterium smegmatis* strains were grown on LA solid media for 3-5 days at 37˚C and then cultured in 7H9 liquid media in a 5 ml starter culture overnight at 37 ˚C with shaking at 200 rpm. *M. smegmatis* was then sub-cultured to a starting OD_{600} of 0.05 and grown to mid-logarithmic phase. *Mycobacterium bovis* BCG strains were grown on 7H10 solid media for 3-5 weeks at 37 ˚C and then cultured in 7H9 liquid media in a 5 ml starter culture for 5-7 days at 37 ˚C without shaking. *M. bovis* BCG was then sub-cultured to a starting OD_{600} of 0.05 and grown to mid-logarithmic phase. Mycobacterial strains were stored as freezer stocks of 1 ml of mid-logarithmic phase culture at -80 ˚C. Standard culturing conditions are used throughout unless otherwise stated.

Table 3.1. Plasmids used in this study

Plasmid	Description	Source
pET-28a(+)	Plasmid designed for expression of recombinant proteins in *E. coli*, N- and C-terminal His-tag, T7 promoter, Thrombin cleavage site, Kan^r	EMD Millipore
pGEX-6P-1	Plasmid designed for expression of recombinant proteins in *E. coli*, N-terminal GST-tag, tac promoter, PreScission protease cleavage site, Amp^r	GE Healthcare Life Sciences
pMAL-c5X	Plasmid designed for expression of recombinant proteins in *E. coli*, N-terminal MBP-tag, tac promoter, Factor Xa protease cleavage site, Amp^r	NEB
pET28a(+)-*csd*	Derivative of pET28a(+) carrying *M. tuberculosis Rv1464 (csd)*, Pmyc1tetO promoter, Hyg^r	This work
pGEX-6P-1-*csd*	Derivative of pGEX-6P-1 carrying *M. tuberculosis Rv1464 (csd)*, tac promoter, Amp^r	This work
pMAL-c5X-*csd*	Derivative of pMAL-c5X carrying *M. tuberculosis Rv1464 (csd)*, tac promoter, Amp^r	This work
pJET 1.2	Cloning vector for cloning of blunt-end PCR products and expression in *E. coli*, T7 promoter, Amp^r	Thermo Scientific
pSE100	Mycobacterial tetracycline-inducible expression vector, derivative of pMS2, Pmyc1tetO promoter, Hyg^r	Sabine Ehrt[131]

pSE100-rbs-*csd*	Derivative of pSE100 carrying *M. tuberculosis Rv1464 (csd)*, modified ribosome binding site, His-tag and thrombin cleavage site, Pmyc1tetO promoter, Hyg^r	This work
pSE100-rbs-*sufc*	Derivative of pSE100 carrying *M. tuberculosis Rv1463* (*sufc*), modified ribosome binding site, His-tag and thrombin cleavage site, Pmyc1tetO promoter, Hyg^r	This work
pSE100-rbs-*mcherry*	Derivative of pSE100 carrying mCherry, modified ribosome binding site, His-tag and thrombin cleavage site, Pmyc1tetO promoter, Hyg^r	This work
pTEK-4S-0X	Vector expressing reverse TetR for use in mycobacteria, derivative of pMS2, psmyc promoter, Kan^r	Sabine Ehrt[131]

*All plasmids used in this study were initially propagated in *E. coli* DH5-alpha.

Table 3.2. Strains and plasmids used in this study

Name	Description	Source
M. bovis		
BCG Pasteur 1173P2	Pasteur 1173P2; avirulent laboratory strain	Laboratory stock
M. smegmatis		
mc^2 155	*ept-1* (efficient plasmid transformation) mutant of mc^{2}6	Laboratory stock
mc^2 155 *attB*::pSE100-rbs-*csd* (pTEK-4S-0X)	Derivative of mc^2 155 carrying *M. tuberculosis Rv1464 (csd)* under the control of Pmyc1tetO promotor, integrated at the *attB* locus, Hyg^r; **	This work
mc^2 155 *attB*::pSE100-rbs-*sufc* (pTEK-4S-0X)	Derivative of mc^2 155 carrying *M. tuberculosis Rv1463* (*sufc*) under the control of Pmyc1tetO promotor, integrated at the *attB* locus, Hyg^r; **	This work
mc^2 155 *attB*::pSE100-rbs-*mcherry* (pTEK-4S-0X)	Derivative of mc^2 155 carrying mCherry, a codon optimised derivative of DsRed isolated from *Discosoma sp*., under the control of Pmyc1tetO promotor, integrated at the *attB* locus, Hyg^r; **	This work
groEL1ΔC	Derivative of mc^2 155 carrying a deletion of the histidine-rich C-terminus in GroEL1	Elke Noens[132]
groEL1ΔC attB::pSE100-rbs-*csd* (pTEK-4S-0X)	Derivative of *groEL1ΔC* carrying *M. tuberculosis Rv1464 (csd)* under the control of Pmyc1tetO promotor, integrated at the *attB* locus, Hyg^r; **	This work
groEL1ΔC sattB::pSE100-rbs-*sufc* (pTEK-4S-0X)	Derivative of *groEL1ΔC* carrying *M. tuberculosis Rv1463* (*sufc*) under the control of Pmyc1tetO promotor, integrated at the *attB* locus, Hyg^r; **	This work
groEL1ΔC attB::pSE100-rbs-*mcherry* (pTEK-4S-0X)	Derivative of *groEL1ΔC* carrying mCherry, a codon optimised derivative of DsRed isolated from *Discosoma sp*., under the control of Pmyc1tetO promotor, integrated at the *attB* locus, Hyg^r; **	This work

E. coli		
DH5-Alpha	Avirulent laboratory cloning strain (high efficiency transformation)	Thermo Fisher Scientific
ArcticExpress	Avirulent strain optimised for protein expression and enhanced solubility of heterologous proteins; derivative of BL21-Gold	Agilent Technologies
ArcticExpress (pET28a(+)-*csd*)	Derivative of ArcticExpress carrying a plasmid expressing a 6xHis-tag fusion with *M. tuberculosis Rv1464 (csd)* from a T7 promotor, Kan^r	This work
ArcticExpress (pGEX-6P-1-*csd*)	Derivative of ArcticExpress carrying a plasmid expressing a GST-tag fusion with *M. tuberculosis Rv1464 (csd)* from a tac promotor, Amp^r	This work
ArcticExpress (pMAL-c5X-*csd*)	Derivative of ArcticExpress carrying a plasmid expressing a MBP-tag fusion with *M. tuberculosis Rv1464 (csd)* from a tac promotor, Amp^r	This work
Rosetta-gami	Avitulent strain optimised for protein expression of proteins containing codons rarely used in *E. coli*; derivative of Origami	Merck Millipore
Rosetta-gami (pET28a(+)-*csd*)	Derivative of Rosetta-gami carrying a plasmid expressing a 6xHis-tag fusion with *M. tuberculosis Rv1464 (csd)* from a T7 promotor, Kan^r	This work
Rosetta-gami (pGEX-6P-1-*csd*)	Derivative of Rosetta-gami carrying a plasmid expressing a GST-tag fusion with *M. tuberculosis Rv1464 (csd)* from a tac promotor, Amp^r	This work
Rosetta-gami (pMAL-c5X-*csd*)	Derivative Rosetta-gami carrying a plasmid expressing a MBP-tag fusion with *M. tuberculosis Rv1464 (csd)* from a tac promotor, Amp^r	This work

**Carries plasmid expressing reverse TetR from the psmyc promoter, Kan^r

Table 3.3. PCR and sequencing primers

Gene / plasmid	Primer name	Sequence (5'-3')	Amplicon (bp)	Description	Primer binding site
Rv1464 (csd)	pET28-pGEX-*csd* F	GAATTCCAT**ATG** ACGGCCTCGGTG AACTCG	1277	*csd* forward primer	5' end of *csd*
Rv1464 (csd)	pET28-*csd* R	GACAAGCTTCAA CG**TCA**CGCTCTT CCAAA		*csd* reverse primer	5 bp downstream of *csd* 3' end
Rv1464 (csd)	pET28-pGEX-*csd* F	GAATTCCAT**ATG** ACGGCCTCGGTG AACTCG	1277	*csd* forward primer	5' end of *csd*

Rv1464 (csd)	pGEX-*csd* R	GACCTCGAGCAACG**TCA**CGCTCTTCCAAA		*csd* reverse primer	5 bp downstream of *csd* 3' end
Rv1464 (csd)	MBP-*csd* F	TCGACGATATC**ATG**ACGGCCTCGGTGAACTC	1264	*csd* forward primer	5' end of *csd*
Rv1464 (csd)	MBP-*csd* R	TCGACGAATTC**TCA**CGCTCTTCCAAAGAAAT		*csd* reverse primer	3' end of *csd*
pET28a(+)	T7 promoter	TAATACGACTCACTATAGGG	N/A	pET28a(+) sequencing forward primer	99 bp upstream of the N-terminal His-tag
pET28a(+)	T7 terminator	GCTAGTTATTGCTCAGCGG	N/A	pET28a(+) sequencing reverse primer	91 bp downstream of the *csd* 3' end
pGEX-6P-1	pGEX 5'	GGGCTGGCAAGCCACGTTTGGTG	N/A	pGEX-6P-1 forward sequencing primer	91 bp upstream of the *csd* 5' end
pGEX-6P-1	pGEX 3'	CCGGGAGCTGCATGTGTCAGAGG	N/A	pGEX-6P-1 reverse sequencing primer	103 bp downstream of the *csd* 3' end
Rv1464 (csd)	*csd* internal F	ACGACGTGGGTCAGGTACTC	N/A	*csd* forward internal primer	194 bp upstream of the *csd* 3' end
Rv1464 (csd)	*csd* internal R	GGTCGCCTCCTCCATCAG	N/A	*csd* reverse internal primer	198 bp downstream of the *csd* 5' end
pJET 1.2	pJET 1.2 F	CGACTCACTATAGGGAGAGCGGC	N/A	pJET 1.2 forward sequencing primer	62 bp upstream of the gene of interest
pJET 1.2	pJET 1.2 R	AAGAACATCGATTTTCCATGGCAG	N/A	pJET 1.2 reverse sequencing	33 bp downstream of the gene of interest
pSE100	pSE100 F	CCGAAATGAGCACGATCC	N/A	pSE100 forward sequencing primer	135 bp upstream of the gene of interest
pSE100	pSE100 R	GACCTCTAGGGTTCCCCAATTA	N/A	pSE100 reverse sequencing primer	19 bp downstream of the gene of interest

Rv1463 (*sufc*)	*sufc* F	GACCAT**ATG**ACC ATTTTGGAAATT AAGGA	816	*sufc* forward primer	5' end of *sufc*
Rv1463 (*sufc*)	*sufc* R	GACAAGCTT**TCA** GGCTCCGGTTGG CGCGG		*sufc* reverse primer	3' end of *sufc*

*Restriction enzyme sites are underlined

*Start and stop codons are in bold

Table 3.4. Growth media and buffer composition

Growth media	**Composition**
Luria Bertani agar (LA)	0.5% Yeast extract, 1% Tryptone, 0.5% Sodium chloride (NaCl), 1.5% agar
Luria Bertani broth (LB)	0.5% Yeast extract, 1% Tryptone, 0.5% NaCl
Middlebrook 7H9 liquid media	0.47% 7H9 powder, 0.2% glucose, 0.2% glycerol, 0.05% Tween-80
Middlebrook 7H10 solid media	1.9% 7H10 powder, 10% oleic albumin dextrose and catalase (OADC), 0.5% glycerol
Sauton's minimal media	2% glycerol, 0.4% L-asparagine, 0.2% glucose, 0.2% citric acid, 0.05% monopotassium phosphate, 0.05% magnesium, 0.05% Tween-80, 0.005% ferric ammonium citrate, 0.00001% zinc sulphate, pH 7.0
Buffers	
Lysis buffer 1	50 mM Tris-HCL pH 7.4, 150 mM NaCl, 0.05% Triton X-100, c0mplete protease inhibitor cocktail (Roche)
Lysis buffer 2	50 mM Tris-HCL pH 7.4, 10 mM $MgCl_2$, c0mplete protease inhibitor cocktail, 20 U/ml DNase I
Lysis buffer 3	50 mM Tris-HCL pH 7.6, 100 mM NaCl, 2 mM $MgCl_2$, 0.5% Triton X-100, c0mplete protease inhibitor cocktail, 25 U/ml Benzoase, 1 mM Dithiotreitol (DTT)
Laemmli 2x sample reducing buffer	6% Sodium dodecyl sulfate (SDS), 20% glycerol, 10% β-Mercaptoethanol (BME), 0.0375% Bromophenol blue, 340 mM Tris pH 6.8, 10 mM Ethylenediaminetetraacetic acid (EDTA)
Laemmli running buffer	25 mM Tris, 192 mM Glycine, 0.1% SDS
Transfer buffer 1	25 mM Tris, 192 mM Glycine, 20% Methanol (MeOH)
TBS-T wash buffer	20 mM Tris pH 7.6, 137 mM NaCl, 0.1% Tween-20
Blocking buffer	TBS-T wash buffer + 10% fat free milk powder
Tris-acetate-edetic acid (TAE)	40 mM Tris pH 7.6, 1 mM EDTA pH 8.0, 20 mM acetic acid

3.2. Preparation of competent cells

3.2.1. Chemically competent *E. coli*

E. coli (DH5-Alpha and ArcticExpress) was cultured in 5 ml of double-strength LB supplemented with 0.2% glucose liquid media overnight with shaking at 200 rpm. A concentration of 20 µg/ml gentamycin was added for ArcticExpress. ArcticExpress was cultured at 30˚C, while DH5-Alpha was cultured at 37 ˚C. The culture was diluted 1:100 in the same media and cultured to an OD_{600} of 0.3-0.4. Cells were harvested by centrifugation in a GSA rotor at 4500 x *g* at 4 ˚C. All subsequent steps were performed on ice. Cells were resuspended in 100 ml of ice-cold 100 mM $CaCl_2$ containing 15% glycerol and incubated on ice for 20-30 min. Cells were then harvested by centrifugation at 4500 x *g* for 10 min at 4˚C. Cells were resuspended in 10 ml of an ice-cold 100 mM $CaCl_2$ solution containing 15% glycerol. Aliquots of 100 µl were made, flash frozen in liquid nitrogen and stored at -80 ˚C until required.

3.2.2. Electrocompetent *E. coli*

E. coli Rosetta-gami was cultured in 50 ml LB media containing 34 µg/ml chloramphenicol, 50 µg/ml streptomycin, and 12.5 µg/ml tetracycline. Eight 500 ml flasks each containing 250 ml LB media (34 µg/ml chloramphenicol, 50 µg/ml streptomycin, and 12.5 µg/ml tetracycline) were incubated overnight at 37 ˚C with shaking at 200 rpm. Five millilitres of *E. coli* culture was inoculated into each of the eight flasks and incubated at 37˚C with shaking at 200 rpm for 3-4 hours. Once an OD_{600} of 0.7 had been reached, cells were harvested by centrifugation in 6x250 ml centrifuge tubes in a GSA rotor for 10 min at 4500 x *g* at 4 ˚C. Cells were washed three times (50 ml, 20 ml and 10 ml washes) in ice-cold 10% glycerol and pooled together in a 50 ml Falcon tube at the final wash step. The pellet was resuspended in 2 ml 10% ice cold glycerol per 1 L of original culture. Aliquots were made of 100 µl each and flash-frozen in liquid nitrogen and stored at -80˚C.

3.2.3. Electrocompetent *M. smegmatis*

M. smegmatis was cultured under standard conditions in 50 ml 7H9 liquid media until an OD_{600} of ~1 was reached. Cells were harvested by centrifugation at 4500 x *g* for 10 min at 4 ˚C. Cells were resuspended in 50ml 10% ice-cold glycerol containing 0.05% Tween-80 and harvested by centrifugation as described above. This wash was repeated with 20 ml

and 10 ml, and cells were finally resuspended in ~2 ml 10% ice-cold glycerol with 0.05% Tween-80.

3.2.4. Electrocompetent *M. bovis* BCG

M. bovis BCG was cultured under standard conditions in 50 ml 7H9 liquid media until an OD_{600} of ~1 was reached. Cells were harvested by centrifugation at 4 500 x *g* for 10 min at room temperature (RT). Cells were washed 3 times in equal volumes of 10% glycerol with 0.05% Tween-80, pre-heated to 37 ˚C. Cells were resuspended in ~2 ml 10% glycerol with 0.05% Tween-80.

3.3. Cloning and molecular biology methods

PCR reactions for gene amplification were performed using ThermoFisher Scientific Phusion High-Fidelity DNA Polymerase according to manufacturer's instructions. Colony PCR was performed using the KAPA Biosystems KAPA2G Robust HotStart ReadyMix PCR Kit according to manufacturer's instructions.

DNA was extracted using the NucleoBond® PC 100 Nucleic Acid and Protein Purification kit (Separations) or the Wizard® plus SV MiniPreps DNA Purification kit (Promega) according to manufacturer's instructions. For restriction enzyme digests, 10 units of enzyme were used to digest 1 µg of plasmid DNA in 3 hours at 37 ˚C in a 20 µl reaction. DNA electrophoresis gels were prepared using 0.8% agarose in TAE buffer containing 0.005% 10 mg/ml Ethidium bromide. Gel slices were purified using the Wizard® SV Gel and PCR Cleanup Kit (Promega) and the standard protocol was followed according to manufacturer's instructions. DNA was quantified using the Nanodrop® ND1000 Spectrophotometer (Thermo Scientific). Ligation reactions were performed using a 1:3 ratio of vector to insert (0.05 pmol ends : 0.15 pmol ends) in a 15 µl reaction with 3 units of T4 ligase (Promega) at 4˚C overnight.

3.4. Transformation

3.4.1. Chemically competent *E. coli*

E. coli competent cells were thawed on ice. A volume of 100 µl of competent cells were added to the entire ligation mix and incubated on ice for 10 minutes. The transformation mixture was then incubated at 37˚C for 10 min and immediately placed back

on ice. A volume of 1 ml of LB liquid media was added to the transformation mixture and incubated for 1 hour at 37 ˚C for cells to recover. The transformation mixture was then concentrated to 100 µl and plated on LA solid media, followed by incubation at 37 ˚C until visible growth was observed.

3.4.2. Electro-competent cells

3.4.2.1. *E. coli*

E. coli competent cells were thawed on ice. A volume of 40 µl of competent cells were added to 2 µl of the ligation mixture or 20 ng of circular plasmid DNA and incubated on ice for 5 min. The transformation mixture was transferred to an electroporation cuvette and electroporation performed using the Bio Rad MicroPulser™ (Setting: Bacteria; Program: Ec2; 2.5 kV; 25 µF; 200 Ω). A volume of 400 µl of LB media was added to the cuvette and transferred to a 1.5 ml Eppendorf tube and incubated for 1 hour at 37˚C for cells to recover. A volume of 100 µl of the transformation mixture was then plated on LA solid media, followed by incubation at 37˚C until visible growth was observed.

3.4.2.2. *M. smegmatis* and *M. bovis* BCG

A volume of 400 µl of competent cells were added to 500-1000 ng of plasmid DNA and incubated on ice for *M. smegmatis* and 37 ˚C for *M. bovis* BCG for 5 min. The transformation mixture was then transferred to an electroporation cuvette and electroporations were performed with the following settings: 2.5 kV, 25 µF and 1000 Ω. After each electroporation, 800 µl of 7H9 liquid media was added to cuvette and then transferred to a 1.5 ml Eppendorf tube and incubated at 37 ˚C cells to recover. *M. smegmatis* cells were incubated for 3 hours and then plated on LA solid media and *M. bovis* BCG cells were incubated for 24 hours and plated on 7H10 solid media.

3.5. Sanger sequencing

All DNA sequencing was performed at the Central Analytical Facility (CAF) at Stellenbosch University and analysed using Sequencher DNA sequence analysis software[133].

3.6. *M. smegmatis* and *M. bovis* BCG growth curves

M. smegmatis and *M. bovis* BCG strains were cultured at a starting OD_{600} of 0.05 under standard conditions and OD_{600} readings recorded every 3 hours for *M. smegmatis* and every 48 hours for *M. bovis* BCG. All growth curves were done in 3 biological replicates.

3.7. Protein over-expression in *E. coli*

Plasmids for the expression of recombinant Csd were transformed into *E. coli* ArcticExpress and Rosetta gami strains. Single colonies were picked and inoculated into 5 ml LB media containing 20 µg/ml gentamycin for ArcticExpress and 34 µg/ml chloramphenicol, 50 µg/ml streptomycin and 12.5 µg/ml tetracycline for Rosetta gami, and cultured under standard conditions. The starter culture was sub-cultured in LB media containing the above mentioned antibiotics and incubated at 30 ˚C for ArcticExpress and 37 ˚C for Rosetta-gami for 2-4 hours to obtain an OD_{600} of 0.4. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 0.4 mM and incubated overnight at 13 ˚C for ArcticExpress and standard conditions for Rosetta gami. Cells were harvested after 24 hours by centrifugation at 4500 x *g* for 10 min at 4 ˚C. Pelleted cells were stored at -80 ˚C overnight or until use. Uninduced and empty vector strains were included as controls.

3.8. Construction of 6xHis-tag fusion proteins

csd along with the pET-28a(+) ribosome binding site (rbs), 6xHis-tag and thrombin cleavage site was excised from pET28a(+)-*csd* with XbaI and HindIII and subsequently ligated into pSE100 at the same restriction sites to generate pSE100-rbs-*csd*. The pSE100-rbs-*csd* vector was then digested with NdeI and HindIII to remove *csd* only. *sufc* and *mcherry* were then ligated into the modified pSE100 vector to generate pSE100-rbs-*sufc* and pSE100-rbs-*mcherry*, respectively.

3.9. Optimisation of 6xHis-tag fusion protein expression

M. smegmatis mc^2 155 derivatives, *M. smegmatis* mc^2 155 *attB*::pSE100-rbs-*csd* (pTEK-4S-0X), *M. smegmatis* mc^2 155 *attB*::pSE100-rbs-*sufc* (pTEK-4S-0X) and *M. smegmatis* mc^2 155 *attB*::pSE100-rbs-*mcherry* (pTEK-4S-0X), were cultured under standard conditions. Starter cultures were grown in the presence of 10 ng/ml anhydrotetracycline (atc). Atc is used to control gene expression in the tetracycline-responsive gene expression system, used here. Starter cultures were subsequently washed twice with culture media and subcultured to an

OD_{600} of 0.05 in media containing either 0 ng/ml atc or 10 ng/ml atc. Aliquots were removed from each culture at 0, 16, 24 and 36 hours. Cells were harvested by centrifugation 4500 x *g* and protein extractions performed as described below. Both the soluble and insoluble protein fractions were collected and analysed by SDS-PAGE using the Bio-Rad 4-15% Mini-PROTEAN® TGX Stain-Free™ Gels. Proteins were transferred to a PVDF membrane. A primary anti-6xHis-tag antibody (Abcam), diluted 1:1000, and a secondary HRP-conjugated goat anti-rabbit antibody, diluted 1:10000, were used for western blot immunodetection. Gels, membranes and chemiluminescent detection was done using the Bio-Rad ChemiDoc™ MP system. The Bio-Rad Image Lab™ software was used to normalise the intensity of the band detected on the Western blot to the total protein loaded in each lane.

3.10. Affinity purification using MagneHis™ beads

Affinity purification of 6xHis-tagged proteins was performed using paramagnetic pre-charged nickel particles (MagneHis™ Ni-Particles) under native conditions. *M. smegmatis* strains were cultured under standard conditions in the absence of atc for 16 hours. Cells were harvested by centrifugation at 4500 x *g*, washed twice with PBS and stored at -80 °C. Cell pellets from 5 ml culture were resuspended in 1 ml of lysis buffer 3 and cells were lysed as described above. MagneHis™ Ni-Particles were added to 1 ml of protein lysate and protocol followed according to manufacturer's instructions. The purification procedure was optimised by testing different volumes of MagneHis™ Ni-Particles per 1 ml protein lysate as well as the influence of high (500 mM) and low (100 mM) NaCl concentrations.

3.11. H_2O_2 and NO exposure of *M. bovis* BCG

M. bovis BCG was cultured under standard conditions until an OD_{600} of ~1. For H_2O_2 exposure, 5 mM or 10 mM H_2O_2 was added to the culture, and aliquots removed at 40 min, 1 hour and 3 hours post addition. Cells were then harvested and proteins extracted as described below. For NO exposure, 0.5 mM or 1 mM diethylenetriamine/nitric oxide adduct (DETA/NO) was added to the culture and aliquots removed at 40 min and 4 hours. Cells were then harvested and proteins extracted as described below.

3.12. Bacterial lysis and protein extraction

Cells were cultured under standard conditions unless otherwise stated. Cells were harvested by centrifugation at 4500 x *g* and washed twice in equal volumes of phosphate buffered saline (PBS). Cell pellets were stored at -80°C overnight or until use.

3.12.1. Sonication

M. bovis BCG, *M. smegmatis* or *E. coli* cell pellets were resuspended in 1 ml lysis buffer and sonicated 3 times at 50 Hz for 20 sec with 2 min incubations on ice between each sonication step. The lysate was then clarified by centrifugation in a microfuge at maximum speed for 30 min at 4°C.

3.12.2. Bead-beating

M. bovis BCG cell pellets were resuspended in 1 ml lysis buffer and transferred to 2 ml cryogenic tubes with O-rings containing glass beads. Cells were mechanically lysed by bead-beating 6 times for 20 sec each at a speed of 4.0 $m.s^{-1}$, with 1 min intervals for cooling on ice. Beat-beating was performed using a Ribolyser (Bio101 Savant, Vista, CA). The lysate was clarified by centrifugation in a microfuge at maximum speed for 30 min at 4°C.

3.13. Buffer exchange

Protein samples were dialysed using the ThermoFisher Scientific Slide-A-Lyzer™ MINI Dialysis Devices with a molecular weight cut-off of 3.5 kDa. A volume of 100 µl of protein samples was dialysed in 1 ml of buffer according to manufacturer's instructions.

3.14. Protein concentration determination

Protein concentrations were determined using the *RC DC*™ Protein Assay (Bio-Rad) according to manufacturer's instructions. Bovine serum albumin was use as the protein standard and standard curve concentrations ranged from 0 – 2 mg/ml. Dilutions were made where needed to ensure concentrations were within the range of the standard curve. The *RC DC*™ Protein Assay is a modified colorimetric Lowry assay compatible with reducing reagents and detergents.

3.15. SDS-PAGE

NuPAGE® Bis-Tris 4-12% Mini Gels were used to run all SDS-PAGE gels, unless otherwise indicated, using the *XCell SureLock*® Mini-Cell electrophoresis system according to

manufacturer's instructions. NuPAGE 2-(N-morpholino)ethanesulfonic acid (MES) buffer was used and NuPAGE Lithium Dodecyl Sulfate (LDS) sample buffer and NuPAGE sample reducing agent were added to samples to a final concentration of 1x and heated for 10 min at 70˚C. Gels were run at a constant 200 V. Prior to loading, the protein concentration of each sample was determined and the volume adjusted to ensure that the same amount of protein was loaded in each lane (~25 µg).

Bio-Rad 4-15% Mini-PROTEAN® TGX Stain-Free™ Gels were used for protein expression optimisation. Stain-free gels contain a trihalo compound that interacts with the tryptophan content in proteins and allows for fluorescent detection of whole protein lysates using the Bio-Rad ChemiDoc™ MP system. Gel images were analysed using Bio-Rad Image Lab™ software. Gels were run with Laemmli sample running buffer and samples were mixed with an equal volume of Laemmli 2x sample reducing buffer prior to loading and heated for 10 min at 70˚C. Gels were run at a constant 200 V.

Bio-rad gels with an acrylamide/bisacrylamide concentration of 12% were made according to manufacturer's instructions and used for the analysis of affinity purified protein samples. Gels were run with Laemmli sample running buffer and samples were mixed with an equal volume of Laemmli 2x sample reducing buffer prior to loading and heated for 10 min at 70˚C. Gels were run at a constant 200 V.

3.16. BN-PAGE and 2D BN/SDS-PAGE

BN-PAGE was performed using the NativePAGE™ Novex® 4–16% Bis-Tris Gels using the *XCell SureLock®* Mini-Cell electrophoresis system according to manufacturer's instructions. Prior to loading, the protein concentration of each sample was determined and the volume adjusted to ensure that the same amount of protein (~25 µg) was loaded in each lane. The NativePAGE™ Bis-Tris non-denaturing buffer system was used and samples were mixed with NativePAGE™ Sample Buffer to a 1x final concentration. Gels were run at 4˚C for 1 hour at 150 V and then for another hour at 250 V. NativeMark™ Unstained Protein Standard was used as a molecular weight marker.

NuPAGE® 4-16% Bis-Tris Gel with a 6.5 cm 2D-well was used for 2D BN/SDS-PAGE according to manufacturer's instructions. The NuPAGE MES buffer system was used. Gels were run at a constant 200 V.

3.17. Western blot

Unless otherwise stated, a primary anti-6xHis-tag antibody (Abcam), diluted 1:1000, and a secondary HRP-conjugated goat anti-rabbit antibody, diluted 1:10000, were used for Western blot immunodetection.

NuPAGE SDS-PAGE gels were transferred to the Amersham Hybond ECL nitrocellulose blotting membrane with a pore size of 0.45 µm using the Bio-Rad Trans-Blot Turbo Transfer System according to manufacturer's instructions. Proteins were transferred using transfer buffer 1 at 100 V at 4 °C. The membrane was then blocked overnight with blocking buffer. The membrane was washed three times for 10 min with TBS-T wash buffer and then incubated with the primary antibody for 1 hour. TBS-T washes were repeated and the membrane incubated with the secondary antibody for 1 hour, followed by another 3 TBS-T washes. The membrane was then incubated for 5 min with Amersham ECL Western Blotting Detection Reagent according to manufacturer's instructions. Chemiluminescent protein banding patterns were detected and analysed using the Thermo Scientific CL-XPosure Film which was developed using the Amersham pharmacia biotech Hyperprocessor.

Bio-Rad Mini-PROTEAN® TGX Stain-Free™ gels were transferred to the Amersham Hybond P PVDF blotting membrane with a pore size of 0.45 µm using the Bio-Rad Trans-Blot Turbo Transfer System according to manufacturer's instructions. All further steps were performed as stated above. Chemiluminescent protein banding patterns were detected and analysed using the Bio-Rad ChemiDoc™ MP system. Total protein lysates were used for normalisation. Transfer and blot images were analysed using Bio-Rad Image Lab™ software.

BN-PAGE gels were transferred to the Amersham Hybond P PVDF blotting membrane with a pore size of 0.45 µm using the Bio-Rad Trans-Blot Turbo Transfer System according to manufacturer's instructions. NuPAGE transfer buffer was used and chemiluminescent protein banding patterns were detected and analysed using the Thermo Scientific CL-XPosure Film which was developed using the Amersham pharmacia biotech Hyperprocessor.

3.18. Mass spectrometry

LC-MS/MS analysis of MagneHis purifications was performed at the Centre for Proteomic and Genomic Research (CPGR). Four 1 ml aliquots of each protein lysate was mixed with 60 µl of MagneHis beads, and the purification performed as described above.

The complexes eluted from these purifications were subsequently pooled together and concentrated using the Millipore Amicon® Ultra 0.5 ml Centrifugal Filters with a molecular weight cut-off of 3 kDa. Approximately 25 µg of sample was loaded per lane and run ~1 cm into a NuPAGE® Bis-Tris 4-12% Mini Gel. The entire sample in each lane was excised and the gel piece cut into 1 mm x 1 mm cubes and stored in HPLC-grade water (Sigma-Aldrich) in 1.5 ml Eppendorf tubes. Three biological replicates were prepared for each sample. Samples were processed according to the standard operating procedures for in-gel digestion at CPGR. In short, proteins were reduced in 2 mM tris-carboxyethyl phosphine and 25 mM ammonium bicarbonate and alkylated in 20 mM iodoacetamide and 25 mM ammonium bicarbonate. In-gel protein digestion was performed using 0.2 mg/ml trypsin and 50 mM ammonium bicarbonate and resulting peptides were extracted from the gel slices. Peptides were then separated by reverse-phase high performance liquid chromatography (R-HPLC) and analysed using the Q-Exactive Quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific, USA) coupled with a Dionex Ultimate 3000 nano-HPLC system. Details of chromatographic separation and LC-MS/MS analysis can be found in supplementary file 1. Details of search parameters can be found in the summary sheet of Byonic output (raw data) for each replicate in supplementary folder 1. Precursor intensity was used for label-free quantification. Spectra obtained from the mass spectrometer were analysed using the Byonic™ Protein Metrics software[134] and a list of identified protein candidates obtained for biological replicates. LC-MS/MS data was filtered to remove known contaminants and to retain only proteins that had two or more unique peptides. Perseus, as part of MaxQuant[135] software was used for further data analysis. The data were $\log_2$ transformed and imputation of missing data was performed to replace missing values with random values drawn from a normal distribution that mimics the signal of low abundant proteins. Matrices containing imputed values were only used for cluster analysis and a Z-score was computed prior to principle component analysis. The original $\log_2$ transformed data, prior to imputation, were subsequently filtered for rows which contained 3 valid values for each group. An ANOVA with Benjamini Hochberg correction was performed to determine significant difference between groups and a fold change of 2 standard deviations from the mean between the test and control was included as statistically enriched. Proteins that were present in all 3 replicates in the Csd or SufC samples and not present in the mCherry control were selected when filtering for the high stringency list. For the intermediate stringency list, we chose

proteins that were present in all 3 replicates in the Csd and SufC affinity purified samples and either not present at all or only detected once in the mCherry control. Differentially enriched proteins were added to both the stringent and intermediate stringency lists. Venny 2.0.2[136] was used to construct a Venn diagram of proteins common to both the Csd and SufC enrichment. Gene ontology terms were retrieved from UniProt[137]. STRING[114] version 10 was used to visualize interaction networks.

LC-MS/MS analysis of the RNA polymerase complex was performed at Stellenbosch University Central analytical facility (CAF) on a Dionex Ultimate 3000 nano-LC system connected to a linear quadropole ion trap–Orbitrap (LTQ-Orbitrap) Fusion Tribrid mass spectrometer equipped with a nanoelectrospray ion source. The gel fraction was excised from the NativePAGE™ Novex® 4–16% Bis-Tris Gel and cut into 1 mm x 1 mm cubes and dehydrated using 100% acetonitrile. This sample was processed according to standard operating procedures of CAF. Proteins were digested and peptides separated by liquid chromatography. MS spectra was analysed using X! Tandem, MyriMatch and Sequest search engines. Details of search parameters can be found in supplementary file 2. Raw data can be found in supplementary file 3. Data was filtered for proteins that have 2 or more unique peptides that map to it. Gene ontology terms were retrieved from UniProt[137].

Chapter 4: Results

4.1. Method development and optimisation

4.1.1. Assessing the growth of *M. bovis* BCG Pasteur 1173P2 under standard conditions

In order to identify the optical density that corresponds to mid-logarithmic (midlog) growth phase for the *M. bovis* BCG Pasteur 1173P2 strain under standard conditions, we measured OD_{600} over a period of 18 days. (Fig. 4.1). An OD_{600} of ~1 was chosen as midlog phase and used for all subsequent experiments.

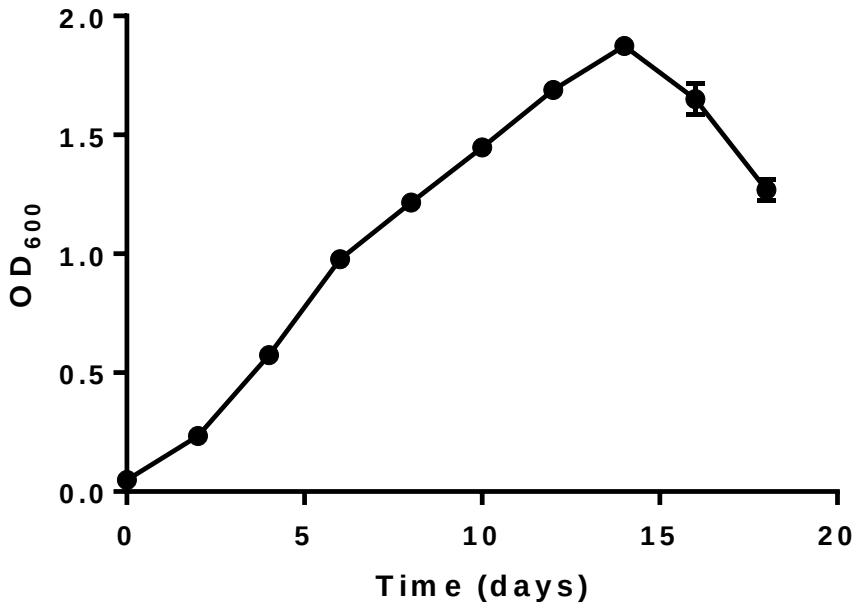


Figure 4.1. Growth of *M. bovis* BCG Pasteur 1173P2 under standard conditions. *M. bovis* BCG Pasteur 1173P2 cultured under standard conditions over 18 days from a starting OD_{600} of 0.05. An average of 3 biological replicates plotted with error bars (standard deviation).

4.1.2. Optimisation of Blue-native PAGE

Blue native polyacrylamide gel electrophoresis (BN-PAGE), originally developed for the analysis of mitochondrial protein complexes[138], is a specialised electrophoresis method which is relatively simple and sensitive and has also been applied for the analysis of cytosolic protein complexes[139]. Here we optimised BN-PAGE for the analysis of mycobacterial cytosolic protein complexes. Bead-beating and sonication were compared as cell lysis methods prior to BN-PAGE analysis (Fig. 4.2) by comparing the protein yield and the

banding patterns of cytoplasmic proteins extracted from *M. bovis* BCG. Sonication yielded the highest concentration of protein and was therefore selected as the lysis method for further analysis. The effect of DNase I treatment (20 U/ml) and dialysis on the banding patterns were also assessed. DNase I treatment appeared to improve the banding pattern by decreasing sample streaking due to DNA-protein association and was included in all subsequent experiments. Buffer-exchange did not appear to have any effect on the banding pattern and was not included in subsequent experiments.

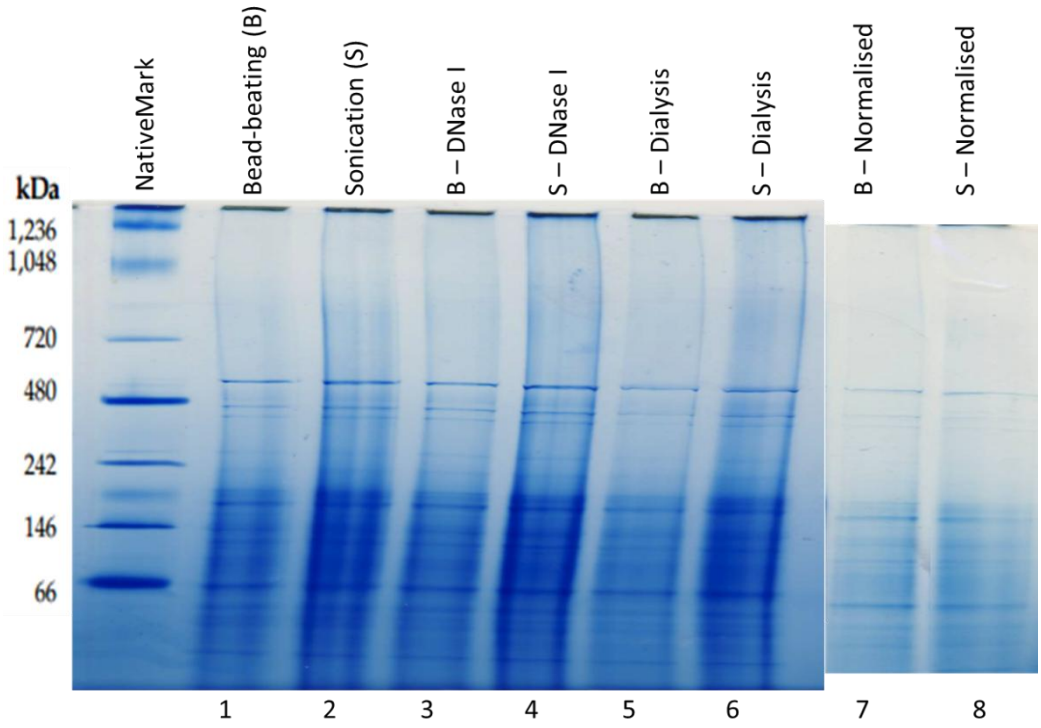


Figure 4.2. Optimisation of extraction procedures for analysis of cytoplasmic complexes by Blue Native PAGE. 4-16% BN-PAGE gel of *M. bovis* BCG cytoplasmic proteins when using bead-beating and sonication as cell-lysis techniques, respectively. Lane 1: bead-beating; lane 2: sonication; lane 3: bead-beating and DNase I treatment; lane 4: sonication and DNase I treatment; lane 5: bead-beating, DNase I treatment and dialysis; lane 6: sonication, DNase I treatment and dialysis; lane 7 and 8: bead-beating and sonication samples normalised by protein concentration.

4.1.3 Immunodetection of the RNA polymerase complex using BN-PAGE and Western blotting

Immunodetection of a well characterised multiprotein complex, RNA polymerase, was used to verify if selected protein complexes could be identified using BN-PAGE.

Cytoplasmic protein complexes were extracted from *M. bovis* BCG and *M. smegmatis* and analysed by BN-PAGE (Fig. 4.3). A Western blot was performed using an RpoB-specific antibody in order to identify the RNA-polymerase complex. The theoretical size of the RNA-polymerase complex is 407.465 kDa, based on the molecular weights of RpoA, RpoB, RpoC, RpoZ, SigA and SigH. On the BN-PAGE, RNA-polymerase appeared to be present in at least 2 complexes, and both these complexes were detected in the region of the 480 kDa band of the Native BN-PAGE marker.

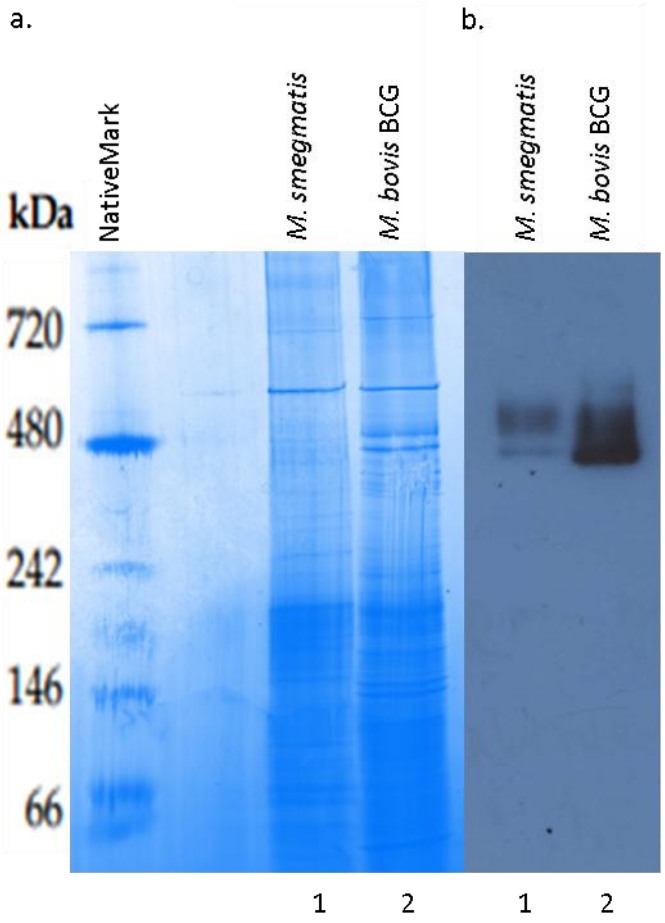


Figure 4.3. Immunodetection of RNA polymerase complex in cytoplasmic fraction separated by BN-PAGE. a) 4-16% BN-PAGE gel showing cytoplasmic protein lysate from *M. smegmatis* in lane 1 and *M. bovis* BCG in lane 2. b) Western blot of a) using an RpoB-specific antibody.

4.1.4. Immunodetection of RpoB following 2D BN/SDS-PAGE

In order to verify the presence of RpoB in the complexes detected in 4.1.3, the *M. bovis* BCG lane from the BN-PAGE gel was the excised and run in the second dimension

under denaturing conditions in order to separate protein complexes into their individual components. A signal for RpoB was detected in the expected region, and corresponded approximately to the size of the RpoB subunit (~130 kDa) (Fig. 4.4).

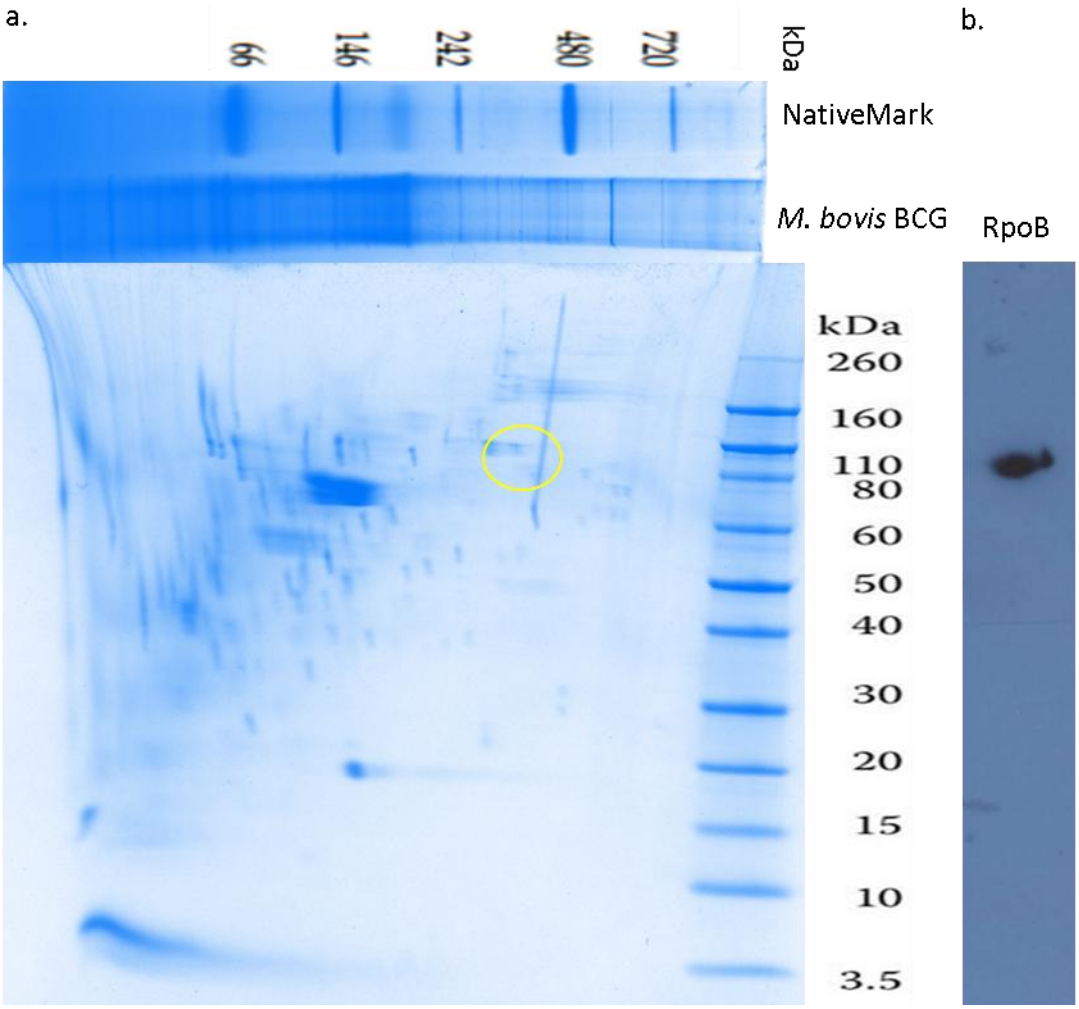


Figure 4.4. Immunodetection of RpoB in cytoplasmic fraction separated by 2D BN/SDS-PAGE. a) 4-16% 2D BN/SDS-PAGE gel of cytoplasmic protein complexes extracted from *M. bovis* BCG. b) Western blot of a) using an RpoB-specific antibody. The region on the second dimension gel where RpoB was detected is indicated by the yellow circle.

4.1.5. Validating BN-PAGE using Mass spectrometry

High-resolution mass spectrometry was used to verify components of the RNA polymerase complex identified by immunodetection in section 4.1.3. The BN-PAGE blot was overlaid with the BN-PAGE gel and the RNA polymerase complex excised and analysed by mass spectrometry (MS). RpoA, RpoB, RpoC and RpoD (SigA) could be detected by MS

(identified proteins with gene ontology can be found in the appendix – Table A1). Additional proteins were identified that could possibly associate with the RNA polymerase complex as detected by experimental and text mining prediction methods using the STRING software (Fig. 4.5). When all the STRING prediction methods were used, only 3 proteins show no predicted association with this complex, however, the presence of multiple complexes migrating to the same position in the gel cannot be ruled out.

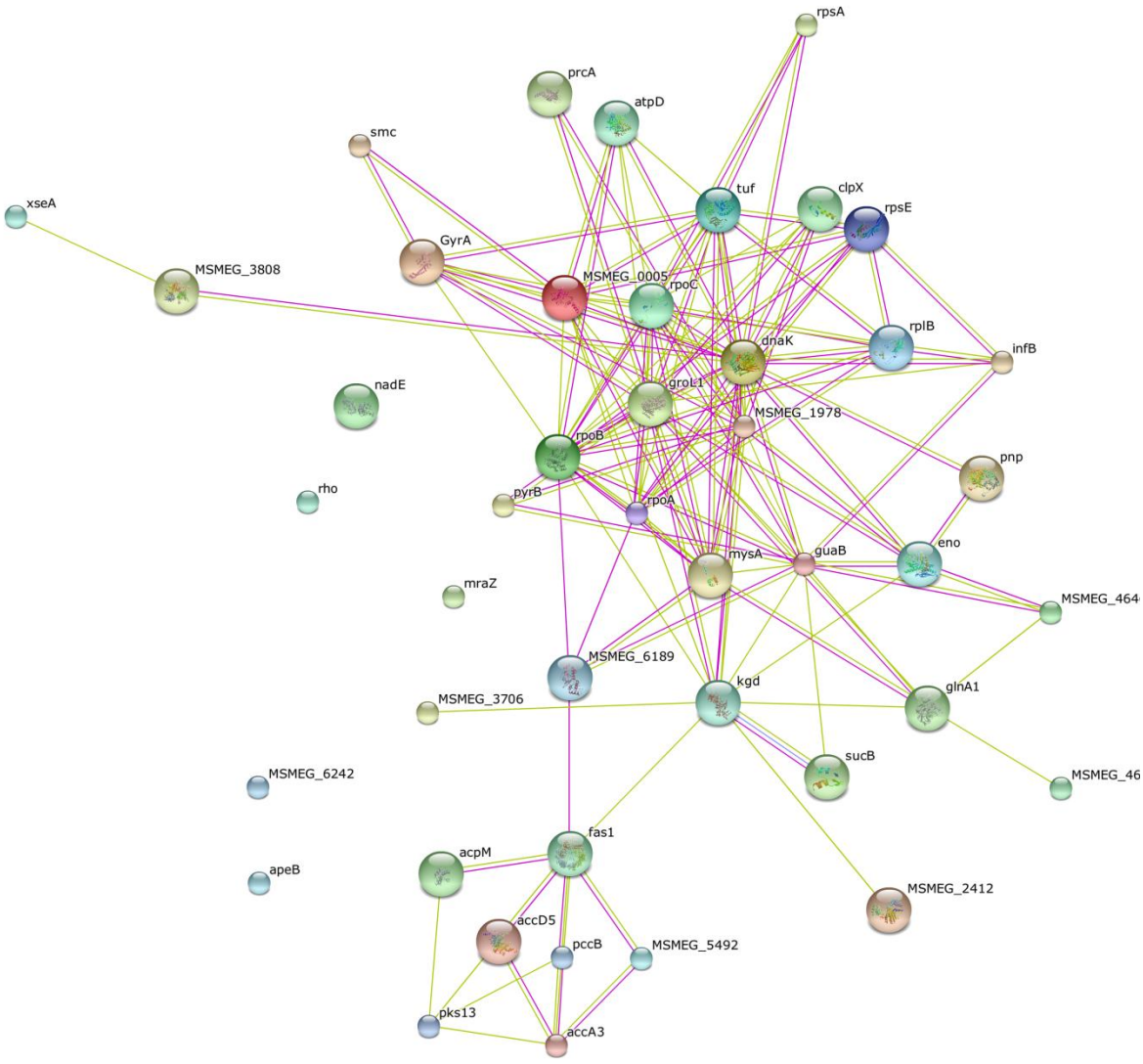


Figure 4.5. STRING protein-protein association network showing possible protein associations with the RNA polymerase complex as identified by BN-PAGE immunodetection and mass spectrometry.

4.2. Identifying the Fe-S cluster biogenesis machinery in the *M. bovis* BCG cytoplasmic fraction

Immunodetection of proteins in mycobacteria is limited by the availability of a suitable antibody to detect the protein of interest. Since no antibody against the components of the Fe-S cluster biogenesis machinery was available, we initially attempted to identify the complex by exposing the organism to conditions that would result in up-regulation of expression of the complex.

4.2.1. Hydrogen peroxide and nitric oxide exposure for inducing expression of the Fe-S cluster biogenesis machinery

Reactive oxygen and nitrogen species have previously been shown to upregulate genes involved in Fe-S cluster assembly[140,141]. Here we exposed *M. bovis* BCG to hydrogen peroxide and nitric oxide in order to evaluate if a visible change in levels of protein complexes could be observed when analysed by BN-PAGE (Fig. 4.6 – 4.9). The different exposure times indicated were chosen due to the fact that upregulation of certain Fe-S cluster assembly proteins was observed at these exposure times in previous studies in *M. tuberculosis* on a transcript level[140,141]. 7H9 liquid media was used as optimal broth for standard culturing. Nutrient starvation was previously shown to induce expression of the *suf* operon[142]. Here we use Sauton's minimal media in an attempt to enhance *suf* operon expression by using nutrient limitation in combination with exposure to oxidative stress. However, no visible difference could be observed.

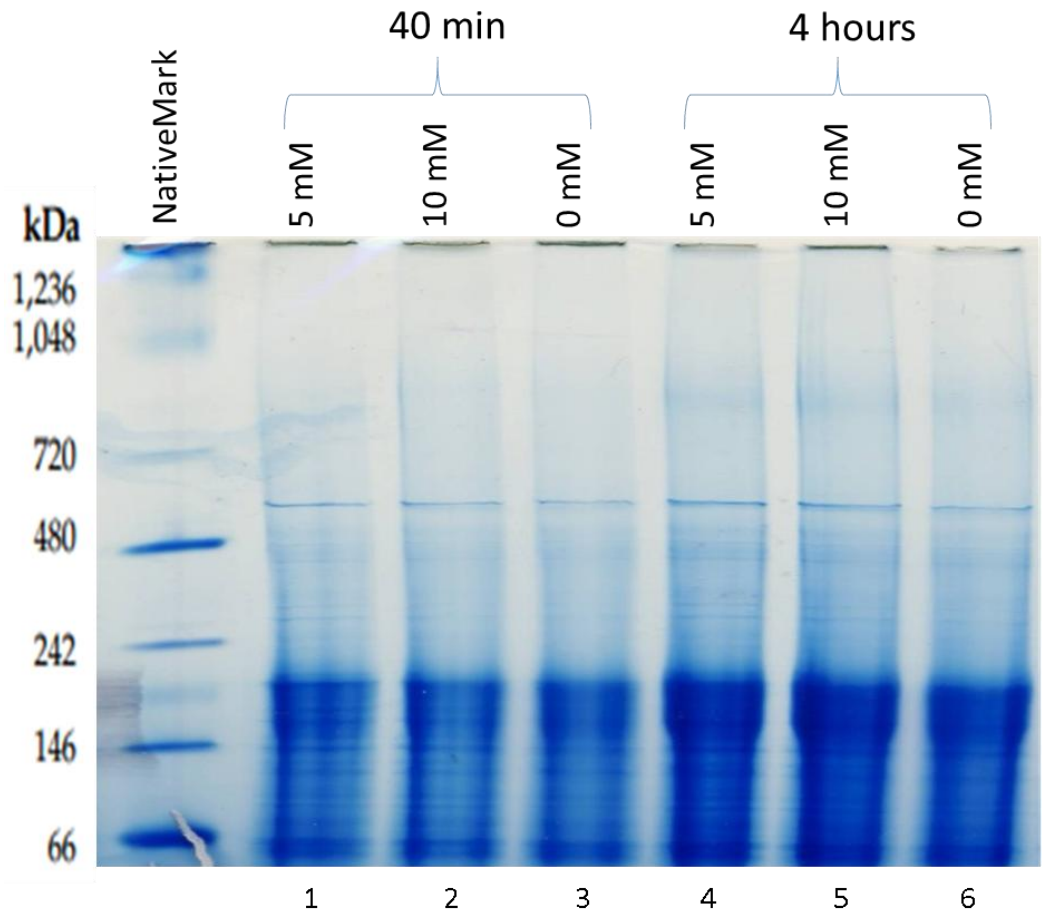


Figure 4.6. Separation of cytoplasmic complexes from *M. bovis* BCG by BN-PAGE following H_2O_2 exposure. 4-16% BN-PAGE gel showing the banding pattern of *M. bovis* BCG cytoplasmic protein complexes after exposure to H_2O_2 for 40 min and 4 hours under standard culture conditions. No visible difference in banding pattern was observed between lanes.

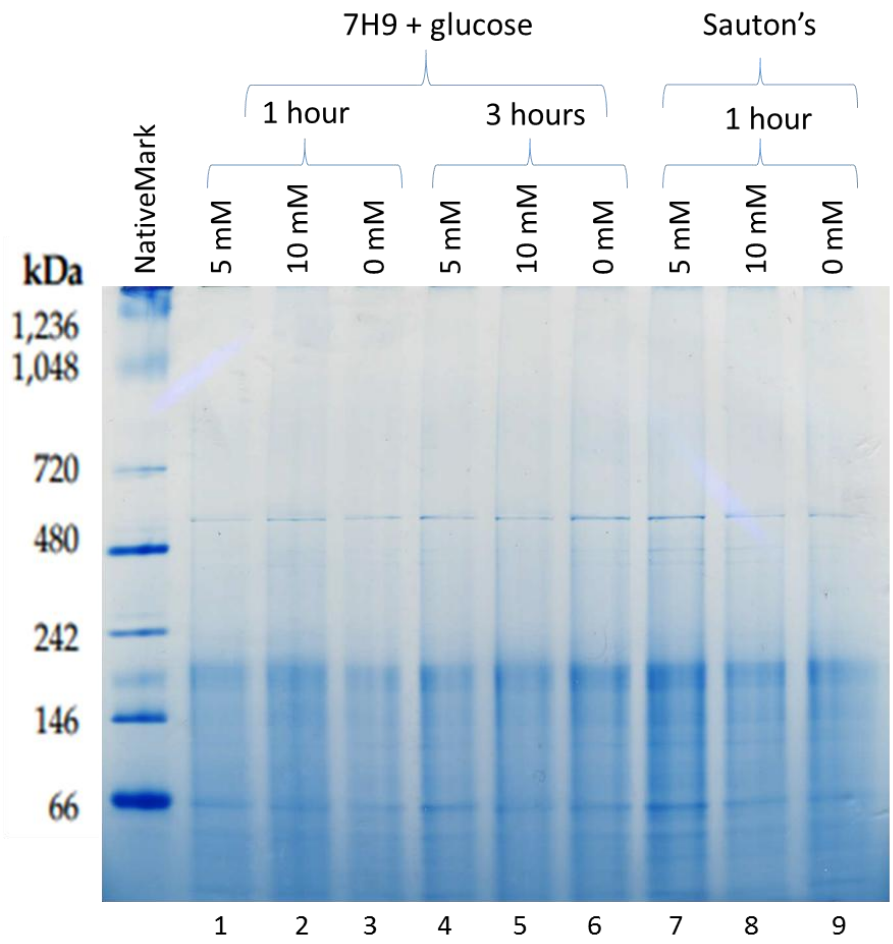


Figure 4.7. Separation of cytoplasmic complexes from *M. bovis* BCG by BN-PAGE following H_2O_2 exposure. 4-16% BN-PAGE gel showing the banding pattern of *M. bovis* BCG cytoplasmic protein complexes after exposure to H_2O_2 for 1 hour and 3 hours under standard conditions and for 1 hour in Sauton's minimal media. No visible difference in banding pattern was observed between lanes.

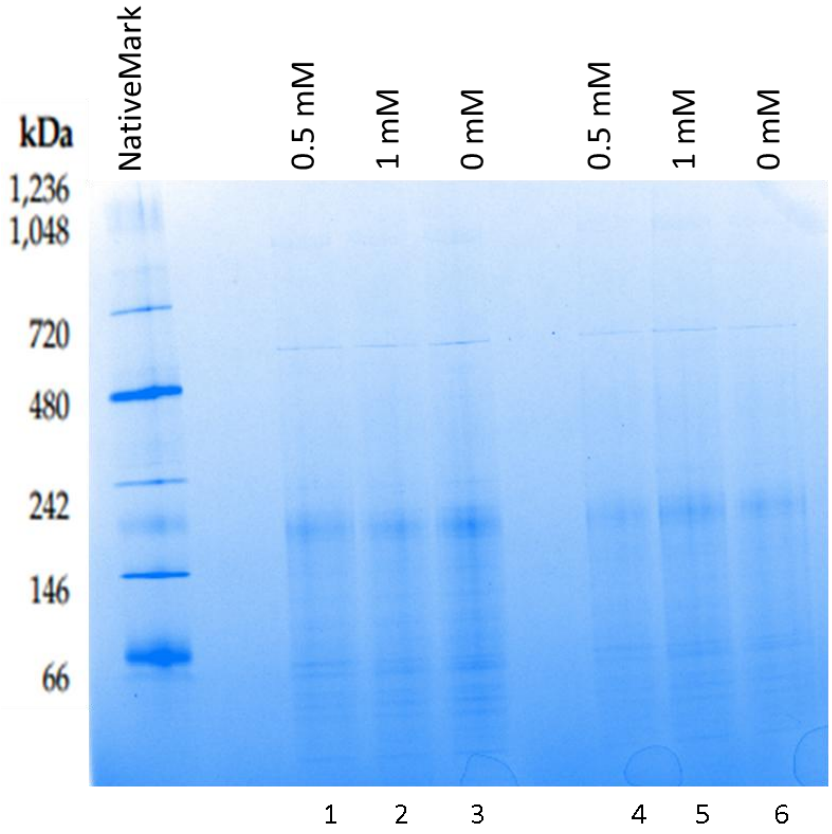


Figure 4.8. Separation of cytoplasmic complexes from *M. bovis* BCG by BN-PAGE following NO exposure. 4-16% BN-PAGE gel showing the banding pattern of *M. bovis* BCG cytoplasmic protein complexes after exposure to DETA/NO for 40 min under standard conditions. No visible difference in banding pattern was observed between lanes.

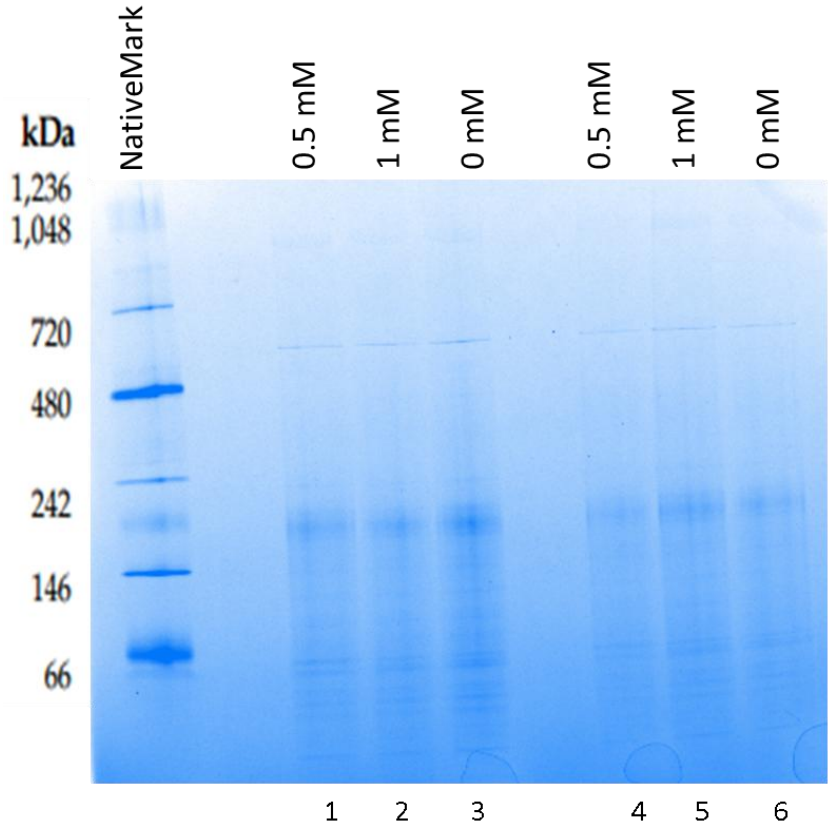


Figure 4.9. Separation of cytoplasmic complexes from *M. bovis* BCG by BN-PAGE following NO exposure. 4-16% BN-PAGE gel showing the banding pattern of *M. bovis* BCG cytoplasmic protein complexes after exposure to DETA/NO for 4 hours under standard conditions. No visible difference in banding pattern was observed between lanes.

4.2.2. Generation of a Csd-specific antibody for identifying the Fe-S cluster biogenesis machinery

Since identification of the Fe-S cluster machinery by hydrogen peroxide and NO exposure was unsuccessful, we attempted to generate an antibody directed against one of the components of the machinery, namely Csd, to detect the complex by immunodetection. In order to do this we needed to express and purify Csd.

4.2.2.1. Cloning of *csd* into *E. coli* expression vectors

csd was successfully PCR amplified from *Mycobacterium tuberculosis* H37Rv genomic DNA and cloned into three expression vectors (pET28a(+), pGEX-6P-1, pMAL-c5X (plasmid maps in appendix). Each construct was transformed into *E. coli* DH5-alpha and transformants plated on LB agar containing the appropriate antibiotic(s). Single colonies

were cultured and DNA extractions were performed followed by restriction enzyme digestion to screen for the presence of *csd* (Fig. 4.10 – 4.12), which was subsequently confirmed by sequencing.

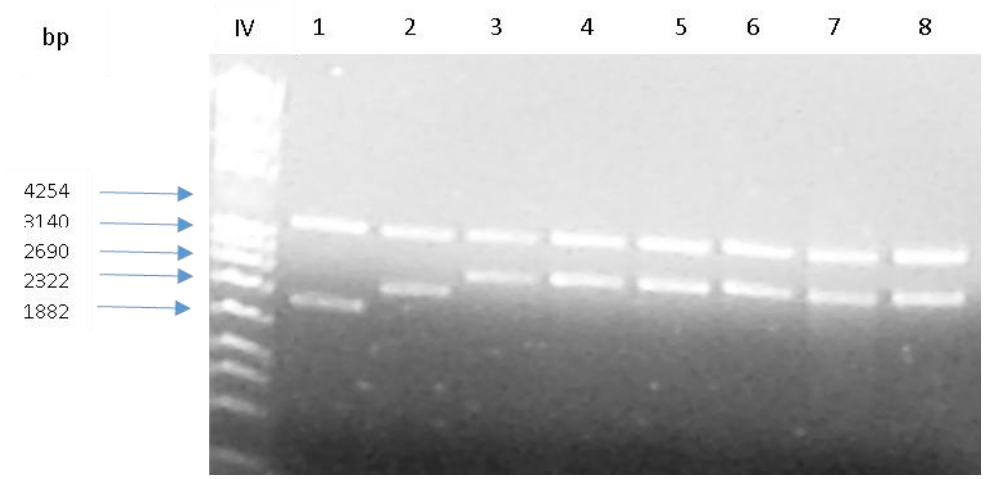


Figure 4.10. Restriction digestion of pET28a(+) clones to confirm presence of *csd*. 0.8% agarose gel of a pET28a(+)-*csd* restriction enzyme double digest with BamHI and EcoRV. Numbers 1 – 8 indicate the number of colonies that were screened. DNA Molecular Weight Marker IV (Sigma).

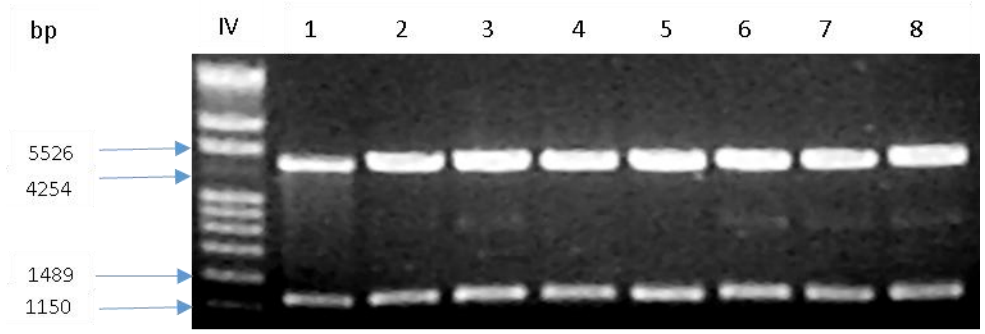


Figure 4.11. Restriction digestion of pGEX-6P-1 clones to confirm presence of *csd*. 0.8% agarose gel of a pGEX-6P-1-*csd* restriction enzyme double digest with EcoRI and XhoI. Numbers 1 – 8 indicate the number of colonies that were screened. DNA Molecular Weight Marker IV (Sigma).

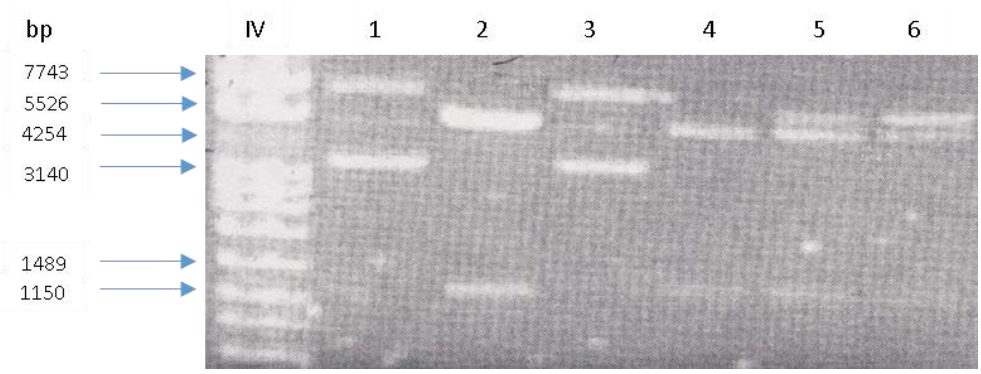


Figure 4.12. Restriction digestion of pMAL-c5X clones to confirm presence of *csd*. 0.8% agarose gel of a pMAL-c5X-*csd* restriction enzyme double digest with EcoRV and EcoRI. Numbers 1 – 6 indicate the number of colonies that were screened. DNA Molecular Weight Marker IV (Sigma).

4.2.2.2. Heterologous expression of Csd in *E. coli* ArcticExpress and Rosetta-gami strains

Three vectors for Csd expression in *E. coli* were generated, namely, pET28a(+)-*csd* for protein purification using a His-tag (17 kDa), pGEX-6P-1-*csd* for purification using a GST tag (26 kDa), and pMAL-c5X-*csd* for purification using a maltose-binding protein tag (42 kDa). Csd over-expression, in all three vectors, was induced for 24 h in *E. coli* ArcticExpress and Rosetta-gami and analyzed by SDS-PAGE (Fig. 4.13 – 4.18). No visible difference was observed for Csd expression in either strain in the soluble fraction (supernatant) compared to the empty vector and uninduced control. Csd expression appears predominant in the insoluble fraction (pellet). This approach was therefore discontinued.

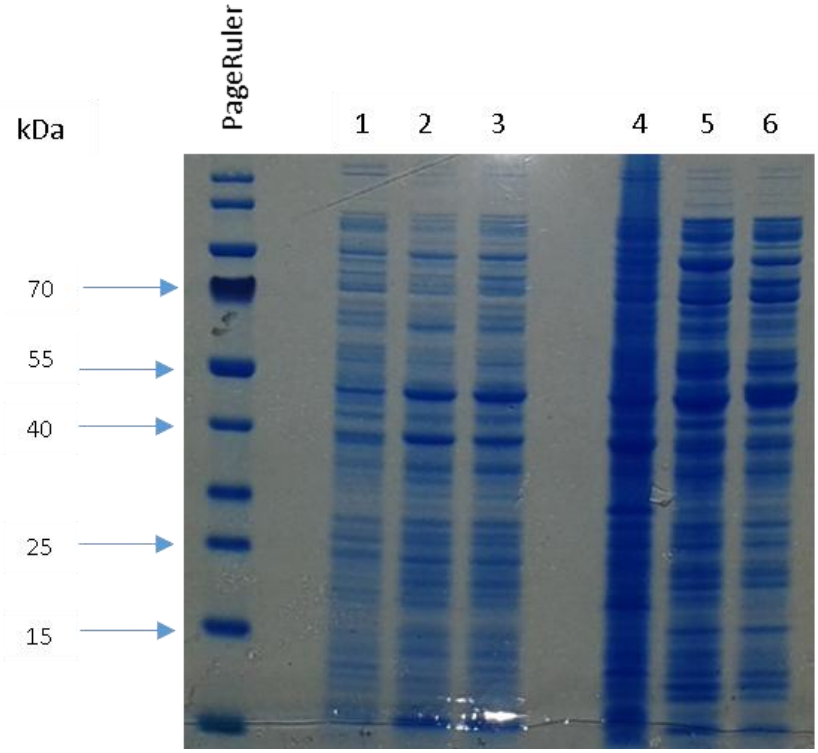


Figure 4.13. Expression of Csd from pET28a(+)-*csd* in *E. coli* ArcticExpress. 4-12% SDS-PAGE of protein fractions obtained after 24 hours of induction of *E. coli* ArcticExpress harbouring either pET28a(+)-*csd* or pET28a(+). (1) supernatant fraction pET28a(+)-*csd*, uninduced; (2) supernatant fraction pET28a(+), induced; (3) supernatant fraction pET28a(+)-*csd*, induced; (4) pellet fraction pET28a(+)-*csd*, uninduced; (5) pellet fraction pET28a(+), induced; (6) pellet fraction pET28a(+)-*csd*, induced. Expected band size: 61 kDa

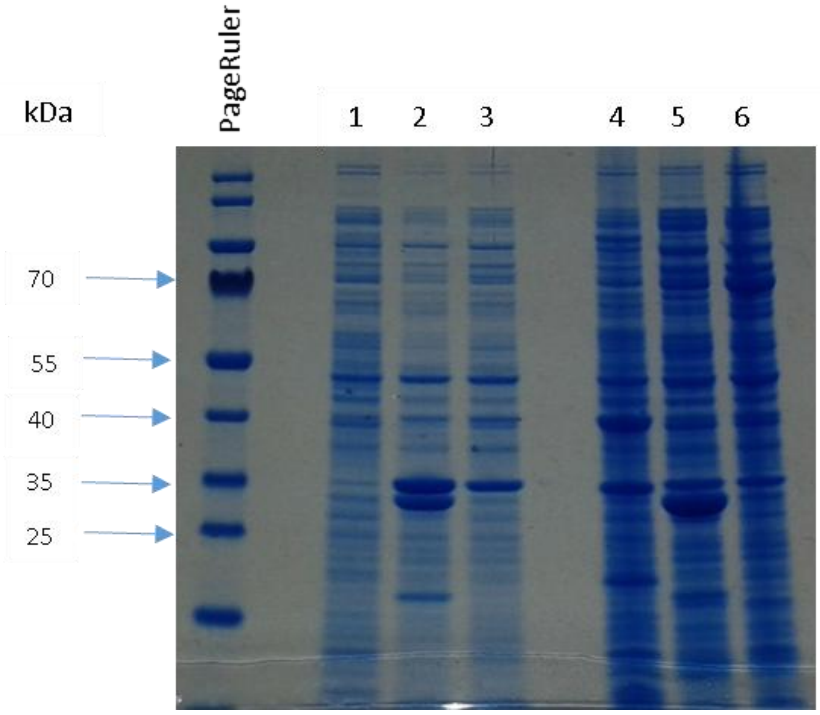


Figure 4.14. Expression of Csd from pGEX-6P-1-*csd* in *E. coli* ArcticExpress. 4-12% SDS-PAGE of protein fractions obtained after 24 hours of induction of *E. coli* ArcticExpress harbouring either pGEX-6P-1-*csd* or pGEX-6P-1. (1) supernatant fraction pGEX-6P-1-*csd*, uninduced; (2) supernatant fraction pGEX-6P-1, induced; (3) supernatant fraction pGEX-6P-1-*csd*, induced; (4) pellet fraction pGEX-6P-1-*csd*, uninduced; (5) pellet fraction pGEX-6P-1, induced; (6) pellet fraction pGEX-6P-1-*csd*, induced. Expected band size: 70 kDa

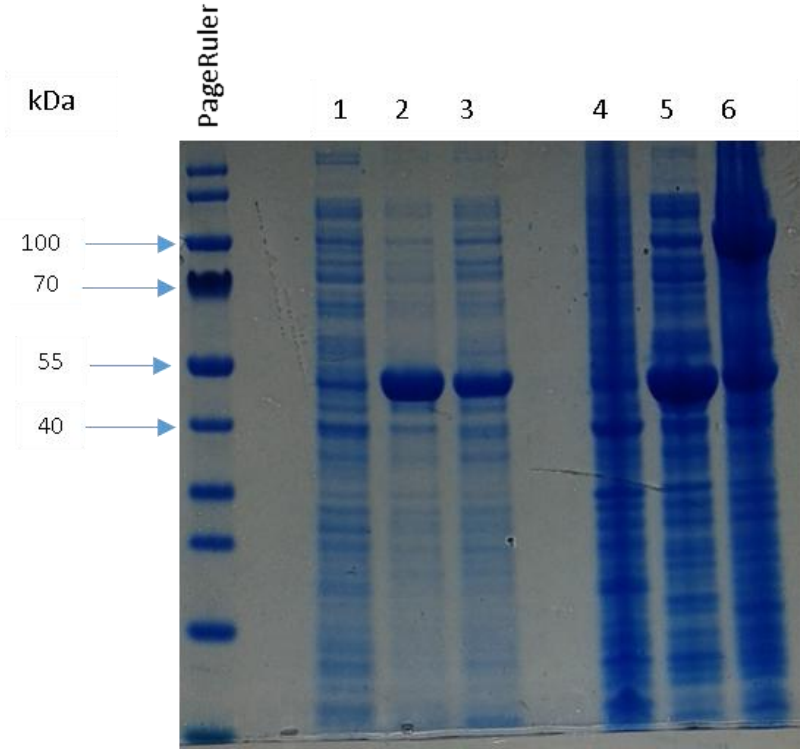


Figure 4.15. Expression of Csd from pMAL-c5X-*csd* in *E. coli* ArcticExpress. 4-12% SDS-PAGE of protein fractions obtained after 24 hours of induction of *E. coli* ArcticExpress harbouring either pMAL-c5X-*csd* or pMAL-c5X. (1) supernatant fraction pMAL-c5X-*csd*, uninduced; (2) supernatant fraction pMAL-c5X, induced; (3) supernatant fraction pMAL-c5X-*csd*, induced; (4) pellet fraction pMAL-c5X-*csd*, uninduced; (5) pellet fraction pMAL-c5X, induced; (6) pellet fraction pMAL-c5X-*csd*, induced. Expected band size: 86 kDa

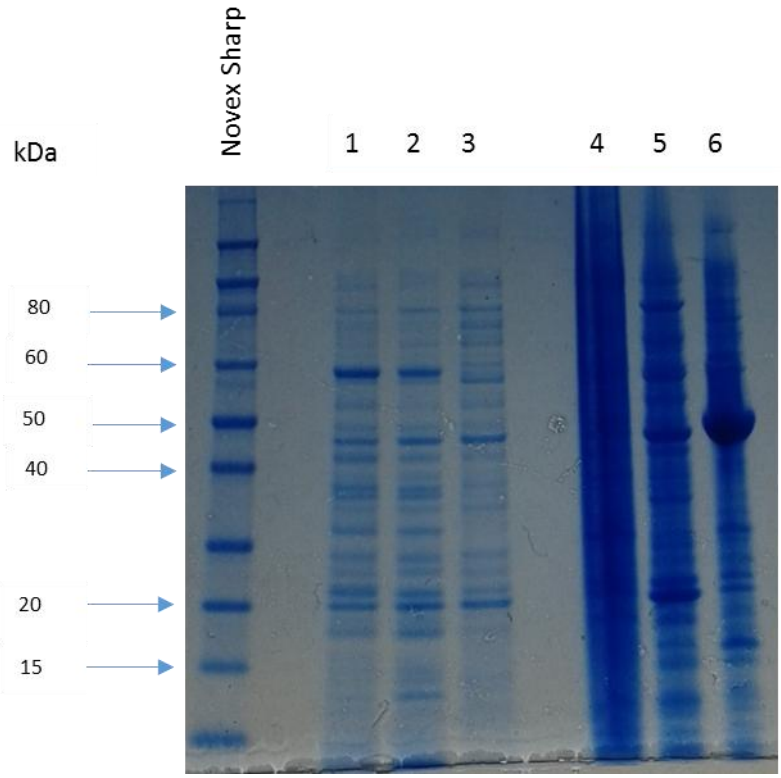


Figure 4.16. Expression of Csd from pET28a(+)-*csd* in *E. coli* Rosetta-gami. 4-12% SDS-PAGE of protein fractions obtained after 24 hours of induction of *E. coli* Rosetta-gami harbouring either pET28a(+)-*csd* or pET28a(+). (1) supernatant fraction pET28a(+)-*csd*, uninduced; (2) supernatant fraction pET28a(+), induced; (3) supernatant fraction pET28a(+)-*csd*, induced; (4) pellet fraction pET28a(+)-*csd*, uninduced; (5) pellet fraction pET28a(+), induced; (6) pellet fraction pET28a(+)-*csd*, induced. Expected band size: 61 kDa

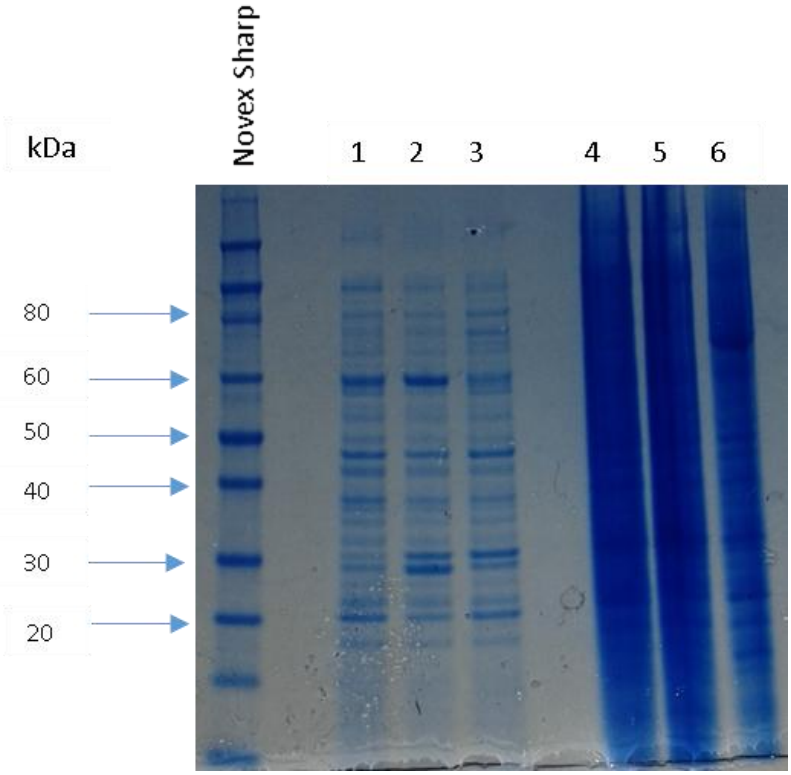


Figure 4.17. Expression of Csd from pGEX-6P-1-*csd* in *E. coli* Rosetta-gami. 4-12% SDS-PAGE of protein fractions obtained after 24 hours of induction of *E. coli* Rosetta-gami harbouring either pGEX-6P-1-*csd* or pGEX-6P-1. (1) supernatant fraction pGEX-6P-1-*csd*, uninduced; (2) supernatant fraction pGEX-6P-1, induced; (3) supernatant fraction pGEX-6P-1-*csd*, induced; (4) pellet fraction pGEX-6P-1-*csd*, uninduced; (5) pellet fraction pGEX-6P-1, induced; (6) pellet fraction pGEX-6P-1-*csd*, induced. Expected band size: 70 kDa

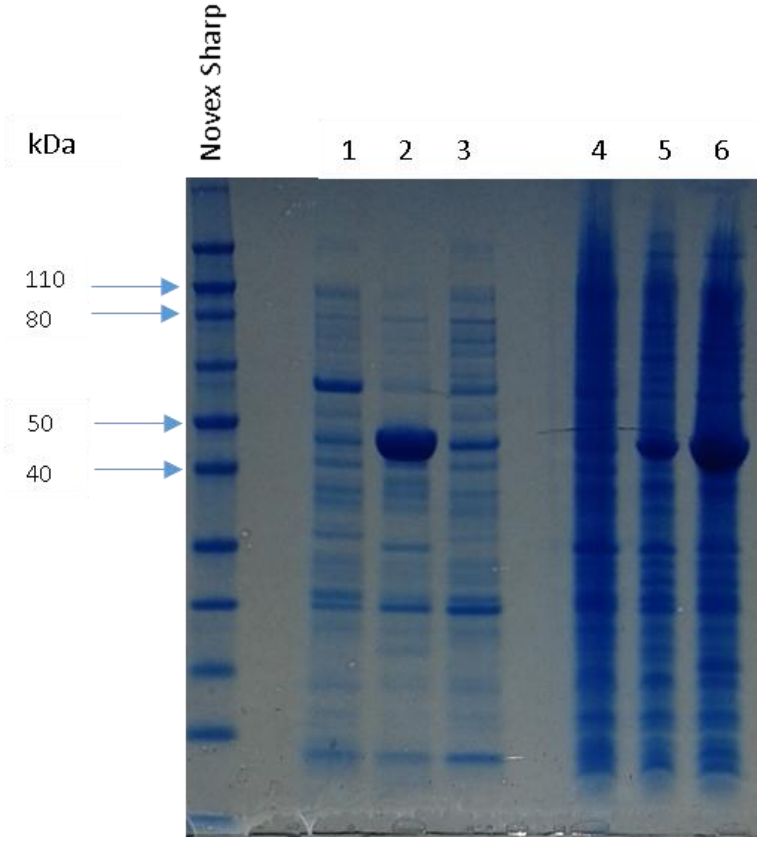


Figure 4.18. Expression of Csd from pMAL-c5X-*csd* in *E. coli* Rosetta-gami. 4-12% SDS-PAGE of protein fractions obtained after 24 hours of induction of *E. coli* Rosetta-gami harbouring either pMAL-c5X-*csd* or pMAL-c5X. (1) supernatant fraction pMAL-c5X-*csd*, uninduced; (2) supernatant fraction pMAL-c5X, induced; (3) supernatant fraction pMAL-c5X-*csd*, induced; (4) pellet fraction pMAL-c5X-*csd*, uninduced; (5) pellet fraction pMAL-c5X, induced; (6) pellet fraction pMAL-c5X-*csd*, induced. Expected band size: 86 kDa

4.2.3. Identifying Fe-S cluster biogenesis machinery by affinity purification and MS

Due to the problem of heterologous mycobacterial protein expression in *E. coli*, experienced in section 4.2.2, we decided to generate 6xHis-tag fusions of components of the Fe-S cluster machinery for expression in mycobacteria. We reasoned that these fusion proteins could then be used to enrich for the complex of interest by affinity purification using a nickel immobilised-metal affinity chromatography (IMAC) resin. Following enrichment, potential interacting partners could then be identified by MS.

4.2.3.1. *csd*, *sufc* and *mcherry* cloning into the pSE100 mycobacterial tetracycline-inducible expression vector

Regulation of gene expression using a tetracycline-responsive expression system

The tetracycline-responsive expression system (Tet system) has been developed and applied in mycobacteria for the controlled regulation of gene expression[143]. Here we use the reverse Tet system, which represses gene expression upon addition of atc, for optimising the expression of 6xHis-tag fusion proteins in *M. smegmatis*.

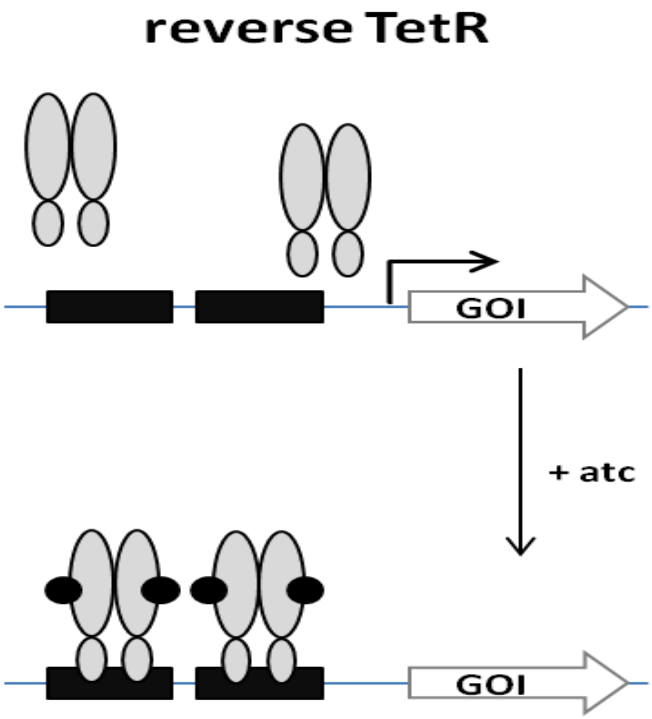


Figure 4.19. Diagram showing the reverse TetR-mediated regulation of gene expression upon addition of atc (adapted from reference 131)[131]. Reverse TetR is provided by the episomal plasmid, pTEK-4S-0X. In the absence of atc, reverse TetR does not bind to the tetO operator sites, enabling transcription. Upon addition of atc, reverse TetR binds to the tetO operator sites, inhibiting transcription.

Two components of the Fe-S cluster biogenesis machinery, namely Csd and SufC, were used for this part of the study. An unrelated protein, mCherry, which has been codon optimised for expression in mycobacteria, was included as a negative control for the affinity purification study. *csd*, *sufc* and *mcherry* were cloned independently into a modified version

of pSE100, which contains a ribosome binding site and N-terminal 6xHis tag. These vectors were propagated in *E. coli* DH5-alpha and screened by restriction enzyme digestion (Fig. 4.20 – 4.22). Plasmids were confirmed to be correct by sequencing and transformed into *M. smegmatis* mc^2 155 along with pTEK-4S-0X to allow for tetracycline-inducible gene expression.

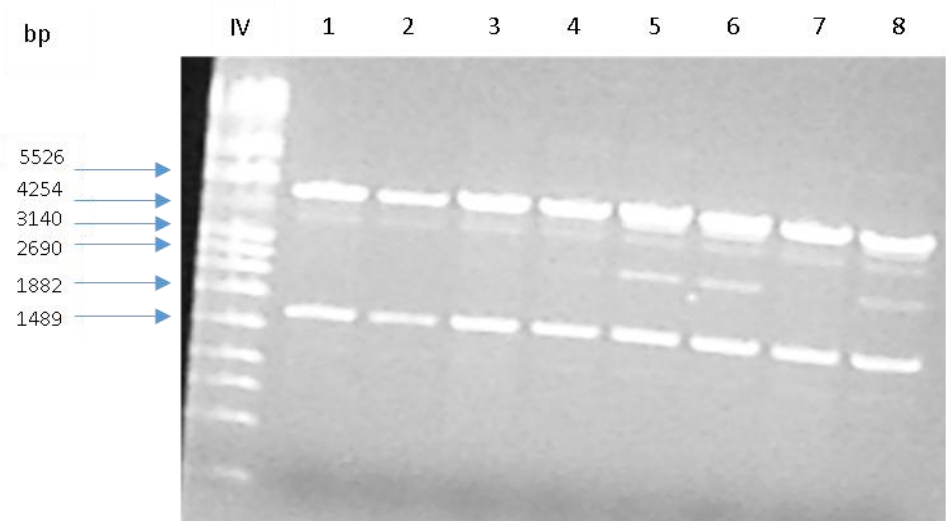


Figure 4.20. Restriction enzyme digest of pSE100-rbs-*csd*. 0.8% agarose gel of a pSE100-rbs-*csd* restriction enzyme double digest with XbaI and HindIII. Numbers 1 – 8 indicate the number of colonies that were screened. DNA Molecular Weight Marker IV (Sigma).

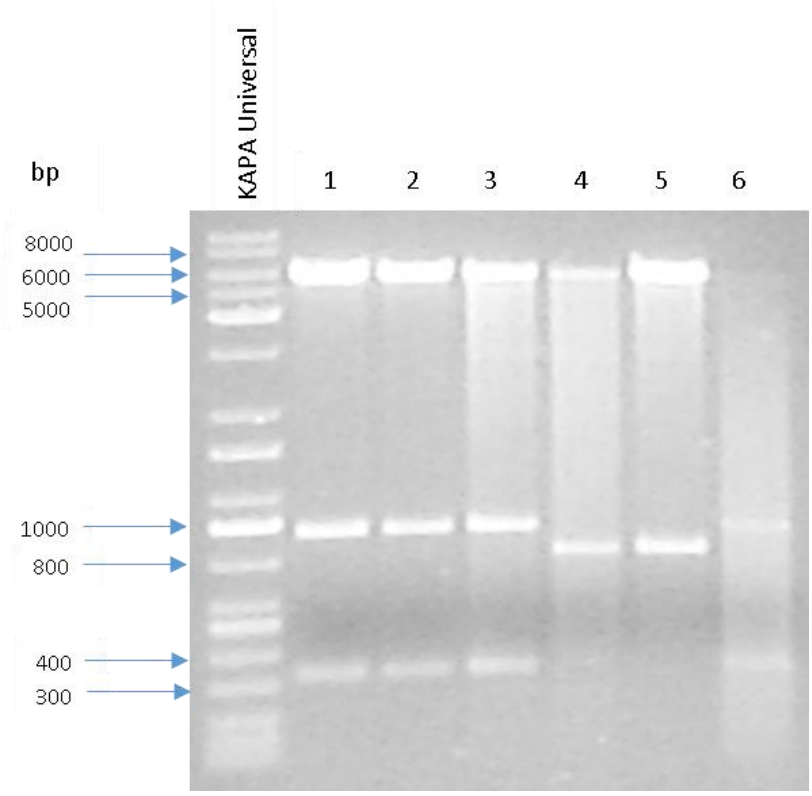


Figure 4.21. Restriction enzyme digest of pSE100-rbs-*sufc*. 0.8% agarose gel of a pSE100-rbs-*sufc* restriction enzyme double digest with NdeI and HindIII. Numbers 1 – 6 indicate the number of colonies that were screened.

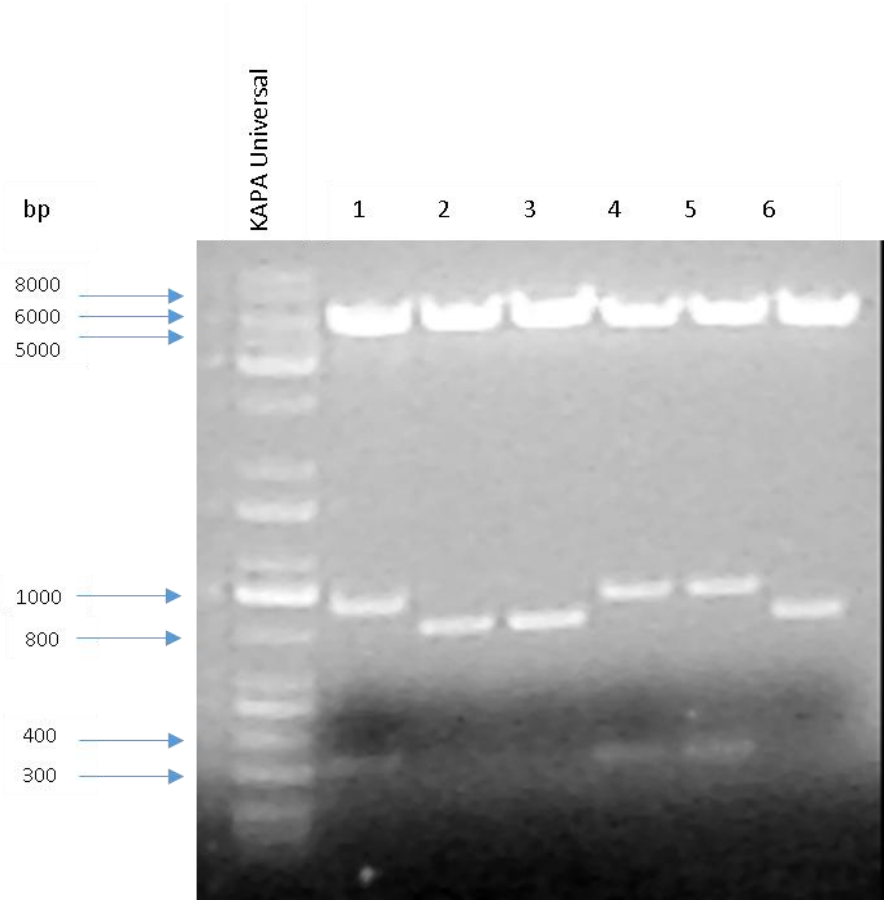


Figure 4.22. Restriction enzyme digest of pSE100-rbs-*mcherry*. 0.8% agarose gel of a pSE100-rbs-*mcherry* restriction enzyme double digest with NdeI and HindIII. Numbers 1 – 6 indicate the number of colonies that were screened.

4.2.3.2. Assessing the growth of *M. smegmatis* strains in the presence and absence of atc

To test the effect of protein over-expression on cell viability, *M. smegmatis* mc^2 155 derivatives generated in this study were cultured under standard conditions and the growth monitored over 36 hours in the presence and absence of atc (Fig. 4.23 – 4.24). No difference in growth between strains was observed.

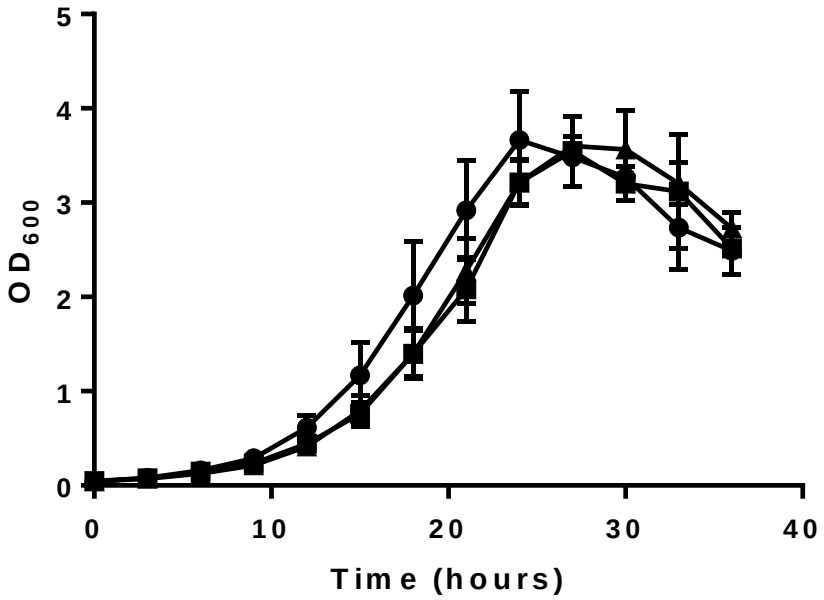


Figure 4.23. Growth curve of *M. smegmatis* mc^2 155 derivatives in the presence of 0 ng/ml atc over 36 hours. Circle: *M. smegmatis* mc^2 155 *attB*::pSE100-rbs-*csd* (pTEK-4S-0X); Square: *M. smegmatis* mc^2 155 *attB*::pSE100-rbs-*sufc* (pTEK-4S-0X); Triangle: *M. smegmatis* mc^2 155 *attB*::pSE100-rbs-*mcherry* (pTEK-4S-0X). An average of 3 biological replicates for each strain plotted with error bars (standard deviation).

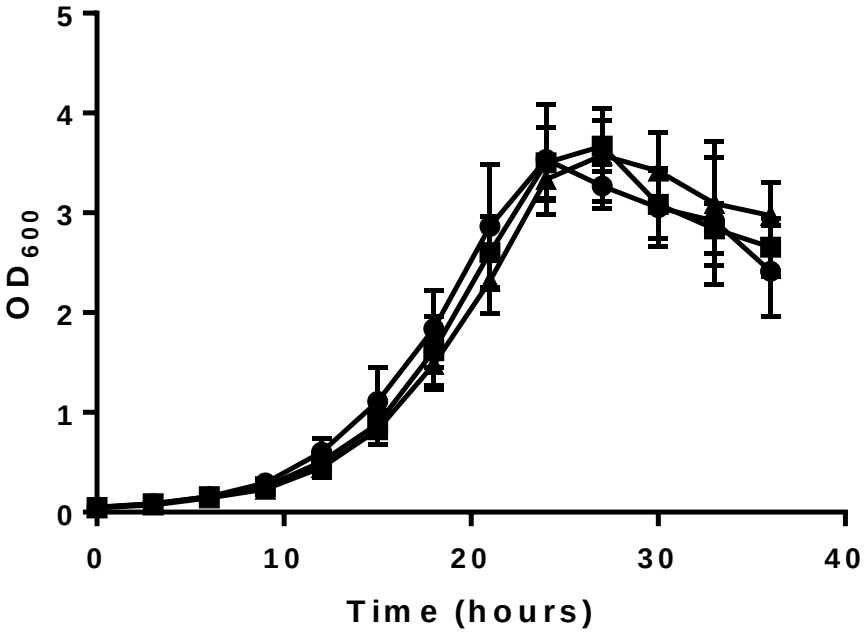


Figure 4.24. Growth curve of *M. smegmatis* mc^2 155 derivatives in the presence of 10 ng/ml atc over 36 hours. Circle: *M. smegmatis* mc^2 155 *attB*::pSE100-rbs-*csd* (pTEK-4S-0X); Square: *M. smegmatis* mc^2 155 *attB*::pSE100-rbs-*sufc* (pTEK-4S-0X); Triangle: *M. smegmatis* mc^2 155 *attB*::pSE100-rbs-*mcherry* (pTEK-4S-0X). An average of 3 biological replicates for each strain plotted with error bars (standard deviation).

4.2.3.3. Expression of Csd, SufC and mCherry – 6xHis fusion proteins in *M. smegmatis*

In order to optimise the expression of 6xHis-tag fusion proteins expression was monitored over time in the absence and presence of anhydrotetracycline (atc). *M. smegmatis* mc^2 155 derivatives, *M. smegmatis* mc^2 155 *attB*::pSE100-rbs-*csd* (pTEK-4S-0X), *M. smegmatis* mc^2 155 *attB*::pSE100-rbs-*sufc* (pTEK-4S-0X) and *M. smegmatis* mc^2 155 *attB*::pSE100-rbs-*mcherry* (pTEK-4S-0X), were cultured under standard conditions. Both the soluble and insoluble protein fractions were collected and analysed by SDS-PAGE (Fig. 4.25 – 4.30). In order to determine the expression of Csd relative to the amount of total protein extracted, the intensity of the band detected on the Western blot was normalised to the total protein detected on the membrane using the Bio-Rad Image Lab™ software (Table 4.1 – 4.4). The majority of 6xHis-tagged proteins were detected in the soluble fraction. Sixteen hour culturing in the absence of atc was selected for all subsequent experiments. An

intermediate atc concentration of 1 ng/ml was also tested (results not shown), but showed a negligible difference from what was observed at 10 ng/ml.

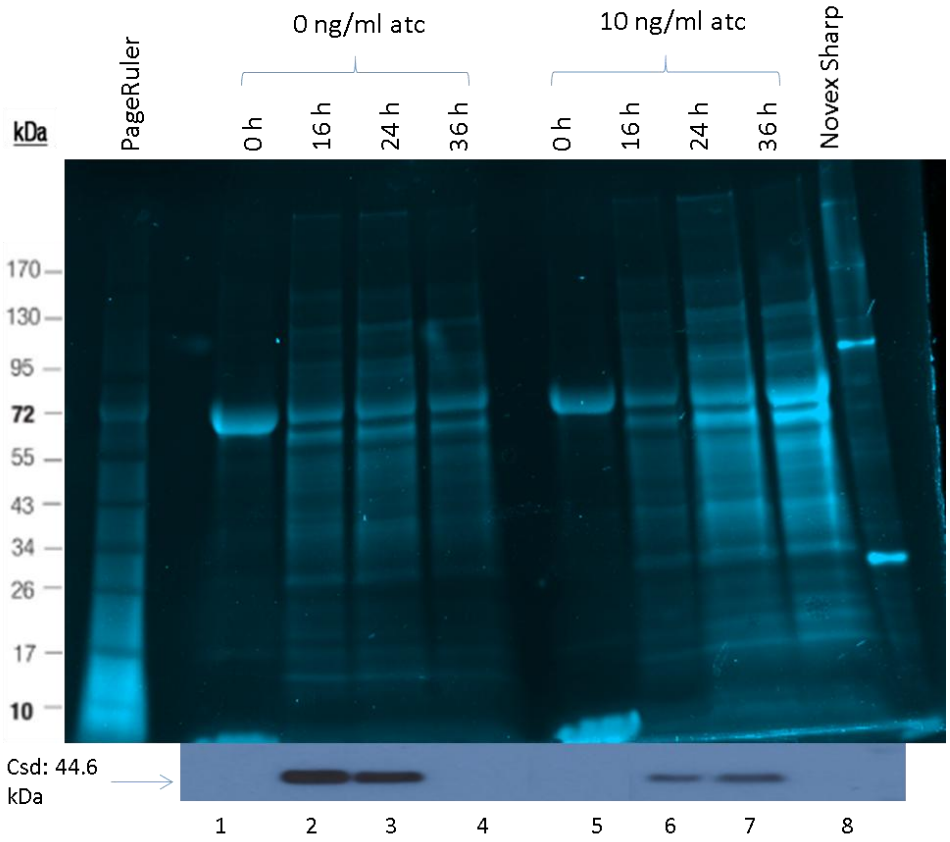


Figure 4.25. Expression of Csd in *M. smegmatis* mc^2 155 *attB*::pSE100-rbs-*csd* (pTEK-4S-0X). 4-15% SDS-PAGE stain-free gel and western blot showing the change in protein expression of *M. smegmatis* mc^2 155 *attB*::pSE100-rbs-*csd* (pTEK-4S-0X) soluble fraction over 36 hours in the presence of 0 ng/ml atc (lane 1-4) and 10 ng/ml atc (lane 5-8).

Table 4.1. Western blot normalisation of *M. smegmatis* mc^2 155 *attB*::pSE100-rbs-*csd* (pTEK-4S-0X) soluble fraction

Lane no.	Volume (intensity)	Normalisation factor	Normalised volume (intensity)
2	1 884 932	1.000	1 884 932
3	1 224 608	1.495	1 831 114
6	463 344	1.270	588 379
7	582 316	0.730	425 145

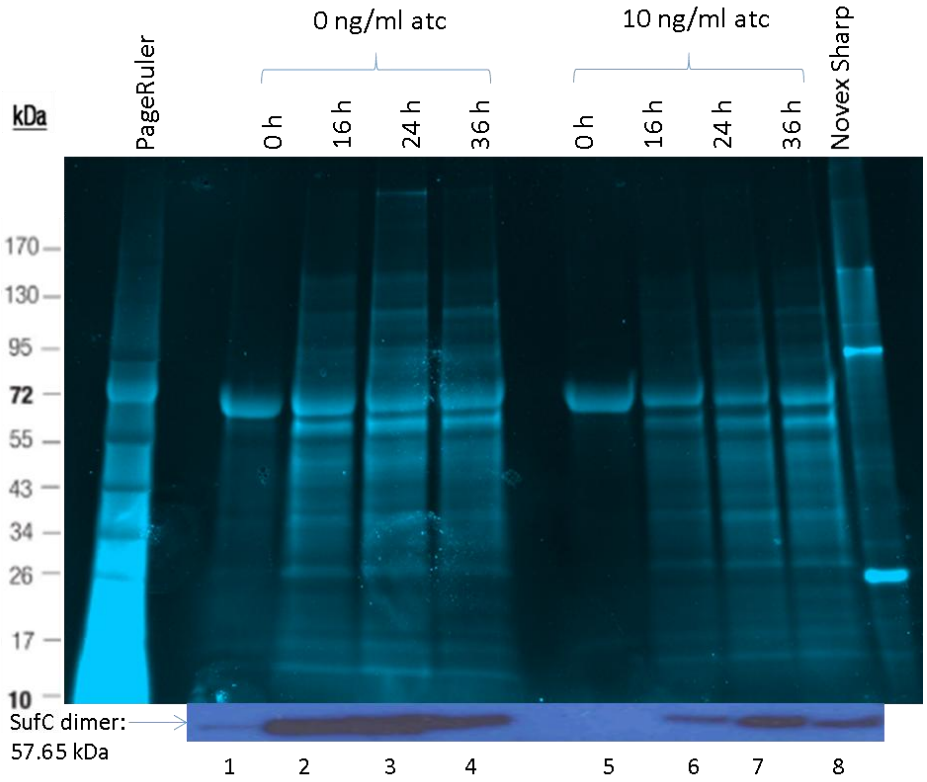


Figure 4.26. Expression of SufC in *M. smegmatis* mc^2 155 *attB*::pSE100-rbs-*sufc* (pTEK-4S-0X). 4-15% SDS-PAGE stain-free gel and western blot showing the change in protein expression of *M. smegmatis* mc^2 155 *attB*::pSE100-rbs-*sufc* (pTEK-4S-0X) soluble fraction over 36 hours in the presence of 0 ng/ml atc (lane 1-4) and 10 ng/ml atc (lane 5-8).

Table 4.2. Western blot normalisation of *M. smegmatis* mc^2 155 *attB*::pSE100-rbs-*sufc* (pTEK-4S-0X) soluble fraction

Lane no.	Volume (intensity)	Normalisation factor	Normalised volume (intensity)
1	136 000	1.000	136 000
2	1 168 400	0.730	853 277
3	1 201 300	0.750	901 126
4	693 400	1.062	736 341
6	212 900	0.464	98 864
7	383 100	1.176	450 544
8	189 600	1.024	194 111

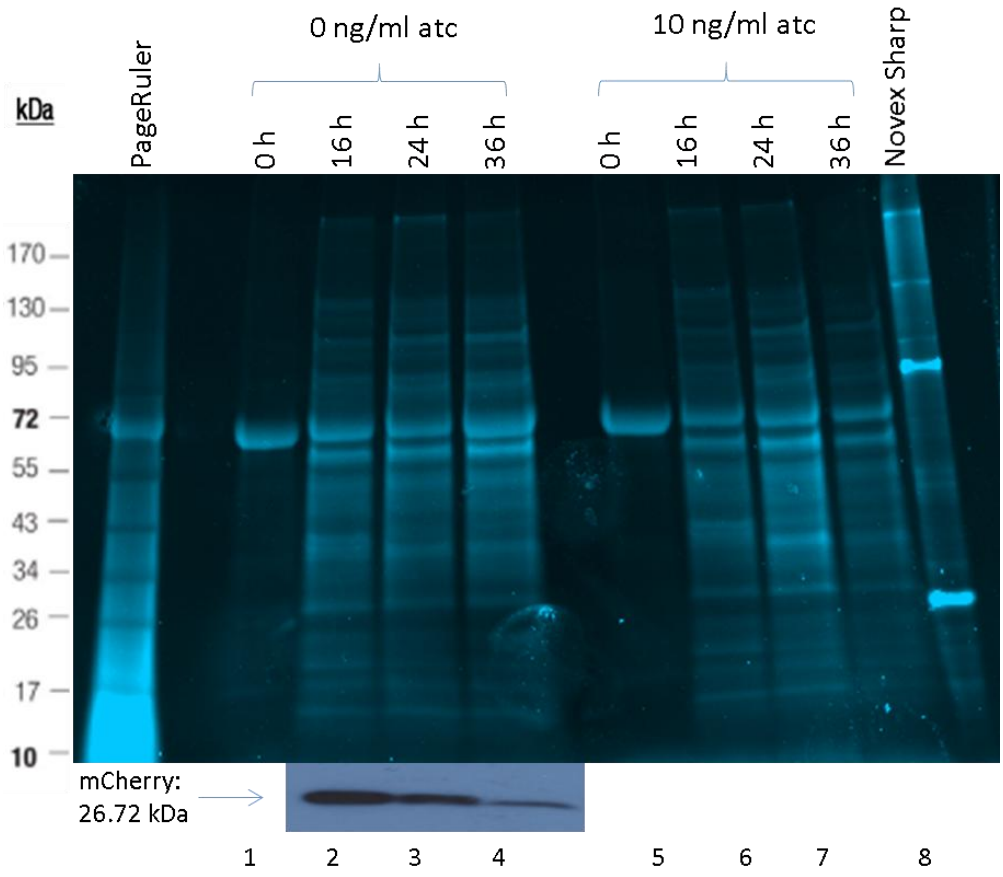


Figure 4.27. Expression of mCherry in *M. smegmatis* mc^2 155 *attB*::pSE100-rbs-*mcherry* (pTEK-4S-0X). 4-15% SDS-PAGE stain-free gel and western blot showing the change in protein expression of *M. smegmatis* mc^2 155 *attB*::pSE100-rbs-*mcherry* (pTEK-4S-0X) soluble fraction over 36 hours in the presence of 0 ng/ml atc (lane 1-4) and 10 ng/ml atc (lane 5-8).

Table 4.3. Western blot normalisation of *M. smegmatis* mc^2 155 *attB*::pSE100-rbs-*mcherry* (pTEK-4S-0X) soluble fraction

Lane no.	Volume (intensity)	Normalisation factor	Normalised volume (intensity)
2	1 808 982	1.000	1 808 982
3	311 346	0.509	158 572
4	158 572	0.492	59 918

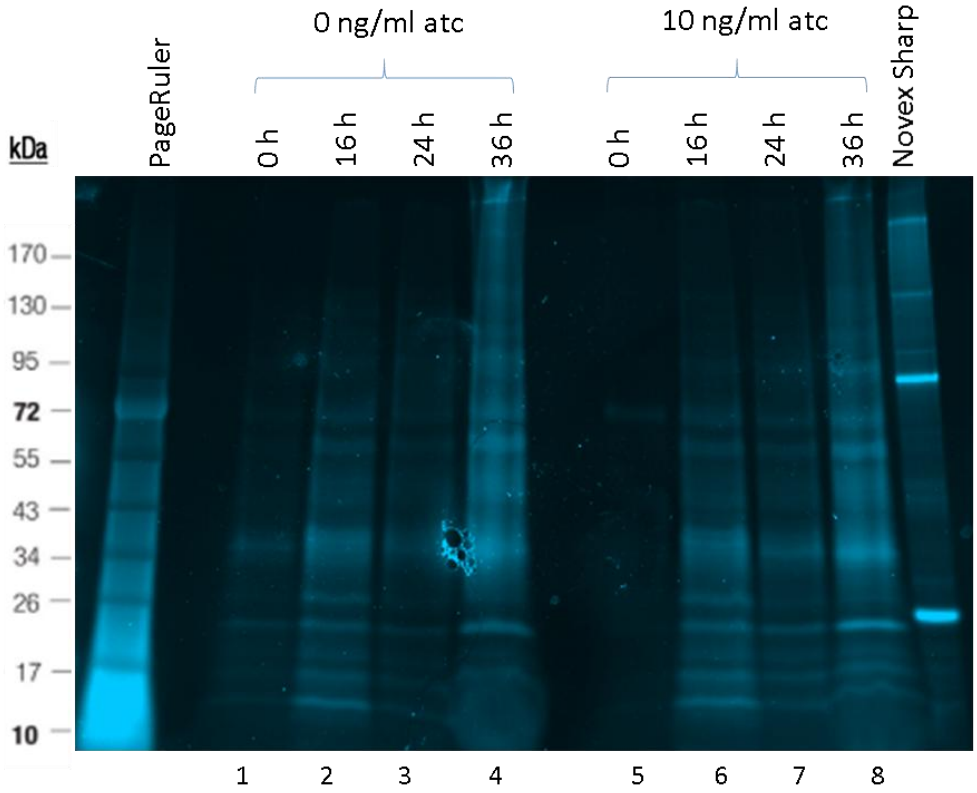


Figure 4.28. Expression of Csd in *M. smegmatis* mc^2 155 *attB*::pSE100-rbs-*csd* (pTEK-4S-0X). 4-15% SDS-PAGE stain-free gel and western blot showing the change in protein expression of *M. smegmatis* mc^2 155 *attB*::pSE100-rbs-*csd* (pTEK-4S-0X) insoluble fraction over 36 hours in the presence of 0 ng/ml atc (lane 1-4) and 10 ng/ml atc (lane 5-8).

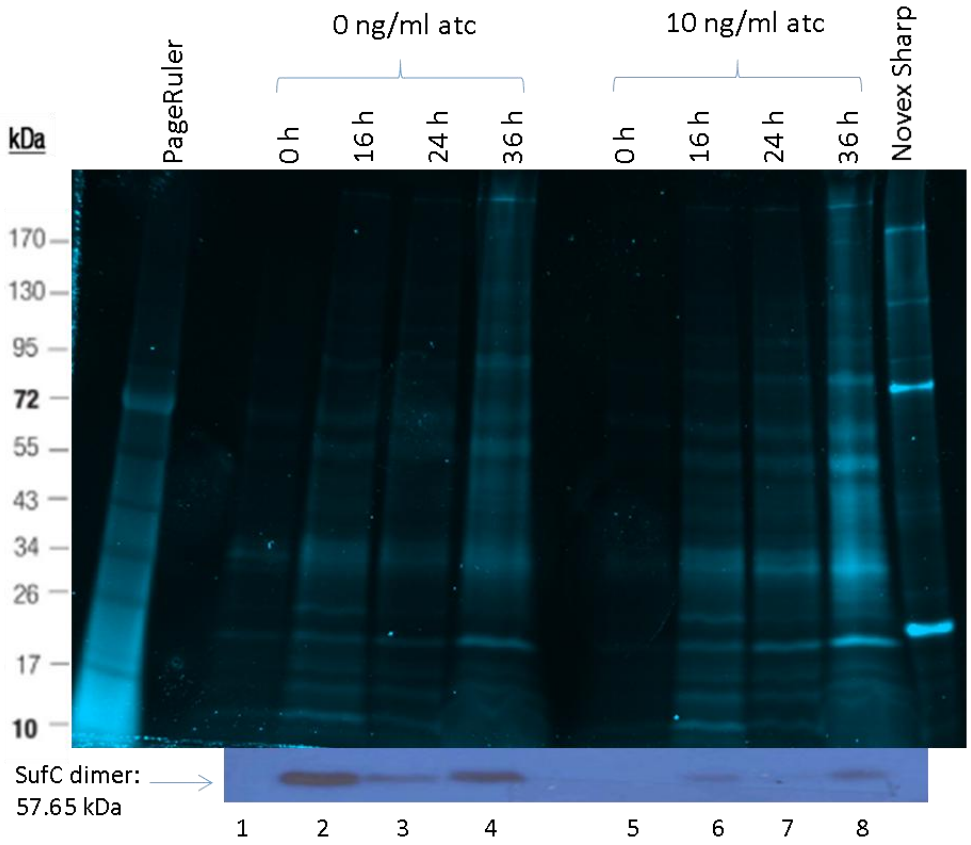


Figure 4.29. Expression of SufC in *M. smegmatis* mc^2 155 *attB*::pSE100-rbs-*sufc* (pTEK-4S-0X). 4-15% SDS-PAGE stain-free gel and western blot showing the change in protein expression of *M. smegmatis* mc^2 155 *attB*::pSE100-rbs-*sufc* (pTEK-4S-0X) insoluble fraction over 36 hours in the presence of 0 ng/ml atc (lane 1-4) and 10 ng/ml atc (lane 5-8).

Table 4.4. Western blot normalisation of *M. smegmatis* mc^2 155 *attB*::pSE100-rbs-*sufc* (pTEK-4S-0X) insoluble fraction

Lane no.	Volume (intensity)	Normalisation factor	Normalised volume (intensity)
2	4 924 500	1.000	4 924 500
3	1 501 900	0 989	1 485 879
4	4 255 200	0.511	2 174 973
6	684 800	0.716	490 564
7	311 000	0.618	192 240
8	1 289 700	0.381	490 812

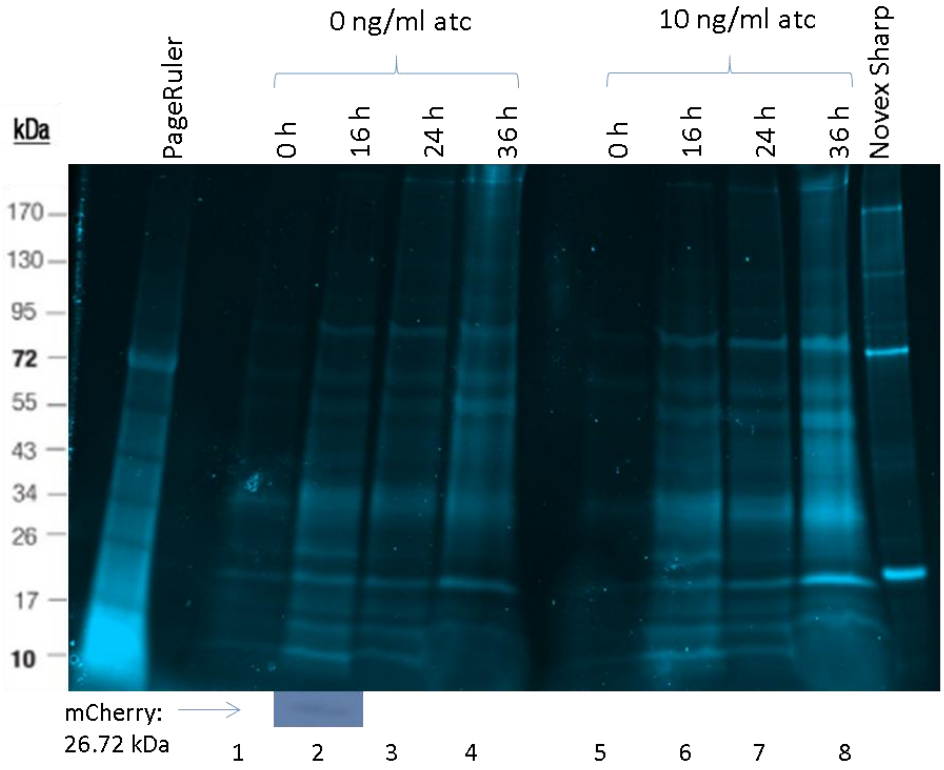


Figure 4.30. Expression of mCherry in *M. smegmatis* mc^2 155 *attB*::pSE100-rbs-*mcherry* (pTEK-4S-0X). 4-15% SDS-PAGE stain-free gel and western blot showing the change in protein expression of *M. smegmatis* mc^2 155 *attB*::pSE100-rbs-*mcherry* (pTEK-4S-0X) insoluble fraction over 36 hours in the presence of 0 ng/ml atc (lane 1-4) and 10 ng/ml atc (lane 5-8).

4.2.3.4. Enriching for Fe-S cluster machinery by Ni-IMAC affinity purification

The Hsp60 chaperone, GroEL, is expressed at relatively high levels in *M. smegmatis*[132]. This is problematic as GroEL has a high affinity for Ni-particles due to its histidine-rich C-terminus. To overcome the problem of non-specific co-purification of GroEL, *M. smegmatis groEL1ΔC* was used for enrichment of Fe-S cluster machinery by Ni-NTA. *M. smegmatis groEL1ΔC* is a derivative of *M. smegmatis* mc^2 155 carrying a deletion of the histidine-rich C-terminus in GroEL1. To confirm that growth of this strain was not influenced by expression of the fusion proteins, growth was monitored under standard conditions in the presence and absence of atc over 36 hours (Fig. 4.31 – 4.32). No difference was observed in growth when the expression of the fusion proteins was induced.

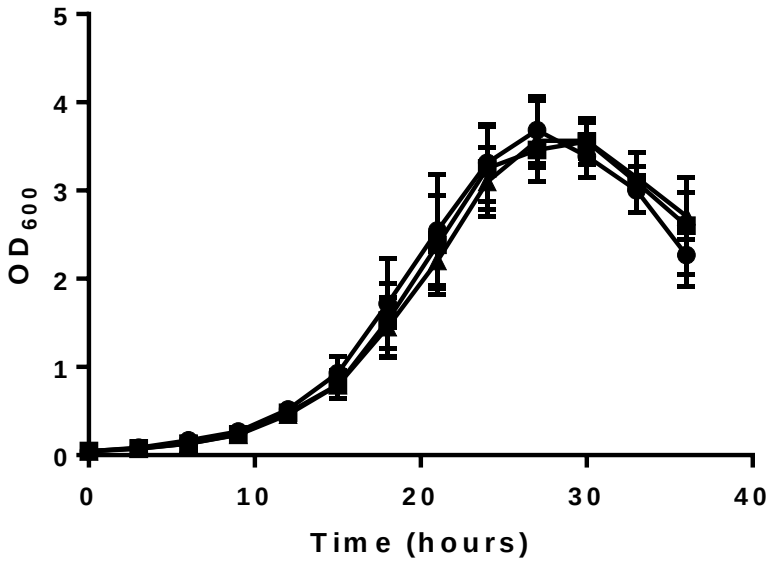


Figure 4.31. Growth curve of *M. smegmatis groEL1ΔC* derivatives in the presence of 0 ng/ml atc over 36 hours. Circle: *M. smegmatis groEL1ΔC attB*::pSE100-rbs-*csd* (pTEK-4S-0X); Square: *M. smegmatis groEL1ΔC attB*::pSE100-rbs-*sufc* (pTEK-4S-0X); Triangle: *M. smegmatis groEL1ΔC attB*::pSE100-rbs-*mcherry* (pTEK-4S-0X). An average of 3 biological replicates for each strain plotted with error bars (standard deviation).

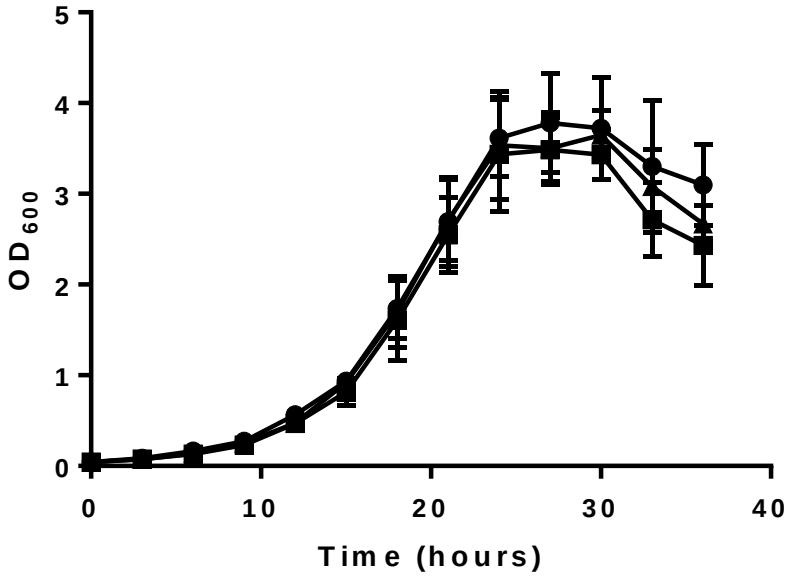


Figure 4.32. Growth curve of *M. smegmatis groEL1ΔC* derivatives in the presence of 10 ng/ml atc over 36 hours. Circle: *M. smegmatis groEL1ΔC attB*::pSE100-rbs-*csd* (pTEK-4S-0X); Square: *M. smegmatis groEL1ΔC attB*::pSE100-rbs-*sufc* (pTEK-4S-0X); Triangle: *M. smegmatis groEL1ΔC attB*::pSE100-rbs-*mcherry* (pTEK-4S-0X). An average of 3 biological replicates for each strain plotted with error bars (standard deviation).

Cytoplasmic extracts from the *M. smegmatis groEL1ΔC attB*::pSE100-rbs-*csd* (pTEK-4S-0X) strain were used to optimise the affinity purification conditions. The influence of the volume of MagnaHis™ beads added, and the concentration of salt in the extraction buffer on the amount of Csd recovered from 5 ml of cytoplasmic extract were evaluated by SDS-PAGE and Western blotting. The amount of Csd eluted from the beads appeared to be constant for all conditions (Fig. 4.33), however the residual Csd that remained in the supernatant fraction decreased as the volume of Ni-particles increased for both the high and low salt conditions (Fig. 4.34 & 4.35). The decrease in Csd remaining in the supernatant was more dramatic for the high salt conditions (Fig. 4.35). No difference in Csd recovery was seen between the 60 µl and 100 µl Ni-particle samples for the low salt supernatant fraction (Fig. 4.36). Therefore, 100 mM NaCl and 60 µl of Ni-particles per 5 ml of culture grown for 16 hours in the absence of atc were used for all subsequent analysis.

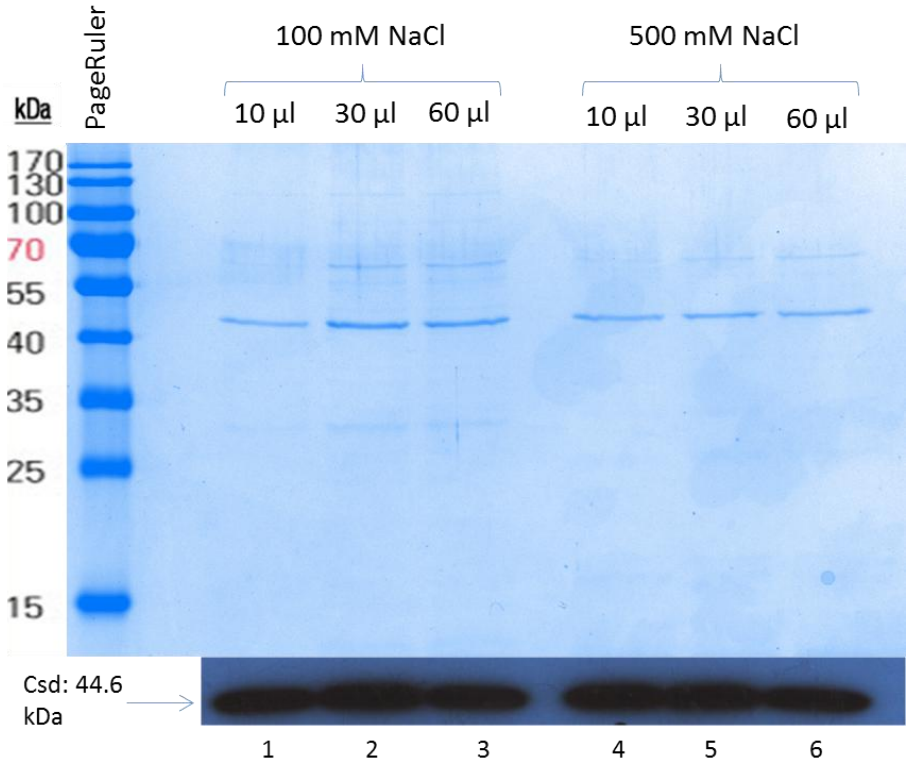


Figure 4.33. Optimisation of the affinity purification conditions. 12% SDS-PAGE gel showing the enrichment for Csd-associated proteins using 10-, 30-, and 60 µl Ni-particles under high (500 mM) or low (100 mM) salt conditions. Immunodetection showing the change in Csd recovered from MagnaHis beads under the above mentioned conditions.

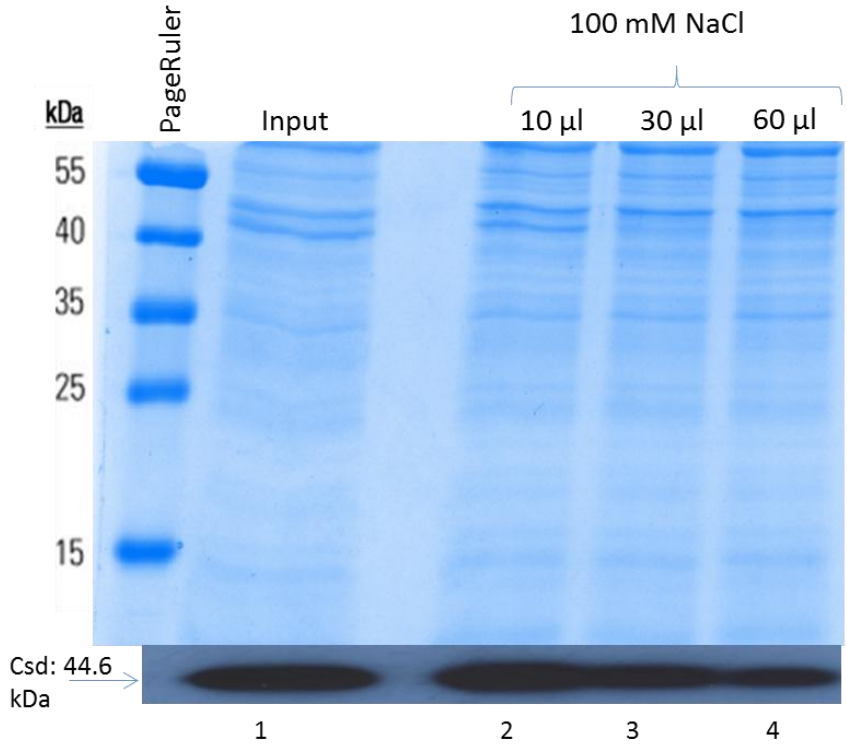


Figure 4.34. Optimisation of the affinity purification conditions. 12% SDS-PAGE gel showing the supernatant fraction (unbound Csd) using 10-, 30-, and 60 µl Ni-particles under low (100 mM) salt conditions. Western blot showing the change in residual Csd under the above mentioned conditions for the supernatant fraction. Lane 1: Input sample (before enrichment)

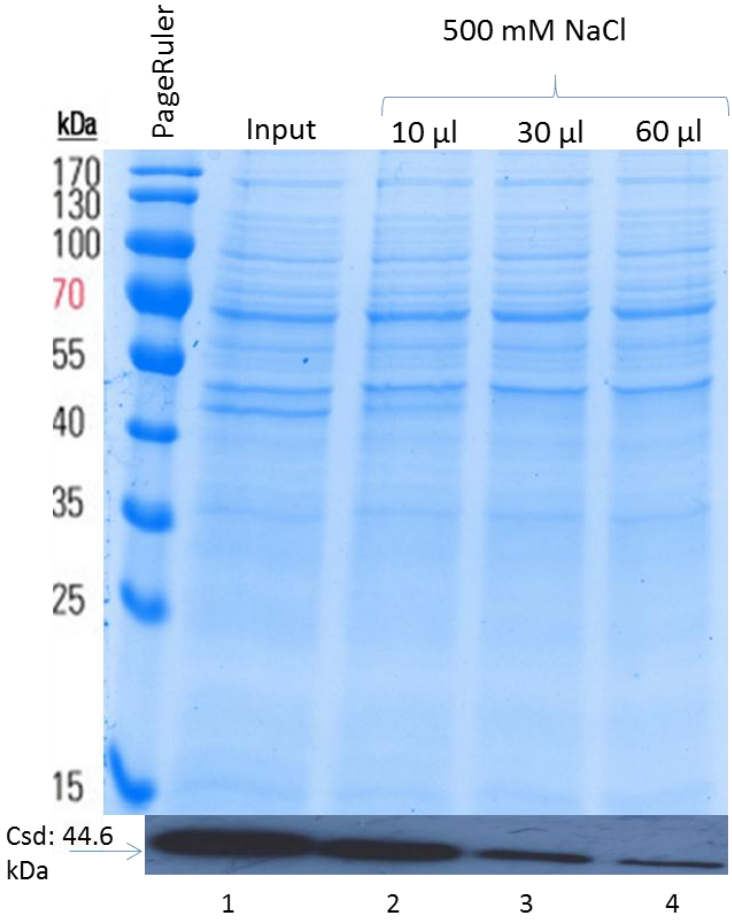


Figure 4.35. Optimisation of the affinity purification conditions. 12% SDS-PAGE gel showing the supernatant fraction (unbound Csd) using 10-, 30-, and 60 µl Ni-particles under high (500 mM) salt conditions (above). Western blot showing the change in residual Csd under the above mentioned conditions for the supernatant fraction (below). Lane 1: Input sample (before enrichment)

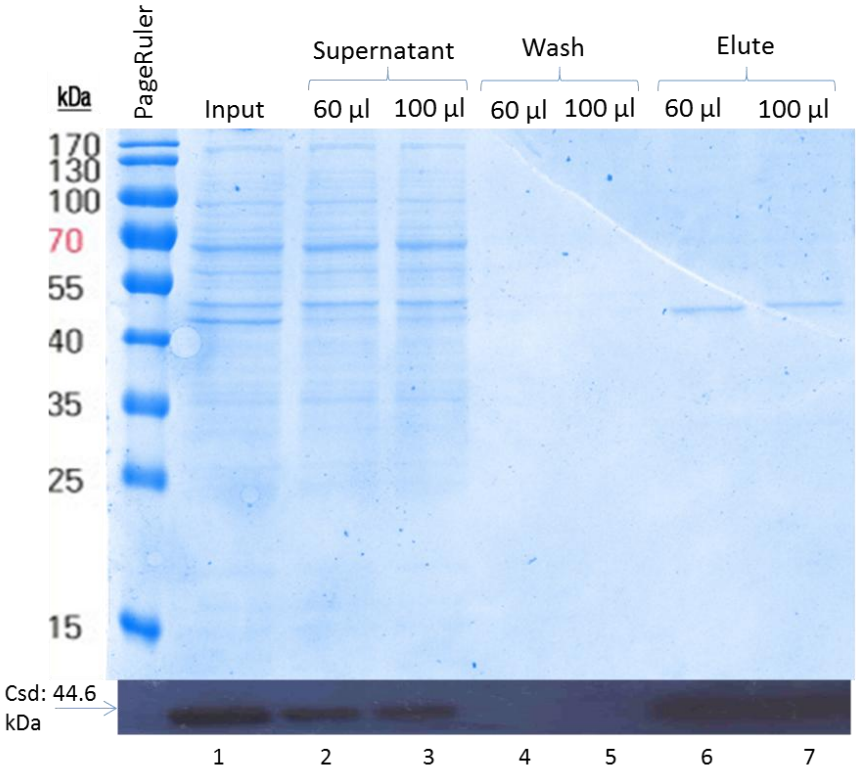


Figure 4.36. Optimisation of the affinity purification conditions. 12% SDS-PAGE gel showing the supernatant, wash and elute after enrichment for Csd-associated proteins using 60- and 100 µl Ni-particles under low (100 mM) salt conditions (above). Western blot showing the change in Csd expression under the above mentioned conditions (below). Lane 1: Input sample (before enrichment)

4.2.3.6. Affinity purification followed by mass spectrometry

M. smegmatis groEL1ΔC derivatives were cultured under standard conditions and affinity purification performed for each strain in 3 biological replicates as optimised in section 4.2.3.4. Samples were subsequently run into a NuPAGE gel in order to remove contaminants, such as polymers leeched from plastics during sample preparation, detergents used in buffers and imidazole used for affinity purification. Gel slices were excised and sent for LC-MS/MS analysis. Correlation between replicates was assessed and positive correlation with a high Pearson correlation coefficient was observed between replicates of the same group. The Pearson correlation coefficient of replicates between different groups was lower than that observed between replicates in the same group, indicating high reproducibility of affinity purification (Fig. 4.37). Furthermore, principle component analysis revealed three distinct clusters, with each biological set clustering

together. This confirms high reproducibility of the method and that affinity purifications, using mCherry, SufC and Csd as bait proteins respectively, can be separated according to the proteins detected by AP-MS (Fig. 4.38).

Venn diagram analysis revealed 17 common identified proteins in the stringent dataset and 18 in the intermediate stringency dataset between proteins identified from the Csd and SufC affinity purifications (Fig. 4.39 – 4.40). The SufC affinity purification accounted for the bulk of the proteins identified. Gene ontology enrichment revealed that SufC associated with many enzymes, some of which are known to be Fe-S cluster containing enzymes, such as *lipA* which encodes a lipoic acid synthetase, *ispG* which encodes 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase and *rlmN* which encodes a probable dual-specificity RNA methyltransferase (Table 4.5). Identified proteins map to a variety of biological processes, the most represented being the regulation of transcription and translation. Other biological processes represented by this dataset include DNA repair, response to oxidative stress, protein refolding and metal ion transport. Many uncharacterised proteins were also identified. STRING networks were constructed using all the prediction methods available which show known or predicted protein-protein associations, with many identified proteins that have no annotated association with other proteins in this network (Fig. 4.41 – 4.42).

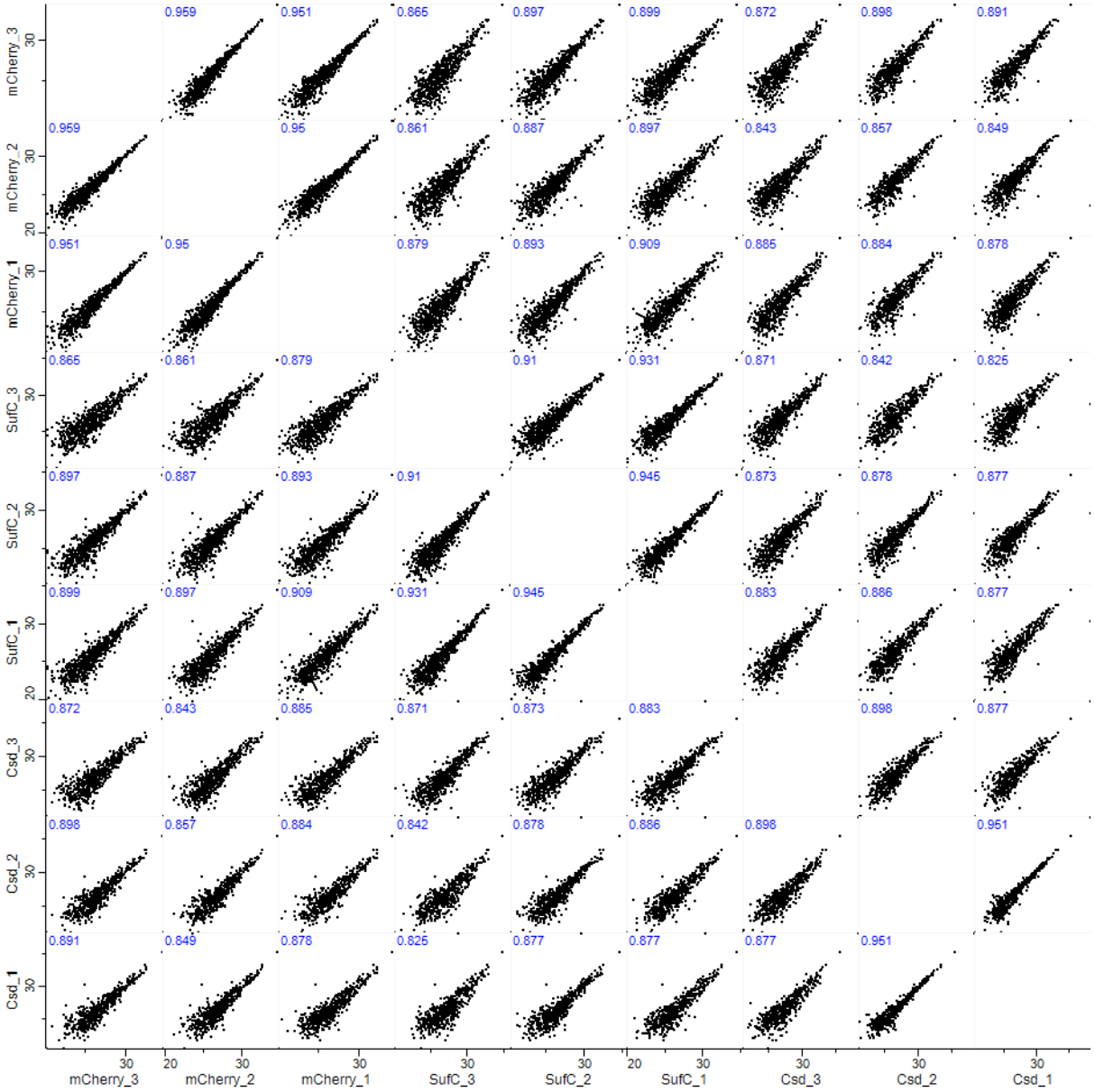


Figure 4.37. Multi-scatter plot showing biological replicates with Pearson correlation coefficient.

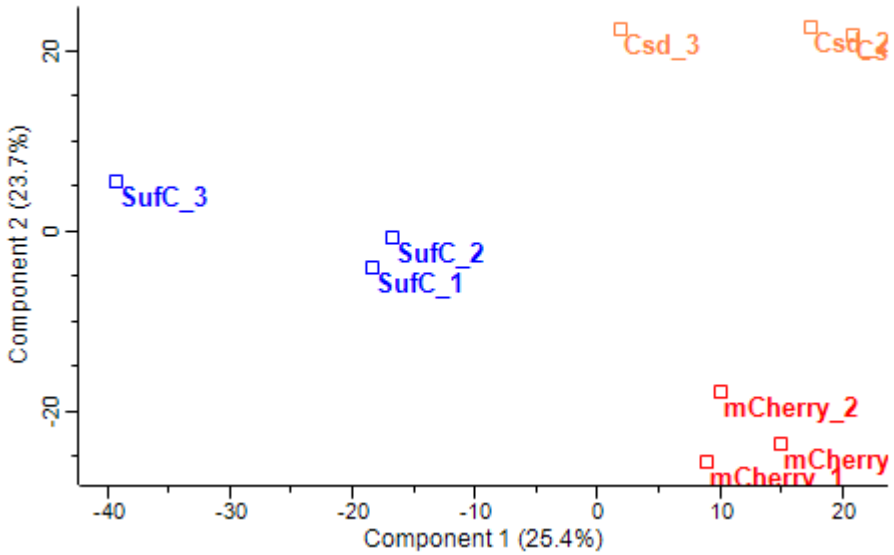


Figure 4.38. Principle component analysis (PCA). PCA of candidate interactors identified by AP-MS from the biological triplicate set, when using Csd, mCherry and SufC as bait proteins for affinity purification, respectively.

Table 4.5. Gene ontology (Molecular function) of proteins enriched by affinity purification

					High stringency		Intermediate stringency	
Entry	**Protein names**	**Gene names**	**Gene ontology (molecular function)**	**Gene ontology (biological process)**	**Csd**	**SufC**	**Csd**	**SufC**
A0QNP9	Rhamnolipids biosynthesis 3-oxoacyl-[acyl-carrier-protein] reductase (Short-chain dehydrogenase/reductase SDR)	MSMEG_0121, MSMEI_0117	3-oxoacyl-[acyl-carrier-protein] reductase (NADPH) activity			✓		✓
A0R075	Sulfur insertion protein LipA	lipA, MSMEG_4286, MSMEI_4185	4 iron, 4 sulfur cluster binding; lipoate synthase activity; metal ion binding	protein lipoylation				✓
A0QVE4	Probable dual-specificity RNA methyltransferase RlmN	rlmN, MSMEG_2545, MSMEI_2485	4 iron, 4 sulfur cluster binding; metal ion binding; rRNA (adenine-C2-)-methyltransferase activity; rRNA binding; tRNA (adenine-C2-)-methyltransferase activity; tRNA binding	rRNA base methylation				✓
I7G6W7	FO synthase	fbiC, MSMEI_4998	4 iron, 4 sulfur cluster binding; transferase activity, transferring alkyl or aryl (other than methyl) groups	coenzyme biosynthetic process				✓
A0QVH9	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase (flavodoxin)	ispG, MSMEG_2580, MSMEI_2518	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase activity; 4 iron, 4 sulfur cluster binding; iron ion binding	isopentenyl diphosphate biosynthetic process, methylerythritol 4-phosphate pathway; terpenoid biosynthetic process		✓		✓
A0QRG9	Aconitate hydratase	MSMEG_1112, MSMEI_1081	aconitate hydratase activity			✓		✓

					High stringency		Intermediate stringency	
Entry	**Protein names**	**Gene names**	**Gene ontology (molecular function)**	**Gene ontology (biological process)**	**Csd**	**SufC**	**Csd**	**SufC**
Q3L887	Acyl-CoA dehydrogenase	fadE5, MSMEG_0406, MSMEI_0396	acyl-CoA dehydrogenase activity; butyryl-CoA dehydrogenase activity			✓		✓
A0QXA4	Acyl-CoA thioesterase II	tesB-1, MSMEG_3228, MSMEI_3146	acyl-CoA hydrolase activity	acyl-CoA metabolic process				✓
A0QQH7	Adenylosuccinate synthetase (AMPSase)	purA, MSMEG_0759, MSMEI_0743	adenylosuccinate synthase activity; GTP binding; magnesium ion binding	'de novo' AMP biosynthetic process		✓		✓
A0QT46	LacI-type transcriptional regulator (Ribose operon repressor, putative)	MSMEG_1708, MSMEI_1668	alanine racemase activity; DNA binding; transcription factor activity, sequence-specific DNA binding	transcription, DNA-templated		✓		✓
Q50441	Alkyl hydroperoxide reductase AhpD	ahpD, MSMEG_4890, MSMEI_4765	alkyl hydroperoxide reductase activity; hydroperoxide reductase activity; peroxidase activity; peroxiredoxin activity	response to oxidative stress		✓		✓
A0R5N8	Aspartokinase	ask, MSMEG_6257, MSMEI_6095	amino acid binding; aspartate kinase activity	lysine biosynthetic process via diaminopimelate; threonine biosynthetic process	✓	✓	✓	✓
A0R221	Homoserine dehydrogenase	MSMEG_4957, MSMEI_4830	amino acid binding; homoserine dehydrogenase activity; NADP binding	isoleucine biosynthetic process; methionine biosynthetic process; threonine biosynthetic process	✓	✓	✓	✓

					High stringency		Intermediate stringency	
Entry	**Protein names**	**Gene names**	**Gene ontology (molecular function)**	**Gene ontology (biological process)**	**Csd**	**SufC**	**Csd**	**SufC**
A0QRA0	Phosphoserine phosphatase (serB2)	serB, serB2, MSMEG_1041, MSMEG_2321, MSMEI_1012, MSMEI_2263	amino acid binding; phosphoserine phosphatase activity	L-serine biosynthetic process		✓		✓
A0QVI1	Uncharacterized protein	MSMEG_2582	antioxidant activity; oxidoreductase activity			✓		✓
A0QQU5	60 kDa chaperonin 1 (GroEL protein 1) (Protein Cpn60 1)	groL1, groEL1, MSMEG_0880, MSMEI_0859	ATP binding	protein refolding	✓	✓	✓	✓
A0QSS4	60 kDa chaperonin 2 (GroEL protein 2) (Protein Cpn60 2)	groL2, groEL2, MSMEG_1583, MSMEI_1545	ATP binding	protein refolding	✓	✓	✓	✓
A0R277	Copper-translocating P-type ATPase	MSMEG_5014, MSMEI_4886	ATP binding; copper-exporting ATPase activity; metal ion binding			✓		✓
I7GAD8	Probable primosomal protein N' (ATP-dependent helicase PriA)	priA, MSMEI_2985	ATP binding; ATP-dependent DNA helicase activity; DNA binding; zinc ion binding	DNA replication, synthesis of RNA primer; DNA unwinding involved in DNA replication			✓	
A0R563	DNA repair protein RadA homolog	radA, MSMEG_6079, MSMEI_5919	ATP binding; damaged DNA binding; DNA-dependent ATPase activity; metal ion binding	DNA repair				✓
A0QQT2	DNA or RNA helicase of superfamily protein II	MSMEG_0866	ATP binding; DNA binding; helicase activity				✓	
A0R4H1	Phosphoribosylformylglycinamidine synthase subunit PurQ	purQ, MSMEG_5831, MSMEI_5673	ATP binding; hydrolase activity; phosphoribosylformylglycinamidine synthase activity	de novo' IMP biosynthetic process; glutamine metabolic process				✓

					High stringency		Intermediate stringency	
Entry	**Protein names**	**Gene names**	**Gene ontology (molecular function)**	**Gene ontology (biological process)**	**Csd**	**SufC**	**Csd**	**SufC**
A0R4C0	Phosphate import ATP-binding protein PstB	pstB, mtp1, pstB1, MSMEG_5779, MSMEI_5626	ATP binding; inorganic phosphate transmembrane transporter activity; phosphate ion transmembrane-transporting ATPase activity				✓	✓
A0QWR0	Shikimate kinase	aroK, MSMEG_3031, MSMEI_2956	ATP binding; magnesium ion binding; shikimate kinase activity	aromatic amino acid family biosynthetic process; chorismate biosynthetic process		✓		✓
I7G934	Histidine kinase	phoR, MSMEI_5712	ATP binding; phosphorelay sensor kinase activity					✓
A0QWZ6	ABC transporter ATP-binding protein	MSMEG_3119, MSMEI_3040	ATPase activity; ATP binding			✓		✓
A0R057	Cytochrome c oxidase subunit 2	ctaC, MSMEG_4268, MSMEI_4167	copper ion binding; cytochrome-c oxidase activity	electron transport chain				✓
A0R5R5	Cyclopropane fatty-acyl-phospholipid synthase	MSMEG_6284, MSMEI_6119	cyclopropane-fatty-acyl-phospholipid synthase activity	lipid biosynthetic process	✓		✓	
A0R3P4	DNA glycosylase (Formamidopyrimidine-DNA glycosylase)	MSMEG_5545, MSMEI_5392	damaged DNA binding; DNA-(apurinic or apyrimidinic site) lyase activity; hydrolase activity, hydrolyzing N-glycosyl compounds; zinc ion binding	base-excision repair; nucleotide-excision repair				✓
A0QQW8	Dihydrolipoyl dehydrogenase	lpdA, lpd, MSMEG_0903, MSMEI_0882	dihydrolipoyl dehydrogenase activity; flavin adenine dinucleotide binding	cell redox homeostasis; glycolytic process				✓

					High stringency		Intermediate stringency	
Entry	**Protein names**	**Gene names**	**Gene ontology (molecular function)**	**Gene ontology (biological process)**	**Csd**	**SufC**	**Csd**	**SufC**
A0R4F1	Transcriptional regulator, TetR family protein	MSMEG_5811	DNA binding	regulation of transcription, DNA-templated; transcription, DNA-templated		✓		✓
A0QR26	TetR-family protein transcriptional regulator	MSMEG_0962, MSMEI_0936	DNA binding	regulation of transcription, DNA-templated; transcription, DNA-templated				✓
I7F6P9	Transcriptional regulator, TetR family	MSMEI_0795	DNA binding	regulation of transcription, DNA-templated; transcription, DNA-templated				✓
A0QSL8	DNA-directed RNA polymerase subunit alpha	rpoA, MSMEG_1524, MSMEI_1488	DNA binding; DNA-directed RNA polymerase activity	transcription, DNA-templated	✓		✓	
A0R4B5	tRNA-dihydrouridine synthase	dus, MSMEG_5774, MSMEI_5620	flavin adenine dinucleotide binding; tRNA dihydrouridine synthase activity					✓
A0R043	GDP-mannose-dependent alpha-(1-6)-phosphatidylinositol monomannoside mannosyltransferase	pimB, MSMEG_4253, MSMEI_4154	glycolipid 6-alpha-mannosyltransferase activity; phosphatidylinositol alpha-mannosyltransferase activity	glycolipid biosynthetic process; mannosylation; pathogenesis; phosphatidylinositol metabolic process; phospholipid biosynthetic process		✓		✓

					High stringency		Intermediate stringency	
Entry	**Protein names**	**Gene names**	**Gene ontology (molecular function)**	**Gene ontology (biological process)**	**Csd**	**SufC**	**Csd**	**SufC**
A0QVC8	D-isomer specific 2-hydroxyacid dehydrogenase NAD-binding protein	MSMEG_2529, MSMEI_2469	glyoxylate reductase activity; NAD binding					✓
A0QS51	Hydrolase, alpha/beta fold family protein (Lipase/esterase lipG)	lipG, MSMEG_1352, MSMEI_1314	hydrolase activity			✓		✓
A0QW04	Inositol-1-monophosphatase	suhB, MSMEG_2762, MSMEI_2692	inositol monophosphate 1-phosphatase activity; inositol monophosphate 3-phosphatase activity; inositol monophosphate 4-phosphatase activity	phosphatidylinositol phosphorylation				✓
A0QS45	50S ribosomal protein L11	rplK, MSMEG_1346, MSMEI_1308	large ribosomal subunit rRNA binding; structural constituent of ribosome	translation	✓		✓	
A0R749	Uncharacterized protein	MSMEG_6779, MSMEI_6597	metal ion binding					✓
A0R2N3	NAD-dependent protein deacylase Sir2	sir2, cobB, npdA, MSMEG_5175, MSMEI_5041	metal ion binding; NAD+ binding; protein-malonyllysine demalonylase activity; protein-succinyllysine desuccinylase activity	cellular response to ionizing radiation; double-strand break repair via nonhomologous end joining		✓		✓
A0QR94	Ribonucleoside-diphosphate reductase subunit beta	nrdF2, MSMEG_1033, MSMEG_2313, MSMEI_1002, MSMEI_2253	metal ion binding; ribonucleoside-diphosphate reductase activity, thioredoxin disulfide as acceptor	deoxyribonucleoside diphosphate metabolic process; deoxyribonucleotide biosynthetic process; DNA replication		✓		✓

					High stringency		Intermediate stringency	
Entry	**Protein names**	**Gene names**	**Gene ontology (molecular function)**	**Gene ontology (biological process)**	**Csd**	**SufC**	**Csd**	**SufC**
A0R3T6	Divalent metal cation transporter MntH	mntH, MSMEG_5589, MSMEI_5438	metal ion binding; symporter activity	metal ion transport			✓	
A0QRH1	Demethylmenaquinone methyltransferase	menG, MSMEG_1115, MSMEI_1083	methyltransferase activity	menaquinone biosynthetic process				✓
A0QRA9	SAM-dependent methyltransferase	MSMEG_1050, MSMEG_2330, MSMEI_1020, MSMEI_2271	methyltransferase activity			✓		✓
A0QRA8	Methyltransferase type 11 (UbiE/COQ5 family protein)	MSMEG_1049, MSMEG_2329, MSMEI_1019, MSMEI_2270	methyltransferase activity			✓		✓
A0QNN7	NAD(P) transhydrogenase subunit beta (Nicotinamide nucleotide transhydrogenase subunit beta)	pntB, MSMEG_0109, MSMEI_0105	NAD(P)+ transhydrogenase (AB-specific) activity; NADP binding			✓		✓
I7FRT7	*Mycobacterium tuberculosis* paralogous family 11	MSMEI_5079	oxidoreductase activity		✓	✓	✓	✓
A0R1G3	Clavaldehyde dehydrogenase	MSMEG_4742	oxidoreductase activity				✓	
I7GA62	Uncharacterized protein	MSMEI_3573	oxidoreductase activity			✓		✓
A0QXW7	Short-chain dehydrogenase/reductase	MSMEG_3451, MSMEI_3370	oxidoreductase activity					✓
A0QQF9	Uncharacterized protein	MSMEG_0741, MSMEI_0725	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen			✓		✓

					High stringency		Intermediate stringency	
Entry	**Protein names**	**Gene names**	**Gene ontology (molecular function)**	**Gene ontology (biological process)**	**Csd**	**SufC**	**Csd**	**SufC**
A0QNG3	Penicillin-binding protein A (PBPA)	pbpA, MSMEG_0031, MSMEI_0033	penicillin binding; transferase activity	cell wall organization; peptidoglycan biosynthetic process; regulation of cell shape				✓
I7FGI9	Pseudouridylate synthase	truA, MSMEI_1490	pseudouridine synthase activity; RNA binding	tRNA pseudouridine synthesis	✓	✓	✓	✓
A0R220	Threonine synthase	thrC, MSMEG_4956, MSMEI_4829	pyridoxal phosphate binding; threonine synthase activity	threonine biosynthetic process				✓
I7FJ98	RNA-binding protein (Contains KH domain)-like protein	MSMEI_2375	RNA binding				✓	
A0R151	50S ribosomal protein L21	rplU, MSMEG_4625, MSMEI_4508	rRNA bindin; structural constituent of ribosome	translation	✓	✓	✓	✓
A0QS46	50S ribosomal protein L1	rplA, MSMEG_1347, MSMEI_1309	rRNA binding; structural constituent of ribosome; tRNA binding	regulation of translation; translation	✓	✓	✓	✓
A0R6N9	Alcohol dehydrogenase GroES-like protein	MSMEG_6616, MSMEI_6438	S-(hydroxymethyl)glutathione dehydrogenase activity; zinc ion binding					✓
A0QWX9	Transaldolase	tal, MSMEG_3102, MSMEI_3024	sedoheptulose-7-phosphate:D-glyceraldehyde-3-phosphate glyceronetransferase activity	carbohydrate metabolic process; pentose-phosphate shunt				✓
I7G2S0	Transcriptional regulator, XRE family	MSMEI_0253	sequence-specific DNA binding		✓	✓	✓	✓

					High stringency		Intermediate stringency	
Entry	**Protein names**	**Gene names**	**Gene ontology (molecular function)**	**Gene ontology (biological process)**	**Csd**	**SufC**	**Csd**	**SufC**
A0QSP9	30S ribosomal protein S9	rpsI, MSMEG_1557, MSMEI_1520	structural constituent of ribosome	translation	✓	✓	✓	✓
A0R3B7	Septum formation initiator subfamily protein, putative	MSMEG_5414		cell cycle		✓		✓
I7G6G4	ChaB family protein	MSMEI_1757		DNA-templated transcription, termination		✓		✓
A0QQP0	Uncharacterized protein	MSMEG_0824, MSMEI_0805			✓	✓	✓	✓
A0QRV5	Uncharacterized protein	MSMEG_1252			✓		✓	
A0R576	Lsr2 protein	lsr2, MSMEG_6092, MSMEI_5934			✓		✓	
A0QSQ8	Uncharacterized protein	MSMEG_1567, MSMEI_1530			✓	✓	✓	✓
A0R595	Uncharacterized protein	MSMEG_6112, MSMEI_5953				✓		✓
A0QRT9	Mpr protein (Uncharacterized protein)	MSMEG_1236, MSMEI_1202				✓		✓
A0R4D0	Uncharacterized protein	MSMEG_5790, MSMEI_5637						✓

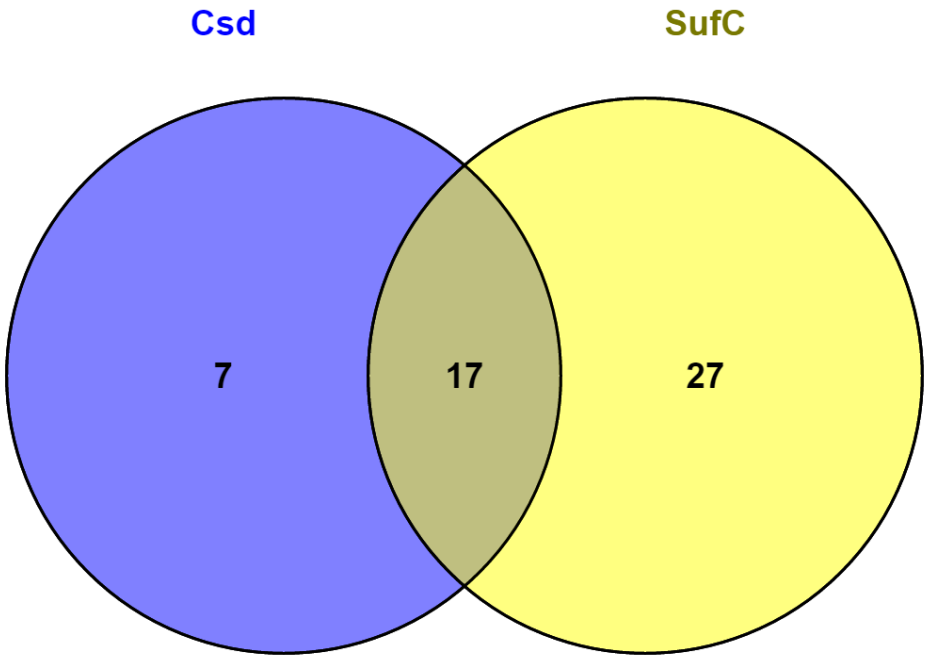


Figure 4.39. Venn diagram showing the proteins that are unique to and common between the Csd and SufC affinity purifications for the high-stringency list.

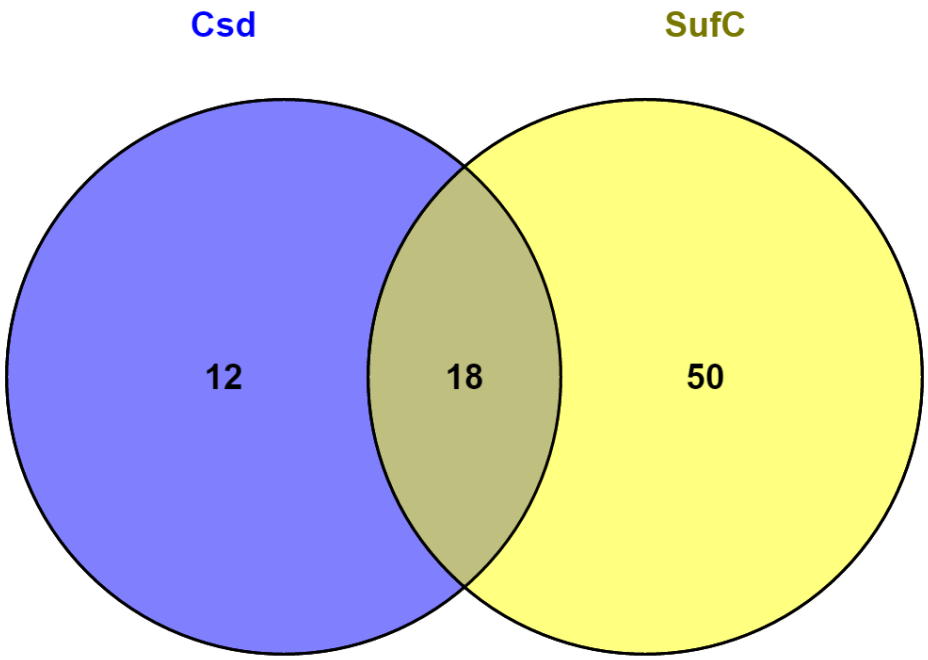


Figure 4.40. Venn diagram showing the proteins that are unique to and common between the Csd and SufC affinity purifications for the intermediate-stringency list.

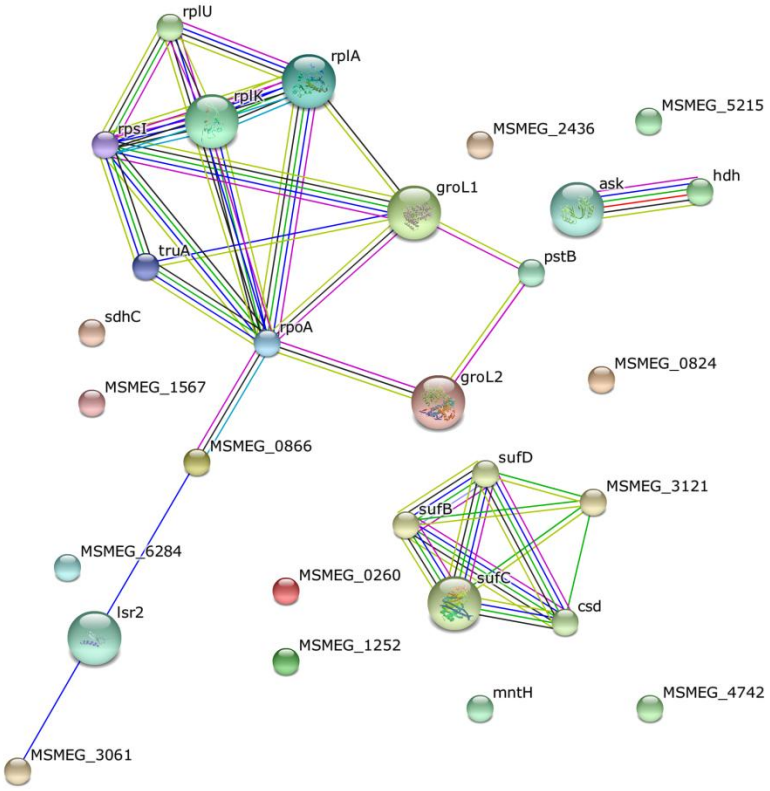


Figure 4.41. Protein association network of proteins identified when using Csd as bait protein for affinity purification (Intermediate stringency).

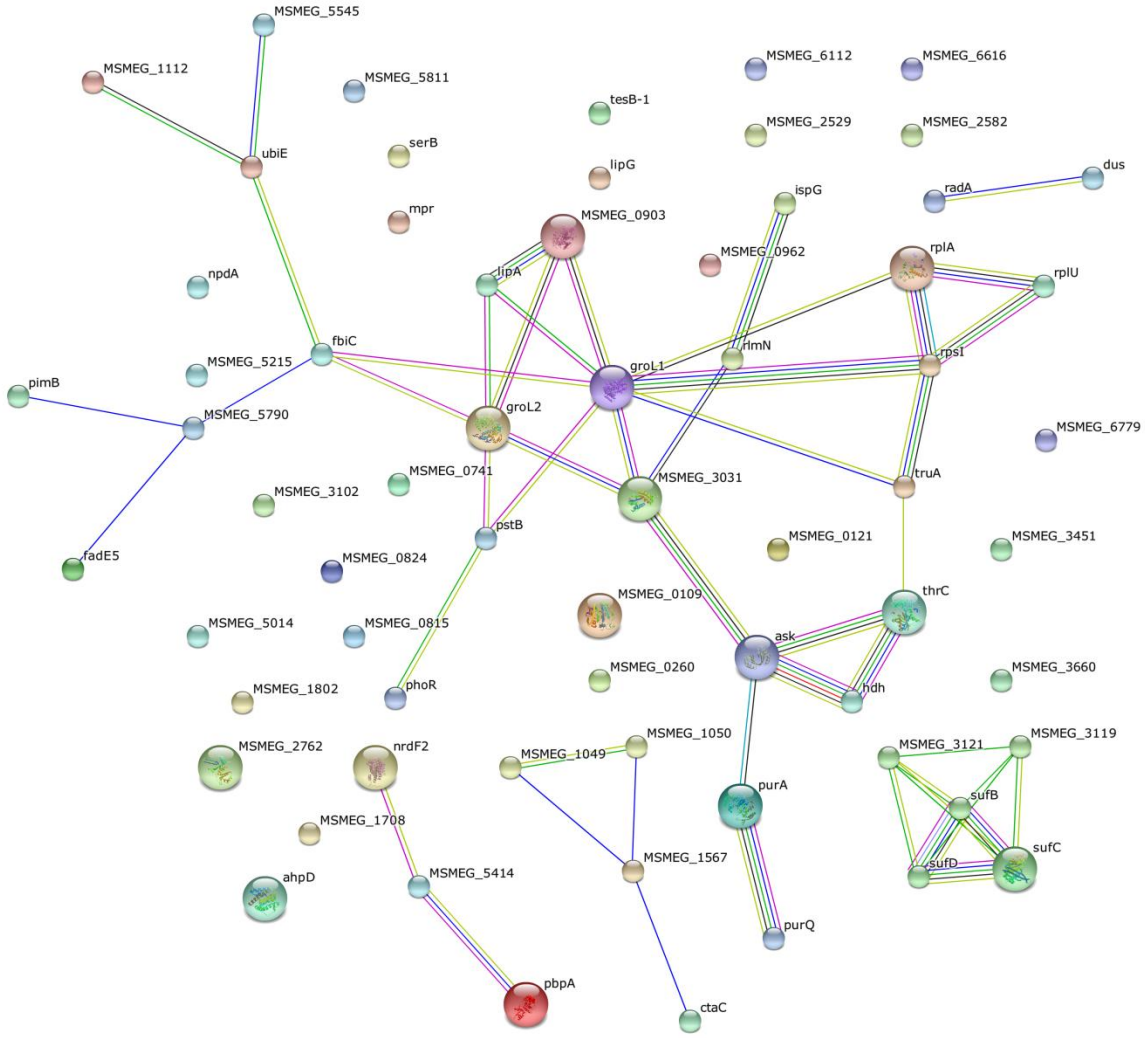


Figure 4.42. Protein association network of proteins identified when using SufC as bait protein for affinity purification (Intermediate stringency).

Chapter 5: Discussion, future work and concluding remarks

5.1. Discussion

Mycobacterium tuberculosis remains a global threat, and our lack of understanding of the physiology of this organism is an ongoing challenge in the fight against TB. Protein-protein interactions are required for a variety of biological processes, such as regulating gene expression, mediating the immune response and maintaining structural stability[5]. New techniques for studying PPIs are needed to gain a holistic view of the physiology of this pathogen. Fe-S clusters are ubiquitous cofactors required by many proteins involved in diverse biological processes[123]. The main machinery responsible for the assembly of Fe-S clusters is predicted to be essential for the survival of *M. tuberculosis*[34,35,126]. However, the exact mechanisms of Fe-S cluster assembly in mycobacteria are poorly understood. This study aimed to develop a method for the analysis of cytoplasmic complexes in mycobacteria and subsequent application of this method for the identification of novel proteins involved in Fe-S cluster assembly. Here we use blue native-PAGE and affinity purification followed by mass spectrometry (AP-MS) for the analysis of mycobacterial cytosolic protein complexes, specifically to identify novel proteins involved in iron-sulphur cluster biogenesis.

The mycobacterial cell wall consists of three primary macromolecules, namely peptidoglycan, arabinogalactan and mycolic acids[144–146]. These three layers are encased in a capsule comprised of proteins and polysaccharides. The arabinan-fused long-carbon-chain mycolic acids form a thick, waxy and lipid-dense cell wall[147]. The complexity of the mycobacterial cell envelope therefore makes it difficult to rupture. BN-PAGE was optimised for the analysis of cytosolic protein complexes extracted from *M. bovis* BCG. We compared two physical lysis methods routinely used for lysis of mycobacteria, namely bead-beating and sonication, to determine their compatibility with BN-PAGE. Bead-beating and sonication are relatively harsh cell lysis techniques and result in a significant amount of foaming which could lead to protein denaturation resulting in protein loss. These harsh cell lysis techniques could also potentially disrupt native complex interactions. Freezing cell pellets prior to lysis weakens the cell wall and facilitates more efficient protein extraction. Pre-lysis lysozyme treatment could also be used to weaken the cell wall due to its ability to hydrolyse polysaccharides in the capsule and break down the peptidoglycan backbone[148], thereby permeabilizing the mycobacterial cell wall and allowing for less harsh sonication to be used.

Due to the complexity of protein samples and the fact that BN-PAGE analysis is a crude, low-resolution technique, it was difficult to assess the effect of these lysis techniques on the integrity of the protein complexes.

Buffer exchange was performed to assess whether the salt concentration in the sample had an effect on BN-PAGE analysis. Although no apparent effect could be observed, again, this is difficult to assess due to the low resolution of BN-PAGE, particularly when evaluating a very complex sample. To overcome this limitation, mass spectrometry can be used to evaluate the integrity of a well characterised protein complex in response to dialysis or different salt concentrations, although this can be costly. Most cell lysis methods result in the release of nucleic acids which associate either specifically or non-specifically with proteins in the sample. DNA present in protein lysates could result in increased viscosity, causing sample streaking during BN-PAGE analysis. This streaking of protein samples during electrophoresis could interfere with BN-PAGE analysis and DNA-protein association can also interfere with proteomic analysis[149]. Sonication is a harsh cell lysis technique and contributes to the shearing of DNA, but does not remove the entire DNA present in the sample. DNase I treatment appeared to eliminate sample streaking, thereby increasing the BN-PAGE resolution. Protein precipitation techniques are also commonly used for the removal of DNA from protein samples; however, protein precipitation could compromise the integrity of complex associations and might not be suitable for BN-PAGE.

Having optimized cell lysis, we then examined whether we were able to recover intact protein complexes. We used the well-characterised RNA polymerase complex for this purpose. We demonstrated that protein complexes analysed by BN-PAGE could be immunodetected and that protein complexes could be separated into individual components by 2D BN/SDS-PAGE, and individual components subsequently immunodetected. Four of the six components of the RNA polymerase complex, namely RpoA (38 kDa), RpoB (129 kDa), RpoC (147 kDa) and SigA (58 kDa) could be identified by LC-MS/MS. RpoZ (12 kDa) and SigH (24 kDa) could however not be identified. This is possibly due to the fact that these proteins are the smallest proteins in the complex. Smaller proteins are less efficiently detected by MS/MS in complex samples since trypsin digestion yields fewer peptides and this signal could be masked by the larger proteins in the complex that have more peptides liberated after proteolysis. This problem could be overcome by

fractionating the complex into its individual components using 2D BN/SDS-PAGE and then analysing these components individually by mass spectrometry. It is important to note that detection of peptides is also dependent on the ionisation properties of peptides. Many other proteins were also identified in the gel fraction. STRING analysis predicted that 41 of the 44 proteins identified form part of the same interaction network. It is tempting to speculate that these proteins are interacting with components of RNA polymerase, however due to the complexity of the sample it is more likely that multiple complexes are present at the same position in the gel. Fractionation techniques such as sucrose density gradient separation, gel filtration or chromatographic separation could be used to decrease sample complexity and decrease possible background. BN-PAGE for the identification of protein complexes is therefore best utilised in conjunction with other techniques, such as affinity purification, to reduce sample complexity and co-migration of complexes.

Fe-S cluster-containing proteins can lose their clusters in the presence of oxidative and nitrosative stress. Fe-S cluster assembly machinery is then upregulated to compensate for this loss[150–153]. Here we attempted to identify Fe-S cluster assembly proteins by upregulation of this complex in response to oxidative and nitrosative stress. We showed that no visible difference in banding pattern could be observed after *M. bovis* BCG exposure to oxidative and nitrosative stress and subsequent analysis of cytosolic protein complexes by BN-PAGE. Previous studies demonstrating an upregulation of Fe-S cluster assembly machinery after exposure to oxidative and nitrosative stress were evaluated at a transcript level using qRT-PCR. Our inability to detect induction could be as a result of post-transcriptional gene regulation. Another major limitation of BN-PAGE is its relatively low resolution, making it exceedingly problematic to visualise variation in the banding pattern of a complex protein sample. For future work, qRT-PCR should be used as a starting point to confirm upregulation of genes of interest on a transcript level. Fractionation techniques mentioned above could also be used to decrease sample complexity. High-resolution mass spectrometry could then be used to analyse proteins present in these complexes, upregulated after exposure to oxidative and nitrosative stress.

An alternate approach for identifying the Fe-S cluster biogenesis machinery, namely immunodetection was then explored. Since no antibodies are available against proteins of the SUF machinery, the first step in this approach was to express and purify one of these

proteins for antibody generation. We chose to express the cysteine desulphurase, Csd, and three fusion tags of different sizes were selected for this purpose. Fusion tags have previously been shown to facilitate protein expression by increasing the solubility of fusion proteins[69,154]. However, heterologous expression of Csd in *E. coli* was unsuccessful. Heterologous expression of *M. tuberculosis* proteins is a common problem possibly due to the GC-rich nature of the DNA in this organism, and the differences in codon usage compared to that of *E. coli*. *M. tuberculosis* utilizes a higher percentage of alanine, proline, arginine and glycine compared to *E. coli*[155,156]. The machinery responsible for *M. tuberculosis* post-translational modifications is also lacking in *E. coli*, contributing to the likelihood of *M. tuberculosis* protein aggregation when expressed in *E. coli*. Heterologous protein expression of *M. tuberculosis* proteins in *E. coli* is, however, still a favourable approach due to the fast growing nature of *E. coli* and the availability of *E. coli* expression vectors which facilitate a high yield of protein production. Protein expression at low temperatures has been shown to improve protein solubility and decrease heat shock protein expression, proteolytic degradation and protein aggregation[157]. Different *E. coli* strains such as Arctic express or expression vectors, such as pCOLD, co-express cold shock proteins which facilitate soluble expression of *M. tuberculosis* proteins[155]. *M. tuberculosis* protein solubility can also be improved by co-expressing *M. tuberculosis* chaperone proteins, such as the *E. coli* GroEL ortholog, Cpn60.1[158]. Co-expression of proteins within the same metabolic pathway has also been shown to improve protein expression[159–161]. In this case, none of these approaches proved useful and attempts at antibody generation ended here.

Due to the above mentioned problems associated with heterologous *M. tuberculosis* protein expression in *E. coli*, we decided to express Csd and SufC as 6xHis-tag fusions in *M. smegmatis*. These fusion proteins could then be used to detect the Fe-S cluster assembly complex using commercial antibodies raised against fusion tags and to purify this complex by affinity purification. Expressing *M. tuberculosis* proteins in a host more closely related to this organism, such as the fast growing *M. smegmatis*, has been shown to be a successful alternative for proteins which cannot be solubly expressed in *E. coli*[132,156,163,164]. 6xHis-tag fusion proteins were expressed under the control of a tetracycline-responsive promoter. The tetracycline-responsive expression system was used as protein over-expression can be toxic at certain levels and can lead to protein aggregation, promoter mutation and plasmid

loss over time[165]. No change in growth was observed for strains expressing Csd and SufC in the presence or absence of atc, indicating that overexpression of these proteins under these conditions was not toxic to *M. smegmatis*. Soluble protein expression was optimised by testing different growth stages and different atc concentrations.

The fusion proteins were then used to enrich for the Fe-S cluster assembly machinery under native conditions using a nickel IMAC resin. Affinity purification was optimised by evaluating different volumes of nickel beads and different salt concentrations. No difference in Csd recovery was seen between the 60 µl and 100 µl Ni-particle samples for the low salt supernatant fraction, even though a significant proportion of residual Csd could still be detected. This indicates that a proportion of Csd does not bind to the Ni-particles under near-physiological salt concentrations, most likely due to the inaccessibility of the 6xHis-tag under native complex conditions. Larger affinity fusion tags could be explored in an attempt to remedy this problem, although, this could potentially interfere with the native complex structure. A protein's structure strongly influences its interaction potential. Protein surface chemistry is incredibly complex and interactions between proteins occur as a result of hydrogen bonding, electrostatic, ionic, hydrophobic and van der Waals forces[166,167]. The direct effect of salt concentration on these forces is poorly understood. Sodium chloride is a chaotropic salt which disrupts the stability of intramolecular forces responsible for protein-protein interactions. High salt concentrations are commonly used in affinity purification protocols to limit the amount of non-specific protein interaction with the proteins of interest and the resin used. However, high salt concentrations weaken the stability of native protein complexes. Near-physiological salt concentrations are therefore more suitable when enriching for native protein complexes.

In order to identify the Fe-S cluster assembly complex, Csd and SufC fusions were used to isolate this complex by affinity purification and subsequently analyse enriched protein samples by high-resolution mass spectrometry. Exploratory data analysis revealed that the technique was reproducible and that protein level differences existed that could be used to discriminate between affinity purifications using different bait proteins. Enzymes identified by SufC affinity purification accounted for the bulk of the identified proteins. SufC forms part of the Fe-S cluster assembly scaffold and interacts stably with the other scaffold proteins, namely SufD and SufB. SufC is an ATPase and displays homology to ATPase

subunits of membrane transporters although SufC is found in the cytosol[168]. SufC is required for *in vivo* assembly of Fe-S clusters and has been implicated in iron acquisition rather than sulphur transfer. SufC is also thought to stabilize SufB for Fe-S cluster formation on SufB. SufC occurs as a homodimer and was found to be associated with itself in our dataset, along with other known interactors, namely Csd, SufB and SufD. A DNA-binding protein that shares homology with Rv1460 transcriptional regulator was also present in the SufC affinity enrichment. Other ATPases were also present in the SufC affinity enrichment, such as RadA, a DNA repair protein, and MSMEG_5014, involved in copper translocation. This suggests that the ATPase activity function of SufC might not be limited to Fe-S cluster assembly.

A cysteine desulphurase is required for the first step of Fe-S cluster assembly, which catalyses the transfer of sulphur from L-cysteine to a scaffold protein[123,124]. In addition to the cysteine desulphurase present in the *suf* operon (*csd*), the genome of *M. tuberculosis* also encodes 2 other genes with homology to this enzyme, namely *iscS*, and *egtE* (*Rv3700c*)[130,169]. IscS does not interact with any of the proteins in the SUF system, but interacts with Fe-S cluster-containing proteins. Deletion of *iscS* in *M. tuberculosis* resulted in impaired activity of enzymes containing Fe-S clusters, suggesting that IscS is involved in Fe-S cluster assembly[38]. *egtE* encodes a pyridoxal-phosphate-dependent protein which probably catalyses the conversion of hercynylcysteine sulfoxide to ergothioneine. Ergothioneine is an antioxidant secreted by mycobacteria and is involved in protection against oxidative stress[37]. The relationship between these 3 cysteine desulphurases is unknown. Csd interacts transiently with the scaffold and functions as the sulphur donor. It was therefore hypothesised that Csd may function promiscuously as a sulphur donor in interaction with other complexes which require sulphur; however, this did not appear to be the case. Proteins responsible for the biosynthesis of other sulphur-containing molecules which require cysteine desulphurases such as molybdenum cofactor, mycothiol and ergothionine were lacking in our analysis, suggesting that Csd might be a Fe-S cluster-specific sulphur donor.

The association between SufC and known or potential Fe-S cluster containing enzymes suggests a direct interaction between the scaffold and apo-proteins for Fe-S cluster insertion. This would contrast the hypothesis of A-type carrier-mediated Fe-S cluster transfer from the scaffold to apo-proteins, but would be consistent with the hypothesis of

direct cluster transfer from scaffold to apo-protein. A-type carrier-mediated Fe-S cluster transfer is well defined in *E. coli*, and is shown to be essential for the maturation of certain Fe-S cluster-containing proteins[170]. It is currently unknown if carrier proteins are required for Fe-S cluster transfer in mycobacteria. There is also evidence in *E. coli* for direct Fe-S cluster transfer from the scaffold to the apoproteins[170]. A similar scenario has been suggested for the IscS of *M. tuberculosis*. IscS was shown to interact directly with Fe-S cluster containing proteins, although no interaction with the SUF machinery was observed[38]. IscS was also shown to be sufficient for the assembly of the Fe-S cluster in WhiB3, a transcriptional regulator[130]. Proteins that map to biological processes related to protein synthesis and transcriptional regulation were largely overrepresented in association with Csd and SufC. The close proximity of Fe-S cluster assembly proteins and proteins involved in protein synthesis suggest that Fe-S clusters could possibly be inserted as proteins are being synthesized. A similar phenomenon is observed in the molybdenum cofactor biosynthesis pathway where molybdenum cofactor is rapidly acquired by molybdenum-dependent enzymes during the early stages of maturation[171].

Ferrous iron is essential to the survival of most living organisms, but can be toxic at high levels due to Fenton reactions[172,173]. Iron acquisition and storage should therefore be tightly regulated. The iron donor for Fe-S cluster assembly is currently unknown[174,175]. Frataxin has been proposed as the iron donor in eukaryotes[176,177] and its homolog, CyaY, proposed for iron donation in the bacterial ISC system[174]. Due to the fact that the SUF system is functional under iron limited conditions, it has been suggested that the SUF system acts as an “iron trap”, using iron binding proteins, called ferritins, for the storage of iron[178]. In mycobacteria iron is acquired from the environment by siderophores. It is suggested that iron-binding carriers transport iron to where it is needed in the cell. In *Erwinia chrysanthemi*, the SUF system has been implicated in the transfer of iron to bacterioferritin where it is stored until incorporation into apo-proteins is required[178]. Metal ion binding proteins and proteins involved in metal ion transport, such as the divalent metal cation transporter identified in association with Csd, MntH, represents an interesting cohort of proteins to investigate as they may play a direct role in Fe-S cluster assembly by providing Fe^{2+} or by regulating Fe-S cluster assembly or repair.

mCherry did not serve as an appropriate control as known interactors of Csd and SufC, such as SufB and SufD which form part of the scaffold were also identified in the control. Non-specific protein association with mCherry makes eliminating background challenging. This non-specific association with mCherry also lead to much difficulty in choosing an appropriate stringency when filtering for proteins that are likely to interact with highest confidence. It is for this reason that 2 levels of stringency were used. Many interacting partners may have been eliminated due to non-specifically associating with mCherry and the criteria used for high stringency filtering. Choosing an appropriate control for affinity purification techniques of this nature is challenging as one would need to find an unrelated protein that is known not to interact with any of the components of the complex of interest, however this is difficult to predict and can often only be determined empirically.

Although a *M. smegmatis groEL1ΔC* carrying a deletion of the histidine-rich C-terminus in GroEL1 was used for affinity purification, non-specific co-purification of GroEL1 and associated proteins were still present in relatively high abundance. Due to the high abundance of GroEL in mycobacteria, its important function in protein folding, and high-background associated with affinity purification techniques, this was not surprising. This is, however, problematic for protein identification by mass spectrometry as higher abundant proteins can mask the signal of proteins with occurs in low abundance, and true interactions could be masked.

Approximately 50 genes encoding putative Fe-S cluster containing proteins in *M. tuberculosis* were previously predicted by using bioinformatics-based amino acid pattern searches and literature curation[179]. The *M. smegmatis* homologs of these proteins are lacking in our dataset. This could be due to inaccurate prediction or alternative mechanisms of Fe-S cluster insertion, such as carrier protein-mediated Fe-S cluster insertion. Another theory is that the cysteine desulphurase, IscS, the housekeeping Fe-S cluster biogenesis sulphur donor in *E. coli*, could be responsible for direct Fe-S cluster insertion without the need for interaction with the SUF machinery[38]. Many factors influence which interactions are identified, including the stage of bacterial growth at which sampling took place, whether or not interactions are stable or transient and the environmental condition in which interactions take place. Certain growth stages and environmental conditions could be favourable for certain interactions and unfavourable for others.

5.2. Future work and concluding remarks

The work done in this study serves only as an initial screen and possible interacting proteins identified by mass spectrometry in association with Csd and SufC need to be validated using other techniques for detecting protein-protein interactions, such as yeast two-hybrid, protein complementation assays or co-immunoprecipitation. Once interactors have been confirmed, their role in Fe-S cluster biogenesis would need to be established. One approach would be to generate deletion mutants of confirmed non-essential interactors and to evaluate the effect on Fe-S cluster biogenesis. Direct measurement of Fe-S cluster abundance is not possible due to cluster instability. Fe-S cluster-dependent enzyme activity is therefore used as an indirect measurement of Fe-S cluster formation. This method is well defined, using succinate dehydrogenase and aconitase for enzyme activity assays[180].

An alternate approach for interacting partners predicted to be essential in mycobacteria, would be the generation of conditional mutants. Recently, the generation of conditional mutants in mycobacteria by protein depletion was described and appears to be more effective than previous methods[181,182]. This work could be expanded by using binary protein-protein interaction techniques in order to define specific interactions occurring within the Fe-S cluster assembly complex. Soluble expression and purification of Fe-S cluster complex components will allow these interactions to be characterised by *in vitro* biochemical assays. A technique has recently been developed which is based on the fusion of potential Fe-S cluster-containing proteins to a fluorescent probe which can be used to track the transfer of Fe-S clusters between proteins[183]. Although this methodology is still in its infancy, it provides an interesting approach for monitoring Fe-S cluster kinetics by tracking Fe-S cluster flux through different branch points within transfer pathways.

Proteomic analysis of multiprotein complexes is possibly one of the most challenging and daunting techniques of modern day molecular biology due to the overwhelming complexity of the proteome. Few methods for achieving this task are available. The lack of specific antibodies against the Fe-S cluster biogenesis machinery, which could be used for immunodetection and affinity purification, limits the application of these techniques. Antibody generation is costly and labour intensive. The use of affinity tag protein fusions provides a feasible solution to this problem; however, non-specific protein interactions with affinity tags and their relevant resins are common. Using affinity purification together with

BN-PAGE, 2D BN/SDS-PAGE and fractionation techniques would be an efficient and sensitive approach for identifying physiologically relevant protein-protein interactions. For a more comprehensive understanding of the mycobacterial Fe-S cluster interaction network, AP-MS analysis should be evaluated at different stages of growth and under different environmental conditions. This work is an important starting point for elucidating the complete Fe-S cluster biogenesis interaction network in mycobacteria.

Appendices

A.1. Plasmid maps

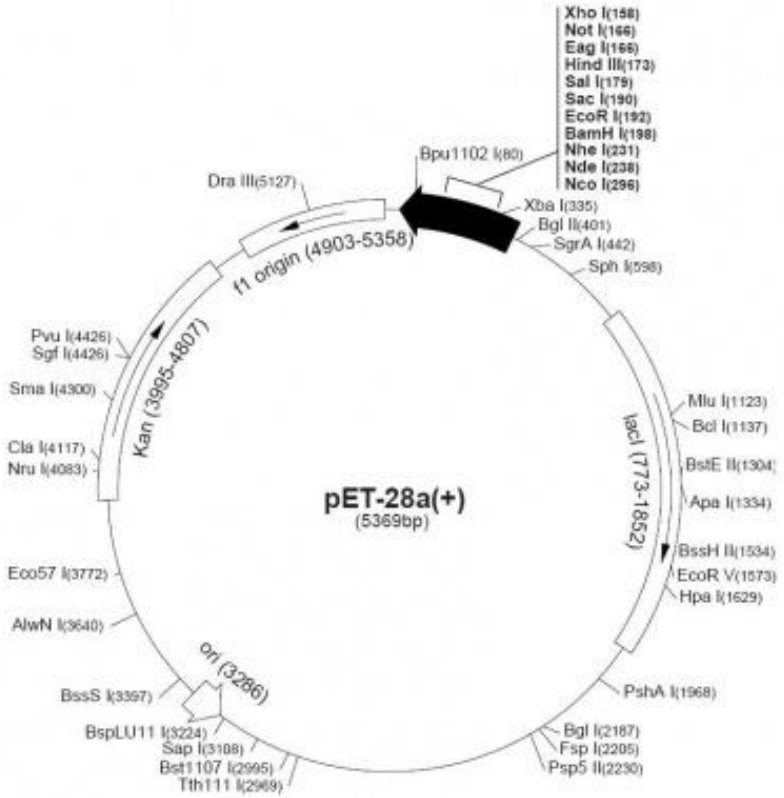


Figure A.1. pET-28a(+) plasmid map (Novagen).

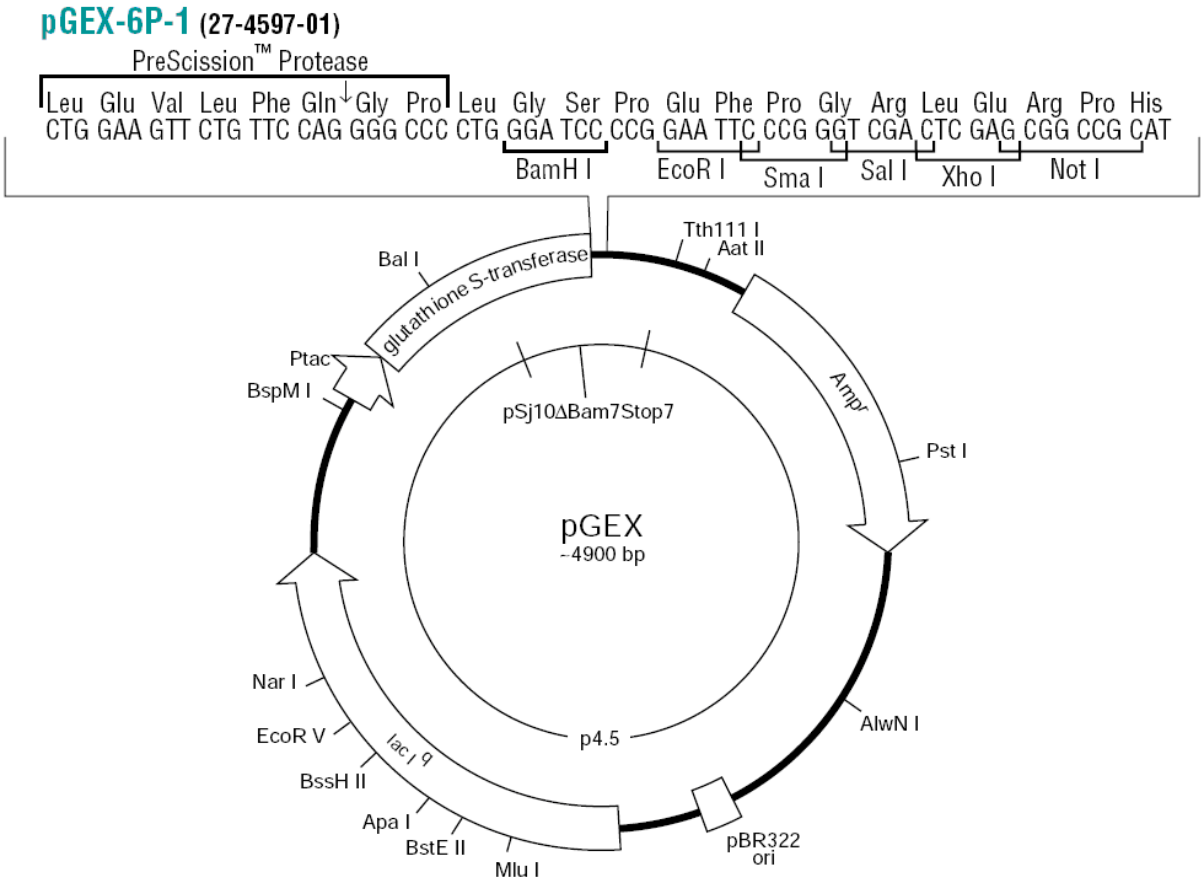


Figure A.2. pGEX-6P-1 vector map (GE Healthcare Life Sciences).

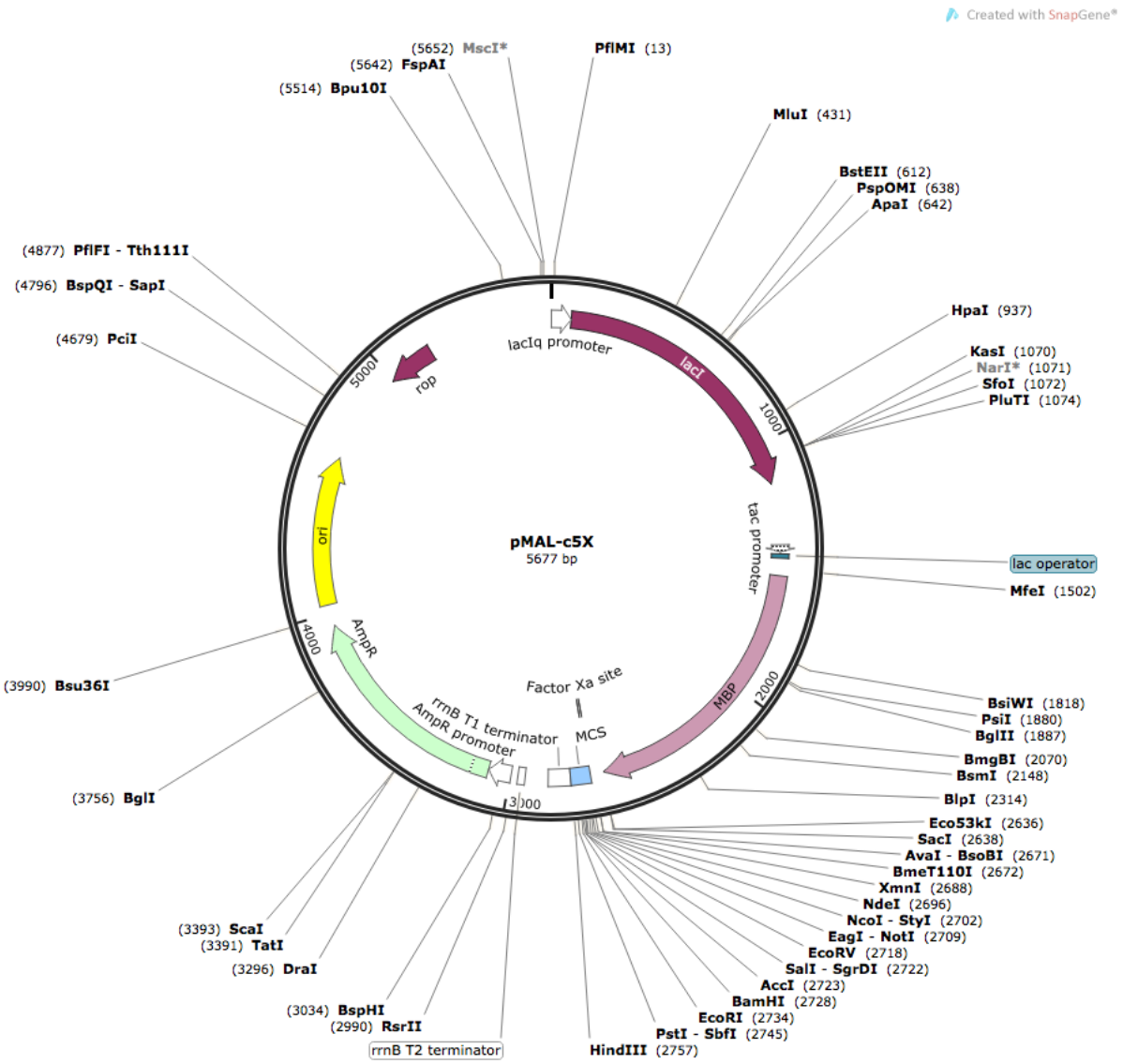


Figure A.3. pMAL-c5X vector map (New England Biolabs).

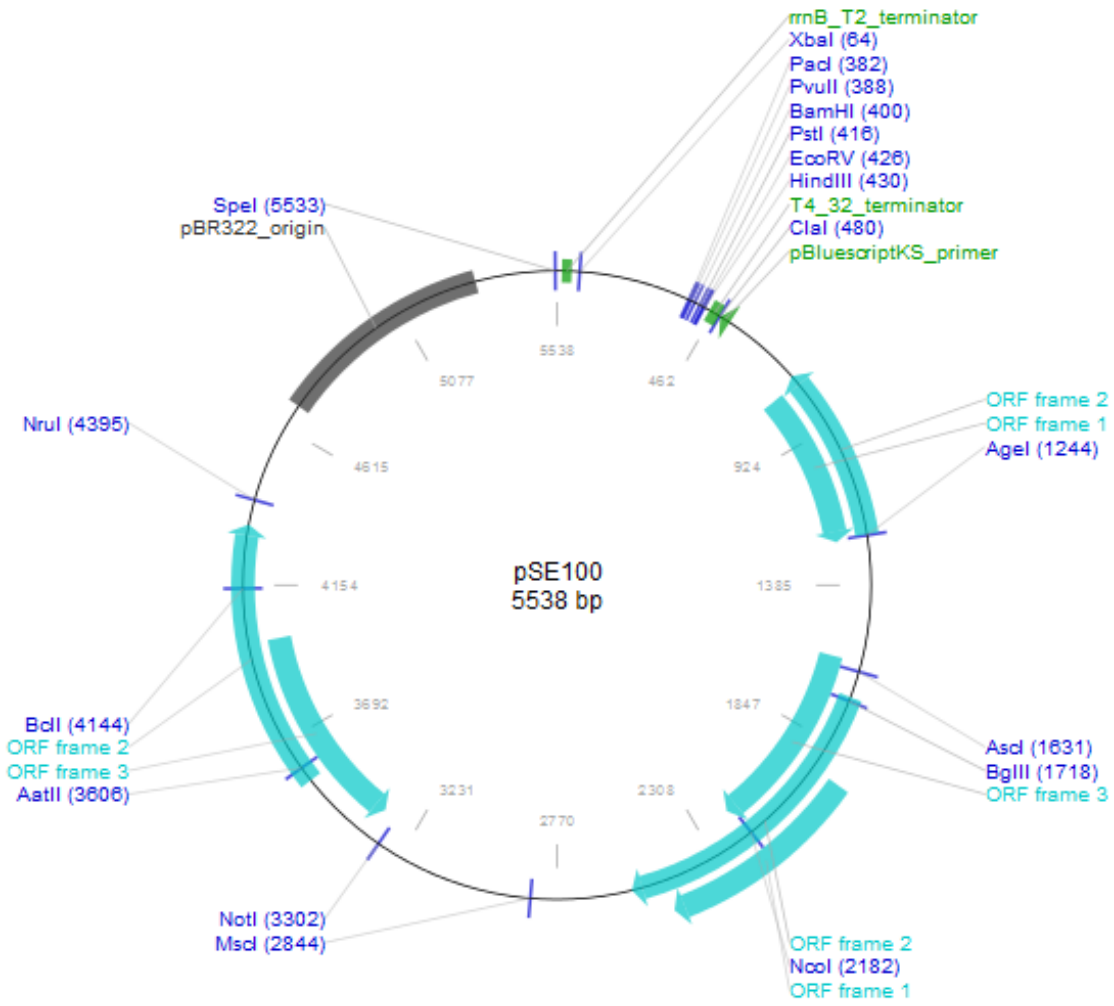


Figure A.4. pSE100 plasmid map (Addgene).

Table A.1. Proteins identified by mass spectrometry from the RNA polymerase gel spot excised from BN-PAGE

				Exclusive unique peptide count	
UniProt Accession	**Gene names**	**Gene ontology (molecular function)**	**Gene ontology (biological process)**	**RNA pol 1**	**RNA pol 2**
A0R5M3	MSMEG_6242, MSMEI_6081	1,3-propanediol dehydrogenase activity; alcohol dehydrogenase (NAD) activity; metal ion binding		2	
A0R2B1	kgd, sucA, MSMEG_5049, MSMEI_4922	2-hydroxy-3-oxoadipate synthase activity; 2-oxoglutarate decarboxylase activity; dihydrolipoyllysine-residue succinyltransferase activity; metal ion binding; oxoglutarate dehydrogenase (succinyl-transferring) activity; thiamine pyrophosphate binding	tricarboxylic acid cycle	7	4
A0R171	MSMEG_4646, MSMEI_4528	2-oxoglutarate synthase activity		3	2
A0QVQ5	pnp, gpsI, MSMEG_2656, MSMEI_2593	3'-5'-exoribonuclease activity; magnesium ion binding; polyribonucleotide nucleotidyltransferase activity; RNA binding	mRNA catabolic process; RNA processing	6	3
A0R0B3	acpM, MSMEG_4326, MSMEI_4226	ACP phosphopantetheine attachment site binding involved in fatty acid biosynthetic process		2	
A0QWS1	pyrB, MSMEG_3043, MSMEI_2969	amino acid binding; aspartate carbamoyltransferase activity	de novo' pyrimidine nucleobase biosynthetic process; 'de novo' UMP biosynthetic process; cellular amino acid metabolic process	2	
A0R4G8	apeB, MSMEG_5828, MSMEI_5670	aminopeptidase activity; metallopeptidase activity; zinc ion binding		2	
A0QTV4	groL, MSMEG_1978, MSMEI_1934	ATP binding	protein refolding	2	
A0QQC8	dnaK, MSMEG_0709, MSMEI_0692	ATP binding	protein folding	2	

				Exclusive unique peptide count	
UniProt Accession	**Gene names**	**Gene ontology (molecular function)**	**Gene ontology (biological process)**	**RNA pol 1**	**RNA pol 2**
A0QQU5	groL1, groEL1, MSMEG_0880, MSMEI_0859	ATP binding	protein refolding	5	9
A0QV14	pyc, pca, MSMEG_2412, MSMEI_2351	ATP binding; biotin binding; biotin carboxylase activity; metal ion binding; pyruvate carboxylase activity	gluconeogenesis; pyruvate metabolic process	2	4
A0QTE1	accA3, MSMEG_1807, MSMEI_1762	ATP binding; biotin carboxylase activity; metal ion binding		4	
A0QV25	smc, MSMEG_2423, MSMEI_2362	ATP binding; DNA binding	chromosome condensation; DNA replication; sister chromatid cohesion		3
P48354	gyrA, MSMEG_0006, MSMEI_0008	ATP binding; DNA binding; DNA topoisomerase type II (ATP-hydrolyzing) activity	DNA-dependent DNA replication; DNA topological change; response to antibiotic	9	5
A0QNE0	gyrB, MSMEG_0005, MSMEI_0007	ATP binding; DNA binding; DNA topoisomerase type II (ATP-hydrolyzing) activity; magnesium ion binding	DNA-dependent DNA replication; DNA topological change	6	5
A0R079	glnA, glnA1, MSMEG_4290, MSMEI_4189	ATP binding; glutamate-ammonia ligase activity	glutamine biosynthetic process; nitrogen fixation	9	11
A0R218	rho, MSMEG_4954	ATP binding; helicase activity; RNA binding; RNA-dependent ATPase activity	DNA-templated transcription, termination	5	4
A0R143	nadE, MSMEG_4617, MSMEI_4500	ATP binding; hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds; NAD+ synthase (glutamine-hydrolyzing) activity	NAD biosynthetic process		3
A0R200	atpD, MSMEG_4936, MSMEI_4809	ATP binding; proton-transporting ATP synthase activity, rotational mechanism	ATP hydrolysis coupled proton transport; plasma membrane ATP synthesis coupled proton transport	2	3

				Exclusive unique peptide count	
UniProt Accession	**Gene names**	**Gene ontology (molecular function)**	**Gene ontology (biological process)**	**RNA pol 1**	**RNA pol 2**
A0R196	clpX, MSMEG_4671, MSMEI_4553	ATP binding; zinc ion binding	protein folding	2	
A0QYW4	uvrA, MSMEG_3808, MSMEI_3720	ATPase activity; ATP binding; DNA binding; excinuclease ABC activity; zinc ion binding	nucleotide-excision repair; SOS response	2	
A0R072	sucB, MSMEG_4283, MSMEI_4182	dihydrolipoyllysine-residue succinyltransferase activity		5	5
A0R5H1	MSMEG_6189, MSMEI_6029	DNA binding	regulation of transcription, DNA-templated; transcription, DNA-templated	2	2
A0QSL8	rpoA, MSMEG_1524, MSMEI_1488	DNA binding; DNA-directed RNA polymerase activity	transcription, DNA-templated	9	12
A0QS66	rpoC, MSMEG_1368, MSMEI_1329	DNA binding; DNA-directed RNA polymerase activity	transcription, DNA-templated	19	25
P60281	rpoB, MSMEG_1367, MSMEI_1328	DNA binding; DNA-directed RNA polymerase activity; ribonucleoside binding	response to antibiotic; transcription, DNA-templated	16	29
A0QW02	rpoD, sigA, MSMEG_2758, MSMEI_2690	DNA binding; sigma factor activity; transcription factor activity, sequence-specific DNA binding	transcription initiation from bacterial-type RNA polymerase promoter	4	2
A0R025	mraZ, MSMEG_4236, MSMEI_4136	DNA binding; transcription factor activity, sequence-specific DNA binding	transcription, DNA-templated	2	2
A0R1H7	MSMEG_4757, MSMEI_4637	enoyl-[acyl-carrier-protein] reductase (NADH) activity	fatty acid biosynthetic process	11	10
A0R2T4	xseA, MSMEG_5226	exodeoxyribonuclease VII activity; nucleic acid binding	DNA catabolic process	2	2
A0R1C2	gdh, MSMEG_4699, MSMEI_4582	glutamate dehydrogenase (NAD+) activity	glutamate catabolic process to 2-oxoglutarate	5	4
A0QS98	tuf, MSMEG_1401, MSMEI_1363	GTPase activity; GTP binding; translation elongation factor activity		2	2

				Exclusive unique peptide count	
UniProt Accession	**Gene names**	**Gene ontology (molecular function)**	**Gene ontology (biological process)**	**RNA pol 1**	**RNA pol 2**
I7FJV9	infB, MSMEI_2565	GTPase activity; GTP binding; translation initiation factor activity		4	2
A0R617	MSMEG_6392	hydrolase activity, acting on ester bonds; phosphopantetheine binding; transferase activity	biosynthetic process	2	4
A0QSU3	guaB, MSMEG_1602	IMP dehydrogenase activity	purine nucleotide biosynthetic process	7	3
A0QYL9	aceA, aceAb, MSMEG_3706, MSMEI_3618	isocitrate lyase activity	carboxylic acid metabolic process	9	10
A0R3B8	eno, MSMEG_5415, MSMEI_5267	magnesium ion binding; phosphopyruvate hydratase activity	glycolytic process	5	4
A0QYY6	rpsA, MSMEG_3833, MSMEI_3743, LJ00_19040	nucleotidyltransferase activity; RNA binding; structural constituent of ribosome	translation	2	2
A0R616	accD4, MSMEG_6391, MSMEI_6223	propionyl-CoA carboxylase activity		2	
A0QTE7	accD5, MSMEG_1813, MSMEI_1769	propionyl-CoA carboxylase activity		5	2
A0R3J2	pccB-1, MSMEG_5492, MSMEI_5340	propionyl-CoA carboxylase activity; transferase activity		3	
A0QSG6	rpsE, MSMEG_1472, MSMEI_1436	rRNA binding; structural constituent of ribosome	translation	3	
A0QSD4	rplB, MSMEG_1439, MSMEI_1403	rRNA binding; structural constituent of ribosome; transferase activity	translation	2	2
A0QZ46	prcA, MSMEG_3894, MSMEI_3804	threonine-type endopeptidase activity	modification-dependent protein catabolic process; proteasomal protein catabolic process	2	

References

1. WHO | Global tuberculosis report 2015. *WHO* at <http://www.who.int/tb/publications/global_report/en/>

2. Tufariello, J. M., Chan, J. & Flynn, J. L. Latent tuberculosis: mechanisms of host and bacillus that contribute to persistent infection. *Lancet Infect. Dis.* **3,** 578–590 (2003).

3. WHO | Latent Tuberculosis Infection (LTBI). *WHO* at <http://www.who.int/tb/challenges/ltbi/en/>

4. Harries, A. D. & Dye, C. Tuberculosis. *Ann. Trop. Med. Parasitol.* **100,** 415–431 (2006).

5. Velasco-García, R. & Vargas-Martínez, R. The study of protein-protein interactions in bacteria. *Can. J. Microbiol.* **58,** 1241–1257 (2012).

6. Braun, P. & Gingras, A.-C. History of protein-protein interactions: From egg-white to complex networks. *PROTEOMICS* **12,** 1478–1498 (2012).

7. Drewes, G. & Bouwmeester, T. Global approaches to protein-protein interactions. *Curr. Opin. Cell Biol.* **15,** 199–205 (2003).

8. De Silva, E. & Stumpf, M. P. H. Complex networks and simple models in biology. *J. R. Soc. Interface R. Soc.* **2,** 419–430 (2005).

9. Lalonde, S. *et al.* Molecular and cellular approaches for the detection of protein-protein interactions: latest techniques and current limitations. *Plant J. Cell Mol. Biol.* **53,** 610–635 (2008).

10. De Las Rivas, J. & Fontanillo, C. Protein-protein interactions essentials: key concepts to building and analyzing interactome networks. *PLoS Comput Biol* **6,** e1000807 (2010).

11. Nooren, I. M. A. & Thornton, J. M. Structural characterisation and functional significance of transient protein-protein interactions. *J. Mol. Biol.* **325,** 991–1018 (2003).

12. Huthmacher, C., Gille, C. & Holzhütter, H.-G. A computational analysis of protein interactions in metabolic networks reveals novel enzyme pairs potentially involved in metabolic channeling. *J. Theor. Biol.* **252,** 456–464 (2008).

13. Reichmann, D., Rahat, O., Cohen, M., Neuvirth, H. & Schreiber, G. The molecular architecture of protein-protein binding sites. *Curr. Opin. Struct. Biol.* **17,** 67–76 (2007).

14. Iwanczyk, J. *et al.* Role of the PDZ Domains in *Escherichia coli* DegP Protein. *J. Bacteriol.* **189,** 3176–3186 (2007).

15. Piehler, J. New methodologies for measuring protein interactions *in vivo* and *in vitro*. *Curr. Opin. Struct. Biol.* **15,** 4–14 (2005).

16. Fields, S. & Song, O. A novel genetic system to detect protein-protein interactions. *Nature* **340,** 245–246 (1989).

17. Deane, C. M., Salwiński, Ł., Xenarios, I. & Eisenberg, D. Protein interactions: two methods for assessment of the reliability of high throughput observations. *Mol. Cell. Proteomics MCP* **1,** 349–356 (2002).

18. Ito, T. *et al.* Roles for the two-hybrid system in exploration of the yeast protein interactome. *Mol. Cell. Proteomics MCP* **1,** 561–566 (2002).

19. Snider, J. *et al.* Detecting interactions with membrane proteins using a membrane two-hybrid assay in yeast. *Nat. Protoc.* **5,** 1281–1293 (2010).

20. Sureka, K. *et al.* Novel role of phosphorylation-dependent interaction between FtsZ and FipA in mycobacterial cell division. *PloS One* **5,** e8590 (2010).

21. Ma, J. & Ptashne, M. Converting a eukaryotic transcriptional inhibitor into an activator. *Cell* **55,** 443–446 (1988).

22. Wang, X., Huang, J., Mukherjee, A., Cao, C. & Lutkenhaus, J. Analysis of the interaction of FtsZ with itself, GTP, and FtsA. *J. Bacteriol.* **179,** 5551–5559 (1997).

23. Din, N., Quardokus, E. M., Sackett, M. J. & Brun, Y. V. Dominant C-terminal deletions of FtsZ that affect its ability to localize in Caulobacter and its interaction with FtsA. *Mol. Microbiol.* **27,** 1051–1063 (1998).

24. Yan, K., Pearce, K. H. & Payne, D. J. A conserved residue at the extreme C-terminus of FtsZ is critical for the FtsA-FtsZ interaction in Staphylococcus aureus. *Biochem. Biophys. Res. Commun.* **270,** 387–392 (2000).

25. Handa, P., Acharya, N., Thanedar, S., Purnapatre, K. & Varshney, U. Distinct properties of *Mycobacterium tuberculosis* single-stranded DNA binding protein and its functional characterization in *Escherichia coli*. *Nucleic Acids Res.* **28,** 3823–3829 (2000).

26. Steyn, A. J. C. *et al. Mycobacterium tuberculosis* WhiB3 interacts with RpoV to affect host survival but is dispensable for *in vivo* growth. *Proc. Natl. Acad. Sci.* **99,** 3147–3152 (2002).

27. Collins, D. M. *et al.* Mutation of the principal sigma factor causes loss of virulence in a strain of the *Mycobacterium tuberculosis* complex. *Proc. Natl. Acad. Sci.* **92,** 8036–8040 (1995).

28. Steyn, A. Role of WhiB3 in *M. tuberculosis* virulence. *Grantome* at <http://grantome.com/grant/NIH/R01-AI058131-07>

29. Singh, A. *et al. Mycobacterium tuberculosis* WhiB3 maintains redox homeostasis by regulating virulence lipid anabolism to modulate macrophage response. *PLoS Pathog* **5,** e1000545 (2009).

30. Stanley, S. A., Raghavan, S., Hwang, W. W. & Cox, J. S. Acute infection and macrophage subversion by *Mycobacterium tuberculosis* require a specialized secretion system. *Proc. Natl. Acad. Sci.* **100,** 13001–13006 (2003).

31. Lin, Y. *et al.* Identification of antituberculosis agents that target ribosomal protein interactions using a yeast two-hybrid system. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 17412–17417 (2012).

32. Tao, J. *et al.* Mycobacterium fluoroquinolone resistance protein B, a novel small GTPase, is involved in the regulation of DNA gyrase and drug resistance. *Nucleic Acids Res.* gks1351 (2012).

33. Huet, G., Daffé, M. & Saves, I. Identification of the *Mycobacterium tuberculosis* SUF machinery as the exclusive mycobacterial system of [Fe-S] cluster assembly: evidence for its implication in the pathogen's survival. *J. Bacteriol.* **187,** 6137–6146 (2005).

34. Sassetti, C. M. & Rubin, E. J. Genetic requirements for mycobacterial survival during infection. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 12989–12994 (2003).

35. Sassetti, C. M., Boyd, D. H. & Rubin, E. J. Genes required for mycobacterial growth defined by high density mutagenesis. *Mol. Microbiol.* **48,** 77–84 (2003).

36. Huet, G., Castaing, J.-P., Fournier, D., Daffé, M. & Saves, I. Protein splicing of SufB is crucial for the functionality of the *Mycobacterium tuberculosis* SUF machinery. at <http://jb.asm.org>

37. Seebeck, F. P. *In vitro* reconstitution of mycobacterial ergothioneine biosynthesis. *J. Am. Chem. Soc.* **132,** 6632–6633 (2010).

38. Rybniker, J. *et al.* The cysteine desulfurase IscS of *Mycobacterium tuberculosis* is involved in iron-sulfur cluster biogenesis and oxidative stress defence. *Biochem. J.* **459,** 467–478 (2014).

39. Loiseau, L., Ollagnier-de-Choudens, S., Nachin, L., Fontecave, M. & Barras, F. Biogenesis of Fe-S cluster by the bacterial Suf system: SufS and SufE form a new type of cysteine desulfurase. *J. Biol. Chem.* **278,** 38352–38359 (2003).

40. Ollagnier-de-Choudens, S. *et al.* Mechanistic studies of the SufS-SufE cysteine desulfurase: evidence for sulfur transfer from SufS to SufE. *FEBS Lett.* **555,** 263–267 (2003).

41. Mihara, H. & Esaki, N. Bacterial cysteine desulfurases: their function and mechanisms. *Appl. Microbiol. Biotechnol.* **60,** 12–23 (2002).

42. Michnick, S. W., Remy, I., Campbell-Valois, F. X., Vallée-Bélisle, A. & Pelletier, J. N. Detection of protein-protein interactions by protein fragment complementation strategies. *Methods Enzymol.* **328,** 208–230 (2000).

43. Johnsson, N. & Varshavsky, A. Split ubiquitin as a sensor of protein interactions *in vivo*. *Proc. Natl. Acad. Sci.* **91,** 10340–10344 (1994).

44. Karimova, G., Pidoux, J., Ullmann, A. & Ladant, D. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 5752–5756 (1998).

45. Dziedzic, R. *et al. Mycobacterium tuberculosis* ClpX interacts with FtsZ and interferes with FtsZ assembly. *PloS One* **5,** e11058 (2010).

46. Singh, A., Mai, D., Kumar, A. & Steyn, A. J. C. Dissecting virulence pathways of *Mycobacterium tuberculosis* through protein-protein association. *Proc. Natl. Acad. Sci.* **103,** 11346–11351 (2006).

47. Tafelmeyer, P., Johnsson, N. & Johnsson, K. Transforming a (beta/alpha)8--barrel enzyme into a split-protein sensor through directed evolution. *Chem. Biol.* **11,** 681–689 (2004).

48. O'Hare, H., Juillerat, A., Dianisková, P. & Johnsson, K. A split-protein sensor for studying protein-protein interaction in mycobacteria. *J. Microbiol. Methods* **73,** 79–84 (2008).

49. Mai, D. *et al.* A screen to identify small molecule inhibitors of protein-protein interactions in mycobacteria. *Assay Drug Dev. Technol.* **9,** 299–310 (2011).

50. Morell, M., Espargaró, A., Avilés, F. X. & Ventura, S. Detection of transient protein-protein interactions by bimolecular fluorescence complementation: the Abl-SH3 case. *Proteomics* **7,** 1023–1036 (2007).

51. Luker, K. E. *et al.* Kinetics of regulated protein-protein interactions revealed with firefly luciferase complementation imaging in cells and living animals. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 12288–12293 (2004).

52. Li, H., Chang, L., Howell, J. M. & Turner, R. J. DmsD, a Tat system specific chaperone, interacts with other general chaperones and proteins involved in the molybdenum cofactor biosynthesis. *Biochim. Biophys. Acta* **1804,** 1301–1309 (2010).

53. Kostecki, J. S., Li, H., Turner, R. J. & DeLisa, M. P. Visualizing interactions along the *Escherichia coli* twin-arginine translocation pathway using protein fragment complementation. *PloS One* **5,** e9225 (2010).

54. Simeone, R. *et al.* Phagosomal rupture by *Mycobacterium tuberculosis* results in toxicity and host cell death. *PLoS Pathog.* **8,** e1002507 (2012).

55. Xu, Y., Piston, D. W. & Johnson, C. H. A bioluminescence resonance energy transfer (BRET) system: Application to interacting circadian clock proteins. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 151–156 (1999).

56. Cui, B. *et al.* Bioluminescence resonance energy transfer system for measuring dynamic protein-protein interactions in bacteria. *mBio* **5,** e01050–14 (2014).

57. Du, Y., Zhang, H., He, Y., Huang, F. & He, Z.-G. *Mycobacterium smegmatis* Lsr2 physically and functionally interacts with a new flavoprotein involved in bacterial resistance to oxidative stress. *J. Biochem. (Tokyo)* **152,** 479–486 (2012).

58. Veyron-Churlet, R., Guerrini, O., Mourey, L., Daffé, M. & Zerbib, D. Protein-protein interactions within the Fatty Acid Synthase-II system of *Mycobacterium tuberculosis* are essential for mycobacterial viability. *Mol. Microbiol.* **54,** 1161–1172 (2004).

59. Veyron-Churlet, R. *et al.* The biosynthesis of mycolic acids in *Mycobacterium tuberculosis* relies on multiple specialized elongation complexes interconnected by specific protein-protein interactions. *J. Mol. Biol.* **353,** 847–858 (2005).

60. Płociński, P. *et al.* Identification of protein partners in mycobacteria using a single-step affinity purification method. *PLoS ONE* **9,** e91380 (2014).

61. Zheng, M. *et al.* DNA microarray-mediated transcriptional profiling of the *Escherichia coli* response to hydrogen peroxide. *J. Bacteriol.* **183,** 4562–4570 (2001).

62. Nachin, L., Hassouni, M. El, Loiseau, L., Expert, D. & Barras, F. SoxR-dependent response to oxidative stress and virulence of *Erwinia chrysanthemi*: the key role of SufC, an orphan ABC ATPase. *Mol. Microbiol.* **39,** 960–972 (2001).

63. Nachin, L., Loiseau, L., Expert, D. & Barras, F. SufC: an unorthodox cytoplasmic ABC/ATPase required for [Fe-S] biogenesis under oxidative stress. *EMBO J.* **22,** 427–437 (2003).

64. Outten, F. W., Djaman, O. & Storz, G. A suf operon requirement for Fe-S cluster assembly during iron starvation in *Escherichia coli*. *Mol. Microbiol.* **52,** 861–872 (2004).

65. Patzer, S. I. & Hantke, K. SufS is a NifS-like protein, and SufD is necessary for stability of the [2Fe-2S] FhuF protein in *Escherichia coli*. *J. Bacteriol.* **181,** 3307–3309 (1999).

66. Bolstad, H. M. & Wood, M. J. An *in vivo* method for characterization of protein interactions within sulfur trafficking systems of *E. coli*. *J. Proteome Res.* **9,** 6740–6751 (2010).

67. Chim, N. *et al.* The TB Structural Genomics Consortium: a decade of progress. *Tuberc. Edinb. Scotl.* **91,** 155–172 (2011).

68. Strong, M. *et al.* Toward the structural genomics of complexes: Crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci.* **103,** 8060–8065 (2006).

69. Arbing, M. A. *et al.* Heterologous expression of mycobacterial Esx complexes in *Escherichia coli* for structural studies is facilitated by the use of maltose binding protein fusions. *PloS One* **8,** e81753 (2013).

70. Brückner, A., Polge, C., Lentze, N., Auerbach, D. & Schlattner, U. Yeast two-hybrid, a powerful tool for systems biology. *Int. J. Mol. Sci.* **10,** 2763–2788 (2009).

71. Rigaut, G. *et al.* A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* **17,** 1030–1032 (1999).

72. Jones, S. & Thornton, J. M. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 13–20 (1996).

73. Huynen, M. A., Snel, B., von Mering, C. & Bork, P. Function prediction and protein networks. *Curr. Opin. Cell Biol.* **15,** 191–198 (2003).

74. Marcotte, E. M. *et al.* Detecting protein function and protein-protein interactions from genome sequences. *Science* **285,** 751–753 (1999).

75. Enright, A. J., Iliopoulos, I., Kyrpides, N. C. & Ouzounis, C. A. Protein interaction maps for complete genomes based on gene fusion events. *Nature* **402,** 86–90 (1999).

76. Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D. & Yeates, T. O. Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles. *Proc. Natl. Acad. Sci.* **96,** 4285–4288 (1999).

77. Overbeek, R., Fonstein, M., D'Souza, M., Pusch, G. D. & Maltsev, N. Use of contiguity on the chromosome to predict functional coupling. *In Silico Biol.* **1,** 93–108 (1999).

78. Dandekar, T., Snel, B., Huynen, M. & Bork, P. Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem. Sci.* **23,** 324–328 (1998).

79. Nourani, E., Khunjush, F. & Durmuş, S. Computational approaches for prediction of pathogen-host protein-protein interactions. *Front. Microbiol.* **6,** (2015).

80. Raman, K. & Chandra, N. *Mycobacterium tuberculosis* interactome analysis unravels potential pathways to drug resistance. *BMC Microbiol.* **8,** 234 (2008).

81. Cui, T., Zhang, L., Wang, X. & He, Z.-G. Uncovering new signaling proteins and potential drug targets through the interactome analysis of *Mycobacterium tuberculosis*. *BMC Genomics* **10,** 118 (2009).

82. Cole, S. T. *et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393,** 537–544 (1998).

83. Wang, Y. *et al.* Global protein-protein interaction network in the human pathogen *Mycobacterium tuberculosis* H37Rv. *J. Proteome Res.* **9,** 6665–6677 (2010).

84. Roberts, G., III, Parrish, J., Mangiola, B. & Finley, R., Jr. in *Two Hybrid Technologies* (eds. Suter, B. & Wanker, E. E.) 39–61 (Humana Press, 2012).

85. Mayer, M. L. & Hieter, P. Protein networks—built by association. *Nat. Biotechnol.* **18,** 1242–1243 (2000).

86. Arifuzzaman, M. *et al.* Large-scale identification of protein-protein interaction of *Escherichia coli* K-12. *Genome Res.* **16,** 686–691 (2006).

87. J C Rain, L. S. The protein-protein interaction map of *Helicobacter pylori*. *Nature* **409,** 211–5 (2001).

88. Fromont-Racine, M., Rain, J. C. & Legrain, P. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nat. Genet.* **16,** 277–282 (1997).

89. Ito, T. *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 4569–4574 (2001).

90. Uetz, P. *et al.* A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* **403,** 623–627 (2000).

91. Giot, L. *et al.* A protein interaction map of *Drosophila melanogaster*. *Science* **302,** 1727–1736 (2003).

92. Rajagopala, S. V. *et al.* The binary protein-protein interaction landscape of *Escherichia coli*. *Nat. Biotechnol.* **32,** 285–290 (2014).

93. Yu, H. *et al.* High quality binary protein interaction map of the yeast interactome network. *Science* **322,** 104–110 (2008).

94. Mehra, A. *et al. Mycobacterium tuberculosis* Type VII secreted effector EsxH targets host ESCRT to impair trafficking. *PLoS Pathog* **9,** e1003734 (2013).

95. Wuchty, S. & Uetz, P. Protein-protein interaction networks of *E. coli* and *S. cerevisiae* are similar. *Sci. Rep.* **4,** 7187 (2014).

96. von Mering, C. *et al.* Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* **417,** 399–403 (2002).

97. Rajagopala, S. V. *et al.* MPI-LIT: a literature-curated dataset of microbial binary protein-protein interactions. *Bioinformatics* **24,** 2622–2627 (2008).

98. Gavin, A.-C., Maeda, K. & Kühner, S. Recent advances in charting protein-protein interaction: mass spectrometry-based approaches. *Curr. Opin. Biotechnol.* **22,** 42–49 (2011).

99. Hu, P. *et al.* Global functional atlas of *Escherichia coli* encompassing previously uncharacterized proteins. *PLoS Biol.* **7,** e96 (2009).

100. Butland, G. *et al.* Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature* **433,** 531–537 (2005).

101. Gavin, A.-C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415,** 141–147 (2002).

102. Ho, Y. *et al.* Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415,** 180–183 (2002).

103. Gavin, A.-C. *et al.* Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440,** 631–636 (2006).

104. Krogan, N. J. *et al.* Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440,** 637–643 (2006).

105. Zybailov, B. L., Glazko, G. V., Jaiswal, M. & Raney, K. D. Large scale chemical cross-linking mass spectrometry perspectives. *J. Proteomics Bioinform.* **6,** 001 (2013).

106. Lougheed, K. E. A., Bennett, M. H. & Williams, H. D. An *in vivo* crosslinking system for identifying mycobacterial protein-protein interactions. *J. Microbiol. Methods*

107. Rappsilber, J. The beginning of a beautiful friendship: Cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J. Struct. Biol.* **173,** 530–540 (2011).

108. Bacteriome.org. at <http://128.100.134.188/bacteriome/>

109. Lehne, B. & Schlitt, T. Protein-protein interaction databases: keeping up with growing interactomes. *Hum. Genomics* **3,** 291–297 (2009).

110. Orchard, S. *et al.* Protein interaction data curation: the International Molecular Exchange (IMEx) consortium. *Nat. Methods* **9,** 345–350 (2012).

111. Kerrien, S. *et al.* Broadening the horizon-level 2.5 of the HUPO-PSI format for molecular interactions. *BMC Biol.* **5,** 44 (2007).

112. Hermjakob, H. *et al.* IntAct: an open source molecular interaction database. *Nucleic Acids Res.* **32,** D452–455 (2004).

113. Goll, J. *et al.* MPIDB: the microbial protein interaction database. *Bioinforma. Oxf. Engl.* **24,** 1743–1744 (2008).

114. Szklarczyk, D. *et al.* STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43,** D447–452 (2015).

115. Xenarios, I. *et al.* DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res.* **30,** 303–305 (2002).

116. Zanzoni, A. *et al.* MINT: a Molecular INTeraction database. *FEBS Lett.* **513,** 135–140 (2002).

117. Stark, C. *et al.* BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* **34,** D535–539 (2006).

118. Bader, G. D., Betel, D. & Hogue, C. W. V. BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Res.* **31,** 248–250 (2003).

119. Snyder, E. E. *et al.* PATRIC: The VBI PathoSystems Resource Integration Center. *Nucleic Acids Res.* **35,** D401–D406 (2007).

120. Wilson, A. J. Inhibition of protein-protein interactions using designed molecules. *Chem. Soc. Rev.* **38,** 3289–3300 (2009).

121. Rao, P. K. S. CCR5 inhibitors: Emerging promising HIV therapeutic strategy. *Indian J. Sex. Transm. Dis.* **30,** 1–9 (2009).

122. Dyer, M. D. *et al.* The human-bacterial pathogen protein interaction networks of *Bacillus anthracis*, *Francisella tularensis*, and *Yersinia pestis*. *PloS One* **5,** e12089 (2010).

123. Roche, B. *et al.* Iron/sulfur proteins biogenesis in prokaryotes: Formation, regulation and diversity. *Biochim. Biophys. Acta BBA - Bioenerg.* **1827,** 455–469 (2013).

124. Py, B. & Barras, F. Building Fe-S proteins: bacterial strategies. *Nat. Rev. Microbiol.* **8,** 436–446 (2010).

125. Fontecave, M. & Ollagnier-de-Choudens, S. Iron-sulfur cluster biosynthesis in bacteria: Mechanisms of cluster assembly and transfer. *Arch. Biochem. Biophys.* **474,** 226–237 (2008).

126. Griffin, J. E. *et al.* High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathog.* **7,** e1002251 (2011).

127. Yano, T. *et al.* Reduction of clofazimine by mycobacterial type 2 NADH:quinone oxidoreductase: a pathway for the generation of bactericidal levels of reactive oxygen species. *J. Biol. Chem.* **286,** 10276–10287 (2011).

128. Shen, G. *et al.* SufR coordinates two [4Fe-4S]2+, 1+ clusters and functions as a transcriptional repressor of the sufBCDS operon and an autoregulator of sufR in cyanobacteria. *J. Biol. Chem.* **282,** 31909–31919 (2007).

129. Wang, T. *et al.* The sufR gene (sll0088 in Synechocystis sp. strain PCC 6803) functions as a repressor of the sufBCDS operon in iron-sulfur cluster biogenesis in cyanobacteria. *J. Bacteriol.* **186,** 956–967 (2004).

130. Singh, A. *et al. Mycobacterium tuberculosis* WhiB3 responds to O2 and nitric oxide via its [4Fe-4S] cluster and is essential for nutrient starvation survival. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 11562–11567 (2007).

131. Guo, X. V. *et al.* Silencing *Mycobacterium smegmatis* by using tetracycline repressors. *J. Bacteriol.* **189,** 4614–4623 (2007).

132. Noens, E. E. *et al.* Improved mycobacterial protein production using a *Mycobacterium smegmatis groEL1ΔC* expression strain. *BMC Biotechnol.* **11,** 27 (2011).

133. Sequencher DNA Sequence Analysis Software from Gene Codes Corporation. at <http://www.genecodes.com/>

134. Bern, M., Kil, Y. J. & Becker, C. Byonic: advanced peptide and protein identification software. *Curr. Protoc. Bioinforma. Ed. Board Andreas Baxevanis Al* **Chapter 13,** Unit13.20 (2012).

135. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26,** 1367–1372 (2008).

136. Oliveros, J.C. (2007-2015) Venny. An interactive tool for comparing lists with Venn's diagrams. http://bioinfogp.cnb.csic.es/tools/venny/index.html

137. Consortium, T. U. UniProt: a hub for protein information. *Nucleic Acids Res.* **43,** D204–D212 (2015).

138. Schägger, H. & von Jagow, G. Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. *Anal. Biochem.* **199,** 223–231 (1991).

139. Dresler, J., Klimentova, J. & Stulik, J. Bacterial protein complexes investigation using blue native PAGE. *Microbiol. Res.* **166,** 47–62 (2011).

140. Voskuil, M. I., Bartek, I. L., Visconti, K. & Schoolnik, G. K. The response of *Mycobacterium tuberculosis* to reactive oxygen and nitrogen species. *Front. Microbiol.* **2,** (2011).

141. Voskuil, M. I. *et al.* Inhibition of respiration by nitric oxide induces a *Mycobacterium tuberculosis* dormancy program. *J. Exp. Med.* **198,** 705–713 (2003).

142. Betts, J. C., Lukey, P. T., Robb, L. C., McAdam, R. A. & Duncan, K. Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. *Mol. Microbiol.* **43,** 717–731 (2002).

143. Ehrt, S. *et al.* Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor. *Nucleic Acids Res.* **33,** e21–e21 (2005).

144. Hett, E. C. & Rubin, E. J. Bacterial growth and cell division: a mycobacterial perspective. *Microbiol. Mol. Biol. Rev.* **72,** 126–156 (2008).

145. Clifton E. Barry, B. S. P., Dean C. Crick, B. S. P. & Michael R. McNeil, B. S. P. Targeting the formation of the cell wall core of *M. tuberculosis*. *Infect. Disord. - Drug Targets* **7,** 182–202 (2007).

146. Sani, M. *et al.* Direct visualization by Cryo-EM of the mycobacterial capsular layer: a labile structure containing ESX-1-secreted proteins. *PLoS Pathog* **6,** e1000794 (2010).

147. Kaur, D., Guerin, M. E., Skovierová, H., Brennan, P. J. & Jackson, M. Chapter 2: Biogenesis of the cell wall and other glycoconjugates of *Mycobacterium tuberculosis*. *Adv. Appl. Microbiol.* **69,** 23–78 (2009).

148. Selvaraj, P., Kannapiran, M., Kurian, S. M. & Narayanan, P. R. Effect of plasma lysozyme on live *Mycobacterium tuberculosis*. *Curr Sci* **81,** 201–203 (2001).

149. Antonioli, P., Bachi, A., Fasoli, E. & Righetti, P. G. Efficient removal of DNA from proteomic samples prior to two-dimensional map analysis. *J. Chromatogr. A* **1216,** 3606–3612 (2009).

150. Hausladen, A., Privalle, C. T., Keng, T., DeAngelo, J. & Stamler, J. S. Nitrosative stress: activation of the transcription factor OxyR. *Cell* **86,** 719–729 (1996).

151. Pomposiello, P. J. & Demple, B. Redox-operated genetic switches: the SoxR and OxyR transcription factors. *Trends Biotechnol.* **19,** 109–114 (2001).

152. Crack, J. C., Green, J., Thomson, A. J. & Le Brun, N. E. Iron-sulfur cluster sensor-regulators. *Curr. Opin. Chem. Biol.* **16,** 35–44 (2012).

153. Chiang, S. M. & Schellhorn, H. E. Regulators of oxidative stress response genes in *Escherichia coli* and their functional conservation in bacteria. *Arch. Biochem. Biophys.* **525,** 161–169 (2012).

154. Fang, Z., van der Merwe, R. G., Warren, R. M., Schubert, W.-D. & Gey van Pittius, N. C. Assessing the progress of *Mycobacterium tuberculosis* H37Rv structural genomics. *Tuberc. Edinb. Scotl.* (2014).

155. Abuhammad, A. *et al.* Improvement of the expression and purification of *Mycobacterium tuberculosis* arylamine N-acetyltransferase (TBNAT) a potential target for novel anti-tubercular agents. *Protein Expr. Purif.* **80,** 246–252 (2011).

156. Daugelat, S. *et al.* The RD1 proteins of *Mycobacterium tuberculosis*: expression in *Mycobacterium smegmatis* and biochemical characterization. *Microbes Infect. Inst. Pasteur* **5,** 1082–1095 (2003).

157. Donovan, R. S., Robinson, C. W. & Glick, B. R. Review: optimizing inducer and culture conditions for expression of foreign proteins under the control of the lac promoter. *J. Ind. Microbiol.* **16,** 145–154 (1996).

158. Batt, S. M. *et al.* Structural basis of inhibition of *Mycobacterium tuberculosis* DprE1 by benzothiazinone inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 11354–11359 (2012).

159. Schnell, R., Sandalova, T., Hellman, U., Lindqvist, Y. & Schneider, G. Siroheme- and [Fe4-S4]-dependent NirA from *Mycobacterium tuberculosis* is a sulfite reductase with a covalent Cys-Tyr bond in the active site. *J. Biol. Chem.* **280,** 27319–27328 (2005).

160. Felnagle, E. A. *et al.* MbtH-like proteins as integral components of bacterial nonribosomal peptide synthetases. *Biochemistry (Mosc.)* **49,** 8815–8817 (2010).

161. McMahon, M. D., Rush, J. S. & Thomas, M. G. Analyses of MbtB, MbtE, and MbtF suggest revisions to the mycobactin biosynthesis pathway in *Mycobacterium tuberculosis*. *J. Bacteriol.* **194,** 2809–2818 (2012).

162. Noens, E. E. *et al.* Improved mycobacterial protein production using a *Mycobacterium smegmatis groEL1ΔC* expression strain. *BMC Biotechnol.* **11,** 27 (2011).

163. Bashiri, G., Squire, C. J., Baker, E. N. & Moreland, N. J. Expression, purification and crystallization of native and selenomethionine labeled *Mycobacterium tuberculosis* FGD1 (Rv0407) using a *Mycobacterium smegmatis* expression system. *Protein Expr. Purif.* **54,** 38–44 (2007).

164. Goldstone, R. M., Moreland, N. J., Bashiri, G., Baker, E. N. & Shaun Lott, J. A new Gateway vector and expression protocol for fast and efficient recombinant protein expression in *Mycobacterium smegmatis*. *Protein Expr. Purif.* **57,** 81–87 (2008).

165. Delogu, G. *et al.* Expression and purification of recombinant methylated HBHA in *Mycobacterium smegmatis*. *FEMS Microbiol. Lett.* **239,** 33–39 (2004).

166. Dumetz, A. C., Snellinger-O'Brien, A. M., Kaler, E. W. & Lenhoff, A. M. Patterns of protein-protein interactions in salt solutions and implications for protein crystallization. *Protein Sci. Publ. Protein Soc.* **16,** 1867–1877 (2007).

167. Zhang, J. in Protein-Protein Interactions - Computational and Experimental Tools (ed. Cai, W.) (InTech, 2012).

168. Saini, A., Mapolelo, D. T., Chahal, H. K., Johnson, M. K. & Outten, F. W. SufD and SufC ATPase activity are required for iron acquisition during *in vivo* Fe-S cluster formation on SufB. *Biochemistry (Mosc.)* **49,** 9402–9412 (2010).

169. Pluskal, T., Ueno, M. & Yanagida, M. Genetic and metabolomic dissection of the ergothioneine and selenoneine biosynthetic pathway in the fission yeast, *S. pombe*, and construction of an overproduction system. *PLoS ONE* **9,** e97774 (2014).

170. Vinella, D., Loiseau, L., Ollagnier de Choudens, S., Fontecave, M. & Barras, F. *In vivo* [Fe-S] cluster acquisition by IscR and NsrR, two stress regulators in *Escherichia coli*. *Mol. Microbiol.* **87,** 493–508 (2013).

171. Vergnes, A., Gouffi-Belhabich, K., Blasco, F., Giordano, G. & Magalon, A. Involvement of the molybdenum cofactor biosynthetic machinery in the maturation of the *Escherichia coli* nitrate reductase A. *J. Biol. Chem.* **279,** 41398–41403 (2004).

172. Pierre, J. L. & Fontecave, M. Iron and activated oxygen species in biology: The basic chemistry. *Biometals* **12,** 195–199 (1999).

173. Imlay, J. A. Pathways of oxidative damage. *Annu. Rev. Microbiol.* **57,** 395–418 (2003).

174. Layer, G., Ollagnier-de Choudens, S., Sanakis, Y. & Fontecave, M. Iron-sulfur cluster biosynthesis: characterization of *Escherichia coli* CYaY as an iron donor for the assembly of [2Fe-2S] clusters in the scaffold IscU. *J. Biol. Chem.* **281,** 16256–16263 (2006).

175. Ding, H., Yang, J., Coleman, L. C. & Yeung, S. Distinct iron binding property of two putative iron donors for the iron-sulfur cluster assembly: IscA and the bacterial frataxin ortholog CyaY under physiological and oxidative stress conditions. *J. Biol. Chem.* **282,** 7997–8004 (2007).

176. Turowski, V. R., Busi, M. V. & Gomez-Casati, D. F. Structural and functional studies of the mitochondrial cysteine desulfurase from *Arabidopsis thaliana*. *Mol. Plant* **5,** 1001–1010 (2012).

177. Foury, F., Pastore, A. & Trincal, M. Acidic residues of yeast frataxin have an essential role in Fe-S cluster assembly. *EMBO Rep.* **8,** 194–199 (2007).

178. Expert, D., Boughammoura, A. & Franza, T. Siderophore-controlled iron assimilation in the enterobacterium *Erwinia chrysanthemi*: evidence for the involvement of bacterioferritin and the Suf iron-sulfur cluster assembly machinery. *J. Biol. Chem.* **283,** 36564–36572 (2008).

179. Saini, V., Farhana, A., Glasgow, J. N. & Steyn, A. J. Iron sulfur cluster proteins and microbial regulation: implications for understanding tuberculosis. *Curr. Opin. Chem. Biol.* **16,** 45–53 (2012).

180. Tian, J., Bryk, R., Itoh, M., Suematsu, M. & Nathan, C. Variant tricarboxylic acid cycle in *Mycobacterium tuberculosis*: identification of alpha-ketoglutarate decarboxylase. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 10670–10675 (2005).

181. Raju, R. M. *et al. Mycobacterium tuberculosis* ClpP1 and ClpP2 function together in protein degradation and are required for viability *in vitro* and during infection. *PLoS Pathog* **8,** e1002511 (2012).

182. Wei, J.-R. *et al.* Depletion of antibiotic targets has widely varying effects on growth. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 4176–4181 (2011).

183. Vranish, J. N. *et al.* Fluorescent probes for tracking the transfer of iron–sulfur cluster and other metal cofactors in biosynthetic reaction pathways. *J. Am. Chem. Soc.* **137,** 390–398 (2015).