
UNIVERSITÄTSKLINIKUM HAMBURG-EPPENDORF

Institut für Biochemie und Signaltransduktion

Prof. Dr. rer. nat. Aymelt Itzen

**Structural and functional insights into a
mycobacterial type VII secretion system**

Dissertation

**zur Erlangung des Doktorgrades eines PhD an der Medizinischen Fakultät der
Universität Hamburg**

vorgelegt von

Christina Ritter

aus Seligenstadt

Hamburg 2022

Angenommen von der Medizinischen Fakultät am: **22.03.2023**

Veröffentlicht mit Genehmigung der Medizinischen Fakultät der Universität Hamburg.

Prüfungsausschuss, der/die Vorsitzende: **PD Dr. Matthias Wilmanns**

Prüfungsausschuss, 2. Gutachter/in: **Prof. Dr. Aymelt Itzen**



This thesis describes work performed under the supervision of Prof. Dr. Matthias Wilmanns at the European Molecular Biology Laboratory (EMBL) in Hamburg, Germany.

Contents

Abstract	iv
Zusammenfassung	vi
List of figures	viii
List of tables	ix
List of abbreviations	xi
Scientific contributions	xiv
1 Introduction	1
1.1 Tuberculosis	1
1.1.1 Tuberculosis - a deadly disease	1
1.1.2 Mycobacterial physiology	2
1.2 Secretion	3
1.2.1 Protein transport machineries	3
1.2.2 Secretion systems in mycobacteria	5
1.3 Aim of this work	19
1.4 Main methods	20
1.4.1 Expression host <i>Mycobacterium smegmatis</i>	20
1.4.2 Electron microscopy	21
1.4.3 Membrane protein reconstitution	21
1.4.4 Protein prediction by AlphaFold2	23
2 Materials and Methods	24
2.1 Common buffers	24
2.2 Molecular biology	25
2.2.1 Preparation of electrocompetent <i>Mycobacterium smegmatis</i> cells	25
2.2.2 <i>Mycobacterium smegmatis</i> electroporation	25
2.2.3 Expression of ESX-5 constructs in <i>Mycobacterium smegmatis</i>	25
2.2.4 Expression of Saposin A in <i>Escheria coli</i>	26
2.2.5 Purification of Saposin A	26
2.2.6 Purification of ESX-5 constructs	27
2.2.7 Direct extraction of ESX-5 by Saposin A	28

2.3	Protein analysis	29
2.3.1	Sodium dodecyl sulfate–polyacrylamide gel electrophoresis	29
2.3.2	Blue native-polyacrylamide gel electrophoresis	29
2.3.3	Western blot	29
2.3.4	Limited proteolysis	30
2.3.5	Surface plasmon resonance analysis	30
2.3.6	Negative stain EM	30
2.3.7	Crosslinking mass spectrometry	30
2.4	Cryo-EM	31
2.4.1	Grid preparation	31
2.4.2	Data collection	32
2.4.3	Processing	33
2.5	Model building and refinement	35
2.6	Integrative modeling of EccB ₅ and EccC ₅	36
2.7	Bioinformatics	38
2.7.1	Analysis of EccC ₅ pore dimensions and EccB ₅ angles	38
2.7.2	Sequence alignments and sequence logo preparation	38
2.7.3	Analysis and visualization of protein structures	38
2.7.4	Molecular dynamics simulations	38
2.7.5	Structure prediction with AlphaFold2	39
2.8	Secretion assay	39
2.8.1	Sample preparation	39
2.8.2	TMT labeling and LC-MS/MS	40
2.8.3	Data analysis	41
3	Results	43
3.1	Structure of the ESX-5 core complex from <i>Mycobacterium xenopi</i>	43
3.1.1	Purification of the ESX-5 core complex	43
3.1.2	Single particle cryo-EM on the ESX-5 core complex	44
3.1.3	The protomer - the basic unit of the ESX-5 core complex	47
3.1.4	Periplasmic interaction network formed by EccB ₅	50
3.1.5	Key interprotomeric interactions in the transmembrane region	51
3.1.6	The central secretion pore formed by EccC ₅ transmembrane helices	53
3.2	Dissecting the role of specific residues by site-directed mutagenesis	56
3.2.1	Targeting the integrity of the pore basket	56
3.2.2	Targeting the secretion pore	56
3.2.3	Impact of mutations on secretion	57
3.2.4	Structural impact of a conserved proline residue	60
3.2.5	Molecular dynamics study on the EccC ₅ pore helices	62

3.3	Understanding conformational flexibility of the ESX-5 core complex	65
3.3.1	Flexibility inside the ESX-5 core complex	65
3.3.2	The ESX-5 core complex in a native lipid environment	66
3.3.3	Binding of ATP γ S to the ESX-5 core complex	70
4	Discussion	76
5	Appendix	89
	Bibliography	96
	Acknowledgment	121
	List of publications	123
	Eidesstattliche Versicherung	124

Abstract

With over 1.5 million deaths per year, mycobacterial pathogens, including *Mycobacterium tuberculosis*, belong to the deadliest pathogens worldwide. Essential to their complex infection mechanism is the secretion of virulence proteins across their otherwise almost impermeable cell envelope. The transport of these heterodimeric protein complexes is mediated by a specialized secretion system, the Type VII or ESX secretion system. Slow-growing, often pathogenic mycobacteria, code for up to five different ESX systems, termed ESX-1 to ESX-5, which fulfill a diverse set of functions in the mycobacterial life cycle. Although the systems are considered promising drug targets, the molecular mechanism by which substrates are transported across the mycobacterial membrane remains elusive so far.

This thesis presents the structure of the inner membrane spanning ESX-5 system from *Mycobacterium xenopi*. Integrating data from cryo-EM, crosslinking mass spectrometry and homology models enabled a detailed view on the hexameric secretion system with central secretion pore. The complex is formed by six protomeric units that comprise the protein components EccB₅:EccC₅:EccD₅ and EccE₅ in a 1:1:2:1 stoichiometry. While the transmembrane segment was overall well defined within the cryo-EM density, the secretion pore, formed by two helices of EccC₅, showed inherent flexibility, which was hypothesized to be caused by a conserved proline residue in the center of the helix. In addition, analysis of the pore diameter suggested a role of phenylalanine residues in a gating mechanism. A mutational analysis carried out in this thesis revealed that the integrity of these pore residues and the integrity of the pore scaffold, formed by EccB₅, is essential for secretion. A cryo-EM structure further confirmed that mutation of the central proline residue leads to the rigidification of the pore helices, indicating that plasticity of the central transmembrane pore is necessary to allow transport of different folded substrates across the cell envelope.

This plasticity was also observed within the cytosol for the three consecutive ATPase domains of EccC₅ and on the periplasmic side for the periplasmic domain of EccB₅. As shown by cryo-EM, addition of ATP γ S slightly stabilized two EccC₅ conformations within the cytoplasm, probably corresponding to different ATP γ S bound states. It is hypothesized that the remaining intrinsic flexibility could bestow dynamics to the system that are essential to carry out huge conformational changes in the translocation cycle.

Along with other recently published structures of ESX-5 and ESX-3 systems and improved

structure prediction methods, the presented results have paved the way for structural comparison of the different systems and for a deeper understanding of the underlying mechanism of secretion.

Zusammenfassung

Mit über 1,5 Millionen Todesfällen pro Jahr gehören Mykobakterien, darunter auch *Mycobacterium tuberculosis*, zu den tödlichsten Erregern weltweit. Für ihren Infektionsmechanismus ist die Sekretion von Virulenzproteinen über ihre sonst äußerst undurchlässige Membran unabdingbar. Der Transport der heterodimeren Proteinkomplexe erfolgt durch ein hochspezialisiertes Sekretionssystem, das Typ VII oder ESX Sekretionssystem. Langsam-wachsende, meist pathogene Mykobakterien kodieren bis zu fünf verschiedene ESX Systeme in ihrem Genom, ESX-1 bis ESX-5 genannt, die verschiedene Aufgaben im mykobakteriellen Lebenszyklus erfüllen. Die ESX Systeme werden als vielversprechende Zielproteine für die Entwicklung neuer Medikamente gehandelt, doch der molekulare Mechanismus mit dem Substrate über die mykobakterielle Membran transportiert werden ist bisher nicht bekannt.

Diese Arbeit präsentiert die Struktur des ESX-5 Systems von *Mycobacterium xenopi* in der inneren Membran. Durch die Kombination von Kryoelektronenmikroskopie, Crosslinking-Massenspektrometrie und homologen Modellen, die für die Modellbildung eingesetzt wurden, wurden neue Einblicke in das hexamere Sekretionssystem mit zentraler Sekretionspore gewonnen. Der Komplex ist aus sechs identischen protomeren Einheiten aufgebaut, die jeweils von den Proteinkomponenten EccB₅:EccC₅:EccD₅ und EccE₅ in einer 1:1:2:1 Stöchiometrie gebildet werden. Während das Segment in der inneren Membran hoch aufgelöst wurde, zeigte die zentrale Sekretionspore, die aus zwei EccC₅-Transmembranhelices besteht, hingegen erhöhte Flexibilität, die durch einen konservierten Prolinrest im Zentrum der zweiten Helix induziert werden könnte. Des Weiteren wurde durch Analyse des Porendurchmessers deutlich, dass Phenylalaninreste eine Rolle in einem möglichen Schließmechanismus der Pore spielen könnten. Eine Mutationsstudie im Rahmen dieser Arbeit konnte zeigen, dass die Integrität der Reste in der Pore sowie die Integrität des Gerüsts der Pore, das von EccB₅ gebildet wird, essentiell sind für die Sekretion. Eine Kryoelektronenmikroskopiestruktur konnte bestätigen, dass die Mutation des zentralen Prolinrests zur Verfestigung der zentralen Helix führt. Dies verdeutlicht, dass die Plastizität der zentralen Transmembranpore essentiell ist, um den Transport verschiedener gefalteter Substrate über die Zellhülle zu ermöglichen.

Diese Flexibilität wurde auch im cytosolischen Bereich für die drei aneinandergereihten ATPase-Domänen von EccC₅ und im periplasmatischen Bereich für die periplasmatische Domäne von EccB₅ beobachtet. Eine Kryoelektronenmikroskopiestruktur konnte zeigen, dass die Zugabe von ATP γ S dazu beiträgt zwei EccC₅ Konformationen zu stabilisieren. Eine gewisse intrin-

sische Flexibilität, könnte jedoch notwendig sein um dem Komplex die erforderliche Dynamik zu verleihen, die für den Transport verschiedener Substrate notwendig ist. Dieser Prozess könnte große Konformationsänderungen erfordern.

Zusammen mit anderen vor Kurzem veröffentlichten Strukturen von ESX-5 und ESX-3 Systemen und verbesserten Methoden zur Vorhersage von Proteinstrukturen, konnten die hier gezeigten Ergebnisse den Weg für strukturelle Vergleiche ebnen und zu einem tieferen Verständnis des Sekretionsmechanismus beitragen.

List of Figures

1.1	<i>Mycobacterium tuberculosis</i> infection cycle	2
1.2	Comparison of the gram-negative, gram-positive and mycobacterial membrane composition	4
1.3	Mycobacterial secretion systems	6
1.4	Architecture of the five different ESX loci	7
1.5	Substrate classes and structures	12
1.6	Known structures of soluble protein domains of the membrane protein complex	17
1.7	Negative stain EM map of the <i>Mycobacterium xenopi</i> ESX-5 core complex .	18
2.1	Overview of the secretion assay workflow	42
3.1	Representative purification of the ESX-5 complex	44
3.2	High resolution structure of the ESX-5 complex	45
3.3	Low resolution EM density and crosslinks obtained by crosslinking MS	46
3.4	Composite ESX-5 EM map and structural model of the ESX-5 system	47
3.5	Cross-sections of the ESX-5 complex	47
3.6	Structural insights into the protomeric composition of the ESX-5 system . .	48
3.7	EccD ₅ cavity and unassigned density	49
3.8	Composition of the periplasmic domain	51
3.9	Interprotomeric interaction observed between EccB ₅ and EccD ₅	51
3.10	Domain swap interaction of EccC ₅ TMH1	52
3.11	Ensemble model of the transmembrane pore from ESX-5	53
3.12	Structural role of specific side chains in the ESX-5 secretion pore	54
3.13	EccB ₅ arrangement within the membrane	54
3.14	Interface of EccB ₅ and EccC ₅ within the membrane section	56
3.15	Impact of selected mutations on the secretion level of ESX-5 substrates . . .	58
3.16	Impact of selected mutations on the secretion level of ESX-5 substrates in comparison to cytosolic and transmembrane components	59
3.17	Purification and characterization of ESX-5 pore mutants P73A and F75C . .	60
3.18	Cryo-EM map of the ESX-5 P73A mutant and difference map	61
3.19	Molecular dynamics simulation of EccC ₅ TMH1 and TMH2 for ESX-5 WT and ESX-5 P73A	63
3.20	Limited proteolysis analysis of ESX-5	65
3.21	Reconstitution of ESX-5 WT into SapNPs	67

3.22	Direct extraction of ESX-5 WT by Saposin A	69
3.23	Sequence logo of Walker A motifs present in EccC	71
3.24	Binding of ATP γ S to the EccC ₅ ATPase domains	72
3.25	Cryo-EM maps of ESX-5 complex in presence of ATP γ S	73
3.26	Low resolution cryo-EM density of the ESX-5 complex in presence of ATP γ S	74
4.1	Available structures of T7SSs	78
4.2	Comparison of the <i>Mycobacterium xenopi</i> ESX-5 and the <i>Mycobacterium tuberculosis</i> ESX-5 structure	79
4.3	Angle analysis of EccB	80
4.4	Prediction of ESX-5 substrate complexes	82
4.5	Proposed model of T7SS mediated substrate translocation	86
5.1	Image processing workflow of an initial cryo-EM dataset collected on the <i>Mycobacterium xenopi</i> ESX-5 complex	93
5.2	Image processing workflow of a final cryo-EM dataset collected on the <i>Mycobacterium xenopi</i> ESX-5 complex	94
5.3	Image processing workflow of a cryo-EM dataset collected on the <i>Mycobacterium xenopi</i> ESX-5 P73A complex	95
5.4	Image processing workflow of a cryo-EM dataset collected on the <i>Mycobacterium xenopi</i> ESX-5 complex in presence of ATP γ S.	95

List of Tables

1.1	Comparison of bacterial secretion systems	5
1.2	Overview on ESX loci encoded in mycobacterial species	8
2.1	Common buffers	24
5.1	Fold-changes of protein abundances of mutants compared to ESX-5 WT	89
5.2	Statistical significance of secretion assay results	90
5.3	Cryo-EM and model building statistics	91
5.4	Cryo-EM and model building statistics	92

List of abbreviations

2D / 3D	Two-dimensional / Three-dimensional
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
ATP _γ S	Adenosine-5'-O-(3-thiotriphosphate)
BCG	Bacillus Calmette–Guérin
BL	Brain lipids
BN-PAGE	Blue native polyacrylamide gel electrophoresis
CFP-10	Culture filtrate protein - 10 kDa
CMC	Critical micelle concentration
CpnT	Channel protein with necrosis-inducing toxin
cryo-EM	Cryogenic electron microscopy
CTF	Contrast transfer function
DDM	n-Dodecyl-β-maltoside
DSS	Disuccinimidyl suberate
DUF	Domain of unknown function
E.coli	<i>Escherichia coli</i>
Ecc	ESX conserved component
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
EMDB	Electron Microscopy Data Bank
ESAT-6	Early secretory antigenic target - 6 kDa
ESCRT	Endosomal sorting complex required for transport
Esp	ESX-1 secretion associated protein
ESX	ESAT-6 secretion system
Esx	ESX associated secreted protein
EsxAB	EsxA and EsxB secreted dimer
FSC	Fourier shell correlation
GSFSC	Gold-standard fourier shell correlation
Hcp	Hemolysin-coregulated protein
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
HIV	Human immunodeficiency virus
HRP	Horseradish peroxidase
IMAC	Immobilized metal affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside

kDa	Kilodalton
LC-MS/MS	Liquid chromatography coupled tandem mass spectrometry
LTQ	Linear ion trap quadrupole
Mabs	<i>Mycobacterium abscessus</i>
MD	Molecular dynamics
MDa	Megadalton
Mlep	<i>Mycobacterium leprae</i>
Mmar	<i>Mycobacterium marinum</i>
MPTR	Major polymorphic tandem repeat family
MS	Mass spectrometry
MSA	Multiple sequence alignment
Msmeg	<i>Mycobacterium smegmatis</i>
MSP	Membrane scaffolding proteins
Mtb	<i>Mycobacterium tuberculosis</i>
MWCO	Molecular weight cut-off
Mxen	<i>Mycobacterium xenopi</i>
NAD ⁺	Nicotinamide adenine dinucleotide
NBD	Nucleotide binding domain
NHS	N-hydroxysuccinimide
NS-EM	Negative stain electron microscopy
o/n	Overnight
OD600	Optical density at 600 nm
PAE	Predicted aligned error
PBS	Phosphate-buffered saline
PDB	Protein Data Bank
PE	Proline-glutamate family protein
PGRS	Polymorphic GC-rich sequences family
pLDDT	Predicted local distance difference test
POPE	1-Palmitoyl-2-oleoylphosphatidylethanolamine
PPE	Proline-proline-glutamate family protein
ppm	Parts per million
RD1	Region of difference 1
RMSD	Root-mean-square deviation
RMSF	Root-mean-square fluctuation
rpm	Revolutions per minute
SapA	Sapasin A
SapNP	Sapasin-derived lipid nanoparticles
SDS-PAGE	Sodium dodecyl sulfate–polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SMA	Styrene maleic anhydride

SMALP	Styrene maleic anhydride lipid particle
Snm	Secretion in mycobacteria
SPR	Surface plasmon resonance
SVP	Serine-valine-proline family
T7SS	Type VII secretion system
Tat	Twin-arginine translocation
TB	Tuberculosis
TiME	Tube-forming proteins in the mycobacterial envelope
TMH	Transmembrane helix
TMT	Tandem mass tag
TNT	Tuberculosis necrotizing toxin
TPR	Tetratricopeptide repeat
Tris	Tris(hydroxymethyl)aminomethane
UBD	Ubiquitin-like domain
UPLC	Ultra performance liquid chromatography
WT	Wild-type

Scientific contributions

In the course of this PhD I performed many different experiments, with the main results described in this thesis. I performed molecular cloning, expression and purification, studied proteins with biophysical and biochemical tools, prepared and screened cryo-EM samples, processed cryo-EM data, was involved in molecular model building and interpreted data using bioinformatic tools. Also I set up crosslinking reactions, learned to use molecular dynamics simulation and prepared samples for MS analysis.

However, the results presented here benefited from the contribution and support of collaborators, whose input will be listed below:

- Katherine Beckham, Wilmanns group, EMBL Hamburg, supervised and supported the work. She performed crosslinking-MS and supported cloning of the mutant constructs as well as purification of the ESX-5 apo and ATP γ S sample.
- Grzegorz Chojnowski, Wilmanns group, EMBL Hamburg, built the final ESX-5 model and performed the angle and HOLE analysis [216].
- Jan Kosinski and Daniel Ziemianowicz, Kosinski group, EMBL Hamburg, built integrative models for the EccB₅ periplasmic domain and EccC₅ ATPase domains.
- Mandy Rettel and Frank Stein, proteomics core facility, EMBL Heidelberg, performed the LC-MS/MS run and supported data analysis.
- Simon Mortensen and Edukondalu Mullapudi, Wilmanns group, EMBL Hamburg, supported and supervised grid freezing and screening of the ESX-5 apo and ATP γ S sample, collected cryo-EM data and supervised cryo-EM processing.
- Felix Weis, cryo-EM facility, EMBL Heidelberg, collected cryo-EM data on the ESX-5 apo sample.

1 Introduction

1.1 Tuberculosis

1.1.1 Tuberculosis - a deadly disease

Despite being known for centuries, tuberculosis (TB) remains a major health threat worldwide until today [176, 17]. In 2021 alone, 10.6 million people fell ill with TB and 1.5 million people died from this infectious disease [176]. Therefore, TB is considered as the leading cause of death by an infectious disease [176].

The causative agent of TB is the bacterial pathogen *Mycobacterium tuberculosis* (Mtb). It is spread via aerosol droplets released by breathing or coughing of a person actively infected with Mtb [134, 39, 177]. The transmission does not necessarily lead to active TB. Instead, the immune system is often able to eliminate the bacteria or to control the infection locally, leading to latent TB [69]. It is estimated that a quarter of the world's population is carrying the pathogen [175]. While people with latent TB are not contagious and do not show clinical manifestations like coughing, chest pain and fever, they are at risk to fall ill with TB later in life [177]. The exact molecular mechanism of the reactivation of latent TB is poorly understood, but risk factors like immune deficiency, for example due to co-infection with human immunodeficiency virus (HIV), diabetes, alcohol and cigarette abuse as well as malnourishment are known to have an impact in that process [177, 175, 134]. In fact, 5% - 15% of people infected with latent TB develop the active disease during their lifetime [175, 177].

Although TB occurs all over the world, the highest burden is observed in the region of South-East Asia, Africa and the Western Pacific accounting for 86% of all TB cases worldwide in 2020 [175]. These include many low and middle income countries, where the high cost of treatments exceeds the income of every second household affected with TB by more than 20% [175]. The standard treatment of TB comprises a combination of four antimicrobial drugs that need to be taken in the course of six months. An interruption of the treatment bears the risk to develop multi-drug resistant TB, which is even more challenging to treat, often with poorer outcomes [177].

A cost effective protection can be provided by vaccination with an attenuated *Mycobacterium bovis* bacillus, Bacillus Calmette–Guérin (BCG), developed more than 100 years ago. Although it provides good protection for infection and disease development for young children, the efficacy of the vaccine is poor for adults and varies even between different geographical

locations [72].

In order to develop new and effective treatments against TB and other mycobacterial diseases, it is necessary to gain a deep understanding of the instruments and mechanisms with which mycobacteria orchestrate and finally disarm the immune response of the host.

1.1.2 Mycobacterial physiology

Mtb evolved a very complex infection cycle, as shown in Figure 1.1. Once inhaled, the bacteria enter the lungs and migrate to the alveoli, where they are sensed and taken up by innate immune cells such as alveolar macrophages. Despite the degrading environment in the (phago-) lysosome of the macrophage, Mtb is able to replicate inside the cell and induces cell lysis to infect further immune cells [134, 177, 39, 188]. The local inflammatory response caused, leads to the migration of a diverse set of immune cells, including B- and T-cells, to the site of infection that will enclose the infected cells in a so-called granuloma, characteristic of Mtb. In this heterogenous arrangement of immune cells the bacteria can be eliminated or contained. A destruction of the granuloma, however, leads to the active disease. At this point in the infection cycle, Mtb is released into the lung, from where the bacteria can further propagate to other organs [134, 177, 39, 188]. While it was long believed that granulomas are in particular beneficial for the host only by locking up Mtb in a restricted area, recent research suggests that granulomas represent a permissive environment for Mtb growth [50].

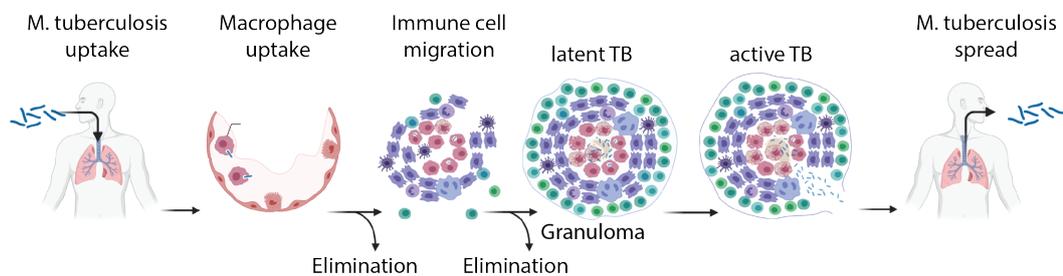


Figure 1.1: Mycobacterium tuberculosis infection cycle Mtb is inhaled and taken up by alveolar macrophages. The local inflammation leads to the recruitment of immune cells that contain Mtb in a closed capsular structure, the granuloma. The orchestration of the immune response allows the bacteria to survive and grow inside the granuloma. When the granuloma breaks down, Mtb can spread in the lung and to other organs, causing symptoms and the active disease. Coughing of a person with active TB releases droplets containing Mtb into the air [134, 177, 39, 188]. The figure was created with elements from BioRender.com.

Evading the chemical and mechanical stress and manipulating the immune response is possible, as Mtb has evolved a diverse set of features and strategies [39].

One of the most intriguing characteristics of Mtb is its unusual cell envelope that is depicted in Figure 1.2 and is a distinct feature of the order *Corynebacteriales*, especially the genus

Mycobacterium [75, 46]. The genus includes more than 170 different species populating different natural and human associated environments [81]. Besides Mtb, this genus comprises also other pathogens like *Mycobacterium leprae* (Mlep) or *Mycobacterium abscessus* (Mabs). Especially the latter one has evoked attention as it causes severe infections that are difficult to treat due to its intrinsic antibiotic resistance [129].

Although officially classified as gram-positive bacteria, *Corynebacteriales* possess, in contrast to gram-positive bacteria and gram-negative bacteria, besides a rather usual lipid bilayer as inner membrane, a peptidoglycan layer linked to a biopolymer composed of arabinose and galactose [240]. This polymer is attached to the so called mycomembrane by covalent bonds. The inner leaflet of the mycomembrane mainly comprises fatty acids with very long fatty acid chains (C60 - C90), the so-called mycolic acids [240, 75, 46]. These fatty acids render the bacteria very hydrophobic and impermeable to many compounds including broad-spectrum antibiotics [46]. The outer leaflet of the mycomembrane in turn is more heterogeneous and is composed of further mycolic acids, lipids and lipoglycans that provide an additional barrier [46]. A loosely attached capsular layer composed of polysaccharides, proteins and a few lipids has been shown to play a role in host pathogen interactions [133, 210]. In total, the amount of lipids accounts for 40% of the dry weight of Mtb and leads to 100 - 1000 times lower permeability of the membrane compared to gram-negative bacteria [127].

Nevertheless, to survive and to establish infection, it is inevitable for Mtb and other mycobacteria to possess a protein machinery that ensures transport of proteins across this otherwise almost impermeable cell envelope. This machinery is an achille's heel for the bacterium as it can be exploited as a potential drug target [28].

1.2 Secretion

1.2.1 Protein transport machineries

The presence of sophisticated protein machineries that allow transport of proteins across the cell membrane is common across all bacteria. These export pathways are essential for many different physiological processes, such as nutrient acquisition, adhesion to host cells, modulation of host-pathogen interactions, virulence and viability, which are mediated by the secreted proteins, also called the effector proteins [203, 107, 56]. Due to the presence of a single lipid bilayer, secretion in gram-positive bacteria requires the presence of pathways that are able to transport proteins across the cytoplasmic membrane only. The Sec and twin-arginine translocation pathway (Tat) are widely spread pathways that are able to fulfill this role [84]. Sec and Tat pathways also exist in gram-negative bacteria. However, further mechanisms are required in gram-negative bacteria to transport proteins across the outer membrane. Thus, Sec and Tat translocation systems are only considered accessory to se-

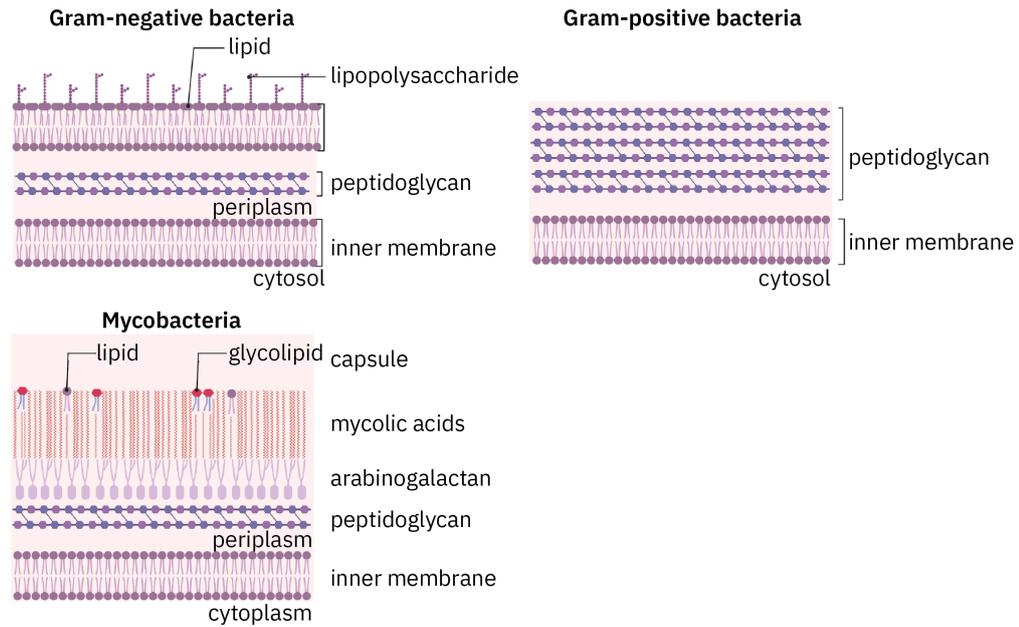


Figure 1.2: Comparison of the gram-negative, gram-positive and mycobacterial membrane composition The mycobacterial membrane is a highly complex structure that contains, in addition to a rather usual inner membrane and peptidoglycan layer, an arabinogalactan layer that is covalently linked to fatty acids with very long fatty acid chains, the so-called mycolic acids, that form the mycomembrane. A loosely attached layer composed of polysaccharides, proteins and lipids forms the outermost barrier of Mycobacteria [240, 75, 46]. The figure was created with elements from BioRender.com.

cretion systems in gram-negative bacteria like the type II or Type V secretion systems [84]. The classification into different types resulted from the conservation of secretion mechanisms between different gram-negative bacteria. So far ten types have been described [84].

The mechanisms of action of these different secretion systems are diverse in many aspects, such as: energy sources, target location of effectors, membranes to cross, substrates, secretion signals, number of steps involved in transport and many more (Table 1.1) [107, 56]. Some systems, like the well-described type III secretion system or the type IV secretion systems, can transport their effectors across the own cell envelope and the host cell membrane in just one step. It is important to mention that bacteria are not limited to one type of secretion system but multiple different secretion system types can coexist in a single bacterial species [107, 56, 84].

As the mycobacterial cell envelope comprises a special outer membrane and, thus, differs from gram-positive and gram-negative bacteria, the question arises, whether specialized mycobacterial pathways exist that allow transport of proteins across the whole mycobacterial cell envelope. Indeed, mycobacteria encode, in addition to a Tat and Sec translocation pathway, two special translocation systems: the SecA2 system that can be considered as paralogue to

the canonical Sec system, and the unique type VII secretion system (T7SS), the only official secretion system type in gram-positive bacteria (Figure 1.3) [83, 35, 84]. While Sec, Tat and also the specialized SecA2 export proteins to the periplasmic space, so far T7SS substrates could not be detected in the periplasm. Thus it is suggested that T7SSs are specialized to transport proteins across the cytoplasmic membrane and the mycomembrane [35].

Table 1.1: Comparison of bacterial secretion systems The table was modified from [107] with data from [158, 168, 113, 111, 222, 103, 180, 55]

Translocation system	Secretion signal	Steps in secretion	Folded substrates	Number of membranes	Gram(+) Gram(-)	Energy source
Sec	N-terminus	1	No	1	Both	ATP/PMF
Tat	N-terminus	1	Yes	1	Both	PMF
T1SS	C-terminus	1	No	2	Gram (-)	ATP
T2SS	N-terminus	2	Yes	1	Gram (-)	ATP
T3SS	N-terminus	1	No	2-3	Gram (-)	PMF
T4SS	C-terminus	1	No	2-3	Gram (-)	ATP
T5SS	N-terminus	2	No	1	Gram (-)	Assembly?
T6SS	Unknown	1	Unknown	2-3	Gram (-)	Unknown
T7SS	C-terminus	1	Yes	2	Gram (+)	ATP
T8SS	N-terminus	2	No	1	Gram (-)	Unknown
T9SS	C-terminus	2	Yes	2	Gram (-)	PMF
T10SS	Unknown	2	Unknown	2	Gram (-)	Unknown
T11SS	Unknown	2	Unknown	1-2	Gram (-)	Unknown

1.2.2 Secretion systems in mycobacteria

Sec, SecA2 and Tat translocation systems

The canonical Sec system is essential for the viability of mycobacteria. It is predicted that around 1000 proteins are targeted for transport by this system. Not all of them remain in the periplasm, some are secreted to the environment via yet unknown mechanisms [32]. The energy for the transport of the unfolded effector proteins through the channel formed by SecY, SecE and SecG (SecYEG) is provided by proton motive force and the ATPase SecA [257]. The proteins are targeted to the system with a positively charged N-terminus, a hydrophobic core and a positively charged C-terminus [167]. Besides transport of proteins, the Sec system is also involved in the insertion of integral membrane proteins in the membrane [151]. The mycobacterial SecA2 system exploits the components of the canonical Sec system. Only a different SecA ATPase, SecA2, is encoded in the genome. It has a high structural similarity to SecA1. Nevertheless, SecA2 has a distinctive role and cannot be replaced by SecA1 alone and vice versa [29]. So far, fewer (15) [236] proteins have been shown to be transported by the SecA2 system and it is suggested that in contrast to the canonical Sec system they could

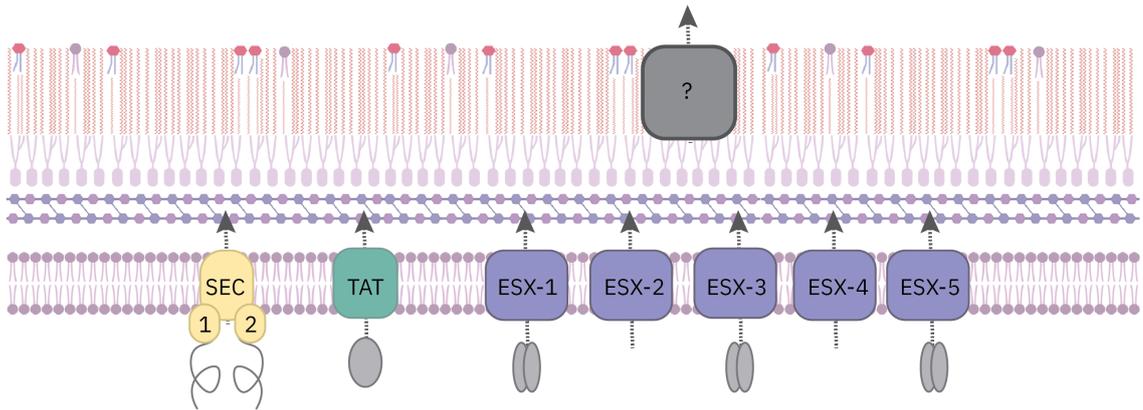


Figure 1.3: Mycobacterial secretion systems Mycobacteria contain a very limited set of protein translocation systems. Besides the canonical Tat and Sec system, mycobacteria encode for a second SecA ATPase and a very specialized secretion system, the ESX or Type VII secretion systems (T7SS) [35]. Up to five different ESX secretion systems can be found in mycobacteria, termed ESX-1 to ESX-5 [120]. Each of these seems to play a distinct role in the mycobacterial life cycle [35]. The components for substrate transport across the inner membrane have been identified [20, 121]. How the folded heterodimeric ESX substrates are transported across the outer membrane is currently unknown.

be even (partially) folded [82]. Overall, the system is not essential for viability, but it has been shown to be required for pathogenesis in pathogenic mycobacteria like *Mycobacterium marinum* (Mmar) and Mtb [30, 249].

The Tat system in turn has been shown to be essential only for Mtb. Other mycobacterial species only show growth defects when knocking out this system [179, 208, 190]. It is reported to transport around 100 folded proteins across the mycobacterial inner membrane via the TatABC machinery [156, 179]. The process is driven by proton motive force [16]. Characteristic for the substrates is their N-terminal secretion signal that comprises two arginines and two hydrophobic residues separated by a polar residue (R-R-X-h-h) [57]. These twin arginine residues are eponyms for the twin-arginine translocase (Tat).

Type VII secretion system

The discovery of the T7SS was initiated by the identification of its secreted targets EsxA (Early secretory antigenic target - 6 kDa (ESAT-6)) and EsxB (culture filtrate protein - 10 kDa (CFP-10)) around 20 years ago. By knocking out these two genes, first evidence arose that the encoded proteins may play a role in virulence [247]. Interestingly, both proteins were lacking the specific Tat and Sec secretion signals [223]. The attention was further drawn to ESAT-6 and CFP-10 when it became evident that the region flanking the two genes (region of difference 1 (RD1)) was deleted in the BCG vaccine strain, leading to its attenuation [155]. Soon afterwards, it was shown that the gene clusters around ESAT-6 and CFP-10 encode for a new, so far undescribed secretion pathway, which was initially named Snm for

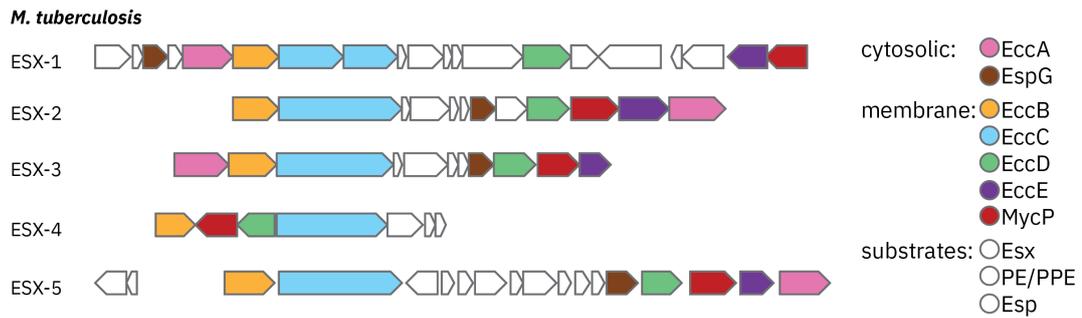


Figure 1.4: Architecture of the five different ESX loci Each ESX locus has a very similar composition and contains all the components needed for substrate transport: cytosolic components including an ATPase and a chaperone, membrane components, that form the secretion machinery across the inner membrane as well as substrates from three different substrate classes [25, 121].

secretion in mycobacteria [223], before later the terms ESX (ESAT-6 secretion system) and type VII secretion system prevailed [25, 1]. In addition to the ESAT-6 gene cluster, genomic and proteomic analysis revealed the presence of four more paralogous loci in Mtb that were numbered ESX-2 to ESX-5 respectively, with the first identified system termed ESX-1 [96, 230]. Each of these loci encodes the proteins, that form the T7SS secretion machinery: membrane proteins, ATPases, chaperones, proteases and substrates, as shown in Figure 1.4 [223, 96, 20, 121]. Conserved components, that are present in at least four of the five loci, were named ecc, for esx conserved component, with the exception of a subtilisin-like protease MycP and the cytosolic chaperone EspG, which was first assigned as an ESX-1 secretion-associated protein (Esp) [25].

The five ESX systems are not exclusively found in Mtb, but are spread across all mycobacterial species and even a few more actinobacterial and corynebacterial species. However, not all mycobacteria possess all five systems. In fact, two to seven gene clusters are observed in different mycobacterial species. These also include plasmid encoded ESX systems [170]. An overview of the distribution of ESX systems can be found in Table 1.2. Phylogenetic analysis suggests that ESX systems evolved by gene duplication from the most ancestral system ESX-4 with ESX-5 being the most recent evolved system [170]. It is striking that the presence of ESX-5 correlates with a slower growth rate of mycobacteria. These so-called slow-growing mycobacteria account for most of the pathogenic species including Mtb [170]. Due to the lack of the membrane porin MspA, their cell envelope is considered even less permeable [12]. Pathogenicity, thus, may have co-evolved with lower permeability of the mycobacterial membrane.

Table 1.2: Overview on ESX loci encoded in mycobacterial species The table was modified from [170, 146, 196]. A circle indicates the presence of the locus in the respective species. e indicates the presence of an ESX-4 duplicate. Plasmid encoded ESX systems are named according to their genetic ancestors. If the species is non-pathogenic (NP), an opportunistic pathogen (OP) or a true pathogen (TP) is indicated in the last column.

Species	ESX-1	ESX-2	ESX-3	ESX-4	ESX-5	Plasmid encoded ESX	Pathogenicity
<i>M. abscessus</i>			•	•			NP
<i>M. africanum</i>	•	•	•	•	•		TP
<i>M. avium</i>		•	•	•	•		OP
<i>M. bovis</i>	•	•	•	•	•		TP
<i>M. bovis</i>	•	•	•	•	•		TP
<i>M. canettii</i>	•	•	•	•	•		TP
<i>M. chubuense</i>	•			•		P2', P2'	NP
<i>M. colombiense</i>		•	•	•	•		OP
<i>M. fortuitum</i>	•		•	• e			OP
<i>M. gilvum</i>	•		•	•		P1	NP
<i>M. indicus pranii</i>		•	•	•	•		NP
<i>M. intracellulare</i>		•	•	•	•		OP
<i>M. kansasii</i>	•	•	•	•	•		OP
<i>M. leprae</i>	•		•		•		TP
<i>M. ulcerans</i>	•		•	•	•		TP
<i>M. marinum M</i>	•		•	•	•		TP
<i>M. massiliense</i>			•	•			OP
<i>M. microti</i>	•	•	•	•	•		TP
<i>M. neoaurum</i>	•		•	•			NP
<i>M. orygis</i>	•	•	•	•	•		TP
<i>M. parascrofulaceum</i>		•	•	•	•	P2, P2', P5	OP
<i>M. phlei</i>	•		•	•			NP
<i>M. rhodesiae</i>	•		•	•			OP
<i>M. smegmatis</i>	•		•	•			NP
<i>M. sp JLS.</i>	•		•	•			NP
<i>M. thermoresistibile</i>	•		•	•			NP
<i>M. tuberculosis</i>	•	•	•	•	•		TP
<i>M. tusciae</i>	•	•	•	•		P2, P2, P3	NP
<i>M. ulcerans</i>			•	•	•		TP
<i>M. vaccae</i>	•		•	• e			NP
<i>M. vanbaalenii</i>	•		•	•			NP
<i>M. xenopi</i>		•	•	•	•		OP
<i>M. yongonense</i>		•	•	•	•	P5	OP

ESX systems

The emergence of several ESX systems suggests that each of these systems plays a distinct role in the mycobacterial life cycle. Indeed, the systems are associated to distinct tasks, such as virulence, nutrient and metabolite uptake [35].

The best studied system, ESX-1, was shown to play a major role in virulence. One of the main pathways ESX-1 is involved in, is the rupture of the phagosomal membrane and the cytosolic escape that was observed in *Mtb* and *Mmar* [215, 119]. This allows the pathogens to replicate in a nutrient rich and less hostile environment [126]. Responsible for the rupture could be the ESX-1 substrate *EsxA*, that was shown to have membrane lysing activity [68, 217]. However, the exact mechanism is far from being understood. Also the induction of necrotic like cell death, which allows the bacteria to spread, can be traced back to the ESX-1 system [215]. This mechanism was further suggested to play an important role in granuloma formation [109, 67, 224]. In addition, in non-pathogenic mycobacteria, ESX-1 seems to play an important role in a special form of horizontal gene transfer, the distributive conjugal transfer. The mechanism, that allows the translocation of large chromosomal fragments is also dependent on the ESX-4 system [106, 105]. A new preprint further describes a role of ESX-1 in nitrogen sensing based on the observation of ESX-1 megacomplexes at cell-cell contacts in low nitrogen environments [115].

The ESX-2 system in contrast is one of the least described systems and evolved more recently. Like the ESX-5 system, this system is associated with slow-growing mycobacteria only, but is not required for the slow growth as it is not encoded in the genome of several slow-growing species like *Mmar* [170]. Until recently, no function could be assigned to the ESX-2 system. However, in 2021, it was reported that in *Mtb* ESX-2, in combination with ESX-4 and ESX-1, is involved in phagosomal membrane rupture needed for the translocation of the tuberculosis necrotizing toxin (TNT), a nicotinamide adenine dinucleotide (NAD⁺) glycohydrolase, to the cytosol [178]. Interestingly, the suggested mechanism contradicts the pathway for TNT translocation in *Mmar* in which no involvement of ESX-2 was observed [125].

Essential for the viability of mycobacteria is the ESX-3 system, due to its crucial role in iron-zinc homeostasis [212, 213]. It was not only shown that ESX-3 is responsible for the uptake of iron [212, 213] but the transcription itself is upregulated under low iron conditions [154, 204]. Interestingly, other substrates than the *EsxA/EsxB* (*EsxAB*) paralogue pair *EsxGH* are involved in the mediation of iron zinc regulation. *EsxGH* instead were shown to interfere with the host immune system by disrupting the function of the endosomal sorting complex required for transport (ESCRT) to impair phagosome maturation. As a result, also ESX-3 plays a major role in virulence [157, 233, 76].

The most ancient system, ESX-4, has a reduced number of genes compared to other ESX-loci. Besides a smaller number of substrates, it is missing two cytosolic components (EspG and EccA) as well as one membrane component (EccE) [25]. The essential role in horizontal gene transfer in *Mycobacterium smegmatis* (Msmeg) was for a long time the only function that could be attributed to ESX-4. Therefore, it was believed that this system is not functional, with the ESX-4 system of *M. abscessus* being the only exception. In this species, it was shown that ESX-4 and its secreted substrates EsxUT are involved in phagosomal rupture comparable to the ESX-1 system. In this context it is noteworthy that the ESX-4 locus in *M. abscessus* comprises a copy of the missing membrane component EccE [144, 145]. In addition, also in other species the knowledge has advanced. In a recent research article, Pajuelo *et al.* [178], investigated the trafficking of the toxin TNT in *Mtb* and suggested that the toxin TNT or rather its precursor CpnT (channel protein with necrosis-inducing toxin) is transported through the membrane via ESX-4. Also the transport of the EsxAB homologues EsxE and EsxF was suggested to depend on ESX-4 [178]. In contrary, a study in *M. marinum* came to the conclusion that CpnT is transported through ESX-5 and grants only a yet unknown supporting role to ESX-4 [125]. However, both studies highlight an interplay between several ESX systems. An interplay was further observed between ESX-4, ESX-1 and ESX-5 in *M. marinum*, where ESX-4 seems to have a regulating effect on ESX-1 and ESX-5 secretion [246]. Understanding the cross-talk between different ESX systems is a growing field of research.

In addition to the functions touched on above, the most recently evolved system, ESX-5, has an essential role for growth in slow-growing mycobacteria. This is due to its function in nutrient uptake, which is mediated by some of the so called PE/PPE substrate pairs [12, 244, 139], which will be introduced in more detail later. These substrate pairs are encoded throughout the genome, but the majority is transported by ESX-5 [12, 120]. It is suggested that some of the PE/PPE proteins form selective channels in the outer membrane to allow import of nutrients like various carbon sources [244, 12, 139]. Interestingly, the importance of ESX-5 for growth can be circumvented when expressing the outer membrane porin MspA, which is encoded in the genome of fast-growing mycobacteria but not in slow-growing bacteria [12]. Besides its role in nutrient uptake, ESX-5 and ESX-5 substrates have been shown to play immune modulatory roles by impacting the cytokine production of human macrophages during infection [2] or by altering the capsular layer, which itself is important for ESX-1 dependent membrane rupture (hemolysis) [13]. Another study further indicated attenuation of *Mtb* ESX-5 mutants in macrophages and mice [27]. Taken together, these findings suggest that a functional ESX-5 system is required for virulence and nutrient uptake of pathogenic mycobacteria.

Substrates

As highlighted in the previous subsection, the function of the different systems is mediated by its secreted effectors. Overall, the substrates of ESX systems can be categorized into three different families: Esx proteins, PE/PPE proteins and Esp proteins. All of them are thought to be transported as folded 4-helix bundles across the membrane [61]. A C-terminal YxxxD/E motif in close proximity to a WXG motif, which is usually located in a loop region between two helices [63, 191, 201, 225, 137, 138] are suggested to form a two-component secretion signal (Figure 1.5) [35, 9]. However, not all substrates encode both signal sequences. Substrates lacking one component, which represent a vast majority of the substrates, were shown to be secreted as heterodimers, in which one protein complements the secretion signal of the other one [191, 201, 225, 137, 138]. These heterodimers are found in bicistronic operons [191].

One example of heterodimeric secreted proteins are Esx proteins. Esx proteins comprise a group of short proteins with a length of around 100 amino acids, that carry the conserved WxG motif and are therefore classified in the WxG100 family [191]. The best studied representatives of this group are EsxA (ESAT-6) and EsxB (CFP-10), which are located in the ESX-1 locus. Both proteins form an antiparallel heterodimer with flexible N- and C termini [191, 201]. While both components are carrying the WxG motif, EsxB comprises in addition the YxxxD/E motif at its C-terminus. Therefore, the WxG proteins are further divided into subclasses related to EsxA and EsxB or to another subgroup based on the sequence similarity to an EsxA protein of *Streptococcus agalactiae* [191]. Some Esx proteins are thought to have pore forming activities [68, 217, 145]. One example is EsxA, which was suggested to dissociate from EsxB at low pH after secretion to mediate phagosomal rupture [68, 217]. However, these results were recently questioned [52]. Another example is EsxEF, which was shown recently to form heteropentameric pores that were suggested to be essential for the secretion of TNT [229].

In Mtb the PE and PPE protein family are the substrates with the highest abundance. Around 10% of the whole Mtb genome encodes for members of this family [51]. Also in other pathogenic mycobacteria, genes encoding PE/PPE proteins are of high abundance [4]. They are characterized by a Proline-Glutamate (PE) or Proline-Proline-Glutamate (PPE) motif near their N-termini, respectively. Those have inspired also the naming of the families [51, 25]. Similar to Esx proteins, PE proteins form antiparallel heterodimers with their corresponding PPE protein [225, 137, 44, 252, 77]. The conserved core of this heterodimer is composed of the around 110 amino acid long N-terminal tail of PE and the around 180 amino acid long N-terminus of its PPE partner [9]. While the YxxxD/E motif is harbored at the C-terminus of the PE protein, the WxG motif is found in the loop region of PPE proteins. Again, both motifs are in close proximity to each other [225, 137, 44, 252, 77]. Albeit the core is quite conserved, the C-termini of PE and PPE proteins are highly variable in size and sequence [9]. These C-terminal extensions are also used to further divide this big group into

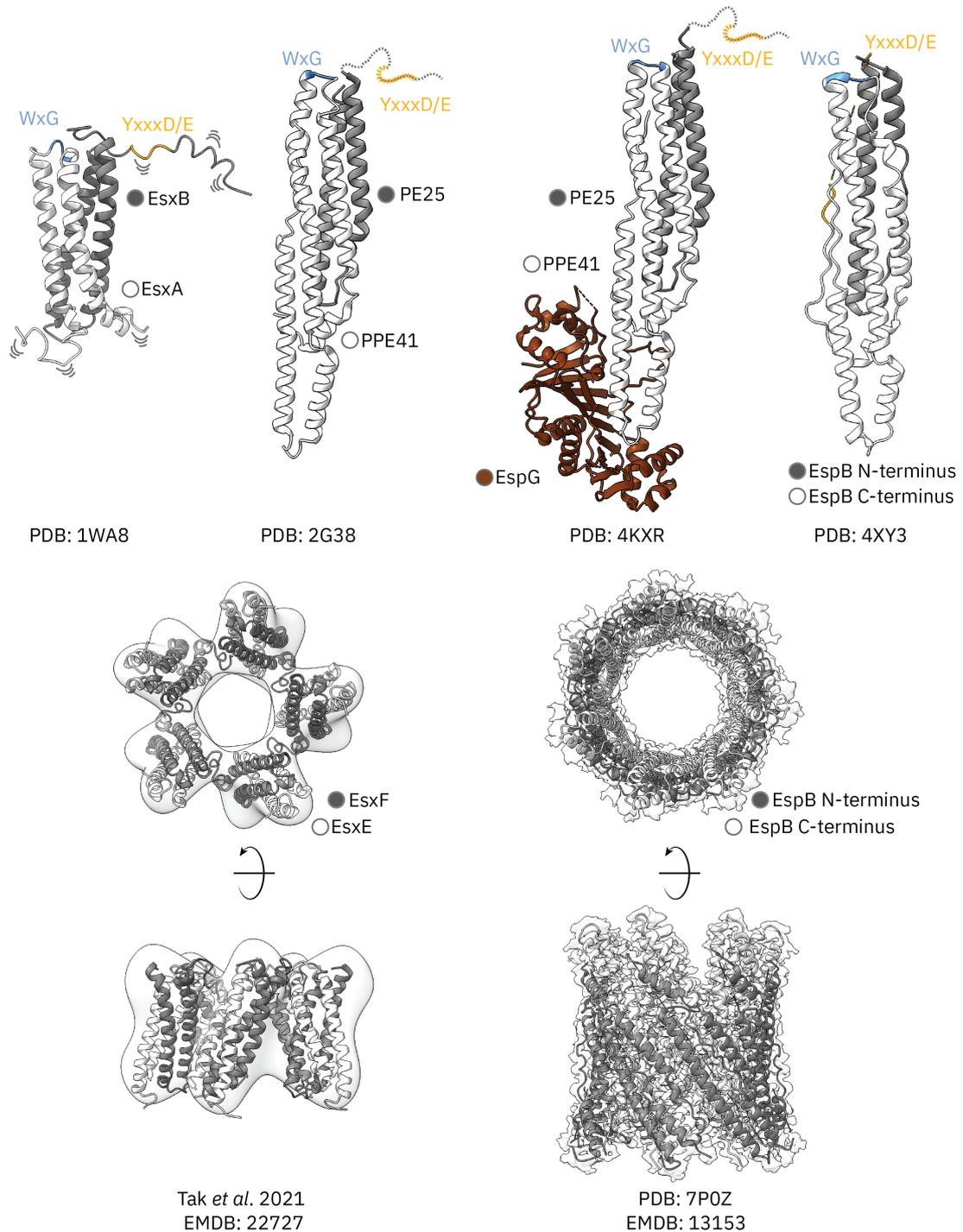


Figure 1.5: Substrate classes and structures Representative structures of ESX substrate classes are shown. Common to all substrates is the formation of 4 helix bundles and the presence of two motifs: the WxG and the YxxxD/E motif, shown in blue and in orange respectively. While PE/PPE and Esx proteins form heterodimers, EspB exists as monomer. PPE proteins can bind to the cytosolic chaperone EspG (brown). Recent cryo-EM studies report the formation of pore forming structures of EsxE and EsxF as well as EspB. PDBs and EMBDs presented were published in: [201, 225, 137, 138, 229, 98]

several subclasses [9]. One of these subclasses are the polymorphic GC-rich PE sequences (PGRS). Due to a very high GC content of 80% or more and long glycine rich sequences of the PE protein (up to 1400 amino acids) studying this subgroup is very challenging [9, 97]. 65% of the PE proteins annotated in the Mtb genome can be assigned to this class. Other classes correspond to the ESX-loci they are encoded in, some are more loosely defined [9, 97].

In contrast, PPE proteins are classified more specifically on sequence motifs. PPE-PPW proteins carry a PxxPxxW motif close to the C-terminus, PPE-SVP proteins contain a SVP motif in the C-terminal domain and PPE-MPTR proteins are characterized by a NxGxGNxG sequence, a major polymorphic tandem repeat (MPTR) [9].

Except for the PPW family, whose members have so far been shown to depend on the ESX-3 system and play a role in iron homeostasis [213, 233, 161, 234, 137] as well as the PE and PPE proteins encoded in the respective loci, PE/PPE dimers are thought to be transported through the ESX-5 system. Interestingly, a few members of the PPE-SVP subfamily have been shown to enhance transport of other substrates. Especially PPE38 is to highlight in this context as it was observed to facilitate the transport of PE PGRS and PPE-MPTR proteins in general [10, 11].

Many PE/PPE proteins are still uncharacterized and the scientific knowledge is limited [9]. Only a few members were assigned a function, but the data is not always very unambiguous. For example, some studies suggest that PE PGRS proteins are involved in virulence mechanisms, others claim a role in virulence attenuation (discussed in [9, 85]). Quite recently, more and more evidence arose that attributes important roles to some PE/PPE pairs in the uptake of nutrients, like glycerol, glucose and Mg^{2+} . Some PE/PPE proteins might thus act similar as outer membrane porins in gram-negative bacteria [244, 15, 139, 171, 70].

Esp proteins are specifically associated with ESX-1 secretion [25]. The only exception is the cytosolic chaperone EspG, which is also found in the ESX-2, ESX-3 and ESX-5 loci and will be introduced later. Not all Esp proteins were shown to be secreted, some, like EspD, were hypothesized to serve as chaperones for other Esp substrates [43, 187]. Others play a dual role in secretion and transcription regulation, like EspR, which is regulating the expression of the espACD operon [206]. A striking feature of this operon is that mutations within these genes lead to general ESX-1 secretion defects [87, 153]. These results highly suggest a strong co-dependence of the Esp substrates with each other and with other secreted substrates. Several mechanisms have been suggested that regulate the expression of this operon in response to certain environmental conditions [92, 143, 181, 132, 104, 26, 206]. Thus, the operon might serve as a molecular switch for ESX-1 secretion.

The only secreted Esp protein structurally characterized is EspB. It is secreted monomeric,

but the 4-helix bundle structure resembles PE/PPE heterodimers [138, 221, 189, 98]. After transport across the inner membrane, the protein is processed by the conserved protease MycP [221, 138]. By now several studies show in addition an oligomerization of EspB into a barrel-like structure that could potentially insert into membranes or bridge the periplasm via a yet unknown mechanism [221, 189, 98]. Also EspC was shown to have the tendency to oligomerize. EM micrographs indicated the formation of large surface exposed filaments, whose role is not elucidated yet [150].

All in all, despite their classification into different groups, PE/PPE, Esx and Esp proteins should not be considered separately. As indicated before, multiple studies point out that there exists a co-dependence and hierarchy of different substrate pairs, which is not fully understood yet [59, 37, 64, 43, 187, 206]. In this context, it is of equally importance to understand the interactions of substrates with the secretion machinery itself and the molecular mechanism of substrate translocation.

Components of the secretion machinery

The conserved components EccA, EccB, EccC, EccD, EccE, MycP and EspG form the core of the ESX secretion machinery. Each locus contains a copy of these proteins with the exception of EccA, EccE and EspG not present in the ESX-4 locus of most mycobacterial species [25].

In the cytosol, EspG serves as specific chaperone for PE/PPE proteins, which have a high tendency to self-aggregate [186]. A hydrophobic pocket in EspG represents a docking platform for a hydrophobic helical tip in the PPE protein [137, 44, 252, 77]. The interaction ensures that the secretion signal at the C-terminal end of the protein remains exposed [63]. By employing mutations and domain swap experiments it was shown that the EspG-PPE interaction further plays an important role in the targeting of PE/PPE proteins through the corresponding ESX system [186, 64]. The low sequence homology of EspG and the structural variation of the binding domain support these findings. Whether EspG itself interacts with the membrane machinery remains unknown [61].

Also for the second conserved cytosolic component, EccA, a lot of open questions remain. The protein comprises two domains, a tetratricopeptide (TPR) domain at its N-terminus and a AAA+ ATPase domain at its C-terminus that are linked via a flexible linker [152]. TPR domains are known to mediate protein-protein interactions [65] and, indeed, yeast 2-hybrid screens revealed interactions of EccA1 with PPE68, EspC [152, 42] while pull-down assays of EccA1 and EccA3 even showed interactions with components of the mycolic acid synthesis pathway [130, 141]. However, none of these interactions could be confirmed *in vitro* so far. The oligomeric state of EccA is expected to be hexameric as the ATPase domain belongs

to the family of AAA+ ATPases that are known to hexamerize to adapt an active state [228, 95]. It was shown that ATPase hydrolysis is required for EccA in order to fulfill its role in secretion [130]. Nonetheless, this role is far from being understood. It was proposed that EccA could act as dissociase by destabilizing the interaction of EspG to its PE/PPE substrates [61]. Another hypothesis is that EccA could act as an instigator, by delivering Esx substrates to the membrane protein complex and providing energy for the first steps in translocation [61]. Interestingly, it is still under debate if EccA is required for secretion. Some data point a pivotal role to EccA in secretion [93, 3, 27, 130], whilst some studies point to a minor role of EccA [53, 121]. These different results could be explained by a study, which attributed these variations to different culture media and growth conditions [185].

EccA is not the only ATPase that powers ESX secretion. Within the membrane machinery, composed of EccB, EccC, EccD, EccE and MycP, the protein EccC comprises 3 consecutive ATPase domains at its C-terminus which are classified as Ftsk/SpoIII ATPases (Figure 1.6, blue structure) [121]. Due to their high flexibility, structural studies of these cytoplasmic domains have been impeded [205, 20, 80, 192]. From the three nucleotide binding domains (NBD), two of them (NBD2 and NBD3) were shown, by mutational analysis, to enhance secretion, but their role is not essential. These results can be explained by a degenerated Walker B motif in both of these domains. Only mutations in the first ATPase domain abolished secretion, suggesting a regulatory role of the two C-terminal ATPases and a driving role of NBD1 [12, 205]. Interestingly, in ESX-1, ATPase domains 2 and 3 are even separately encoded as a two domain protein. Based on a crystal structure of a homologous EccC in *T. curvata*, a second regulatory role was assigned to the linker between ATPase1 and ATPase2. This linker (linker2) is binding to a pocket of ATPase2, presumably leading to a lower hydrolysis rate [205]. By domain swap experiments, it was observed that this linker is a determinant in ESX-5 for the species specific secretion of a PE/PPE substrate pair [38]. Hence, the release of linker2 might enable binding of PE/PPE substrates. Similar to EccA, EccC belongs to a family of ATPases that is known to oligomerize into hexamers. Induction of this oligomerization event was suggested to be mediated by binding of the YxxxD/E motif of an EsxB homologue in an empty binding pocket of NBD3 [245, 205]. As a consequence, the regulatory linker domain could be released from the NBD1 binding pocket to enable transport. In this model it is ensured that energy is only consumed in the presence of substrate [205]. Thus, secretion could be modulated by the expression of the substrates of the corresponding ESX system [205]. Beside the ATPase domains, the 150 kDa protein comprises at its N-terminus two transmembrane helices, followed by a long helical linker to a domain of unknown function (DUF). Although the classical catalytic residues are missing, it was suggested that this domain could function as an additional ATPase. Mutations in the predicted catalytic cavity, indeed, prevented secretion [80]. Whether this is due to abolished ATP hydrolysis or if the DUF domain plays a rather structural role in secretion needs to be clarified.

No obvious function was assigned to the protein EccB yet, although EccB₅, for example, has been shown to be essential for the viability of Mtb [71]. The 50 kDa protein is composed of a single transmembrane helix at its N-terminus followed by a big periplasmic domain. Two crystal structures of this periplasmic domain of EccB₁, from Msmeg and Mtb, revealed a pseudo two-fold symmetry of the protein, formed by a central domain and four repeat domains, which flank the central domain and imply a slight S-Shape (1.6, orange structure) [242, 256]. By using ATPase assays, a study suggested that the periplasmic domain of EccB could be involved in ATP hydrolysis [256]. As the periplasmic space is free of ATP, these results are questioned [35]. While structural homology search revealed no similar structures of the whole periplasmic domain, the repeat domains show low sequence homology to the phage protein PlyCB [242]. This protein is known to bind to peptidoglycan [169]. Thus, it was suggested that EccB could tether the complex to the peptidoglycan layer [242], but evidence for this hypothesis is still missing.

Also the role of EccD is not understood yet. With 11 predicted transmembrane helices, it is estimated to form an essential part of the membrane protein complex [242, 121]. Transposon mutations in EccD₅ abolished secretion and impeded assembly of the complex, which supports the hypothesis of a central role of EccD [121]. Besides the transmembrane domain, the N-terminus of EccD is located in the cytosol and is connected via a long linker with the transmembrane domain [242]. By close analysis of a crystal structure of the soluble domain of Mtb EccDB₁, a high similarity to ubiquitin-like domains (UBD) was observed (Figure 1.6, green structure) [242]. The function of this domain in secretion has not been elucidated so far.

In contrast to all other components, experimental structural information of single EccE domains are missing. As a deletion of Mxen EccE₅ did not affect the assembly of the secretion system, it is estimated that the 50 kDa protein is located at the periphery of the complex [20]. This finding is also supported by the fact that EccE is not encoded in the ESX-4 locus [120]. Interestingly, an Msmeg ESX-3 construct, lacking EccE, was incapable to secrete the corresponding Esx proteins EsxG and EsxH [80]. Also in Mtb ESX-1, deletion of EccE abolished secretion and had a negative impact on stability of the membrane complex [219]. In addition, in the exceptional case of Mabs ESX-4, which encodes an EccE₄, ESX-4 was active in contrast to other ESX-4 systems [144]. These observations grant EccE an important role in secretion.

Although MycP is thought to interact only transiently with the complex [239, 238, 121], the role of this protease was investigated in more detail. It belongs to the family of subtilisin-like proteases that are serine proteases, which comprise a catalytic domain with a catalytic triad formed by aspartate, histidine and serine [33]. Crystal structures of Mtb and Msmeg MycP₁

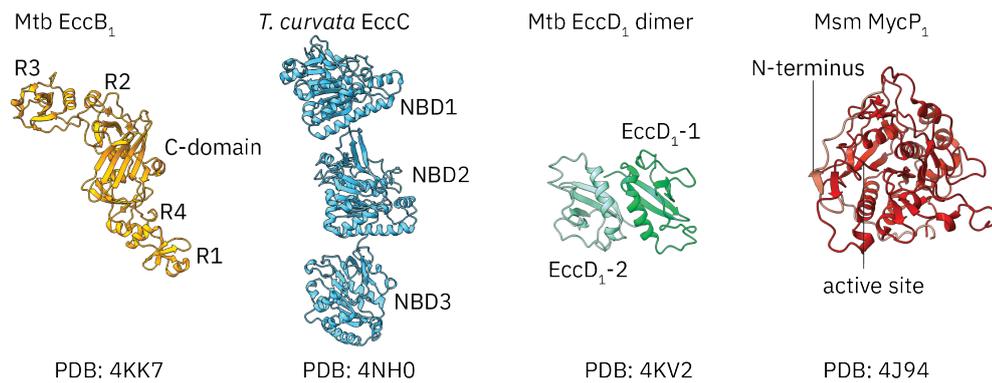


Figure 1.6: Known structures of soluble protein domains of the membrane protein complex The periplasmic domain of EccB, the three consecutive ATPase domains of EccC, a dimer of the UBD of EccD and the protease domain of MycP are shown. So far no structure of the soluble EccE domain could be obtained and it is not clear how the proteins are assembled within the complex. Structures presented were published in [242, 205, 220].

as well as Mtb MycP₃ [220, 243] confirmed a subtilisin like fold (Figure 1.6, red structure). In contrast to other subtilisin-like proteases, however, the N-terminal domain, which is thought to regulate protease activity, is not cleaved during maturation of the enzyme but rather binds around the active site instead of inside the cavity. The interaction of the N-terminal propeptide is even stabilized by a disulphide bond [220, 243]. Two detailed analyses of the propeptide of MycP revealed a structural role of the propeptide, rather than a functional role [226, 238]. It seems obvious that an involvement of a protease domain in secretion is the processing of substrates, but so far, cleavage was only observed for the ESX-1 substrate EspB, which is processed at its C-terminus. In addition, while it was shown that MycP is required for secretion through ESX systems, the protease activity of MycP is not. This was demonstrated by mutations of the active site of MycP₁ and MycP₅ [239, 174]. It was further reported that the transmembrane domain and the protease domain are stabilizing the membrane protein complex via a transient interaction [238, 239, 174].

Membrane protein complex

To decipher the molecular details and mechanisms of the membrane transport, it is not sufficient to focus on single components. Instead, it is essential to consider the interplay of all components.

As introduced before, each of the membrane components EccB, EccC, EccD, EccE and MycP are required to form a functional secretion system [121, 12, 205, 71, 80, 219, 239, 238]. However, it is unclear how they are assembled and concerted to enable transport. Deletions of single components suggested a peripheral position of EccE and MycP, while EccB, EccC and EccD were suggested to form the core of the secretion system [20, 239, 238].

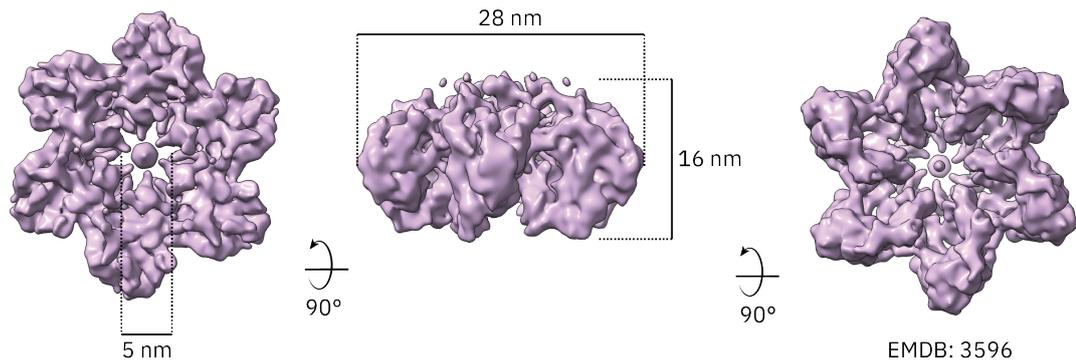


Figure 1.7: Negative stain EM map of the *Mycobacterium xenopi* ESX-5 core complex

The low-resolution map revealed a hexameric complex with a central secretion pore with dimensions that suggest the position in the inner membrane. The map presented was obtained by Beckham *et al.* [20].

First conclusions of the size of the complex were drawn from blue native polyacrylamide gel electrophoresis (BN-PAGE). The ESX-1 and ESX-5 complexes studied were formed by EccB, EccC, EccD and EccE and migrated at a size of >1.5 MDa indicating the formation of a high oligomeric complex [121]. A negative stain EM (NS-EM) structure of ESX-5 from *Mycobacterium xenopi* (Mxen) gave first insights in the overall dimensions and composition of the membrane protein complex formed by the Ecc components [20]. The structure revealed a C₆-symmetric assembly that formed around a central pore of estimated 5 nm diameter, shown in Figure 1.7. Mass spectrometry (MS) analysis revealed an equimolar stoichiometry of EccB, EccC, EccD and EccE, indicating a hexameric assembly of six identical protomers with a size of around 1.8 MDa. In addition, gold-labeling of EccC as well as fitting of the crystal structure of the soluble domain of EccB into the NS-EM map enabled the localisation of non-membrane embedded protein domains either into the cytoplasm or the periplasm. The findings suggest that the three characteristic ATPase (FtsK/SpoIIIE) domains of EccC are located in the cytoplasm and EccB in the periplasm. The location of MycP was investigated by a study, which used a chimeric EccB-MycP fusion to conclude that the periplasmic domain of EccB and the protease domain of MycP work in close proximity [237]. Although EccC could be located inside the cytoplasm, the ATPase domains are not included in the NS-EM map due to the high flexibility. It was speculated that this flexibility could be a special feature of a yet not characterized substrate binding mode [20].

Surprisingly, based on the overall dimensions of the NS-EM map of Mxen ESX-5, it was further concluded that the complex is only spanning the inner membrane and does not reach to the mycomembrane [20]. Thus, additional components are needed to mediate protein translocation across the outer membrane. Given the co-dependence of several ESX substrates as well as their tendency to oligomerize, it has been speculated that a substrate

could mediate transport [13, 221, 229]. Several substrates have been proposed: EspB, EspC, EsxEF and recently also a group called TiME proteins (tube-forming proteins in the mycobacterial envelope) that harbour either Tat, Sec or T7SS secretion signals [40]. Clear evidence for a facilitating role in transport across the outer membrane is, however, still missing. For EspB, for example, there are reasonable doubts that an EspB oligomer could insert into the membrane, since structures of an EspB oligomer revealed a charged surface on the outside of the pore [98, 189]. Instead, it was suggested that it could bridge the periplasmic space. The existence of such a bridge is also not proven yet. As a consequence, it is not evident whether type VII secretion is a one or two-step process.

1.3 Aim of this work

Since the discovery of the ESX/ T7SS, a lot of effort has been made to understand the molecular mechanism of secretion in mycobacteria. However, little is known about how the virulence factors in the cytosol are recognized by the T7SS, how effectors are translocated to the periplasm or how effectors cross the outer membrane. To unravel the molecular details of the translocation process mediated by T7SSs, high resolution structures of the entire complex are needed.

This thesis focuses on the structural elucidation of the ESX-5 complex in the inner membrane of mycobacteria. As this complex is the core of the T7SS and inevitable for viability and virulence of slow-growing, pathogenic mycobacteria, it could represent an ideal target for new treatments against TB or other mycobacterial diseases [28]. A structural and molecular understanding could, thus, advance the discovery of novel anti-TB drugs. In addition, given the high conservation of ESX-loci, it is expected that based on an ESX-5 structure, conclusions about a general T7SS architecture and mechanism can be drawn.

To study the ESX-5 structure, the Mxen ESX-5 complex was selected, as it showed superior stability and a good expression level [20]. In addition, the protein components share a high sequence identity (55% - 85%) to other ESX-5 systems. For the recombinant expression the non-pathogenic and fast-growing mycobacterial species *Msmeg* will be used as expression host. This ensures a fast growth and the integration of the membrane complex in an almost natural environment, compared to other bacterial expression systems like *Escherichia coli* (*E.coli*). Due to the size of the complex of over 1.8 MDa, single particle cryo EM was chosen as main method for the structural characterization. Complemented with biophysical and biochemical assays, the data presented aims to provide a deeper understanding on the overall secretion pathway.

1.4 Main methods

1.4.1 Expression host *Mycobacterium smegmatis*

For the recombinant expression of bacterial genes, often *E. coli* is considered as suitable expression host. However, only one third of the mycobacterial proteins were reported to be expressed, soluble and properly folded in *E.coli* [102, 18].

This might be, because mycobacteria evolved and adapted to their environmental niche and developed specific physiological needs and pathways. The absence of specific chaperones, post-translational modifications, metabolites and ligands but also the absence of a transcription and translation machinery adapted to the high GC-content of the mycobacterial genes could impair protein production in *E.coli* [18, 22]. The unique cell envelope of mycobacteria is an additional mycobacterial feature that differs tremendously from *E. coli*, as discussed before. As a consequence, proper insertion and folding of membrane proteins in *E.coli* could be impaired. Therefore it is likely, that the use of an expression host as closely related as possible increases the chance to obtain properly folded membrane proteins and membrane protein complexes as the T7SS.

The non-pathogenic, fast-growing, mycobacterial species Msmeg has been shown to be a suitable expression host for mycobacterial genes. It was demonstrated, that five of eight proteins from Mtb that were insoluble when expressed in *E.coli* were found soluble in Msmeg [102]. The doubling time of Msmeg of around three hours is significantly higher than in *E. coli* (20 minutes) [18] but seems worthwhile, given the higher quality and yields expected.

The commonly used expression strain Msmeg mc²155 was derived from the wild-type (WT) ATCC 607 strain and offers increased transformation efficiency compared to the wild-type strain [218]. A deletion of the histidine-rich C-terminus of GroEL, that caused contaminations after immobilized metal affinity chromatography (IMAC), further improved the strain for purification purposes [172].

By now, there is a great variety of vectors available for mycobacterial expression, including different inducible and constitutive promoters with an assortment of tags and antibiotic resistance genes, which give scope for different applications [18, 22, 162].

Based on the flexibility offered by the Msmeg expression system and the similarity of Msmeg to Mxen and other mycobacterial species, Msmeg was selected as expression host for the production of the Mxen ESX-5 system.

1.4.2 Electron microscopy

The main method applied for structure determination of purified T7SS was single particle cryogenic electron microscopy (cryo-EM). Due to many technical advances within the recent years, cryo-EM has been shown to be a powerful tool for the 3D reconstruction of macromolecular complexes to resolutions below 4 Å, in some cases even to atomic resolution [142, 166, 255]. In contrast to X-ray crystallography, cryo-EM does not require the growth of crystals. Instead, proteins are imaged in a thin layer of amorphous ice. This is achieved by rapidly freezing the sample on a cryo-EM grid in liquid ethane, also referred to as plunge-freezing or vitrification [74, 6].

At cryogenic temperatures the sample is stable in the column vacuum of the electron microscope and the low temperatures reduce the radiation damage induced by the electron beam used for imaging. Nevertheless, the imaging dose needs to be kept low (10 - 50 e⁻ per Å²) to preserve high resolution details [99]. However, since biological samples are composed of elements with weak intrinsic scattering potential, their phase contrast is rather low, leading to low signal to noise ratios [100]. Therefore, to achieve high resolution reconstructions, images of thousands of particles are needed. These images represent the two dimensional projections of each particle. With the help of computer software the individual images are aligned and averaged. In Fourier space, the projections of different orientations, correspond to different slices through the 3D volume. Modern computer software is, thus, able to reconstruct the 3D volume of the particles based on 2D images [88, 62]. The obtained resolution is estimated based on the Fourier Shell Correlation (FSC or GSFSC). This function describes the correlation of two halves of a dataset as a function of the spatial frequency [207]. Excellent introductions into the principle of single-particle cryo-EM are provided in [45, 41, 142].

As no crystals are needed, cryo-EM requires less sample and allows higher heterogeneity of the sample, since aggregated or distorted particles can be sorted out during data processing. This is beneficial, especially for membrane proteins and macromolecular complexes like the T7SS. Purification yields are likely to be low and detergents or lipid micelles used to extract the protein from the native lipid environment could impair crystal contacts and often add additional heterogeneity to the sample.

Thus, it is expected that the chances to achieve high-resolution structural information on the ESX-5 T7SS is highest when applying cryo-EM.

1.4.3 Membrane protein reconstitution

As mentioned before, for single particle cryo-EM, the T7SS needs to be extracted from the lipid membrane. This step is critical, as replacing the natural environment by detergent could lead to a loss of stability, denaturation and/ or aggregation. Most often the extraction

is achieved by the use of amphiphilic molecules, like detergents [173]. Above their critical micelle concentration (CMC) these molecules self-assemble into small particles with a hydrophobic core and a polar shell called micelles [254, 94]. When added to lipid membranes, the membrane disassembles and micelle belts form around the membrane proteins, shielding their hydrophobic core from the aqueous buffer. The choice of detergent, therefore plays an important role [254, 94].

After extraction of the membrane protein, a detergent concentration above the CMC needs to be maintained in all following steps. The presence of detergents in cryo-EM, however, hampers structural analysis as it reduces contrast and introduces further heterogeneity to the sample [173, 14].

To overcome this problem, several further membrane mimetics were developed. These can be divided into two classes: The first class allows the incorporation of the membrane protein in a lipid environment, while in the second class, the hydrophobic core of the membrane protein is shielded by synthetic polymers or peptides.

In the first class, most common approaches include membrane scaffold-protein (MSP)-based or synthetic nanodiscs [19, 254]. MSP is an optimized human serum apolipoprotein and is designed to wrap around a phospholipid bilayer, generating a disc shaped proteo-lipid particle. The reconstitution of the membrane into the nanodisc requires the extraction of the protein from the lipid environment with detergents, but once reconstituted, detergents can be omitted [19, 254, 14, 199]. Styrene maleic anhydride (SMA) is used to obtain synthetic nanodiscs. In contrast to MSP-based nanodiscs, the synthetic, amphiphilic polymer can extract membrane proteins directly from native membranes without the need for detergent [136, 199, 254, 14]. This allows the characterization of the membrane protein in the lipid environment of the expression host. However, for the reconstitution of the T7SS both methods have a key disadvantage: the diameter of the nanodisc is highly dependent on the length of the polymer or MSP protein chain. Commercially available nanodiscs are produced in different sizes, but so far, the diameters are limited to around 17 nm for MSP or synthetic polymers [108, 227].

As the diameter of the hexameric T7SS is estimated to be 28 nm, by NS-EM, neither SMA nor MSP based-nanodiscs can be applied for structural studies on ESX-5. An alternative approach makes use of a small lipid binding protein of the saposin family, that is used as a scaffolding protein. In this approach, multiple Saposin A (SapA) molecules assemble to form a saposin lipid nanoparticle (SapNP) [90, 86]. Initially, the method was based on the reconstitution of the membrane protein after detergent solubilization. However, it was reported, that SapA molecules are also able to extract proteins directly from the membrane [149]. Due to the modular principle of SapA, there is, in theory, no limit in diameter [86]. However, so far it has not been demonstrated that SapNPs can be applied for a successful reconstitution

of proteins of sizes comparable to the T7SS.

The most widely used approach in the second class are amphipols. These synthetic amphiphatic polymers directly bind to the hydrophobic core of the membrane protein [232, 14, 199]. Amphipols are added to the protein after detergent solubilisation. They are very versatile and were shown to stabilize a number of different proteins, also the T7SS [20, 14]. However, some conformations might be only obtained in a lipid environment, so both approaches are useful.

Given the successful reconstitution of ESX-5 into amphipols shown before, amphipols were applied as main membrane mimetic, but to investigate the impact of a native lipid environment on the T7SS the reconstitution into SapNPs was tested as well.

1.4.4 Protein prediction by Alphafold2

Within the four years of this PhD, a scientific breakthrough was achieved by Jumper *et al.* from Deepmind [131]. They developed a machine learning approach able to accurately predict protein coordinates with (near-) atomic precision, based on a single protein sequence provided. The method is based on contact predictions from a multiple sequence alignment (MSA) and has triggered high excitement in the structural biology field - which can be expressed by more than 4000 citations within less than 1.5 years. The quality of the prediction of a protein sequence by Alphafold2 can be judged by two measures [131]:

- 1) The predicted local-distance difference test (pLDDT) that gives an estimate on the confidence of the prediction per-residue on a scale from 0-100. While a structure with a score of above 70 can be considered as relatively well modeled, a score above 90 corresponds to high accuracy of the prediction [131].
- 2) The predicted alignment error (PAE) which provides information on the confidence of the relative positioning of individual domains. The expected position error is given in Å. The PAE as quality matrix becomes especially important in the context of multi-domain proteins or multimeric proteins that AlphafoldMultimer was trained on [131, 79].

While Alphafold2 predicts individual domains of the ESX inner membrane components with a relatively high accuracy (pLDDT > 70), due to the size, the number of components and a relative low structural conservation resulting in weak MSAs, a reliable prediction of a full ESX secretion complex can not be standardly obtained by Alphafold2 [47]. However, the method is especially useful to explore unknown, but conserved contacts and was applied in this thesis to predict substrate complexes.

2 Materials and Methods

This chapter includes protocols that were applied to obtain results presented in Chapter 3. As described in Scientific Contributions, the results part is complemented with contributions from collaborators. Method parts to which I was not contributing are cited here for completeness and are highlighted in italics.

2.1 Common buffers

Table 2.1: Common buffers Common solutions and buffers used in the course of this thesis. For the preparation of the ESX-5 ATP γ S sample, buffers marked with asterisk were supplemented with 100 μ M ATP γ S and 2 mM MgCl₂.

Buffer	Composition
7H9 complete medium	7H9 medium, 13.9 mM NaCl, 0.5% BSA, 0.2% glucose, 0.2 % glycerol, 0.05% Tween 80
7H9 expression medium	7H9 medium, 0.2% glucose, 0.2 % glycerol, 0.05% Tween 80
Saposin lysis buffer	20 mM HEPES pH 7.5, 300 mM NaCl, 15 mM imidazole, 0.01 mg/ml DNase, 1 mg/ml Lysozyme, 1:100 protease inhibitor
Saposin wash 1 buffer	20 mM HEPES pH 7.5, 300 mM NaCl, 15 mM imidazole
Saposin wash 2 buffer	20 mM HEPES pH 7.5, 300 mM NaCl, 30 mM imidazole
Saposin elution buffer	20 mM HEPES pH 7.5, 150 mM NaCl, 400 mM imidazole
Saposin dialysis/ SEC buffer	PBS
ESX lysis buffer	20 mM Tris pH 8, 300 mM NaCl, 10 % glycerol, 0.01 mg/ml DNase, 1:100 protease inhibitor
ESX Solubilization/ Wash buffer*	20 mM Tris pH 8, 300 mM NaCl, 10 % glycerol
ESX elution buffer*	20 mM Tris pH 8, 300 mM NaCl, 10 % glycerol 10 mM desthiobiotin, 0.003% DDM
ESX SEC/ Gradient buffer*	20 mM Tris pH 8, 150 mM NaCl
ESX Gradient 10% buffer	20 mM Tris pH 8, 150 mM NaCl, 10% glycerol
ESX Gradient 40% buffer	20 mM Tris pH 8, 150 mM NaCl, 40% glycerol
PBS-T	PBS, 0.05% Tween 20

2.2 Molecular biology

2.2.1 Preparation of electrocompetent *Mycobacterium smegmatis* cells

To prepare electrocompetent Msmeg mc²155 groEL1 Δ C, cells were scraped on an agar plate and incubated for 3 days at 37 °C. One colony was picked to inoculate 5 ml 7H9 complete medium in a 50 ml Falcon tube. After 3 days incubation in a rotating incubator, 300 μ l of the saturated preculture were added to 50 ml complete medium in a 300 ml flask and grown overnight to an optical density at 600 nm (OD₆₀₀) of 0.5 to 1.0. The culture medium was then filled into a 50 ml tube and spun for 10 minutes at 4,000 rpm at 4 °C. While the supernatant was discarded, the pellet was resuspended in 50 ml ice-cold 10% glycerol and spun for another 10 minutes at 4,000 rpm at 4 °C. This step was repeated twice. The final pellet was resuspended in 2 ml ice-cold 10% glycerol and stored in 100 μ l aliquots at -80 °C.

2.2.2 *Mycobacterium smegmatis* electroporation

One 100 μ l aliquot of electrocompetent Msmeg was thawed on ice to transform one expression construct. Subsequently the cells were transferred to pre-chilled electroporation cuvettes (Biozym) and 1 μ l of the expression plasmid was added. The suspension was mixed thoroughly and incubated on ice for 30 minutes before an electroporation was carried out three times using the Electroporator 1000 (Stratagene) at 2.7 kV. After electroporation the cuvette was placed on ice immediately and 1 ml pre-chilled 7H9 medium was added. The suspension was transferred to a 2 ml centrifuge tube and shaken at 37 °C for 2 hours. The cells were spun down in a microcentrifuge at 6,000 rpm and the pellet was plated on an agar plate with appropriate antibiotics. The plate was placed in an incubator at 37 °C. Colonies could be observed after 3-4 days.

2.2.3 Expression of ESX-5 constructs in *Mycobacterium smegmatis*

To set up precultures, 10 ml 7H9 expression medium containing appropriate antibiotics were inoculated with a single colony in a 50 ml Falcon tube and shaken for 3 days in an incubator at 37 °C. When saturated, 3-4 ml were used to inoculate 500 ml of 7H9 expression medium with appropriate antibiotics. One ml of the remaining cell suspension was used to prepare a glycerol stock by mixing it with 1 ml of 50% glycerol and subsequent storage at -80 °C. Cells were grown in a shaker incubator at 37 °C and 120 rpm for approximately one day or when cells reached an OD₆₀₀ >1.5. In case of inducible vectors, like pMyNT or pMyC [22], protein expression was induced by the addition of 2 mg/ml acetamide (Sigma) to the culture. Cells were incubated for another 24 hours before they were pelleted for 40 minutes at 7,000 rpm and 20 °C in a floor centrifuge. In case of non-inducible constructs, cells were pelleted with the same settings after an OD₆₀₀ of >1.5 was reached. The pellets were stored at -20 °C.

2.2.4 Expression of Saposin A in *Escheria coli*

Precultures of SapA transformed into *E.coli* Rosetta-gami-2 (DE3) cells, were set up using a glycerol stock for inoculation, kindly provided by the group of Christian Löw, EMBL Hamburg. The cells were set up in 50 ml LB medium supplemented with kanamycin, tetracyclin and chloramphenicol (all Roth) and shaken overnight with 120 rpm at 37 °C.

Four ml of the saturated preculture were then used to inoculate 500 ml of TB medium and the cells were grown in a shaker incubator set to 142 rpm and 37 °C to an OD600 of 1. Once the cells reached the aimed OD, the temperature was lowered to 18 °C and the cells were induced with 1 mM Isopropyl β -D-1-thiogalactopyranoside (IPTG) (Roth). The cells were further shaken for overnight expression and spun down the next day at 7,000 rpm for 20 minutes. The pellets were stored for further use at -20 °C.

2.2.5 Purification of Saposin A

Per 1 g of cells 5 ml of Saposin lysis buffer were added and incubated under stirring for at least 40 min at 4°C. The cells were lysed running the suspension twice through an emulsifier (Avestin Emulsiflex) set to a pressure of 15,000 psi. The lysate was collected on ice and subsequently spun down in a preparative ultracentrifuge (Beckmann Optima) at 30,000 rpm to remove cell debris. The pellet was discarded and the supernatant was added to IMAC beads (Machery-Nagel), preequilibrated in Saposin lysis buffer. The IMAC beads were further incubated on a rotating wheel for 1 hour before transferred to a gravity flow column (Bio-Rad). Before opening the column, beads were allowed to settle down. The flowthrough was collected and applied a second time. The column was first washed with Saposin wash 1 buffer and with Saposin wash 2 buffer afterwards. The protein was then eluted with two column volumes of Saposin elution buffer. Flowthrough, wash and elution fractions were analyzed via SDS PAGE to assess protein purity. Eluted fractions were pooled and transferred to a dialysis bag with a molecular weight cut-off (MWCO) of 3.5 kDa (Roth). To cleave the His-tag, TEV-protease was added to a final concentration of 0.1 mg/ml. The dialysis bag was placed in Saposin dialysis buffer and stirred overnight at 4°C. On the dialysed sample a negative IMAC was performed, adding it to IMAC beads and collecting the flow through. This step ensures removal of the TEV protease as well as uncleaved protein material. The Saposin sample was concentrated to 1 ml in a 5 kDa MWCO centrifugal filter (Merck) and applied with a flow rate of 1 ml/min to an S75 16 60 column equilibrated in Saposin buffer. Peak fractions were collected, concentrated to a concentration of 1.5-3 mg/ml and stored at -80 °C until further use.

2.2.6 Purification of ESX-5 constructs

Cell lysis and membrane preparation

For cell lysis, the cell pellet was weighed and 20 ml of ESX lysis buffer were added to 5 g of cells. The cells were stirred at 4 °C for at least 1 hour. Once the cells were dissolved, they were applied to an emulsifier at an operating pressure of 20,000 psi. To ensure sufficient cell lysis, this step was repeated three times. The homogenized cell suspension was collected in falcon tubes on ice and subsequently spun down for 12 minutes at 8000 x g at 4 °C. The pelleted unlysed cells were discarded and the supernatant was transferred to ultracentrifuge tubes to pellet the membranes at 35,000 rpm for one hour at 4 °C in an ultracentrifuge. The supernatant was discarded and the pelleted membranes were removed from the ultracentrifuge tube with a plastic spoon and transferred to a potter homogenizer. Ten ml to 20 ml ESX Solubilization buffer were added to resuspend the membranes. The suspension was aliquoted, flash frozen in liquid nitrogen and stored at -80 °C.

Membrane solubilization and strep-tag affinity purification

To proceed with the purification membrane, aliquots were thawed on ice and resuspended in ESX Solubilization buffer to a final concentration of 5-10 mg/ml. 0.25% n-dodecyl- β -maltoside (DDM) powder (Anatrace) was added for solubilisation of the membranes and the suspension was stirred for 1-2 hours before ultracentrifugation at 30,000 rpm for 30 minutes at 4°C to remove insoluble material. The supernatant was applied to either loose Streptactin beads (iba) or a 5 ml StrepTrap column (cytiva), that was equilibrated with ESX Wash buffer beforehand. In case of loose beads, the beads were incubated for 10-20 min on a rotating wheel and transferred to a gravity flow column (Bio-Rad). The flowthrough, of the gravity flow column was collected and applied a second time. After sample application, in both cases the column was washed with 10 times bed volume in ESX Wash buffer. Afterwards, the sample was eluted with two column volumes ESX Elution buffer. To assess quality and purity, flowthrough, wash and elution fractions were analyzed via sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) and the concentration was estimated using a Nanodrop (Thermo Scientific).

Amphipol exchange

Eluted fractions were exchanged into Amphipol A8-35 (Anatrace) by adding amphipols to the sample in a 1:5 (w/w) ratio. The mixture was incubated on a rotating wheel for 20-30 minutes, before biobeads were added to remove DDM. The amount of biobeads (Bio-Rad) was calculated based on the weight of residual detergent present in the solution, which was estimated to be 0.03%. The final amount of biobeads added accounted for 20 times of the total weight of DDM. The solution was further incubated overnight on a rotating wheel at 4 °C.

SapA exchange

Mxen ESX-5 was reconstituted into SapA and lipids to mimic an environment closer to the mycobacterial inner membrane. For that purpose, fractions eluting from the StrepTrap column or StrepTactin beads were incubated with SapA and brain lipids (Avanti Polar Lipids). Different protein:SapA:lipid molar ratios were tested. After 30 minutes incubation, biobeads were added as described for amphipol exchange, followed by an overnight incubation at 4 °C on a rotating wheel.

2.2.7 Direct extraction of ESX-5 by SapA

Direct extraction of ESX-5 by SapA was used to preserve the natural lipid environment of ESX-5 inside the Msmeg cell envelope and to circumvent the addition of a suboptimal saposin lipid ratio. Instead of DDM, SapA was added to a final concentration of 8 µM to the suspended membranes and the solution was stirred overnight. In the following, the standard affinity purification protocol for ESX-5 was followed. The use of detergent was, however, omitted.

Size exclusion chromatography

After the exchange into amphipols or SapA, Biobeads were removed with a filter and the sample was concentrated to 150 µl in a 100 kDa MWCO concentrator (Merck). The sample was filtered again before applying it on a Superose 6 10/300 column (GE/ cytiva) using an Äkta Pure or an Äkta Explorer at a flow rate of 0.15 ml/min and ESX SEC buffer as running buffer. Void peak fractions, that typically contained hexameric ESX-5 species were collected and analyzed via SDS-PAGE and BN-PAGE.

For analysis of small sample volumes <100 µl or low concentrated samples, instead of the Äkta system, the Agilent HPLC system was used in combination with a Superose 6 3.2/300 column (GE/ cytiva) column or with a Superose 6 5/150 column.

Gradient centrifugation

Gradient centrifugation was used interchangeably with size exclusion chromatography to achieve highest possible sample purity and homogeneity. Biobeads were removed as described before and the sample was applied to a 3.5 kDa MWCO dialysis bag (Roth) and dialysed for at least 2 hours in ESX gradient buffer. The sample was concentrated to 100 µl in a 100 kDa MWCO concentrator. ESX Gradient 10% and ESX Gradient 40% buffer were prepared and mixed with a GradientMaster (Biocomp) to form a linear 10%-40% gradient. The sample was applied on top of the gradient and the centrifuge was started at 25,000 rpm for 17.5 hours. Afterwards the centrifuge tube was poked with a sharp needle and the sample was fractionated, collecting 3 drops per fraction in microcentrifuge tubes. The sample was analyzed by SDS-PAGE and BN-PAGE.

2.3 Protein analysis

2.3.1 Sodium dodecyl sulfate–polyacrylamide gel electrophoresis

To control the protein purity SDS-PAGE analysis was performed, using 4–12% precast NuPAGE Bis-Tris Gels (Invitrogen). Sample buffer was prepared by adding 100 μ L water and 20 μ L 0.5 M TCEP (Sigma) to 20 μ L of NuPAGE LDS sample buffer (Invitrogen). The protein sample and the buffer were mixed in a 1:2 (v/v) ratio and applied to the gel. To determine the migration height, a molecular weight marker was loaded. Gels were run for 45 minutes at a constant voltage of 180 V in MES SDS Running buffer. After the run, gels were stained with InstantBlue (Expedeon) or Quick Coomassie Stain (Serva).

2.3.2 Blue native-polyacrylamide gel electrophoresis

BN-PAGE was carried out to assess the size of protein complexes. Either the NativePAGE buffer system from Invitrogen or the NativeGel buffer system from SERVA were used with the corresponding 3%-12% precasted gels, according to the protocol. Before sample loading the wells were washed with cathode buffer to remove residual storage buffer. As molecular weight standard, the ThermoFisher native mark was loaded. Gels were run at 100 V for the first 30 minutes, before the voltage was increased to 200 V to keep a minimal current of 2 mA for approximately 3 hours. The gel was fixed and destained in 10% acetic acid and 40% ethanol for one hour. Afterwards the gel was washed with water and stained again with InstantBlue (Expedeon) or Quick Coomassie Stain (Serva).

2.3.3 Western blot

For Western blot an SDS-PAGE gel was run as described above. Instead of staining, the gel was transferred on a nitrocellulose membrane using the Trans-blot turbo setup (BioRad). To minimize unspecific binding, the membrane was incubated in PBS-T supplemented with 5% milk with agitation at room temperature for 1 hour. After the blocking procedure, the membrane was washed twice in PBS-T. For analysis of strep-tagged and His-tagged proteins primary antibodies with conjugated horseradish peroxidase (HRP) were used. The antibodies were diluted (1:5000 for anti-strep-tag HRP (Iba) and 1:5000 for anti His-tag HRP (ThermoFisher)) in PBS-T and incubated with the membrane for one hour. Remaining antibodies were removed by three 15 minute washing steps with PBS-T. The PBS-T wash buffer was discarded and excess of wash buffer was removed by pipeting. The SuperSignal West Pico Chemiluminescent Substrate kit (Thermo Fisher) was used for detection of the antibodies. The working solution was prepared by mixing equal parts of the Stable Peroxide Solution and the Luminol/Enhancer Solution. Two ml of this working solution was applied to the membrane and incubated 5 minutes, before imaging it in the gel imaging system from Bio-Rad.

2.3.4 Limited proteolysis

A stock of 0.1 mg/ml purified Mxen ESX-5 complex in SEC buffer was used to add three different proteases (Elastase, Trypsin and Proteinase K (all Sigma)) to a final concentration of 100 ng/ml and 10 ng/ml. Samples were incubated 30 min, 60 min, 120 min, 240 min and overnight respectively, before they were flash frozen. For analysis, LDS sample buffer was added and the samples were loaded on an SDS-PAGE gel for analysis as described before.

2.3.5 Surface plasmon resonance analysis

To assess binding of ATP γ S to the three ATPase domains of EccC₅ a surface plasmon resonance (SPR) analysis was carried out using a T200 instrument (Biacore/ Cytiva). For preparation, 1 μ M EccC₅ V429 construct, provided by Katherine Beckham, was immobilized on a CM5 chip (GE/cytiva) after activating the surface with 0.4 M 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) and 0.1 M N-hydroxysuccinimide (NHS) (GE/cytiva). After immobilization the surface was deactivated using 1 M ethanolamine (GE/cytiva). PBS was used as running buffer. For binding analysis, ATP γ S (Sigma) was injected to the flow cell in a concentration range of 0 μ M to 50 μ M for 90 s, before the dissociation was recorded for 240 s, respectively. The response was calculated by subtracting the response of an empty flow cell from the signal of the flow cell with immobilized EccC₅. For preparing a dose-response-curve endpoints of the association curve were plotted against the concentration using the BIAevaluation software (GE/cytiva).

2.3.6 Negative stain EM

For NS-EM 400 mesh carbon coated grids (microtonano) were glow discharged using a GloQube (Quorum) at 25 mA for 30 s in negative mode. 4 μ l of sample at a concentration of 0.01 mg/ml to 0.3 mg/ml were applied directly on a carbon-coated copper grid and incubated for 1 minute before the liquid was removed with blotting paper. Afterwards, the grid was washed twice with water and once with 2% uranyl acetate stain, by dipping the grid into 20 μ l drops and subsequent blotting, respectively. Finally, the grid was plunged into uranyl acetate and incubated 30 seconds before blotting. Negative stained EM samples described in this thesis were analyzed at the XFEL XBI facility at a JEOL 200 kV cryo-TEM or at the CSSB cryo-EM facility using a Talos L120C (Thermo Fisher) microscope with a 120 kV accelerating voltage.

2.3.7 Crosslinking mass spectrometry

A protocol for crosslinking MS can be found in Beckham *et al.* [21]. For completeness the procedure, outlined in this paper is cited in the following paragraphs:

"Fifty micrograms of purified ESX-5 complex was cross-linked by addition of an iso-stoichiometric mixture of H12/D12 isotope-coded, di-succinimidyl-suberate (DSS) (Creative Molecules). The cross-linking reaction (final concentration of 1 mM) was incubated for 30 min at 37°C and quenched by addition of ammonium bicarbonate to a final concentration of 50 mM for 10 min at 37°C. Cross-linked proteins were denatured using urea and RapiGest (Waters) [final concentration of 4 M and 0.05% (w/v)], respectively. Samples were reduced using 10 mM dithiothreitol (30 min at 37°C), and cysteines were carbamidomethylated with 15 mM iodoacetamide (30 min in the dark). Protein digestion was performed using 1:100 (w/w) LysC (Wako Chemicals) for 4 hours at 37°C and then finalized with 1:50 (w/w) trypsin (Promega) overnight at 37°C after the urea concentration was diluted to 1.5 M. Samples were then acidified with 10% (v/v) trifluoroacetic acid and desalted using OASIS HLB μ Elution Plate (Waters). Cross-linked peptides were enriched using size exclusion chromatography (SEC) [147]. Collected SEC fractions were analyzed by LC-coupled tandem MS (MS/MS) using a nanoACQUITY Ultra Performance LC (UPLC) system (Waters) connected online to linear ion trap quadrupole (LTQ)–Orbitrap Velos Pro instrument (Thermo Fisher Scientific). Peptides were separated on a BEH300 C18 (75 mm by 250 mm by 1.7 mm) nanoACQUITY UPLC column (Waters) using a stepwise 60-min gradient between 3 and 85% (v/v) acetonitrile in 0.1% (v/v) fusaric acid. Data acquisition was performed using a top-20 strategy, where survey MS scans (mass/charge ratio range of 375 to 1600) were acquired in the Orbitrap ($R = 30,000$), and up to 20 of the most abundant ions per full scan were fragmented by collision-induced dissociation (normalized collision energy = 40, activation $Q = 0.250$) and analyzed in the LTQ. To focus the acquisition on larger cross-linked peptides, charge state 1, charge state 2, and unknown were rejected. Dynamic exclusion was enabled with a repeat count = 1, exclusion duration = 60 s, list size = 500, and mass window of ± 15 parts per million (ppm). Ion target values were 1,000,000 (or 500-ms maximum fill time) for full scans and 10,000 (or 50-ms maximum fill time) for MS/MS scans. The sample was analyzed in technical duplicates. To assign the fragment ion spectra, raw files were converted to centroid mzXML format using a raw converter and then searched using xQuest [148] against a FASTA database containing the sequences of the cross-linked proteins. Posterior probabilities were calculated using xProphet [148], and results were filtered using the following parameters: false discovery rate = 0.05, minimum Δ score = 0.95, MS1 tolerance window of -4 to +7 ppm, and identity (Id) score > 36."

2.4 Cryo-EM

2.4.1 Grid preparation

Mycobacterium xenopi ESX-5 and Mycobacterium xenopi ESX-5 ATP γ S

For single particle cryo-EM on Mxen ESX-5 WT in presence and absence of ATP γ S, 3.6 μ l of the ESX-5 void peak fraction was applied on freshly glow-discharged Quantifoil R2/1

Cu 200 mesh grids with 2-nm continuous carbon. Sample vitrification was performed using a Vitrobot Mark IV at 100 % humidity and 10° C. The liquid was blotted for 2 s and the sample was plunge-frozen in a liquid propane/ ethane mix.

Mycobacterium xenopi ESX-5 P73A

For cryo-EM analysis of the Mxen ESX-5 P73A mutant, 4 µl of fraction 9 dialysed into SEC buffer was applied on a freshly glow-discharged Quantifoil R1.2/1.3 Cu 200 mesh grid. Sample vitrification was performed using a Leica GP2 plunge freezer at 99 % humidity and 10° C. The liquid was blotted away for 2 s and the sample was plunge-frozen in a liquid propane/ ethane mix.

2.4.2 Data collection

Initial cryo-EM dataset ESX-5 WT

An initial cryo-EM dataset of the ESX-5 WT, including 6,067 movies was collected at the Center for Structural Systems Biology (CSSB), Hamburg, using a Titan Krios at an acceleration voltage of 300kV with a K3 direct electron detector (Gatan) and a Bio-Quantum K3 energy filter (Gatan). The data acquisition was performed using EPU (Thermo Fisher Scientific). Movies were recorded with a pixel size of 0.67 Å, 50 frames and a total dose of 50.00 e⁻/Å² in electron counting mode.

Cryo-EM dataset ESX-5 WT

The final cryo-EM dataset of the ESX-5 WT was collected at the Cryo-Electron Microscopy Service Platform, EMBL Heidelberg, on a Titan Krios operated at 300 kV (Thermo Fisher) equipped with a K3 direct detection camera (Gatan) and a Bio-Quantum K3 energy filter (Gatan) operated by SerialEM (Mastronarde). In total 27.873 movies with 40 frames were recorded in electron counting mode. The total dose accounted for 49.34 e⁻/Å² and the pixel size was set to 0.645 Å.

Cryo-EM dataset Mycobacterium xenopi ESX-5 P73A

Data on the ESX-5 P73A mutant was collected at the CSSB electron microscopy facility on a Titan Krios equipped, as described for the initial ESX-5 dataset. Data acquisition was performed using EPU. 10,233 movies were collected with 40 frames and a total dose of 40.92 e⁻/Å² in electron counting mode. The pixel size of the movies was set to 0.85 Å.

Cryo-EM dataset Mycobacterium xenopi ESX-5 ATP γ S

The dataset of the ESX-5 complex in presence of ATP γ S was collected at the CSSB electron microscopy facility in Hamburg using the same microscope setup as described for the initial

dataset. With the help of EPU 12,789 movies were collected with a pixel size of 0.68 Å and a total dose of 48.80 e⁻/Å², using 50 frames in electron counting mode.

2.4.3 Processing

Cryo-EM processing was performed in cryoSPARC [194] and RELION [211]. As the maps included in this thesis were obtained using only the cryoSPARC workflow, the steps in RELION are not further described. Also further processing strategies in cryoSPARC aimed to achieve higher resolution globally or of specific regions were omitted, as long as they did not improve the map quality.

Initial cryo-EM dataset ESX-5 WT

As a first step, beam-induced motion was corrected using cryoSPARC's own implementation of patch-motion correction. Afterwards the landscape of the contrast transfer function (CTF) was estimated by patch CTF estimation. Templates generated from an initial dataset, collected at a Talos Arctica (Thermo Fisher) at the CSSB cryo-EM facility, were used to pick 249,802 particles that were subjected to several rounds of 2D-classification. After each round, particle classes, corresponding to distorted particles were excluded. Finally, the best 69,564 particles were used to build an *ab-initio* model. As the *ab-initio* model revealed a clear hexamer, the data was refined using the nonuniform refinement algorithm [195] without imposed symmetry. As the C1 symmetric map showed clear C6 symmetry and the resulting resolution of 6.3 Å was too low for model building, the non-uniform refinement [195] was performed also in C6 resulting in a 4.4 Å map.

Final cryo-EM dataset ESX-5 WT

The cryo-EM processing was similarly described in Beckham *et al.* [21] but was rephrased in the following paragraph:

Motion correction and CTF estimation were performed as described for the initial ESX-5 dataset. Afterwards the exposures were curated using CTF-fit parameters and local-motion distances as selection criteria. Templates were generated, from the initial dataset and the remaining 18,598 micrographs were used for template-based particle picking. After particle inspection, 635,219 particles were extracted and binned four times. The resulting particles were aligned and classified in four rounds of 2D classification. Visual inspection of the 2D classes after each round allowed the identification of distorted particles and artifacts. These particles were excluded for further processing, ensuring that only the best particles were included in the *ab-initio* reconstruction using three classes. One of the resulting classes, showed a hexameric map, while the two others corresponded to dimeric or monomeric reconstructions. Further processing of the monomeric and dimeric classes using non-uniform refinement

[195], however, did not reveal the presence of secondary structure elements. Thus, processing of these classes was not followed up. The 123,759 particles of the hexameric class were further sorted using an additional round of 2D classifications. After curating the 2D classes, the remaining 121,974 particles were reextracted and binned two times. Following another *ab-initio* model, non-uniform refinement [195] was carried out without imposed symmetry. This yielded a map with a resolution of 3.9 Å. Since initial model building showed a clear indication of C6 symmetry, the data was reprocessed in C6 symmetry, resulting in a reconstruction at 3.4 Å, in which the transmembrane part was resolved to up to 2.8 Å resolution and the cytosolic part was less well resolved with 4.0 Å to 5.5 Å resolution. A resolution of 3.0 Å was achieved after C6 symmetry expansion of the particle stack and a subsequent local refinement with a mask around the cytoplasmic region of only one protomeric unit. For resolving the periplasmic domains, the 121,974 particles were subjected to heterogenous refinement that classified the particles into three different classes. One class, comprising 94,163 particles showed more density on the periplasmic side of the complex. A 3D variability analysis [193] was run on these particles, that separated the particles into three clusters. One cluster revealed elongated, C2-symmetric density of the periplasmic part and was refined by applying the non-uniform refinement [195] algorithm in C1 and C2 symmetry. The resulting models were used as input model for the refinement of the 121,974 particles. For improved separation of the elongated density in the periplasm, another round of heterogenous refinement was carried out after which one class of 52,015 particles was further refined using non-uniform refinement [195]. This yielded a reconstruction of 4.6 Å resolution. The resolution of the periplasmic domain, however, accounted for less than 10 Å.

Cryo-EM dataset ESX-5 P73A

Motion correction and CTF estimation were performed as described for the initial ESX-5 cryo-EM dataset. Using CTF-fit parameters and local-motion distances as quality criteria, only 9,275 of the 10,233 micrographs were selected for further processing. Initial templates for template based particle picking were generated, applying the blob picker and 2D classes generated from the extracted particles picked. After inspection of the particles picked by template based particle picking, 985,743 particles were extracted from the micrographs, binned two times and several rounds of 2D classifications were performed. As described for other datasets, the 2D classes were curated after each round of classification, ensuring that only classes corresponding to intact protein were included for further processing. With the remaining 175,048 particles an *ab-initio* reconstruction was performed, requesting two classes. Only one of these classes yielded hexameric features and to clean up the particle stack further, another 2D classification was performed. The best 93,495 particles were then used for another *ab-initio* reconstruction, resulting in an even clearer hexameric reconstruction. Particles were re-extracted without binning and further rounds of 2D classifications were performed, leading to a final set of 78,699 particles that were used to generate a new *ab-initio* model. This model was used for non-uniform refinement [195] applying C1 and C6

symmetry, resulting in maps of 3.9 Å and 5.3 Å resolution, respectively.

Cryo-EM dataset ESX-5 WT in presence of ATP γ S

Processing was performed similarly to the ESX-5 P73A dataset. After curation of the micrographs 8,610 micrographs were selected for further processing. The template based particle picking resulted in 320,185 particles that were further classified using several rounds of 2D classifications. The 81,739 particles remaining were used to reconstruct three *ab-initio* models, of which only one showed hexameric features. This class was refined using non-uniform refinement [195]. While the C6 symmetric map reached a resolution of 3.9 Å, the C1 map yielded a resolution of 5.3 Å. Low resolution density of the ATPase domains in upward and downward conformation was observed in both maps.

2.5 Model building and refinement

Initial model building was performed in Coot [78] using homology models derived from Swiss-model [248]. The final atomic models of the ESX-5 WT structure shown in this thesis, however, were built *de novo*. The procedure was precisely documented in Beckham *et al.* [21]. It is cited here for completeness:

"As there are no reliable, high-resolution structures of any of the ESX-5 components or their homologs available in the protein data bank (PDB), we built de novo a model of the TM and nearby cytoplasmic regions of the complex (EccB₅ 18-73; EccC₅ 12-417; EccD₅-1 23-502; EccD₅-2 18-494; and EccE₅ 95-332). An initial model was traced into a masked, focused refinement map using ARP/wARP cryo-EM module with default parameters [48]. Next, domains for which we solved the high-resolution crystal structure (EccD₅, residues 17 to 107) were fitted into the focused refinement map as rigid bodies using a Jiggle Fit tool from Coot [78]. The resulting model was completed manually using Coot in regions with local resolution allowing for unambiguous de novo model tracing. The interpretation of poorly resolved map regions was aided by alternative blurring and sharpening of the map in Coot. We used an iterative approach, where each manual model building step was followed with sequence assignment using findMySequence program [49], which allowed for an identification and correction of tracing errors (insertions and deletions). Loops that were resolved in the density but difficult to trace manually were built using the RosettaES density-guided enumerative-sampling algorithm from the Rosetta suite [91]. The complete protomer model built into a focused refinement map was expanded to a complex using symmetry operations derived directly from the C6-symmetrized map using phenix.find_ncs_from_density [231] and completed manually in Coot. Apart from solving minor symmetry conflicts, we traced the model fragments that were resolved in the symmetrized map only. These included two TMHs of EccC₅ (TMH1 and TMH2; residues 37 to 94). First, TMH1 was built de novo in Coot into the better resolved density and assigned to the sequence using findMySequence

program [49], which allowed for an unambiguous determination of the helix direction and sequence register. Subsequently, the second TM helix (TMH2) was built using the RosettaES density-guided enumerative-sampling algorithm followed with refinement with C6 symmetry. We also added to the model the most distant to the central pore EccD₅-2 helix (TMH11) based on a model of the corresponding helix in EccD₅-1. Geometry of the models was improved in ChimeraX [101] using ISOLDE [58] tool. Last, models of the protomer and full complex were refined against corresponding maps using phenix.real_space_refine [7], with nonbonded restraints weight increased to 200. For the complex, additional restraints to the initial model coordinates and strict rotamer matching were used."

Atomic model building of the ESX-5 P73A mutant was performed based on the model obtained for ESX-5 WT. Single protein components were aligned in ChimeraX [101] and the model was refined using restraints in Coot [78]. The single domains were expanded into the C6-symmetric map by applying phenix.find_ncs_from_density [231]. Inner pore helices were built applying additional helical restraints on the model. The final refinement of the structure is still ongoing.

2.6 Integrative modeling of EccB₅ and EccC₅

Integrative modeling combined the data derived from cryo-EM, crosslinking MS and homology modeling. The detailed procedure is described in Beckham *et al.* [21] and is cited here for completeness:

"The models of the hexameric assembly of the periplasmic domains of EccB₅ (amino acids 74 to 490) and cytoplasmic ATPase domains of EccC₅ (amino acids 431 to 1390) were built using an integrative modeling protocol similar to what was previously used by us [66, 8, 140]. The modeling procedure described in more detail below is implemented as a custom software based on Integrative Modeling Platform [250] version 2.13 and Python Modeling Interface [209] further described in [197]. All additional code and input files necessary to reproduce the steps will be released on Zenodo repository upon publication. The TM region of the ESX-5 structure built *de novo* as above and a homology model of the monomeric EccB₅ periplasmic domain and EccC₅ ATPase domains were used as input for modeling. The homology models of EccB₅ and EccC₅ were built using Modeller [209] based on the crystal structure of EccB₁ of *Mtb* (PDB ID: 3X3M [256]) and EccC of *Thermomonospora curvata* (PDB ID: 4NH0 [205]) and using the sequence alignment obtained from the HHpred server (50). The non-symmetrized (C1) EM map and available cross-links were used as modeling restraints. Owing to the low resolution (<10 Å) of the periplasmic and cytoplasmic regions, the high-frequency noise in the EM map was removed using a Gaussian filter with a SD of 3 Å for EccB₅ and 5 Å for EccC₅. In addition, to limit the conformational space, the fitting was performed using only a segment of the EM map not yet occupied by the TM region of the ESX-5 structure.

The models were additionally restrained using high-confidence cross-links above an xQuest [148] *ld* score of 36. At this threshold, two and six cross-links could be mapped to the EccB₅ and EccC₅ sequences, respectively, and used for modeling. As the first step of the modeling, large libraries of alternative fits to the EM map of the monomeric EccB₅, and EccC₅ structures were generated using the FitMap tool of the UCSF Chimera [184]. The fitting was performed using 100,000 random initial placements, cross-correlation about the mean as the fitting score [Chimera's "cam" score [184] equivalent to Pearson correlation coefficient], and the requirement of at least 80% of the input structure being covered by the EM map envelope defined at a permissive density threshold. This resulted in 9268 unique alternative fits for EccB₅ and 5068 fits for EccC₅ after clustering. Second, the resulting alternative fits of the monomeric structures and the TM region of the ESX-5 structure were built *de novo* as above and used as input for the simultaneous fitting of six copies of EccB₅ using the EM map and cross-link restraints and likewise for six copies of EccC₅. The fitting was performed through simulated annealing Monte Carlo optimization that generates alternative configurations of the fits precalculated as above. The optimization was performed independently 4000 times with 12,000,000 Monte Carlo steps for each run for EccB₅ and 2500 times with 12,000,000 Monte Carlo steps for each run for EccC₅. The sampling exhaustiveness was assessed by ensuring that (i) the score converges in individual runs, (ii) no new better scoring models appear with extra runs, and (iii) the score distributions in two random samples of the models are statistically similar. The scoring function for the optimization was a sum of the EM fit restraint represented as the *P* values of the precalculated domain fits [calculated as described in [66, 8, 140]], cross-linking restraints, clash score, connectivity distance between neighboring domains, a term preventing overlap of the protein mass with the TM region, and a two- or sixfold symmetry restraint for EccB₅ and EccC₅, respectively. During the optimization, the structures were simultaneously represented at two resolutions—in C α -atom representation and a coarse-grained representation—in which each 10-residue stretch was converted into one bead. The 10-residue bead representation was used for all restraints to increase computational efficiency except for the domain connectivity and cross-link restraints, for which the C α -only representation was used for reasons of accuracy. Last, top-scoring models from the previous step were subjected to a refinement coupled to an analysis of exhaustiveness of conformational sampling and estimation of model precision using a procedure proposed by Viswanath et al. [241]. To this end, the models from the first modeling stage (simultaneous fitting based on the alternative fits) were split into two random subsets. The top 30 models from each subset were refined using a Monte Carlo simulated annealing optimization in which the structures were moved in the EM map with small rotational and translational increments. The scoring function consisted of cross-correlation to the EM map, domain connectivity restraint, clash score, a term preventing overlap of the protein mass with the TM region, and a two- or sixfold symmetry restraint as above. For EccC₅, the monomeric homology model was split into two rigid bodies at the boundary between the ATPase domains 1 and 2, and an elastic network restraint was applied to enable limited flexibility between these domains.

Each of the 30 models was refined with 200 independent runs with 260,000 steps. The top-scoring models from each of the two runs were selected, leading to two independent samples of refined models (about 1000 models in each sample). For EccC₅, the multiple fitting step converged to a single top-scoring model; thus, only a single model was selected for the refinement, and the samples were generated by splitting the resulting refined models. The scores of the two samples were compared to each other to ensure convergence. The highest sampling precision at which sampling was exhaustive was determined on the basis of the root mean square deviation (RMSD) comparisons between all models and clustering at incremental RMSD thresholds using the statistical tests provided by Viswanath et al. [241]. The two samples were then clustered at the resulting precision level, and for each cluster, the model precision, defined as the average RMSD distance to cluster centroid, was calculated. The top 10 scoring models from all refined models were taken as the final ensemble model of the ESX-5 with the EccB₅. All the top 10 models satisfied both EccB₅ cross-link restraints (with a distance threshold of 30 Å)."

2.7 Bioinformatics

2.7.1 Analysis of EccC₅ pore dimensions and EccB₅ angles

The analysis of the pore dimensions was performed for the ensemble models of the EccC₅ pore helices obtained by Rosetta [91] using the HOLE software suite [216]. For the angle analysis of EccB₅, the ensemble models, resulting from integrative modeling, were analyzed by applying cctbx scripts [110].

2.7.2 Sequence alignments and sequence logo preparation

Sequence alignments were performed with ClustalOmega and analysed with MView [214, 34]. Sequence logos were prepared using the WebLogo online application [60]. For the inspection of sequences, sequences of Mtb, Msmeg, Mmar, Mxen and Mabs were used.

2.7.3 Analysis and visualization of protein structures

Analysis of protein structures was mainly carried out in ChimeraX [101]. Especially volume data, sequence analysis and structure analysis tools were widely used. Proteins were coloured in the following colour code if not stated otherwise: EccC: # 66ccffff, EccB: orange, EccD-1: # 21ff80ff, EccD-2: # 95fdccff, EccE: # 8061FF, MycP: red

2.7.4 Molecular dynamics simulations

Molecular dynamics (MD) simulations were prepared using the membrane builder implemented in the CHARMM-GUI [128, 253]. As WT input model the transmembrane helices of one single EccC₅ molecule were chosen from the Mxen ESX-5 WT cryo-EM model. For the

P73A mutant input model, the WT input model was mutated at position 73 from a proline to an alanine in ChimeraX [101]. Each input models were oriented within the membrane, respectively. The length of the X and Y axis of the systems were set to 80 Å, whereas the Z dimension was calculated based on the water thickness on each side of the membrane to 22.5 Å. As phosphatidylethanolamine is one of the most abundant lipids in the mycobacterial inner membrane, a 1-palmitoyl-2-oleoylphosphatidylethanolamine (POPE) bilayer was built as model membrane. The final number of lipids in the upper leaflet accounted for 103 lipids and in the lower leaflet for 105 lipids. Further on, KCl was added in a concentration of 0.15 to the system. Energy minimization, equilibration and production were performed in GROMACS [5] using the CHARMM36 force field [24]. After energy minimization, that removes steric clashes and wrong geometries, five steps of equilibration were carried out, to relaxate the system introducing it to a temperature of 25°C and a pressure of 1 bar. Finally, a 100 ns production run was carried out for both systems using 50,000,000 2 fs timesteps. The RMSD and RMSF of the C α -backbone was analysed using MDAnalysis [159] in jupyter notebooks [135].

2.7.5 Structure prediction with AlphaFold2

Structures of multimeric substrates were predicted applying Alphafold 2 multimer [131, 79] without AMBER [118] relaxation.

2.8 Secretion assay

2.8.1 Sample preparation

For MS based secretion assays, constructs were transformed into Msmeg mc²155 groEL1 Δ C. For each construct four colonies were picked and precultures were set up as described before. 0.5 ml of the saturated precultures were used to inoculate 50 ml 7H9 medium. Cells were grown to an OD of 1, before they were spun down at 4,000 rpm for 30 minutes. Pellet and supernatant were separated, respectively. For the whole cell fraction, pellets were dissolved in PBS supplemented with 0.01 mg/ml DNase and incubated on a rotating wheel for 1 hour. Afterwards, the sample was lysed using a sonicator for 135 s with a Bandelin MS73 probe attached and 45% amplitude. To ensure sufficient cell lysis and to keep membrane proteins soluble Triton was added to a final concentration of 1% and SDS to a concentration of 0.1%. The suspension was incubated for one hour. Afterwards 20 μ l samples were taken for analysis by western blot, to check for expression of each construct and colony. Further on, 2 ml of the sample were transferred to a microcentrifuge tube and spun at 1,000 x g to remove residual cell debris. The concentration of the supernatant was adjusted to 0.4 mg/ml by a Bradford assay according to the manufacturers protocol (Roth). For the secretome fraction: The culture supernatant was filtered through a 0.2 μ m filter and the secreted proteins were precipitated adding TCA (Roth) to a concentration of 5%. Samples

were frozen afterwards to ensure complete precipitation. After thawing, the sample was spun down with 7,000 rpm at 4 °C. The supernatant was discarded and the pellet was resuspended in acetone (Roth) and transferred to a microcentrifuge tube. After a centrifugation run at 13,000 rpm in a benchtop centrifuge, the supernatant was again discarded and the sample was dried. The pellet was then resuspended in 1 M Tris and 0.1% SDS (Sigma). For MS analysis the sample was adjusted to a concentration of 0.4 mg/ml using a Bradford assay (Roth).

2.8.2 TMT labeling and LC-MS/MS

This step was performed by Mandy Rettel from the proteomics core facility at EMBL Heidelberg. Her protocol is included for completeness:

"Reduction of disulphide bridges in cysteine containing proteins was performed with dithiothreitol (56°C, 30 min, 10 mM in 50 mM HEPES, pH 8.5). Reduced cysteines were alkylated with 2-chloroacetamide (room temperature, in the dark, 30 min, 20 mM in 50 mM HEPES, pH 8.5). Samples were prepared using the SP3 protocol [122, 123] and trypsin (sequencing grade, Promega) was added in an enzyme to protein ratio 1:50 for overnight digestion at 37°C. Next day, peptide recovery in HEPES buffer by collecting supernatant on magnet and combining with second elution wash of beads with HEPES buffer. Peptides were labelled with TMT10plex [251] Isobaric Label Reagent (ThermoFisher) according the manufacturer's instructions. In short, 0.8mg reagent was dissolved in 42ul acetonitrile (100%) and 4 ul of stock was added and incubated for 1h room temperature. Followed by quenching the reaction with 5% hydroxylamine for 15min. RT. Samples were combined for the TMT10plex and for further sample clean up an OASIS® HLB µElution Plate (Waters) was used. Offline high pH reverse phase fractionation was carried out on an Agilent 1200 Infinity high-performance liquid chromatography system, equipped with a Gemini C18 column (3 µm, 110 Å, 100 x 1.0 mm, Phenomenex) [200].

"An UltiMate 3000 RSLC nano LC system (Dionex) fitted with a trapping cartridge (µ-Precolumn C18 PepMap 100, 5µm, 300 µm i.d. x 5 mm, 100 Å) and an analytical column (nanoEase™ M/Z HSS T3 column 75 µm x 250 mm C18, 1.8 µm, 100 Å, Waters). Trapping was carried out with a constant flow of 0.05% trifluoroacetic acid at 30 µL/min onto the trapping column for 6 minutes. Subsequently, peptides were eluted via the analytical column with a constant flow of solvent A (0.1% formic acid in water) at 0.3 µL/min with increasing percentage of solvent B (0.1% formic acid in acetonitrile). The outlet of the analytical column was coupled directly to an Orbitrap Fusion Lumos (Thermo) mass spectrometer using the Nanospray Flex™ ion source in positive ion mode. The peptides were introduced into the Orbitrap Fusion Lumos via a Pico-Tip Emitter 360 µm OD x 20 µm ID; 10 µm tip (CoAnn Technologies) and an applied spray voltage of 2.4 kV. The capillary temperature was set at 275°C. Full mass scan was acquired with mass range 375-1500 m/z in profile mode in the

orbitrap with resolution of 120000. The filling time was set at maximum of 50 ms with a limitation of 4×10^5 ions. Data dependent acquisition (DDA) was performed with the resolution of the Orbitrap set to 30000, with a fill time of 94 ms and a limitation of 1×10^5 ions. A normalized collision energy of 38 was applied. MS2 data was acquired in profile mode."

"IsobarQuant [89] and Mascot (v2.2.07) were used to process the acquired data, which was searched against an Uniprot Msmeg (UP000000757) proteome database containing expressed sequences, common contaminants and reversed sequences. The following modifications were included into the search parameters: Carbamidomethyl (C) and TMT10 (K) (fixed modification), Acetyl (Protein N-term), Oxidation (M) and TMT10 (N-term) (variable modifications). For the full scan (MS1) a mass error tolerance of 10 ppm and for MS/MS (MS2) spectra of 0.02 Da was set. Further parameters were set: Trypsin as protease with an allowance of maximum two missed cleavages: a minimum peptide length of seven amino acids; at least two unique peptides were required for a protein identification. The false discovery rate on peptide and protein level was set to 0.01."

2.8.3 Data analysis

Statistical analysis of the secretion assay data was performed using scripts in R and applying the *limma* package [202]. This package uses an empirical Bayes approach for statistical tests. Fold-changes were calculated on a \log_2 scale after normalizing protein abundance to ESX-5 WT values by calculating the difference of the mean between mutant and ESX-5 WT measurements.

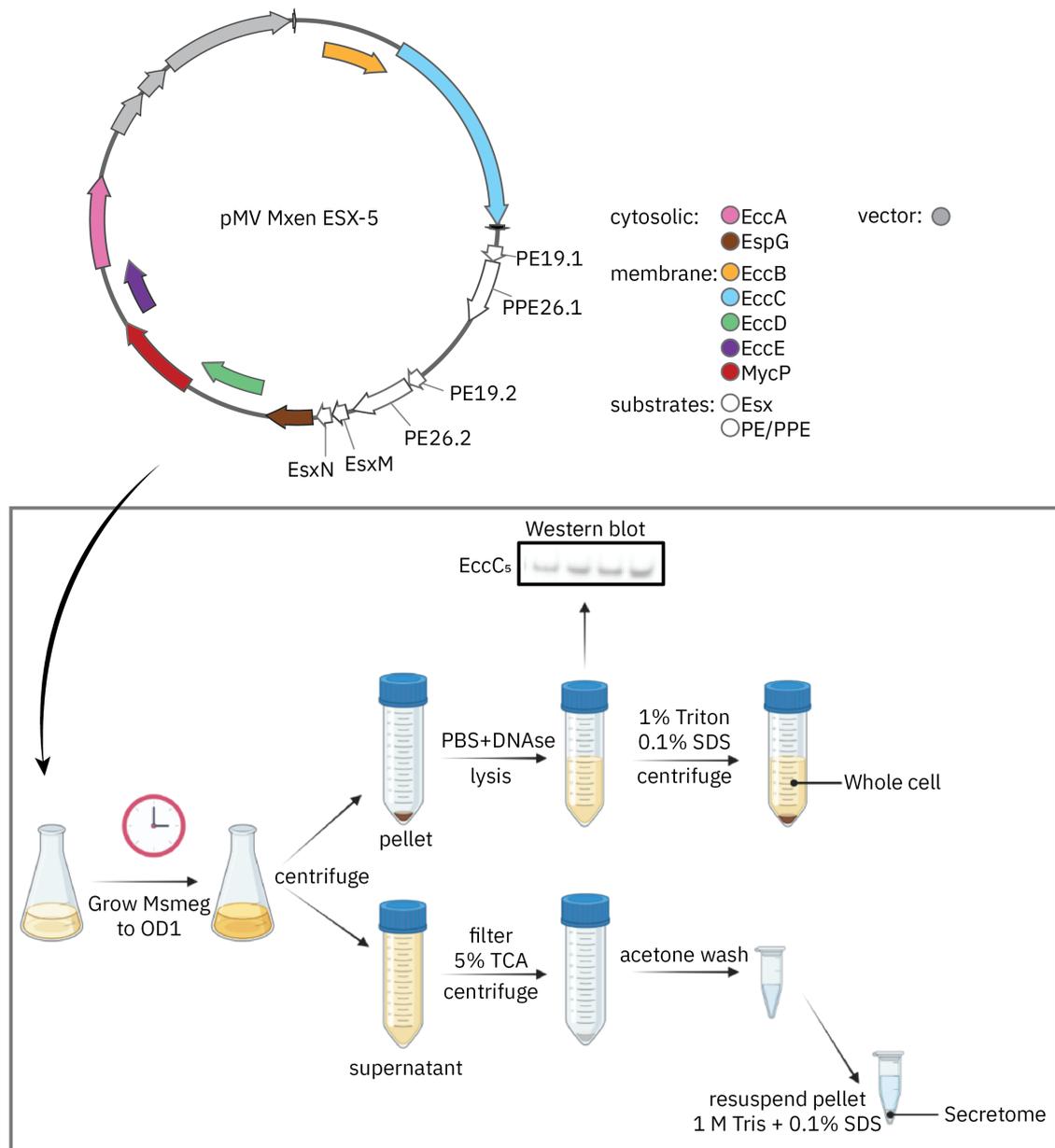


Figure 2.1: Overview of the secretion assay workflow A schematic representation of the secretion assay is highlighted. The figure was created with elements from BioRender.com

3 Results

The work described in this chapter focuses on the elucidation of the Mxen ESX-5 complex. If not further specified, the data in the following chapter refers to this system.

3.1 Structure of the ESX-5 core complex from *Mycobacterium xenopi*

Most of the work presented in this section was published in Beckham, Ritter, Chojnowski *et al.* [21]. The following subsections will thus resemble the published manuscript.

3.1.1 Purification of the ESX-5 core complex

For structural studies, the ESX-5 locus from Mxen was recombinantly expressed in Msmeg. Following membrane solubilization with DDM the membrane complex was isolated by strep-tag affinity purification. The localization of the strep-tag at the C-terminus of EccC₅ allowed, as expected, the pull-down of EccC₅, EccB₅, EccD₅ and EccE₅, as shown by SDS-PAGE (Figure 3.1, A) and confirmed by MS. These results indicate that a stable complex was formed by these four components. The fifth membrane component, MycP₅, was not detected in the elution fraction, reinforcing previous observations that MycP₅ is only forming transient interactions with the complex [239, 238, 121]. The eluted fraction was exchanged into Amphipol A8-35, which has been shown before to be crucial for homogeneity and the hexameric assembly of ESX-5 [20].

By size exclusion chromatography, smaller subcomplexes and lower oligomeric species could be separated from the protein complex. SDS-PAGE analysis revealed that fractions eluting in the void peak of the Superose 6 column contained all four components to a high purity (Figure 3.1 B, D). On BN-PAGE, the same fractions migrated in a single band at a size >1236 kDa, indicating the formation of a homogeneous, high-oligomeric complex, as shown in Figure 3.1, C.

To further judge the size and homogeneity of the sample, NS-EM was performed (Figure 3.1, E). The vast majority of particles had a size of 25 nm - 30 nm with a few aggregations or smaller particles in the background. Both can be explained by the coelution of aggregates in the void volume of the Superose 6 column.

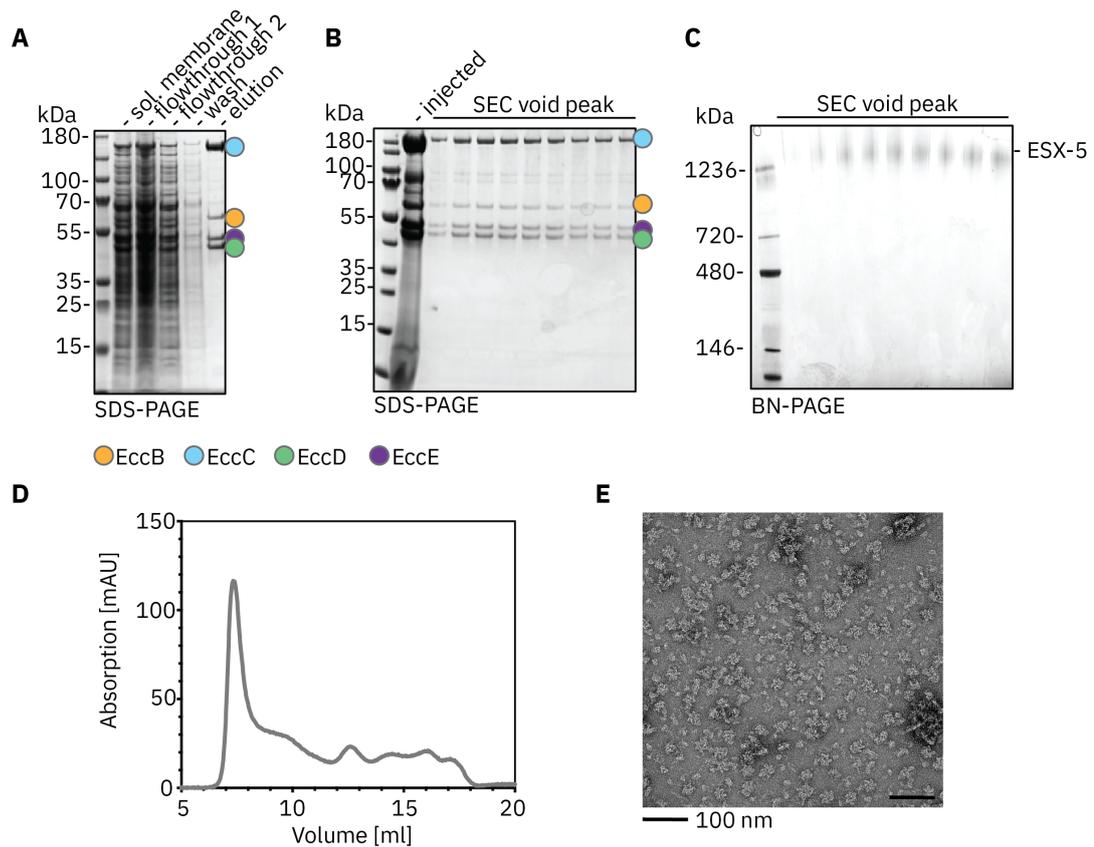


Figure 3.1: Representative purification of the ESX-5 complex (A) SDS-PAGE of the strep-tag affinity purification. The migration height corresponding to the proteins forming the ESX-5 core is indicated by colored circles. The tag at the C-terminal end of EccC₅ enabled the pulldown of EccC₅, EccB₅, EccE₅ and EccD₅. (B) SDS-PAGE of the void peak from size exclusion chromatography. The protein components that were pulled down coeluted in the void peak fraction. (C) BN-PAGE of the void peak fractions from size exclusion chromatography. Bands at a size >1236 kDa indicate the formation of a high oligomeric complex. (D) Size exclusion chromatogram with UV280-absorption signal. Void peak fractions correspond to the hexameric complex. (E) NS-EM micrograph of void peak fraction showing hexameric particles at a size of around 30 nm.

3.1.2 Single particle cryo-EM on the ESX-5 core complex

Following NS-EM, the most homogeneous fraction was chosen and vitrified for cryo-EM analysis. An initial dataset including 6,067 movies was collected at the CSSB cryo-EM facility, Hamburg. Processing of this dataset allowed to judge the overall quality of the data (Figure 5.1). The maps processed in C1 and later in C6 unveiled a hexameric complex with promising density in the membrane region and the cytoplasmic regions close to the membrane. First secondary structure information became visible, but the resolution of 6.3 Å of the C1 map and 4.4 Å of the C6 map was not sufficient to allow model building, especially since only parts of the maps achieved these resolutions, what can be deduced by the GSFSC curves shown in Figure 5.1.

To enable more homogenous and bigger particle sets, a second dataset with 27,873 movies was collected at the EM facility at EMBL Heidelberg (Figure 5.2). The C1 map built by 121,974 particles reached a resolution of 3.9 Å. From this map, it became evident that the membrane protein complex is built by six identical protomeric units that follow a clear C6 symmetry.

Reprocessing the data in C6 further improved the resolution of the hexameric complex to 3.4 Å. Focusing the refinement after symmetry expansion on the cytoplasmic density of the protomeric unit further increased the resolution to 3.0 Å.

Based on these maps, it was possible to build 88% of the membrane region and the membrane adjoining regions of the hexameric ESX-5 complex *de novo* (Figure 3.2, Figure 3.6).

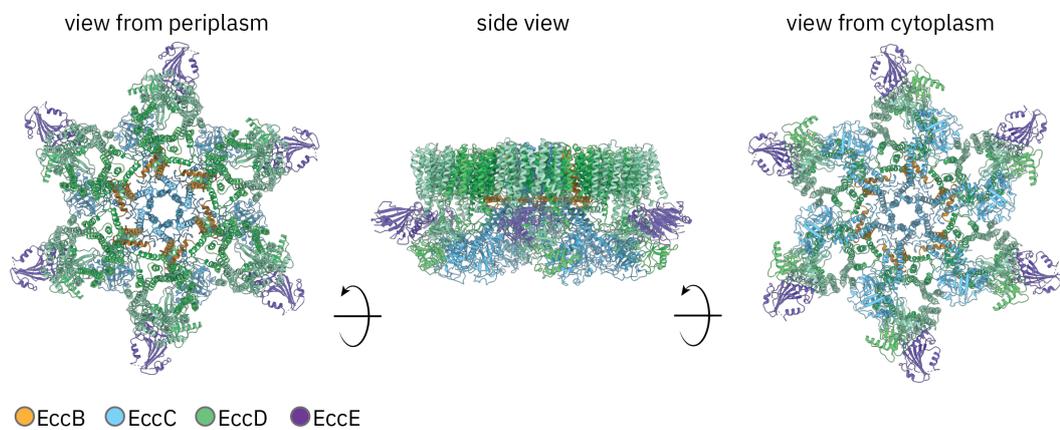


Figure 3.2: High resolution structure of the ESX-5 complex High resolution structural information was obtained for the membrane part and the upper cytoplasmic part. The model built is shown from the the side, the periplasm and the cytoplasm. The figure was modified from Beckham *et al.* [21], Figure 1.

In contrast, the periplasmic and distal cytosolic part of the complex was poorly resolved in both maps (Figure 3.3, A, B). Therefore, it was necessary to sort and clean the dataset further by heterogenous refinement and 3D variability analysis. The resulting map, processed in C1, showed low resolution features in the periplasm and, when lowering the threshold of the map, also in the cytoplasm. Interestingly, in the periplasm, the density did not follow C6 symmetry but rather C2 symmetry. As the resolution of the periplasmic region and the cytoplasmic region (>10 Å) did not allow model building, an integrative structural biology approach was applied, combining the EM map information with distance restraints from crosslinking MS (Figure 3.3) and homology models of the periplasmic EccB domain and the cytoplasmic ATPase domains of EccC. While the resulting ensemble model for the cytoplasmic domains gave rough estimates on conformational flexibility in the cytoplasm, EccB₅ ensemble models could be placed with high precision (7 Å) into the map. Taken together,

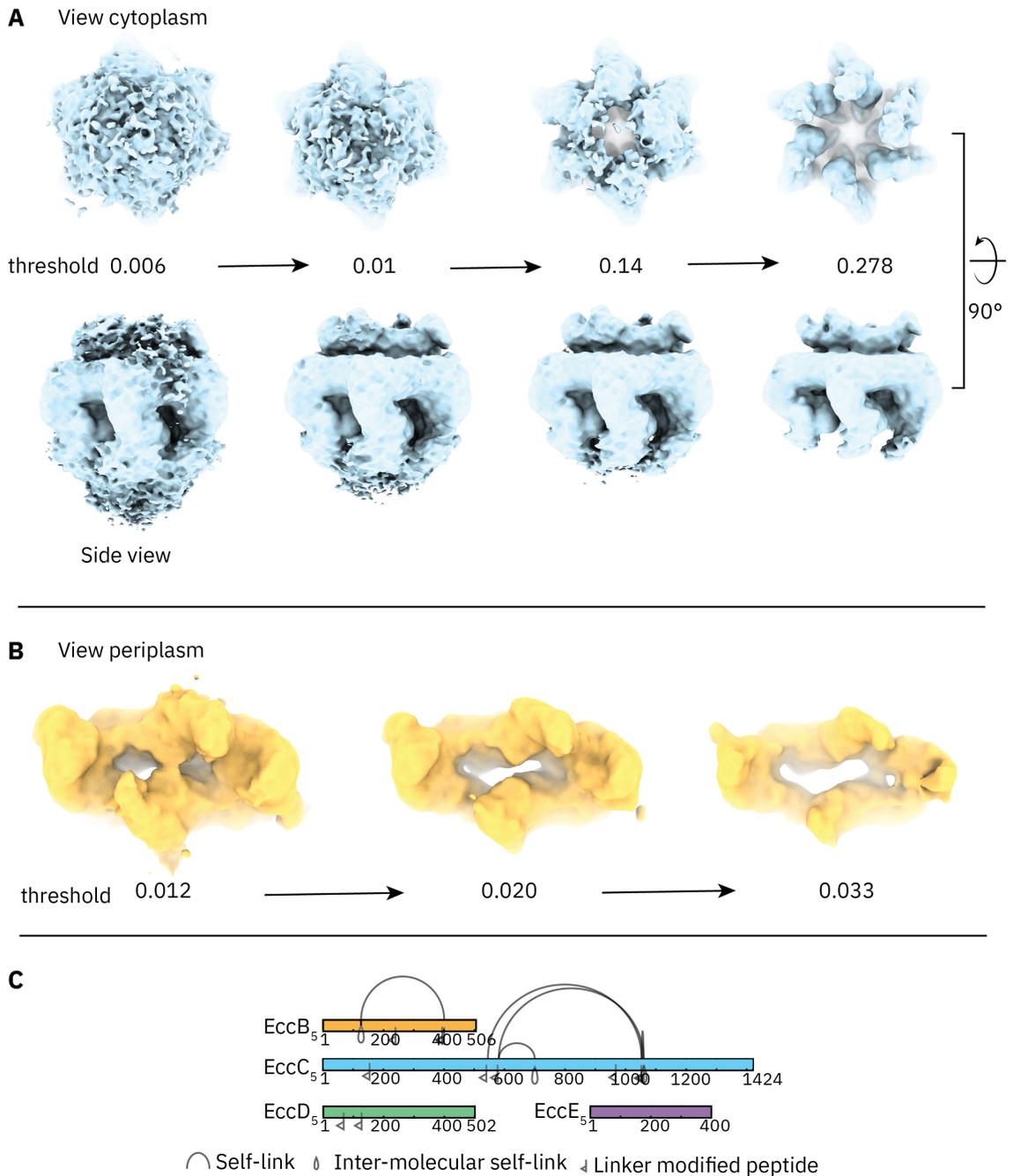


Figure 3.3: Low resolution EM density and crosslinks obtained by crosslinking MS The blue map (A) shows the map used for integrative modeling of Ecc₅ and the orange (B) map was used for integrative modeling of EccB₅. In both cases density only became visible at low resolution. (C) Crosslinks were mainly obtained within Ecc₅ and EccB₅. The figure was modified from Beckham *et al.* [21], Figure S1 and S6.

these models enable the interpretation of the complete ESX-5 core in molecular detail and allow the interpretation of intermolecular interactions in different regions of the complex (Figure 3.4, Figure 3.5).

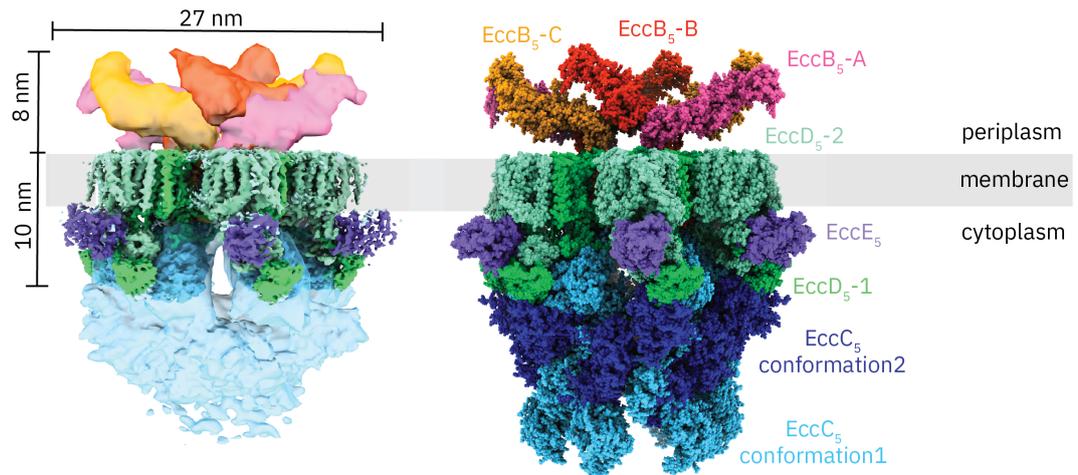


Figure 3.4: Composite ESX-5 EM map and structural model of the ESX-5 system The map is colored according to the model shown on the right. For flexible parts on the distal cytosolic side and the periplasmic side, integrative models were built based on the cryo-EM maps, crosslinking mass spectrometry and homology models. The highest scoring models are displayed on the right. Integrative modeling of EccC₅ revealed two major conformations that are displayed in light blue and dark blue colors, respectively. The figure was modified from Beckham *et al.* [21], Figure 1.

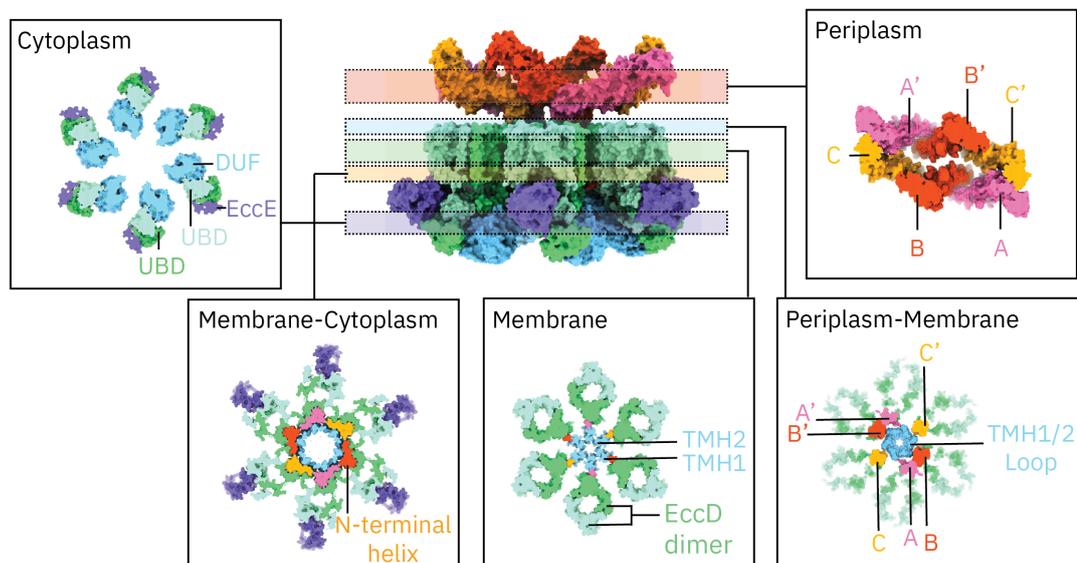


Figure 3.5: Cross-sections of the ESX-5 complex The cross-sections highlight interactions and regions at the cytoplasmic, membrane-cytoplasmic, membrane, periplasmic-membrane and at the periplasmic site shown as separate insets. The figure was modified from Beckham *et al.* [21], Figure 2.

3.1.3 The protomer - the basic unit of the ESX-5 core complex

Six identical protomeric units form the basic building blocks of the ESX-5 secretion system, as they are arranged symmetrically around the central pore (Figure 3.1, Figure 3.10, right).

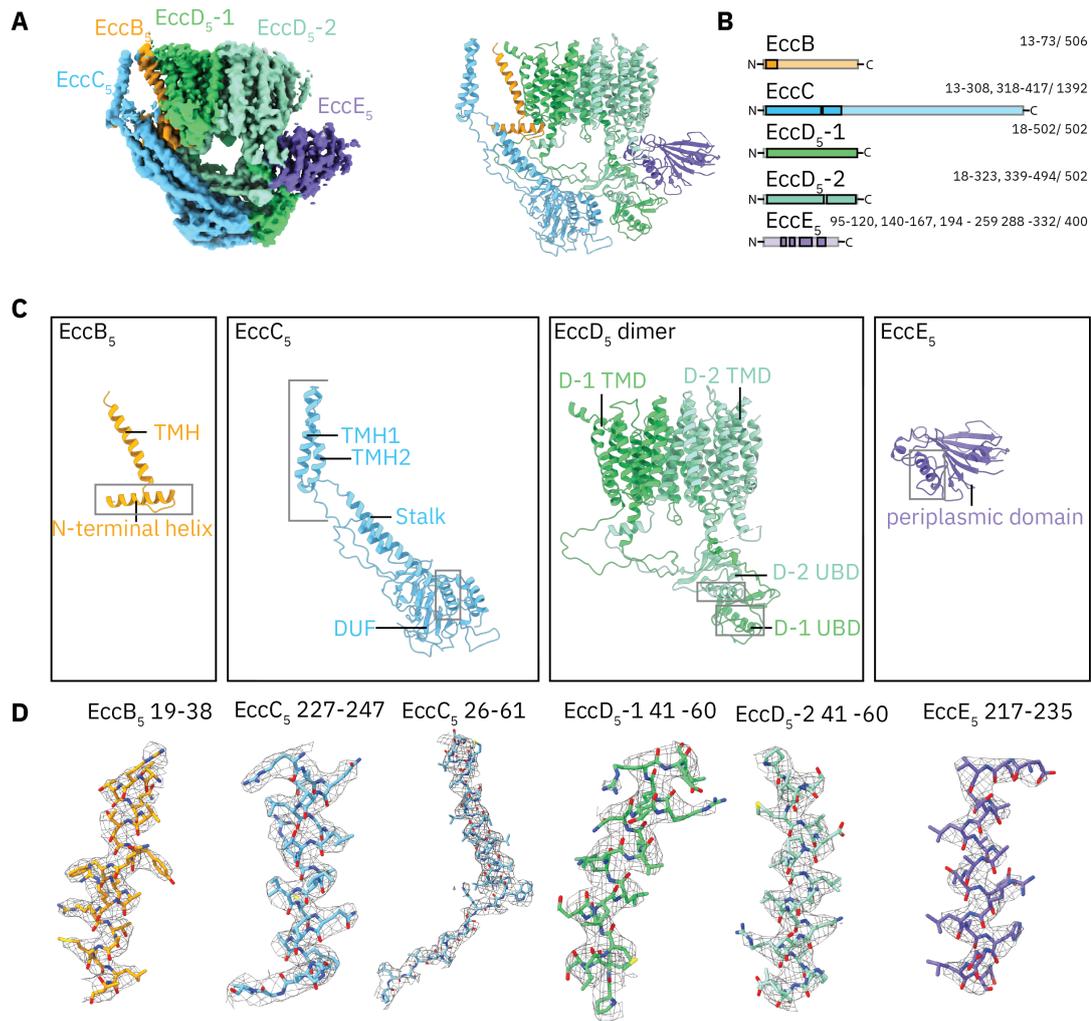


Figure 3.6: Structural insights into the protomeric composition of the ESX-5 system.

(A) cryo-EM map and model build *de novo* into the density (B) Schematic showing the sequence coverage of the high resolution model. Light colors indicate regions that could not be reliably modelled (C) Structures of the individual components that form the core of the secretion system. Functional domains and key features are labeled. (D) Representative density of the local refined map that was used for model building. The location of the helices selected is indicated in (C). The figure was modified from Beckham *et al.* [21], Figure S4.

Each protomer is composed of EccC₅, EccB₅, EccD₅ and EccE₅ in a 1:1:2:1 stoichiometry (Figure 3.6, A). Within the membrane, the center of the protomer is formed by an EccD₅ dimer. The 11 transmembrane helices (TMH) of each EccD₅ molecule combined form an elliptic ring in the membrane. The arrangement follows a quasi twofold symmetry with a dimerization interface between helix 2 of one EccD₅ molecule and helices 9 and 10 of the other EccD₅ molecule, respectively (Figure 3.7, A). Within the cytoplasm, the ring is further stabilized by interactions of a linker domain between TMH6 and TMH7 and a cytoplasmic helix in the linker domain between the transmembrane domain and cytoplasmic domain. Interestingly, the ring shaped arrangement is not empty, but is filled with elongated densities

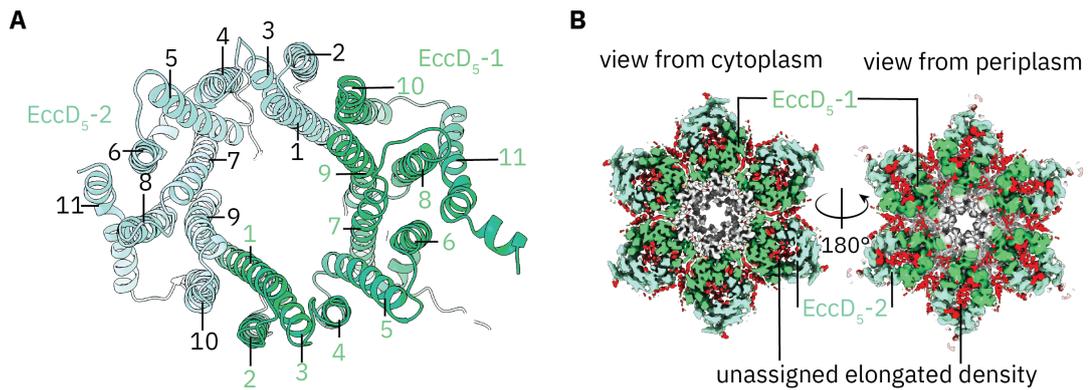


Figure 3.7: EccD₅ cavity and unassigned density (A) Membrane cavity formed by the 22 TMHs of two EccD₅ molecules. The helices of each EccD₅ molecule are numbered from 1 to 11. (B) Cryo-EM density of ESX-5. Density that could not be assigned to protein is shown in red. Unassigned elongated density is accumulated in the cavity formed by EccD₅-1 and EccD₅-2. The figure was modified from Beckham *et al.* [21], Figure S3.

that could not be assigned to protein. It is conceivable that this density accounts for lipids that copurify with the complex (Figure 3.7, B). A preliminary MS analysis confirmed that despite detergent solubilization and amphipol reconstitution, lipids are still bound to the complex [21]. TMH11 of each EccD₅ molecule was observed to adapt a diagonal shape within the membrane. This helix corresponds to the helix closest or most distal to the central pore. The EccD₅ molecule with TMH11 closest to the pore is referred to in the following as EccD₅-1 and the molecule with TMH11 most distal to the central pore as EccD₅-2. EccD₅ TMHs could not be modeled with high confidence into the EM density, but based on the data it seems reasonable that TMH11 of EccD₅-2 interacts similarly with EccE₅ as TMH11 of EccD₅-1 does with the TMH of EccB₅. The interaction defined by the TMH of EccB₅ and TMH11 of EccD₅-1 define a central scaffold for the EccC₅ TMHs that are forming the central pore.

In the cytosol, the TMHs of EccC₅ are followed by a stalk domain that is encompassed by the N-terminal EccB₅ helix oriented parallel to the membrane (Figure 3.1.3). This interaction is highly stabilized by hydrogen bonds as well as salt bridges between two aspartate residues on EccC₅ (D101, D112) and arginine residues on EccB₅ (R31, R32). The stalk domain is further flanked by the extended loop between TMH6 and TMH7 of EccD₅-1 and the cytoplasmic linker of EccD₅-2. Especially the latter one needs to be highlighted in that context, as the interaction is not only mediated by hydrogen bonds but also through salt bridges between three residue pairs. The stalk domain is continued by the DUF domain. A foldseek search [235] revealed similarity of the DUF domain to the ATPase FtsK, suggesting EccC₅ comprises not only three, but four consecutive FtsK/SpoIII ATPase domains. However, the canonical Walker motif is not conserved in the DUF domain (Figure 3.23). The two EccD₅ UBD domains dimerize and interact both with the DUF domain. Thus, EccD₅

also forms a scaffold in the cytoplasm. The EccD₅-1 UBD is located closer to the membrane compared to the EccD₅-2 UBD domain. Interactions of the UBDs with the DUF domain are mediated only by a few hydrogen bonds and a salt bridge. On the opposite side of the two UBD domains, the soluble domain of EccE₅ is binding. The main interaction to EccE₅ in the cytoplasm is however mediated by the cytosolic linker of EccD₅-1 and the extended loop between TMH-6 and TMH-7 of EccD₅-2. The cytosolic domain of EccE₅ could not be completely resolved, but is composed by a central multiple-stranded β -sheet, surrounded by α -helices. Structural homology search for EccE₅ revealed no hits.

While the DUF domain, UBD domains and the uncharacterized EccE₅ domains could be assigned to well-defined density in the map, density more distant to the membrane, corresponding to the three consecutive ATPase domains of EccC₅ could not be modelled confidently. This suggests high flexibility of these domains. The integrative ensemble model shows that six EccC₅ domains can be fitted into the density in multiple conformations - ranging from closed arrangements with extended EccC₅ domains to more open conformations. Adapting multiple conformations of the ATPase domains could be a key to the function of T7SS.

3.1.4 Periplasmic interaction network formed by EccB₅

Periplasmic density in the structure was assigned to EccB₅. The resulting ensemble model of EccB₅ is composed of three different dimer pairs formed by EccB₅-A and EccB₅-B, EccB₅-B and EccB₅-C, and EccB₅-C and EccB₅-A, as shown in Figure 3.8. The components follow an approximate twofold symmetry with their respective symmetry mates (EccB₅-B', EccB₅-C', EccB₅-A'). Overall, the arrangement resembles the keel of a boat, with the dimensions of 20 nm in length, 10 nm in width, and 8 nm in height, separated by an undulated cleft.

On the long side of the keel, the dimer composed by EccB₅-A and EccB₅-B forms a V-shape, in which the repeat domains R1 and R4 are interacting, respectively. The wide opening angle of the V-shape accounts for approximately 113°. In contrast, the dimer composed by EccB₅-B and EccB₅-C forms a parallel arrangement, in which both EccB₅ molecules are oriented in a similar way, but are shifted in respect to each other. Based on the ensemble model, contacts between EccB₅-B and EccB₅-C are mediated by the R3 and central domain of EccB₅-B with the Central domain and R1 domain of EccB₅-C, respectively. At the tip of the keel, EccB₅-C and EccB₅-A' are interacting with each other via their domains most distal to the membrane (R2 and R3).

As EccB₅ is a very modular protein, consisting of repetitive domains (repeat domains) [242], it is conceivable that it mediates further interactions. In addition, the presence of different conformations of EccB₅ in the periplasm suggests a high flexibility of the hinge region between the transmembrane domain and the periplasmic domain. An angle analysis of the six EccB₅ molecules in the ensemble models revealed high variability of the angles between

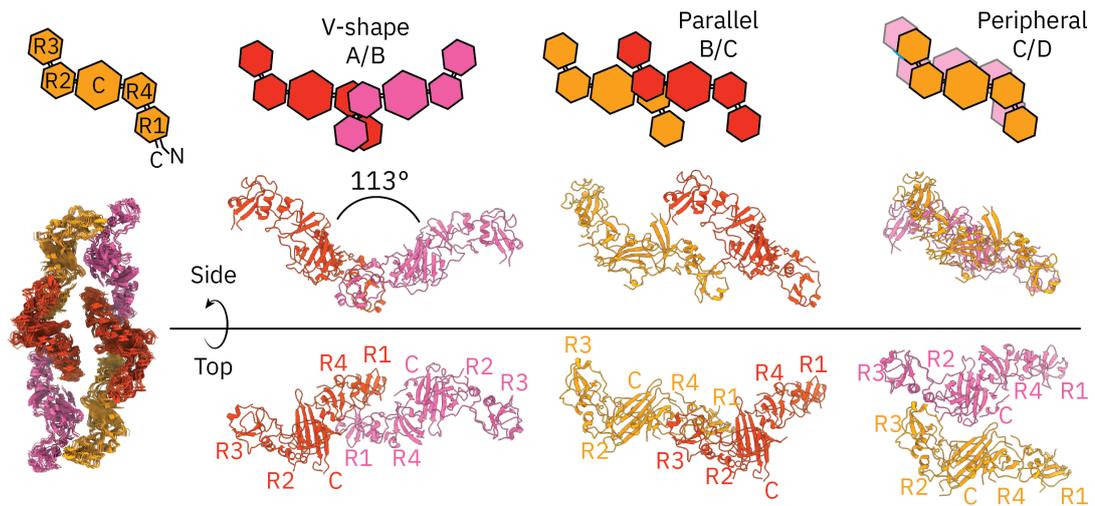


Figure 3.8: Composition of the periplasmic domain The modular organization of the individual EccB₅ domains is shown in the schematic on the top left. Three different EccB₅ dimer pairs form the periplasmic domain of ESX-5, shown on the bottom left. A schematic on the top depicts the three different dimers. The corresponding atomic model is shown from side and top in cartoon representation. The figure was modified from Beckham *et al.* [21], Figure 2.

transmembrane domain and periplasmic domain ranging from 53.4° for EccB₅-B' to 79.0° for EccB₅-C. This high flexibility could be key for conformational changes of EccB₅ that allow closing and opening of the transmembrane and periplasmic pore.

3.1.5 Key interprotomeric interactions in the transmembrane region

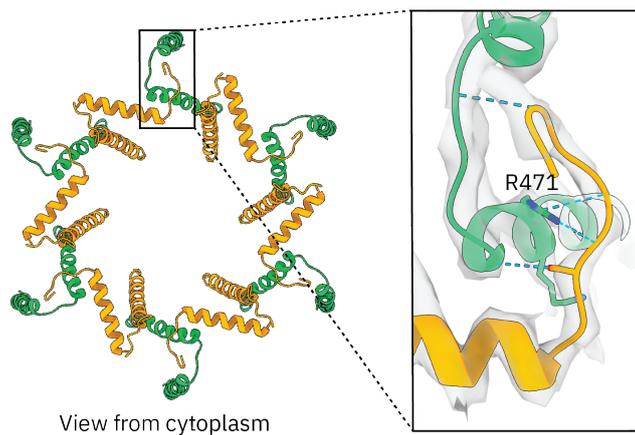


Figure 3.9: Interprotomeric interaction observed between EccB₅ and EccD₅ The EccB₅ N-terminus (orange) hooks into a loop between EccD₅ TMH10 and TMH11 (green). The figure was modified from Beckham *et al.* [21], Figure S4.

Besides the interprotomer interactions of EccB₅ molecules in the periplasm, few further in-

teresting interprotomeric interactions could be observed in the transmembrane region that could be involved in the assembly of the hexameric complex. One of the interactions occurred between the N-terminus of EccB₅ and the loop between TMH10 and TMH11 of EccD₅ (Figure 3.9). The EccB₅ N-terminus seems to hook into the EccD₅ loop. A hydrogen bond network involving EccD₅ R471 as key residue could be crucial to stabilize the interaction in between different protomers. In that regard, the variability of the length and sequence of the N-termini in different ESX loci could play a regulatory role in the assembly of a hexameric complex and could be essential for the formation of non-hybrid complexes.

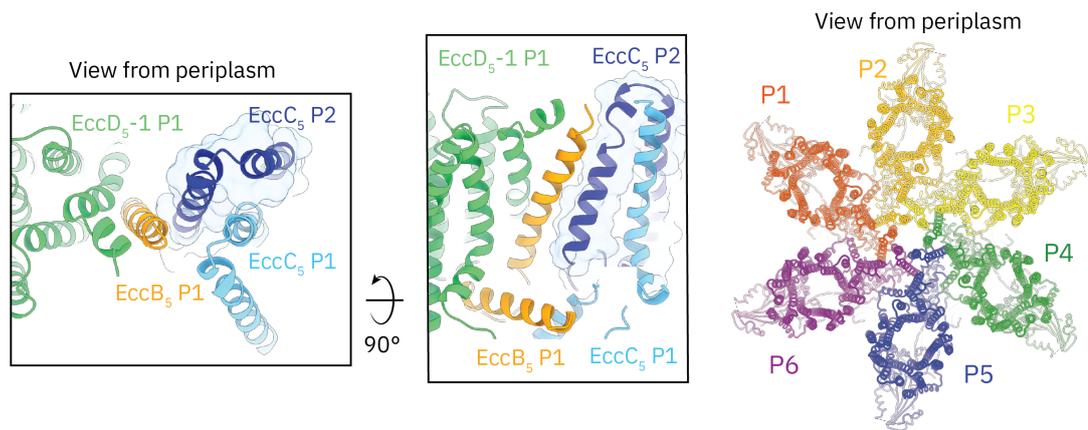


Figure 3.10: Domain swap interaction of EccC₅ TMH1 TMH 1 is sandwiched between the TMH of EccB₅ and TMH2 of EccC₅ from the next protomer as shown on the left from the cytoplasm and from side view. On the right, each protomer is colored differently to highlight the domain swap interaction of the pore. The figure was modified from Beckham *et al.* [21], Figure 2.

The transmembrane domains of EccB₅ and EccC₅ seem to contribute to the hexamerization as well, as EccC₅ TMH1 of one protomer is swapping to the next protomer, where it is sandwiched in between TMH2 and the EccB₅ TMH of the next protomer (Figure 3.10). This domain swapping occurs in an anticlockwise fashion from the periplasmic perspective and is stabilized by hydrophobic interactions.

The hierarchy of events leading to the assembly of the hexameric complex is not clear. One hypothesis is that first interprotomeric interactions occur by the periplasmic domains before the hexameric assembly is locked by the N-terminal EccB₅ hook. As a consequence, the EccB₅ hook interaction could drive the domain swap of the EccC₅ helices. These events are crucial, since they enable the formation of the central transmembrane pore that is required for the secretion of substrates.

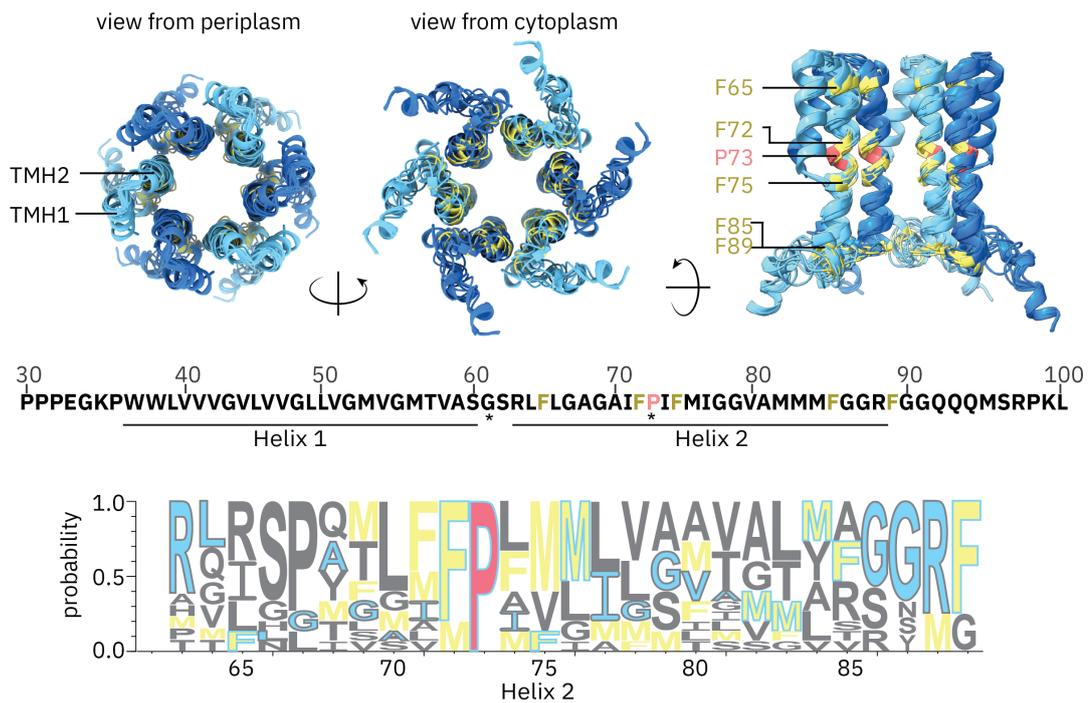


Figure 3.11: Ensemble model of the transmembrane pore from ESX-5 The ensemble model of the ESX-5 pore helices is shown from the periplasmic and cytosolic side as well as cross section from the side. Helices from separate protomers are colored alternatingly in blue and dark blue. The inner face of the pore is formed by EccC₅ TMH2 and is surrounded by EccC₅ TMH1. The sequence of the two TMHs reveals the presence of a conserved proline residue and conserved bulky side chains like phenylalanine. The position of these residues is indicated in the sequence and in the model by red and yellow color. A sequence logo on the bottom highlights the conservation of bulky side chains (yellow) and the proline residue (red) in all ESX systems. The sequence of Mxen ESX-5 is highlighted in the logo in blue or with a blue frame, in case it coincided with a yellow or red residue. Further residues are colored gray. The figure was modified from Beckham *et al.* [21], Figure 3.

3.1.6 The central secretion pore formed by EccC₅ transmembrane helices

The central transmembrane pore of the ESX-5 core complex is composed by the EccC₅ TMHs, especially TMH2, that is forming the inner face of the pore (Figure 3.11). While the density for the transmembrane part of the whole complex was very well defined, the density for TMH2 could only be interpreted when blurring the map. The lower resolution of this part is caused by flexibility that is reflected in the ensemble model built by the 100 highest scoring Rosetta [91] models. Notably, a central, conserved proline residue induces a kink in TMH2 after which the resolution decreased (Figure 3.12, A, B). This has been similarly observed in other membrane transporters, in which proline residues were reported to play key roles in regulation and function [54, 116]. In the context of protein transport by the T7SS, P73 could act as hinge point, enabling the adaption of multiple conformations that may be necessary to transport a range of different substrates across the cell membrane.

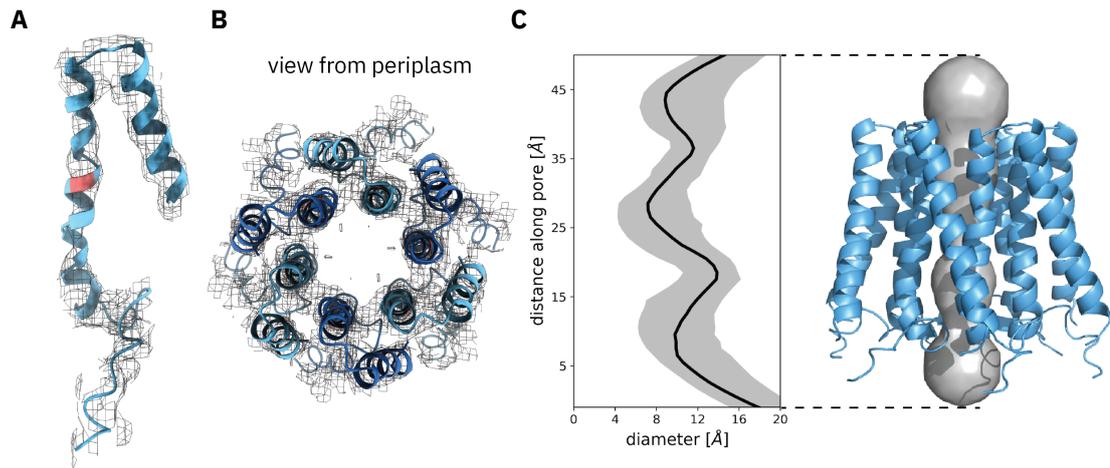


Figure 3.12: Structural role of specific side chains in the ESX-5 secretion pore (A) Cryo-EM density and model of the pore helices reveals lower resolution after a conserved proline residue, shown in the top scoring Rosetta model (B) Cryo-EM density and model of the pore helices, shown as top view, gives insights into the dimensions of the central membrane pore (C) Analysis of the pore diameter with HOLE [216] performed on the ensemble model of the transmembrane pore. The median pore diameter is shown in black and the 90% confidence interval is shown in gray. A graphical visualization of the pore diameter is shown for the top scoring model. The figure was modified from Beckham *et al.* [21], Figure 3.

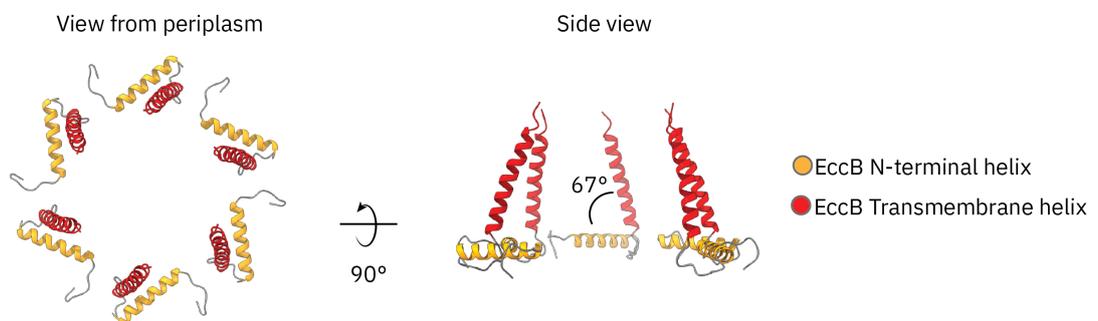


Figure 3.13: EccB₅ arrangement within the membrane The N-terminal and TMH of EccB₅ form a basket-like arrangement that encloses the EccC₅ pore helices. The figure was modified from Beckham *et al.* [21], Figure S7.

Analysis of the diameter of the central pore with HOLE [216] revealed three constriction sites in the 100 highest scoring ensemble models, with the narrowest constriction resulting in a diameter of less than 1 nm (Figure 3.12, C). These constriction sites correspond to phenylalanine residues F66, F72 and F75 distributed in EccC₅ TMH2. The presence of bulky side chains like phenylalanine and methionine within the pore helices are conserved within all five ESX loci (Figure 3.11). As folded substrates, like the Esx dimer pairs, have an approximate size of 2.2 nm [201, 225, 137, 138] and, thus, would not be able to pass the transmembrane pore in the present model, it can be concluded that ESX-5 adapts a gated,

secretion competent conformation in this structure. This might be due to the absence of substrates and cofactors, as the system needs to restrict the pore diameter to prevent leakage of molecules. This concept is not unique to ESX-5. Other transport systems show similarities, like the T3SS in which a conserved Met-Met-Met loop similarly limits the pore diameter to less than 1 nm [124, 160]. This so called M-gasket was reported to play a dual role during secretion: it maintains the membrane barrier during the secretion cycle and serves as a gate that allows secretion [124]. Other examples are A/B type toxins as the anthrax toxin from *Bacillus anthracis*. In this channel, a phenylalanine ring called Φ -clamp restricts the pore diameter to 6 Å and opening and closing of it underlies an allosteric regulation mechanism [114]. Both examples do not share any sequence similarity with each other or the T7SS, but they highlight fundamental common principles of protein transport in bacteria.

How pore opening is regulated in ESX systems is not clear, but interesting observations could be made in the arrangement of EccB₅ in the membrane: The N-terminal helix of EccB₅ and the TMH encompass the EccC₅ pore helices like a basket in which the angle between both helices defines the space for the EccC₅ TMHs (Figure 3.13). Conformational changes that lead to a higher angle between EccB₅ helices could be key to open the central pore. In turn, lower angles could be crucial to close the pore completely.

3.2 Dissecting the role of specific residues by site-directed mutagenesis

Structural analysis of the Mxen ESX-5 core complex highlighted residues that could play a role in mediating secretion. The role of these residues were probed by generating several site directed mutants and by testing the impact of these residues on the structure and function of the ESX-5 complex. The results of these experiments are provided in this section.

3.2.1 Targeting the integrity of the pore basket

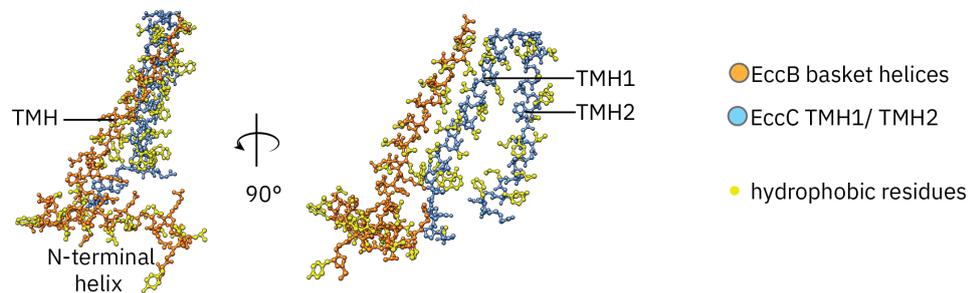


Figure 3.14: Interface of EccB₅ and EccC₅ within the membrane section TMH1 and TMH2 of Ecc₅ are shown in blue, TMH and N-terminal helix of EccB₅ are shown in orange. Side chains of hydrophobic residues are colored yellow. The interface between the TMH of EccB₅ and Ecc₅ TMH1 is hydrophobic.

The EccB₅ basket includes the N-terminal and TMH of EccB₅ that are enclosing the EccC₅ pore helices (Figure 3.13). Based on structural analysis described in the previous section, the interaction between EccB₅ and EccC₅ could play a role in a domain swap of EccC₅ TMH1 (Figure 3.10). In addition, EccB₅ TMH could serve as a scaffold for the pore helices, in which conformational changes in the relative positioning of the EccB₅ TMH in respect to the EccB₅ N-terminal helix could lead to pore opening and closing (Figure 3.13). To decipher the importance of the EccB₅-EccC₅ interaction, three mutants were designed aiming to alter it. As the interface between TMH1 of EccC₅ and the TMH of EccB₅ is hydrophobic (Figure 3.14), the strategy for destabilizing the interaction was to introduce bulky hydrophobic side chains like phenylalanine that may force the two helices apart. This should ensure proper insertion of the TMHs into the membrane. Based on these thoughts, the following mutants were generated: EccC₅ V41F, EccB₅ V57F and EccC₅ V40F V41F V44F.

3.2.2 Targeting the secretion pore

The inner membrane pore is thought to play a central role in the process of substrate translocation. The presence of bulky side chains within EccC₅ TMH2 is a conserved motif between ESX systems and even between species. As described in Subsection 3.1.6 and shown in Figure

3.1.6, similar to other transport systems, phenylalanine residues or methionine residues could be not only actively involved in the transport of substrates but also in gating of the pore (Figure 3.11, Figure 3.12, C). To test the importance of these residues, first, one mutation was obtained exchanging a phenylalanine to a cysteine at EccC₅ position 75. A cysteine is considerably smaller than phenylalanine and might, thus, alter substrate translocation efficiency. In addition, cysteine could form disulfide bonds that could lock the system in a closed state. If a disulfide bond is formed, substrates could get trapped within the secretion system. A second mutation introduced an alanine at position 72 that replaced the conserved phenylalanine.

Another interesting feature observed, was the conservation of a proline residue in the center of EccC₅ TMH2, position 73 (Figure 3.11, Figure 3.12, A). The proline residue is hypothesized to introduce a kink and substantial flexibility to this helix. As this flexibility might be needed to accommodate a range of substrates in the pore during secretion, it could be essential for secretion. To probe this hypothesis, the proline was replaced by an alanine, which is supposed to support the integrity of TMH2.

3.2.3 Impact of mutations on secretion

To assess the effect of mutations on secretion quantitatively, a proteomics approach was used. The mutations were introduced into the pMV expression plasmid by site directed mutagenesis and the vectors were transformed for expression into Msmeg. Whole cell fraction and secreted fractions were analyzed separately (see Figure 2.1). As Msmeg does not encode an ESX-5 secretion system, it can be expected that the Mxen ESX-5 substrates are solely secreted by the Mxen ESX-5 system.

An initial anti-strep western blot analysis of different mutant colonies confirmed the presence of EccC₅ in the whole cell fraction in colonies of five of the six mutants, indicating successful transformation and expression. Only the EccC₅ F72A mutant was not detected by western blot and was subsequently excluded from analysis. A reason for the low expression was a technical problem with the expression plasmid, which needs to be addressed in the future.

A robust quantification of the sample by MS was achieved by isobaric labeling with tandem mass tag. These are labels that have an identical mass, but differ in the presence of different heavy isotopes [251]. The labeling enabled the simultaneous measurement of all secreted fractions and whole cell fractions in one LC-MS/MS run, respectively.

To assess the level of secretion, the analysis focused on the abundance of secreted proteins in the secretome and in the whole cell fraction. Besides the Esx pair, EsxMN, two PE/PPE pairs are encoded within the Mxen ESX-5 locus. Both are homologues to PE19 and PPE26 in Mtb, respectively, and are named PE19.1, PE19.2, PPE26.1 and PPE26.2. It is, however,

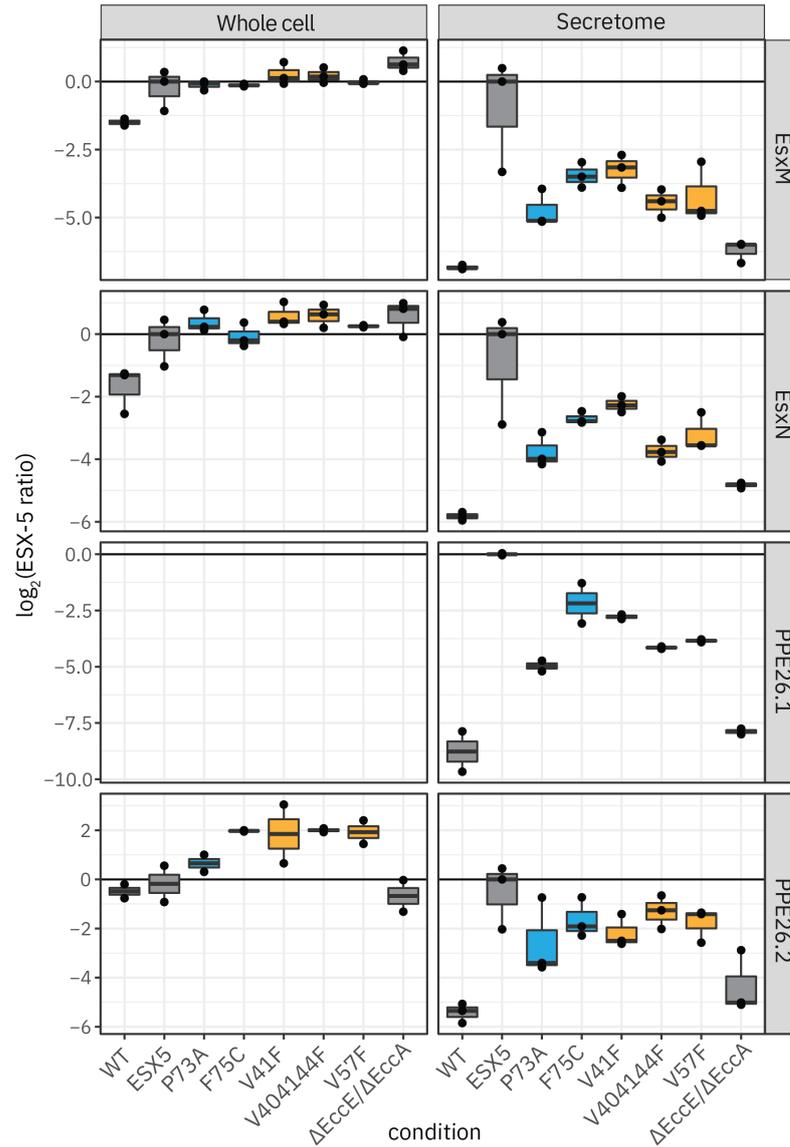


Figure 3.15: Impact of selected mutations on the secretion level of ESX-5 substrates Protein abundance of four secreted substrates, EsxM, EsxN, PPE26.1 and PPE26.2 are shown for 8 different ESX-5 constructs, normalized to ESX-5 levels. Results for controls are colored in gray, for pore mutants in blue and for pore basket mutants in orange. The figure shows the result of three independent measurements (N=3).

of note that PPE26.2 belongs to the family of PPE-SVP proteins whereas PPE26.1 cannot be clearly assigned to a class. Thus, it is expected that both proteins fulfill different functional roles.

By normalizing the LC-MS/MS results to ESX-5 WT levels, it becomes evident that secretion of EsxM and EsxN is significantly ($p < 0.06$) decreased in all analyzed mutants (Figure 3.15, Table 5.2). With an EsxM decrease of -2.3 on a \log_2 -fold scale, the effect was least distinct for the V41F mutant, whereas the P73A mutant showed the strongest effect with

-3.8 also on a \log_2 -fold scale. Calculated p-values and fold-changes can be found in Table 5.2 and Table 5.1. A $\Delta\text{EccE}_5/\Delta\text{EccA}_5$ mutant was introduced as control, since it was shown before that EccE_5 is a functional part of the membrane complex and required for secretion [80]. As expected, secretion was decreased to -5.3 times (\log_2), which can be considered as the absolute zero or abolished secretion.

Within the lysate, levels of EsxM and EsxN were stable for all mutations, which indicates a fast degradation of substrates or might even point to a regulation mechanism. Overall, it is striking, that EsxN and EsxM levels within the cell are very low, even in the ESX-5 WT, as the fold-change to Msmeg WT is far less than 2.5 on a \log_2 scale.

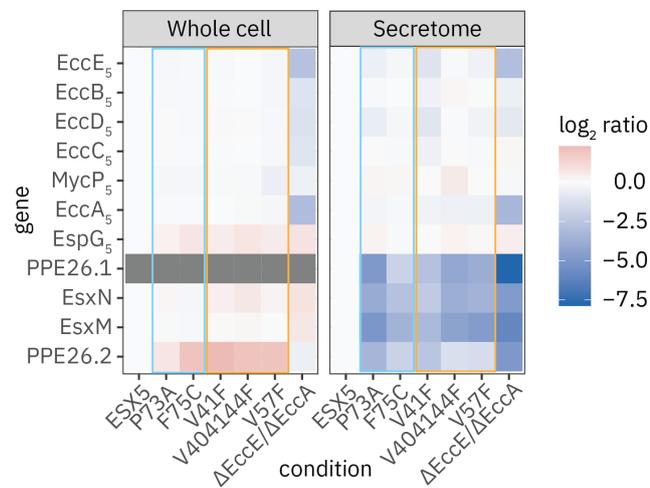


Figure 3.16: Impact of selected mutations on the secretion level of ESX-5 substrates in comparison to cytosolic and transmembrane components The level of cytosolic, membrane and secreted components is normalized to ESX-5 levels in the whole cell fraction and the secretome. Changes are indicated on a \log_2 scale by red (increasing concentration) and blue (decreasing concentration). Gray color indicates, that the protein was not detected at all, as it was the case for PPE26.1 in the whole cell fraction. Pore mutants are highlighted with a blue frame, basket mutants with an orange frame. The figure shows the result of three independent measurements ($N=3$).

PPE26.1 was not detected in the whole cell fraction and also PPE26.2 was not very abundant. Interestingly, compared to the ESX-5 WT, a slight, around twofold accumulation of substrate within the cell was observed for F75C, V41F, V404144F and S57F, but not for P73A and $\Delta\text{EccE}_5/\text{EccA}_5$. In the secretome, however, PPE26.1 and PPE26.2 follow a similar trend compared to the Esx proteins, although not all differences are significant.

PE19.1 and PE19.2 proteins were not detected at all, which might be due to the absence of suitable Trypsin cleavage sites as described in Ates 2020 [9].

To make sure the observed decrease of secretion was not a result of different expression levels, the abundance of other locus components was analyzed as well and is presented in Figure 3.16. The comparison between the different mutants revealed that the protein levels of all components, from membrane components to cytosolic components, was stable for all mutants with the exception of the $\Delta\text{EccE}_5/\Delta\text{EccA}_5$ mutant, which showed, as expected, a strong decrease in EccE_5 and EccA_5 in the whole cell and the secretomic fraction. Thus, it could be concluded that the expression of the pMV plasmid was not affected by the point mutations.

3.2.4 Structural impact of a conserved proline residue

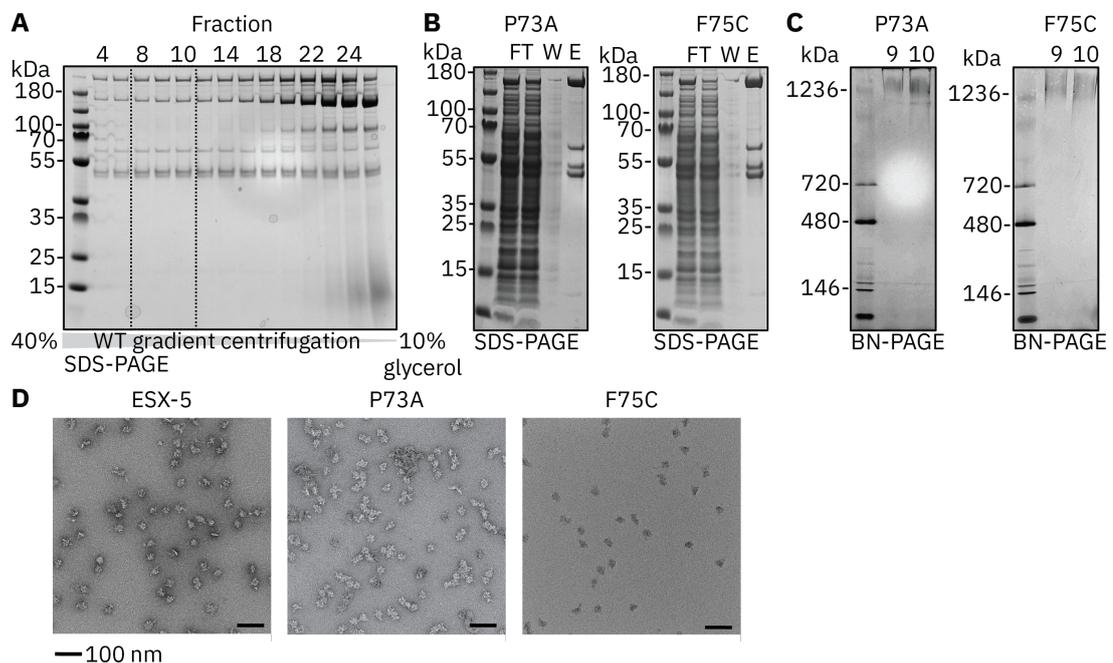


Figure 3.17: Purification and characterization of ESX-5 pore mutants P73A and F75C
 (A) Exemplary result of a 10% to 40% glycerol gradient centrifugation. Every second fraction is run on an SDS-PAGE for analysis. Fractions 8-10 can be considered as hexameric ESX-5 complex. (B) SDS-PAGE of the P73A and F75C affinity purification, showing samples from flowthrough (FT), wash (W) and elution (E). The pull-down of the four protein components confirms that the protomeric assembly of ESX-5 is intact. (C) BN-PAGE of P73A and F75C after gradient centrifugation. Fraction 9 and 10 migrate at a size >1236 kDa, suggesting that the hexameric assembly is not affected. (D) NS-EM micrographs of ESX-5 WT, P73A and F75C. Micrographs show particles of around 30 nm in size for all constructs, implying that changes introduced by mutations are subtle on the macroscopic scale.

As the pore mutant P73A had the strongest impact on secretion, it was analyzed if the pore mutations could hamper the assembly of the protomeric unit or the formation of a hexameric complex. Both mutants were purified as described in Subsection 3.1.1 by pull-down with a strep-tag at the EccC_5 C-terminus. Similar to the WT, all four components, EccB_5 , EccC_5 , EccD_5 and EccE_5 , could be pulled down, as shown by SDS-PAGE (Figure 3.17, B). This

suggests a stable formation of the protomeric unit, which is not affected by the mutations. After amphipol exchange and density gradient centrifugation (Figure 3.17, A), high molecular weight fractions were applied on BN-PAGE. Both mutants migrated at a size of >1200 kDa in accordance with the ESX-5 WT complex (Figure 3.17, C; compare to Figure 3.1). It could be concluded that also the assembly into a hexamer was not affected by the point mutations. This was further confirmed by NS-EM, where both mutants showed similar sized particles of around 30 nm in diameter. These particles had a distinct hexameric conformation (Figure 3.17, D).

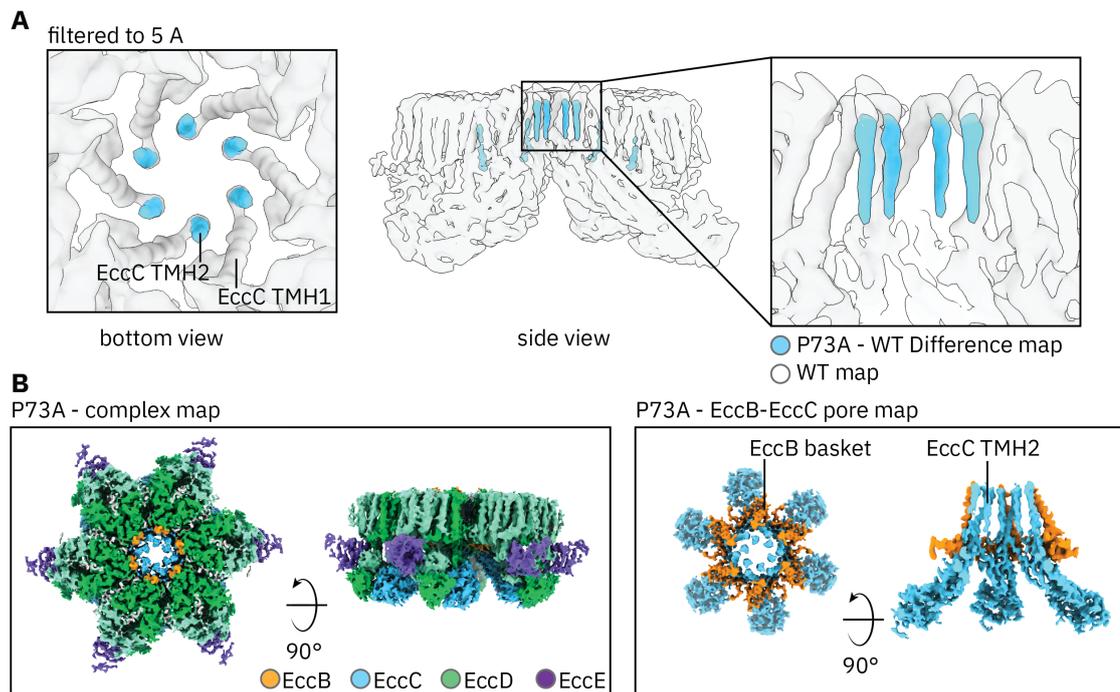


Figure 3.18: Cryo-EM map of the ESX-5 P73A mutant and difference map (A) ESX-5 WT shown in white and difference map of P73A and ESX-5 WT shown in blue. Before subtraction, maps were filtered to 5 Å. The difference map is displayed at a threshold level of 0.1. Main differences can be observed within the pore region, where the density of TMH2 is visible in the difference map. (B) Cryo-EM map of ESX-5 P73A colored according to the underlying protein model (EccB₅ orange, EccC₅ blue, EccD₅₋₁ green, EccD₅₋₂ light green, EccE₅ purple). Despite the mutation, the complex assembles into a hexamer with gated pore. On the right, the map of the inner section of EccC₅ and EccB₅ is shown in blue and orange, respectively.

Since a defect in protein assembly could be excluded for the pore mutants, it can be assumed that the impact of the mutation is manifested on the atomic-scale. Thus, the impact of the P73A mutation on the ESX-5 structure was assessed using cryo-EM. For this purpose, high molecular weight fractions were screened via NS-EM and the most homogeneous fraction was selected for plunge-freezing. After screening, the best cryo-EM grid was selected for data collection. In total 10,233 movies were collected that were processed in C1 and C6

symmetry as shown in Figure 5.3. While the resolution of the C1 map was limited to 6.2 Å, the C6 map reached a global resolution of 3.9 Å. However, the local resolution was worse in some regions of the map as indicated by the GSFSC curve. This could be due to a lower particle number and, as a result, a lower population of rare viewing angles.

A model was built into the density by refining the WT structure into the EM density of the mutant map. Interestingly, the overall structural arrangement of the P73A mutant was comparable to WT ESX-5 as it was built by six identical protomeric units formed by the protein components EccB₅, EccC₅, EccD₅ and EccE₅ in a 1:1:2:1 stoichiometry (Figure 3.18, B). Even the arrangement of pore helices was similar compared to the WT structure. However, the resolution of the second TMH of EccC₅ was improved to around 4.5 Å - 5 Å. As a result, more details became visible. This was expected since the mutation was designed to stabilize a helical secondary structure and to rigidify the pore. After filtering the WT and mutant map to 5 Å resolution and calculating a difference map, the stabilization of EccC₅ TMH2 became even more explicit (Figure 3.18, A). In the resulting difference map, main density can be assigned to EccC₅ TMH2, confirming that the increased resolution of this helix is a main difference between both maps. Furthermore, the helix is straightened compared to the WT structure, which showed a substantial kink introduced by the proline residue. This is in accordance to other studies that demonstrated that proline residues introduce kink-swivel motions to helices crucial for the function of the respective transmembrane proteins [31]. The conservation of the proline residue within all ESX systems suggests that these motions contribute to a general secretion mechanism.

Taken together, the results indicate that P73 is inducing a helix break that enables intrahelical dynamics within TMH2, as similarly described for other membrane transporters [54, 116, 73]. The results suggest a fundamental role of plasticity within the inner membrane pore for ESX-mediated protein translocation across the inner membrane.

3.2.5 Molecular dynamics study on the EccC₅ pore helices

The flexibility of the TMHs of EccC₅ was further assessed by MD simulations. To reduce the complexity of the system, the TMHs of EccC₅ were considered separately in a POPE bilayer. POPE was selected as phosphatidylethanolamine represents one of the most abundant lipids in the mycobacterial membrane [46]. Still, it is important to keep in mind that this choice represents only a simplification of a complex lipid bilayer composition in mycobacteria.

MD simulations are often limited by the time-scales in which a reaction or conformational change occurs. This is due to the fact that the procedure requires a lot of computing resources [117]. However, to analyse the flexibility of helices, it can be expected that significant changes will be observed already within a few nanoseconds [112, 31]. Indeed, when following the RMSD of the C_α backbone over a time of 100 ns (1000 frames), after 20 ns, the RMSD

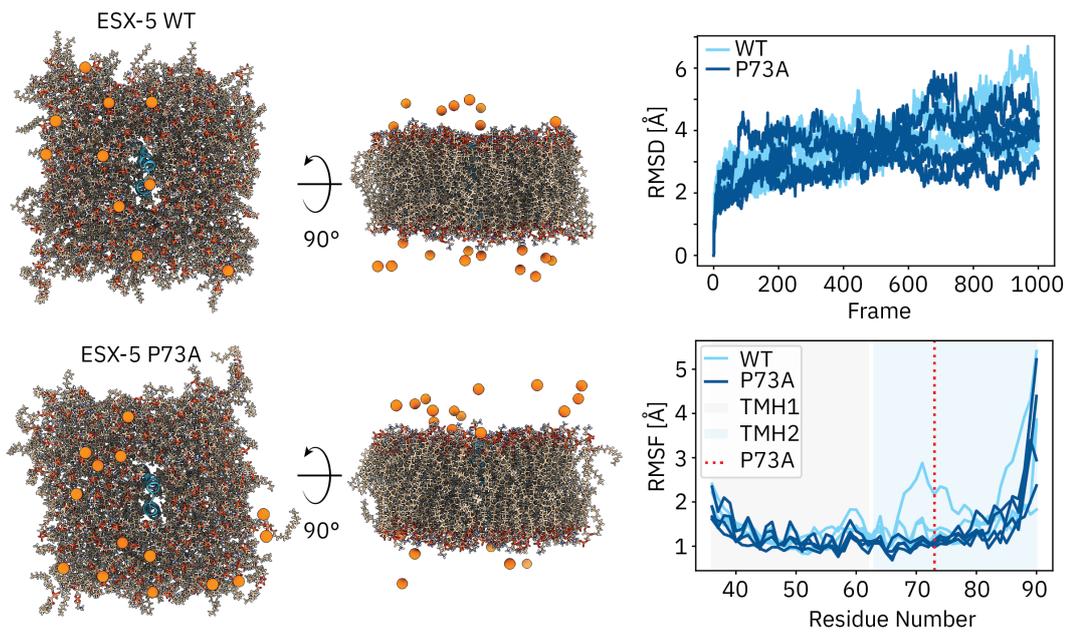


Figure 3.19: Molecular dynamics simulation of EccC₅ TMH1 and TMH2 for ESX-5 WT and ESX-5 P73A On the left, input models are shown for ESX-5 WT and ESX-5 P73A from the periplasmic side and as side view. The helices were embedded in a double membrane composed of POPE (beige tails and red heads) and ions were added to both sides of the membrane (orange). Results of the 100 ns simulation in GROMACS [23] are shown on the right. The first panel shows the time dependent RMSD of the C α backbone for the ESX-5 WT (light blue) and the ESX-5 mutant (blue). In the lower panel, the RMSF is shown in relation to the position of the residue. Position 73 is highlighted with a dotted red line. Except for one outlier, the MD simulations could not prove higher flexibility of ESX-5 WT helices.

was stable in a range of about 2-6 Å for both, the WT helices and the P73A mutant helices (Figure 3.19). This implies that main movements were completed within the first 200 nanoseconds.

The overall dynamics of the backbone, indicated by the RMSD of both helices, did not differ significantly between mutant and WT. To get further insights into the contribution of different regions to the flexibility, the RMSF (root mean square fluctuation) per residue was calculated (Figure 3.19). With less than 2 Å the fluctuation of residues 40 - 79 was rather low. Only one simulation of the WT helices showed an increased fluctuation around P73, that was however not reproducible by three further MD simulation runs of the WT helices. Thus, the simulation could not prove the flexibility introduced by the proline residue at position 73. Starting at residue 80, the fluctuation increased for mutant and WT, confirming a high variability of the loop region that connects the stalk domain with the TMHs. This flexibility was also observed in both WT and mutant structures. However, as these loops are located

at the C-terminal end in this simulation setup, it can be expected that the overall flexibility of this region is even further increased compared to the full length EccC₅ structure or in context of the full complex. This could have a crucial impact on the simulation results in the P73 region. Thus, the simplification introduced might be insufficient to describe dynamics in the context of a big complex.

3.3 Understanding conformational flexibility of the ESX-5 core complex

The cryo-EM structure of the Mxen ESX-5 system revealed substantial flexibility of the periplasmic domains and the cytosolic domains. The flexibility observed could be a fundamental feature of the system in order to adjust to effector proteins of different sizes. Nevertheless, substrates and co-factors could stabilize the system or could induce conformational changes. This section will assess the flexibility and will give insights into additional components that might be crucial to stabilize the system.

3.3.1 Flexibility inside the ESX-5 core complex

In addition to the analysis of the cryo-EM structure, flexibility was assessed by a limited proteolysis analysis using three different proteases. The use of three different proteases ensured the coverage of different cleavage sites leading to reduced bias. Each protease was incubated with purified ESX-5 complex in a high (1:55) and a low (1:550) concentration. Samples were taken at different time points after addition of the proteases.

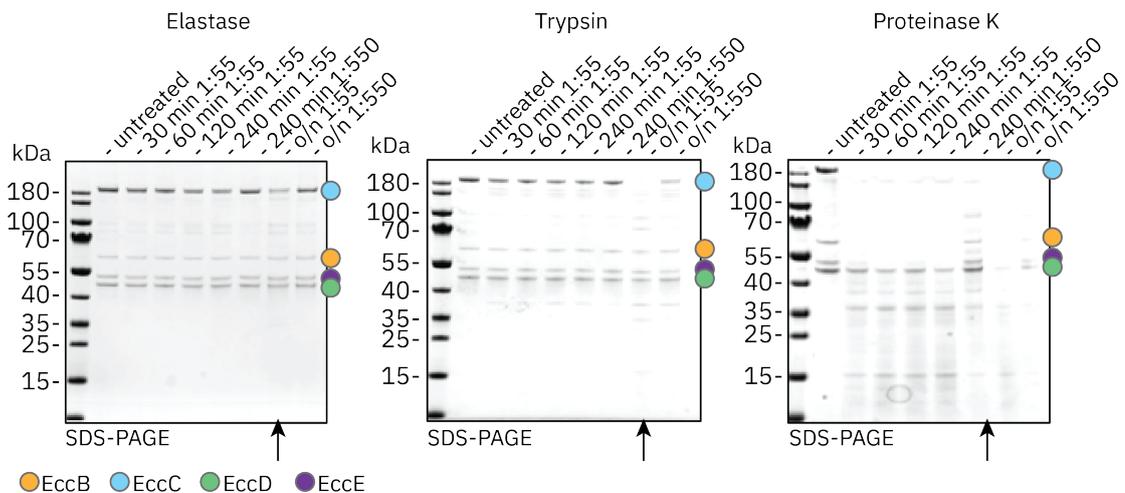


Figure 3.20: Limited proteolysis analysis of ESX-5 The analysis was carried out with three different proteases: Elastase, Trypsin and ProteinaseK. SDS-PAGEs are shown of ESX-5 complex samples that were taken at different time points after incubation with the proteases. The migration height of the individual intact proteins are indicated with colored circles.

Assessment of these samples on SDS-PAGE, shown in Figure 3.20, confirmed a high accessibility of EccC₅ by all proteases. In all samples, the EccC₅ band was the first or, in the case of Elastase, even the only band that showed a reduced intensity after addition of the proteases. Although EccB₅ showed higher stability compared to EccC₅, the intensity of the EccB₅ band decreased after overnight (o/n) treatment with trypsin and vanished after overnight treat-

ment with a low concentration of ProteinaseK. Also EccE₅ levels were affected by incubation with a high concentration of ProteinaseK as the bands disappeared after 30 minutes. Only EccD₅ levels were unaffected by all proteases, except after overnight treatment with a high concentration of ProteinaseK. These observations give rise to the notion that EccD₅ is the least accessible and flexible component.

The conclusions drawn by the limited proteolysis analysis can be explained in the context of the structure, in which the cytosolic ATPase domains of EccC₅ showed long, flexible protrusions into the cytoplasm. Thus, these domains are very accessible to proteases, leading to their fast degradation. Also EccB₅ could not be resolved to high resolution, indicating flexibility of this domain. The fast degradation in the limited proteolysis experiment strengthened this hypothesis. A reason for the degradation of EccE₅ could be its position at the periphery of the complex. However, the compact fold of the cytoplasmic domain of EccE₅ and the binding to EccD₅ makes it less accessible compared to EccB₅ and EccC₅. The long transmembrane domains of EccD₅ and its cytoplasmic domains tightly bound to EccC₅, EccE₅ and EccD₅ render these domains very stable and unaccessible for proteases in solution.

3.3.2 The ESX-5 core complex in a native lipid environment

Within the ESX-5 WT map, elongated density became visible in the cavity formed by the TMHs of EccD₅-1 and EccD₅-2 molecules (Figure 3.7, Subsection 3.1.3). This density could not be assigned to protein. An initial lipidomics analysis confirmed that lipids are still bound to the complex [21]. These observations raised the question if lipids are essential for stability and for the function of the membrane protein complex.

To analyze the impact of the lipid environment on the structure and stability of the complex, ESX-5 needs to be reconstituted in a lipid environment. Traditional MSP-based nanodiscs, however, are limited by a size of up to 17 nm in diameter [108, 227]. Due to the diameter of 27 nm and the 162 TMHs of ESX-5 (Figure 3.2, Figure 3.4) these nanodiscs are not suitable. Also the use of SMALPS (SMA lipid particles) is limited to smaller membrane proteins. New reconstitution methods like SapNPs are modular and can adapt to the size and shape of the membrane protein for reconstitution [86]. However, so far it has not been demonstrated that SapNPs can be applied on proteins with a size comparable to ESX-5.

To test if the ESX-5 system can be successfully reconstituted into SapNPs, DDM solubilized complex, following a strep-tag affinity purification, was incubated for 30 minutes with SapA and lipids in a 1:140:800 ratio. To start with, a lipid mixture (brain lipids) was selected to mimic a heterogenous lipid membrane. Similar to the amphipol reconstitution, detergent was removed by the addition of biobeads.

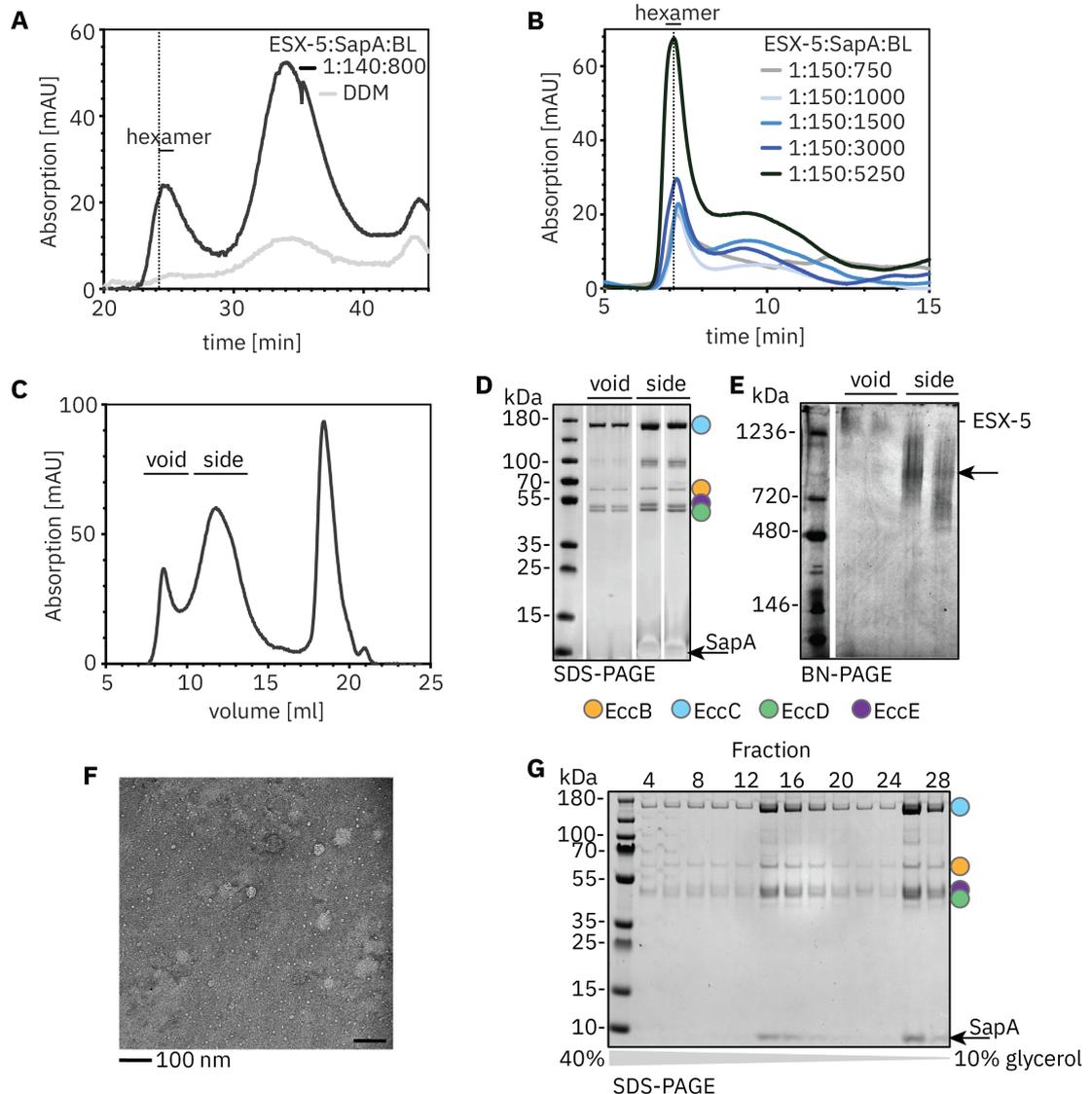


Figure 3.21: Reconstitution of ESX-5 WT into SapNPs (A) SEC profile of ESX-5 reconstituted into SapA and brain lipids (BL) in a 1:140:800 ratio and ESX-5 in DDM after detergent removal by biobeads. The SEC run was performed on a Superose 6 3.2/100 column. The location of the void peak is indicated by a dotted line. Only the reconstituted sample shows significant absorption at 280 nm for the hexamer, indicating successful rescue of the protein complex by SapNPs. (B) SEC profile of reconstituted ESX-5 applying different ESX-5 SapA lipid ratios. A Superose 6 5/150 column was used. (C) SEC profile of an upscaled purification of ESX-5 reconstituted in SapA and brain lipids. A Superose 6 10/300 increase column was applied. (D) SDS-PAGE of void and side peak fractions retrieved from the upscaled purification. The migration height of SapA is indicated by an arrow, the migration height of membrane protein components is indicated by colored circles. (E) BN-PAGE of void and side peak fractions. In the void peak fractions ESX-5 migrates at a size comparable to ESX-5 in amphipols. In the side peak fractions, bands are diffuse, but a main band >720 kDa is visible, which could correspond to a dimeric ESX-5 complex. (F) NS-EM micrograph of reconstituted ESX-5 void peak fractions. High heterogeneity can be observed. (G) SDS-PAGE of ESX-5 reconstituted into SapNPs after 10%-40% glycerol gradient centrifugation.

By comparing the protein complex that was incubated with SapA and lipids with the complex in DDM after detergent removal on a size exclusion column, it became evident that only the reconstituted protein showed the characteristic absorption profile, while in DDM a flat line was observed (Figure 3.21, A).

This was expected, as the removal of detergents by biobeads likely leads to aggregation of membrane proteins, in case other membrane mimetics are missing. These results indicate that the complex can be rescued by SapA and lipids after detergent removal.

Next, to optimize the reconstitution into SapNPs, different complex:saposin:lipid ratios were screened. For this purpose the complex was incubated with different ratios of SapA and lipids and subsequently applied on an analytical size exclusion column (Figure 3.21, B). The signal of the void peak fraction, corresponding to the hexameric complex, was highest for the 1:150:5250 complex:saposin:lipid ratio representing the highest amount of lipids used for reconstitution. However, the peak appeared broadened and slightly shifted to higher molecular weight. For this reason, a complex:saposin:lipid ratio of 1:140:800 was selected for an upscaled purification that allowed the analysis of the samples on SDS-PAGE and BN-PAGE (Figure 3.21, D, E).

The protein eluting in the void peak corresponded, as expected, to the four core components EccB₅, EccC₅, EccD₅ and EccE₅. The side peak could be assigned to the same proteins, although a degraded EccC₅ band became visible at a size of roughly 100 kDa, as similarly observed for amphipol reconstituted ESX-5 and confirmed by MS. A band at 10 kDa corresponding to SapA was, however, absent for the void peak or barely visible for the side peak. On BN-PAGE the void peak fraction migrated at a size of >1236 kDa, comparable to the hexameric amphipol reconstituted ESX-5. The sample from the side peak showed a diffuse band with main intensity at a size of 720 kDa - 900 kDa. This could correspond to a dimeric or trimeric complex.

To check whether SapA was bound to the sample, an MS analysis was performed with the bands extracted from BN-PAGE. The analysis confirmed the presence of SapA in the side peak, whereas in the void peak only traces of SapA were detected. This can be explained by the high number of lipids present in the solution. ESX-5 could have been reconstituted particularly into liposomes or micelles instead of SapNPs. This also explains the increase of the void peak fractions for samples with higher lipid concentration as observed before (Figure 3.1). Analysing the void peak fractions by NS-EM (Figure 3.21, F) confirmed the presence of micelles and liposomes in the sample.

To improve the separation of lipid micelles and ESX-5 reconstituted in SapNPs, size exclusion chromatography was replaced by density gradient centrifugation (Figure 3.21, G). A SapA band was clearly visible in fraction 12, 14, 16 and 18, indicating successful reconstitution.

As the intact ESX-5 complex in amphipols got separated in fraction 9 or 10, main fractions reconstituted in SapNPs seem to correspond to smaller subcomplexes, like dimers or trimers. However, the yields were too low for further characterization by NS-EM. A limiting step is the removal of glycerol after gradient centrifugation that needs to be optimized in the future. It is conceivable, that also earlier fractions, e.g. fraction 10 contained SapA, but due to the low concentration, bands were not visible. Future work will aim to optimize yields and complex:SapA:lipid ratios to obtain hexameric reconstituted samples, but, in addition, also subcomplexes in a lipid environment could expand the knowledge on complex-lipid interactions of the T7SS.

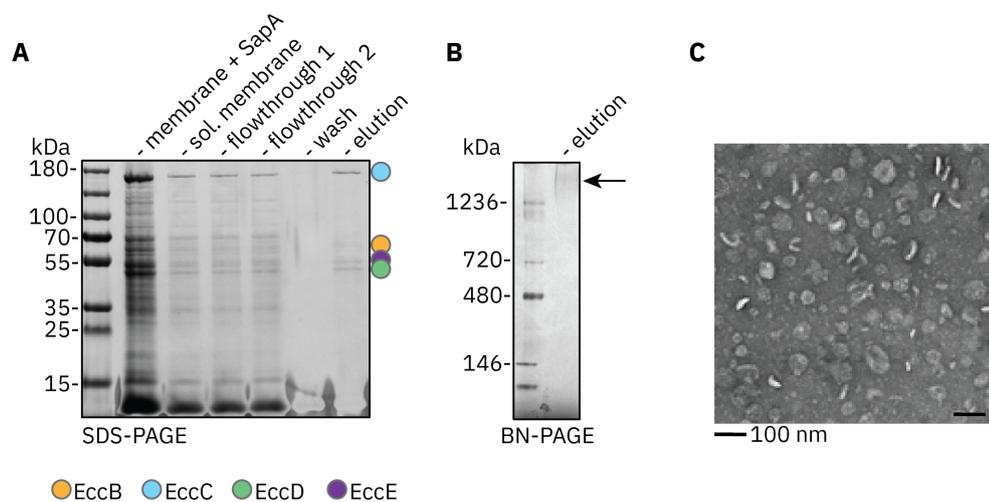


Figure 3.22: Direct extraction of ESX-5 WT by Saposin A (A) SDS-PAGE of the strep-tag affinity purification of ESX-5 after direct extraction by SapA the migration height of membrane protein components is indicated by colored circles. All four protein components could be successfully extracted from the membrane. (B) BN-PAGE of SapA extracted ESX-5, a faint blurred band at a size >1236 kDa is visible indicating the extraction of hexameric ESX-5 (C) NS-EM micrograph of ESX-5 directly extracted with SapA. High heterogeneity is observed in the sample indicating the presence of lipids within the sample.

To circumvent the addition of excess lipids at the outset, an alternative protocol for reconstitution was established. It was reported before that by addition of SapA to purified membranes, membrane proteins can be extracted directly from the lipid bilayer without addition of detergents [149]. Beside the advantage that excess lipids can be removed during affinity purification, this method enables the extraction of the protein directly from the membrane of the expression host. In the case of the Mxen ESX-5 system, expressed in Msmeg, the complex could get reconstituted in mycobacterial lipids that are more similar to the Mxen lipid environment compared to the brain lipid mixture used in the traditional reconstitution approach. In addition, the method allows to skip the addition of detergent increasing the chance to preserve structure and function of the protein.

It was tested, whether this approach could be also applied for the purification of ESX-5. After the incubation of membranes with SapA overnight, strep-tag affinity purification was carried out. The SDS-PAGE of the eluted fraction revealed that all four core protein components could be pulled down (3.22, A). Also a faint SapA band became visible, indicating a successful reconstitution of the protein complex. However, the obtained yields were very low. Therefore the quality of the sample was analyzed on BN-PAGE and NS-EM before size exclusion (3.22 B, C). On BN-PAGE the sample migrated at a height >1236 kDa similar to amphipol reconstituted ESX-5, although the band appeared more diffuse. This is not surprising given the fact that no size exclusion chromatography step was performed prior to the analysis. On NS-EM, liposomes, micelles and disc shaped particles of different sizes were observed, suggesting a high remaining concentration of lipids. It is conceivable that gradient centrifugation could be used to get a suitable separation of protein and lipids. However, yields need to be increased to obtain a homogeneous protein solution suitable for further analysis.

Despite further optimization that is needed to reduce the heterogeneity caused by high amounts of lipids, it is remarkable that the ESX-5 core and subcomplexes can be reconstituted by SapNPs, which is indicated by SapA coeluting with the protein complex during size exclusion chromatography and gradient centrifugation. Given the fact that each protomer consists of 27 TMHs (see Subsection 3.1.3), even the reconstitution of single protomeric units alone would highlight the versatility of SapNPs, not even to mention dimeric, trimeric or even hexameric protomers. 56 TMHs has been so far the maximum number of TMHs that could be reconstituted [165].

Once higher yields and higher homogeneity is achieved for either the traditional or the direct extraction technique, structure and activity should be compared to the amphipol reconstituted sample to assess the role of a lipid environment on structure and function. Such a structure could drastically improve the understanding of complex-lipid interactions and could give further insights into T7SS mechanisms.

3.3.3 Binding of ATP γ S to the ESX-5 core complex

The apo cryo-EM structure as well as the limited proteolysis analysis presented in this thesis revealed high flexibility of the cytoplasmic EccC₅ ATPase domains. The three consecutive ATPase domains belong to the family of P-loop ATPases in which the Walker A motif is responsible for the binding and coordination of ATP [205]. The DUF domain follows the typical fold of P-loop ATPases as well, as described in Subsection 3.1.3. However, its Walker A motif is highly degenerated or even hardly recognisable, making it likely that the DUF domain is incapable of binding ATP as discussed in Subsection 3.1.3 (Figure 3.23).

During the purification process, bound nucleotides from the expression host could be stripped

from the complex. As a result, it can be assumed that the majority of particles isolated are not binding nucleotides. In consequence, due to the missing capacity of binding ATP, the DUF domain is unaffected by the absence of nucleotides, whereas in the three consecutive ATPase domains, lack of nucleotides could induce flexibility. In contrast, the binding of nucleotides could lock the three consecutive ATPase domains in a certain conformation and reduce flexibility in between the individual domains. This could allow the binding of substrates, which could then lead to a multimerization and activation of the ATPase domains as described by Rosenberg *et al.* [205].

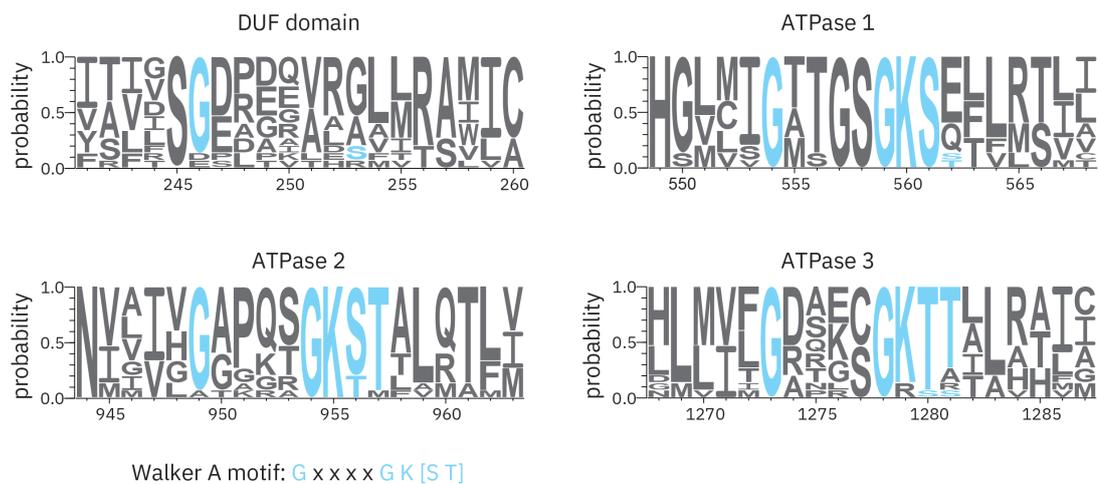


Figure 3.23: Sequence logo of Walker A motifs present in EccC Sequence logos were calculated based on a sequence alignment of all ESX-systems from Mtb, Mmar, Mabs, Msmeg and Mxen with WebLogo [60]. Residues corresponding to the Walker A motif are colored in blue. The ATP-binding motif of the DUF domain is highly degenerated.

A surface plasmon resonance experiment confirmed the binding of the non-hydrolyzable ATP analogue ATP γ S to the three consecutive ATPase domains. For that purpose, a construct of the EccC₅ C-terminus, including the three ATPase domains (provided by Katherine Beckham, Wilmanns group, EMBL Hamburg) was immobilized on a gold surface via capture-coupling to a carboxymethylated dextran matrix (CM5). The sensorgram showed a successful annealing and coupling of the construct to the sensor surface (Figure 3.24, A).

Different concentrations of ATP γ S were injected to the flow cell and for each concentration clear association and disassociation was observed (Figure 3.24, B). The sensorgrams show that fast on and off rates are superimposed by lower rates, indicating different binding rates for the ATPase domains. In these cases, kinetic fitting is not the method of choice. Instead, the association endpoints were analyzed as a function of concentration (Figure 3.24, C). The resulting dose-response curve shows a saturation at a concentration of 0.78 μ M, but an increasing signal starting at 3.13 μ M, indicates different (at least 2) binding rates for the ATPase domains. The dose-response curve was, thus, fitted assuming two-site specific binding

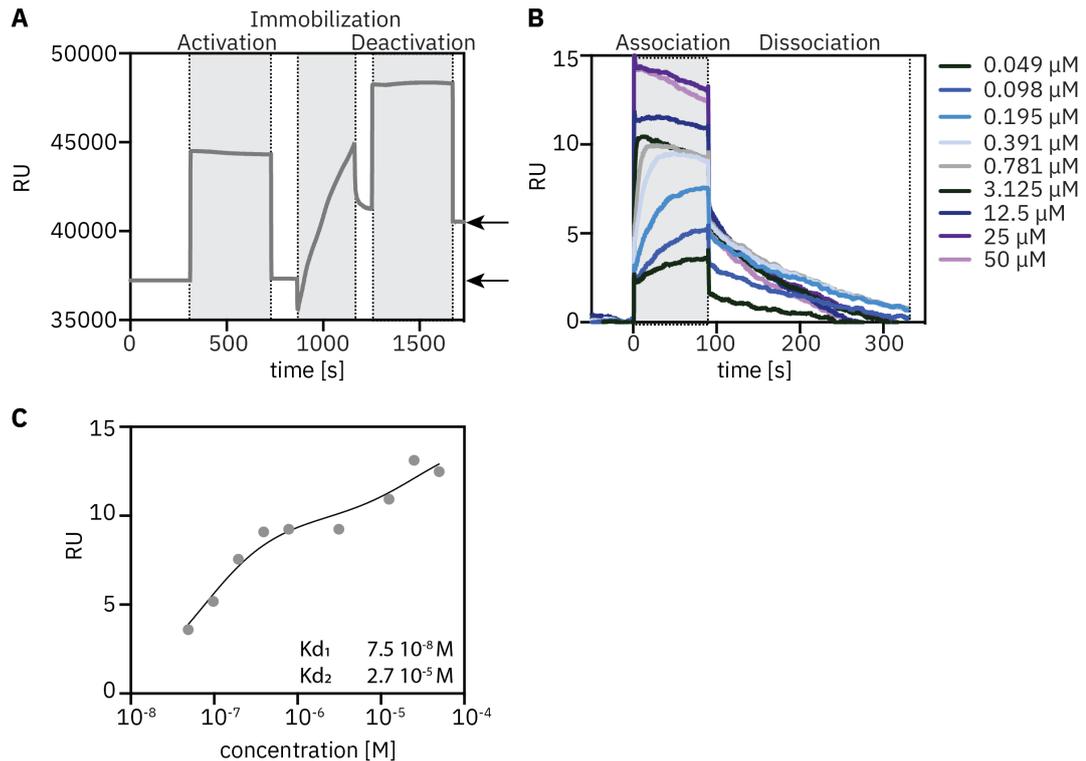


Figure 3.24: Binding of ATP γ S to the EccC₅ ATPase domains The response is measured in resonance units (RU) (A) Immobilization of the EccC₅ ATPase domains on the CM5-Chip. (B) SPR sensorgram of ATP γ S binding to the EccC₅ ATPase domains. ATP γ S was injected at different concentrations. (C) Dose-response curve of ATP γ S binding to EccC₅. The points were fitted assuming two-site specific binding.

using Graphpad Prism. The R^2 of 0.96 was superior for two-site specific binding compared to 0.89 and 0.95 obtained by fitting assuming one-site specific binding or one-site specific and nonspecific binding, although not significant. The resulting two association constants (K_d) accounted for 75 nM and 27 μ M. It is conceivable that all three ATPase domains bind ATP or ATP γ S with different affinities. A third binding affinity, however, cannot be retrieved reliably from the resulting data. To obtain more concise values for ATP affinity of the individual domains, all three domains should be analyzed separately. This would also improve the assignments of binding constants to the three ATPases. In the context of the translocation process, it seems reasonable that the ATP hydrolysis at ATPase3 is the time-limiting step to avoid accumulation of substrates close to the pore. Hence, the affinity for ATP or ATP analogues should be highest for ATPase1 and lowest for ATPase3. This hypothesis needs to be confirmed in the future.

To judge the impact of bound nucleotides on the structure of Mxen ESX-5, the membrane protein complex was purified in presence of 100 μ M ATP γ S during all purification steps. After quality control, the sample was applied on a cryo-EM grid and plunge-frozen. 12,789 movies were collected and processed in C6 and C1 symmetry as shown in Figure 5.4. Within the

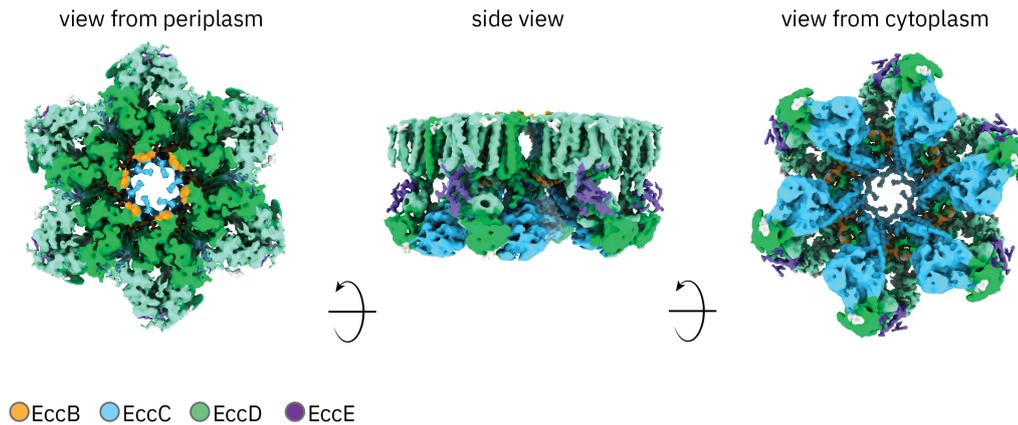


Figure 3.25: Cryo-EM maps of the ESX-5 complex in presence of $\text{ATP}_{\gamma}\text{S}$ The high-resolution map of ESX-5 is shown from the periplasm, the cytoplasm and as side view. The map is colored according to the scheme used in the ESX-5 model.

high-resolution part of the structure, the resulting maps resembled those of the WT map. The cross-correlation between both maps accounted for >0.8 . Further on, the individual protein domains exhibited the same orientation and position and also the pore showed the gated, secretion competent conformation as observed for the apo structure. Only density for EccE_5 was less well defined, which could be a result of the lower particle number that was used for the 3D reconstruction (Figure 3.25).

When lowering the threshold of the C6 and C1 map and filtering the maps with a gaussian filter, additional low-resolution density became visible in the cytoplasmic part, presumably corresponding to the ATPase domains of EccC_5 (Figure 3.26, A). Fitting the crystal structure of the three ATPase domains, obtained by Katherine Beckham (Wilmanns group, EMBL Hamburg), into the density confirmed this hypothesis, since the EM density could accommodate the complete model (Figure 3.26, B). However, instead of one conformation, two conformations were observed within the maps, which could not be separated - despite intensive effort in reducing particle heterogeneity in the dataset by 2D and 3D classifications, 3D variability analysis [193], focused refinements and heterogeneous refinements. In the first conformation, the EccC_5 domains are in a widened upward position. In this position, the individual ATPase domains are located in far distance from each other without any notable contacts. In the second conformation, the ATPase domains are in a downward position, forming a widened translocation tunnel. In this conformation, the ATPase domains are in closer proximity, especially the third ATPase domain that limits the opening of the tunnel to a diameter of around 7.8 nm. The fit of the crystal structure to the downward conformation, however, did not reveal contacts in between the different ATPase domains. A multimerization of the ATPase domains is crucial for ATP processing, as ATP can only be hydrolyzed, when an arginine residue of the ATPase domain (R-finger) completes the active site of the neighbouring ATPase [205]. It is not evident, whether the missing contact is due

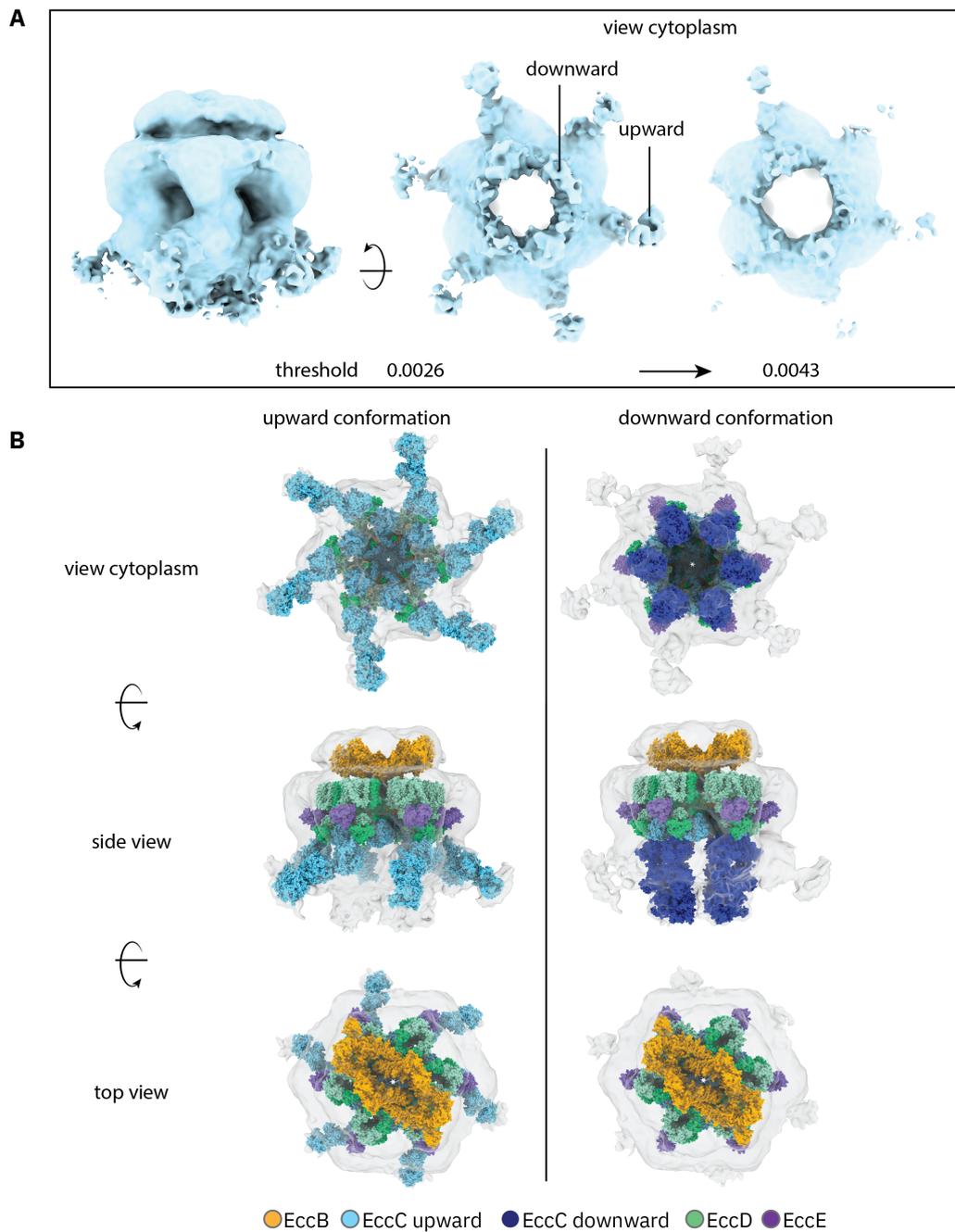


Figure 3.26: Low resolution cryo-EM density of the ESX-5 complex in presence of $\text{ATP}\gamma\text{S}$. (A) At low threshold values, low resolution features for the three ATPase domains of Ecc₅ become visible. Two conformations were observed that could not be separated from each other during cryo-EM processing: the upward and the downward conformation (B) The low resolution map allowed in addition to the high-resolution model the fitting of an Ecc₅ crystal structure from Mxen into the density, which is shown from the cytoplasm, from the side and from the periplasm.

to the fit in the low resolution map or if the downward conformation corresponds still to a pre-secretion state. It is, however, possible that the addition of substrate could lead to a further restriction of the diameter that would induce the final multimerization crucial for the

transport of proteins, as it was similarly suggested by Rosenberg *et al.* [205].

The results demonstrate that ATP analogues are, indeed, able to stabilize EccC₅, although the flexibility and heterogeneity observed by EM remains high. Especially the presence of two conformations within the map is intriguing. Due to the low resolution, no final conclusions can be drawn on how these two different conformations are induced. One possibility could be the presence of different ATP binding states within the maps. ATP γ S was used in a concentration five times higher than the binding constant throughout the purification, but sample handling or slow hydrolysis and or degradation of ATP γ S could have lowered the actual nucleotide concentration in the sample. Binding to substrates could have initialized a different conformation as well. To draw final conclusions on the mechanism, higher resolution maps of the cytoplasmic part of the complex are required. These maps could give insights into orientation and ATP binding state of the three ATPases.

4 Discussion

The results provided in this thesis give first insights into the molecular architecture of a T7SS, its hexameric assembly and the central secretion pore. The structure of the Mxen ESX-5 complex revealed the core machinery, built by six identical protomeric units, each composed of EccB₅:EccC₅:EccD₅ and EccE₅ in a 1:1:2:1 stoichiometry (Subsection 3.1.2, Subsection 3.1.3). Key mediators of the hexameric assembly were identified in the periplasm and the membrane. Within the periplasm, the six periplasmic domains of EccB₅ form a dimer of trimer arrangement, composed of three different dimeric interfaces - the V-shaped, parallel and peripheral dimer (Subsection 3.1.4, Figure 3.8). The resulting arrangement revealed C₂ symmetry, which differs from the C₆ symmetry in the membrane section. In the membrane section, main contacts between protomeric units were observed between the N-terminus of EccB₅ and an extended loop of EccD₅-1 (Figure 3.9) as well as by an unusual domain swap of the first pore helix, TMH1 of EccC₅ (Figure 3.10). The pore helices are enclosed in a basket-like structure formed by the N-terminal helix of EccB₅ and the TMH of EccB₅ (Figure 3.13). It is hypothesized that small changes in the angle between both helices could lead to pore opening and closing. Within the central secretion pore, EccC₅ TMHs 1 and 2 show increased flexibility compared to the remaining transmembrane part (Figure 3.11, Figure 3.12, A, B). Especially TMH2, which forms the inner face of the helix, could not be modeled reliably and was, thus, presented as an ensemble model. Analyzing the ensemble model drew the attention to restriction sites formed by bulky side chains like phenylalanine that limit the pore diameter to a size that would not allow transport of folded heterodimers across the membrane (Figure 3.12, C). It is suggested that these residues could play an important role in a pore gating mechanism. The observed flexibility could be introduced by a conserved proline residue in the center of the helix, which introduces a helix break and therefore induces kink-swivel movements (Figure 3.11, Figure 3.12, A, B).

To test the hypotheses made, several mutants were designed that impact the integrity of the pore dynamics, pore gating or the scaffold of the pore helices formed by the EccB₅ basket. All of these mutants were proven to reduce secretion efficiency significantly, emphasizing that the highlighted regions, indeed, play an important role within the translocation cycle (Figure 3.15, Figure 3.16). Especially the proline residue in TMH2 was shown to be essential for secretion. A cryo-EM structure of a P73 mutant demonstrated that mutation of this residue leads to a rigidification of the pore helix, giving rise to the notion that a certain degree of plasticity or flexibility is needed in the course of secretion to transport a range of different

substrates across the mycobacterial membrane (Figure 3.18).

Plasticity or rather flexibility was also observed in other regions of the secretion machinery, especially on the periplasmic and distal cytosolic side. For these sections integrative ensemble models were built (Figure 3.3). Co-factors like substrates, lipids or nucleotides were hypothesized to transiently stabilize certain conformations. For this reason, alternative purification protocols were established, that will allow the characterization of the complex in a lipid environment in the future (Subsection 3.3.2). In addition, the structure was analyzed in presence of ATP γ S. Indeed, within the resulting map, low resolution density of the distal cytosolic side became visible, indicating slight stabilization of the three consecutive ATPase domains of EccC₅ (Figure 3.26). Intriguingly, not a single conformation, but two conformations were observed, which could correspond to different ATP bound states. Further stabilization is, however, required for structural studies to draw detailed conclusions on the induction of both states. An intrinsic flexibility in the relative positioning of the individual ATPase domains, could, however, be a key to mediate huge conformational changes, required for the transport of folded heterodimers of varied sizes.

During the course of this PhD, three other structural studies on the inner membrane complex of ESX systems were published. In 2019 Famelis *et al.* and Poweleit *et al.* [80, 192] both obtained cryo-EM maps of the Msmeg ESX-3 system. Intriguingly, both structures showed the complex in a dimeric state. The protomeric architecture of the ESX-3 systems clearly resembled the protomeric arrangement in the hexameric ESX-5 system: Each protomer was built in a 1:1:2:1 stoichiometry of EccB₃, EccC₃, EccD₃ and EccE₃ and in the cytosol, the two ubiquitin like domains interacted with the DUF domain and the EccE₃ cytoplasmic domain. As in the ESX-5 maps, the periplasmic and cytosolic side was less well resolved, confirming the flexible nature of these regions as common feature of type VII secretion. Despite the lower resolution in the periplasm, it became evident, that the periplasmic domain of EccB₃ forms a V-shaped dimer as observed in ESX-5. The absence of a central secretion pore in both dimeric structures suggests, that the dimeric structure is a building block of the hexameric T7SS. ESX systems could preassemble in a dimeric form before hexamerisation or could disassemble into defined dimers after translocation.

Another interesting structure was published by Bunduc *et al.* [36], who obtained a cryo-EM map of Mtb ESX-5 in its hexameric form in presence of the fifth membrane component MycP₅. Three MycP₅ molecules are anchored in the membrane and bind with their protease domain on top of three EccB₅ V-shaped dimers, with the active site facing the central pore of the complex. The EccB₅ V-shaped dimers further multimerize in a hexamer with peripheral dimer interactions as similarly observed in the Mxen ESX-5 (Figure 4.2, B, D). Interestingly, in presence of MycP₅ the resolution of the periplasmic region is highly improved implying a stabilizing role of MycP₅ on the periplasmic part. While the overall architecture of

the transmembrane domain and the upper cytoplasmic side resembled the architecture of the hexameric Mxen ESX-5 and the dimeric MsmeG ESX-3, a major difference can be observed in the central secretion pore that appears to be closed completely in presence of MycP₅. Thus, MycP₅ could induce or stabilize a closed state of the secretion pore (see Figure 4.2, C). While MycP₅ induces conformational changes in the secretion pore, the cytoplasmic domain of EccC₅ remains very flexible. Low resolution density in the cytoplasm of the Mtb ESX-5 map confirms the presence of an upward and downward state of the ATPase domains in presence of ATP analogues, here ADP-AIF₃ [36], as similarly observed for the Mxen ESX-5 complex (see Figure 3.26).

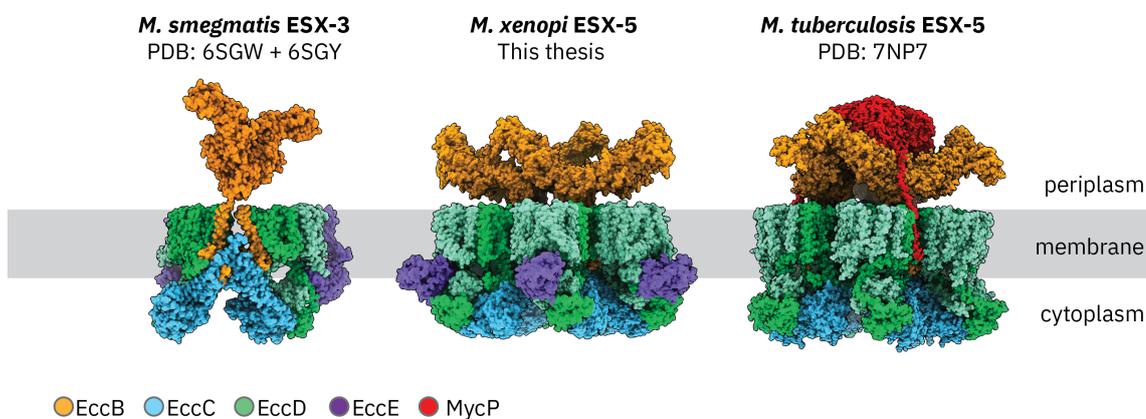


Figure 4.1: Available structures of T7SSs. Atomic models are shown for the dimeric complexes of MsmeG ESX-3, the hexameric complex of Mxen ESX-5 and the hexameric structure of Mtb ESX-5 in presence of MycP₅.

The availability of structures in three different states (shown in Figure 4.1) allows the comparison of their architecture. Common to all structures is the protomeric arrangement and the presence of three segments, including (1) the periplasmic domain that tends to be flexible in the absence of MycP, (2) the transmembrane and upper cytosolic part that is most rigid and exhibits high similarity in between all structures and (3) the distal cytosolic part that shows high intrinsic flexibility.

While the basic architecture is comparable in most regions, the pore basket and the pore conformation vary in between the different structures. In the dimeric models, the pore helices are absent [80, 192]. Most likely, the two EccC helices require the scaffold formed by the hexameric EccB basket, comprised by the N-terminal helix and the TMH of EccB in order to reduce the dynamic space EccC helices can adopt (compare Figure 3.13). The absence of the helices in the dimer, however, confirms the dynamic nature of the helices, as observed in the Mxen ESX-5 complex. Dynamics are also required to convert the closed conformation of the transmembrane helices observed in the Mtb ESX-5 structure into the gated structure observed for Mxen ESX-5 (Figure 4.2, C). In the Mtb structure, the pore is formed by a trimer

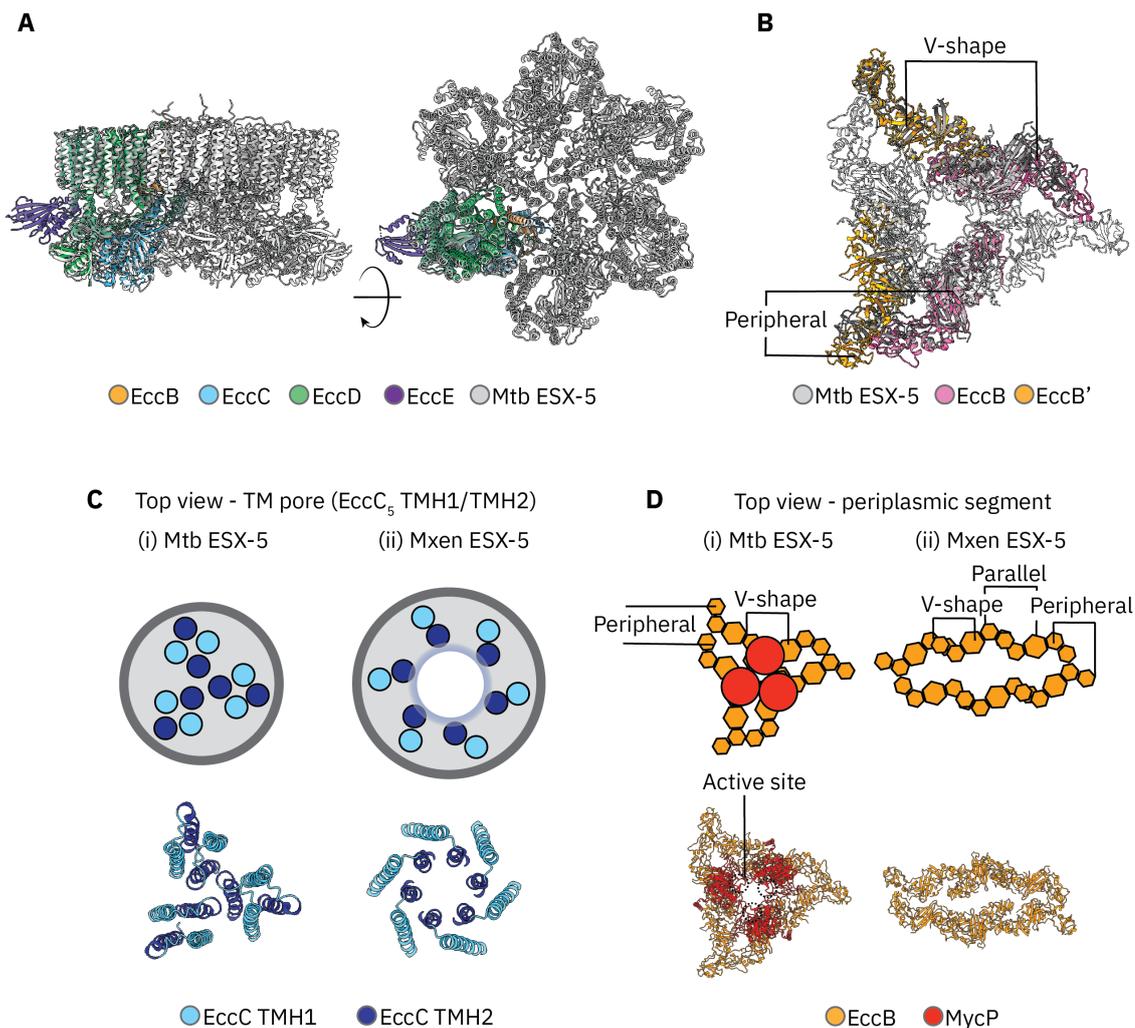


Figure 4.2: Comparison of the *Mycobacterium xenopi* ESX-5 and the *Mycobacterium tuberculosis* ESX-5 structure (A) The protomeric structure of Mxen ESX-5 was fitted to the hexameric structure of Mtb ESX-5 (white). The fit revealed a comparable protomeric arrangement of Mtb ESX-5 compared to Mxen ESX-5. (B) The peripheral and the V-shape Mxen EccB₅ dimers were fitted to the periplasmic EccB₅ arrangement of Mtb shown in white. While the two fitted dimer pairs were identified in the Mtb ESX-5 arrangement, the parallel dimer pair is absent. (C) Schematic view on the transmembrane pore of Mtb ESX-5 and Mxen ESX-5 and corresponding structure. While the pore is completely closed in the Mtb ESX-5 structure, the pore in the Mxen ESX-5 complex is in a gated conformation. (D) Schematic view on the periplasmic segment of ESX-5 from Mtb and ESX-5 from Mxen and the corresponding structures. In presence of MycP the periplasmic domain adapts a 3-fold symmetry formed by two dimeric interfaces compared to the 2-fold symmetry in absence of MycP formed by three dimeric interfaces, as indicated in the schematics. The active site of Mtb MycP₅ is indicated by a dashed circle. The figure was modified from Beckham *et al.* [21], Figure 4.

of four-helix bundles, in which a trimer of EccC₅ TMH2 forms the inner face of the pore, while on the other side TMH2 also corresponds to the outermost helix in each 4-helix bundle.

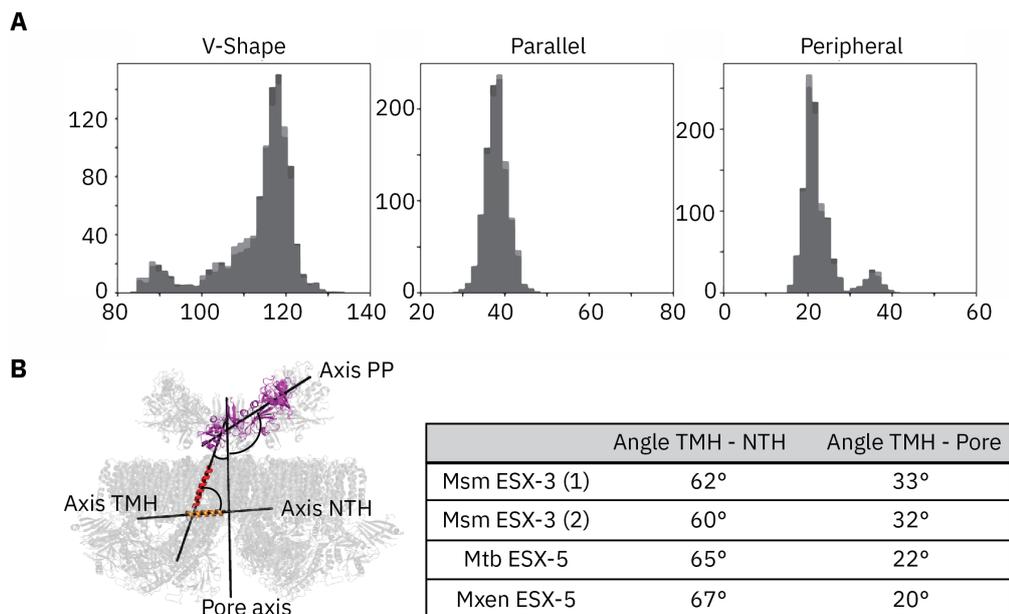


Figure 4.3: Angle analysis of EccB. (A) Analysis of the periplasmic domain of the Mxen ESX-5 ensemble model. Angles of the different dimer pairs are shown. The distribution of angles differs significantly between the different dimers. (B) Left visualization of axis used for the calculation of angles of the four different structures. Right, table showing the approximate angles calculated between the EccB transmembrane helix and the N-terminal helix and in between the transmembrane helix and the pore axis. The figure was modified from [21], Figure S7.

A conformational change could be triggered by the dissociation of MycP₅ and the following rearrangement of EccB₅, which converts the C3 symmetric arrangement of Mtb built by only peripheral and V-shape dimer pairs of EccB₅ to the C2 symmetric arrangement of Mxen EccB₅ (Figure 4.2, B, D). That EccB can undergo huge conformational changes, becomes evident when analyzing the angles between the different periplasmic domains or the angle between the transmembrane domain and the periplasmic domain of EccB₅ in the Mxen ESX-5 system alone (Figure 4.3, A). The ensemble model shows huge variation in the angles between different dimers ranging from around 120° for the V-shape to around 40° of the parallel dimer and around 20° for the peripheral dimer. Also the hinge of the periplasmic domains was shown to be flexible with angles ranging from around 53.4° to 79.0°, as discussed in Subsection 3.1.4. In the context of the other published structures, the dynamic range of EccB becomes even more obvious. In the ESX-3 dimer structure, for example, the angle of the EccB₃ V-shaped dimer accounts only for 85°, which highly differs from the larger angles (113°) obtained in the hexameric structures. In addition, in the Mtb ESX-5 system, the parallel dimer pair is missing from the periplasmic arrangement. Differences in EccB are also transmitted to the transmembrane domain, or more precisely, the EccB pore basket (Figure 4.3, B; compare to Figure 3.13). Although the N-terminal helix is situated parallel to the

membrane and is followed by a transmembrane helix in all structures available, the angles between both helices differ and as a result, also the angles of the transmembrane helix to the pore axis differ, especially between dimeric and hexameric structures. The difference in angles between dimeric and hexameric structures accounts for up to 13° in case of the pore axis and the transmembrane helix. The wide angle observed in the ESX-3 structure could prevent the formation of a hexameric arrangement, as a resulting EccB basket would presumably not offer enough space to accommodate the EccC transmembrane helices. The hinge between the N-terminal helix and the transmembrane helix of EccB could thus play an important role in the assembly process of the hexameric ESX system. The angle difference between both hexameric structures was only subtle, but even small changes in the angles could lead to a significant rearrangement of the pore helices as seen in both structures. This mechanism could be induced by binding or dissociation of MycP in the periplasm.

A question that the experimentally available structures could not answer so far is, how substrate specificity is achieved and which factors determine substrate hierarchy. In the Mxen ESX-5 locus, three different substrate pairs are encoded: EsxM and EsxN, PE19.1 and PPE26.1 and another PE19/PPE26 pair (referred to as PE19.2 and PPE26.2 in this thesis). A recent preprint pointed out that the secretion of PE/PPE and EsxMN proteins in this locus is co-dependent [37]. The co-dependence could be a result of a multiple recognition event in the cytoplasm or due to the formation of specific pores in the periplasm and the mycomembrane by the secreted substrates. Even a combination of both possibilities seems reasonable.

However, it was not discussed yet, whether different substrate classes can multimerize prior to secretion. Interestingly, when running AlphaFoldMultimer [131, 79] for the PE/PPE substrate pairs, respectively, with EsxMN, the resulting prediction shows the interaction of EsxMN and the PE/PPE proteins in a head to tail fashion in which an intermolecular β -sheet is formed by the N-terminal ends of EsxN and PE19.2 (Figure 4.4, A). This β -sheet is further flanked by the conserved WxG motifs of EsxM and PPE26.2. Although H-bonds have been observed between the newly formed β -sheet and the backbone of the WxG motif, specific interactions of the tryptophan side chains could not be retrieved from the predicted model.

The PAE plot shows high confidence for the relative positioning of the helix bundles of EsxMN and PE/PPE. Only for the C-terminal domain of the PPE protein high uncertainty of the localization is observed from the PAE plot. This is not surprising as the C-terminal domain was predicted with low pLDDT score (<50). The filament formed by the four proteins reaches a length of about 19.5 nm.

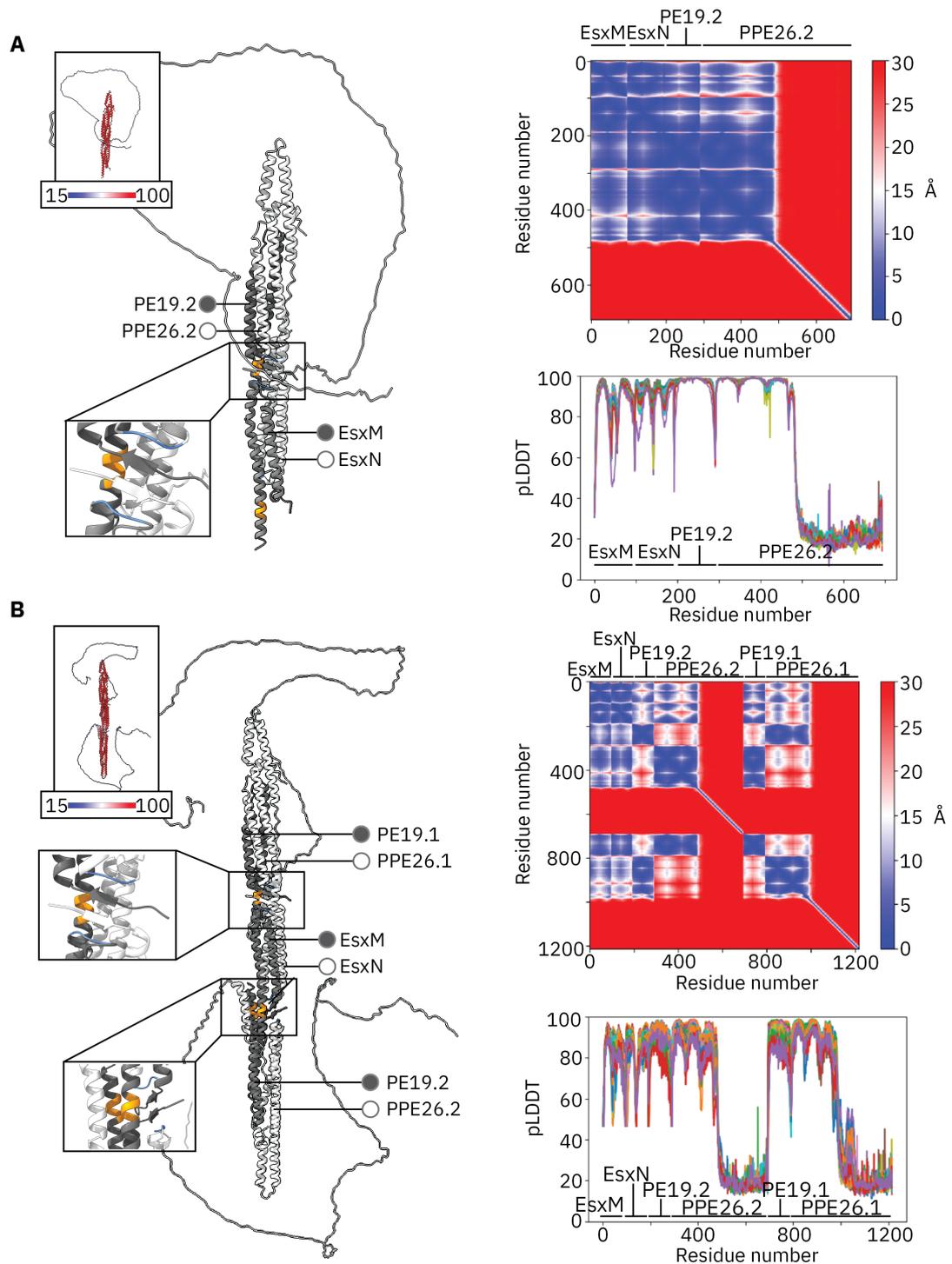


Figure 4.4: Prediction of ESX-5 substrate complexes (A) Alphafold2 models of a Mxen EsxM/ EsxN/ PE19.2/ PPE26.2 tetramer. The pLDDT score is plotted on the structure in the upper left corner. The inset highlights the intermolecular β -sheet formed by the two dimer pairs, which is flanked by the WxG motif, colored in blue. The PAE plot of the prediction is shown in the upper right corner and the PLDDT plot is shown below. (B) Prediction of a hexamer composed of EsxM and EsxN, PE19.1, PPE26.1, PE19.2 and PPE26.2. The insets show the respective contact sites of EsxM and EsxN with the PE and PPE proteins. PAE and pLDDT plots are shown on the right. The pLDDT plot shows the result of six output models.

Even more intriguing than the EsxMN and PE/PPE tetramer is the prediction of a hexamer built by two PE/PPE pairs and EsxMN. Again, C-terminal ends are predicted with low confidence, but the relative positioning of the helical domains seems reliable based on the PAE plot (Figure 4.4, B). As for the tetramer, the helical bundles form an elongated structure with a length of about 29 nm.

The interaction between EsxMN and the first PE/PPE pair is unaffected compared to the tetramer. The second PE/PPE pair is connected on the opposite site of EsxMN, however the contacts are very similar. Here, EsxM and PE19 form the intermolecular β -sheet that is again flanked by WxG motifs of EsxN and PPE26. In this arrangement EsxMN serves as connector between the two PE/PPE bundles.

The hydrophobic tip, that was shown to bind to EspG, is exposed at both ends on the filamentous arrangement of the tetramer or hexamer. This tip was previously shown to be a determinant for system specificity. When swapping the hydrophobic tip of an ESX-1 PE/PPE pair with the hydrophobic tip of an ESX-5 PE/PPE pair, the ESX-1 related PE/PPE pair was rerouted through ESX-5 in Mmar [64]. The additional observation of the authors that the ESX-1 substrate EsxAB was similarly rerouted through ESX-5 without any additional modification was particularly striking [64]. The results can be easily explained by the tetrameric or hexameric preassembly predicted, as the PE/PPE substrate pair piggybacks the Esx substrate pair. The hydrophobic tip could interact with the linker 2 domain of EccC, which was shown to be of importance for system specific secretion in a similar domain swap experiment [35].

The dimensions of the PE/PPE substrate pair further suggest, PE/PPE could interact not only with the linker 2 domain but also with ATPase3 via its YxxxD/E motif, which is located opposite of the hydrophobic tip. It was shown before that binding of the YxxxD/E motif to the third nucleotide binding domain of EccC could induce multimerization of the ATPase domains [205]. The YxxxD/E motif of the Esx pair could be important in a later step, when induced by the ATP hydrolysis in one or some of the consecutive ATPase domains the substrate filament is transported closer to the central pore. The YxxxD/E motif of EsxMN could bind to ATPase3 and maintain the hexamerized state of the ATPase domains, as similarly suggested by Rosenberg *et al.* [205]. Although multimerization of the substrates could explain substrate dependencies, experimental evidence of heterotetramers, heterohexamers or even higher multimeric states is missing so far.

Taking the results from this thesis and further literature together, it is possible to discuss a new model for substrate translocation across the mycobacterial membrane, which is sketched in Figure 4.5. It is conceivable that substrate transport across the inner membrane occurs through a central pore, as it is the case for other structurally characterized secretion systems [198, 107]. A dimer, as observed by Famelis *et al.* and Poweleit *et al.* [80, 192], could, thus,

represent a preassembly of the complex (Figure 4.5, step 1). Initial hexamerization could be induced by binding of MycP and a rearrangement and stabilization of the periplasmic domain (Figure 4.5, step 2). The binding may induce a widening of the EccB V-shaped dimer, as the angle of the V-shaped dimer is considerably bigger in both hexameric structures compared to the dimeric structures as discussed before in this chapter. In a next step, MycP could mediate contacts in between three different MycP bound dimers. Main interactions could include loop1, the N-terminus and loop3 of MycP as reported by [36]. Conformational changes of the periplasmic EccB domains could enhance the formation of a hexamer. A conformational rearrangement in EccB is further necessary to widen the hinge between the N-terminal and transmembrane domain (Figure 4.3). This would allow the accommodation of pore helices in the EccB basket (compare Figure 3.13). Binding of ATP to the three cytosolic ATPase domains could then stabilize the flexible domains and induce an upward conformation of the EccC domain as observed in this thesis (Figure 3.26, B) or by Bunduc *et al.* [36] (Figure 4.5, step 3).

In the cytosol, PE/PPE proteins, bound to EspG, could interact with the TPR domain of EccA, which induces a closed conformation of EccA, as suggested by Crosskey *et al.* [61] (Figure 4.5, step a). Based on the Alphafold2 prediction mentioned above (Figure 4.4), EsxM and EsxN could bind to a PE/PPE pair (Figure 4.4, A; Figure 4.5, step b). The order of events is not clear. Binding could either occur before or after the interaction with EccA. In a next step, the EccA, EspG, PE, PPE complex could be delivered to EccC in the upward conformation (Figure 4.5, step c and 3). Interactions of substrates with the membrane complex could trigger a conformational change in EccA that enables the hydrolysis of ATP and triggers the dissociation of the substrates from EspG (Figure 4.5, step 4) [61]. These steps are very hypothetical as there is no experimental evidence available so far. However, similar dissociation mechanisms have been observed for EccA homologues such as ClpX and CbbX [61, 164, 163]. Given the high affinity of EspG to the PE/PPE substrate pair of 50 nM [137], discussions about a dissociase that provides energy for the disassembly seem reasonable. In the future, experimental evidence is required to prove this hypothesis.

The initial interaction of the substrates with the complex could be mediated by the YxxxD/E motif (Figure 4.5, step 4). Binding of the YxxxD/E motif was shown previously to induce multimerization of the ATPase domains, which in turn enabled ATP hydrolysis [205]. In addition, the binding could induce the release of the linker2 domain which connects ATPase1 and ATPase2 [205] (Figure 4.5, step 5). The linker2 domain was shown to be involved in species-specific secretion of Esx and PE/PPE proteins [35] and so was the hydrophobic tip of the PPE protein, that interacts with EspG [64]. The dimensions of the PE/PPE dimer suggest that the PE protein could serve as a donor for the YxxxD/E motif while at the same time, the hydrophobic tip of the corresponding PPE protein is recognized by linker2, ensuring species and probably also system specific transport. As the EsxMN dimer piggy-backs on the

PE/PPE protein dimer in this model (Figure 4.4), the recognition of the hydrophobic tip of the PE/PPE protein is sufficient to guide also EsxMN through the respective system. This dependency was already reported in the literature [35, 64] (Figure 4.5, step 5).

The multimerization of the ATPase domains in a downward conformation could cause further conformational changes in the membrane and in the periplasm. The EccB basket could open up further, allowing the pore to adapt an open, but gated conformation (Figure 3.11, Figure 4.3, Figure 3.13). Induced by this rearrangement, MycP could dissociate from the complex allowing the rearrangement of the periplasmic EccB domains into the flexible C2-symmetric arrangement observed in this thesis (Figure 4.2, D; Figure 3.8; Figure 4.5, step 5). It is not known yet, if EccC ATP hydrolysis, needed to transport the substrate through the pore, occurs sequentially or simultaneously in the three consecutive ATPase domains. ATP binding seems to follow a hierarchy, as presented in Subsection 3.3.3. This might be needed to ensure proper guidance of the substrate by ATPase1 to avoid accumulation of substrates within the translocation tunnel formed by all three ATPase domains of EccC in downward conformation (compare Figure 3.26). To secure the integrity of the translocation tunnel after moving the substrate closer to the pore, the YxxxD/E motif of EsxM and EsxN could bind to ATPase3 (Figure 4.5, step 6, Figure 4.4). This could prevent the disassembly of the translocation tunnel and keeps the ATP hydrolysis active.

Within the membrane, bulky side chains, like phenylalanine residues in the case of ESX-5 or methionine residues limit the pore diameter as discussed in Subsection 3.1.6 and 3.2.2 (Figure 3.11, Figure 3.12). This was similarly observed for the type III secretion system, in which the so-called methionine-gasket forms a deformable barrier that prevents uncontrolled efflux or influx of molecules important for bacterial fitness [124, 160]. A dilation of the M-gasket of the type III secretion system allows the secretion of the substrate, giving rise to a dual role of the methionine residues in gating and in adapting a conformational space that mediates substrate translocation. Also SecY transport was suggested to follow this mechanistic principle by using isoleucines as gate keeping residues [182]. This can be similarly proposed for type VII secretion. Once the substrates reach the membrane pore, a small dilation or deformation of the EccC TMH2 leads to a rearrangement of phenylalanine side chains sufficient to allow the passage of substrates (Figure 4.5, step 7). Energy for this dilation could be provided by the ATPase domains. The presence of the conserved proline residue at position 73 in case of Mxen ESX-5 could be the key to endow the complex, the flexibility and plasticity needed for the dilation of the pore, once the substrates enter the transmembrane section as discussed in Subsection 3.1.6 and Section 3.2. The flexibility of the periplasmic EccB domains in absence of MycP could ensure similar dilation of the periplasmic cleft observed in the Mxen ESX-5 structure (Figure 3.3, Figure 3.8).

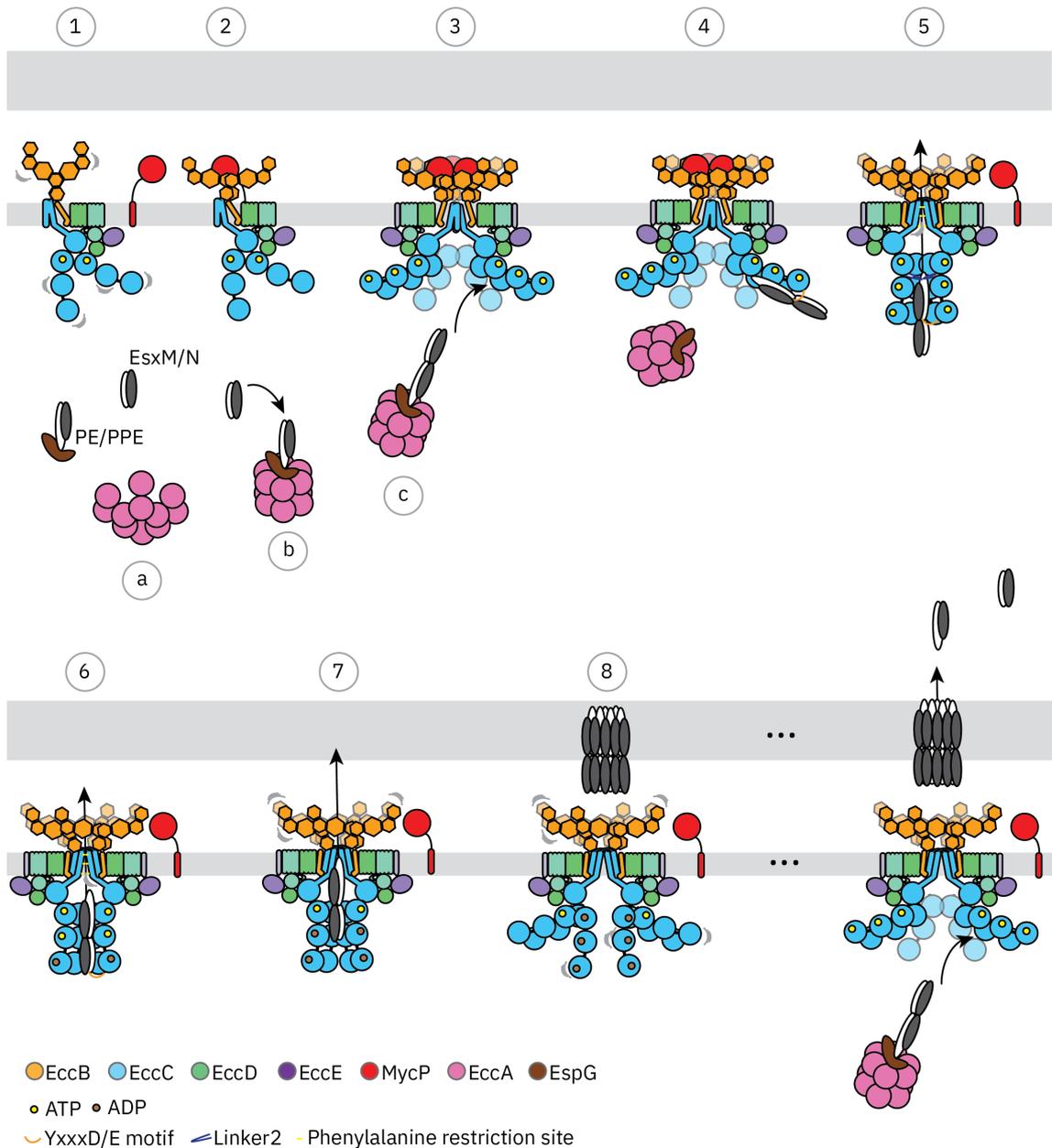


Figure 4.5: Proposed model of T7SS mediated substrate translocation First steps include the hexamerization of the dimeric ESX systems induced by MycP binding and further mediated by the rearrangement of the EccB periplasmic domains and the EccB basket (compare Figure 3.13, Figure 4.2, D) (1)(2). Heterooligomeric substrate complexes formed by EsxM/N and PE/PPE (as shown in Figure 4.4) are delivered by EccA and EspG to EccC (a)(b)(c), which adopts an upward conformation in presence of bound ATP (3). EccA acts as dissociase by inducing the dissociation of substrates and EspG as described in [61]. This allows binding of the YxxxD/E motif of the PE/PPE pair to EccC ATPase3 (4), the hexamerization of ATPase domains and the subsequent release of linker2 domain that coordinates the hydrophobic tip of PPE (5). Induced by this hexamerization, MycP dissociates, leading to a rearrangement of EccB periplasmic domains and the opening of the central secretion pore (compare Figure 4.2, D) (5). Driven by ATP the substrate complex is transported closer to the pore. The YxxxD/E motif on EsxM/N ensures the hexamerized state of the ATPase domains (6). Further ATP hydrolysis moves the substrates through the flexible pore, gated by phenylalanine residues (see Figure 3.11, Figure 3.12) (7). In the periplasm, substrates multimerize into pore like structures (8), which eventually penetrate the mycomembrane and facilitate their own transport across the outer membrane.

However, a possible gating mechanism and the involvement of the periplasmic domains need to be analyzed in more detail in the future. It cannot be excluded, that further rearrangements of the periplasmic domain is necessary to mediate the exit of the substrate.

How substrates cross the periplasmic space and the outer membrane remains a key open questions in the type VII secretion field. It has been shown that substrates like EsxE and EsxF or EspB have a tendency to oligomerize and it was suggested that some PE/PPE substrates could form pores in the outer membrane for nutrient transport [221, 189, 98, 229]. It is conceivable that other PE/PPE pairs could form outer membrane pores that allow substrate transport. The question of when and how exactly these higher oligomeric structures form has not been resolved yet. For EspB, for example, a concentration dependence has been observed [98]. The secretion of substrates across the inner membrane could increase a local concentration of substrates that oligomerize in a concentration dependent manner in the periplasm or in the outer membrane and may enable self-transport or the transport of further substrates (Figure 4.4, step 8). The hydrophobic tip of the PPE proteins could, for example, trigger an insertion into the outer membrane. Substrates could, thus, play a dual role in type VII secretion by forming an integral part of the secretion machinery as well as serving as effector proteins. This concept is not new. The hemolysin-coregulated protein (Hcp), an integral part of the type VI secretion system, was also shown to play a dual role in type VI secretion. Hcp hexamerizes and the resulting hexameric rings are stacked into tube like structures that bind and stabilize effectors to enable their transport. The Hcp protein itself also acts as an effector and fulfills very diverse roles including, for example, immune modulation by impairing phagocytosis [183]. Whether this highly hypothetical concept of concentration dependent oligomerization and secretion of T7SS substrates holds true needs to be tested in the future.

The structural and functional data provided in this thesis contributes to a comprehensive, molecular understanding of the T7SS. The transport model suggested integrates available functional and structural data and reveals several possible similarities to other secretion systems, implying that protein secretion is based on several common fundamental principles. Furthermore, the model also revealed many open questions that need to be tackled in order to obtain full insights into the molecular mechanism of type VII secretion. Further research is needed to understand the recognition and interaction of substrates with the complex and previous steps in the cytosol. How substrates interact could further shed light on the system and species specificity observed, which is poorly understood so far. Finally, it is necessary to identify the mechanistic of periplasmic and outer membrane transport, which allow effector proteins to reach their respective target location.

A detailed molecular and mechanistic understanding of the T7SS could facilitate the development of new and specific drugs against TB and further mycobacterial diseases. Increasing

case numbers and the emergence of multidrug-resistant mycobacterial strains highlight the urgent need for new anti-TB therapeutics [176]. The location at the host pathogen interface and the involvement of most ESX systems in virulence make T7SSs interesting drug targets [28]. The structural data provided in this thesis could help to identify type VII specific interactions that could be targeted in order to impair a functional assembly of the system, to interfere with the activation of transport or to block protein transport. Further research is needed to identify the best possible strategy.

5 Appendix

Table 5.1: Fold-changes of protein abundances of mutants compared to ESX-5 WT The fold-change describes the difference of the mean protein abundances normalized to ESX-5 on a \log_2 scale. Fold-changes are shown for all proteins encoded in the locus that were detected by LC-MS/MS.

fold-change (\log_2)	Secretome					
	P73A	F75C	V41F	V404144F	V57F	$\Delta\text{EccE}/\Delta\text{EccA}$
EccA ₅	-0.3	-0.3	-0.5	-0.6	-0.6	-3.5
EccB ₅	-0.2	0.0	-0.3	0.2	-0.1	-0.7
EccC ₅	0.2	0.0	-0.4	0.2	0.0	0.1
EccD ₅	-0.4	-0.2	-0.7	0.0	-0.4	-0.8
EccE ₅	-0.5	-0.3	-0.8	-0.1	-0.5	-3.1
EspG ₅	0.3	-0.1	-0.1	0.4	0.1	0.6
MycP ₅	0.2	0.2	0.1	0.5	0.2	0.0
EsxM	-3.8	-2.5	-2.3	-3.5	-3.3	-5.3
EsxN	-2.9	-1.9	-1.4	-2.9	-2.4	-4.0
PPE26.1	-4.8	-2.1	-2.7	-4.1	-3.8	-7.7
PPE26.2	-2.0	-1.1	-1.6	-0.8	-1.3	-3.8
fold-change (\log_2)	Whole cell					
	P73A	F75C	V41F	V404144F	V57F	$\Delta\text{EccE}/\Delta\text{EccA}$
EccA ₅	0.0	0.0	0.0	0.0	-0.1	-3.0
EccB ₅	0.3	0.2	0.4	0.2	0.0	-0.8
EccC ₅	0.3	0.2	0.3	0.2	0.2	-0.7
EccD ₅	0.3	0.2	0.4	0.4	0.3	-1.1
EccE ₅	0.2	0.2	0.3	0.3	0.2	-2.4
EspG ₅	0.5	0.7	0.7	0.7	0.6	0.9
MycP ₅	0.2	0.2	0.5	0.2	-0.2	-0.1
EsxM	0.1	0.1	0.5	0.5	0.2	1.0
EsxN	0.6	0.1	0.8	0.8	0.4	0.8
PPE26.2	0.8	2.1	2.0	2.1	2.1	-0.5

Table 5.2: Statistical significance of secretion assay results p-values are shown for the test of the null hypothesis "There is no difference between protein abundances in ESX-5 and mutant samples". * p<0.06; ** p<0.01; *** p<0.001

p-value	Secretome					
	P73A	F75C	V41F	V404144F	V57F	$\Delta\text{EccE}/\Delta\text{EccA}$
EccA ₅	1.3E-01	1.5E-01	1.4E-02*	4.3E-03**	4.9E-03**	1.4E-13***
EccB ₅	5.4E-01	8.8E-01	2.3E-01	3.7E-01	7.5E-01	1.6E-02*
EccC ₅	4.2E-01	9.3E-01	8.9E-02	4.7E-01	8.5E-01	5.0E-01
EccD ₅	3.3E-01	5.7E-01	7.9E-02	9.9E-01	3.3E-01	3.7E-02*
EccE ₅	2.2E-01	4.2E-01	8.4E-02	7.8E-01	2.9E-01	8.5E-07***
EspG ₅	1.7E-01	5.9E-01	7.6E-01	9.7E-02	7.7E-01	2.2E-02*
MycP ₅	5.3E-01	7.0E-01	7.6E-01	2.5E-01	5.8E-01	9.5E-01
EsxM	4.2E-05***	2.4E-03**	4.5E-03**	1.0E-04***	2.2E-04***	6.1E-07***
EsxN	4.2E-05***	3.2E-03**	1.9E-02*	4.5E-05***	3.9E-04***	7.7E-07***
PPE26.1	1.8E-03**	1.3E-01	5.6E-02*	6.8E-03**	1.1E-02*	1.5E-05***
PPE26.2	1.7E-02*	1.7E-01	4.9E-02*	3.3E-01	1.3E-01	1.1E-04***
p-value	Whole cell					
	P73A	F75C	V41F	V404144F	V57F	$\Delta\text{EccE}/\Delta\text{EccA}$
EccA ₅	9.4E-01	9.8E-01	1.0E+00	9.4E-01	4.7E-01	2.5E-12***
EccB ₅	4.8E-01	5.6E-01	2.3E-01	5.1E-01	9.0E-01	3.2E-02*
EccC ₅	2.5E-01	4.5E-01	3.1E-01	4.6E-01	5.6E-01	1.8E-02*
EccD ₅	5.1E-01	6.2E-01	3.3E-01	3.4E-01	5.2E-01	1.3E-02*
EccE ₅	4.8E-01	4.8E-01	3.2E-01	2.8E-01	5.2E-01	1.8E-07***
EspG ₅	1.2E-01	4.7E-02*	5.3E-02*	3.2E-02*	7.4E-02	1.5E-02*
MycP ₅	4.9E-01	4.9E-01	1.7E-01	4.5E-01	4.9E-01	8.7E-01
EsxM	7.0E-01	7.2E-01	1.1E-01	1.4E-01	4.6E-01	4.6E-03**
EsxN	2.0E-01	7.8E-01	9.1E-02	8.9E-02	3.2E-01	9.7E-02
PPE26.2	1.6E-01	1.4E-03**	2.3E-03**	1.2E-03**	1.7E-03**	4.1E-01

Table 5.3: Cryo-EM and model building statistics

	ESX-5 apo C6 full (1)	ESX-5 apo C1 full (2)	ESX-5 apo C6 local refined (3)
Magnification	130,000	130,000	130,000
Voltage (kV)	300	300	300
Electron exposure (e- Å ²)	49.34	49.34	49.34
Pixel size (Å)	0.645	0.645	0.645
Symmetry imposed	C6	C1	C6
Initial particle images (no.)	635,219	635,219	635,219
Final particle images (no.)	121,974	52,015	731,844
Map resolution (Å)	3.41	4.6	3.0
FSC threshold	0.143	0.143	0.143
Map resolution range (Å)	2.8 - 8.5	2.8 - 18.0	2.8 - 14.0
Refinement			
Initial model used (PDB code)	7b9f (3)	-	7b7j / de novo
PDB deposition id	7b9s		7b9f
EMDB deposition id	EMD-12105	EMD-12674	EMD-12103
Map sharpening Bfactor (Å ²)	-20	0	0
Model composition			
Non-hydrogen atoms	71622		11251
Protein residues	9240		1450
Ligands	0		0
B factors (Å²)			
Protein	152		128
Ligand	-		-
R.m.s. deviations			
Bond lengths (Å)	0.01		0.01
Bond angles (°)	1.06		1
Validation			
MolProbity score	1.44		1.42
Clashscore	3.91		4.11
Poor rotamers (%)	0.64		0
Ramachandran plot			
Favored (%)	96.12		96.56
Allowed (%)	3.87		3.44
Disallowed (%)	0.01		0
CC-mask (phenix score)	0.8		0.81

Table 5.4: Cryo-EM and model building statistics * indicates that model building is on-going.

	ESX-5 ATPyS C1 (4)	ESX-5 ATPyS C6 (5)	ESX-5 P73A C1 (6)	ESX-5 P73A C6 (7)
Magnification	130,000	130,000	105,000	105,000
Voltage (kV)	300	300	300	300
Electron exposure (e ⁻ Å ²)	48.8	48.8	40.9	40.9
Pixel size (Å)	0.68	0.68	0.85	0.85
Symmetry imposed	C1	C6	C1	C6
Initial particle images (no.)	320,185	320,185	985,743	985,743
Final particle images (no.)	38,516	38,516	78,699	78,699
Map resolution (Å)	4.8	3.87	5.4	3.9
FSC threshold	0.143	0.143	0.143	0.143
Map resolution range (Å)	3.0 - 57.265	3.0 - 35.2	43.0 - 59.8	2.5 - 44.6
Refinement				
Initial model used (PDB code)				7b9s (1)
PDB deposition id				
EMDB deposition id				
Map sharpening Bfactor (Å ²)	0	0	0	0
Model composition				
Non-hydrogen atoms				70182
Protein residues				9204
Ligands				0
B factors (Å²)				
Protein				148
Ligand				
R.m.s. deviations				
Bond lengths (Å)				0.01
Bond angles (°)				1.47
Validation				
MolProbity score				1.72
Clashscore				8.62
Poor rotamers (%)				0.08
Ramachandran plot				
Favored (%)				96.17
Allowed (%)				3.59
Disallowed (%)				0.24
CC-mask (phenix score)				0.71

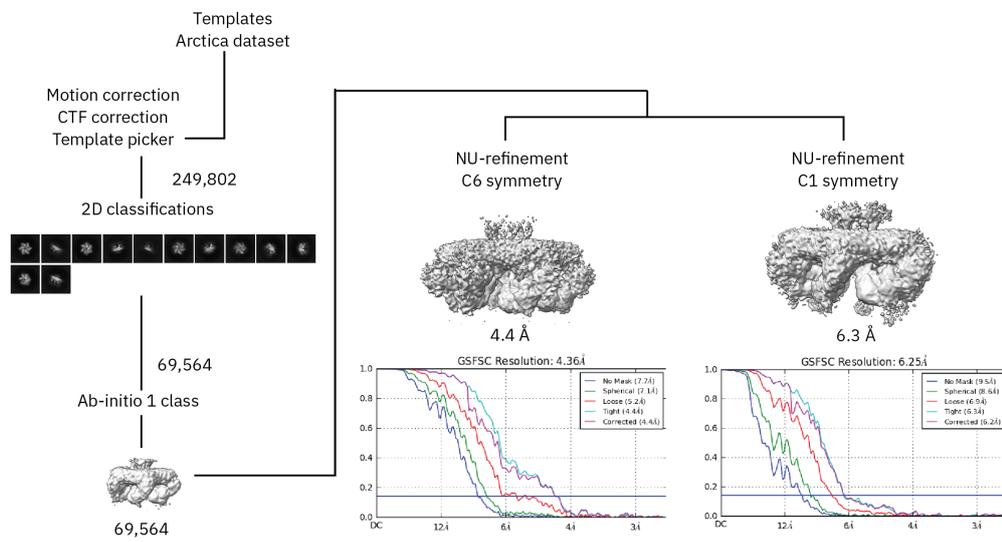


Figure 5.1: Image processing workflow of an initial cryo-EM dataset collected on the *Mycobacterium xenopi* ESX-5 complex. GSFSC curves are shown for a refined model in C6 symmetry and a refined model in C1 symmetry. The resolution obtained did not allow the building of a molecular model.

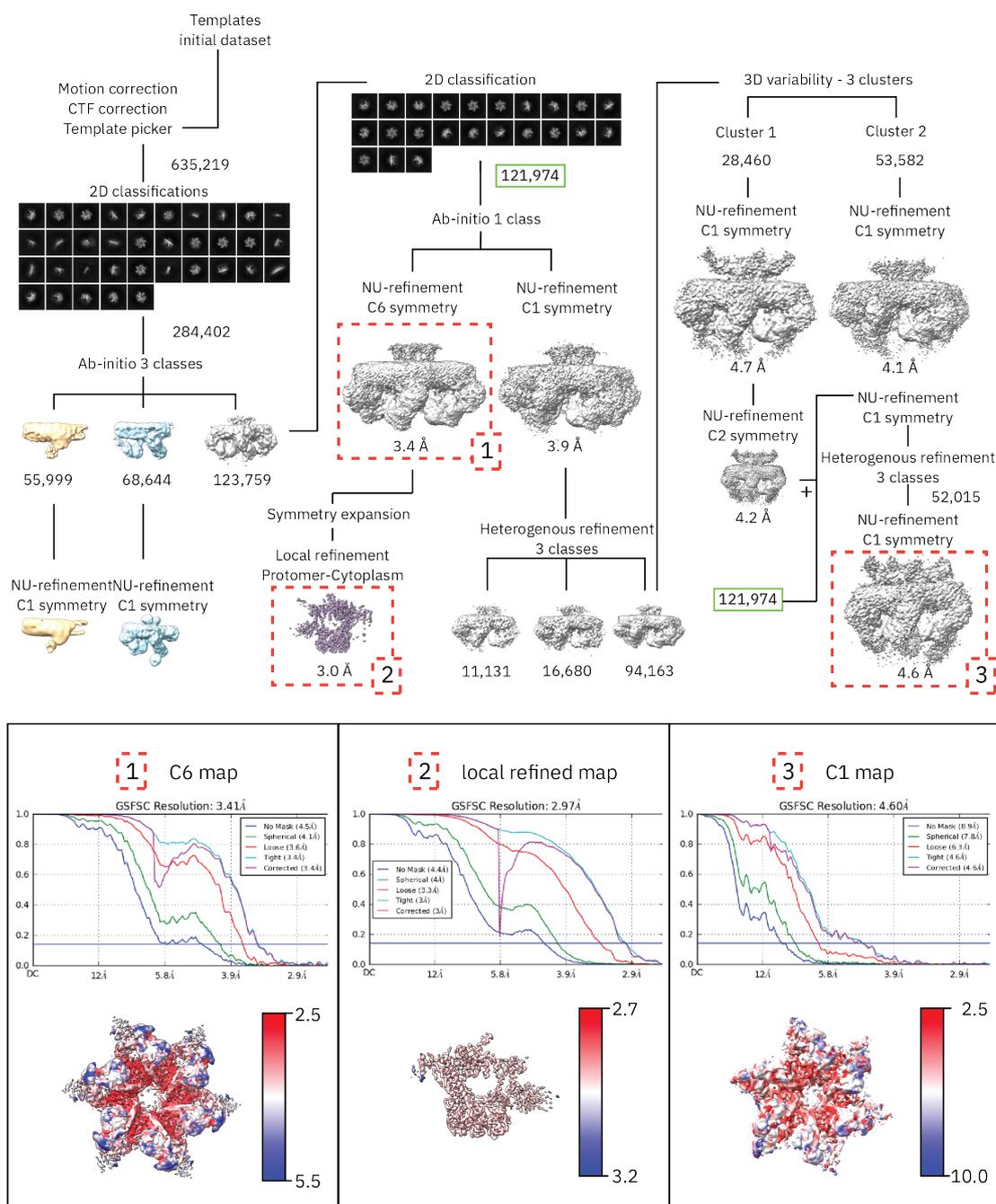


Figure 5.2: Image processing workflow of a final cryo-EM dataset collected on the *Mycobacterium xenopi* ESX-5 complex. The resolution of the maps could be improved compared to the initial dataset. Maps that were used for model building and interpretation are highlighted in orange boxes and are numbered 1, 2 and 3, respectively. The corresponding GSFSC curves and local resolution maps are shown below. The local resolution maps indicate by a red-white-blue colour scale which parts of the structure were resolved best and which regions are flexible. The figure is modified from [21], Figure S2

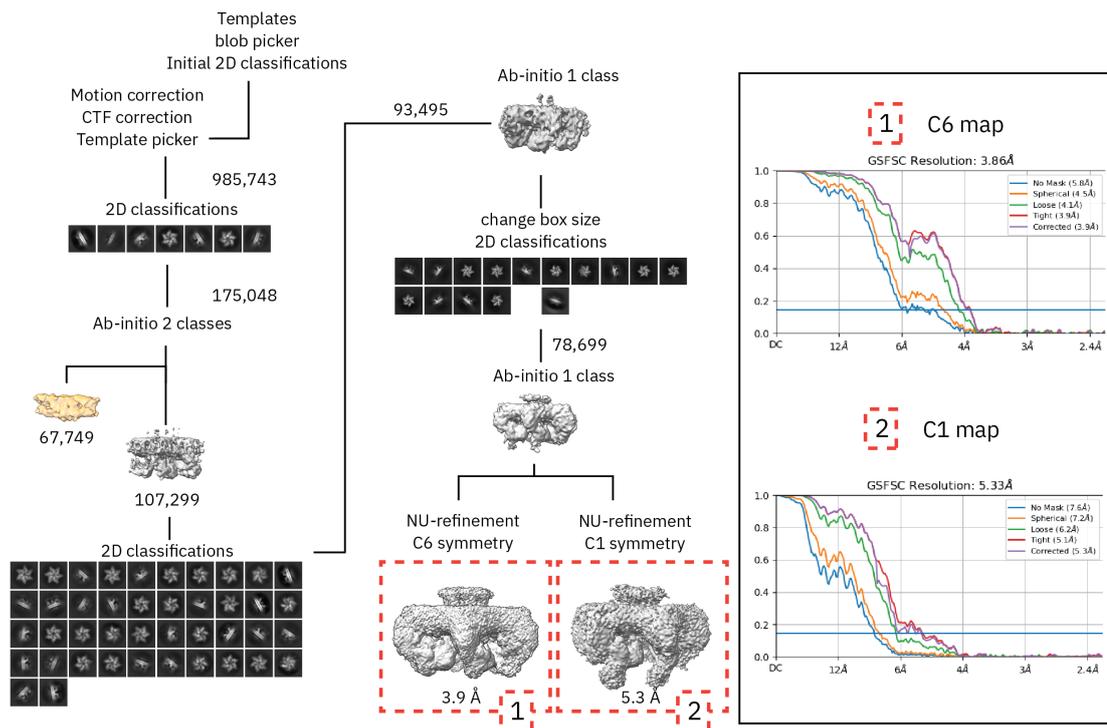


Figure 5.3: Image processing workflow of a cryo-EM dataset collected on the *Mycobacterium xenopi* ESX-5 P73A complex Maps that were used for interpretation are highlighted in orange boxes and are numbered 1, 2 respectively. The corresponding GSFSC curves are shown on the right. Workflows for further classification strategies are omitted.

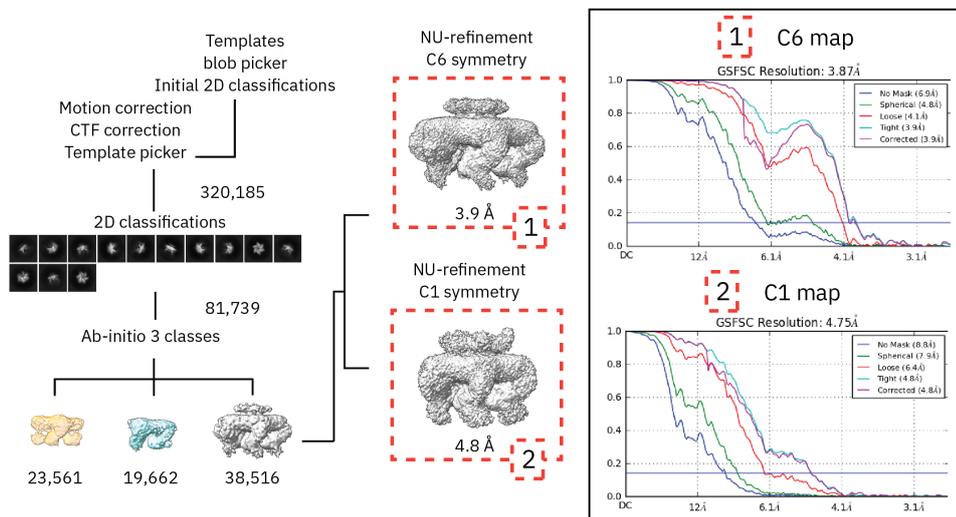


Figure 5.4: Image processing workflow of a cryo-EM dataset collected on the *Mycobacterium xenopi* ESX-5 complex in presence of ATP γ S. Maps that were used for interpretation are highlighted in orange boxes and are numbered 1, 2 respectively. The corresponding GSFSC curves are shown on the right. Workflows for further classification strategies are omitted.

Bibliography

- [1] A. M. Abdallah, N. C. Gey van Pittius, P. A. DiGiuseppe Champion, J. Cox, J. Luirink, C. M. J. E. Vandenbroucke-Grauls, B. J. Appelmelk, and W. Bitter. Type VII secretion — mycobacteria show the way. *Nature Reviews Microbiology*, 5:883–891, 2007.
- [2] A. M. Abdallah, N. D. L. Savage, M. van Zon, L. Wilson, C. M. J. E. Vandenbroucke-Grauls, N. N. van der Wel, T. H. M. Ottenhoff, and W. Bitter. The ESX-5 secretion system of *Mycobacterium marinum* modulates the macrophage response. *The Journal of Immunology*, 181(10):7166–7175, 2008.
- [3] A. M. Abdallah, T. Verboom, F. Hannes, M. Safi, M. Strong, D. Eisenberg, R. J. Musters, C. M. Vandenbroucke-Grauls, B. J. Appelmelk, J. Luirink, and W. Bitter. A specific secretion system mediates PPE41 transport in pathogenic mycobacteria. *Molecular Microbiology*, 62(3):667–679, 2006.
- [4] A. M. Abdallah, T. Verboom, E. M. Weerdenburg, N. C. Gey Van Pittius, P. W. Mahasha, C. Jiménez, M. Parra, N. Cadieux, M. J. Brennan, B. J. Appelmelk, and W. Bitter. PPE and PE-PGRS proteins of *Mycobacterium marinum* are transported via the type VII secretion system ESX-5. *Molecular Microbiology*, 73(3):329–340, 2009.
- [5] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19–25, 2015.
- [6] M. Adrian, J. Dubochet, J. Lepault, and A. W. McDowell. Cryo-electron microscopy of viruses. *Nature*, 308(5954):32–36, 1984.
- [7] P. V. Afonine, B. K. Poon, R. J. Read, O. V. Sobolev, T. C. Terwilliger, A. Urzhumtsev, and P. D. Adams. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallographica Section D: Structural Biology*, 74(6):531–544, 2018.
- [8] M. Allegretti, C. E. Zimmerli, V. Rantos, F. Wilfling, P. Ronchi, H. K. Fung, C. W. Lee, W. Hagen, B. Turoňová, K. Karius, M. Börmel, X. Zhang, C. W. Müller, Y. Schwab, J. Mahamid, B. Pfander, J. Kosinski, and M. Beck. In-cell architecture of the nuclear pore and snapshots of its turnover. *Nature*, 586(7831):796–800, 2020.
- [9] L. S. Ates. New insights into the mycobacterial PE and PPE proteins provide a framework for future research. *Molecular Microbiology*, 113(1):4–21, 2020.

-
- [10] L. S. Ates, A. Dippenaar, R. Ummels, S. R. Piersma, A. D. Van Der Woude, K. Van Der Kuij, F. Le Chevalier, D. Mata-Espinosa, J. Barrios-Payán, B. Marquina-Castillo, C. Guapillo, C. R. Jiménez, A. Pain, E. N. Houben, R. M. Warren, R. Brosch, R. Hernández-Pando, and W. Bitter. Mutations in ppe38 block PE-PGRS secretion and increase virulence of *Mycobacterium tuberculosis*. *Nature Microbiology*, 3(2):181–188, 2018.
- [11] L. S. Ates, F. Sayes, W. Frigui, R. Ummels, M. P. Damen, D. Bottai, M. A. Behr, J. W. van Heijst, W. Bitter, L. Majlessi, and R. Brosch. RD5-mediated lack of PE_PGRS and PPE-MPTR export in BCG vaccine strains results in strong reduction of antigenic repertoire but little impact on protection. *PLoS Pathogens*, 14(6):e1007139, 2018.
- [12] L. S. Ates, R. Ummels, S. Commandeur, R. van de Weerd, M. Sparrius, E. Weerdenburg, M. Alber, R. Kalscheuer, S. R. Piersma, A. M. Abdallah, M. Abd El Ghany, A. M. Abdel-Haleem, A. Pain, C. R. Jiménez, W. Bitter, and E. N. G. Houben. Essential role of the ESX-5 secretion system in outer membrane permeability of pathogenic mycobacteria. *PLoS genetics*, 11(5):e1005190–e1005190, 2015.
- [13] L. S. Ates, A. D. van der Woude, J. Bestebroer, G. van Stempvoort, R. J. Musters, J. J. Garcia-Vallejo, D. I. Picavet, R. van de Weerd, M. Maletta, C. P. Kuijl, N. N. van der Wel, and W. Bitter. The ESX-5 system of pathogenic mycobacteria is involved in capsule integrity and virulence through its substrate PPE10. *PLoS Pathogens*, 12(6):e1005696, 2016.
- [14] H. E. Autzen, D. Julius, Y. Cheng, S. Francisco, S. Francisco, and S. Francisco. Proteins in their Native Environment. *Curr Opin Struct Biol.*, 58:259–268, 2019.
- [15] M. R. Babu Sait, H. Koliwer-Brandl, J. A. Stewart, B. M. Swarts, M. Jacobsen, T. R. Ioerger, and R. Kalscheuer. PPE51 mediates uptake of trehalose across the mycomembrane of *Mycobacterium tuberculosis*. *Scientific Reports*, 12(1):1–12, 2022.
- [16] U. K. Bageshwar and S. M. Musser. Two electrical potential-dependent steps are required for transport by the Escherichia coli Tat machinery. *Journal of Cell Biology*, 179(1):87–99, 2007.
- [17] I. Barberis, N. L. Bragazzi, L. Galluzzo, and M. Martini. The history of tuberculosis: From the first historical records to the isolation of Koch’s bacillus. *Journal of Preventive Medicine and Hygiene*, 58(1):E9–E12, 2017.
- [18] G. Bashiri and E. N. Baker. Production of recombinant proteins in *Mycobacterium smegmatis* for structural and functional studies. *Protein Science*, 24(1):1–10, 2015.
- [19] T. H. Bayburt and S. G. Sligar. Membrane protein assembly into Nanodiscs. *FEBS Letters*, 584(9):1721–1727, 2010.
-

-
- [20] K. S. Beckham, L. Ciccarelli, C. M. Bunduc, H. D. Mertens, R. Ummels, W. Lugmayr, J. Mayr, M. Rettel, M. M. Savitski, D. I. Svergun, W. Bitter, M. Wilmanns, T. C. Marlovits, A. H. Parret, and E. N. Houben. Structure of the mycobacterial ESX-5 type VII secretion system membrane complex by single-particle analysis. *Nature Microbiology*, 2: 17047, 2017.
- [21] K. S. Beckham, C. Ritter, G. Chojnowski, D. S. Ziemianowicz, E. Mullapudi, M. Rettel, M. M. Savitski, S. A. Mortensen, J. Kosinski, and M. Wilmanns. Structure of the mycobacterial ESX-5 type VII secretion system pore complex. *Science Advances*, 7(26), 2021.
- [22] K. S. Beckham, S. Staack, M. Wilmanns, and A. H. Parret. The pMy vector series: A versatile cloning platform for the recombinant production of mycobacterial proteins in *Mycobacterium smegmatis*. *Protein Science*, 29(12):2528–2537, 2020.
- [23] H. J. Berendsen, D. van der Spoel, and R. van Drunen. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, 91(1-3):43–56, 1995.
- [24] R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, and A. D. J. MacKerell. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone Φ , ψ and side-chain χ_1 and χ_2 dihedral angles. *Journal of Chemical Theory and Computation*, 8(9):3257–3273, 2012.
- [25] W. Bitter, E. N. Houben, D. Bottai, P. Brodin, E. J. Brown, J. S. Cox, K. Derbyshire, S. M. Fortune, L. Y. Gao, J. Liu, N. C. Van Pittius, A. S. Pym, E. J. Rubin, D. R. Sherman, S. T. Cole, and R. Brosch. Systematic genetic nomenclature for type VII secretion systems. *PLoS Pathogens*, 5(10):8–13, 2009.
- [26] B. Blasco, J. M. Chen, R. Hartkoorn, C. Sala, S. Uplekar, J. Rougemont, F. Pojer, and S. T. Cole. Virulence regulator EspR of *Mycobacterium tuberculosis* is a nucleoid-associated protein. *PLoS Pathogens*, 8(3), 2012.
- [27] D. Bottai, M. di Luca, L. Majlessi, W. Frigui, R. Simeone, F. Sayes, W. Bitter, M. J. Brennan, C. Leclerc, G. Batoni, M. Campa, R. Brosch, and S. Esin. Disruption of the ESX-5 system of *Mycobacterium tuberculosis* causes loss of PPE protein secretion, reduction of cell wall integrity and strong attenuation. *Molecular Microbiology*, 83(6):1195–1209, 2012.
- [28] D. Bottai, A. Serafini, A. Cascioferro, R. Brosch, and R. Manganelli. Targeting type VII/ESX secretion systems for development of novel antimycobacterial drugs. *Current pharmaceutical design*, 20(27):4346–4356, 2014.
-

-
- [29] M. Braunstein, A. M. Brown, S. Kurtz, and J. Jacobs. Two nonredundant SecA homologues function in mycobacteria. *Journal of Bacteriology*, 183(24):6979–6990, 2001.
- [30] M. Braunstein, B. J. Espinosa, J. Chan, J. T. Belisle, and W. R. Jacobs. SecA2 functions in the secretion of superoxide dismutase A and in the virulence of *Mycobacterium tuberculosis*. *Molecular Microbiology*, 48(2):453–464, 2003.
- [31] J. N. Bright and M. S. Sansom. The flexing/twirling helix: Exploring the flexibility about molecular hinges formed by proline and glycine motifs in transmembrane helices. *Journal of Physical Chemistry B*, 107(2):627–636, 2003.
- [32] M. Brittany, K. Zulauf, and M. Braunstein. The Sec pathways and exportomes of *Mycobacterium tuberculosis*. *Microbiology Spectrum*, 5(2):5.2.24, 2017.
- [33] G. D. Brown, J. A. Dave, N. C. Gey Van Pittius, L. Stevens, M. R. Ehlers, and A. D. Beyers. The mycosins of *Mycobacterium tuberculosis* H37Rv: A family of subtilisin-like serine proteases. *Gene*, 254(1-2):147–155, 2000.
- [34] N. P. Brown, C. Leroy, and C. Sander. MView: A web-compatible database search or multiple alignment viewer. *Bioinformatics*, 14(4):380–381, 1998.
- [35] C. M. Bunduc, W. Bitter, and E. N. G. Houben. Structure and function of the mycobacterial type VII secretion systems. *Annual Review of Microbiology*, 2020.
- [36] C. M. Bunduc, D. Fahrenkamp, J. Wald, R. Ummels, W. Bitter, E. N. Houben, and T. C. Marlovits. Structure and dynamics of a mycobacterial type VII secretion system. *Nature*, 593(7859):445–448, 2021.
- [37] C. M. Bunduc, C. Kuijl, R. Ummels, T. C. Marlovits, W. Bitter, and E. N. G. Houben. Reconstitution of a minimal ESX-5 type VII secretion systems uncovers the essential role of the PPE proteins in secretion. *bioRxiv*, page 2022.09.05.506643, 2022.
- [38] C. M. Bunduc, R. Ummels, W. Bitter, and E. N. Houben. Species-specific secretion of ESX-5 type VII substrates is determined by the linker 2 of EccC5. *Molecular Microbiology*, 114(1):66–76, 2020.
- [39] C. Bussi and M. G. Gutierrez. *Mycobacterium tuberculosis* infection of host cells in space and time. *FEMS Microbiology Reviews*, 43(4):341–361, 2019.
- [40] X. Cai, L. Liu, C. Qiu, C. Wen, Y. He, Y. Cui, S. Li, X. Zhang, L. Zhang, C. Tian, L. Bi, Z. H. Zhou, and W. Gong. Identification and architecture of a putative secretion tube across mycobacterial outer envelope. *Science Advances*, 7(34), 2021.
- [41] M. Carroni and H. R. Saibil. Cryo electron microscopy to determine the structure of macromolecular complexes. *Methods*, 95(2016):78–85, 2016.
-

-
- [42] P. A. Champion, M. M. Champion, P. Manzanillo, and J. S. Cox. ESX-1 secreted virulence factors are recognized by multiple cytosolic AAA ATPases in mycobacteria. *Molecular Microbiology*, 73(5):950–962, 2009.
- [43] J. M. Chen, S. Boy-Röttger, N. Dhar, N. Sweeney, R. S. Buxton, F. Pojer, I. Rosenkrands, and S. T. Cole. EspD is critical for the virulence-mediating ESX-1 secretion system in *Mycobacterium tuberculosis*. *Journal of Bacteriology*, 194(4):884–893, 2012.
- [44] X. Chen, H. F. Cheng, J. Zhou, C. Y. Chan, K. F. Lau, S. K. W. Tsui, and S. W. ngor Au. Structural basis of the PE–PPE protein interaction in *Mycobacterium tuberculosis*. *Journal of Biological Chemistry*, 292(41):16880–16890, 2017.
- [45] Y. Cheng. Single-particle Cryo-EM at crystallographic resolution. *Cell*, 161(3):450–457, 2015.
- [46] L. Chiaradia, C. Lefebvre, J. Parra, J. Marcoux, O. Burlet-Schiltz, G. Etienne, M. Tropis, and M. Daffé. Dissecting the mycobacterial cell envelope and defining the composition of the native mycomembrane. *Scientific Reports*, 7(1):12807, 2017.
- [47] G. Chojnowski. Prediction of type VII secretion systems by AlphaFold2. *unpublished*.
- [48] G. Chojnowski, J. Pereira, and V. S. Lamzin. Sequence assignment for low-resolution modelling of protein crystal structures. *Acta Crystallographica Section D: Structural Biology*, 75:753–763, 2019.
- [49] G. Chojnowski, A. J. Simpkin, D. A. Leonardo, W. Seifert-Davila, D. E. Vivas-Ruiz, R. M. Keegan, and D. J. Rigden. FindMySequence: a neural-network-based approach for identification of unknown proteins in X-ray crystallography and cryo-EM. *IUCrJ*, 9:86–97, 2022.
- [50] S. B. Cohen, B. H. Gern, and K. B. Urdahl. The Tuberculous Granuloma and Preexisting Immunity. *Annual Review of Immunology*, 40:589–614, 2022.
- [51] S. T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, Eiglmeier, S. Gas, C. E. Barry, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, and B. B. Jagels. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, 393:537–544, 1998.
- [52] W. H. Conrad, M. M. Osman, J. K. Shanahan, F. Chu, K. K. Takaki, J. Cameron, D. Hopkinson-Woolley, R. Brosch, and L. Ramakrishnan. Mycobacterial ESX-1 secretion system mediates host cell lysis through bacterium contact-dependent gross membrane disruptions. *Proceedings of the National Academy of Sciences*, 114(6):1371–1376, 2017.
-

-
- [53] S. E. Converse and J. S. Cox. A protein secretion pathway critical for *Mycobacterium tuberculosis* virulence is conserved and functional in *Mycobacterium smegmatis*. *Journal of bacteriology*, 187(4):1238–1245, 2005.
- [54] F. S. Cordes, J. N. Bright, and M. S. Sansom. Proline-induced distortions of transmembrane helices. *Journal of Molecular Biology*, 323(5):951–960, 2002.
- [55] R. A. Corey, E. Pyle, W. J. Allen, D. W. Watkins, M. Casiraghi, B. Miroux, I. Arechaga, A. Politis, and I. Collinson. Specific cardiolipin–SecY interactions are required for proton-motive force stimulation of protein secretion. *Proceedings of the National Academy of Sciences of the United States of America*, 115(31):7967–7972, 2018.
- [56] T. R. Costa, C. Felisberto-Rodrigues, A. Meir, M. S. Prevost, A. Redzej, M. Trokter, and G. Waksman. Secretion systems in Gram-negative bacteria: Structural and mechanistic insights. *Nature Reviews Microbiology*, 13(6):343–359, 2015.
- [57] S. Cristóbal, J. W. De Gier, H. Nielsen, and G. Von Heijne. Competition between Sec- and TAT-dependent protein translocation in *Escherichia coli*. *EMBO Journal*, 18(11):2982–2990, 1999.
- [58] T. I. Croll. ISOLDE: A physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallographica Section D: Structural Biology*, 74(6):519–530, 2018.
- [59] R. M. Cronin, M. J. Ferrell, C. W. Cahir, M. M. Champion, and P. A. Champion. Proteo-genetic analysis reveals clear hierarchy of ESX-1 secretion in *Mycobacterium marinum*. *Proceedings of the National Academy of Sciences of the United States of America*, 119(24):e2123100119, 2022.
- [60] G. Crooks, G. Hon, J. Chandonia, and S. Brenner. Weblogo: a sequence logo generator. *Genome Res*, 14:1188–1190, 2004.
- [61] T. D. Crosskey, K. S. Beckham, and M. Wilmanns. The ATPases of the mycobacterial type VII secretion system: Structural and mechanistic insights into secretion. *Progress in Biophysics and Molecular Biology*, 152:25–34, 2020.
- [62] R. A. Crowther, L. A. Amos, J. T. Finch, D. J. De Rosier, and A. Klug. Three dimensional reconstructions of spherical viruses by Fourier synthesis from electron micrographs. *Nature*, 226(5244):421–425, 1970.
- [63] M. H. Daleke, R. Ummels, P. Bawono, J. Heringa, C. M. Vandenbroucke-Grauls, J. Luirink, and W. Bitter. General secretion signal for the mycobacterial type VII secretion pathway. *Proceedings of the National Academy of Sciences of the United States of America*, 109(28):11342–11347, 2012.
-

-
- [64] X. M. P. Damen, T. H. Phan, R. Ummels, A. Rubio-Canalejas, W. Bitter, and X. E. N. Houben. Modification of a PE/PPE substrate pair reroutes an Esx substrate pair from the mycobacterial ESX-1 type VII secretion system to the ESX-5 system. *Journal of Biological Chemistry*, 295(18):5960–5969, 2020.
- [65] L. D. D’Andrea and L. Regan. TPR proteins: The versatile helix. *Trends in Biochemical Sciences*, 28(12):655–662, 2003.
- [66] M. I. Dauden, J. Kosinski, O. Kolaj-Robin, A. Desfosses, A. Ori, C. Faux, N. A. Hoffmann, O. F. Onuma, K. D. Breunig, M. Beck, C. Sachse, B. Séraphin, S. Glatt, and C. W. Müller. Architecture of the yeast Elongator complex. *EMBO reports*, 18(2):264–279, 2017.
- [67] J. M. Davis and L. Ramakrishnan. The role of the granuloma in expansion and dissemination of early Tuberculosis infection. *Cell*, 136(1):37–49, 2009.
- [68] M. I. De Jonge, G. Pehau-Arnaudet, M. M. Fretz, F. Romain, D. Bottai, P. Brodin, N. Honoré, G. Marchal, W. Jiskoot, P. England, S. T. Cole, and R. Brosch. ESAT-6 from *Mycobacterium tuberculosis* dissociates from its putative chaperone CFP-10 under acidic conditions and exhibits membrane-lysing activity. *Journal of Bacteriology*, 189(16):6028–6034, 2007.
- [69] M. de Martino, L. Lodi, L. Galli, and E. Chiappini. Immune response to *Mycobacterium tuberculosis*: A narrative review. *Frontiers in Pediatrics*, 7:350, 2019.
- [70] S. Dechow, J. Baker, M. Murto, and R. Abramovitch. ppe51 Variants enable growth of *Mycobacterium tuberculosis* at acidic pH by selectively promoting glycerol uptake. *Journal of Bacteriology*, 204(11):e00212–22, 2022.
- [71] M. Di Luca, D. Bottai, G. Batoni, M. Orgeur, A. Aulicino, C. Counoupas, M. Campa, R. Brosch, and S. Esin. The ESX-5 associated eccB5-eccC5 locus is essential for *Mycobacterium tuberculosis* viability. *PLoS ONE*, 7(12), 2012.
- [72] H. M. Dockrell and S. G. Smith. What have we learnt about BCG vaccination in the last 20 years? *Frontiers in Immunology*, 8:1–10, 2017.
- [73] R. S. G. D’Rozario and M. S. P. Sansom. Helix dynamics in a membrane transport protein: comparative simulations of the glycerol-3-phosphate transporter and its constituent helices. *Molecular Membrane Biology*, 25(6-7):571–583, 2008.
- [74] J. Dubochet and A. W. McDowell. Vitrification of pure water for electron microscopy. *Journal of Microscopy*, 124(3):3–4, 1981.
- [75] C. L. Dulberger, E. J. Rubin, and C. C. Boutte. The mycobacterial cell envelope — a moving target. *Nature Reviews Microbiology*, 18(1):47–59, 2020.
-

- [76] M. Ekansh, S. M. L., U. Grace, T. Emir, M. Alka, K. Stefan, H. P. I., and P. J. A. *Mycobacterium tuberculosis* type VII secretion system effectors differentially impact the ESCRT endomembrane damage response. *mBio*, 9(6):e01765–18, 2018.
- [77] D. C. Ekiert and J. S. Cox. Structure of a PE-PPE-EspG complex from *Mycobacterium tuberculosis* reveals molecular specificity of ESX protein secretion. *Proceedings of the National Academy of Sciences*, 111(41):14758–14763, 2014.
- [78] P. Emsley and K. Cowtan. Coot: Model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography*, 60(Pt 12 Pt 1):2126–2132, 2004.
- [79] R. Evans, M. O'Neill, A. Pritzel, N. Antropova, A. Senior, T. Green, A. Žídek, R. Bates, S. Blackwell, J. Yim, O. Ronneberger, S. Bodenstein, M. Zielinski, A. Bridgland, A. Potapenko, A. Cowie, K. Tunyasuvunakool, R. Jain, E. Clancy, P. Kohli, J. Jumper, and D. Hassabis. Protein complex prediction with AlphaFold-Multimer. *bioRxiv*, page 2021.10.04.463034, 2022.
- [80] N. Famelis, A. Rivera-Calzada, G. Degliesposti, M. Wingender, N. Mietrach, J. M. Skehel, R. Fernandez-Leiro, B. Böttcher, A. Schlosser, O. Llorca, and S. Geibel. Architecture of the mycobacterial type VII secretion system. *Nature*, 576(7786):321–325, 2019.
- [81] T. Fedrizzi, C. J. Meehan, A. Grottola, E. Giacobazzi, G. Fregni Serpini, S. Tagliazucchi, A. Fabio, C. Bettua, R. Bertorelli, V. De Sanctis, F. Rumpianesi, M. Pecorari, O. Jousson, E. Tortoli, and N. Segata. Genomic characterization of Nontuberculous Mycobacteria. *Scientific Reports*, 7:45258, 2017.
- [82] M. E. Feltcher, H. S. Gibbons, L. S. Ligon, and M. Braunstein. Protein export by the mycobacterial SecA2 system is determined by the preprotein mature domain. *Journal of Bacteriology*, 195(4):672–681, 2013.
- [83] M. E. Feltcher, J. T. Sullivan, and M. Braunstein. Protein export systems of *Mycobacterium tuberculosis*: Novel targets for drug development? *Future Microbiology*, 5(10):1581–1597, 2010.
- [84] A. Filloux. Bacterial protein secretion systems: Game of types. *Microbiology (United Kingdom)*, 168(5):1–16, 2022.
- [85] S. Fishbein, N. van Wyk, R. M. Warren, and S. L. Sampson. Phylogeny to function: PE/PPE protein evolution and impact on *Mycobacterium tuberculosis* pathogenicity. *Molecular Microbiology*, 96(5):901–916, 2015.
- [86] A. Flayhan, H. D. Mertens, Y. Ural-Blimke, M. Martinez Molledo, D. I. Svergun, and C. Löw. Saposin lipid nanoparticles: A highly versatile and modular tool for membrane protein research. *Structure*, 26(2):345–355.e5, 2018.

- [87] S. M. Fortune, A. Jaeger, D. A. Sarracino, M. R. Chase, C. M. Sasseti, D. R. Sherman, B. R. Bloom, and E. J. Rubin. Mutually dependent secretion of proteins required for mycobacterial virulence. *Proceedings of the National Academy of Sciences of the United States of America*, 102(30):10676–10681, 2005.
- [88] J. Frank. *Three-dimensional electron microscopy of macromolecular assemblies: visualization of biological molecules in their native state*. Oxford university press, 2006.
- [89] H. Franken, T. Mathieson, D. Childs, G. M. Sweetman, T. Werner, I. Tögel, C. Doce, S. Gade, M. Bantscheff, G. Drewes, F. B. Reinhard, W. Huber, and M. M. Savitski. Thermal proteome profiling for unbiased identification of direct and indirect drug targets using multiplexed quantitative mass spectrometry. *Nature Protocols*, 10(10):1567–1593, 2015.
- [90] J. Frauenfeld, R. Löving, J.-P. Armache, A. F.-P. Sonnen, F. Guettou, P. Moberg, L. Zhu, C. Jegerschöld, A. Flayhan, J. A. G. Briggs, H. Garoff, C. Löw, Y. Cheng, and P. Nordlund. A saposin-lipoprotein nanoparticle system for membrane proteins. *Nature methods*, 13(4):345–351, 2016.
- [91] B. Frenz, A. C. Walls, E. H. Egelman, D. Veessler, and F. Di Maio. RosettaES: A sampling strategy enabling automated interpretation of difficult cryo-EM maps. *Nature Methods*, 14(8):797–800, 2017.
- [92] W. Frigui, D. Bottai, L. Majlessi, M. Monot, E. Josselin, P. Brodin, T. Garnier, B. Gicquel, C. Martin, C. Leclerc, S. T. Cole, and R. Brosch. Control of *M. tuberculosis* ESAT-6 secretion and specific T cell recognition by PhoP. *PLoS Pathogens*, 4(2):e33, 2008.
- [93] L. Y. Gao, S. Guo, B. McLaughlin, H. Morisaki, J. N. Engel, and E. J. Brown. A mycobacterial virulence gene cluster extending RD1 is required for cytolysis, bacterial spreading and ESAT-6 secretion. *Molecular Microbiology*, 53(6):1677–1693, 2004.
- [94] R. M. Garavito and S. Ferguson-Miller. Detergents as Tools in Membrane Biochemistry. *Journal of Biological Chemistry*, 276(35):32403–32406, 2001.
- [95] S. N. Gates and A. Martin. Stairway to translocation: AAA+ motor structures reveal the mechanisms of ATP-dependent substrate translocation. *Protein Science*, 29(2):407–419, 2020.
- [96] N. C. Gey Van Pittius, J. Gamielien, W. Hide, G. D. Brown, R. J. Siezen, and A. D. Beyers. The ESAT-6 gene cluster of *Mycobacterium tuberculosis* and other high G+C Gram-positive bacteria. *Genome biology*, 2(10):1–18, 2001.
- [97] N. C. Gey Van Pittius, S. L. Sampson, H. Lee, Y. Kim, P. D. Van Helden, and R. M. Warren. Evolution and expansion of the *Mycobacterium tuberculosis* pe and

- ppe multigene families and their association with the duplication of the *esat-6* (*esx*) gene cluster regions. *BMC evolutionary biology*, 6, 2006.
- [98] A. Gijbsbers, V. Vinciauskaite, A. Siroy, Y. Gao, G. Tria, A. Mathew, N. Sánchez-Puig, C. López-Iglesias, P. J. Peters, and R. B. Ravelli. Priming mycobacterial ESX-secreted protein B to form a channel-like structure. *Current Research in Structural Biology*, 3:153–164, 2021.
- [99] R. M. Glaeser. Retrospective: Radiation damage and its associated "Information Limitations". *Journal of Structural Biology*, 163(3):271–276, 2008.
- [100] R. M. Glaeser. Invited Review Article: Methods for imaging weak-phase objects in electron microscopy. *Review of Scientific Instruments*, 84(11), 2013.
- [101] T. D. Goddard, C. C. Huang, E. C. Meng, E. F. Pettersen, G. S. Couch, J. H. Morris, and T. E. Ferrin. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Science*, 27(1):14–25, 2018.
- [102] R. M. Goldstone, N. J. Moreland, G. Bashiri, E. N. Baker, and J. Shaun Lott. A new Gateway® vector and expression protocol for fast and efficient recombinant protein expression in *Mycobacterium smegmatis*. *Protein Expression and Purification*, 57(1):81–87, 2008.
- [103] D. G. Gorasia, P. D. Veith, and E. C. Reynolds. The type IX secretion system: Advances in structure, function and organisation. *Microorganisms*, 8(8):1–9, 2020.
- [104] B. R. Gordon, Y. Li, L. Wang, A. Sintsova, H. Van Bakel, S. Tian, W. W. Navarre, B. Xia, and J. Liu. Lsr2 is a nucleoid-associated protein that targets AT-rich sequences and virulence genes in *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences of the United States of America*, 107(11):5154–5159, 2010.
- [105] T. A. Gray and K. M. Derbyshire. Blending genomes: distributive conjugal transfer in mycobacteria, a sexier form of HGT. *Molecular Microbiology*, 108(6):601–613, 2018.
- [106] C. Gray, Todd A., Clark, Ryan, R., Boucher, Natalie, Lapierre, Pascal, Smith and K. M. Derbyshire. Secretion Systems in Mycobacteria. *Science*, 354(6310):347–350, 2016.
- [107] E. R. Green and J. Meccas. Bacterial secretion systems: An overview. *Virulence Mechanisms of Bacterial Pathogens*, 4(1):213–239, 2016.
- [108] Y. V. Grinkova, I. G. Denisov, and S. G. Sligar. Engineering extended membrane scaffold proteins for self-assembly of soluble nanoscale lipid bilayers. *Protein Engineering, Design and Selection*, 23(11):843–848, 2010.

-
- [109] M. I. Gröschel, F. Sayes, R. Simeone, L. Majlessi, and R. Brosch. ESX secretion systems: Mycobacterial evolution to counter host immunity. *Nature Reviews Microbiology*, 14(11):677–691, 2016.
- [110] R. W. Grosse-Kunstleve, N. K. Sauter, N. W. Moriarty, and P. D. Adams. The Computational Crystallography Toolbox: Crystallographic algorithms in a reusable software framework. *Journal of Applied Crystallography*, 35(1):126–136, 2002.
- [111] A. S. Grossman, T. J. Mauer, K. T. Forest, and H. Goodrich-Blair. A widespread bacterial secretion system with diverse substrates. *mBio*, 12(4), 2021.
- [112] A. Grottesi, C. Domene, B. Hall, and M. S. Sansom. Conformational dynamics of M2 helices in KirBac channels: Helix flexibility in relation to gating via molecular dynamics simulations. *Biochemistry*, 44(44):14586–14594, 2005.
- [113] S. Hamsanathan and S. M. Musser. The Tat protein transport system: Intriguing questions and conundrums. *FEMS Microbiology Letters*, 365(12):1–11, 2018.
- [114] N. J. Hardenbrook, S. Liu, K. Zhou, K. Ghosal, Z. Hong Zhou, and B. A. Krantz. Atomic structures of anthrax toxin protective antigen channels bound to partially unfolded lethal and edema factors. *Nature Communications*, 11(1):840, 2020.
- [115] N. Herrera, P. D. Odermatt, M. Voorhies, R. Nakagawa, A. Sil, F. Chang, and O. S. Rosenberg. The ESX-1 secretion system senses bacterial contact and prepares mycobacteria for environmental adaptation. *bioRxiv*, page 2021.10.17.464699, 2021.
- [116] S. Hofmann, D. Janulienė, A. R. Mehdipour, C. Thomas, E. Stefan, S. Brüchert, B. T. Kuhn, E. R. Geertsma, G. Hummer, R. Tampé, and A. Moeller. Conformation space of a heterodimeric ABC exporter under turnover conditions. *Nature*, 571(7766):580–583, 2019.
- [117] S. A. Hollingsworth and R. O. Dror. Molecular dynamics simulation for all. *Neuron*, 99(6):1129–1143, 2018.
- [118] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics*, 65(3):712–725, 2006.
- [119] D. Houben, C. Demangel, J. van Ingen, J. Perez, L. Baldeón, A. M. Abdallah, L. Caleechurn, D. Bottai, M. van Zon, K. de Punder, T. van der Laan, A. Kant, R. Bossers-de Vries, P. Willemsen, W. Bitter, D. van Soolingen, R. Brosch, N. van der Wel, and P. J. Peters. ESX-1-mediated translocation to the cytosol controls virulence of mycobacteria. *Cellular Microbiology*, 14(8):1287–1298, 2012.
-

- [120] E. N. Houben, K. V. Korotkov, and W. Bitter. Take five - Type VII secretion systems of Mycobacteria. *Biochimica et Biophysica Acta - Molecular Cell Research*, 1843(8):1707–1716, 2014.
- [121] E. N. G. Houben, J. Bestebroer, R. Ummels, L. Wilson, S. R. Piersma, C. R. Jiménez, T. H. M. Ottenhoff, J. Luirink, and W. Bitter. Composition of the type VII secretion system membrane complex. *Molecular Microbiology*, 86(2):472–484, 2012.
- [122] C. S. Hughes, S. Foehr, D. A. Garfield, E. E. Furlong, L. M. Steinmetz, and J. Krijgsveld. Ultrasensitive proteome analysis using paramagnetic bead technology. *Molecular Systems Biology*, 10(10):757, 2014.
- [123] C. S. Hughes, S. Moggridge, T. Müller, P. H. Sorensen, G. B. Morin, and J. Krijgsveld. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nature Protocols*, 14(1):68–85, 2019.
- [124] S. Hüsing, M. Halte, U. van Look, A. Guse, E. J. Gálvez, E. Charpentier, D. F. Blair, M. Erhardt, and T. T. Renault. Control of membrane barrier during bacterial type-III protein secretion. *Nature Communications*, 12(1):1–12, 2021.
- [125] B. Izquierdo Lafuente, R. Ummels, C. Kuijl, W. Bitter, and A. Speer. *Mycobacterium tuberculosis* toxin cpnt is an esx-5 substrate and requires three type vii secretion systems for intracellular secretion. *mBio*, 12(2):1–16, 2021.
- [126] S. V. Jamwal, P. Mehrotra, A. Singh, Z. Siddiqui, A. Basu, and K. V. S. Rao. Mycobacterial escape from macrophage phagosomes to the cytoplasm represents an alternate adaptation mechanism. *Scientific Reports*, 6(1):23089, 2016.
- [127] V. Jarlier and H. Nikaido. Permeability barrier to hydrophilic solutes in *Mycobacterium chelonae*. *Journal of Bacteriology*, 172(3):1418–1423, 1990.
- [128] S. Jo, T. Kim, V. G. Iyer, and W. Im. CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry*, 29(11):1859–1865, 2008.
- [129] M. D. Johansen, J. L. Herrmann, and L. Kremer. Non-tuberculous mycobacteria and the rise of *Mycobacterium abscessus*. *Nature Reviews Microbiology*, 18(7):392–407, 2020.
- [130] S. A. Joshi, D. A. Ball, M. G. Sun, F. Carlsson, B. Y. Watkins, N. Aggarwal, J. M. McCracken, K. K. Huynh, and E. J. Brown. EccA1, a component of the *Mycobacterium marinum* ESX-1 protein virulence factor secretion pathway, regulates mycolic acid lipid synthesis. *Chemistry and Biology*, 19(3):372–380, 2012.
- [131] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Židek, A. Potapenko, A. Bridgland, C. Meyer, S. A. Kohl, A. J.

- Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstern, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, 2021.
- [132] C. Kahramanoglou, T. Cortes, N. Matange, D. M. Hunt, S. S. Visweswariah, D. B. Young, and R. S. Buxton. Genomic mapping of cAMP receptor protein (CRPMt) in *Mycobacterium tuberculosis*: Relation to transcriptional start sites and the role of CRPMt as a transcription factor. *Nucleic Acids Research*, 42(13):8320–8329, 2014.
- [133] R. Kalscheuer, A. Palacios, I. Anso, J. Cifuentes, J. Anguita, W. R. Jacobs, M. E. Guerin, and R. Prados-Rosales. The *Mycobacterium tuberculosis* capsule: A cell structure with key implications in pathogenesis. *Biochemical Journal*, 476(14):1995–2016, 2019.
- [134] R. L. Kinsella, D. X. Zhu, G. A. Harrison, A. E. Mayer Bridwell, J. Prusa, S. M. Chavez, and C. L. Stallings. Perspectives and Advances in the Understanding of Tuberculosis. *Annual Review of Pathology: Mechanisms of Disease*, 16(1):377–408, 2021.
- [135] T. Kluyver, B. Ragan-Kelley, F. Pérez, B. Granger, M. Bussonnier, J. Frederic, K. Kelley, J. Hamrick, J. Grout, S. Corlay, P. Ivanov, D. Avila, S. Abdalla, and C. Willing. Jupyter Notebooks—a publishing format for reproducible computational workflows. *Positioning and Power in Academic Publishing: Players, Agents and Agendas - Proceedings of the 20th International Conference on Electronic Publishing, ELPUB 2016*, pages 87–90, 2016.
- [136] T. J. Knowles, R. Finka, C. Smith, Y. P. Lin, T. Dafforn, and M. Overduin. Membrane proteins solubilized intact in lipid containing nanoparticles bounded by styrene maleic acid copolymer. *Journal of the American Chemical Society*, 131(22):7484–7485, 2009.
- [137] N. Korotkova, D. Freire, T. H. Phan, R. Ummels, C. C. Creekmore, T. J. Evans, M. Wilmanns, W. Bitter, A. H. Parret, E. N. Houben, and K. V. Korotkov. Structure of the *Mycobacterium tuberculosis* type VII secretion system chaperone EspG5 in complex with PE25-PPE41 dimer. *Molecular Microbiology*, 94(2):367–382, 2014.
- [138] N. Korotkova, J. Piton, J. M. Wagner, S. Boy-Röttger, A. Japaridze, T. J. Evans, S. T. Cole, F. Pojer, and K. V. Korotkov. Structure of EspB, a secreted substrate of the ESX-1 secretion system of *Mycobacterium tuberculosis*. *Journal of Structural Biology*, 191(2):236–244, 2015.
- [139] M. Korycka-Machała, J. Pawełczyk, P. Borówka, B. Dziadek, A. Brzostek, M. Kawka, A. Bekier, S. Rykowski, A. B. Olejniczak, D. Strapagiel, Z. Witczak, and J. Dziadek. PPE51 is involved in the uptake of disaccharides by *Mycobacterium tuberculosis*. *Cells*, 9(3):603, 2020.

- [140] J. Kosinski, S. Mosalaganti, A. von Appen, R. Teimer, A. L. DiGuilio, W. Wan, K. H. Bui, W. J. H. Hagen, J. A. G. Briggs, J. S. Glavy, E. Hurt, and M. Beck. Molecular architecture of the inner ring scaffold of the human nuclear pore complex. *Science*, 352(6283):363–365, 2016.
- [141] N. L. Kriel, M. Newton-Foot, O. T. Bennion, B. B. Aldridge, C. Mehaffy, J. T. Belisle, G. Walzl, R. M. Warren, S. L. Sampson, and N. C. Gey van Pittius. Localization of EccA3 at the growing pole in *Mycobacterium smegmatis*. *BMC Microbiology*, 22(1):1–12, 2022.
- [142] W. Kühlbrandt. The resolution revolution. *Science*, 343(6178):1443–1444, 2014.
- [143] V. A. Kumar, R. Goyal, R. Bansal, N. Singh, R. R. Sevalkar, A. Kumar, and D. Sarkar. Espr-dependent ESAT-6 protein secretion of *Mycobacterium tuberculosis* requires the presence of virulence regulator phoP. *Journal of Biological Chemistry*, 291(36):19018–19030, 2016.
- [144] L. Laencina, V. Dubois, V. Le Moigne, A. Viljoen, L. Majlessi, J. Pritchard, A. Bernut, L. Piel, A.-L. Roux, J.-L. Gaillard, B. Lombard, D. Loew, E. J. Rubin, R. Brosch, L. Kremer, J.-L. Herrmann, and F. Girard-Misguich. Identification of genes required for *Mycobacterium abscessus* growth in vivo with a prominent role of the ESX-4 locus. *Proceedings of the National Academy of Sciences*, 115(5):E1002–E1011, 2018.
- [145] M. Lagune, V. Le Moigne, M. D. Johansen, F. V. Sotomayor, W. Daher, C. Petit, G. Cosentino, L. Paulowski, T. Gutschmann, M. Wilmanns, F. P. Maurer, J. L. Herrmann, F. Girard-Misguich, and L. Kremer. The ESX-4 substrates, EsxU and EsxT, modulate *Mycobacterium abscessus* fitness. *PLoS Pathogens*, 18(8):e1010771, 2022.
- [146] M. Lagune, C. Petit, F. V. Sotomayor, M. D. Johansen, K. S. Beckham, C. Ritter, F. Girard-Misguich, M. Wilmanns, L. Kremer, F. P. Maurer, and J. L. Herrmann. Conserved and specialized functions of type vii secretion systems in non-tuberculous mycobacteria. *Microbiology (United Kingdom)*, 167(7), 2021.
- [147] A. Leitner, R. Reischl, T. Walzthoeni, F. Herzog, S. Bohn, F. Förster, and R. Aebersold. Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography. *Molecular and Cellular Proteomics*, 11(3):1–12, 2012.
- [148] A. Leitner, T. Walzthoeni, and R. Aebersold. Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline. *Nature Protocols*, 9(1):120–137, 2014.
- [149] P. Lloris-Garcerá, S. Klintner, L. Chen, M. J. Skynner, R. Löving, and J. Frauenfeld. DirectMX – One-step reconstitution of membrane proteins from crude cell membranes into Salipro Nanoparticles. *Frontiers in Bioengineering and Biotechnology*, 8:215, 2020.

-
- [150] Y. Lou, J. Rybniker, C. Sala, and S. T. Cole. EspC forms a filamentous structure in the cell envelope of *Mycobacterium tuberculosis* and impacts ESX-1 secretion. *Molecular Microbiology*, 103(1):26–38, 2017.
- [151] J. Luirink, G. Von Heijne, E. Houben, and J. W. De Gier. Biogenesis of inner membrane proteins in *Escherichia coli*. *Annual Review of Microbiology*, 59:329–355, 2005.
- [152] A. Luthra, A. Mahmood, A. Arora, and R. Ramachandran. Characterization of Rv3868, an essential hypothetical protein of the ESX-1 secretion system in *Mycobacterium tuberculosis*. *Journal of Biological Chemistry*, 283(52):36532–36541, 2008.
- [153] J. A. MacGurn, S. Raghavan, S. A. Stanley, and J. S. Cox. A non-RD1 gene cluster is required for Snm secretion in *Mycobacterium tuberculosis*. *Molecular Microbiology*, 57(6):1653–1663, 2005.
- [154] A. Maciag, E. Dainese, G. M. Rodriguez, A. Milano, R. Provvedi, M. R. Pasca, I. Smith, G. Palù, G. Riccardi, and R. Manganelli. Global analysis of the *Mycobacterium tuberculosis* Zur (FurB) regulon. *Journal of Bacteriology*, 189(3):730–740, 2007.
- [155] G. G. Mahairas, P. J. Sabo, M. J. Hickey, D. C. Singh, and C. K. Stover. Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. *Journal of Bacteriology*, 178(5):1274–1282, 1996.
- [156] J. A. McDonough, J. R. McCann, E. M. E. Tekippe, J. S. Silverman, N. W. Rigel, and M. Braunstein. Identification of functional Tat signal sequences in *Mycobacterium tuberculosis* proteins. *Journal of Bacteriology*, 190(19):6428–6438, 2008.
- [157] A. Mehra, A. Zahra, V. Thompson, N. Sirisaengtaksin, A. Wells, M. Porto, S. Köster, K. Penberthy, Y. Kubota, A. Dricot, D. Rogan, M. Vidal, D. E. Hill, A. J. Bean, and J. A. Philips. *Mycobacterium tuberculosis* type VII secreted effector EsxH targets host ESCRT to impair trafficking. *PLoS Pathogens*, 9(10):e1003734, 2013.
- [158] I. Meuskens, A. Saragliadis, J. C. Leo, and D. Linke. Type V secretion systems: An overview of passenger domain functions. *Frontiers in Microbiology*, 10:1163, 2019.
- [159] N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, and O. Beckstein. MDAAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry*, 32(10):2319–2327, 2011.
- [160] S. Miletic, D. Fahrenkamp, N. Goessweiner-Mohr, J. Wald, M. Pantel, O. Vesper, V. Kotov, and T. C. Marlovits. Substrate-engaged type III secretion system structures reveal gating mechanism for unfolded protein translocation. *Nature Communications*, 12(1):1546, 2021.
-

- [161] A. Mitra, A. Speer, K. Lin, S. Ehrt, and M. Niederweis. PPE surface proteins are required for heme utilization by *Mycobacterium tuberculosis*. *mBio*, 8(1):e01720–16, 2017.
- [162] F. Movahedzadeh and W. Bitter. Ins and Outs of Mycobacterial Plasmids BT - Mycobacteria Protocols: Second Edition. *Methods in molecular biology (Clifton, N.J.)*, 465:217–228, 2009.
- [163] O. Mueller-Cajar. The diverse AAA+ machines that repair inhibited Rubisco active sites. *Frontiers in Molecular Biosciences*, 4:1–17, 2017.
- [164] O. Mueller-Cajar, M. Stotz, P. Wendler, F. U. Hartl, A. Bracher, and M. Hayer-Hartl. Structure and function of the AAA + protein CbbX, a red-type Rubisco activase. *Nature*, 479(7372):194–199, 2011.
- [165] R. Nagamura, M. Fukuda, A. Kawamoto, K. Matoba, N. Dohmae, R. Ishitani, J. Takagi, and O. Nureki. Structural basis for oligomerization of the prokaryotic peptide transporter PepT_{So2}. *Acta Crystallographica Section F*, 75(5):348–358, 2019.
- [166] T. Nakane, A. Kotecha, A. Sente, G. McMullan, S. Masiulis, P. M. Brown, I. T. Grigoras, L. Malinauskaite, T. Malinauskas, J. Miehl, T. Uchański, L. Yu, D. Karia, E. V. Pechnikova, E. de Jong, J. Keizer, M. Bischoff, J. McCormack, P. Tiemeijer, S. W. Hardwick, D. Y. Chirgadze, G. Murshudov, A. R. Aricescu, and S. H. Scheres. Single-particle cryo-EM at atomic resolution. *Nature*, 587(7832):152–156, 2020.
- [167] P. Natale, T. Brüser, and A. J. Driessen. Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane-Distinct translocases and mechanisms. *Biochimica et Biophysica Acta - Biomembranes*, 1778(9):1735–1756, 2008.
- [168] F. Navarro-Garcia, F. Ruiz-Perez, Á. Cataldi, and M. Larzábal. Type VI secretion system in pathogenic escherichia coli: Structure, role in virulence, and acquisition. *Frontiers in Microbiology*, 10(1965), 2019.
- [169] D. Nelson, R. Schuch, P. Chahales, S. Zhu, and V. A. Fischetti. PlyC: A multimeric bacteriophage lysin. *Proceedings of the National Academy of Sciences of the United States of America*, 103(28):10765–10770, 2006.
- [170] M. Newton-Foot, R. M. Warren, S. L. Sampson, P. D. van Helden, and N. C. Gey van Pittius. The plasmid-mediated evolution of the mycobacterial ESX (Type VII) secretion systems. *BMC Evolutionary Biology*, 16(1):62, 2016.
- [171] K. R. Nicholson and P. A. Champion. Bacterial secretion systems : Networks of pathogenic regulation and adaptation in mycobacteria and beyond. *PLoS pathogens*, 18(7):e1010610, 2022.

-
- [172] E. E. Noens, C. Williams, M. Anandhakrishnan, C. Poulsen, M. T. Ehebauer, and M. Wilmanns. Improved mycobacterial protein production using a *Mycobacterium smegmatis* groEL1 Δ C expression strain. *BMC Biotechnology*, 11(1):27, 2011.
- [173] R. Nygaard, J. Kim, and F. Mancina. Cryo-electron microscopy analysis of small membrane proteins. *Current Opinion in Structural Biology*, 64:26–33, 2020.
- [174] Y. M. Ohol, D. H. Goetz, K. Chan, M. U. Shiloh, C. S. Craik, and J. S. Cox. *Mycobacterium tuberculosis* MycP1 protease plays a dual role in regulation of ESX-1 secretion and virulence. *Cell Host and Microbe*, 7(3):210–220, 2010.
- [175] W. H. Organization. *Global tuberculosis report 2021*. World Health Organization, 2021.
- [176] W. H. Organization. *Global tuberculosis report 2022*. World Health Organization, 2022.
- [177] M. Pai, M. A. Behr, D. Dowdy, K. Dheda, M. Divangahi, C. C. Boehme, A. Ginsberg, S. Swaminathan, M. Spigelman, H. Getahun, D. Menzies, and M. Raviglione. Tuberculosis. *Nature Reviews Disease Primers*, 2:16076, 2016.
- [178] D. Pajuelo, U. Tak, L. Zhang, O. Danilchanka, A. D. Tischler, and M. Niederweis. Toxin secretion and trafficking by *Mycobacterium tuberculosis*. *Nature Communications*, 12(1):6592, 2021.
- [179] T. Palmer and B. C. Berks. The twin-arginine translocation (Tat) protein export pathway. *Nature Reviews Microbiology*, 10(7):483–496, 2012.
- [180] T. Palmer, A. J. Finney, C. K. Saha, G. C. Atkinson, and F. Sargent. A holin/peptidoglycan hydrolase-dependent protein secretion system. *Molecular Microbiology*, 115(3):345–355, 2021.
- [181] X. Pang, B. Samten, G. Cao, X. Wang, A. R. Tvinnereim, X. L. Chen, and S. T. Howard. MprAB regulates the espA operon in *Mycobacterium tuberculosis* and modulates ESX-1 function and host cytokine response. *Journal of Bacteriology*, 195(1):66–75, 2013.
- [182] E. Park and T. A. Rapoport. Mechanisms of Sec61SecY-mediated protein translocation across membranes. *Annual Review of Biophysics*, 41(1):21–40, 2012.
- [183] Y. Peng, X. Wang, J. Shou, B. Zong, Y. Zhang, J. Tan, J. Chen, L. Hu, Y. Zhu, H. Chen, and C. Tan. Roles of Hcp family proteins in the pathogenesis of the porcine extraintestinal pathogenic *Escherichia coli* type VI secretion system. *Scientific Reports*, 6:26816, 2016.
-

- [184] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin. UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, 25(13):1605–12, 2004.
- [185] T. H. Phan and E. N. G. Houben. Bacterial secretion chaperones: the mycobacterial type VII case. *FEMS Microbiology Letters*, 365(18):fny197, 2018.
- [186] T. H. Phan, R. Ummels, W. Bitter, and E. N. Houben. Identification of a substrate domain that determines system specificity in mycobacterial type VII secretion systems. *Scientific Reports*, 7:42704, 2017.
- [187] T. H. Phan, L. M. van Leeuwen, C. Kuijl, R. Ummels, G. van Stempvoort, A. Rubio-Canalejas, S. R. Piersma, C. R. Jiménez, A. M. van der Sar, E. N. Houben, and W. Bitter. EspH is a hypervirulence factor for *Mycobacterium marinum* and essential for the secretion of the ESX-1 substrates EspE and EspF. *PLoS Pathogens*, 14(8):e1007247, 2018.
- [188] J. A. Philips and J. D. Ernst. Tuberculosis pathogenesis and immunity. *Annual Review of Pathology: Mechanisms of Disease*, 7:353–384, 2012.
- [189] J. Piton, F. Pojer, S. Wakatsuki, C. Gati, and S. T. Cole. High resolution CryoEM structure of the ring-shaped virulence factor EspB from *Mycobacterium tuberculosis*. *Journal of Structural Biology: X*, 4:100029, 2020.
- [190] J. E. Posey, T. M. Shinnick, and F. D. Quinn. Characterization of the twin-arginine translocase secretion system of *Mycobacterium smegmatis*. *Journal of Bacteriology*, 188(4):1332–1340, 2006.
- [191] C. Poulsen, S. Panjekar, S. J. Holton, M. Wilmanns, and Y. H. Song. WXG100 protein superfamily consists of three subfamilies and exhibits an α -helical C-terminal conserved residue pattern. *PLoS ONE*, 9(2):e89313, 2014.
- [192] N. Poweleit, N. Czudnochowski, R. Nakagawa, D. Trinidad, K. C. Murphy, C. Sasseti, and O. S. Rosenberg. Title: The structure of the endogenous ESX-3 secretion system. *eLife*, 8:e52983, 2019.
- [193] A. Punjani and D. J. Fleet. 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *Journal of Structural Biology*, 213(2):107702, 2021.
- [194] A. Punjani, J. L. Rubinstein, D. J. Fleet, and M. A. Brubaker. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods*, 14(3):290–296, 2017.

-
- [195] A. Punjani, H. Zhang, and D. J. Fleet. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nature Methods*, 17(12):1214–1221, 2020.
- [196] S. A. Rahman, Y. Singh, S. Kohli, J. Ahmad, N. Z. Ehtesham, A. K. Tyagi, and S. E. Hasnain. Comparative analyses of nonpathogenic, opportunistic, and totally pathogenic mycobacteria reveal genomic and biochemical variabilities and highlight the survival attributes of *Mycobacterium tuberculosis*. *mBio*, 5(6):e02020, 2014.
- [197] V. Rantos, K. Karius, and J. Kosinski. Integrative structural modeling of macromolecular complexes using Assemblin. *Nature Protocols*, 17(1):152–176, 2022.
- [198] C. Rapisarda, M. Tassinari, F. Gubellini, and R. Fronzes. Using Cryo-EM to investigate bacterial secretion systems. *Annual Review of Microbiology*, 72:231–254, 2018.
- [199] G. Ratkeviciute, B. F. Cooper, and T. J. Knowles. Methods for the solubilisation of membrane proteins: The micelle-aneous world of membrane protein solubilisation. *Biochemical Society Transactions*, 49(4):1763–1777, 2021.
- [200] M. Reichel, Y. Liao, M. Rettel, C. Ragan, M. Evers, A. M. Alleaume, R. Horos, M. W. Hentze, T. Preiss, and A. A. Millar. In planta determination of the mRNA-binding proteome of arabidopsis etiolated seedlings. *Plant Cell*, 28(10):2435–2452, 2016.
- [201] P. S. Renshaw, K. L. Lightbody, V. Veverka, F. W. Muskett, G. Kelly, T. A. Frenkiel, S. V. Gordon, R. G. Hewinson, B. Burke, J. Norman, R. A. Williamson, and M. D. Carr. Structure and function of the complex formed by the tuberculosis virulence factors CFP-10 and ESAT-6. *EMBO Journal*, 24(14):2491–2498, 2005.
- [202] M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, and G. K. Smyth. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43(7):e47–e47, 2015.
- [203] A. Rivera-Calzada, N. Famelis, O. Llorca, and S. Geibel. Type VII secretion systems: structure, functions and transport models. *Nature Reviews Microbiology*, 19(9):567–584, 2021.
- [204] G. M. Rodriguez, M. I. Voskuil, B. Gold, G. K. Schoolnik, and I. Smith. ideR, an essential gene in *Mycobacterium tuberculosis*: Role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response. *Infection and Immunity*, 70(7):3371–3381, 2002.
- [205] O. S. Rosenberg, D. Dovala, X. Li, L. Connolly, A. Bendebury, J. Finer-Moore, J. Holton, Y. Cheng, R. M. Stroud, and J. S. Cox. Substrates control multimerization and activation of the multi-domain ATPase motor of type VII secretion. *Cell*, 161(3):501–512, 2015.
-

- [206] O. S. Rosenberg, C. Dovey, M. Tempesta, R. A. Robbins, J. S. Finer-Moore, R. M. Stroud, and J. S. Cox. EspR, a key regulator of *Mycobacterium tuberculosis* virulence, adopts a unique dimeric structure among helix-turn-helix proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 108(33):13450–13455, 2011.
- [207] P. B. Rosenthal and J. L. Rubinstein. Validating maps from single particle electron cryomicroscopy. *Current Opinion in Structural Biology*, 34:135–144, 2015.
- [208] B. Saint-Joanis, C. Demangel, M. Jackson, P. Brodin, L. Marsollier, H. Boshoff, and S. T. Cole. Inactivation of Rv2525c, a substrate of the twin arginine translocation (Tat) system of *Mycobacterium tuberculosis*, increases β -lactam susceptibility and virulence. *Journal of Bacteriology*, 188(18):6669–6679, 2006.
- [209] D. Saltzberg, C. H. Greenberg, S. Viswanath, I. Chemmama, B. Webb, R. Pellarin, I. Echeverria, and A. Sali. Modeling biological complexes using integrative modeling platform BT. In M. Bonomi and C. Camilloni, editors, *Biomolecular Simulations: Methods and Protocols*, pages 353–377. Springer New York, New York, NY, 2019.
- [210] M. Sani, E. N. Houben, J. Geurtsen, J. Pierson, K. De Punder, M. Van Zon, B. Wever, S. R. Piersma, C. R. Jiménez, M. Daffé, B. J. Appelmek, W. Bitter, N. D. Van Wel, and P. J. Peters. Direct visualization by Cryo-EM of the mycobacterial capsular layer: A labile structure containing ESX-1-secreted proteins. *PLoS Pathogens*, 6(3):e1000794, 2010.
- [211] S. H. W. Scheres. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*, 180(3):519–530, 2012.
- [212] A. Serafini, F. Boldrin, G. Palù, and R. Manganeli. Characterization of a *Mycobacterium tuberculosis* ESX-3 conditional mutant: Essentiality and rescue by iron and zinc. *Journal of Bacteriology*, 191(20):6340–6344, 2009.
- [213] M. S. Siegrist, M. Unnikrishnan, M. J. McConnell, M. Borowsky, T.-Y. Cheng, N. Siddiqi, S. M. Fortune, D. B. Moody, and E. J. Rubin. Mycobacterial ESX-3 is required for mycobactin-mediated iron acquisition. *Proceedings of the National Academy of Sciences*, 106(44):18792–18797, 2009.
- [214] F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J. D. Thompson, and D. G. Higgins. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7:539, 2011.
- [215] R. Simeone, A. Bobard, J. Lippmann, W. Bitter, L. Majlessi, R. Brosch, and J. Enninga. Phagosomal rupture by *Mycobacterium tuberculosis* results in toxicity and host cell death. *PLoS pathogens*, 8(2):e1002507–e1002507, 2012.

- [216] O. S. Smart, J. G. Neduvellil, X. Wang, B. A. Wallace, and M. S. Sansom. HOLE: A program for the analysis of the pore dimensions of ion channel structural models. *Journal of Molecular Graphics*, 14(6):354–360, 1996.
- [217] J. Smith, J. Manoranjan, M. Pan, A. Bohsali, J. Xu, J. Liu, K. L. McDonald, A. Szyk, N. LaRonde-LeBlanc, and L. Y. Gao. Evidence for pore formation in host cell membranes by ESX-1-secreted ESAT-6 and its role in *Mycobacterium marinum* escape from the vacuole. *Infection and Immunity*, 76(12):5478–5487, 2008.
- [218] S. B. Snapper, R. E. Melton, S. Mustafa, T. Kieser, and W. R. Jr. Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. *Molecular Microbiology*, 4(11):1911–1919, 1990.
- [219] P. Soler-Arnedo, C. Sala, M. Zhang, C. Stewart T., and J. Piton. Polarly localized EccE1 is required for ESX-1 function and stabilization of ESX-1 membrane proteins in *Mycobacterium tuberculosis*. *Journal of Bacteriology*, 202(5):e00662–19, 2020.
- [220] M. Solomonson, P. F. Huesgen, G. A. Wasney, N. Watanabe, R. J. Gruninger, G. Prehna, C. M. Overall, and N. C. Strynadka. Structure of the mycosin-1 protease from the mycobacterial ESX-1 protein type VII secretion system. *Journal of Biological Chemistry*, 288(24):17782–17790, 2013.
- [221] M. Solomonson, D. Setiাপutra, K. A. Makepeace, E. Lameignere, E. V. Petrotchenko, D. G. Conrady, J. R. Bergeron, M. Vuckovic, F. Dimaio, C. H. Borchers, C. K. Yip, and N. C. Strynadka. Structure of EspB from the ESX-1 type VII secretion system and insights into its export mechanism. *Structure*, 23(3):571–583, 2015.
- [222] O. Spitz, I. N. Erenburg, T. Beer, K. Kanonenberg, I. B. Holland, and L. Schmitt. Type I secretion systems—one mechanism for all? *Protein Secretion in Bacteria*, 7(2):215–225, 2019.
- [223] S. A. Stanley, S. Raghavan, W. W. Hwang, and J. S. Cox. Acute infection and macrophage subversion by *Mycobacterium tuberculosis* require a specialized secretion system. *Proceedings of the National Academy of Sciences*, 100(22):13001–13006, 2003.
- [224] E. J. Stoop, T. Schipper, S. K. Rosendahl Huber, A. E. Nezhinsky, F. J. Verbeek, S. S. Gurcha, G. S. Besra, C. M. Vandenbroucke-Grauls, W. Bitter, and A. M. Van Der Sar. Zebrafish embryo screen for mycobacterial genes involved in the initiation of granuloma formation reveals a newly identified ESX-1 component. *DMM Disease Models and Mechanisms*, 4(4):526–536, 2011.
- [225] M. Strong, M. R. Sawaya, S. Wang, M. Phillips, D. Cascio, and D. Eisenberg. Toward the structural genomics of complexes: Crystal structure of a PE/PPE protein complex

- from *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences*, 103(21):8060–8065, 2006.
- [226] D. Sun, Q. Liu, Y. He, C. Wang, F. Wu, C. Tian, and J. Zang. The putative propeptide of MycP1 in mycobacterial type VII secretion system does not inhibit protease activity but improves protein stability. *Protein and Cell*, 4(12):921–931, 2013.
- [227] D. J. Swainsbury, S. Scheidelaar, N. Foster, R. van Grondelle, J. A. Killian, and M. R. Jones. The effectiveness of styrene-maleic acid (SMA) copolymers for solubilisation of integral membrane proteins from SMA-accessible and SMA-resistant membranes. *Biochimica et Biophysica Acta - Biomembranes*, 1859(10):2133–2143, 2017.
- [228] T. A. Sysoeva. Assessing heterogeneity in oligomeric AAA+ machines. *Cellular and Molecular Life Sciences*, 74(6):1001–1018, 2017.
- [229] U. Tak, T. Dokland, and M. Niederweis. Pore-forming Esx proteins mediate toxin secretion by *Mycobacterium tuberculosis*. *Nature Communications*, 12(1):1–17, 2021.
- [230] F. Tekaiia, S. V. Gordon, T. Garnier, R. Brosch, B. G. Barrell, and S. T. Cole. Analysis of the proteome of *Mycobacterium tuberculosis* in silico. *Tubercle and Lung Disease*, 79(6):329–342, 1999.
- [231] T. C. Terwilliger. Finding non-crystallographic symmetry in density maps of macromolecular structures. *Journal of Structural and Functional Genomics*, 14(3):91–95, 2013.
- [232] C. Tribet, R. Audebert, and J. L. Popot. Amphipols: Polymers that keep membrane proteins soluble in aqueous solutions. *Proceedings of the National Academy of Sciences of the United States of America*, 93(26):15047–15050, 1996.
- [233] J. A. M. Tufariello, J. R. Chapman, C. A. Kerantzas, K. W. Wong, C. Vilchèze, C. M. Jones, L. E. Cole, E. Tinaztepe, V. Thompson, D. Fenyö, M. Niederweis, B. Ueberheide, J. A. Philips, and W. R. Jacobs. Separable roles for *Mycobacterium tuberculosis* ESX-3 effectors in iron acquisition and virulence. *Proceedings of the National Academy of Sciences of the United States of America*, 113(3):e348–e357, 2016.
- [234] M. Tullius, S. Nava, and M. Horwitz. PPE37 is essential for *Mycobacterium tuberculosis* heme-iron acquisition (HIA), and a defective PPE37 in *Mycobacterium bovis* BCG prevents HIA. *Infection and Immunity*, 87(2):e00540–18, 2019.
- [235] M. van Kempen, S. S. Kim, C. Tumescheit, M. Mirdita, J. Söding, and M. Steinegger. Foldseek: fast and accurate protein structure search. *bioRxiv*, page 2022.02.07.479398, 2022.

- [236] V. J. van Winden. *The mycosins of pathogenic mycobacteria: digesting the role of conserved proteases in type VII secretion*. PhD thesis, Vrije Universiteit Amsterdam, 2021.
- [237] V. J. van Winden, C. M. Bunduc, R. Ummels, W. Bitter, and E. N. Houben. A chimeric EccB-MycP fusion protein is functional and a stable component of the ESX-5 type VII secretion system membrane complex. *Journal of Molecular Biology*, 432(4):1265–1278, 2020.
- [238] V. J. van Winden, M. P. Damen, R. Ummels, W. Bitter, and E. N. Houben. Protease domain and transmembrane domain of the type VII secretion mycosin protease determine system-specific functioning in mycobacteria. *Journal of Biological Chemistry*, 294(13):4806–4814, 2019.
- [239] V. J. C. van Winden, R. Ummels, S. R. Piersma, C. R. Jiménez, K. V. Korotkov, W. Bitter, and E. N. G. Houben. Mycosins Are Required for the Stabilization of the ESX-1 and ESX-5 Type VII Secretion Membrane Complexes. *mBio*, 7(5):e01471–16, 2016.
- [240] A. T. Vincent, S. Nyongesa, I. Morneau, M. B. Reed, E. I. Tocheva, and F. J. Veyrier. The mycobacterial cell envelope: A relict from the past or the result of recent evolution? *Frontiers in Microbiology*, 9:2341, 2018.
- [241] S. Viswanath, I. E. Chemmama, P. Cimermancic, and A. Sali. Assessing exhaustiveness of stochastic sampling for integrative modeling of macromolecular structures. *Biophysical Journal*, 113(11):2344–2353, 2017.
- [242] J. M. Wagner, S. Chan, T. J. Evans, S. Kahng, J. Kim, M. A. Arbing, D. Eisenberg, and K. V. Korotkov. Structures of EccB1 and EccD1 from the core complex of the mycobacterial ESX-1 type VII secretion system. *BMC Structural Biology*, 16(1):5, 2016.
- [243] J. M. Wagner, T. J. Evans, J. Chen, H. Zhu, E. N. Houben, W. Bitter, and K. V. Korotkov. Understanding specificity of the mycosin proteases in ESX/type VII secretion by structural and functional analysis. *Journal of Structural Biology*, 184(2):115–128, 2013.
- [244] Q. Wang, H. I. Boshoff, J. R. Harrison, P. C. Ray, S. R. Green, P. G. Wyatt, and C. E. Barry. PE/PPE proteins mediate nutrient transport across the outer membrane of *Mycobacterium tuberculosis*. *Science*, 367(6482):1147–1151, 2020.
- [245] S. Wang, K. Zhou, X. Yang, B. Zhang, Y. Zhao, Y. Xiao, X. Yang, H. Yang, L. W. Guddat, J. Li, and Z. Rao. Structural insights into substrate recognition by the type VII secretion system. *Protein and Cell*, 11(2):124–137, 2020.

- [246] Y. Wang, Y. Tang, C. Lin, J. Zhang, J. Mai, J. Jiang, X. Gao, Y. Li, G. Zhao, L. Zhang, and J. Liu. Crosstalk between the ancestral type VII secretion system ESX-4 and other T7SS in *Mycobacterium marinum*. *iScience*, 25(1):103585, 2022.
- [247] B. J. Wards, G. W. De Lisle, and D. M. Collins. An *esat6* knockout mutant of *Mycobacterium bovis* produced by homologous recombination will contribute to the development of a live tuberculosis vaccine. *Tubercle and Lung Disease*, 80(4-5):185–189, 2000.
- [248] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. De Beer, C. Rempfer, L. Bordoli, R. Lepore, and T. Schwede. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research*, 46(W1):W296–W303, 2018.
- [249] B. Y. Watkins, S. A. Joshi, D. A. Ball, H. Leggett, S. Park, J. Kim, C. D. Austin, A. Paler-Martinez, M. Xu, K. H. Downing, and E. J. Brown. *Mycobacterium marinum* SecA2 promotes stable granulomas and induces tumor necrosis factor alpha in Vivo. *Infection and Immunity*, 80(10):3512–3520, 2012.
- [250] B. Webb and A. Sali. Comparative protein structure modeling using MODELLER. *Current Protocols in Bioinformatics*, 2016(June):5.6.1–5.6.37, 2016.
- [251] T. Werner, G. Sweetman, M. F. Savitski, T. Mathieson, M. Bantscheff, and M. M. Savitski. Ion coalescence of neutron encoded TMT 10-plex reporter ions. *Analytical Chemistry*, 86(7):3594–3601, 2014.
- [252] Z. A. Williamson, C. T. Chaton, W. A. Ciocca, N. Korotkova, and K. V. Korotkov. PE5–PPE4–EspG3 heterotrimer structure from mycobacterial ESX-3 secretion system gives insight into cognate substrate recognition by ESX systems. *Journal of Biological Chemistry*, 295(36):12706–12715, 2020.
- [253] E. L. Wu, X. Cheng, S. Jo, H. Rui, K. C. Song, E. M. Dávila-Contreras, Y. Qi, J. Lee, V. Monje-Galvan, R. M. Venable, J. B. Klauda, and W. Im. CHARMM-GUI membrane builder toward realistic biological membrane simulations. *Journal of Computational Chemistry*, 35(27):1997–2004, 2014.
- [254] V. Yeh, A. Goode, and B. B. Bonev. Membrane protein structure determination and characterisation by solution and solid-state nmr. *Biology*, 9(11):1–23, 2020.
- [255] K. M. Yip, N. Fischer, E. Paknia, A. Chari, and H. Stark. Atomic-resolution protein structure determination by cryo-EM. *Nature*, 587(7832):157–161, 2020.
- [256] X. L. Zhang, D. F. Li, J. Fleming, L. W. Wang, Y. Zhou, D. C. Wang, X. E. Zhang, and L. J. Bi. Core component EccB1 of the *Mycobacterium tuberculosis* type VII secretion system is a periplasmic ATPase. *FASEB Journal*, 29(12):4808–4814, 2015.

- [257] J. Zimmer, Y. Nam, and T. A. Rapoport. Structure of a complex of the ATPase SecA and the protein-translocation channel. *Nature*, 455(7215):936–943, 2008.

Acknowledgment

This thesis would not have been possible without the support of many people along the way. I am very grateful that I met many different enthusiastic, talented and inspiring scientists who helped to shape me as a scientist, but also as a person.

First of all, I would like to express my gratitude to Matthias Wilmanns for his guidance throughout the four years. He entrusted me this exciting but also very challenging project and encouraged me to pursue my ideas and present my research. I would also like to thank Annabel Parret, for welcoming me in the TB subgroup and for introducing me to the TB field. Her supervision within the first year of my PhD has paved the way for many discoveries presented in this thesis.

I am especially thankful for my supervisor, teammate and friend Kate Beckham. Together, we have gone with the project through ups and downs, but always managed to keep a good spirit up. I could always count on her support scientifically and non-scientifically and cannot imagine a better team-player and supervisor than her. I am very happy that I won't be the last person she will guide through a PhD.

I would also like to thank my thesis advisory committee Jan Kosinski, Aymelt Itzen and Eva Kowalinski who guided me to focus on the important scientific questions and shared their ideas.

Of course, I would also like to acknowledge all the fantastic collaborators that contributed to the project: Grzegorz Chojnowski, for his support in model building and refinement and his enthusiasm about new bioinformatic tools, Jan Kosinski and Daniel Ziemianowicz for building integrative models and providing new ideas and Edukondalu Mullapudi, Anthony Chen and Simon Mortensen for their assistance regarding all EM related topics and their patience when teaching me at the microscope. Thanks also to Mandy Rettel and Frank Stein from the proteomics core facility, for setting up the secretion assay and for the numerous fruitful discussions with Joel Selkrig.

I have been extremely lucky to be part of different subgroups and I would like to thank all the members and subgroup leaders, past and present, for the many productive brainstorming and troubleshooting sessions or just informal chats. Also, I would like to emphasize the support of

different facilities, the SPC, PCF, the cryo-EM facility but of course also the Wetlab-Team, especially Claudia Litz and Sonja Staack, who did not only keep the wetlab up and running but also supported me whenever I needed help.

There are many more people, who I would like to thank as they made my 4-year long journey fun and special: fellows and staff, group members (former and present), administration, IT, lunch group members (former and present) and office mates (only former). Too numerous, to mention everyone by name but without all of them, I would not have been able to present the project as it is.

Finally I want to express my deepest gratitude to my family and friends, especially my parents, my brother and of course Malte for their support, love, understanding and patience not only during the last four years.

List of publications

Within the course of this PhD, I contributed to the following publications:

Conserved and specialized functions of Type VII secretion systems in non-tuberculous mycobacteria.

Lagune M*, Petit C,* Sotomayor FV*, Johansen MD*, Beckham KSH, Ritter C, Girard-Misguich F, Wilmanns M, Kremer L, Maurer FP, Herrmann JL *Microbiology (Reading, England)*, 2021

doi:10.1099/mic.0.001054.

Structure of the mycobacterial ESX-5 type VII secretion system pore complex.

Beckham KSH*, Ritter C*, Chojnowski G*, Ziemianowicz DS, Mullapudi E, Rettel M, Savitski MM, Mortensen SA, Kosinski J, Wilmanns M *Science advances*, 2021

doi:10.1126/sciadv.abg9923.

* = equally contributing authors

Eidesstattliche Versicherung

Ich versichere ausdrücklich, dass ich die Arbeit selbständig und ohne fremde Hilfe verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die aus den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen einzeln nach Ausgabe (Auflage und Jahr des Erscheinens), Band und Seite des benutzten Werkes kenntlich gemacht habe. Ferner versichere ich, dass ich die Dissertation bisher nicht einem Fachvertreter an einer anderen Hochschule zur Überprüfung vorgelegt oder mich anderweitig um Zulassung zur Promotion beworben habe. Ich erkläre mich einverstanden, dass meine Dissertation vom Dekanat der Medizinischen Fakultät mit einer gängigen Software zur Erkennung von Plagiaten überprüft werden kann.

Hamburg, den _____ Unterschrift: _____