



UNIVERSITÄT ZU LÜBECK

From the Institute of Physics
of the University of Lübeck
Director: Prof. Dr. Christian G. Hübner

**Biophysical characterization and analysis of different
mycobacterial WXG100 proteins to study possible inter-loci
complex formation.**

Dissertation
for Fulfillment of
Requirements for the
Doctoral Degree
of the University of Lübeck

from the Department of Natural Sciences

Submitted by

Christina Jana Luise Kallenberg
from Neumarkt i. d. Oberpfalz

Lübeck 2025

First referee: Prof. Dr. rer. nat. Christian Hübner
Second referee: Prof. Dr. rer. nat. Thomas Gutschmann
Date of oral exam: 29th of September 2025

Approved for printing.
Lübeck, 6th of October 2025

Acknowledgements

First and foremost, I would like to thank my doctoral supervisor, Prof. Dr. Hübner, very much. I am deeply indebted to him for his competent supervision, motivating scientific suggestions, and active support. His confidence in my work and valuable advice have contributed significantly to the success of this dissertation.

I would also like to thank Dr. Young-Hwa Song for her support, professional expertise, and valuable advice. Her constructive feedback and open ear helped me a lot during my doctoral studies. Her encouragement was crucial in persuading me to pursue a doctorate.

I would also like to thank Prof. Dr. Taube from the Institute of Virology and Prof. Dr. Peters from the Institute of Chemistry, who both gave me the opportunity to work as a guest at their institutes. The use of their infrastructure and the scientific exchange were invaluable for my research.

My special thanks go to Prof. Dr. Gutschmann for the interesting and enriching collaborations. Also, I greatly appreciate that he agreed to act as a second reviewer for my dissertation.

I would like to thank the entire Institute of Physics for the pleasant working atmosphere. I would particularly like to emphasize my colleagues Dr. Verena Hirschfeld, Christian Magaña, Christian Rönnau, Joana Weiler, and Kim Reiter, who were a great support to me not only professionally but also personally. The friendly cooperation, the stimulating discussions, and their encouraging words have enriched my doctoral period and made it unforgettable.

Thank you very much!

Abstract

Tuberculosis has remained a global public health concern for the past 30 years. The causative agent *Mycobacterium tuberculosis* has developed several strategies to evade the host immune system and for intracellular survival. Extensive genomic and proteomic studies have revealed hundreds of genes involved in the bacterium's resistance and survival. The first known secreted virulence factors, EsxA and EsxB, are the paragons of a group of heterologous proteins, termed the "WXG100 family". *M. tuberculosis* has 23 WXG100 proteins (EsxA-EsxW), five of such protein pairs are found within a genomic cluster making up a type-VII-secretion system (ESX-1 to ESX-5). Mycobacterial WXG100 proteins appear to share a similar secondary structure and form complexes with neighboring co-expressed proteins. The question arises as to whether Esx proteins from different genomic loci, i.e. not within the same operon, can form complexes. If this were the case, even only for some, the recombination of individual complex partners could greatly increase the versatility of the bacterium in reacting to the numerous host defense mechanisms.

In this work, ten different mycobacterial Esx proteins were recombinantly expressed and purified in order to evaluate the possibility of inter-loci complex formation and their structural properties. The results suggest that inter-loci complex formation between several Esx proteins is possible. Special focus was put on the complex formation between EsxB and EsxT, which was evaluated with chemical crosslinking studies and single-molecule FRET experiments. Kinetic evaluations suggest a rather low binding constant between EsxB and EsxT (mM range), compared to the high binding constant between the native partners EsxB and EsxA (nM range). The proteins EsxF and EsxE were found to be highly heat resistant, both in complex and monomeric form, with a high structural recovery of over 90 % after heating to 96 °C. In addition, evidence of homo-dimer or homo-oligomer formation was found for EsxF and EsxU, which has not been reported previously.

Zusammenfassung

Die Tuberkulose ist seit mehr als 30 Jahren ein Problem der globalen öffentlichen Gesundheit. Der Erreger *Mycobacterium tuberculosis* hat mehrere Strategien entwickelt, um dem Immunsystem des Wirts zu entgehen und intrazellulär zu überleben. Umfangreiche genomische und proteomische Studien haben zahlreiche Genen aufgedeckt, die an der Resistenz und dem Überleben des Bakteriums beteiligt sind. Die ersten bekannten sekretierten Virulenzfaktoren, EsxA und EsxB, sind die Paradebeispiele einer Gruppe heterologer Proteine, die als "WXG100-Familie" bezeichnet werden. *M. tuberculosis* besitzt 23 WXG100-Proteine (EsxA-EsxW), wobei fünf dieser Proteinpaaire in einem genomischen Cluster zu finden sind, welches ein Typ-VII-Sekretionssystem bildet (ESX-1 bis ESX-5). Mykobakterielle WXG100-Proteine scheinen eine ähnliche Sekundärstruktur zu haben und bilden Komplexe mit benachbarten und co-exprimierten Proteinen. Es stellt sich die Frage, ob Esx-Proteine von verschiedenen genomischen Loci, d. h. nicht innerhalb desselben Operons, Komplexe bilden können. Wenn dies der Fall wäre, und sei es auch nur für einige, könnte die Rekombination einzelner Komplexpartner die Vielseitigkeit des Bakteriums bei der Reaktion auf die zahlreichen Abwehrmechanismen des Wirts erheblich steigern.

In dieser Arbeit wurden zehn verschiedene mykobakterielle Esx-Proteine rekombinant exprimiert und aufgereinigt, um deren strukturelle Eigenschaften zu untersuchen und die Möglichkeit der inter-loci Komplexbildung zu evaluieren. Die Ergebnisse deuten darauf hin, dass eine Komplexbildung auch zwischen Esx-Proteinen von unterschiedlichen ESX-Clustern möglich ist. Besonderes Augenmerk wurde auf die Komplexbildung zwischen EsxB und EsxT gelegt,

die mit chemischen Crosslinkingstudien und Einzelmolekül-FRET-Experimenten untersucht wurde. Kinetische Untersuchungen deuten auf eine eher niedrige Bindungskonstante zwischen EsxB und EsxT hin (mM-Bereich), verglichen mit der hohen Bindungskonstante zwischen den nativen Partnern EsxB und EsxA (nM-Bereich). Die Proteine EsxF und EsxE erwiesen sich als sehr hitzebeständig, sowohl im Komplex als auch in monomerer Form, mit einer hohen Rückfaltungsrate von über 90 % nach Erhitzen auf 96 °C. Darüber hinaus wurden für EsxF und EsxU Hinweise für die Bildung von Homo-Dimeren oder Homo-Oligomeren gefunden, was bisher nicht berichtet wurde.

Table of Contents

1	Summary	1
2	Introduction	3
2.1	Mycobacteria and Tuberculosis	3
2.1.1	Mycobacteria	4
2.1.2	The Mycobacterial Cell Wall	4
2.1.3	The Type-VII-Secretion System	7
2.1.4	Esx Proteins	11
2.1.5	Tuberculosis	11
2.2	Protein Biochemistry	16
2.2.1	Protein Structure and Protein Folding	16
2.2.2	Protein Interaction and Protein Binding	17
2.2.3	Thermodynamics of Proteins	19
2.2.4	Protein Bioconjugation	20
2.3	Förster Resonance Energy Transfer	23
2.3.1	Principles of Fluorescence	23
2.3.2	Characteristics of Fluorophores	25
2.3.3	Resonance Energy Transfer	25
2.4	Fluorescence Correlation Spectroscopy	27
2.4.1	Correlation Analysis	27
2.5	Circular Dichroism Spectroscopy	29
2.5.1	Principles of CD Spectroscopy	30
2.5.2	Interpretation of a CD Spectrum	33
3	Materials and Methods	35
3.1	Esx Protein Production	35
3.1.1	Cloning and Production of pMyCA Expression Vectors	35
3.1.2	Protein Expression in <i>Mycobacterium smegmatis</i>	37
3.1.3	Protein Purification	37
3.1.4	Protein Analysis	38
3.2	Protein Processing	38
3.2.1	Crosslinking Studies	38
3.2.2	Fluorescent Labeling	39
3.3	CD Analysis	41
3.3.1	Sample Preparation	41
3.3.2	Experimental Setup and Measurement Parameters	41
3.3.3	Data Analysis	42
3.4	FRET Analysis	44
3.4.1	Sample Preparation	44

3.4.2	Experimental Setup and Measurement Parameters	45
3.4.3	Titration Measurements	46
3.4.4	smFRET under Crosslinking Conditions	46
3.4.5	Data Analysis	47
4	Results	51
4.1	Protein Purification	51
4.1.1	Detection of Apparent Homo-multimers Consisting Solely of EsxF	52
4.1.2	Indication of Membrane-association of EsxU/T and of Formation of Homo-multimers by EsxU	53
4.2	Crosslinking Experiments	54
4.3	Characterization of EsxF/E	57
4.3.1	CD Spectroscopy Reveals Thermodynamics and a High Heat-resistance of the EsxF/E Complex and its Monomers	57
4.3.2	smFRET: Fluorescently Labeled EsxF Suggests Structured Monomer	62
4.4	Characterization of EsxU/T from ESX4	64
4.4.1	The Complex EsxU/T and its Monomers are Sensitive to Heat Treatment	64
4.5	Singe Molecule FRET (smFRET) Experiments	69
4.5.1	EsxA Titration to Study Binding Kinetics	69
4.5.2	EsxT Titration to Study Inter-loci Complex Formation	74
4.5.3	Combining Crosslinking with smFRET	76
4.6	Effect of pH on the EsxB/A Complex and its Monomers	78
4.7	Phylogenetic Assessment of EsxU and EsxF	80
5	Discussion	83
5.1	Kinetic Binding Studies with smFRET	83
5.1.1	EsxA Titration	83
5.1.2	EsxT Titration	86
5.2	Characterization of EsxF/E and EsxU/T	87
5.2.1	Biophysical and Structural Characterization of EsxF/E and EsxU/T	88
5.2.2	Indications for Homodimers of EsxF and EsxU	91
5.3	Conclusions and Outlook	93
A	Supplementary Figures	i
B	List of Figures	iv
C	List of Tables	vi
D	Curriculum Vitae	vii

List of Abbreviations

AF488	Alexa Fluor [®] 488
AF647	Alexa Fluor [®] 647
APD	avalanche photodiode
APS	ammonium persulfate
CD (spectroscopy)	circular dichroism (spectroscopy)
CFP-10	10 kDa culture filtrate protein, EsxB
CV	column volume
DMSO	dimethyl sulfoxide
DTT	dithiothreitol
EDC	N-(3-Dimethylaminopropyl)-N'-Ethylcarbodiimide
ESAT-6	6 kDa early secretory antigen target, EsxA
ESX	ESAT-6 secretion system
FCS	fluorescence correlation spectroscopy
FRET	Förster resonance energy transfer
IC	internal conversion
ISC	intersystem crossing
MB	molecular brightness
MBTC	<i>Mycobacterium tuberculosis</i> complex
MRE	mean residue ellipticity
MRW	mean residue weight
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PMMA	polymethyl methacrylate
PTFE	polytetrafluoroethylene
RD	region of difference
RGM	rapidly-growing mycobacteria
SDS	sodium dodecyl sulfate
SGM	slow-growing mycobacteria

TB	tuberculosis
TCEP	Tris(2-carboxyethyl)phosphin
TEV (protease)	Tobacco Etch Virus (protease)
TRIS	Tris(hydroxymethyl)aminomethane
VR	vibrational relaxation
sNHS	sulfo-N-hydroxysuccinimide
sm	single molecule

Chapter 1

Summary

Although treatable with antibiotics, tuberculosis (TB) is still one of the three leading causes of death from a single infectious agent, causing even twice as many deaths as HIV/AIDS [1]. The disease burden is highest in low-income countries with inadequate health care. With currently around 10 million new infections every year and a high mortality of around 50 % if left untreated [1], TB has remained a global public health emergency for the past 30 years [2]. Caused by the bacterium *Mycobacterium tuberculosis*, the disease affects primarily the lungs and is spread by pathogen-laden aerosols expelled by coughing. It is estimated that only about 5 % of all infected people develop active tuberculosis within two years of infection [1]. Although some patients will clear the infection, most will develop the latent form of tuberculosis (LTB). This poses a problem in the fight against the disease, as latency can go undetected for decades before an outbreak occurs. Despite significant advances in diagnostic tests in recent years, sputum smear microscopy is still widely used in low- and middle-income countries. A drawback of this method is that it only detects active TB. The latent disease is diagnosed mainly with the tuberculin skin test (TST) or an interferon- γ release assay (IGRA). However, these two immune-based approaches cannot accurately differentiate between latent infection, resolved infection with acquired immunity and active tuberculosis, and often have low sensitivity [3].

The quest for adequate and effective prevention, treatment, and diagnosis calls for a thorough understanding of the pathogenesis. In particular, the host-pathogen interaction is a central focus of TB research, as *M. tuberculosis* has developed several strategies to evade the host immune system. As the bacteria are able to switch between replication during active TB and a non-replicating (NRP) state during LTB, modulation of gene expression is another area of great research interest. Extensive genomic and proteomic studies have revealed hundreds of genes involved in the intracellular survival of the bacteria [4]. Among the first known secreted virulence factors are the proteins EsxA and EsxB. They belong to the group of WXG100 proteins, many of which have been shown to be essential for bacterial growth and survival within the host [5]. In total, *M. tuberculosis* has 23 such heterologous WXG100 proteins (EsxA-EsxW). Five of these bi-cistronically encoded protein pairs are found within a genomic cluster containing genes encoding a type-VII-secretion system (T7SS), which is responsible for the secretion of the respective protein pair. For those WXG100 proteins from *M. tuberculosis*, whose structure has been determined, it has been shown that they form heterodimers. Each monomer adopts a helix-turn-helix conformation and their antiparallel association results in a four-helical bundle (EsxB/A [5], EsxG/H [6], EsxO/P [7]). The proteins EsxR/S exist in a hetero-dimeric and a hetero-tetrameric form [8]. Such multi-helical

heteromers are thought to be the functional conformation adopted by all Esx homologs. While WXG100 homodimers are found in other organisms (*sagEsxA* in *Streptococcus agalactiae* [5] and *sauEsxA* in *Staphylococcus aureus* [9]), none have been detected in *M. tuberculosis* so far. The fact that the WXG100 proteins appear to share a similar secondary structure and form complexes with neighboring co-expressed proteins makes them an intriguing target for research, given their involvement in a range of essential growth functions and virulence. As of July 2024, the function of most Esx proteins remains unknown, especially those that do not belong to one of the five secretion clusters (ESX-1 to ESX-5). Some of the WXG100 proteins are known to be involved in the adaptation of the pathogen to intracellular survival in the host and in the transition between dormancy and replication [10,11], making them potential targets for new diagnostic approaches, drugs or vaccines.

Assuming that Esx proteins are so abundant due to their importance for bacterial survival and growth, the ability to up- and down-regulate 11 protein pairs provides *M. tuberculosis* with a wide range of environmental adaptations. The question arises as to whether Esx proteins from different genomic loci, i.e. not within the same operon, interact and may even be able to form complexes. If this were the case, even for only some of the Esx proteins, the recombination of individual complex partners could greatly increase the versatility of the bacterium.

The aim of this work was to recombinantly express and purify different mycobacterial Esx proteins and to evaluate their structural properties and the possibility of inter-loci complex formation. The well-known proteins EsxB and EsxA were used to establish and adapt the method of single-molecule FRET (smFRET) for the measurement of binding kinetics. This method was then used to explore the possibility of inter-loci interactions and complex formation between EsxB and EsxT. Biophysical analyses of the less well-characterized proteins EsxU, EsxT, EsxF and EsxE were performed to assess secondary structure and heat resistance. These methods were also used to study the binding dynamics between EsxA and EsxB under different pH conditions, mimicking environmental conditions that may be encountered during the host immune response.

The results of this work suggest that inter-loci complex formation between a variety of Esx proteins is possible. Special focus was put on the complex formation between EsxB and EsxT, which was evaluated with chemical crosslinking studies and smFRET. Kinetic evaluations suggest a rather low binding constant between EsxB and EsxT, compared to the high binding constant between the native partners EsxB and EsxA. The proteins EsxF and EsxE were found to be highly heat resistant, both in complex and monomeric form, with a high structural recovery of over 90 % after heating to 96 °C. In addition, evidence of homo-dimer or homo-oligomer formation was found for EsxF and EsxU, which has not been reported previously. The detailed information in Chapter **2 Introduction** is intended to familiarize the reader with the biological background of mycobacteria and tuberculosis, as well as the biophysical and biochemical concepts underlying the methods and theoretical models used in this work. The steps of the individual experiments and details of their evaluation are given in Chapter **3 Material and Methods**. The findings of this work are presented in Chapter **4 Results** and are evaluated in the light of the current state of knowledge in Chapter **5 Discussion**. Throughout this work, the nomenclature of the proteins will follow that proposed by Bitter et al. in 2009 [12].

Chapter 2

Introduction

2.1 Mycobacteria and Tuberculosis

In the late 19th century, the prevalent miasma theory suffered a great setback when the German physician and microbiologist Robert Koch published his findings on the Etiology of Tuberculosis in 1882 [13]. In this important work, Koch describes thin, rod-shaped bacteria, named *Mycobacterium tuberculosis* (Mtb), to cause tuberculosis; a deadly infectious disease which, according to him, is to be lined up even before the plague and cholera. Today, more than 140 years after Koch's groundbreaking publication, tuberculosis (TB) still ranks as one of the leading causes of death by a single pathogen worldwide¹, even surpassing HIV/AIDS. Meanwhile, the disease burden has declined greatly in western, developed countries in central Europe and North America (incidence below 10 per 100,000 population per year) [14]. High-incidence countries are found mainly in sub-Saharan Africa and south-east Asia, with India contributing just short of a third of the world's incident cases in 2021 [14]. While pulmonary TB (PTB) is the dominant form of the disease, it can disseminate and infect other parts of the body. This manifestation is called extrapulmonary TB (EPTB) and represented around 16 % of TB incident cases in 2019 [15]. It affects the lymph nodes most frequently, but an EPTB can also be established in the central nervous system, the spine, the genitourinary system, abdominal tissues and organs, and the bones and muscles [16].

To this day, the only available preventive immunization is the live attenuated *Mycobacterium bovis* bacillus Calmette-Guérin (BCG) bacterial vaccine strain, first used almost a century ago [17]. The World Health Organization (WHO) recommends neonatal vaccination as an important part of the TB prevention strategy. The vaccination's efficiency has been analyzed in a number of studies and varies considerably between populations. Depending on the study, protection of the BCG vaccine against pulmonary TB ranges from 44 % to 99 %. Generally, it was observed that the protection against infection with Mtb and against the development of PTB during childhood is highest when individuals are immunized as neonates [17]. The protection duration also varies strongly. While protection against PTB and EPTB is known to decline over time, some studies suggest that it lasts for up to ten years [18]. The treatment of active pulmonary tuberculosis, as recommended by the WHO, is

¹The 2022 Global Tuberculosis Report by the WHO states that TB was the leading cause of death by a single pathogen worldwide up until the coronavirus (COVID-19) pandemic. It is estimated that during the years 2020 and 2021, TB ranked as the second leading cause of death from a single infectious agent after COVID-19. Further, it is explained that the estimation of the TB disease burden during the pandemic was especially difficult in 2020 and 2021 since the reporting of new TB infections has declined drastically during these years.

a six-month long regime with four of the first-line anti-TB drugs isoniazid, rifampicin, ethambutol, and pyrazinamide [19]. However, the long duration of treatment with four different drugs makes it prone to noncompliance and misuse or mis-administration of drugs, which often results in the development of drug-resistant Mtb strains. The emergence of multidrug-resistant TB (MDR-TB) and even extensively drug-resistant TB (XDR-TB) has become a major concern in the past years².

2.1.1 Mycobacteria

The numerous different mycobacterial species are traditionally grouped into two groups according to their growth characteristics. Those species that grow colonies in less than seven days on solid medium were grouped together as rapidly-growing mycobacteria (RGM) [20], whereas the others are called slow-growing mycobacteria (SGM). All major human pathogens, *M. leprae*, *M. ulcerans* and tuberculosis-causing mycobacteria belong to the slow-growing group. Tuberculosis in humans and other mammals is caused by members of the so-called *Mycobacterium tuberculosis* complex (MTBC). This is a group of closely related SGM and includes the human-adapted pathogens *M. tuberculosis* and *M. africanum* and the animal-adapted species *M. bovis*, *M. orygis*, *M. microti*, *M. caprae* and others. These bacilli have most likely evolved in Africa from a common environmental ancestor, closely resembling the present strain of *M. canettii* [21, 22]. Those species that neither cause tuberculosis nor leprosy are referred to as non-tuberculous mycobacteria (NTMs) [23]. They can also cause disease, albeit only in immune-compromised individuals, and include *M. avium*, *M. marinum*, *M. xenopi*, *M. goodnae*, *M. kansasii* (all SGM) and the fast-growing *M. abscessus*.

Members of the genus *Mycobacterium* have a number of unique characteristics, both genetically as well as phenotypically. The genome's high G+C content, ranging from 62 % to 70 %, is a clear difference from most other bacteria [24]. Regarding the phenotypical characteristics, it was observed early on that mycobacteria are not to be stained with the usual Gram-method. Instead, in the early 1880s, a new staining method for these bacilli was proposed by Robert Koch [13] and improved further by the physician Paul Ehrlich, the bacteriologist Franz Ziehl and the pathologist Friedrich Neelsen [25–28]. The refined staining technique, called Ziehl-Neelsen staining, is still widely used today, remaining a standard method in TB diagnosis in low- and middle-income countries [14]. Sputum smears or tissue samples are first stained with a solution of carbolic acid with fuchsin and heated, destained with acid alcohol and counter-stained with a solution of methylene blue. Mycobacteria retain the red color from the first staining step during the destaining step with acidic solutions; hence they are also referred to as acid-fast³ bacteria. This unusual staining property can be explained by the distinctiveness of the mycobacterial cell envelope.

2.1.2 The Mycobacterial Cell Wall

Mycobacteria are classified as Gram-positive since they lack the typical outer lipid bilayer membrane found in Gram-negative species; however, they are neither fully Gram-positive

²MDR-TB strains are resistant to at least two of the most powerful first-line anti-TB drugs isoniazid and rifampicin. XDR-TB strains are MDR with additional resistance to at least one of the second-line agents: fluoroquinolones (e.g., moxifloxacin) and injectable drugs (e.g., kanamycin).

³To avoid any misunderstanding, it should be noted that the term “fast” refers to the noun fastness, in the sense of stability or resistance; acid-fastness meaning resistance to destaining with acidic solutions.

nor Gram-negative with regards to Gram-staining results and envelope structure [29,30] (see Figure 2.1). The mycobacterial cell membrane (CM) was long considered to be a conventional plasma membrane; however, lipid studies have revealed that besides the typical phospholipids cardiolipin (CL), phosphatidylinositol (PI) and phosphatidylethanolamine (PE), the most abundant lipids found in the inner membrane are phosphatidylinositol mannosides (PIMs) [31]. PIMs are thought to increase the membrane's stability and reduce its permeability to drugs [32]. The cell membrane is surrounded by a relatively thin layer of peptidoglycan (PG), comparable to that of Gram-negative bacteria. It is followed by a layer of arabinogalactan (AG), which is covalently linked to the PG layer [33]. Similar to Gram-negative species, mycobacteria possess an outer membrane, albeit of a very different structure. The inner leaflet of the mycobacterial outer membrane (OM) is made up of unusually long-chain fatty acids called mycolic acids (MA). There are three different types of MAs found in members of the MTBC, namely α -mycolates, methoxycolates and ketomycolates [32]. Most MAs are, in turn, covalently linked to the cell wall polysaccharide AG. The three linked components make up the cell wall core, also called the mycolyl arabinogalactan-peptidoglycan (mAGP) complex [33]. The outer leaflet of the OM is comprised of various free glycolipids like di-, tri- and pentaacyl trehalose (DATs, TATs and PATs), sulfated trehalose glycolipids (SGLs) and large (> 90 carbons) hydrophobic phthiocerol dimycocerosates (PDMIs) and phenolic glycolipids (PGLs) [30,32,34]. Further, lipoglycans like lipomannan (LM) and lipoarabinomannan (LAM) are anchored in the outer leaflet of the inner and outer membranes over PIMs [31,35]. In pathogenic mycobacterial species, the outer membrane is further surrounded by a loosely bound capsular structure consisting mostly of polysaccharides and proteins, containing hardly any lipids [36,37]. Structurally and functionally, these surface-layer glycoproteins can be compared to the lipopolysaccharides (LPS) of Gram-negative bacteria [38].

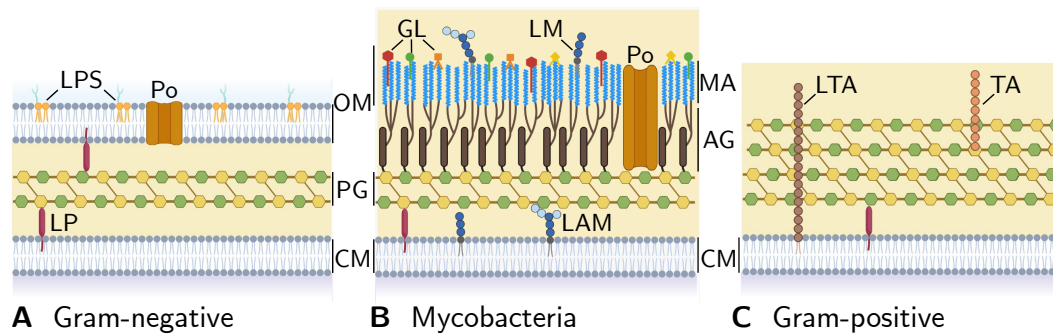


Figure 2.1: Simplified cell wall structures of Gram-negative and Gram-positive bacteria and mycobacteria. **A** shows the Gram-negative cell wall structure with a thin peptidoglycan (PG) layer between the cell membrane (CM) and the outer membrane (OM). The outer lipid bilayer contains porins (Po) spanning the complete bilayer and lipopolysaccharides (LPS) in the outer leaflet. Lipoproteins (LP) can be found in the periplasmic spaces, anchored in either membrane. **B** is a schematic representation of the mycobacterial cell envelope. The outer membrane (OM) is made up of long-chained mycolic acids (MA), which are covalently linked to an arabinogalactan (AG) layer. AG is, in turn, linked to the thin peptidoglycan (PG) layer. Free glycolipids (GL) make up the outer leaflet of the OM. Lipoarabinomannan (LAM) and Lipomannan (LM) are anchored in the outer leaflets of the CM or OM by diacylglycerol. Lipoproteins (red, oval-shaped) are found in the periplasmic space between the cell membrane (CM) and the PG layer. The capsule structure found in pathogenic mycobacteria is not shown in this representation. **C** shows the structure of Gram-positive cell walls. The cell membrane (CM) is surrounded by a thick peptidoglycan layer (green-yellow), containing teichoic acids (TA) as well as lipoteichoic acids (LTA), the latter being anchored in the CM by diacylglycerol. Lipoproteins of the cell membrane (CM) reach into the periplasmic space between the CM and PG layers. Models reproduced from [30, 35, 39], created using BioRender.com

Due to the large amounts of MAs, which can be as long as 70-90 carbons, the mycobacterial cell envelope is of a waxy nature and, therefore, especially robust [30]. For instance, the routine decontamination treatment of patient sputum with 1 M NaOH, which would kill most microorganisms, leaves viable *M. tuberculosis* cells [36]. The unique cell wall not only shields the bacterium from harsh environmental conditions, it also presents an efficient hydrophobic permeability barrier against most broad-spectrum antibiotics and other hydrophilic molecules [40–42]. Studies by Jarlier and Nikaido [40] showed that the permeability for certain cephalosporins was up to 1000-fold lower in *Mycobacterium chelonae* than in the most resistant Gram-negative bacteria *E. coli* and *Pseudomonas aeruginosa*. As the direct interface between bacterium and host, the cell wall also plays an important role in virulence, pathogen survival and host-pathogen-interaction [37, 41, 43–45]. This aspect will be taken up again in 2.1.5, where the establishment of an Mtb infection will be explained in more detail.

Up to a third of bacterial proteins are designated for extra-cytoplasmic locations; either to be embedded in the plasma membrane or secreted to the periplasm or further [46]. While the unique structure of the mycobacterial cell wall provides extraordinary protection, it also poses a barrier regarding the uptake of hydrophilic molecules and nutrients (e.g., glucose [40]), as well as hindering the excretion of proteins, such as virulence factors. The typical Gram-positive bacteria relies on the Sec or Tat pathways to transport proteins into or over the plasma membrane [47, 48]. Gram-negative bacteria have a range of additional pathways,

namely six different types of secretion systems (types I-VI) to transport molecules and proteins to or across their two membranes [49]. In the last decades, researchers have uncovered a novel type of secretion system, first found in *M. tuberculosis*, which allows them to secrete proteins. It joins the ranks of bacterial secretion systems as the seventh type, consequently named the Type-VII Secretion System.

2.1.3 The Type-VII-Secretion System

The Type-VII Secretion System (T7SS) was first discovered in *M. tuberculosis* but it is now known that analogous systems are also present in other mycobacteria, as well as in different other high-G+C Actinobacteria, like *Corynebacterium diphtheriae* or *Streptomyces coelicolor* [50]. A similar, yet clearly distinct system has been described in low-G+C Firmicutes like *Staphylococcus aureus*, *Streptococcus agalactiae* and *Listeria monocytogenes*, subsequently referred to as Type-VIIIb Secretion System [51].

Incitement and central player in the discovery of this novel bacterial secretion system is a low molecular mass, highly potent T-cell antigen, named ESAT-6 (6 kDa early secretory antigen target) or EsxA, first described in 1995 by Andersen et al. [52]. Genetic analysis of EsxA and its co-secreted partner CFP-10 (10 kDa culture filtrate protein), later renamed to EsxB, revealed a lack of any known N-terminal secretion signals characteristic for proteins secreted by the Sec or Tat pathways [50], suggesting the existence of a different, specialized secretion pathway. In 1998 the complete genome sequence of *M. tuberculosis* H37Rv was determined and analyzed by Cole et al. [53], in the course of which at least eleven additional genes encoding proteins showing sequence similarities with EsxA were identified. These were grouped in the *esxA* gene family and encode small proteins of around 100 amino acids. Furthermore, several small genes similar to *esxB* were found, situated directly adjacent to the *esxA*-like genes, thus forming an *esxA-esxB* operon. Today, eleven such operons are known, five of which are localized within a specific gene cluster, which includes a membrane-associated ATPase and several other putative components of the specialized T7SS. These paralogous gene clusters are named *esx-1* to *esx-5* (see Figure 2.2 **A**) and encode the respective ESX system (ESAT-6 secretion system). Special interest was directed to the ESX systems after it was discovered that the primary attenuating deletion of the vaccine strain *M. bovis* BCG, the so-called region of difference 1 (RD1), includes *esxA*, *esxB* and other genes of ESX-1 [54, 55]. While ESX systems are found widely among both RGM and SGM, the number of encoded systems differs between the strains [56]. Those mycobacteria that cause tuberculosis encode all five ESX systems but other well-known human pathogens like *M. leprae* and *M. abscessus* do not (see Table 2.1). Large-scale pan-genomic analysis of mycobacteria has further uncovered plasmid-encoded ESX systems in numerous species [57].

Table 2.1: ESX systems in different mycobacteria and their roles. The distribution of the five ESX systems among different mycobacterial species (human pathogens in bold) [56, 58]. The role of each ESX in Mtb is indicated on the right [59].

Species	ESX					
<i>M. abscessus</i>			3	4		
<i>M. africanum</i>	1	2	3	4	5	
<i>M. avium</i>		2	3	4	5	
<i>M. bovis</i>	1	2	3	4	5	ESX
<i>M. kansasii</i>	1	2	3	4	5	1 Virulence
<i>M. leprae</i>	1		3		5	2 Unknown
<i>M. marinum</i>	1		3	4	5	3 Zinc and iron uptake
<i>M. smegmatis</i>	1		3	4		4 Conjugation
<i>M. tuberculosis</i>	1	2	3	4	5	5 PE/PPE secretion
<i>M. ulcerans</i>			3	4		
<i>M. xenopi</i>		2	3		5	

The importance of ESX systems seemed clear; however, their heterogeneous distribution among both pathogenic and non-pathogenic species hinders intuitive attribution of specific roles. Extensive research has uncovered the purpose of most ESX systems in different mycobacterial species; however, the molecular structure of the system’s components and the exact mechanism of secretion remain partly unclear. While each system is responsible for the secretion of the encompassed EsxA- and EsxB-like proteins, they are involved in a number of different functions.

While ESX-1 is essential for full virulence and intracellular survival of *M. tuberculosis* [59], it is further involved in haemolysis in the fish pathogen *M. marinum* [60] and in conjugation in the non-pathogenic *M. smegmatis* [61]. ESX-3 was shown to be required for iron and zinc uptake in both the human pathogen *M. tuberculosis* as well as the non-pathogenic *M. smegmatis*. The most ancestral system, ESX-4, from which the other systems probably evolved by a series of duplications [50], is critical for intracellular growth of *M. abscessus* [59] and plays a role in DNA transfer between bacteria in *M. smegmatis*. ESX-5 is thought to be the most recently evolved system [50] and is exclusive to SGMs. It is involved in nutrient uptake and essential for growth in *M. tuberculosis* and responsible for the secretion of PE/PPE (Pro-Glu and Pro-Pro-Glu) proteins.

During the past seven years, different groups have uncovered the molecular organization of the systems ESX-3 of *M. smegmatis*, ESX-5 of *M. tuberculosis* H37Rv and *M. xenopi* at atomic resolution [62–65]. These findings resulted in a T7SS model as seen in Figure 2.2 **B**. The current model of a T7SS is a large membrane complex consisting of three protomers containing the four conserved membrane components EccB, EccC, EccD and EccE in a 1:1:2:1-stoichiometry [63, 65]. The conserved MycP protease is also part of the ESX machinery but is not tightly associated with the core complex. With one MycP protease per protomer, the total complex stoichiometry observed in ESX-5 of *M. tuberculosis* and ESX-3 of *M. smegmatis* of EccB:EccC:EccD:EccE:MycP is 6:6:12:6:3 [65]. Three EccB dimers and three MycP proteases reach into the periplasmic space and are anchored in the inner membrane by one transmembrane helix (TMH) per protein. Each EccB dimer forms a triangle,

together creating a central cavity with a diameter of about 5 nm [62], which is capped by the three MycP proteases arranged in a dome-like structure [65] (see Figure 2.2 **C**). The active site of MycP is directed into the central lumen of the cavity, suggesting the translocation and perhaps the processing of potential substrates at this location [65]. With 22 TMHs each, the six EccD dimers form a circular raft around the central cavity inside the cell membrane. Unlike other members of the FtsK/SpoIIIE ATPase family, EccC has two to three enzymatic domains instead of only one and is mostly encoded by a single gene [66]. The ATPase domains extend into the cytoplasm as a string of four. These flexible arms of EccC can adopt an extended or a contracted conformation and are thus thought to be involved in the initial steps of transmembrane transport, such as substrate selection and recognition [65].

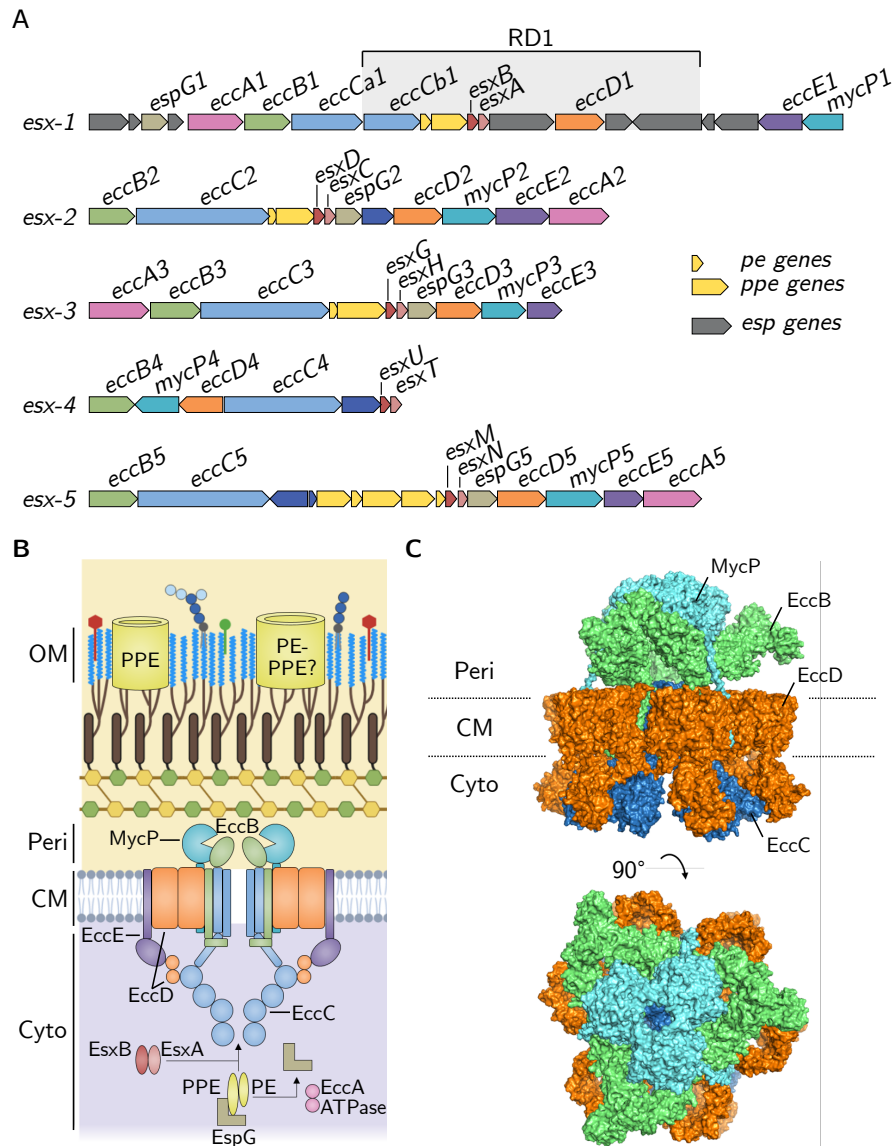


Figure 2.2: The ESX gene loci and the ESX secretion system model. Colors of genes and proteins correspond throughout the figure. **A** shows the gene clusters of the five ESX secretion systems of *M. tuberculosis*. The region of difference 1 (RD1, shaded gray) in *esx-1* is the primary attenuating deletion of the vaccine strain *M. bovis* BCG. Esp: ESX-associated proteins, Ecc: ESX-conserved-components. **B** shows the model of a type VII secretion system. EccB, EccC, EccD, EccE and MycP are conserved membrane components and form the secretion complex in the cell membrane (CM). Proteins of the Esx- and PE-PPE family are secreted by this machinery. The cytoplasmic ATPase EccA plays a role in the secretion of Esx proteins and helps dissociate the cytoplasmic chaperone EspG from its PE-PPE protein substrates prior to secretion. PE and PPE proteins are known to form pores in the outer membrane (OM). Cyto: Cytoplasm, Peri: Periplasm. **C** shows the side and top view of the core membrane complex of ESX-5 from *M. tuberculosis*, consisting of EccB5, EccD5, EccC5 and MycP5 (PDB 7NP7). This model is missing EccE5. Figures reproduced from [59] and [65], created using BioRender.com and PyMol [67].

2.1.4 Esx Proteins

As briefly mentioned before, the ESX substrates, EsxA and EsxB and the EsxA- and EsxB-like proteins are encoded within five different *esx*-loci. Additionally, there are six further *esxA/esxB*-like tandem genes in *M. tuberculosis* and a singleton, resulting in a total of 23 genes, named *esxA* to *esxW* encoding for proteins named accordingly (see Table 2.2). Despite a low sequence similarity of around 15 %, the Esx proteins are grouped into the so-called WXG100 protein family. They share the common characteristics of a size of approximately 100 amino acids (aa), a conserved central Trp-Xaa-Gly (W-X-G) motif and a similar secondary structure of two helical regions connected by a loop and flexible termini [68]. The mostly conserved C-terminal YXXXD/E motif was proposed to act as a secretion signal [69]. Despite a shared general secretion signal, the protein pairs are specifically secreted by their respective ESX system, meaning there are further yet unknown recognition factors [69].

Table 2.2: Esx proteins of *M. tuberculosis*. Esx proteins with the corresponding ESX system; EsxB paralogs are bold [70].

		Protein			
		inside of locus		outside of locus	
ESX-1	EsxA	EsxB			
ESX-2	EsxC	EsxD			
ESX-3	EsxH	EsxG	EsxR	EsxS	
			EsxQ		
ESX-4	EsxT	EsxU			
			EsxI	EsxJ	
ESX-5	EsxN	EsxM	EsxL	EsxK	
			EsxO	EsxP	
			EsxV	EsxW	
none			EsxE	EsxF	

2.1.5 Tuberculosis

Historically known under the names “white plague”, “phthisis (pulmonaris)”⁴ or simply called “consumption”, Tuberculosis is as old as mankind [71]. The latter names refer to the deterioration of the patient’s physical condition as the disease progresses. Transmission is airborne and occurs by inhalation of aerosol droplets containing the pathogen, making the lungs the primary site of infection [13, 72]. Typical symptoms of pulmonary TB (PTB) include coughing, fever, chest pain, fatigue, night sweats and rapid, unintentional weight loss [72]. However, only 5 % to 10 % of exposed, immunocompetent individuals actually develop active primary tuberculosis after infection [73]. In the majority of cases, TB develops into a latent form, meaning the infection is contained by an acquired immune response, and the disease does not break out. Latent TB infection (LTBI) can last for decades or even a lifetime, depending on the individual’s immune status and numerous environmental factors [73, 74]. LTBI is defined

⁴from the ancient Greek phthísis: wasting away

by the evidence of immunological sensitization by mycobacterial proteins while showing no clinical signs and symptoms of an active disease [75].

In this chapter, infection and development of the disease will be explained with special regard to immune reactions, host-pathogen interactions and the immune-evasive mechanisms of Mtb.

Host Immune Response and Manifestation of Disease

The human immune system is far too complex to do it any justice in a single introductory chapter and going into too much detail would go beyond the scope of this work. Therefore, this section will focus on the most important aspects of the human immune system in the face of an Mtb infection, without being intended to be exhaustive.

The innate – or unspecific – immune system represents the first line of defense after infection. The most important innate immune cells in an Mtb infection are macrophages, dendritic cells (DC), neutrophils (all of which are antigen-presenting cells (APCs)), and natural killer (NK) cells [76]. All of these lymphocytes are able to unspecifically recognize microorganisms, which elicits direct and indirect unspecific immune responses. Upon activation by a pathogen, these cells secrete immuno-stimulatory cytokines to stimulate immune cell maturation and to recruit other immune cells to the site of infection. As a direct defense strategy, they can release cytotoxic molecules, hydrolytic enzymes or reactive oxygen species (ROS) to attack the intruder. However, this also causes substantial damage to the host tissue, which exacerbates the inflammation [77]. APCs are capable of taking up pathogens by phagocytosis, killing the microorganisms and presenting the intruders' peptides on their own surface with the help of major histocompatibility complex (MHC) molecules. Subsequently, APCs travel to the draining lymph nodes and present the processed peptides and antigens of the eliminated invader to cells of the adaptive – or specific – immune system: B cells and T cells. DCs are the most potent APCs as they show the best peptide presenting capability and, therefore, are the prime T cell- and B cell-stimulators [78]. Activated DCs make long-lasting and stable contact with T cells in the lymph nodes, priming them on the specific antigen [79]. Depending on the T cell subset, this creates either cytotoxic T cells (CD8⁺), which can specifically kill infected cells, or T helper cells (CD4⁺), which aid in the activation of antigen-specific B cells. Primed B cells can develop into plasma cells, producing antibodies specific to the respective antigens of the pathogen. Activated B cells and T cells migrate to the site of infection to support the fight against the intruder in a more specific fashion.

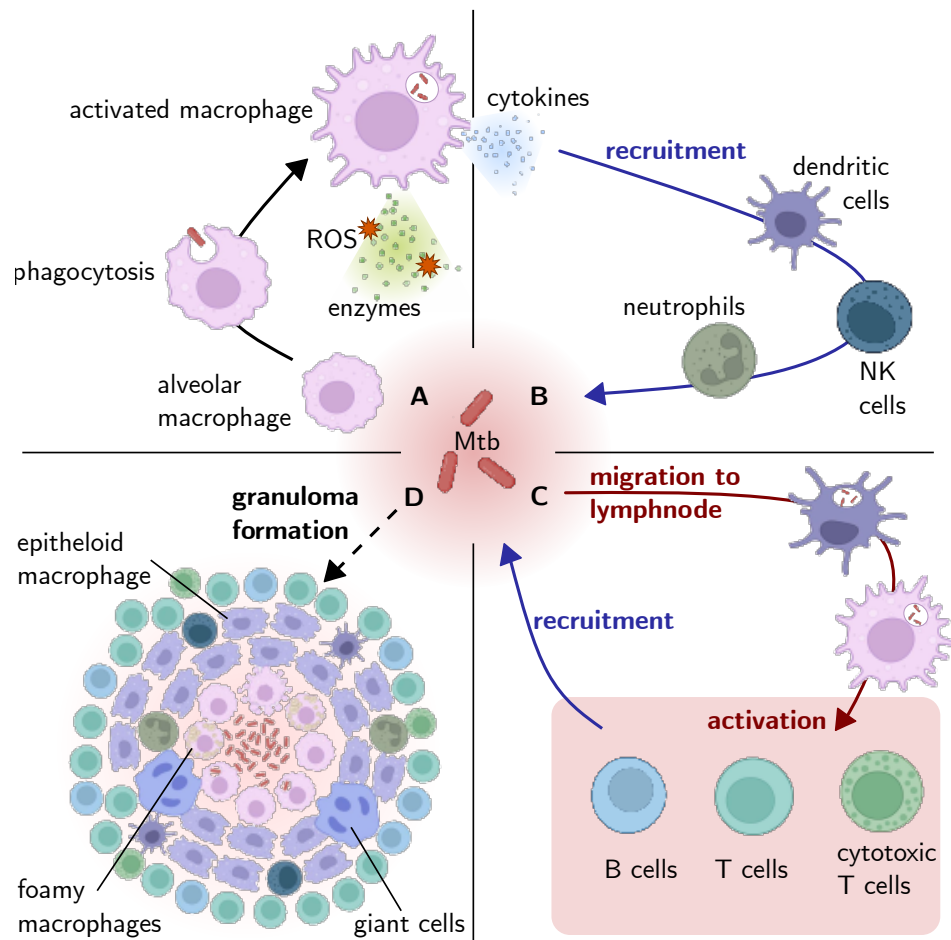


Figure 2.3: Schematic representation of cellular immune response and granuloma formation after infection with *M. tuberculosis* (Mtb). **A** Alveolar macrophages are usually the first cells at the site of infection (red area in the center). They take up the bacteria by phagocytosis, which activates the macrophage and elicits the secretion of bactericidal hydrolytic enzymes and reactive oxygen species (ROS). **B** Activated macrophages secrete pro-inflammatory cytokines, which help to recruit other innate immune cells like dendritic cells, natural killer cells (NK cells) and neutrophils to the site of infection. These cells play an important role in decimating and eliminating infectious microorganisms. **C** After maturation, macrophages and dendritic cells migrate to the draining lymph node of the lung and prime B and T cells. After activation, (cytotoxic) T and B cells migrate to the site of infection, which marks the onset of the adaptive immune response. **D** If neither the innate nor the adaptive immune system is able to eliminate the intruding bacteria, the primary lesion will develop into a granuloma. The center of a granuloma is often caseous and contains free bacilli surrounded by (infected) macrophages, some of which have developed into foamy cells or are succumbing to necrosis. They are surrounded by epithelioid macrophages and giant cells, as well as NK cells and neutrophils. The outer layer of the granuloma is made up of B cells and (cytotoxic) T cells. Models reproduced from [30,35,39], created using BioRender.com

Infection with Mtb is airborne, meaning it occurs by inhalation of infectious droplets. When the bacteria reach the lower respiratory tract, macrophages are typically the first immune cells they encounter. Alveolar macrophages (AM) and interstitial macrophages (IM) are the two major populations of lung macrophages [80], responsible for clearing the airway surfaces

of dust particles and microorganisms [81]. They continuously scavenge tissues and when detecting bacteria, they engulf them by phagocytosis, thereby trapping the intruders inside an organelle-like structure: the phagosome. After taking up microorganisms, macrophages then produce and secrete bactericidal molecules, elicit a pro-inflammatory immune response, and recruit other immune cells to the site of infection [81] (see Figure 2.3 A). During phagocytosis, AMs release oxygen radicals like $\bullet\text{OH}$, $\text{O}_2^{\bullet-}$ and other oxygen-derived reactive molecules like H_2O_2 and HOCl [81, 82]. Such reactive molecules can kill pathogens through oxidation and decarboxylation of the bacterial membrane. Further, macrophages stimulated by $\text{TNF-}\alpha$ or $\text{INF-}\gamma$ can produce nitrogen oxides like NO . During NO production and in the presence of other radicals (e.g. ROS), other toxic reactive nitrogen intermediates (RNI) can form, such as $\text{NO}\bullet$, NO^+ , HNO , NO_2^- and ONOO^- , which can lead to enzymatic inhibition [83–85]. ROS and RNI molecules are either secreted directly into the external environment or enter the phagosome. RNI can act in concert with superoxide radicals within acidic phagosomes to generate toxic products capable of limiting the survival and growth of *Mtb* [86, 87]. Over time, the phagosome “matures”, meaning it acquires microbicidal properties by a series of sequential fusions with various endocytic vesicles, eventually fusing with lysosomes to become a phagolysosome [88]. Lysosomes contain a diverse group of enzymes, including acid hydrolases and lysozyme, with the pH typically being at 4.7 [89], which is the favorable environment for the hydrolytic enzymes. Additionally, during the maturation process, proton-ATPase molecules accumulate in the phagosome’s membrane, resulting in further acidification of the phagolysosome [90]. However, *Mtb* – among others – has developed several strategies to evade and survive these otherwise effective defensive mechanisms. Like other intracellular pathogens, *Mtb* can survive inside the macrophage phagosome and even replicate in this protected niche while inhibiting the phagosome’s acidification and fusion with lysosomes [91, 92]. There is even strong evidence that *Mtb* can escape the phagosome and replicate in the macrophage’s cytosol [93, 94]. Further, the tubercle bacilli inhibit or hold up immune cell apoptosis, ensuring their protected replication within the host cell. By this and by interfering with several important signaling pathways in macrophages and other immune cells, the onset of the more specific and more effective adaptive immune system is delayed.

Nevertheless, the encounter between AMs and mycobacteria triggers an initial inflammatory response, namely the release of chemokines, like Interleukin-8 (IL-8), to recruit inflammatory cells such as neutrophils or natural killer (NK) cells to the site of infection [77] (see Figure 2.3 B).

Polymorphonuclear neutrophils (PMNs) are professional phagocytic cells and are attracted by high levels of IL-8. They are one of the first additional immune cells to arrive at the site of infection [95]. A massive influx of PMNs can be observed in the early stages of an *Mtb* infection and are the predominant cell type in lesions of patients with active TB [77]. Typically, PMNs contribute to the destruction of bacteria through different mechanisms, such as the release of antimicrobial peptides (AMPs), ROS and a variety of hydrolytic enzymes released from different types of granules. However, in the case of mycobacterial infection, these otherwise highly efficient immune cells fail to eliminate the invader, while the release of potent cytotoxic molecules causes severe tissue damage in their immediate vicinity [77, 95]. Eventually, necrotic cell death is initiated in PMNs by the high presence of ROS [95]. This, in turn, is beneficial for mycobacteria, as macrophages take up necrotic PMN debris, thereby being infected with *Mtb* and again allowing the survival and replication of the pathogen [96].

At about the same time as neutrophils are recruited to the site of infection, an influx of DCs can be observed [78]. DCs are considered to be the most efficient APCs and T-cell activators and are therefore regarded as the bridge between the innate and the adaptive immune system [97]. Immature DCs reside in all tissues and are specialized in capturing microorganisms and processing their antigens. After phagocytosis, DCs mature, which changes their specialty to the presentation of antigens. During maturation, the capabilities of capturing decline and MHC-molecules needed for antigen presentation are upregulated [78]. Activated DCs start to migrate to lymphatic organs such as lymph nodes (see Figure 2.3 C). However, DCs are also readily infected by Mtb, meaning the bacteria can survive within these immune cells [98,99]. By migrating to the lymph nodes while failing to kill the engulfed bacteria, DCs, as well as macrophages, involuntarily contribute to the dissemination of Mtb to deeper parenchymal lung tissue [100].

NK cells are granular lymphocytes that also, when activated, release potent cytotoxic molecules and induce apoptosis of infected cells, including infected macrophages. Cytokines like IL-2 and IL-12 increase and activate direct and indirect defense mechanisms of NK cells, respectively [101]. The cytoplasmic granules of NK cells contain perforin and granulysin, proteins that can cause pores in cellular membranes or disrupt lipid metabolism, respectively [101], which can solely destroy extracellular Mtb upon release [102]. Since macrophages serve as a replication niche for Mtb, NK cell-induced apoptosis of infected macrophages helps to reduce the viability of the bacteria [103]. NK cells further stimulate the immune response indirectly by releasing cytokines like IFN- γ and TNF- α .

The focal accumulation of macrophages, DCs and PMNs at the site of bacterial infection forms a primary tuberculous lesion. Over time, this leads to the development of organized structures known as granulomas, which are considered to be the hallmark of tuberculosis [104,105] (see Figure 2.3 D). About nine days after infection, bacteria can be detected in the draining lymph nodes, leading to the differentiation of naive T cells. Within a further ten to twelve days, antigen-specific effector T cells have migrated to the lesions in the lung tissue and further enhanced the pro-inflammatory response by activating infected phagocytes [97]. The lesions now remodel into organized granulomas with sheets of lymphocytes and fibroblasts enclosing infected (partly foamy and apoptotic) phagocytes in their center. Although the granuloma controls the focus of infection and prevents bacterial spread, it also provides a relatively protected niche for the pathogen, with reduced access for antibiotics. This arrangement is maintained in a dynamic way, with continuous immune cell recruitment replenishing dead cells. As long as the inflammatory response is balanced, the granulomas are maintained and contain the infection, resulting in latent TB [77]. It is estimated that between 5 % and 10 % of infected individuals will develop active TB [106], which is the result of an imbalance in the host immune response and is characterized by granuloma necrosis. This leads to the dissemination of bacteria to other parts of the lung or expectoration from the body, facilitating host-to-host transmission [77]. Risk factors for a TB infection include a prior HIV infection, under-nutrition, air pollution, type 2 diabetes mellitus, alcohol abuse, and smoking [72].

2.2 Protein Biochemistry

2.2.1 Protein Structure and Protein Folding

Proper folding of a protein is essential for its function and efficient interaction with substrates or other proteins. Principally, the final folding and three-dimensional structure of a protein is already largely determined by the primary structure - the linear sequence of amino acids. The order of amino acids defines the formation of regular structures in the polypeptide chain, the two main types known as α -helices or β -sheets. These two components make up around 50 % of a protein's secondary structure [107] and are developed and stabilized by hydrogen bonds between main groups of peptides.

An α -helix is a rod-shaped structure with a diameter of five to six Å. The polypeptide backbone is tightly wound right-handedly around the helix axis, whereby all side chains are directed outward. The helix completes a full turn every 3.6 amino acids and is stabilized by hydrogen bonds between the primary amine hydrogen of each amino acid (a) and the primary carbonyl oxygen of the fourth amino acid down from it (a+4). Typically, an α -helical region in a globular protein spans over 10 or 11 residues, equal to around 17 Å, and is located on the protein surface. While any of the twenty amino acids can contribute to the formation of an α -helix, Alanine (Ala), Glutamine (Glu), Leucine (Leu) and Methionine (Met) are most often found and Glycine (Gly), Tyrosine (Tyr), Serine (Ser) and Proline (Pro) are more uncommon [107]. α -helices make up around 30 % of the secondary structure found in an average globular protein [108]. The formation of an α -helical hairpin is a very common structure of two subsequent helices connected by a loop or turn region, resulting in a tight cluster of the two helices annealed in an antiparallel manner to each other. This structural component is known to play an essential role in DNA-binding proteins and, hence, in transcription regulation [109]. Multiple helices can form so-called coiled-coil domains, which are made up of two to five helix strands in either parallel or antiparallel orientation [110]. Such multi-helical domains are often found in transmembrane regions of membrane proteins and often play a role in cellular signaling and transport [111].

β -sheets are plate-like structures with the polypeptide chains being rather elongated than coiled. The stabilizing hydrogen bonds in β -sheets are also found between primary amines and carbonyl groups but occur between lateral regions of a polypeptide chain, lying either parallel or anti-parallel to each other. A typical β -sheet consists of 2 to 15 strands, each between 5 and 10 residues long. The three-dimensional arrangement of α -helices and β -sheets together with loops and unstructured or random coiled parts of the peptide chain is called the tertiary structure. This is greatly influenced by interactions between amino acid side chains either between each other or with the surrounding solvent. Hydrophilic side chains are found on the protein's surface, whereas hydrophobic side chains lie within the center of the protein. The tertiary structure is stabilized mostly by ionic bonds, hydrogen bonds and van der Waals bonds.

Lastly, a protein's quaternary structure refers to the three-dimensional assembly of multiple polypeptide chains, forming one macromolecule or functional unit. The stability of the quaternary structure is mainly due to inter-chain hydrogen and ionic bonds, as well as disulfide bridges.

2.2.2 Protein Interaction and Protein Binding

The functionality of a protein is highly dependent on its correct three-dimensional conformation and often on the binding to ligands like metal ions or other proteins. Protein-protein interactions are involved in a wide variety of biological processes, including signal transduction, regulation of gene expression, metabolic pathways and cell death [112]. Thus, it is hardly surprising that a number of diseases have been attributed to pathological misfolding of proteins, e.g. Alzheimer’s and Parkinson’s diseases [113], or to interrupted protein-protein interactions, some of which are believed to play a role in cancer development [114]. Knowing the importance of proteins to rapidly and efficiently fold in a crowded cellular environment, this process has been intensively studied. While small proteins often fold rapidly and spontaneously into their native conformations, larger proteins are more prone to misfolding and take longer to fold correctly [115]. One factor that supports protein folding *in vivo* is molecular chaperones, which substantially improve folding efficiency.

While some proteins fold in the timescale of microseconds [116], others take several minutes [117]. When studying protein-protein interactions, one essential question regarding the mechanisms of folding and binding often remains unanswered: Are both proteins correctly folded before association, or does the binding of one partner influence the folding of the other? There are two main models for coupled binding and folding mechanisms, called “induced-fit binding” and “conformational-selection binding” (see Figure 2.4). The first model assumes that the unfolded protein (U) binds to its ligand (L) and, upon this, gains the correct (or “folded”, F) conformation, while the latter model requires the protein to acquire the folded form before it can bind to its ligand.

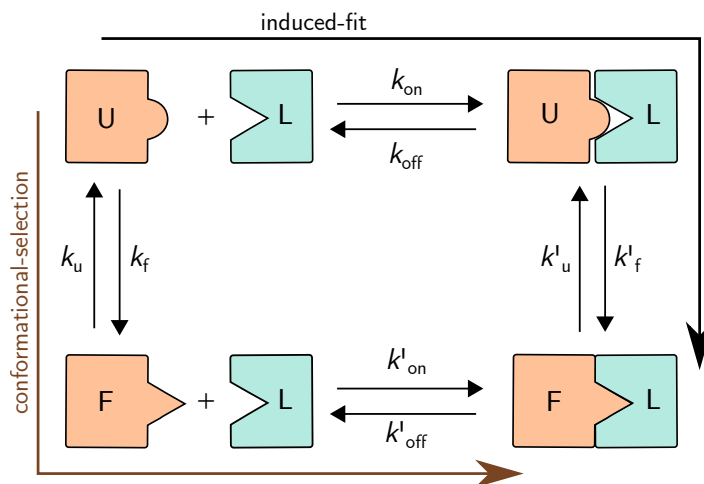
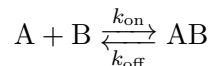


Figure 2.4: Models of coupled protein binding and folding. According to the induced-fit binding model, the unfolded protein (U) binds to its ligand (L) and, upon this, folds into the correct conformation (F). In the conformational-selection binding mechanism, the protein first folds into its native form (F) before it binds to its ligand (L). The rates k_{on} and k_{off} represent the rates of association and dissociation, and the rates k_f and k_u represent the rates of folding and unfolding, respectively.

Assuming the simplest one-step binding mechanism between two different proteins (A and

B) forming a 1:1 complex (AB), the reaction would be written as



where k_{on} and k_{off} are the rates for association and dissociation, respectively. As the association reaction involves two reactants, each with a first-order rate law, the overall reaction order is described with second-order binding kinetic [118]. The association and dissociation rates are defined as

$$\begin{aligned} k_{\text{on}} &= k_+ \cdot [A][B] \\ k_{\text{off}} &= k_- \cdot [AB] \end{aligned} \quad (2.1)$$

with the rate constants k_+ in units of $\text{M}^{-1} \cdot \text{s}^{-1}$ and k_- in units of s^{-1} . These rate constants are independent of the concentrations but do depend on the temperature. When the reaction has reached an equilibrium (eq), both rates k_{on} and k_{off} are equal. This allows the calculation of the equilibrium dissociation constant K_d in units of $\text{mol} \cdot \text{l}^{-1}$, which is defined as

$$K_d = \frac{k_-}{k_+} = \frac{[A]_{\text{eq}}[B]_{\text{eq}}}{[AB]_{\text{eq}}} \quad (2.2)$$

The concentration of the reactants $[A]_{\text{eq}}$ and $[B]_{\text{eq}}$ refer to the concentration of free molecules at equilibrium ($[R]_{\text{free}}$) in units of $\text{mol} \cdot \text{l}^{-1}$, while the concentration of $[AB]_{\text{eq}}$ refers to the concentration of the complex at equilibrium in units of $\text{mol} \cdot \text{l}^{-1}$ (see Equation (2.3)). The K_d value indicates the interaction affinity of the respective molecules. The lower the K_d , the lower the dissociation rate constant, the higher the affinity. At any time point, the concentration of each reactant (R) can be described as

$$[R]_{\text{total}} = [R]_{\text{free}} + [R]_{\text{complex}} \quad (2.3)$$

The differential equation describing the second-order rate law with initial concentrations of both reactants not being equal ($[A]_0 \neq [B]_0$) is

$$\frac{d[A]}{dt} = -k[A][B] \quad (2.4)$$

In the special case where the concentration of one component is significantly higher than of the other ($[A] \gg [B]$), the change of the excess reactant's concentration becomes negligible (as $[A] \approx [A]_0$). In this case, the reaction rate would depend solely on the concentration of the trace component ($[B]$), similar to a first-order reaction. The integrated form of the rate law would be

$$[B] = [B]_0 e^{-[A]kt} \quad (2.5)$$

where $[A]$ and $[B]$ are the concentrations of A and B at time t and $[B]_0$ is the initial concentration of B. Because of the similar behavior to a first-order reaction, this type of reaction is called "pseudo-first-order". The second-order rate constant k (in units of $\text{M}^{-1}\text{s}^{-1}$) can be multiplied with the concentration of the reactant assumed to remain constant ($[A]$) to create the pseudo-first-order rate constant k^* (in M), which can then be used in writing the

integrated form of the pseudo-first-order rate law:

$$[B] = [B]_0 e^{-k^* t} \quad (2.6)$$

Designing titration experiments in such a way that a pseudo-first-order reaction can be assumed greatly facilitates the analysis of binding kinetics. Keeping one monomer (B) at a constant concentration throughout the experiments and varying the concentration of the excess protein partner (A), allows to determine the (A-)concentration-dependent rate constant k_{obs} . For a binding mechanism as described and with the concentration conditions stated above, the observable equilibrium rate constant is defined as

$$k_{\text{obs}} = k_+ \cdot [A] + k_- \quad (2.7)$$

In This Work

In this work, the titration experiments with EsxB and EsxA or EsxT were performed in a way that allowed pseudo-first-order kinetics to be assumed, meaning the binding partners EsxA or EsxT were in great excess over EsxB. The complex formation was detected through an increase in FRET efficiency as the fluorophores came closer together, and EsxB was tightly folded in the complex with its binding partner.

2.2.3 Thermodynamics of Proteins

When regarding the processes of protein folding and binding from a physical chemist's point of view, thermodynamic principles become important. For experiments done in this work, the transition from a folded (F) to an unfolded (U) form (and vice versa) due to changing temperature needs to be addressed. For most proteins, the change in heat capacity between those two conformations at constant pressure ΔC_p is quite large [119]. The essential thermodynamic equations for protein unfolding (transition from F to U, termed "FU") under isobaric conditions are:

$$\begin{aligned} \Delta H_{FU} &= \Delta H_0 + \Delta C_p (T - T_r), \\ \Delta S_{FU} &= \Delta S_0 + \Delta C_p \cdot \ln \left(\frac{T}{T_r} \right) \\ &= \frac{\Delta H_0}{T_r} + \Delta C_p \cdot \ln \left(\frac{T}{T_r} \right), \\ \Delta G_{FU} &= \Delta H_0 - T \Delta S_0 + C_p \left(T - T_r - T \cdot \ln \left(\frac{T}{T_r} \right) \right). \end{aligned}$$

T_r can be any reference temperature. However, it is often convenient to set it equal to the melting temperature T_m , which is the midpoint temperature of unfolding. With the parameters of the conformational enthalpy ΔH_0 and entropy ΔS_0 , as well as with ΔC_p , the Gibbs free energy can be determined with respect to temperature (see Equation (2.8)). This equation can also be viewed as a measure of the stability of the protein as a function of temperature. Assuming a protein in its native state (F) is most ordered, and the free energy is at a minimum, the process of unfolding due to an increase in temperature would lead to an increase in both enthalpy and entropy.

Thermal unfolding of proteins was observed with circular dichroism (CD) spectroscopy in this work. In contrast to using differential scanning calorimetry (DSC), it was not possible to determine the change in constant-pressure heat capacity (ΔC_p) [120]. However, the data of temperature-dependent ellipticity change can be fit with equations, revealing T_m and ΔH_0 , while assuming $\Delta C_p = 0$. With these terms, ΔS_0 and ΔG_{FU} can be calculated and plotted to summarize the thermal stability of a protein.

2.2.4 Protein Bioconjugation

Some biophysical methods require modifications of the proteins prior to analysis. Depending on the experimental setup, it might be necessary to immobilize the proteins to a surface, label them with a nano-gold particle or a fluorescent dye, or use crosslinking agents to study protein interaction. Such bioconjugation reagents – crosslinkers – and modification reagents can be described by their chemical reactivity, molecular properties, or by their applications. Depending on the experimental technique, proteins, or rather their amino acid side chains, can be targeted specifically by using labeling reagents with specific targets or tagged unspecifically by targeting common and naturally present groups. Regarding the reactive groups of bioconjugates, the most commonly used reagents target amine-, carboxyl- or thiol groups. Typically, two of such reactive groups are linked by a spacer region, mostly an alkyl chain of a certain length. If both reactive groups target the same functional group, they are called homobifunctional, otherwise they are termed heterobifunctional. In this work, the well-characterized reagents EDC and sNHS were used to study a possible complex formation of different Esx proteins by unspecifically crosslinking the interaction partners in solution. Subsequent analysis by SDS-PAGE and mass spectrometry were used to identify the linked proteins. For FRET-analysis, specific fluorescent labeling of different Esx proteins was achieved by using fluorescent dyes linked to a reactive group targeting sulfhydryls (-SH).

Crosslinking Studies

Chemical crosslinking is a fairly simple method to stabilize protein conformation, non-covalent complexes and protein-protein interactions. Simply put, a solution containing one or more protein species is mixed with crosslinking reagents, and after incubation, the linked products are detected. This can be done by SDS-PAGE, where the linked complexes have a higher molecular weight than the individual monomers or more thoroughly in combination with mass spectrometry [121].

1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) is a zero-length crosslinker, meaning it will not be part of the final complex but rather orchestrate a direct conjugation between target molecules. EDC induces isopeptide bonds between primary amines and carboxylates of proteins in solution [122] (see Figure 2.5 **A** and **D**). Using a zero-length crosslinker for the chemical crosslinking studies increases the chances of linking residues that are located at the actual protein-protein interaction site(s) [123]. This means that if populations of one or more cross-linked individual proteins are, in fact, detected, they have, with great certainty, interacted. In order for the crosslink to be formed, the reactive sites must be in very close proximity for a sufficient amount of time, and this must happen for a substantial fraction of the protein population to be detected. Random interactions are either too short-lived to be linked or will be too few and fall below the detection limit [121]. In the first step,

EDC reacts with a carboxyl group and forms an adduct, an O-acylisourea intermediate, which is fairly unstable in aqueous solutions (see Figure 2.5 **A**). Next, it can be displaced by nucleophilic attack from a primary amino group, resulting in a stable amide bond between the original carboxylic acid group and the primary amino group (see Figure 2.5 **D**). However, EDC crosslinking is most efficient at low pH values of around 4 to 5 [124]. When using buffers at neutral pH, which is commonly more suitable for proteins, the reaction efficiency is much lower. Furthermore, the unstable O-acylisourea intermediate is prone to undergo hydrolysis, resulting in the original carboxyl group (see Figure 2.5 **C**) and the release of a isourea byproduct of EDC. Therefore, to increase the crosslinking reaction, a second reagent can be introduced to the coupling protocol, namely N-hydroxysulfosuccinimide (sNHS). It can form a stable, highly amine-reactive ester with the EDC-activated carboxylic group and react with primary amines, resulting in stable amide bonds (see Figure 2.5 **B**). NHS ester crosslinking is typically performed at neutral pH, allowing work at physiological pH [124].

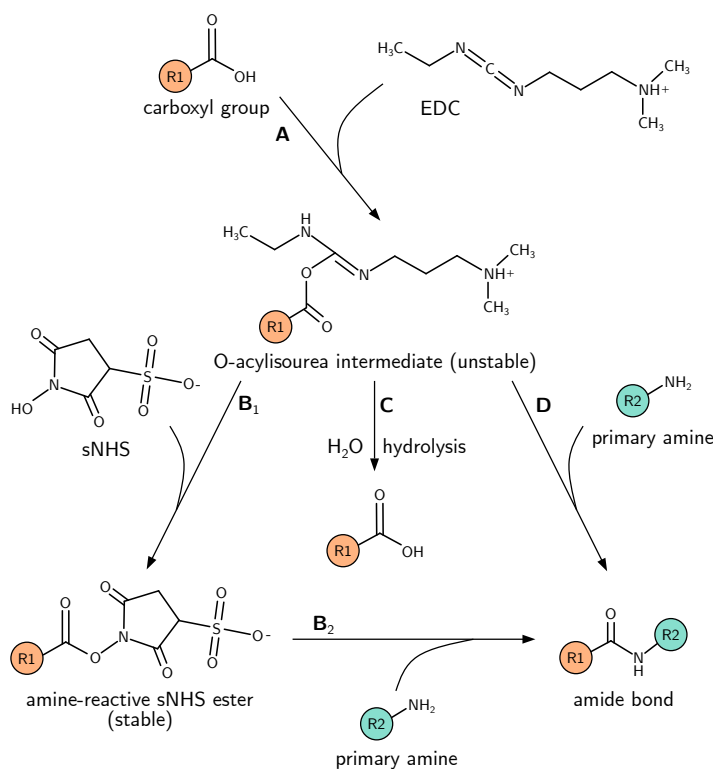


Figure 2.5: Reaction scheme of EDC/sNHS-mediated crosslinking. **A** EDC can react with the carboxylic acid of a molecule or protein (“R1”), forming an unstable O-acylisourea intermediate. **B₁** In the presence of sNHS, EDC can couple sNHS to the carboxyl of “R1”, forming a stable and amine-reactive sNHS ester. **B₂** In a second step, the sNHS ester can react with primary amines of other molecules or proteins (“R2”) to form a stable amide bond between “R1” and “R2”. **C** In absence of primary amines or sNHS, the O-acylisourea intermediate can rapidly hydrolyze, regenerating the original carboxylic acid of “R1”. **D** If primary amine groups of a molecule or protein (“R2”) are present in the solution, they can displace EDC from the unstable intermediate by nucleophilic attack, forming a stable amide bond between “R1” and “R2”.

Fluorescent Labeling

Chemically attaching fluorescent dyes to peptides, proteins or other biomolecules is another form of bioconjugation. In this case, the respective fluorophore is joined over a linker to a reactive group, typically targeting thiol- or amine groups. Amine-reactive probes are often used for unspecific labeling. They typically target the primary amino group in the side chain of lysine, as well as the N-termini of proteins. In contrast to the relatively high prevalence of lysines in proteins (6 %), cysteines are less abundant (2 %) and therefore present a suitable target for specific labeling [121]. Maleimides are one type of reactive group that targets sulfhydryl groups (-SH), used to link the dye to the protein by forming a stable thioether bond to the side chains of cysteines (see Figure 2.6). For the labeling reaction, it is necessary to ensure that the cysteines do not form disulfide bonds among each other since this would make a reaction with maleimides impossible. As thiols are susceptible to oxidation and subsequent formation of disulfides or higher oxidation products, reducing agents should be used in the labeling process [124]. There are several reagents that can be used for the reduction of cysteine disulfides, e.g., dithiothreitol (DTT) or 2-mercaptoethanol (β -MEtOH) [125]. However, the reagents themselves contain a thiol group and would interfere with the maleimide-labeling [126]. For this purpose, another strong reducing agent for disulfide linkage, tris(2-carboxyethyl)phosphine (TCEP) [127], is better suited, as it does not contain a sulfhydryl group, thus not competing in the labeling reaction.

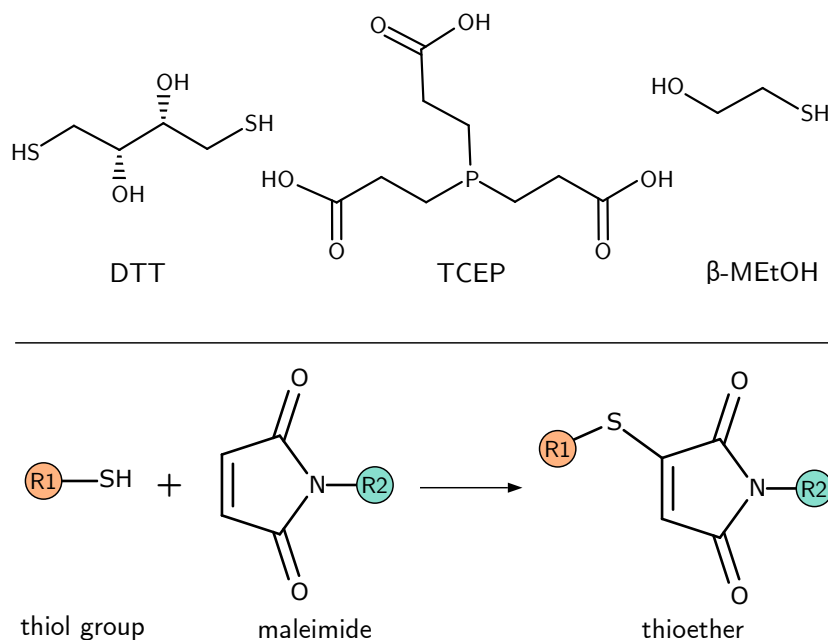


Figure 2.6: Sulfhydryl reductants and maleimide reaction. **Top:** Three common sulfhydryl reducing agents: dithiothreitol (DTT), tris(2-carboxyethyl)phosphine (TCEP) and 2-mercaptoethanol (β -MEtOH). In contrast to DTT and β -MEtOH, TCEP does not contain any sulfhydryl groups, meaning it will not compete for coupling sites in a maleimide reaction. **Bottom:** The maleimide reaction. Maleimide molecules react specifically with thiol groups, forming a stable thioether linkage. In terms of fluorescent labeling of proteins, “R1” represents a protein’s cysteine and “R2” a fluorophore.

2.3 Förster Resonance Energy Transfer

Förster resonance energy transfer, or FRET for short, describes the mechanism where, upon excitation, one molecule transfers energy to a different one nearby and hence excites the latter. It is named after the German physical chemist Theodor Förster, who contributed greatly to the understanding of nontrivial energy transfer between molecules in the mid-20th century [128–130]. This energy transfer depends especially on the distance between the two molecules. When using two light-sensitive molecules, like fluorophores, attached to the molecules of interest, FRET becomes a useful method to analyze conformational changes in proteins or interactions of nucleic acids and proteins. In the course of this work, FRET between two fluorophores conjugated to different mycobacterial WXG-100 proteins was used to determine a change in protein structure. Hence, this chapter will focus on FRET with fluorescent dyes.

It should be mentioned at this point that when two fluorophores are used, the acronym FRET is often transcribed as “Fluorescence Resonance Energy Transfer”, which is, however, misleading since the energy transfer happens before fluorescence and is not initiated, let alone accomplished by it. Nevertheless, it encompasses the two essential topics that will be described in this chapter: **F**luorescence and **R**esonance **E**nergy **T**ransfer.

2.3.1 Principles of Fluorescence

Before going more into depth, it is essential to recapitulate the mechanism of photoluminescence and, most of all, fluorescence. The processes involved in fluorescence can be schematically illustrated with the help of the so-called Jabłoński-Diagram (see Figure 2.7), named after the Polish physicist Aleksander Jabłoński, who first proposed it in 1933 [131].

It is used to describe transitions of electrons between different electronic energy levels. The ground state is termed S_0 , whereas n -higher electronic energy levels are labeled as S_n . In each of the electronic states, there are various vibrational energy levels in which the molecules can exist. Molecules, or rather their electrons, can be excited, meaning transferred to a higher energetic state, by absorption of energy E . In the case of fluorescence, this is achieved by photons of the energy

$$E = h\nu_{\text{ex}} \quad (2.8)$$

with ν_{ex} being the frequency of the exciting photons and h is Planck’s constant. The energy absorption excites the electron to a vibrational energy level of S_1 or of higher levels (S_n). Subsequently, the electron relaxes by non-radiative vibrational energy relaxation (VR) to the lowest vibrational energy level of the current electronic energy level S_n . This occurs on a time scale of 10^{-12} s, whereby the excess energy is given off to the surrounding solvent. By internal conversion (IC), the electron can change to a higher vibrational level of a lower energy level S_{n-1} and is followed by VR. The processes of VR and IC can alternate until the electron has returned to the ground state S_0 . Depending on the molecular structure of the excited molecule, relaxation from S_1 to the ground state S_0 can also occur by emission of a photon. This process is called fluorescence. It occurs on a time scale of several nanoseconds. Depending on the fluorescent molecules, the spectra of both absorption and emission are individual and depend on the chemical molecular structure. Due to the mentioned energy loss during relaxation by VR and IC and the interaction with surrounding solvent molecule

[132,133], the energy of the emitted photons during fluorescence is lower than the energy of the exciting photons ($h\nu_{\text{ex}} > h\nu_{\text{emF}}$). This phenomenon is also termed “Stokes shift”, named after the Irish-English mathematician and physicist George Gabriel Stokes, who described this observation in 1852 [134].

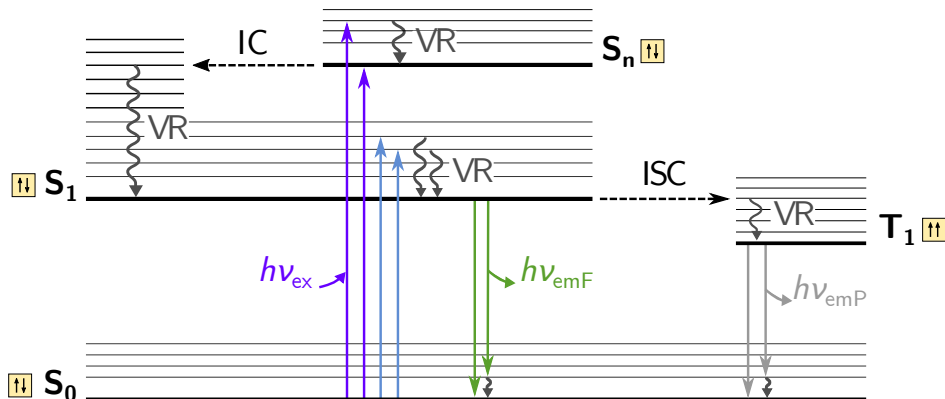


Figure 2.7: Jablonski diagram. This term diagram after A. Jablonski shows the possible electronic transitions between energy levels of a molecule. Electronic energy levels are shown as bold horizontal lines; thin horizontal lines depict vibrational energy levels. Levels are labeled according to their spin multiplicity either as S_n (S for singlet) or T_n (T for triplet), with the index n indicating the height of the energy level and S_0 representing the ground state. Spin orientation is indicated in small yellow boxes. Absorption of a photon with the energy $h\nu_{\text{ex}}$ leads to the elevation of an electron to a higher energy level S_n (solid-line purple vertical arrows). The excited electron can relax by radiation-less vibrational energy relaxation (VR, vertical grey curved arrows) and by internal conversion (IC horizontal dashed black arrow) followed by VR to the lowest vibrational level of S_1 . When returning to the ground state S_0 , a photon of the energy $h\nu_{\text{emF}}$ is emitted - this process is called fluorescence (solid-line green vertical arrows). An excited electron from a singlet state can also undergo intersystem crossing (ISC, horizontal dashed black arrow) to a triplet state. This involves a forbidden spin conversion, making this transition less likely. Phosphorescence occurs during the return to the ground state S_0 from a triplet state T_1 , with a photon of the energy $h\nu_{\text{emP}}$ being emitted (solid-line light-grey vertical arrows). The relationship between the photon’s energy of the depicted processes is as follows: $h\nu_{\text{ex}} > h\nu_{\text{emF}} > h\nu_{\text{emP}}$.

For the sake of completeness, the topic of the other mechanism of luminescence, phosphorescence, will be briefly touched as well. Once excited, there are yet other possibilities for the electron to return to the ground state: that of intersystem crossing (ISC) accompanied by non-radiative VR or relaxation by emittance of light. In the electronic energy levels described previously (S_n), the “S” stands for *singlet* state. This refers to the state of all electrons being paired, each partner of a pair having an opposite spin. This means that the net angular momentum of electrons in any S_n state is zero, and the multiplicity of states is one (hence *singlet*). In contrast to that, ISC causes an excited electron to switch into a *triplet* state (T_1), whereby the excited electron has the same spin orientation as the electron in the ground state, resulting in two unpaired electrons and in a multiplicity of three (hence *triplet*). Phosphorescence occurs when the electron relaxes directly from the triplet state T_1 to the ground state S_0 by emitting a photon. However, this type of transition is spin-forbidden. Hence the emission rate of phosphorescent photons is clearly slower (10^3 s^{-1} to 10^0 s^{-1}) than that of fluorescence (10^8 s^{-1}).

2.3.2 Characteristics of Fluorophores

The outermost electron orbitals of a molecule are involved in the fluorescence process. Therefore, fluorophores typically contain several aromatic groups with π -bonds, as such ring structures distribute the electrons of the outer orbitals over a wide area [135]. The characteristics of a fluorophore are mostly determined by its chemical structure, although the solvent plays an important role too. Next to the individual spectral properties, i.e., the absorption and emission spectra as well as the molar absorption coefficient, the fluorescence lifetime, and the quantum yield are probably the most important features of a fluorophore. The quantum yield (Φ) is an indicator for the fluorescence intensity. It is defined as the ratio of the number of emitted photons to the number of absorbed photons. Recalling the Jabłoński-Diagram in Figure 2.7, there are other processes next to fluorescence that are responsible for the return of electrons from the excited state S_1 to the ground state S_0 , namely IC and ISC. These non-radiative events can be grouped into a single rate constant k_{nr} to represent all radiation-less ways, i.e. decay rates, that depopulate the excited state. Defining the emission rate of the fluorophore as k_{rad} , the quantum yield of such a molecule can be expressed as

$$\Phi = \frac{k_{rad}}{k_{rad} + k_{nr}} \quad (2.9)$$

The lifetime (τ) of a fluorophore is defined as the average time spent in the excited state following excitation. Employing the aforementioned rate constants, the fluorescence lifetime can be expressed as

$$\tau = \frac{1}{k_{rad} + k_{nr}} \quad (2.10)$$

It is important to note that fluorescence emission is a random process, with every excited fluorophore having the same probability of emitting a photon in a given time period after excitation. This means that τ is not a concrete time value but a statistical average. Typical fluorescence lifetimes of aromatic fluorochromes are in the range of 10^{-8} s to 10^{-9} s [136]. Another characteristic of fluorophores is their photostability. Depending on the molecule, the number of excitation-relaxation cycles (μ) prior to its destruction can vary greatly. Before succumbing to photobleaching, fluorophores can cycle from 700 to 3 million times between excitation and relaxation, depending on their molecular structure [137]. The resistance of a fluorophore to photo-destruction is described by its bleaching quantum yield (Φ_B). While less stable fluorophores emit less than 1000 photons, good fluorophores can emit millions of photons before irreversible bleaching. The process of photobleaching most probably occurs from the S_1 or T_1 state.

2.3.3 Resonance Energy Transfer

Having elaborated on the basic principles of absorption and emission that fluorophores can undergo, there is yet another important process to address in this chapter, namely that of resonance energy transfer (RET). This describes a process by which energy is transferred from a molecule in the excited state, called the donor (D), to a molecule in the ground state, termed the acceptor (A). It is important to note at this point, that the process of energy transfer addressed in this chapter does not involve the appearance of a photon, meaning it is non-radiative. It happens by means of long-range, intermolecular dipole-dipole coupling

between D and A molecules, which depends on the distance and the relative orientation between the dipoles. Under optimal conditions, RET can span distances between A and D (r_{AD}) from between 10 Å to 100 Å [138]. The rate of energy transfer (k_T) from a donor D to an acceptor A separated by a distance r is given by

$$k_T(r) = \frac{1}{\tau_D} \left(\frac{R_0}{r} \right)^6 \quad (2.11)$$

where τ_D is the fluorescence lifetime of D (in the absence of A). The term R_0 is also called “Förster distance” and is a distance parameter that depends on the DA-pair in use. It can be calculated from the spectral properties of the donor and acceptor molecules and the donor quantum yield and will be defined later on (see equation (2.15)). The strong dependence of the energy transfer on the distance between the molecules becomes evident when looking at equation (2.11). Further prerequisites for efficient RET to occur are that the emission spectrum of the donor molecule must overlap sufficiently with the absorption spectrum of the acceptor and therein, both the quantum yield of D (Φ_D) (or the donor fluorescence intensity, F_D) and the absorption coefficient of A (ε_A) must be adequately high. The overlap integral $J(\lambda)$ is given by

$$J(\lambda) = \int_0^\infty F_D(\lambda) \varepsilon_A(\lambda) \lambda^4 d\lambda \quad (2.12)$$

with $F_D(\lambda)$ being a normalized, corrected factor for the fluorescence intensity of D at the wavelength λ [139].

The energy transfer can also be described in terms of transfer efficiency (E_T). Here, the number of quanta transferred from D to A in relation to the quanta absorbed by D are evaluated. This can be done by using the relative fluorescence intensity of the donor in the absence (F_D) and presence (F_{DA}) of the acceptor or the lifetimes under the respective conditions (τ_D and τ_{DA}).

$$E_T = 1 - \frac{F_{DA}}{F_D} = 1 - \frac{\tau_{DA}}{\tau_D} \quad (2.13)$$

Given that τ_D is the lifetime of the donor, $1/\tau_D$ equals the decay rate of the donor (see Equation (2.10)). If the transfer rate $k_T(r)$ is faster than the donor-decay rate $1/\tau_D$, energy transfer will be more efficient than vice versa. From equation (2.11) it is evident that when $r = R_0$, $k_T = 1/\tau_D$. Putting this into words means that when the D-A distance is equal to the Förster distance, the rate of transfer is equal to the decay rate of the donor. In terms of transfer efficiency (E_T), one would determine the ratio of the transfer rate to the total decay rate of the donor in the presence of the acceptor, which includes the donor decay rate and the energy transfer rate:

$$E_T = \frac{k_T(r)}{\tau_D^{-1} + k_T(r)} \quad (2.14)$$

This means that at a distance $r = R_0$, the transfer efficiency equals 0.5. As briefly mentioned before, the Förster distance R_0 is DA-pair-specific. It depends on the relative spatial orientation of the transition dipoles of D and A (given by the term κ^2), on the quantum yield of D in the absence of A (Φ_D) and on the spectral overlap ($J(\lambda)$). Further, the refractive index of the medium (n) is taken into account. Thus, the Förster distance R_0 in Å is given by

$$R_0 = 0.211 \sqrt[6]{\kappa^2 \cdot n^{-4} \cdot \Phi_D \cdot J(\lambda)} \quad (2.15)$$

with the wavelength being expressed in nm and $J(\lambda)$ in units of $\text{M}^{-1} \cdot \text{cm}^{-1} \cdot (\text{nm})^4$. The orientation factor κ^2 is generally assumed equal to the dynamically averaged value of $2/3$ and the refractive index n is typically assumed to be either close to that of water ($n = 1.33$) or 1.39 for biomolecules in aqueous solutions [137].

2.4 Fluorescence Correlation Spectroscopy

Fluorescence correlation spectroscopy (FCS) is a method used to observe single fluorescent molecules in solution and can give information about diffusion and reaction kinetics. It is based on the time-dependent analysis of fluorescence intensity fluctuations. These typically result from molecules diffusing out of and into focus but can also occur due to macromolecule-ligand-binding, internal macromolecule dynamics, and ISC. The small focal volume is defined by a focused laser beam and a confocal aperture [137] and is typically of an ellipsoidal shape. The fluctuation of a fluorescent signal recorded in an experiment is then analyzed regarding the signal correlation between varying delay times τ (see Figure 2.8 top).

2.4.1 Correlation Analysis

Correlation analysis can be performed as a measure of self-similarity (autocorrelation) or to compare different species of fluorescent particles (crosscorrelation). It is important to adjust the number of molecules in the focus at any given time to a minimum. The temporal average of particle number should be between 0.1 and 1000 [140]. A focal volume of 0.1 fl would correspond to a molecule concentration in the nanomolar range. Considering the Poisson distribution, the possibility P of having n fluorophores in the focal volume with an average number of molecules N is

$$P(n, N) = \frac{N^n}{n!} e^{-N} \quad (2.16)$$

The fluorescence fluctuations are defined as the deviations from the temporal average signal:

$$\begin{aligned} \delta F(t) &= F(t) - \langle F(t) \rangle \quad \text{with} \\ \langle F(t) \rangle &= \frac{1}{T} \int_0^T F(t) dt \end{aligned} \quad (2.17)$$

with T being the total data accumulation time. The autocorrelation results from multiplying the fluorescence intensity at time t , $F(t)$, with the intensity after a delay time τ , $F(t + \tau)$. With varying τ – usually in the range of 10^{-2} ms to 10^2 ms – and averaging the results over a large number of measurements, the autocorrelation curve is created. The normalized autocorrelation function for fluorescence intensities is defined as

$$G(\tau) = \frac{\langle \delta F(t) \cdot \delta F(t + \tau) \rangle}{\langle F(t) \rangle^2} \quad (2.18)$$

with the autocorrelation amplitude $G(0)$ being the normalized variance of the fluctuating fluorescence signal $\delta F(t)$ and equal to the inverse of the average number of molecules N . Given that $G(0) = 1/N$ and considering the Poisson distribution (equation (2.16)), it becomes clear that the possibility of having more than one molecule in the focal volume decreases with increasing amplitudes (see Figure 2.8 bottom right table). As a result, the analysis of FCS

data can give information about the molecule concentration⁵. Another interpretation that can be made from autocorrelation analysis is the diffusion rates of molecules in the solution. For measuring translational diffusion in a three-dimensional space, the diffusion coefficient D and the geometry of the ellipsoidal focus volume need to be included in the calculations. The autocorrelation function for three-dimensional diffusion is defined as

$$G(\tau) = G(0) \left(1 + \frac{4D\tau}{s^2}\right)^{-1} \left(1 + \frac{4D\tau}{u^2}\right)^{-1/2} \quad (2.19)$$

where $G(0)$ is the amplitude at $\tau = 0$, s is the radius, and u is the half-length of the ellipsoid representing the observed volume. The influence of the concentration and the diffusion rate of the molecules on the autocorrelation function can be seen in Figure 2.8.

⁵or rather the number of observed particles

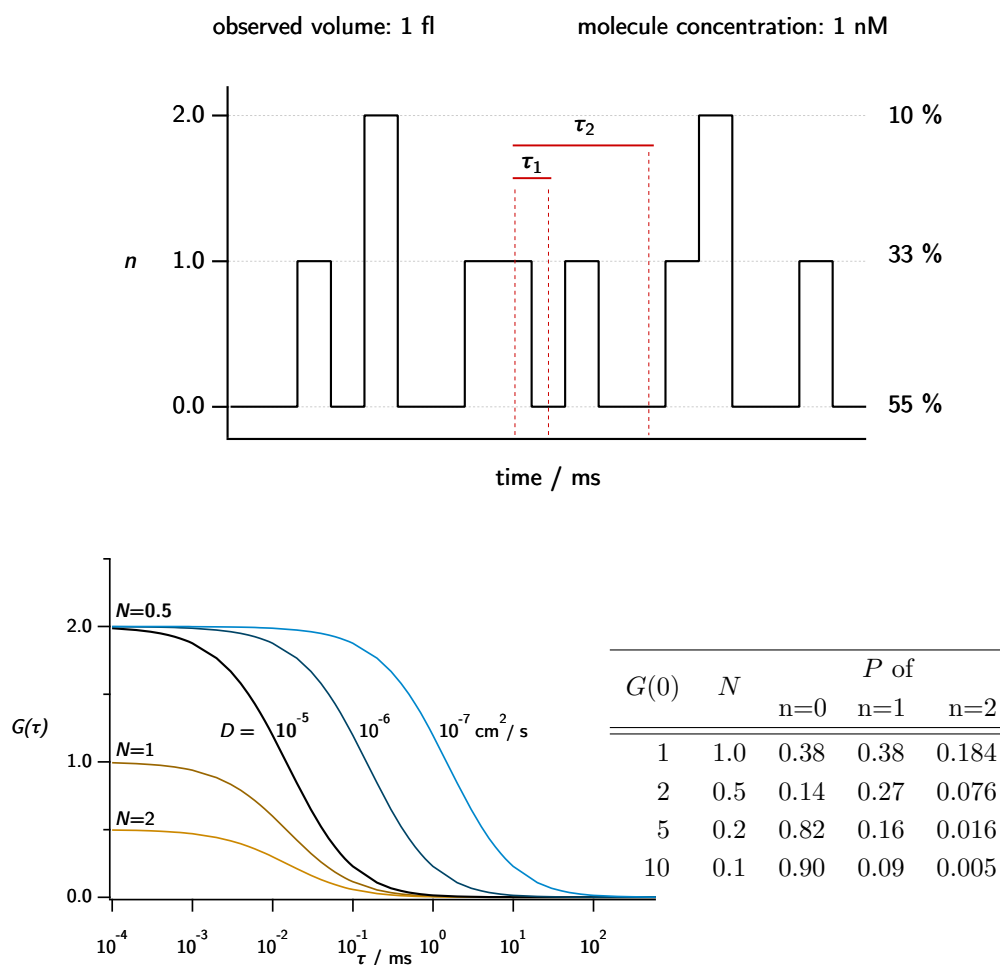


Figure 2.8: Schematic representation of fluctuations in fluorescent signal and creation of an auto-correlation function. **Top:** The number of fluorophores n changes over time as fluorophores diffuse into and out of focus. In this example, the observed focal volume is assumed to be 1 fl, and the molecule concentration is 1 nM. Assuming Poisson distribution (see Equation (2.16)), the average number of molecules in focus (N) would be 0.6, leading to the probability of having no molecule, one molecule, or two molecules in the observed volume being 55 %, 33 % and 10 %, respectively. The correlation of a fluorescent signal is calculated for a range of delay times τ_1, τ_2, \dots and results in an auto-correlation function. **Bottom:** Simulated auto-correlation functions (left) and table (right) showing the influence of different molecule concentrations (N) and different diffusion coefficients (D) on the correlation curves (left) and on the probability of observing no, one or two fluorophores in a focal volume of about 0.26 fl. Recreated from [137].

2.5 Circular Dichroism Spectroscopy

The method of Circular Dichroism (CD) spectroscopy can be used to analyze the secondary structure of peptides and proteins in solution. It uses the chiral nature of polypeptides and their structure-dependent differential absorption behavior of oppositely circularly polarized light to acquire structural information of the sample [141].

2.5.1 Principles of CD Spectroscopy

Even though light can not exclusively be described as an electromagnetic wave, this simplification will facilitate the explanation of the physical background of CD spectroscopy in the following section.

Light waves can be defined by their intensity or the wave's amplitude, their propagation direction, their wavelength or frequency, and their polarization. The latter refers to the direction of the wave's electric field vectors, which oscillates perpendicular to the direction of propagation. When the electromagnetic field oscillates in only one plane, the light is considered to be linearly polarized (Figure 2.9 A). Circularly polarized light is the result of two orthogonal electromagnetic wave planes with a 90° phase shift (Figure 2.9 B), which can result in clockwise (right-handed or R) or counter-clockwise (left-handed or L) rotation of the electric field vector along the direction of propagation [142]. Since CD spectroscopy can be considered a special type of absorption spectroscopy, the fundamental Lambert-Beer law should also be mentioned at this point

$$A_{\lambda} = \varepsilon_{\lambda} \cdot l \cdot c \quad (2.20)$$

where A_{λ} is the wavelength-specific absorbance, ε_{λ} is the wavelength-specific molar absorption coefficient, l the optical path length in cm and c the molar concentration of the absorbing molecule.

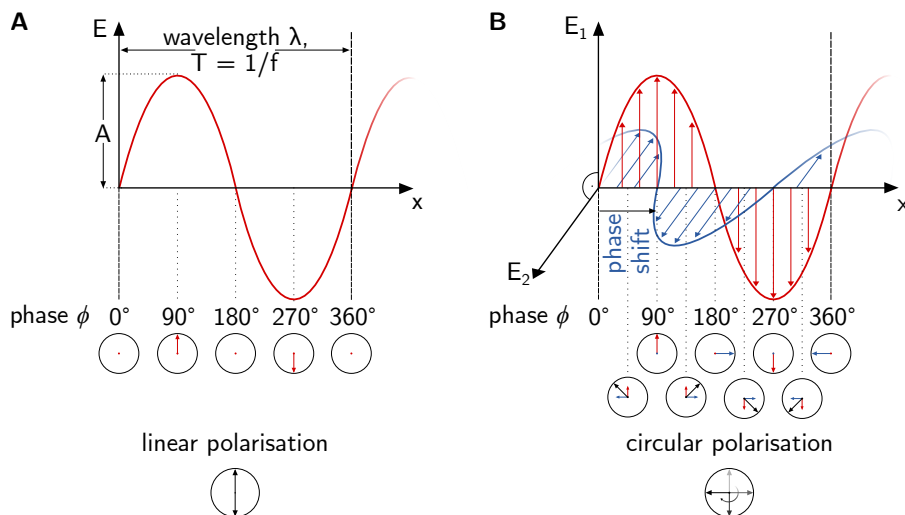


Figure 2.9: Polarized light. Diagram **A** shows linear polarization of light. The electric field E (red) with an amplitude A is oscillating (here: vertically) perpendicularly to the direction of propagation (x). The wavelength λ is indicated, and five points of the wave's phase ϕ are marked. Small sketches of the direction of E vectors are given below to illustrate the changes along x (with x being the center of the circles). **B** shows circular polarization of light. Two waves (red and blue) are shown to oscillate, both perpendicularly to the direction of propagation (x). Their amplitude and wavelength shall be equal. The electric field of the blue wave (E_2) is orthogonal to that of the red wave (E_1). Colored arrows are depicted to illustrate the different directions of field vectors for E_1 and E_2 . The blue wave is phase-shifted by 90° to the red wave. Eight points of the wave's phase ϕ are marked, and small sketches of the direction of E are given below to illustrate the changes (with x being the center of the circles). The upper row of these sketches shows the total vector coming from either wave (colored accordingly), while the lower row shows each wave's contribution (colored accordingly) to the total vector (black). The rotation in this example is clockwise or right-handed (R).

CD spectropolarimeters measure the difference in absorbance (ΔA) of L and R within a sample at a specific wavelength λ :

$$\Delta A_\lambda = A_{\lambda,L} - A_{\lambda,R} \tag{2.21}$$

If the absorption of L and R within a molecule were zero or of equal value, i.e. $\Delta A = 0$, the two components would create radiation of linear polarization (Figure 2.10 **A**). However, if L and R are absorbed by a sample each to a different extent, the result would be elliptically polarized light (Figure 2.10 **B**).

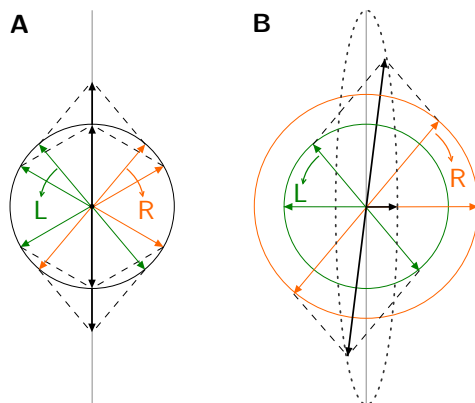


Figure 2.10: The CD effect. As seen in **A** plane-polarized light can also result from a left and a right circularly polarized component which are of equal amplitude. The colored arrows depict field vectors of the electric field of right circularly polarized light (R, orange) and of left circularly polarized light (L, green). The black arrows show the resulting total vector, which moves vertically along the center axis (light gray). Sketch **B** shows the effect of different absorption of R and L on circularly polarized light. In this example, L is absorbed to a greater extent than R, resulting in a smaller amplitude of L (green circle) than R (orange circle). The sum vector of both contributors (black arrow) follows an ellipse (dashed line), meaning the resulting light is elliptically polarized.

The ellipticity (θ) resulting from circular dichroism at a specific wavelength λ is given in degrees and is defined as:

$$\theta_{\lambda} = \tan^{-1} \frac{b}{a} \quad (2.22)$$

where a and b are the minor and major axis of the resulting ellipse. The relationship between the difference in absorbance and the ellipticity is

$$\theta = 32.98 \cdot \Delta A \quad (2.23)$$

The ellipticity is recorded at different wavelengths of the far UV spectrum, usually in the range of 190 nm to 250 nm, and given as a function of the wavelength in a CD spectrum. When combining equations (2.21) and (2.23), it becomes clear that positive values of ellipticity θ in the CD spectrum imply a stronger absorption of L, while negative values arise from a stronger absorption of R.

Circular dichroism in the wavelength range of 190 nm to 240 nm is mostly induced by absorption from peptide bonds, with strong absorption at 190 nm and weaker absorption at 210 nm to 220 nm [143]. The absorption spectra of polypeptides are long known to be influenced by their secondary structure [144], which, in turn, results in distinct CD spectra for α -helices, β -sheets and random coils (Figure 2.11). The typical CD signals corresponding to these three secondary structures, as shown in Figure 2.11, are summarized in Table 2.3.

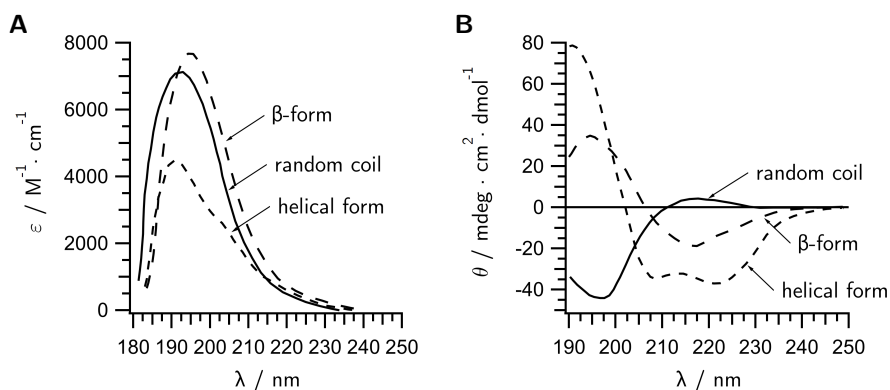


Figure 2.11: Wavelength-dependent absorption of peptides with different secondary structures and their distinct CD spectra. **A** shows the typical absorption spectra of poly-L-lysine hydrochloride in an aqueous solution. This polypeptide is known to have different conformations depending on its environment: a random coil or disordered conformation at room temperature and a pH of 6.0; an α -helical conformation at a basic pH value of 10.8 at room temperature and a β -form when heated to 52°C at a pH value of 10.8. Reproduced from [144]. Corresponding to this, **B** shows the typical CD spectrum of poly-L-lysine in the different conformations. Reproduced from [145].

Table 2.3: Typical features of CD spectra of proteins in different conformations. The typical signals of three different secondary structure elements are listed alongside their causative electronic transition, their position within the far UV spectrum, and their typical magnitude. Values taken from [146].

Secondary structure	Signal	Position / λ nm	Electron transition	Typical magnitude / $\text{deg} \cdot \text{cm}^2 \cdot \text{dmol}^{-1}$	
α -helix	positive	190 – 195	$\pi \rightarrow \pi^*$	60 000	to 80 000
	negative	208	$\pi \rightarrow \pi^*$	-36 000	\pm 3000
	negative	222	$n \rightarrow \pi^*$	-36 000	\pm 3000
β	positive	195 – 200	$\pi \rightarrow \pi^*$	30 000	to 50 000
	negative	215 – 220	$n \rightarrow \pi^*$	-10 000	to -20 000
random	positive	220	$n \rightarrow \pi^*$	< 10 000	
	negative	200	$\pi \rightarrow \pi^*$	-20 000	

2.5.2 Interpretation of a CD Spectrum

As mentioned previously, most CD spectra are recorded in the wavelength range of 190 nm to 250 nm, which holds information about the secondary structure of the sample. Near UV spectra, ranging from 260 nm to 320 nm can give information about the tertiary structure or structural changes thereof in changing environments. Dichroism in peptide and protein samples arises mostly from the peptide bond, aromatic amino acid chains, and disulfide bonds. In the most simple approach, a CD spectrum represents a composite of the protein's fractional secondary structures. Therefore, it is technically possible to express the spectrum by a linear

combination of the spectra of those elements plus a noise term:

$$\theta_\lambda = \sum \varepsilon_i \cdot S_{\lambda i} + \text{noise} \quad (2.24)$$

where θ_λ is the ellipticity of the protein at wavelength λ , ε_i is the fraction of each secondary structure (i) and $S_{\lambda i}$ is the ellipticity at each wavelength λ for each i^{th} secondary structural element and the term “noise” includes the contributions of aromatic chromophores and prosthetic groups [147]. Several different methods to analyze a CD spectrum to estimate the secondary structures have been developed so far, and they are all based on equation (2.24). In most methods, the experimentally obtained CD spectrum is compared to spectra of polypeptides or proteins with known conformations using different approaches. The goodness of the resulting fit can be rated according to the root mean square deviation (RMSD).

Chapter 3

Materials and Methods

The first major part of this work was to produce sufficient quantities of ten different Esx proteins to carry out experiments with. To get a first overview of possible inter-loci complex formations, chemical crosslinking studies were performed by combining monomers of ESX-1 to ESX-4 and the “orphan” proteins EsxF/E. Single- and double-cysteine mutants of the EsxB-like EsxF and EsxU were produced for subsequent fluorescent labeling. Focusing on the phylogenetically most ancient cluster ESX-4, the EsxA-like protein EsxT was used for titration experiments with EsxB using smFRET. For further biophysical characterization of EsxB/A, EsxF/E and EsxU/T, CD spectroscopy was used to evaluate the secondary structure of the complexes and their monomers. Thermodynamic, chemical and pH stability were also evaluated using CD spectroscopy.

3.1 Esx Protein Production

Ten different Esx proteins from *M. tuberculosis* were produced, namely EsxB and EsxA from ESX-1, EsxD and EsxC from ESX-2, EsxG and EsxH from ESX-3, EsxU and EsxT from ESX-4 and EsxF and EsxE (“orphan” proteins). All EsxB-like proteins are equipped with a His₆-tag, and are encoded together with the respective EsxA-like protein partner in the pMyCA-expression vectors.

3.1.1 Cloning and Production of pMyCA Expression Vectors

The respective *esx* gene pairs were available in the formerly used pMyNT-expression vector. This originally used vector was shortened in the promotor region to make the vector smaller and allow possible longer constructs to be inserted. This was done by performing a PCR using the respective pMyNT-vector as a template and a specially designed primer pair, which complementarily binds to the vector as well as introduces a BsrGI restriction site to both of the amplicon ends for subsequent vector circularization, as described previously [148]. The created pMyCA expression vectors were verified by sequencing using the service of LGC Genomics GmbH, Germany. All vectors produced and used in this work are listed in Table 3.1.

Introducing Cysteine Mutations with QuikChange Mutagenesis

To specifically introduce cysteine mutations in *esx* genes within the pMyCA-expression vectors, QuikChange site-directed mutagenesis was performed. This is a one-step PCR-based

Table 3.1: Overview of expression vectors. Expression vectors created (bold) and used in the course of this work. The genes encoded within each vector are given, as well as the corresponding ESX-cluster.

Vector name	encoded genes	ESX Cluster
pMyCA-EsxB/EsxA	His ₆ -EsxB, EsxA	ESX-1
pMyCA-EsxB-TE/EsxA	His ₆ -EsxB-T10C-E88C, EsxA	ESX-1
pMyCA-EsxD/EsxC	His ₆ -EsxD, EsxC	ESX-2
pMyCA-EsxG/EsxH	His ₆ -EsxG, EsxH	ESX-3
pMyCA-EsxU/T	His ₆ -EsxU, EsxT	ESX-4
pMyCA-EsxU-L/T	His ₆ -EsxU-L33C, EsxT	ESX-4
pMyCA-EsxU-LQ/T	His ₆ -EsxU-L33C-Q112C, EsxT	ESX-4
pMyCA-EsxF/E	His ₆ -EsxF, EsxE	-
pMyCA-EsxF-V/EsxE	His ₆ -EsxF-V13C, EsxE	-
pMyCA-EsxF-VT/EsxE	His ₆ -EsxF-V13C-T91C, EsxE	-

method that is especially efficient for introducing point mutations or mutating single codons. PCR is performed using the expression vector with the wildtype genes as a template and primers containing the desired mutation that are complementary to the gene of interest. The primers for introducing cysteine mutations used in this work are listed in Table 3.2 and were obtained from Thermo Fisher (Thermo Fisher Scientific Inc, USA) using their service of custom standard DNA oligo synthesis. The elongation step of each temperature cycle is adjusted according to the length of the vector. If two cysteine mutations are wanted in one gene, the second cysteine was introduced in a subsequent QuikChange PCR step using the verified expression vector with the first mutation as a template. After template digestion by DpnI,

Table 3.2: Primers used for QuikChange site-directed mutagenesis. The position of the cysteine mutation is highlighted by bold lettering. The sequences of all primers are given in 5' → 3' direction.

Primer Name	Sequence
EsxF V13C F	cgcgtagagcctgcg TGC atgcagggtttcgcc
EsxF V13C R	ggcgaaacctgcat GCA cgaggctctacgcg
EsxF T91C F	ggttatcaacaacagag TGC gcgtcggcgaggtgc
EsxF T91C R	gcacctgcgccgacgc GCA ctcgttggttgataacc

the resulting nicked expression vector is introduced into ultra-competent *E. coli* XL10 Gold cells (Agilent, USA) by heat shock transformation for vector repair. Bacteria are plated on LB-agar supplemented with 50 µg/µl Hygromycin B and grown at 37 °C overnight. For vector production, a liquid culture of a single colony from a plate is grown, and pMyCA expression vectors are harvested by using the NucleoSpin® Plasmid Easy Pure Kit (Macherey-Nagel, Germany). Since the structures of EsxF and EsxU have not yet been determined, the posi-

tions for the cysteine mutations were chosen based on the location of mutations in EsxB-TE by mere alignment of the central “WXG”-motifs (see Figure 3.1).

```

EsxB      MAEMKTDAAATLAQEAGNFERISGDLKTQIDQVESTAGSLQQQWFCAAGTAAQAAVVRFQEAANKQKQELDEISTNIRQAGVQYSRADEEQQQALSSQMGF
EsxF      MGADDTLRVFAVMQGFASLDGAAEHLAVQLAELDAQVQMLGGWRGASGSAYGSAWELWHRGAGEVQLGLSMLAAAIAHAGAGYQHNETASAQVLRVGGG
EsxU      ...DVSTPNTLNADFLLMRSVAGITDARNEEIRAMLQAFIRMSGVPPSVWGLAAARFQDVVDRWNAESTRLYHVLHAIADTIRHNEAALREAGQIHARHIAAAGDDL

```

Figure 3.1: Positioning of cysteine mutations in EsxF and EsxU according to EsxB-TE. Alignment of EsxB, EsxF and EsxU by the central WXG-motif (orange) served as a guide to determine the position of cysteine mutations (red) in EsxF and EsxU. Helical regions of EsxB are marked with boxes.

3.1.2 Protein Expression in *Mycobacterium smegmatis*

The different Esx proteins were recombinantly expressed in *Mycobacterium smegmatis* mc2155 using the respective pMyCA-expression vectors (see Section 3.1.1).

Transformation

Electrocompetent *M. smegmatis* mc2155 were transformed with the respective pMyCA expression vector (approx. 100 ng) by electroporation using 1 mm electroporation cuvettes (peqlab/VWR, USA) and setting the electroporator to a capacitance of 25 μ F, a resistance of 1000 Ω and a voltage of 2.5 V and administering one electrical pulse. After subsequent incubation on ice for ten minutes, bacteria were plated on a Middlebrook 7H10 agar plate containing 50 μ g/ml Hygromycin B. The plates were incubated for at least three days at 37 $^{\circ}$ C until colonies became visible.

Bacterial Growth and Lysis

Transformed bacteria were grown in Middlebrook 7H9 broth base supplemented with 0.05 % Tween80, 1.11 mM D-Glucose and 50 μ g/ml Hygromycin B at 37 $^{\circ}$ C. For protein expression, 1 l of the medium was inoculated with 2.5 ml of dense overnight culture and incubated for 22 hours before induction with 35 mM acetamide. After induction, the culture was grown for another 24 hours before being harvested by centrifugation at 3000 *timesg*. Bacteria were harvested and resuspended in Lysis Buffer (50 mM Tris-HCl, 300 mM NaCl, 20 mM imidazole, 1 M urea, pH 8 at 4 $^{\circ}$ C, supplemented with a self-made protease-inhibitor cocktail containing 25 μ g/ml AEBSF, 5 μ g/ml Leupeptin and 0.5 μ g/ml E-64) and lysed using a Microfluidizer[®] LM10 (Microfluidics[™], USA) with a pressure of 20,000 psi. The lysate was centrifuged at 47,810 \times g at 4 $^{\circ}$ C for one hour, and the clear supernatant was additionally filtered using a 0.45 μ m pore size filter.

3.1.3 Protein Purification

Protein purification steps were performed using the respective purification columns and an ÄKTA-pure system (GE Healthcare, USA). All EsxB-like proteins are expressed with a cleavable N-terminal His₆-tag, allowing the complexes to be purified using immobilized metal affinity chromatography (IMAC). Purification of the complexes was done at 4 $^{\circ}$ C using a HiTrap

IMAC FF 1 ml column (GE Healthcare, USA) for Ni²⁺-affinity purification. Subsequent size exclusion chromatography (SEC) was performed using a HiLoad 16/60 Superdex 75 pg column (GE Healthcare, USA). All IMAC steps were done using Tris-HCl buffers with a pH of 8 at 4 °C (Binding-Buffer: 50 mM Tris-HCl, 300 mM NaCl, 20 mM imidazole; Elution Buffer: 50 mM Tris-HCl, 150 mM NaCl, 500 mM imidazole) and SEC was performed with PBS (10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, 137 mM NaCl, 2.7 mM KCl, HCl-adjusted to pH 7.4). For the purification of the complex EsxT/EsxU from ESX-4 (wildtype and cysteine-mutants thereof), the Tris-HCl buffers were supplemented with an additional 0.1 % Triton-X-100 and 3 M urea was used in the lysis buffer. After purification of the complexes under native conditions, the dimers are dissociated into monomers using a denaturing environment by adding 8 M urea. The pure monomers are obtained by using another IMAC step to separate the His₆-tagged EsxB-like from the tagless EsxA-like proteins. After dissociation, the monomers are brought into PBS, pH 7.4, using a PD-10 desalting column (GE Healthcare, USA). Protein concentration was measured with a Nanodrop 1000™ (Thermo Scientific, USA).

3.1.4 Protein Analysis

The purification process and monomerization as well as the purity of the final protein product was monitored and analyzed by SDS-PAGE. The protein concentration was measured with a Nanodrop 1000™ (Thermo Scientific, USA) by using the appropriate extinction coefficients (ϵ) and molecular weights (MW) of the proteins, as seen in Table 3.3. The folding ability of the cysteine mutants was compared to that of wild-type proteins with CD spectroscopy.

3.2 Protein Processing

3.2.1 Crosslinking Studies

For the crosslinking studies, the well-characterized reagents EDC and sulfo-NHS were used. N-hydroxy-sulfosuccinimide (sNHS) is highly amine-reactive. The solutions were prepared just before the experiment, as these chemicals are not long-term stable in aqueous solutions. For the crosslinking studies, wild-type complexes EsxB/A, EsxD/C, EsxG/H, EsxU/T and EsxF/E and their monomers were used. The monomers were mixed together roughly in equimolar amounts in a total volume of 50 μ l and incubated at 80 °C for five minutes to evoke unfolding and monomerization. Then, the crosslinking reagents EDC and sNHS were added to the samples at a final concentration of 5 mM and 2 mM, respectively. As a positive control control, each wildtype complex was likewise incubated with EDC and NHS, however these samples were not heat-treated. To evaluate homo-oligomer formation, EsxE, EsxF, EsxU and EsxT were also incubated individually with and without the addition of crosslinking reagents. After a reaction time of 4 hours at room temperature, 10 μ l samples were taken from each reaction mixture and transferred into 10 μ l SDS-loading buffer. This terminates the crosslinking reaction because the buffer contains β -mercaptoethanol, which quenches EDC. The crosslinking reaction was analyzed by SDS-PAGE to visualize the appearance of apparent crosslinking bands with a higher molecular weight. In the experiments, EsxB-like proteins were combined with different EsxA-like proteins to evaluate possible complex formation.

Table 3.3: Parameters of different Esx-protein monomers and complexes. The molecular weight (MW) and the molar extinction coefficient at 280 nm (ϵ_{280}) of each protein are used for the spectrophotometric determination of protein concentration. The mean residue weight (MRW) is calculated by dividing the MW by the number of amino acids and is used for the analysis of CD data (see section 3.3.3). The isoelectric point pI is noted for each protein.

	Protein	MW / Da	$\epsilon_{280}/M^{-1} \cdot \text{cm}^{-1}$	MRW / Da	pI
ESX-1	His ₆ -EsxB/EsxA	22,997.2	26,470	114.41	5.1
	His ₆ -EsxB-TE/EsxA	22,897.3	26,470	113.92	5.2
	His ₆ -EsxB	13,111.3	8,480	123.69	5.6
	His ₆ -EsxB-TE	13,101.4	8,480	123.60	5.8
	EsxA	9,903.9	17,990	104.50	4.5
ESX-2	His ₆ -EsxD/EsxC	23,309.8	11,460	112.07	5.1
	His ₆ -EsxD	13,407.9	8,480	118.65	5.8
	EsxC	9,919.9	2,980	104.42	4.4
ESX-3	His ₆ -EsxG/EsxH	22,397.0	33,920	112.55	5.8
	His ₆ -EsxG	12,024.4	4,470	116.74	6.5
	EsxH	10,390.6	29,450	108.24	4.6
ESX-4	His ₆ -EsxU/EsxT	26,914.2	34,950	116.51	6.3
	His ₆ -EsxU	15,811.8	13,980	120.70	6.6
	EsxT	11,120.4	20,970	111.20	5.9
—	His ₆ -EsxF/EsxE	22,258.8	38,960	111.85	5.6
	His ₆ -EsxF-V13C/EsxE	22,003.5	38,960	110.57	5.6
	His ₆ -EsxF-VT/EsxE	22,076.6	38,960	110.94	5.6
	His ₆ -EsxF	12,706.1	20,970	116.57	5.7
	His ₆ -EsxF-V13C	12,450.7	20,970	114.23	5.7
	His ₆ -EsxF-VT	12,523.8	20,970	114.90	5.7
	EsxE	9,570.8	17,990	106.34	5.4

3.2.2 Fluorescent Labeling

Fluorophores

The fluorophores used in this work were Alexa FluorTM 488 (AF488), which acted as a donor dye and either Alexa FluorTM 647 (AF647) (both from InvitrogenTM by Thermo Fisher Scientific, USA) or Atto 647N (ATTO-TEC GmbH, Germany) acting as an acceptor dye. The excitation and emission spectra of all dyes, as well as the molecular structures of the fluorophores, are shown in Figure 3.2. The quantum yield and lifetime, as well as other features of each fluorophore are provided by the manufacturers and are summarized in Table 3.4.

Fluorescence Labeling

Fluorescent labeling of the proteins was performed via crosslinking of fluorescent maleimide-dyes to cysteine-mutants (see Figure 2.6 bottom). Alexa FluorTM 488 C₅ maleimid, Alexa FluorTM 647 C₂ maleimid and Atto 647N maleimid were used to label the EsxB-like proteins

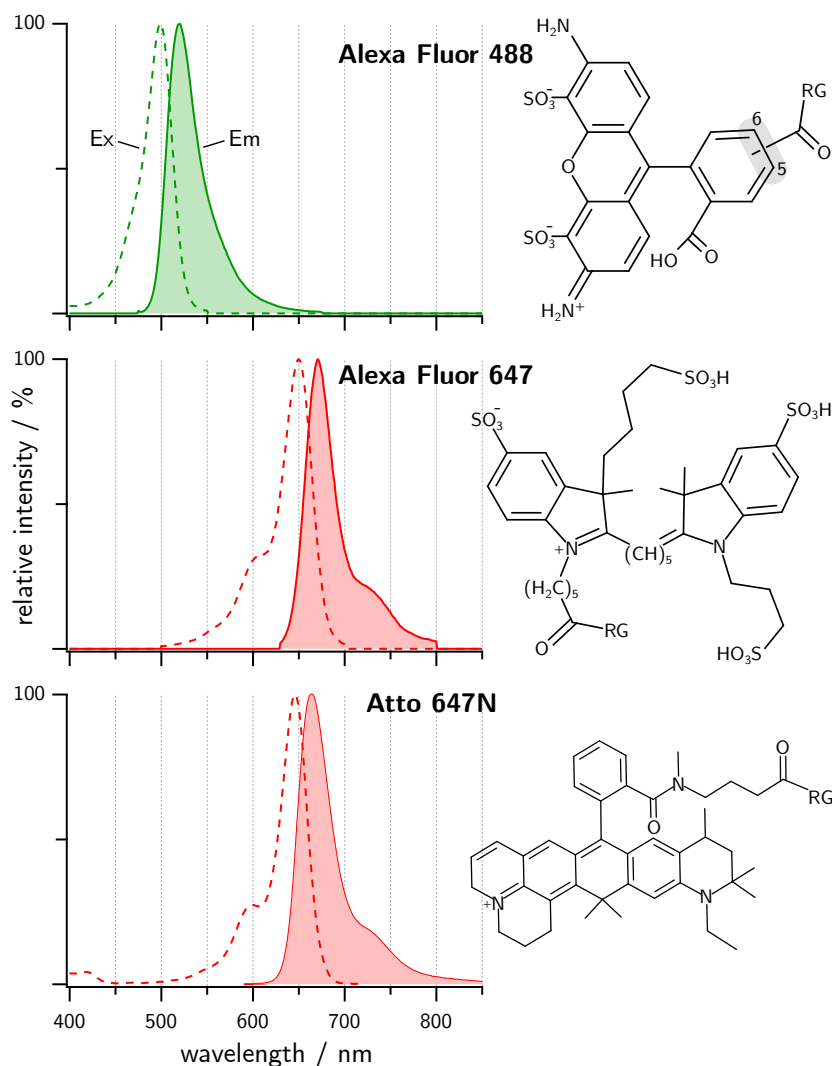


Figure 3.2: Fluorescent dyes used in this work. The excitation spectra (dashed line), emission spectra (solid line) and molecular structure of all fluorescent dyes used in this work are shown. Regarding the use as a FRET pair, AF488 acted as a donor dye (green), and either AF647 or Atto647N acted as an acceptor dye (red). In the chemical structures, RG marks the location of the respective reactive group used for protein labeling.

in complex with their respective wildtype EsxA-like partner. For this, the cysteine mutant complex (in PBS, pH 7.4) was mixed with a ten-fold molar excess of TCEP to reduce any disulfide bonds and a four-fold molar excess of fluorescent dye in a total reaction volume of 0.5 ml to 1 ml. The mixture was incubated, slowly rotating, at 4 °C for at least 12 hours. Subsequently, the protein complex was dissociated into monomers as described before (see Section 3.1.3). The monomerization process, as well as the subsequent buffer change with a desalting column (PD-10 column, GE Healthcare), ensured a clearance of any free dye from the protein solution. Successful fluorescent labeling was visualized by SDS-PAGE and gel documentation in a custom-built device prior to gel staining. Labeling efficiencies were estimated by recording a UV-Vis spectrum with a Nanodrop 1000TM and calculating the protein-dye ratio. For the calculation of the labeling efficiencies, the absorption contributions of the fluorescent dyes to (A_{280}) were subtracted from the recorded value of A_{280} and the molar

Table 3.4: Fluorophore characteristics. Fluorescence quantum yield (Φ), fluorescence lifetime (τ), wavelength of maximum absorption (λ_{ex}) and emission (λ_{em}), extinction coefficient at maximum absorption ($\varepsilon(\lambda_{ex})$), the absorption at 280 nm (A_{280}) and molecular weight (MW) are listed for each fluorophore with the respective maleimide modification. Data provided by the manufacturers [149–151].

Fluorophore	Φ	τ/ns	λ_{ex}/nm	λ_{em}/nm	$\varepsilon(\lambda_{ex})/\text{M}^{-1} \cdot \text{cm}^{-1}$	A_{280}	$\text{MW}/\text{g} \cdot \text{mol}^{-1}$
AF488	0.92	4.1	493	516	72,000	$0.10 \cdot A_{493}$	721
AF647	0.33	1.0	651	671	265,000	$0.04 \cdot A_{651}$	1300
Atto647N	0.65	3.5	646	664	150,000	$0.05 \cdot A_{646}$	868

extinction coefficients of maximum absorption ($\varepsilon(\lambda_{ex})$) were used accordingly to calculate the molar concentrations (see Table 3.4). The concentration ratio of dye to protein was used as an estimate for labeling efficiency.

3.3 CD Analysis

CD spectroscopy was used to analyze the folding behavior of the Esx monomers and complexes under different pH conditions as well as temperature-dependent unfolding and refolding. A J-715 spectropolarimeter from Jasco (JASCO Corporation, Japan) was used for the experiments.

3.3.1 Sample Preparation

The concentration of the protein solutions was measured using a Nanodrop 1000TM, and samples of 0.1 mg/ml to 1 mg/ml were used, depending on the sample. The protein concentration was chosen in such a way that the High Tension (HT) voltage did not reach the limit of 1000 V above a wavelength of 200 nm. Protein samples were in PBS (pH 7.4) or phosphate buffer (25 mM, pH 7.5).

3.3.2 Experimental Setup and Measurement Parameters

Nitrogen gas was used to purge air from the instrument optics. At least 20 minutes before starting the experiments and for the duration of measurements, a nitrogen flow with a flow rate of around 4 l/min was applied. Immediately before protein spectra were recorded, a spectrum of the respective buffer was obtained, covering the same wavelength range as the protein spectrum. Spectra were recorded between 190 nm and 240 nm or 250 nm with increments of 0.1 nm in the continuous scanning mode. The scanning speed was set to 50 nm/min, and each spectrum resulted from an accumulation of three scans. Quartz glass cuvettes from Hellma Analytics (Hellma GmbH & Co. KG, Germany) with 1 mm path length and around 350 μl volume were used for all measurements. For temperature-dependent measurements, the spectropolarimeter was equipped with the F 25 refrigerated & heating circulator from Julabo (JULABO GmbH, Germany) and the PTC-348WI Peltier temperature controller from Jasco (JASCO Corporation, Japan). The cuvettes were closed with a PTFE stopper to prevent evaporation. The ellipticity at 222 nm was recorded during a temperature

increase from 6 °C to 96 °C and during a sequential temperature decrease from 96 °C to 6 °C (both in increments of 2 °C and with a temperature slope of 1 °C/min), representing a complete temperature cycle. Before and after a temperature cycle, a complete spectrum from 190 nm to 250 nm was obtained in order to investigate the refolding capacity of the protein sample after heating and cooling.

3.3.3 Data Analysis

Data Processing

Raw data recorded by the spectropolarimeter were processed using custom-made scripts in Igor Pro 9 (by WaveMetrics, USA). First, the values of the respective blank measurements were subtracted from the spectra obtained from protein samples. Then, the curves were converted from the recorded ellipticity at a specific wavelength (θ_λ) in mdeg to the mean residue ellipticity (MRE or $\theta_{\text{MRW},\lambda}$) in $\text{deg} \cdot \text{cm}^2 \cdot \text{dmol}^{-1}$, using the mean residue weight (MRW) in g/mol (see Table 3.3), θ_λ in deg, the optical pathway d of 0.1 cm and the protein concentration c of the solution in mg/ml in Equation (3.1). This normalizes the data regarding the different protein concentrations, protein sizes, and amino acid compositions, allowing different curves to be directly compared with each other.

$$\text{MRE} = \frac{\text{MRW} \cdot \theta_\lambda \cdot 100}{d \cdot c} \quad (3.1)$$

Secondary Structure Determination

Two different web-based programs were used to interpret the obtained CD spectra for the determination of the secondary structure. Using the DichroWeb server [152], the CD data was analyzed using the “Self-Consistent” method, version 3 (SELCON3) [153]. This method works with a basis set of proteins with known structure and CD spectra [147, 154]. In this work, reference set 4 was used, as it was optimized for the wavelength range used in the experiments (190 nm to 250 nm), and results showed the best fit with the lowest normalized root mean square deviation (NRMSD). This reference set includes 43 soluble globular proteins and results in the assignment of six structural components: regular helix, disordered helix, regular sheet, disordered sheet, turn and unordered [155].

The BeStSel server [156] uses the “Beta Structure Selection” method, developed to distinguish between parallel and antiparallel β -sheets. With this method, eight different secondary structure components are defined: regular helix, distorted helix, parallel β -sheets, three types of antiparallel β -sheets with three different twists (left-handed, slightly right-handed and right-handed twisted), turn and others. The analysis is based on an optimized set of 73 reference proteins. Subsets of reference proteins and wavelength ranges are optimized for each secondary structure separately to provide the best prediction on the entire reference dataset. While being able to distinguish different β -sheet conformations, this method also shows high reliability for α -helices and low NRMSD values compared with other methods [156].

Temperature-dependent Measurements

The results of the temperature-dependent measurements were fit with complex functions based on the Gibbs-Helmholtz equation in order to determine the thermodynamics of un-

folding [157]. The fit of the ellipticity of a protein as a function of temperature allows the determination of the *van't Hoff* enthalpy ΔH and entropy ΔS of unfolding, the midpoint of the unfolding transition T_m and the free energy ΔG of folding. Assuming there is only one transition from a folded to an unfolded state, the data points can be fitted with the following equations:

$$\theta_T = \alpha \cdot (\theta_F - \theta_U) + \theta_U, \quad \text{with} \quad (3.2a)$$

$$\alpha = \left(\frac{K}{1 + K} \right), \quad (3.2b)$$

$$K = e^{(-\Delta G/(RT))}, \quad (3.2c)$$

$$\Delta G = \Delta H \left(1 - \frac{T}{T_m} \right) - \Delta C_p \left((T_m - T) + T \cdot \ln \left(\frac{T}{T_m} \right) \right). \quad (3.2d)$$

Here, ΔC_p is the change in heat capacity between the folded and unfolded state of the protein at constant pressure, θ_F and θ_U are the ellipticity of the fully folded and fully unfolded form, respectively. The temperature T is given as an absolute temperature in Kelvin, and R is the gas constant ($8.31 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$). While the method of differential scanning calorimetry (DSC) allows direct estimation of the change in a protein's heat capacity during denaturation, it is not possible to estimate this parameter from one single CD denaturing curve [120, 158]. Therefore, for initial calculations, ΔC_p is set equal to zero, simplifying the term (3.2d). If a protein has (un)folding intermediates, meaning there are two or more transitions from the folded to the unfolded form, the equations need to be extended. For two unfolding transitions, the data points are fit with the sum of two transitions as follows:

$$\theta_T = ((u_1 \cdot y_1) + (u_2 \cdot y_2)) \cdot (\theta_F - \theta_U) + \theta_U, \quad (3.3)$$

where the term $(u \cdot y)$ describes the fraction of either form with $u_2 = 1 - u_1$ while $u_1 < u_2$. The factors y_1 and y_2 are defined similar to α in (3.2) as

$$y_1 = \frac{K_1}{1 + K_1}, \quad y_2 = \frac{q - f}{d}, \quad \text{with} \quad (3.4)$$

$$d = 4cK_2, \quad f = \sqrt{8cK_2 + 1}, \quad q = 4cK_2 + 1.$$

The folding constant of the first transition K_1 is defined as in Equation (3.2), while K_2 is extended by the molar protein concentration c :

$$K_1 = \exp \left(\frac{\Delta H_1}{RT} \left(\frac{T}{T_{m1}} - 1 \right) \right), \quad K_2 = \exp \left(\frac{\Delta H_2}{RT} \left(\frac{T}{T_{m2}} - 1 \right) \right) - \ln(c). \quad (3.5)$$

ΔH_1 , T_{m1} and ΔH_2 , T_{m2} are associated with the first and second transition, respectively. If employed for the data of a dimer, it is assumed that complex dissociation occurs at the higher T_m , which would make $T_{m1} < T_{m2}$ in Equation (3.5).

van't Hoff Analysis of Thermodynamic Parameters

A different approach to analyzing CD data of thermal unfolding is the *van't Hoff* approach. From the data points, it is relatively easy to obtain the fraction of folded protein (α) at any temperature, using θ_F and θ_U , the ellipticity of the fully folded (F) and fully unfolded (U) protein, respectively. Calculating the fraction by

$$\alpha = \frac{[F]}{[F] + [U]} = \frac{\theta_T - \theta_U}{\theta_F - \theta_U} \quad (3.6)$$

with θ_T being the recorded ellipticity at temperature T and considering that the binding constant K is defined as

$$K = \frac{[F]}{[U]} = -\frac{\alpha}{\alpha - 1} \quad (3.7)$$

the binding constant at any temperature can be determined from the data. Combining the *van't Hoff* equation

$$\frac{d}{dT} \ln K = \frac{\Delta H}{RT^2} \quad (3.8)$$

with the equation for the Gibbs free energy isotherm reaction

$$\Delta G = -RT \ln K \quad (3.9)$$

and the relationship of

$$\Delta G = \Delta H - T \Delta S \quad (3.10)$$

we obtain

$$\ln K = -\frac{\Delta H}{RT} + \frac{\Delta S}{R}, \quad (3.11)$$

with R being the natural gas constant ($8.31 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$). This means, by calculating the binding constant K at different temperatures and plotting the natural logarithm of it as a function of $1/T$ in a so-called *van't Hoff* plot, the enthalpy ΔH and entropy ΔS of unfolding can be calculated by a linear fit. Looking at Equation (3.11) it becomes clear that the fit's slope multiplied with R equals the enthalpy and the fit's intercept multiplied with R equals the entropy of unfolding.

3.4 FRET Analysis

FRET experiments were performed using a custom-built confocal setup in the Institute of Physics. Experiments were performed under "single-molecule" (sm) conditions to analyze the binding, or rather folding behavior, of EsxB in the presence of different EsxA-like proteins.

3.4.1 Sample Preparation

The ideal sample concentration of the fluorescently labeled protein was determined by viewing the live auto-correlation curve. The protein concentration was adjusted so that the amplitude of the correlation curve was between seven and ten. This precaution ensures that the probability of having two fluorescent proteins in focus simultaneously is extremely low (see Section 2.4). Protein samples were diluted to the appropriate concentration using sm-PBS

buffer (PBS with 1 mM TCEP, 100 $\mu\text{g}/\text{ml}$ BSA, 0,001 % Tween, pH 7.4). Within each experiment session, the concentration of the labeled protein was kept constant. When using newly labeled batches of protein, however, the concentration was adjusted anew.

For the pH-dependent FRET measurements, phosphate-citrate buffers were prepared according to Table 3.5 and likewise supplemented with 1 mM TCEP, 100 $\mu\text{g}/\text{ml}$ BSA and 0,001 % Tween.

Table 3.5: Composition of phosphate-citrate buffers used for pH-dependent smFRET measurements.

pH	concentration / mM	
	Na ₂ HPO ₄	citric acid
4	38.6	30.7
5	51.4	24.3
6	64.2	17.9
7	87.2	6.5

3.4.2 Experimental Setup and Measurement Parameters

The schematic construction of the confocal FRET setup can be seen in Figure 3.3. Two lasers were used for the measurements: a NEWPORT 20 mW Cyan Laser CDRH (Spectra-Physics, Canada) with the wavelength of 488 nm (the “donor laser”) and either a 640 OBIS 60 mW (Coherent®, USA) with the wavelength of 641 nm or a laser diode LDH-P-C-635 (PicoQuant, Germany) with a wavelength of 635 nm (the “acceptor lasers”). The beam of the acceptor laser was filtered with a 635/10 cleanup filter (ThorLabs, USA). After coupling out of a single mode fiber (Schäfter+Kirchhoff, Hamburg), the cumulative laser beam entered an apochromatic objective (UPlanApo 4x/0,16, Olympus) and after passing through an aperture is redirected by a z405/470/633 dichroic beam splitter (AHF, Analysetechnik AG, Germany). The beam then passes a telecentric lens system with a 50 μm pinhole in the focal point into the water immersion objective (CFI Plan Apo VC 60XC WI, Nikon). Protein solutions were placed on PMMA-coated coverslips (22x22 mm, Carl Roth GmbH, Germany), and focusing was achieved using a PZT-Servo controller (Physik Instrumente, Germany). Fluorescence was collected by the same objective, the fluorescence signals were spectrally filtered using a 640DCXR dichroic beam splitter (Chroma Technology, USA) and band pass filters ET525/50M and ET685/70M (both from ThorLabs, USA) for detecting the donor and the acceptor signal, respectively. The signals were each focused onto the active area of an avalanche photodiode (APD) (SPCM-AQR-14, Perkin-Elmer™, Canada). The setup was controlled by a custom-made LabView software (National Instruments, USA). Throughout the FRET experiments, the intensity of the donor laser was adjusted to be around 40 μW , and the acceptor laser was set to between 5 μW and 10 μW (depending on the laser in use). This was achieved by using a power meter that could be inserted into the light path just before the objective, and the intensity was measured according to the respective wavelength. Measurements were conducted using “pulsed overlaid excitation” (POE) to ensure the use of double-labeled protein species in the FRET-histogram in data analysis. For this, the PicoQuant laser diode was pulsed at 10 MHz with a pulse width of < 10 ns using a Picosecond

Pulsed Diode Laser Driver (Sepia PDL 808, PicoQuant, Germany). When using the OBIS laser, pulsing was created using a frequency generator setting a pulse length of 25 ns at 2 MHz. The donor laser of 488 nm was operated in continuous mode.

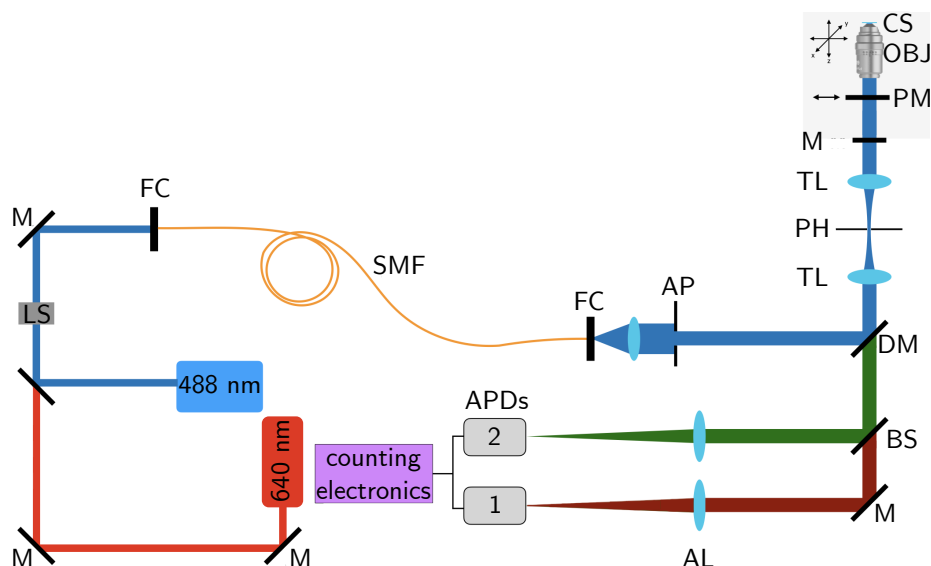


Figure 3.3: Schematic representation of the confocal FRET setup. “Donor laser”: 488 nm, blue and “Acceptor laser”: 640 nm, red are shown towards the bottom left. M: mirror, LS: laser shutter, FC: fiber coupling, SMF: single-mode fiber, AP: aperture, DM: dichroic mirror, TL: telecentric lens, PH: pinhole, PM: power meter, OBJ: objective, CS: cover slip, BS: beam splitter, AL: achromatic lens, APD: avalanche photodiodes. Figure used by courtesy of Kim C. Reiter.

3.4.3 Titration Measurements

Double-labeled EsxB-AF488-AF647 and the unlabeled monomers EsxA and EsxT were brought into smPBS. The concentration of EsxB-AF488-AF647 remained constant throughout an experiment session while increasing amounts of EsxA-like binding partners were added to each new mixture. Prior to the measurement, the mixture was heated at 72 °C for two minutes in a heat block to provoke unfolding. Depending on the experiment, photons were recorded for between 30 minutes and 8 hours.

3.4.4 smFRET under Crosslinking Conditions

Interaction and/or binding of two proteins is strongly dependent on the concentration of either partner and their affinity (see Section 2.2.2). Under “single-molecule” (sm) conditions with very low protein concentrations of at least one partner, the analysis of binding kinetics can be challenging. Therefore, chemical crosslinking was combined with FRET measurements in order to visualize protein interaction in concentration ranges below its K_d value. For this, the respective proteins were heated at 72 °C for two minutes in a heat block to provoke unfolding and mixed with EDC and NHS as described in Section 3.2.1. Right after mixing the protein solution with the crosslinking reagents, the FRET measurements were performed over the course of four hours. The native binding partner EsxA, as well as the EsxA-like protein EsxT from ESX-4, were each mixed with the double-labeled EsxB prior to crosslinking and FRET

measurement. To analyze the intramolecular crosslinking capability of the monomer itself, EsxB was mixed with NHS and EDC and measured as a control.

3.4.5 Data Analysis

Donor and Acceptor emission (DonEm and AccEm) is detected by one avalanche photon diode (APD) each. Each detected photon has two “time tags”: a Microtime-tag, which refers to the delay time after the red laser pulse, and a Macrotime-tag, which refers to the time point of the experiment when the photon was detected. This information is saved in a “.t3r” file, which is loaded and analyzed with custom-made macros in Igor Pro 9 (WaveMetrics, USA). The data analysis and processing is schematically represented in Figure 3.4 and shall be briefly explained here. With the custom-made macros, every detected photon is binned and histogrammed in two different aspects, one concerning the Microtime and the other concerning the Macrotime. For the Microtime-histogram, the photons are sorted into bins of 35 ns within a time window of 100 ns, which is the time window between two pulses of 10 MHz. For the Macrotime-histogram, the photons are binned in 1 ms increments over the whole duration of the measurement. The Microtime window (100 ns) is divided into two time-ranges, one representing the acceptor excitation (AccEx) and the other the donor excitation (DonEx). Then, each photon is sorted into one of four time-traces, depending on the excitation time window (Microtime-tag) and the emission detection as either donor or acceptor signal: DonEx/AccEm (DA), DonEx/DonEm (DD), AccEx/AccEm (AA), AccEx/DonEm (AD). All four time traces are synchronized, meaning every time point of the Macrotime is present in each time trace. Next, each of the time traces is corrected for the background signal by subtracting the average signal of a buffer measurement. Furthermore, the influence of the donor excitation caused by continuous wave donor laser during POE needs to be addressed. If FRET occurs, acceptor emission can be elicited by donor excitation. When using POE, there is a continuous donor excitation, even during the acceptor laser pulse. Hence, the acceptor signal detected in the “acceptor excitation” time window can theoretically not fully be attributed to the acceptor laser pulse if FRET occurs. Hence, the AA time trace is corrected to better represent the acceptor emission evoked by acceptor excitation. Since the window length of the acceptor excitation is only about one-third of that of the donor excitation, only a fraction of 0.34 of DA is subtracted from the AA time trace. Lastly, if the quantum yield of the fluorophores and the detection efficiencies of the optical and electronic setup differ, this, too, needs to be corrected. For this, each fluorophore has a factor γ , consisting of its quantum yield and the measured detection efficiency. The time traces DD and AD are multiplied with the ratio of these factors γ_A/γ_D to correct for these differences. Now that the data has been thoroughly corrected, there are two thresholds to be set in order to select the correct – meaning relevant – photon bins for FRET calculations. The relevant protein species for FRET experiments are only double-labeled molecules carrying both an acceptor and a donor fluorophore. To ensure the presence of an acceptor dye in the detected focused molecule, thereby eliminating donor-only forms, an acceptor threshold is selected and applied in the AA time trace. Further, to eliminate very low signals and background signals, a sum threshold is set, which refers to the sum of photons detected in the traces DD and DA. Only if both thresholds are passed will the corresponding signals of DD and DA be used to

calculate the FRET Efficiency E_T with

$$E_T = \frac{DA}{DA + DD} \quad (3.12)$$

and visualized in a histogram. The corresponding bins in DD, DA and AA are further used to calculate the stoichiometry S with

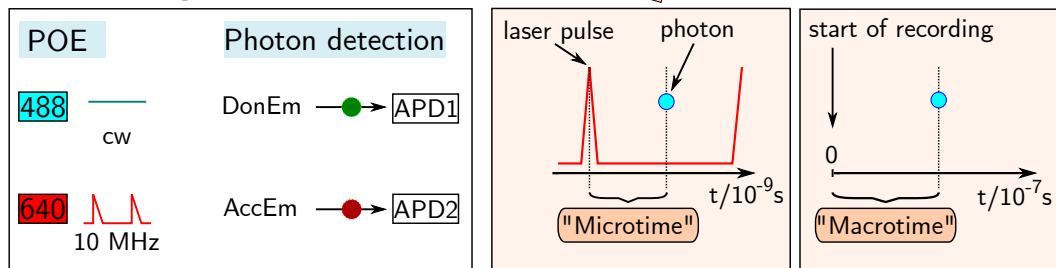
$$S = \frac{DA + DD}{AA + DA + DD} \quad (3.13)$$

to verify the use of double-labeled protein species in the FRET histogram. If a protein carries both fluorophores, in an ideal case, each fluorophore will emit signals of the same intensity (DonEm = AccEm). As the donor signal is decreased if FRET occurs, the true donor signal consists of DD + DA. If the observed species were D-only, the value of S would be 1 according to Equation (3.13). In the ideal case of DonEm = AccEm, the signal of AA would be the same as the true donor signal (DA+DD), resulting in a value of 0.5 for S. The apparent distance between the fluorescent dyes can be calculated from the resulting FRET efficiency. For this, a Gauss curve is fit onto the peaks in the FRET histograms. The x-value of the fit curves maximum is used as the average FRET efficiency and used in the following equation to calculate the distance between dyes r :

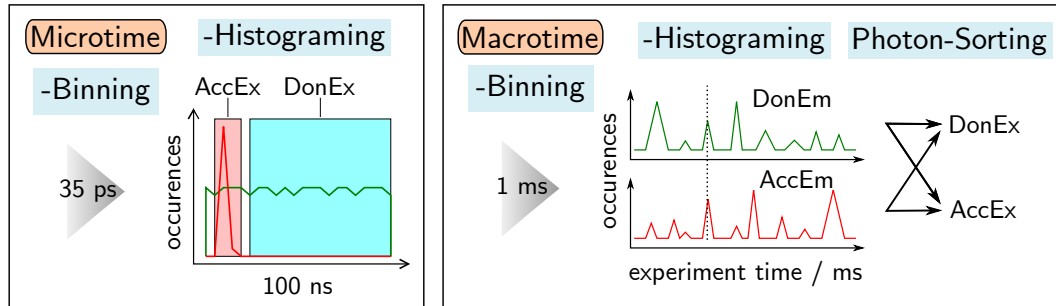
$$r = \sqrt[6]{\frac{1}{E_T} - 1} \cdot R_0 \quad (3.14)$$

with R_0 being the FRET-pair dependent Förster radius (see Figure 3.5).

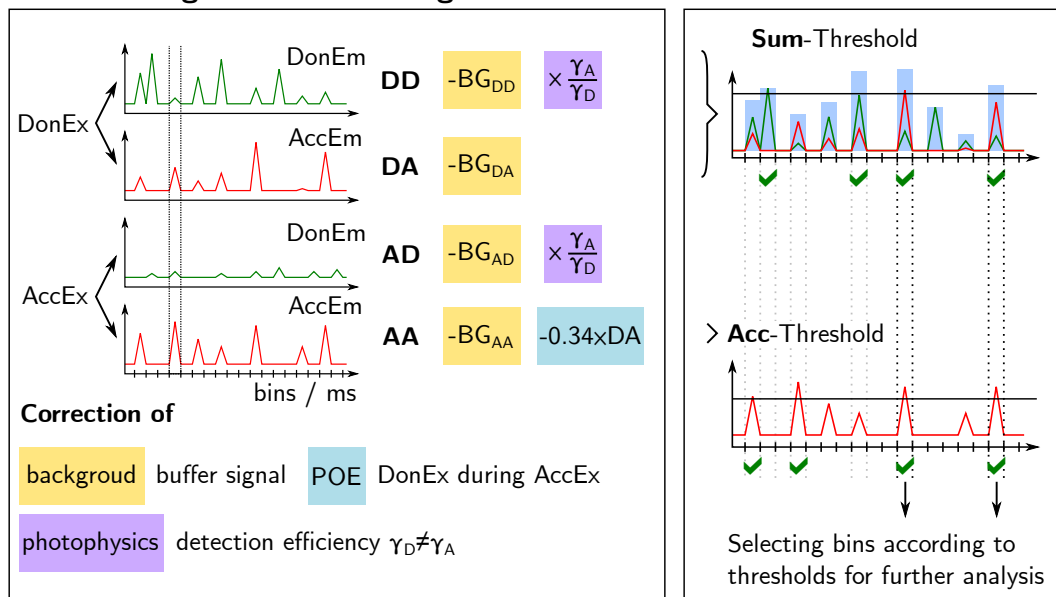
A Recording Data



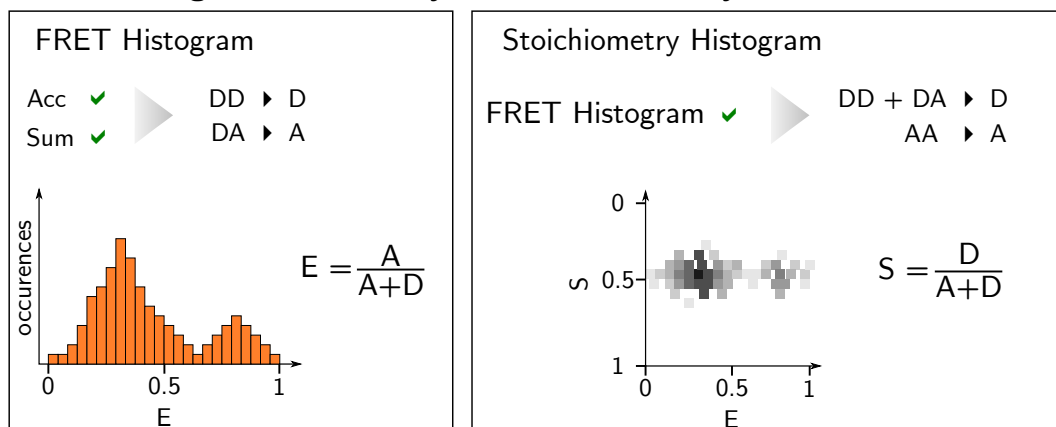
B Processing Data



C Correcting Data and Setting Thresholds



E Calculating FRET Efficiency E and Stoichiometry S



◀ **Figure 3.4: Overview of FRET data recording and analysis.** **A** Data is recorded with POE: “donor laser” (cyan) is operated under continuous wave mode, “acceptor laser” (red) is pulsed at 10 MHz. Donor (Don) and Acceptor (Acc) emission (Em) are detected by one APD each. Each photon has two “time tags”: the Microtime-tag refers to the delay time after the red laser pulse, and the Macrotime-tag refers to the photon arrival time during the experiment. **B** Every detected photon is sorted into 35 ps bins and histogrammed according to its Microtime-tag and sorted into 1 ms bins for Macrotime-histogramming. The Microtime window is divided into two time-ranges: one representing the acceptor excitation (AccEx) and the other the donor excitation (DonEx). Photons are then sorted according to their Microtime-tags from two time-traces (DonEm, AccEm) into four (DonEx/AccEm, DonEx/DonEm, AccEx/AccEm, AccEx/DonEm). **C** Data is corrected for background signal (yellow), for DonEx during POE (blue), and for the difference in quantum yields of fluorophores and different detection efficiencies by the setup (purple). An acceptor threshold is selected, which is applied in the AccEx/AccEm time trace, and a sum threshold is set, which refers to the sum of photons detected in the traces DonEx/DonEm and DonEx/AccEm. **D** The FRET efficiency (E) is calculated by using only those bins that fulfill both thresholds and calculating the ratio between acceptor (A) and donor (D) signal, which are taken from both DonEx traces. The value of the stoichiometry S corresponds to the fluorophore labeling of the protein species included in the FRET histogram, with a value of 1 representing donor-only (D) labeled protein and a double-labeled (AD) protein showing a value of 0.5.

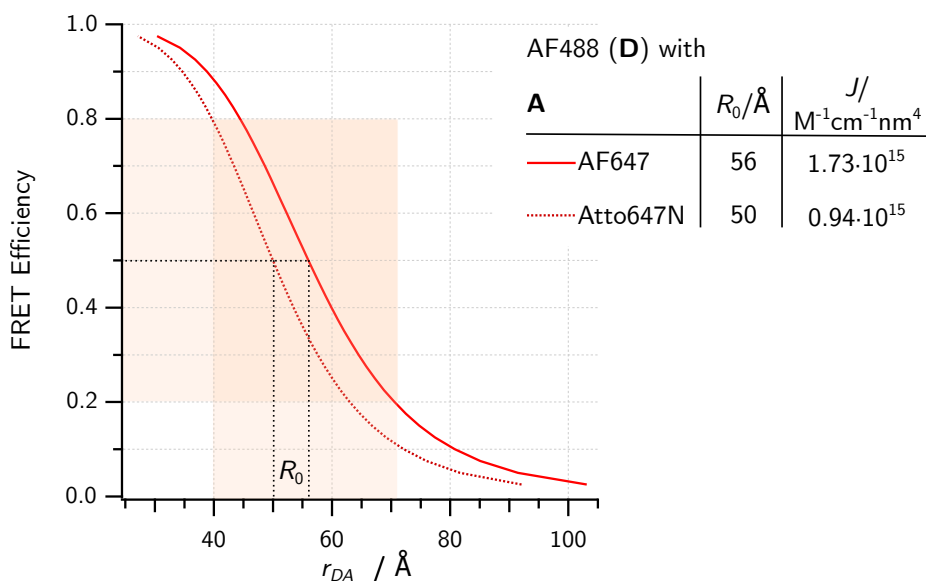


Figure 3.5: Characteristics of each FRET-pair using AF488 as a donor (D) and either AF647 or Atto647N as an acceptor (A). The graph on the left shows the FRET Efficiency (E) depending on the distance between D and A (r_{DA}). The shaded area marks the region where E is almost a linear function of r_{DA} , meaning that the detection of distance changes in this area is most sensitive. The table on the right lists the Förster distance (R_0) and the overlap integral (J) for each DA-pair. R_0 values provided by manufacturers [149–151], J calculated using the FPbase FRET Calculator online tool [159].

Chapter 4

Results

4.1 Protein Purification

Five different Esx protein pairs were purified in the course of this work, as listed in Table 4.1. While some could be abundantly expressed and easily purified (EsxB/A and EsxF/E), others proved to be more challenging (EsxU/T).

Table 4.1: List of purified Esx-complexes and yield. The purified Esx-complexes are listed alongside the approximate yield of protein in mg per liter of bacterial culture.

Protein complex	ESX-cluster	yield
EsxB/A	ESX-1	10.0 mg/l
EsxD/C	ESX-2	1.0 mg/l
EsxG/H	ESX-3	0.5 mg/l
EsxU/T	ESX-4	1.5 mg/l
EsxF/E	- -	3.0 mg/l

This section will focus on insights gained during the purification process of certain complexes. Specifically, the analysis of the steps of lysate preparation suggested a membrane-association of EsxU/T. Further, the analysis of purification steps with SEC of EsxF/E and EsxT/U hinted at the presence of homo-multimers of the EsxB-like proteins EsxF and EsxU. For comparison, Figure 4.1 shows the typical results of purification of His₆-EsxB/A (ESX-1) with IMAC and SEC. Figure 4.2 shows an exemplary SDS-PAGE result of three different purified complexes and their respective monomers. The result shows a very distinctive pattern and running behavior of the proteins, despite being of almost identical molecular weight.

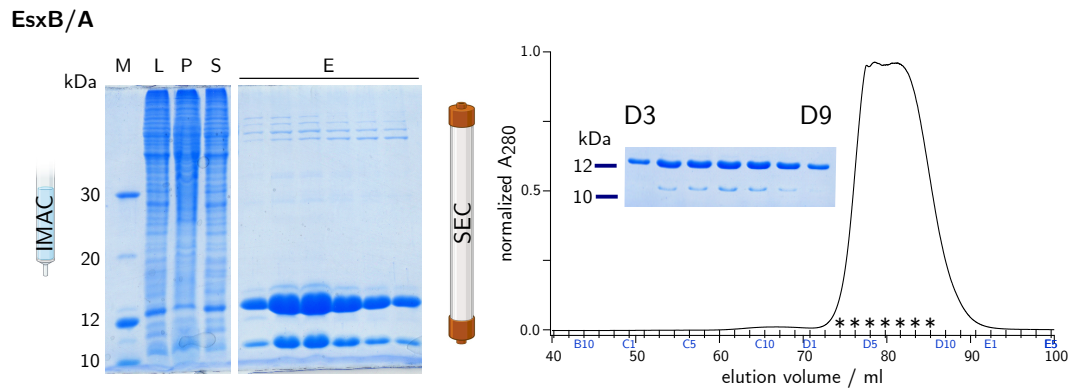


Figure 4.1: Typical purification process of EsxB/A from ESX-1. The left part shows the results of IMAC purification of His₆-EsxB/A. L: bacterial lysate; S and P: supernatant and pellet of centrifuged lysate, respectively; E: elution fractions. The right part shows a typical chromatogram of SEC purification of His₆-EsxB/A. A₂₈₀: absorption at 280 nm, elution fractions are indicated in blue, asterisks mark the fractions analyzed with SDS-PAGE, and the results of the SDS-PAGE of samples corresponding to the marked fractions are included.

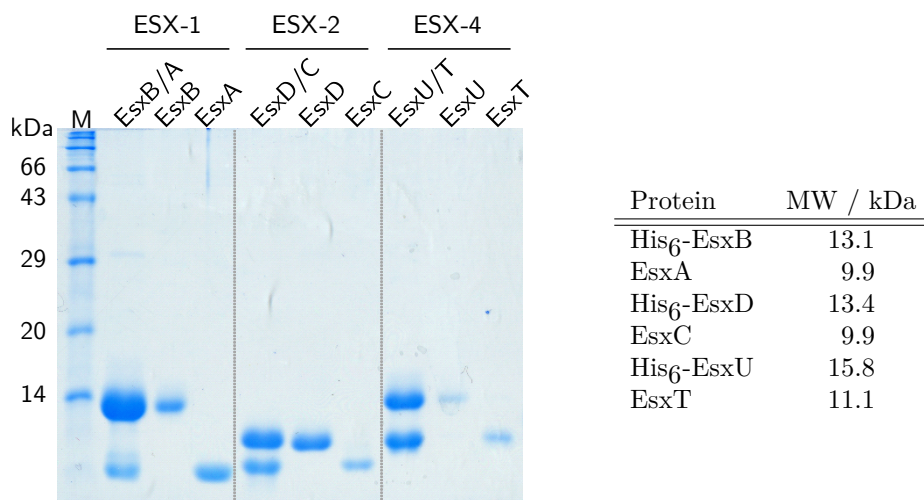


Figure 4.2: Different purified Esx complexes and their respective monomers after dissociation. 17 % SDS-Page gel with lane M being the marker of annotated apparent sizes in kDa. The actual molecular weight (MW) of the monomers from ESX-1, ESX-2 and ESX-4 are listed in the table on the right.

4.1.1 Detection of Apparent Homo-multimers Consisting Solely of EsxF

A typical purification of EsxF/E using IMAC and SEC columns is shown in Figure 4.3. Purification of His₆-EsxF/E proved to be comparatively high-yield (approximately 3 mg protein/l of bacterial culture). The IMAC purification routinely showed a contamination of around 60 kDa. However, this could be detected regardless of the Esx-complex (see Figures 4.1 and 4.3) and was identified by mass spectrometry as the *M. smegmatis* chaperone GroEL. Performing SEC with the HiLoad 16/60 Superdex 75 pg column, it was observed that elution of the His₆-EsxF/E complex peaked at an elution volume (V_e) of around 73 ml (see Figure 4.3;

SEC purification, fractions D2-D6). This V_e corresponds to globular proteins of the size of around 50 kDa, according to the calibration curve (see Appendix Figure A.1). A distinct peak of a much lower amplitude was observed just before the elution of the complex, with its center at a V_e of around 63 ml (see Figure 4.3; SEC purification, fractions C9-C11). According to the calibration curve of the column, any peaks appearing below 66 ml contain proteins with sizes above the upper separation limit of 70 kDa (referring to globular proteins). When analyzing samples taken from these fractions with SDS-PAGE, only one band could be detected, corresponding to the band of EsxF (EsxB-like). A subsequent mass-spectrometric analysis showed the band consists solely of His₆-EsxF (Appendix).

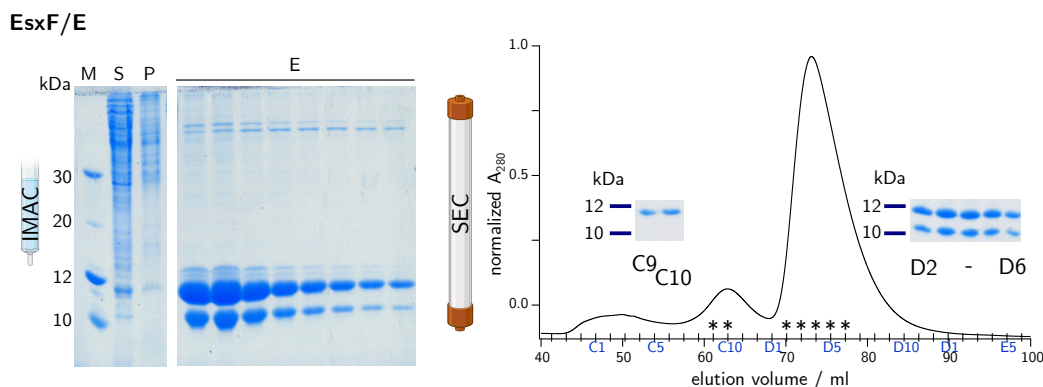


Figure 4.3: Purification of the orphan complex EsxF/E. The left part shows the results of IMAC purification of His₆-EsxF/E. S and P: supernatant and pellet of centrifuged lysate, respectively; E: elution fractions. The right part shows a typical chromatogram of SEC purification of His₆-EsxF/E. A_{280} : absorption at 280 nm, elution fractions are indicated in blue, asterisks mark the fractions analyzed with SDS-PAGE, and the results of the SDS-PAGE of samples corresponding to the marked fractions are included.

4.1.2 Indication of Membrane-association of EsxU/T and of Formation of Homo-multimers by EsxU

The typical purification process of EsxT/U is shown in Figure 4.4. The purification of recombinant His₆-EsxU/T was relatively low yield, with around 1 mg to 2 mg of protein per liter of bacterial culture. In Figure 4.4 it can be clearly seen that during the preparation of the bacterial lysate, a notably large amount of EsxU/T remains in the pellet after centrifugation when using the standard Lysis Buffer, which was used for purification of all other Esx-complexes (Figure 4.4; IMAC purification, left). However, increasing the concentration of urea from 1 M to 3 M and supplementing the buffer with 0.1 % Triton-X-100 visibly increased the amount of soluble EsxU/T in the supernatant (Figure 4.4; IMAC purification, right) and almost doubled the amount of purified protein from around 0.8 mg protein/l bacterial culture to about 1.8 mg protein/l bacterial culture. Performing SEC with the HiLoad 16/60 Superdex 75 pg column, it was observed that elution of the His₆-EsxU/T complex peaked at a V_e of around 75 ml (Figure 4.4; SEC purification, fractions D1-D7). This V_e is comparable with globular proteins of the apparent size of around 45 kDa, according to the calibration curve (see Appendix Figure A.1). A distinct peak of a lower amplitude was observed before the elution of the complex, with its center at a V_e of around 46 ml (see Figure 4.3; SEC purification, fractions B10-C2). As mentioned before, any peaks appearing below 66 ml contain proteins

above the upper separation limit of 70 kDa (referring to globular proteins). When analyzing samples taken from this peak with SDS-PAGE, only one band became visible, corresponding to that of the EsxB-like His₆-EsxU.

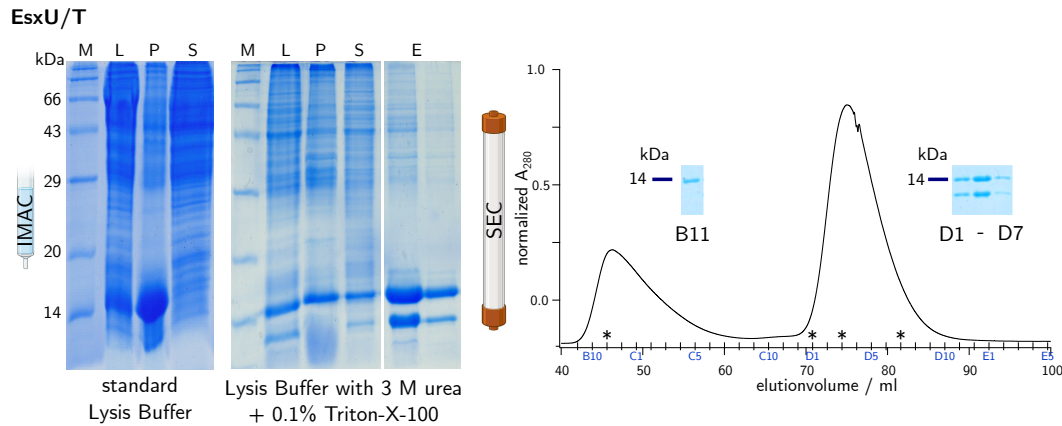


Figure 4.4: Purification of the complex EsxU/T from ESX4. The left part shows the results of IMAC purification of His₆-EsxU/T using either the standard Lysis Buffer (left) or a Lysis Buffer containing 3 M urea and 0.1 % Triton-X-100 (right). L: bacterial lysate; S and P: supernatant and pellet of centrifuged lysate, respectively; E: elution fractions. The right part shows a typical chromatogram of SEC purification of His₆-EsxU/T. A₂₈₀: absorption at 280 nm, elution fractions are indicated in blue, asterisks mark the fractions analyzed with SDS-PAGE, and the results of the SDS-PAGE of samples corresponding to the marked fractions are included.

4.2 Crosslinking Experiments

Chemical crosslinking experiments were conducted to get an overview of the possibilities of inter-loci complex formation. Purified monomers of ESX-1 (EsxB and EsxA), ESX-2 (EsxD and EsxC), ESX-3 (EsxG and EsxH), and the orphan proteins EsxF and EsxE were combined in roughly equimolar amounts and mixed with the crosslinking reagents NHS and EDC (for detailed experiment method see Section 3.2.1). The presence of a complex resulting from interacting monomers was analyzed by SDS-PAGE. Figure 4.5 shows that all conducted inter-loci combinations of EsxB-like with EsxA-like monomers treated with the crosslinking reagents show some form of covalently linked complex. For the correct interpretation of these results, it is important to examine the behavior of single monomers when mixed with crosslinking reagents without a potential binding partner to evaluate homo-complex formation. Given the insights gained from the purification steps, the focus was on the orphan monomers EsxF and EsxE (Figure 4.6) and EsxU and EsxT from ESX-4 (Figure 4.7).

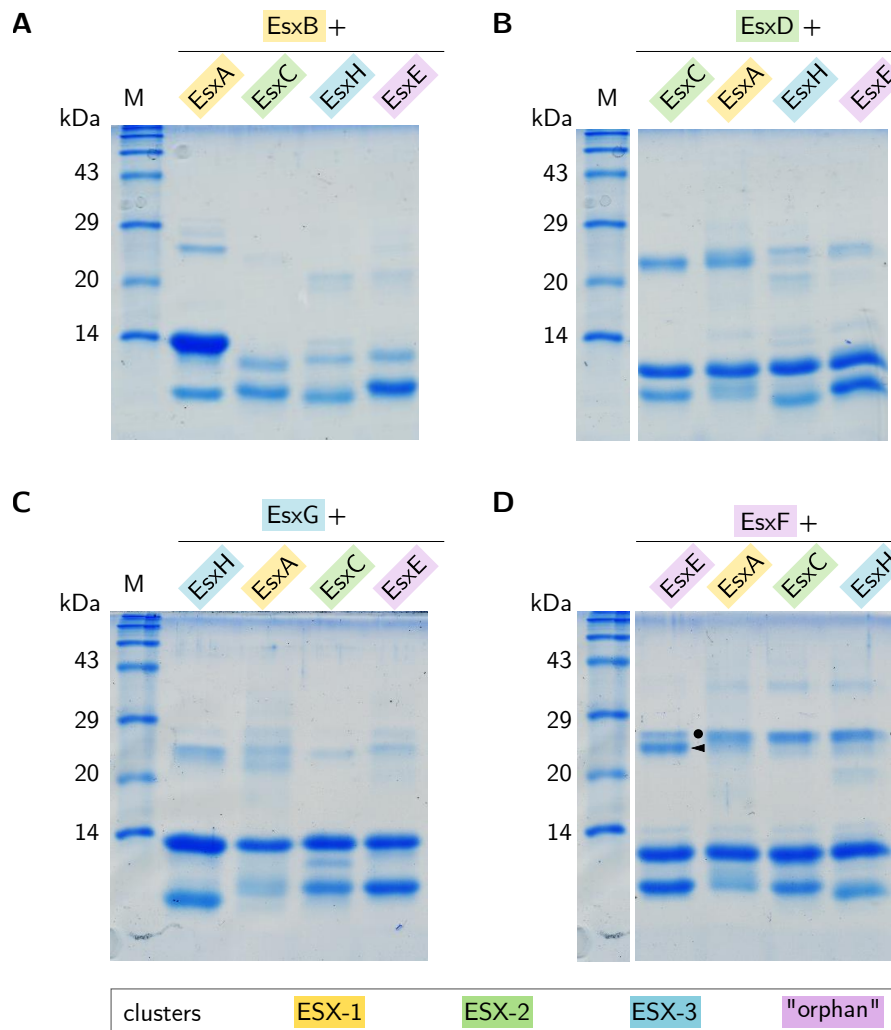


Figure 4.5: Crosslinking experiments with different Esx proteins. 17 % SDS-Page gels with lanes M being the marker of annotated apparent sizes in kDa. For the chemical crosslinking experiments, **A** EsxB from ESX-1 was combined with its wildtype binding partner EsxA and EsxA-like proteins from ESX-2, ESX-3 and the orphan EsxE, **B** the EsxB-like protein EsxD from ESX-2 was combined with its wildtype binding partner EsxC and the EsxA(-like) proteins from ESX-1, ESX-3 and the orphan EsxE, **C** the EsxB-like protein EsxG from ESX-3 was combined with its wildtype binding partner EsxH and the EsxA(-like) proteins from ESX-1, ESX-2 and the orphan EsxE and **D** the EsxB-like protein EsxF was combined with its wildtype binding partner EsxE and the EsxA(-like) proteins from ESX1, ESX-2 and ESX-3. The black dot marks the higher band; the arrowhead marks the lower band in the lane, showing EsxF cross-linked with EsxE, with the blank dot marking the crosslinking bands in the other lanes.

The crosslinking results shown in Figures 4.5 **D** and 4.6 suggest that the EsxB-like EsxF forms both, a heterodimeric complex with EsxE as well as a homo-oligomer with a slightly higher apparent weight (arrowhead and black dot, respectively). Since the EsxB-like protein partners are always slightly larger (EsxF: ≈ 12.7 kDa, EsxE: ≈ 9.5 kDa), this seems plausible. These results support the findings of the SEC presented in the preceding section, where, under native conditions, a protein species consisting of only EsxF of a larger size than the heterodimeric EsxF/E complex was observed, suggesting the presence of a homodimeric or homo-oligomeric EsxF complex. Thus, extra bands of higher apparent weight observed in the

combination of EsxF with EsxA-like proteins (as seen in Figure 4.5 **D**, blank dot) are likely the result of homo-oligomeric EsxF-complexes and would need to be investigated further.

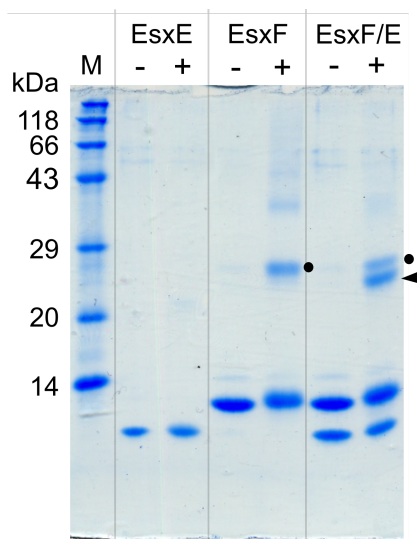


Figure 4.6: Crosslinking experiments with EsxF/E complex and monomers. 17 % SDS-Page gel with lane M being the marker of annotated apparent sizes in kDa. The monomers and complexes either underwent a one-hour incubation with crosslinking reagents NHS and EDC (+) or without the addition of these reagents (-). The black dot marks the band of an apparent EsxF-EsxF homodimer, while the arrowhead marks the apparent EsxF-EsxE heterodimeric complex band.

The crosslinking results presented in Figure 4.7 suggest that the EsxB-like EsxU can form a homo-oligomeric complex when the pure monomer is incubated with crosslinking reagents. However, in contrast to EsxF/E, the complex EsxU/T shows only one extra band of higher apparent molecular weight when treated with NHS and EDC, instead of two. Compared to the band of the EsxU/T complex, the apparent homodimeric EsxU/U complex runs slightly higher in the SDS-PAGE gel and also appears in the other combinations of EsxU with EsxA-like proteins.

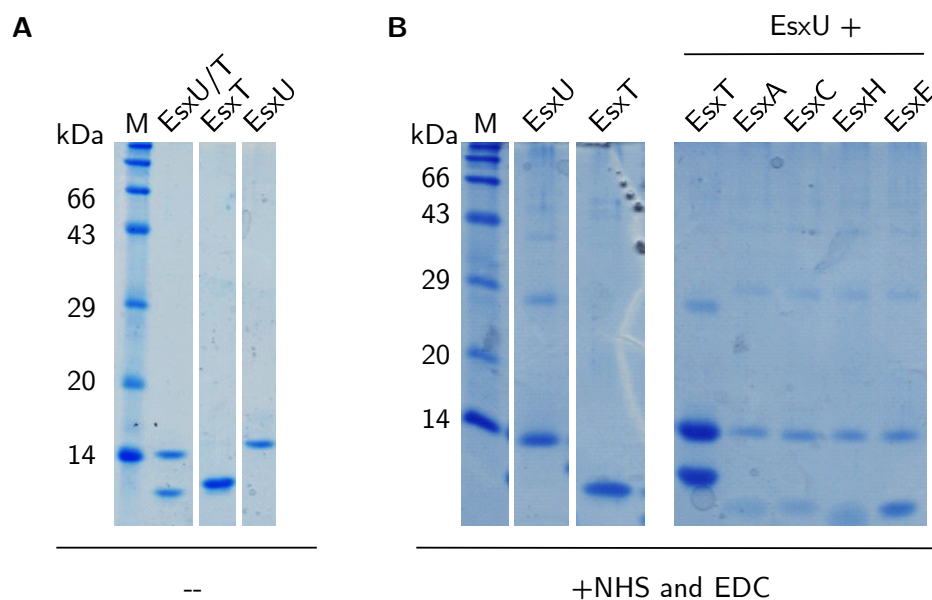


Figure 4.7: Crosslinking experiments with EsxT/U complex and monomers. 17 % SDS-Page gel with lane M being the marker of annotated apparent sizes in kDa. **A** shows the complex and monomers without the addition of crosslinking reagents, **B** shows the monomers and complex, as well as the EsxB-like protein EsxU combined with EsxA-like proteins from other ESX loci, after 2 hours of incubation with NHS and EDC at 30 °C.

4.3 Characterization of EsxF/E

This chapter is dedicated to the structural and thermodynamic characterization of the “orphan” complex EsxF/E and its monomers. The complex and monomer samples were analyzed by CD spectroscopy to evaluate the secondary structure in solution as well as in the context of thermal denaturation. As chemical crosslinking experiments suggested a homodimer formation of EsxF, smFRET measurements with fluorescently labeled EsxF were performed to evaluate this possibility further.

4.3.1 CD Spectroscopy Reveals Thermodynamics and a High Heat-resistance of the EsxF/E Complex and its Monomers

Using CD spectroscopy to analyze the structure and temperature-dependent structural changes of the wildtype complex EsxF/E and its monomers revealed a relatively high heat resistance. As seen in part **A** of Figures 4.8, 4.9 and 4.10, the CD spectra before and after the temperature cycle (TC) hardly differ. The apparent high structural recovery after heat treatment can also be observed when looking at the temperature-dependent curves at a wavelength of 222 nm. Both curves, that of the heating and that of the cooling step of the TC (part **B** of the respective figures) are practically mirror images and have almost the same ellipticity value before and after the TC.

In order to analyze the thermodynamic stability of the EsxF/E complex and its monomers in solution, the data of thermal denaturation was evaluated to compare the temperatures at the midpoint of denaturation (T_m), enthalpy ΔH and entropy ΔS of unfolding and the free energy ΔG of either sample. Assuming either one or two transitions (1T or 2T, respectively) from the folded to the unfolded state, different fits were applied to the data to yield one or two transition temperatures. With the respective fit, the enthalpy ΔH_0 of unfolding was determined (see Table 4.2), given that $\Delta C_p = 0$. The *van't Hoff* plots shown in part **C** of Figures 4.8, 4.9 and 4.10 are given to visualize the linear behavior of $\ln(K)$ as a function of T^{-1} in the transition region, meaning Equations (3.6)-(3.11) hold true. Assuming that ΔG_0 equals zero at T_m , the Gibbs free energy as a function of temperature can be calculated using ΔH_0 (see Part **D** of respective figures). Using ΔH_0 and T_m determined from the 1T or 2T fit, the entropy of unfolding ΔS_0 is calculated. The thermodynamic parameters of unfolding are summarized in Table 4.2.

The EsxF/E complex shows the highest T_m of the three samples with (65.02 ± 0.55) °C for the 1T-fit. Assuming two transitions from the folded to the unfolded form, the fit reveals T_{m1} of (30.92 ± 2.32) °C and T_{m2} of (66.18 ± 0.34) °C. The 2T model seems to fit better than the 1T model (see fit residuals in Figure 4.8 **B**). The EsxB-like monomer EsxF has the lowest T_m at (44.94 ± 0.43) °C. For the EsxE monomer, a T_m of (53.64 ± 0.69) °C was determined by the 1T-fit.

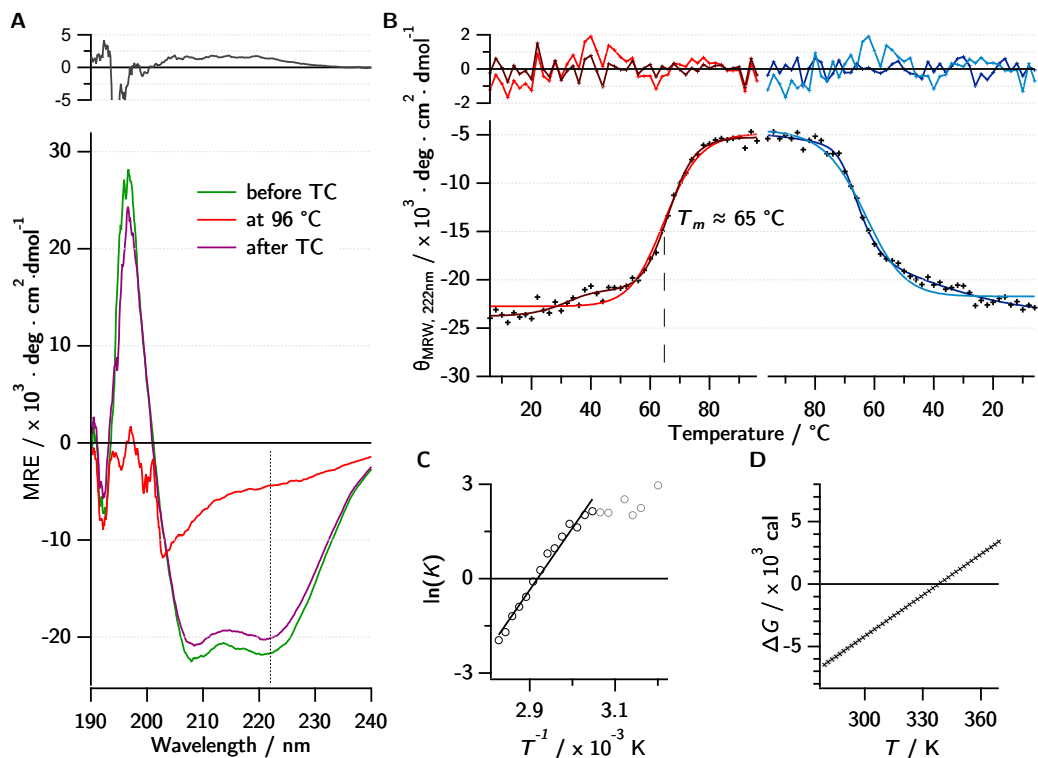


Figure 4.8: Temperature-dependent CD spectroscopy analysis of the EsxF/E complex. Graph **A** shows the CD spectra before (green) and after (magenta) the temperature cycle (TC) consisting of a heating step (from 6 °C to 96 °C) and a cooling step (from 96 °C to 6 °C) and at 96 °C (red). The divergence of the curve after TC from the curve before TC is shown in black above the graph. The wavelength of 222 nm, which is used for temperature-dependent observation of ellipticity, is indicated by a dotted line. **B** shows the temperature-dependent ellipticity at 222 nm in the course of the heating step (left) and the cooling step (right) of the TC. The graphs show the fits for a one-step transition from folded to unfolded protein during the heating step (red) and vice versa during the cooling step (blue). An additional fit for a two-step transition for each curve is included (darker colors). The approximate melting temperature T_m is indicated as determined by the fit. The fits' residuals are shown in the respective colors above each graph. **C** shows the *van't Hoff* plot. The data points used for the linear fit (saturated color) correspond to the temperature span where the transition in the melting curve in **B** shows a linear behavior. **D** shows the free energy of folding ΔG for the complex EsxF/E as a function of temperature. The standard deviation is indicated in shaded color.

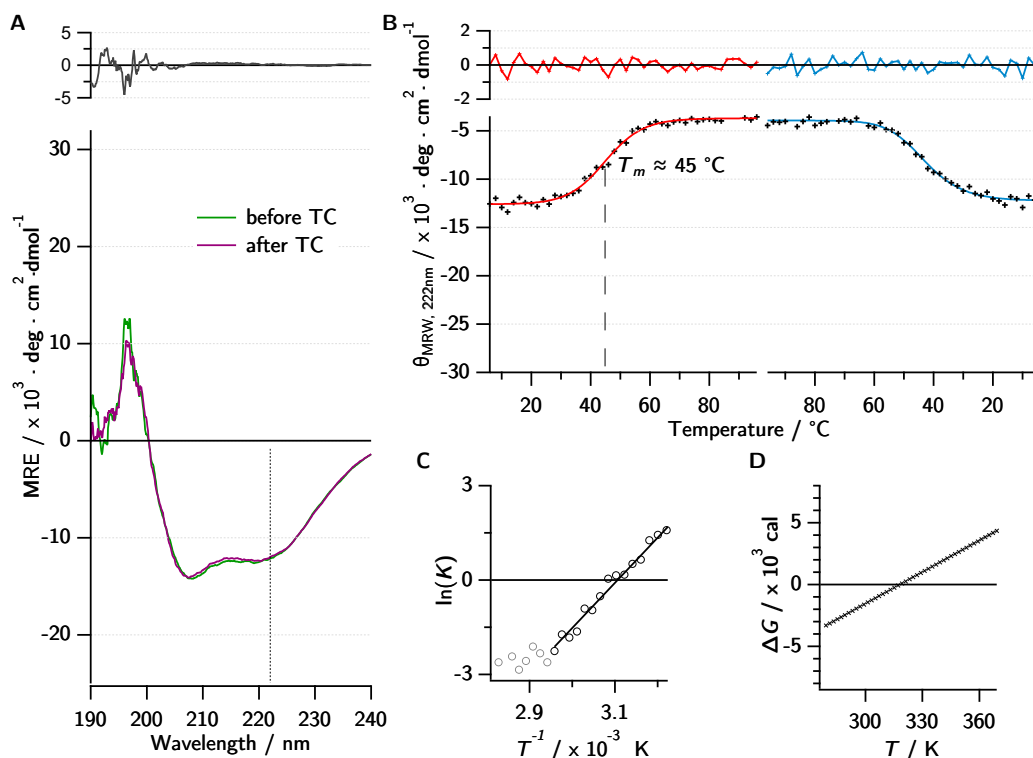


Figure 4.9: Temperature-dependent CD spectroscopy analysis of the EsxB-like monomer EsxF. Graph **A** shows the CD spectra before (green) and after (magenta) the temperature cycle (TC) consisting of a heating step (from 6 $^{\circ}\text{C}$ to 96 $^{\circ}\text{C}$) and a cooling step (from 96 $^{\circ}\text{C}$ to 6 $^{\circ}\text{C}$). The divergence of the curve after TC from the curve before TC is shown in black above the graph. The wavelength of 222 nm, which is used for temperature-dependent observation of ellipticity, is indicated by a dotted line. **B** shows the temperature-dependent ellipticity at 222 nm in the course of the heating step (left) and the cooling step (right) of the TC. The graphs show the fits for a one-step transition from folded to unfolded protein during the heating step (red) and vice versa during the cooling step (blue). The approximate melting temperature T_m is indicated as determined by the fit. The fits' residuals are shown in the respective colors above each graph. **C** shows the *van't Hoff* plot. The data points used for the linear fit (saturated color) correspond to the temperature span where the transition in the melting curve in **B** shows a linear behavior. **D** shows the free energy of folding ΔG for the EsxF monomer as a function of temperature. The standard deviation is indicated in shaded color.

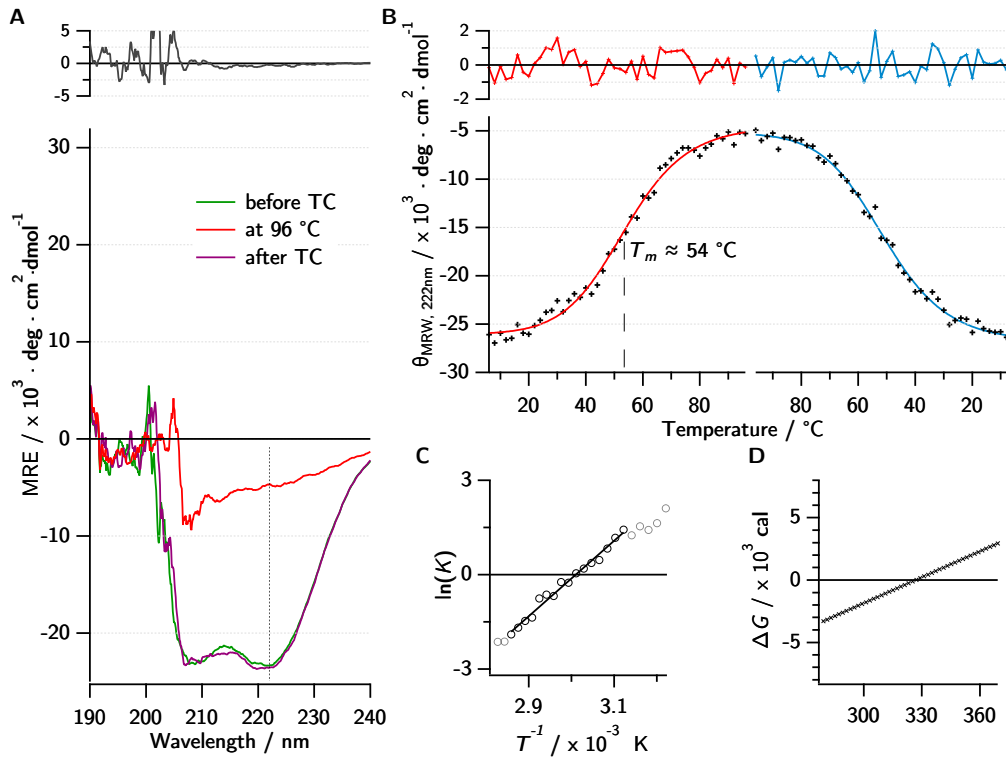


Figure 4.10: Temperature-dependent CD spectroscopy analysis of the EsxA-like monomer EsxE. Graph **A** shows the CD spectra before (green) and after (magenta) the temperature cycle (TC) consisting of a heating step (from 6 °C to 96 °C) and a cooling step (from 96 °C to 6 °C) and at 96 °C (red). The divergence of the curve after TC from the curve before TC is shown in black above the graph. The wavelength of 222 nm, which is used for temperature-dependent observation of ellipticity, is indicated by a dotted line. **B** shows the temperature-dependent ellipticity at 222 nm in the course of the heating step (left) and the cooling step (right) of the TC. The graphs show the fits for a one-step transition from folded to unfolded protein during the heating step (red) and vice versa during the cooling step (blue). The approximate melting temperature T_m is indicated as determined by the fit. The fits' residuals are shown in the respective colors above each graph. **C** shows the *van't Hoff* plot. The data points used for the linear fit (saturated color) correspond to the temperature span where the transition in the melting curve in **B** shows a linear behavior. **D** shows the free energy of folding ΔG for the EsxE monomer as a function of temperature. The standard deviation is indicated in shaded color.

Table 4.2: Thermodynamic parameters from thermal unfolding of EsxF/E and its monomers. The transition temperature T_m and the change in enthalpy ΔH_0 as determined by 1T or 2T fits. The entropy ΔS_0 is calculated at T_m , where $\Delta G_0 = 0$. R indicates the recovery in % after TC, determined from the ellipticity at 222 nm.

Sample	fit	T_m / K	$\Delta H_0 / \text{kcal mol}^{-1}$	$\Delta S_0 / \text{cal K}^{-1}$	R / %	
EsxF	1T	318.09 ± 0.43	31.83 ± 1.95	100.07 ± 6.24	96.6	
EsxE	1T	326.79 ± 0.69	19.48 ± 1.22	59.61 ± 3.88	101.2	
EsxF/E	1T	338.17 ± 0.55	38.28 ± 3.12	113.20 ± 9.41	95.5	
	2T	T1	304.07 ± 2.32	52.19 ± 3.56	171.64 ± 12.98	96.3
		T2	339.33 ± 0.34	51.37 ± 1.15	151.39 ± 3.48	

The secondary structure composition was estimated by using two different web-based programs: BeStSel [156,160] and DichroWeb [152,161]. The secondary structure compositions calculated with either program of either spectrum are presented in Figure 4.11. The mean values of the α -helical content calculated by DichroWeb and BeStSel are 51.35 % \pm 0.25 % for the heterodimer EsxF/E, 39.95 % \pm 3.85 % for EsxE and 29.50 % \pm 1.00 % for EsxF. The fractions of the secondary structures before and after TC remain mostly within the range of \pm 10 % with both monomers. The complex EsxF/E shows a decrease of helical content of 7 % after TC, in favor of an increase in “turn”- and “other” structures, when looking at the analysis done by BeStSel (Figure 4.11 **B**). In contrast to this, the analysis done by DichroWeb results in an increase of helical content of the complex after TC by over 5 % in favor of a decrease in β -sheet structure.

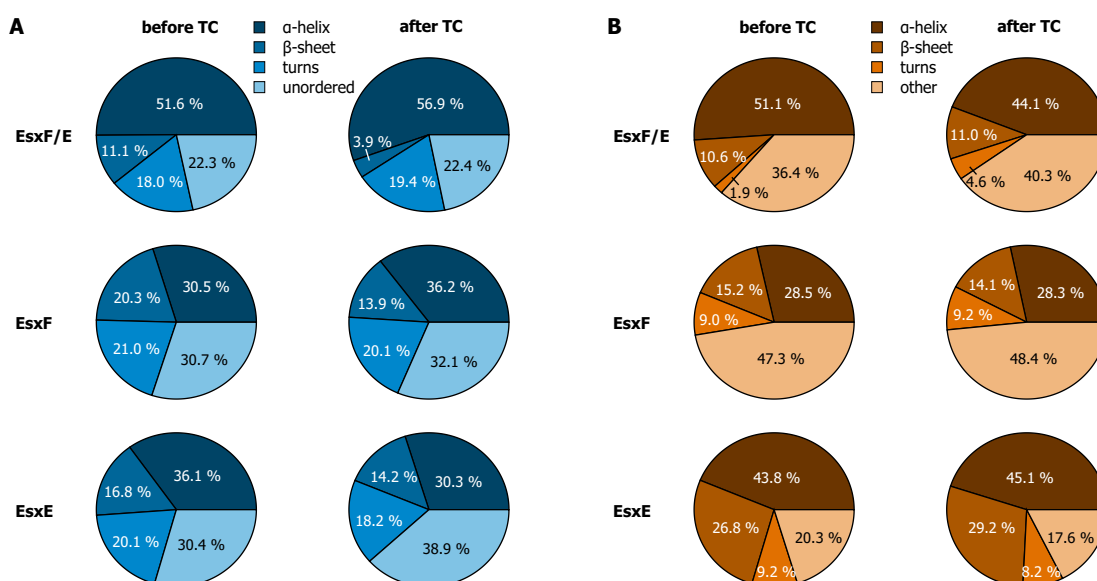


Figure 4.11: Secondary structure composition of EsxF/E and its monomers. The composition of the secondary structures of the EsxF/E complex and its monomers, as determined by deconvolution of the CD spectra before and after the temperature cycle (TC). Spectral deconvolution was done with **A** the web-service of DichroWeb [152] using the method Selcon3 and reference set 4 and **B** the web-services of BeStSel [156].

Taking together the curves of ΔG (parts **D** of Figures 4.8, 4.9 and 4.10), the change in entropy ΔS_0 and enthalpy ΔH_0 of unfolding (Table 4.2), it seems that the complex EsxF/E is most structured and stable. Further, the complex shows the highest T_m , making it more thermostable than its single monomers. EsxE has the lowest change in entropy and enthalpy while unfolding, suggesting a less structured monomeric conformation. Further, it shows the lowest change in free energy, indicating less stability of the monomer than the other two samples. Additionally, it was determined to have the lowest transition temperature, making it the least thermostable of the three samples. According to both, BeStSel and DichroWeb, the EsxE monomer is more helically structured than its binding partner EsxF (Figure 4.11).

4.3.2 smFRET: Fluorescently Labeled EsxF Suggests Structured Monomer

In order to analyze the homo-oligomer formation of EsxF suggested by chemical crosslinking studies, single- and double-cysteine mutants of EsxF (EsxF-V13C and EsxF-VT, respectively) were fluorescently labeled and used for smFRET measurements. The results are presented in Figure 4.12. In addition to the FRET histograms, the stoichiometry graphs are shown for each histogram.

Following the approach used with the kinetics studies of EsxB/A, the double-cysteine mutant EsxF-V13C-T91C (EsxF-VT) was simultaneously labeled with the fluorophores AF488 and Atto647N. The FRET histogram of this measurement is shown in Figure 4.12 **A**. The result shows high transfer efficiencies from 0.5 to 1 in a relatively broad peak with a maximum at $E = 0.93$. The corresponding stoichiometry graph suggests double-labeled protein species, meaning the histogram could result either from double-labeled monomers, complexes of such, or complexes of single-labeled monomers. Thus, it can not be determined whether the high transfer efficiency is the result of an EsxF homo-oligomer or of a tightly structured monomer. In order to investigate this further, a titration study was done, increasing the concentration of urea to observe the denaturation process (Figure 4.12 **A**). At a concentration of 7 M urea, two distinct peaks are visible: the first having its maximum at a transfer efficiency E of around 0.35, the second close to E of 1. Considering the strong denaturing characteristics of urea and the stoichiometry showing double-labeled protein species suggests that the results shown in Figure 4.12 **A** stem mostly from double-labeled EsxF or complexes of such. In order to evaluate the denaturation process, the histograms were divided into two parts, representing the fraction of folded protein (E of 0.8 - 1.0) and the fraction of unfolded protein (E of 0.0 - 0.8). The increase in the fraction of unfolded protein in dependence of urea concentration is shown in Figure 4.12 **B** (blue triangles, right y-axis). To bring this result into perspective with the secondary structure of EsxF under denaturing conditions, **B** also shows the corresponding values for MRE at 222 nm of the EsxF monomer in PBS buffer with varying concentrations of urea (black squares, left y-axis). Assuming the MRE value at 0 M urea corresponds to the completely folded form of EsxF, the blue squares represent the fraction of unfolded EsxF at the different urea concentrations in compliance with the CD data. Fitting both data sets with equation (4.2) results in a very similar value for the midpoint concentration of urea, with (2.49 ± 0.28) M for the FRET data and (2.29 ± 0.61) M for the CD data.

Furthermore, double-cysteine mutants EsxF-VT were separately labeled with either the donor or the acceptor dye, mixed together, and heated at 72 °C for two minutes to promote unfolding prior to measuring. The result of this experiment is shown in **C** of Figure 4.12. The histogram shows that the double-labeled species detected here present a rather low FRET efficiency (0-0.2). Taking into account the stoichiometric information, this result suggests complexes of donor-labeled and acceptor-labeled EsxF monomers, showing a low transfer efficiency. As there are two possible labeling sites on each monomer, leading to a relatively large number of possible oligomer-combinations, the measurement was repeated with single-labeled single-cysteine mutants EsxF-V13C (Figure 4.12 **D**). This resulted in a similar histogram as seen with the separately labeled double-cysteine mutants, showing a low FRET efficiency (0 to 0.2) predominantly.

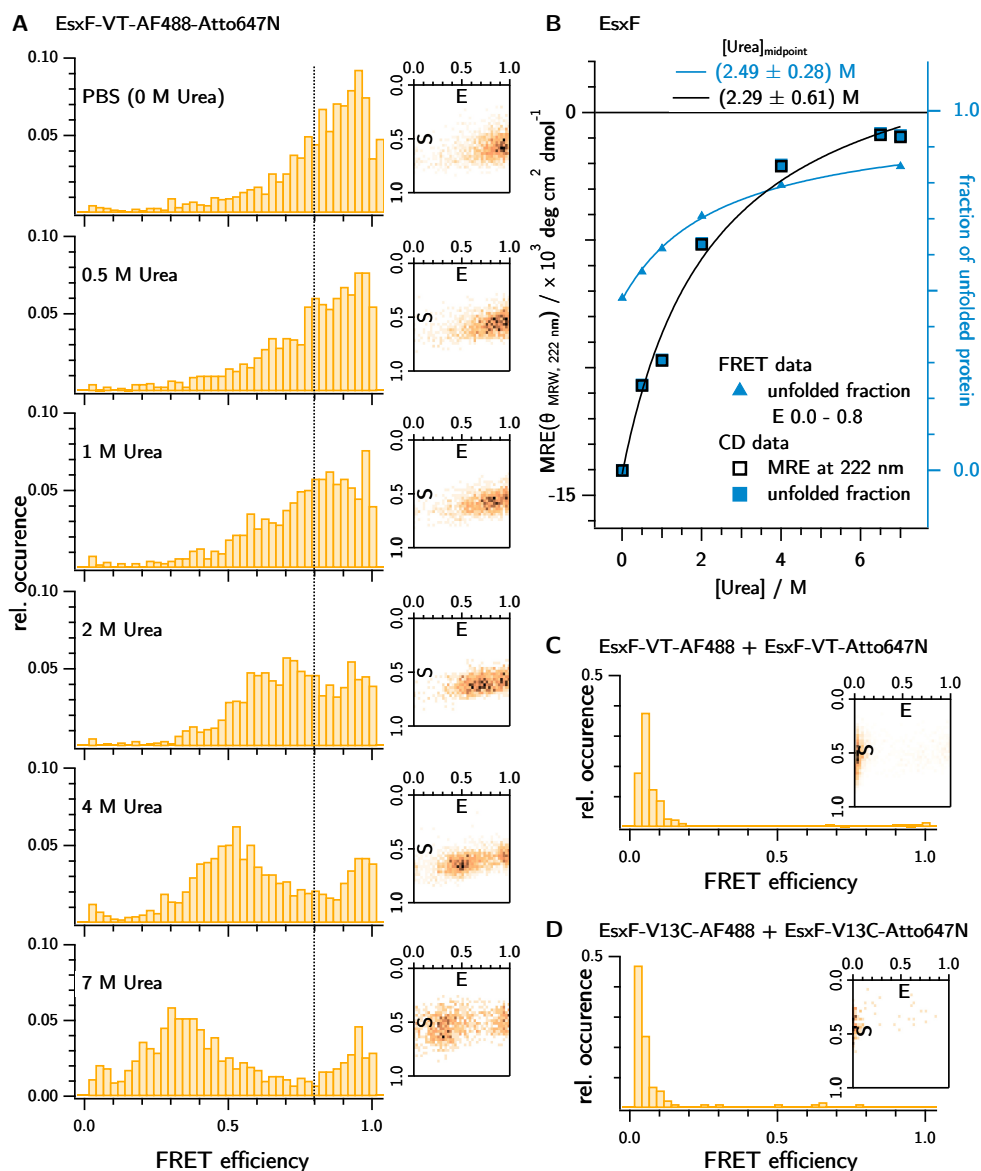


Figure 4.12: FRET measurements with double- and single-labeled EsxF monomers. The results of different FRET measurements are presented. The respective stoichiometric (S) property of the data shown in the histogram is shown in the square graph next to the histogram. **A** FRET histograms of double-cysteine mutant EsxF-VT labeled with both fluorophores simultaneously in normal buffer conditions (**top**) and with increasing amounts of urea (to **bottom**). **B** shows the normalized ellipticity of EsxF at 222 nm (left axis) in dependence of urea concentration. The scale of the right axis (blue), describing the fraction of unfolded protein, is chosen in such a way that the MRE value of EsxF at 0 M urea represents 0 % unfolded protein. The fraction of the unfolded double-labeled EsxF-VT from **A** (E 0.0 - 0.8) is shown as blue triangles. Both data sets are fit with Equation (4.2) (shown as lines) and the respective midpoint urea concentration is noted at the top. **C** FRET histogram of a mixture of double-cysteine mutant EsxF-VT labeled with either of the fluorophores. **D** FRET histogram of a mixture of single-cysteine mutant EsxF-V13C labeled with either of the fluorophores.

4.4 Characterization of EsxU/T from ESX4

The complex EsxU/T from the ESX-4 cluster and its monomers were analyzed by CD spectroscopy to evaluate the secondary structure in solution as well as in the context of thermal denaturation. As the protein purification process and chemical crosslinking experiments suggested a possible homodimer formation of EsxU, CD spectroscopy was performed and cystein-mutants of EsxU were produced for smFRET measurements to analyze this possibility further.

4.4.1 The Complex EsxU/T and its Monomers are Sensitive to Heat Treatment

Using CD spectroscopy to analyze the structure and temperature-dependent structural changes of the wildtype complex EsxT/U and its monomers revealed that the protein's structure is partially irreversibly changed by heat treatment. As seen in part **A** of Figures 4.13, 4.14 and 4.15, the CD spectra before and after the temperature cycle (TC) clearly differ – with the EsxB-like protein EsxU showing the least divergence between the recorded curve before and after the TC. This could be due to partial denaturation of protein, which does not refold to its native state. Looking at the temperature-dependent curves at a wavelength of 222 nm, the ellipticity at the end of the cooling step is clearly above the ellipticity recorded before the heating step.

In order to analyze the thermodynamic stability of the EsxU/T complex and its monomers in solution, the data of thermal denaturation was evaluated to compare the temperatures at the midpoint of denaturation (T_m), enthalpy ΔH and entropy ΔS of unfolding and the free energy ΔG of either sample. Application of the 2T-fit did not result in useful fit parameter or a better fit of the data points, therefore only the 1T fit was used. By using the fit, the enthalpy ΔH_0 of unfolding was determined (see Table 4.3), given that $C_p = 0$. The van't Hoff plots shown in part **C** of Figures 4.13, 4.14 and 4.15 are given to visualize the linear behavior of $\ln(K)$ as a function of T^{-1} in the transition region, meaning Equations (3.6)-(3.11) hold true. Assuming that ΔG_0 equals zero at T_m , the Gibbs free energy as a function of temperature can be calculated using ΔH_0 (see part **D** of respective Figures). Using ΔH_0 and T_m determined from the fit, the entropy of unfolding ΔS_0 is calculated. The thermodynamic parameters of unfolding are summarized in Table 4.3. The fit of the temperature-dependent curves at 222 nm of the complex EsxT/U under the assumption of one transition reveals a T_m of (54.05 ± 0.85) °C. For the EsxA-like protein EsxT a T_m of (46.06 ± 0.38) °C was determined. In contrast to the other temperature-dependent curves, the curves of the EsxB-like protein EsxU do not reach a clear plateau at high temperatures. This might explain why the T_m resulting from the direct fit shows a relatively large standard deviation with (70.60 ± 6.76) °C. Despite the relatively large error, EsxU seems to be the most thermostable of the three samples. Further, the CD spectrum after the TC shows the least divergence from the spectrum recorded before heating (see Figure 4.14 **A**).

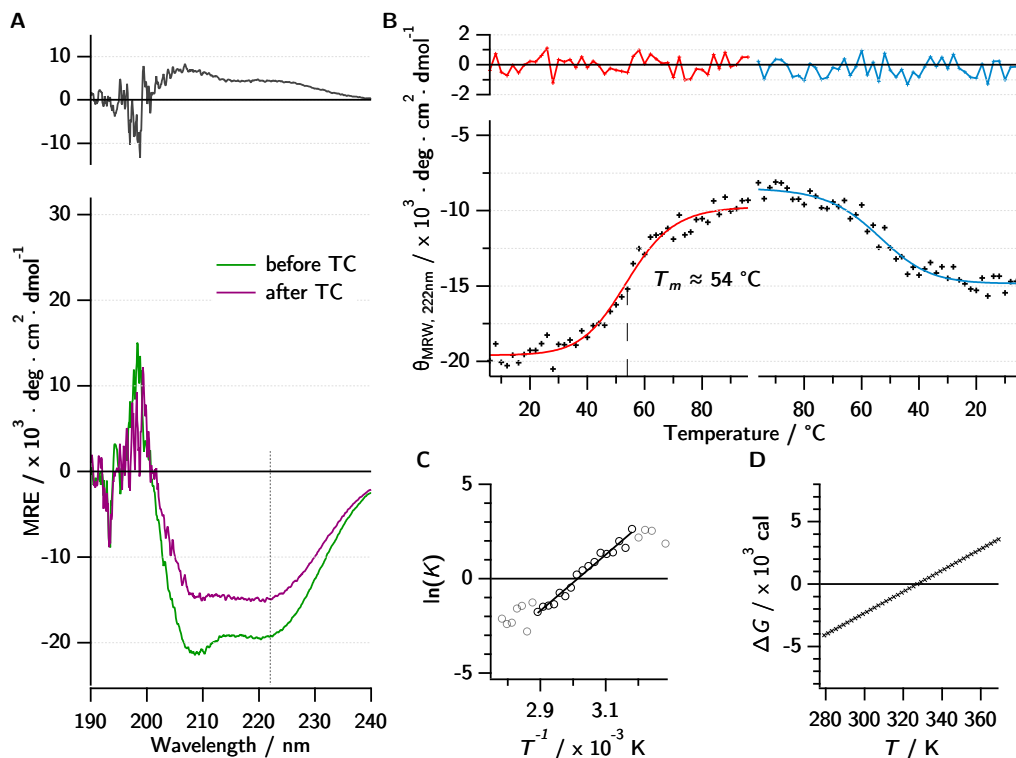


Figure 4.13: Temperature-dependent CD spectroscopy analysis of the EsxU/T complex from ESX-4. Graph **A** shows the CD spectra before (green) and after (magenta) the temperature cycle (TC) consisting of a heating step (from 6 $^{\circ}\text{C}$ to 96 $^{\circ}\text{C}$) and a cooling step (from 96 $^{\circ}\text{C}$ to 6 $^{\circ}\text{C}$). The divergence of the curve after TC from the curve before TC is shown in black above the graph. The wavelength of 222 nm, which is used for temperature-dependent observation of ellipticity, is indicated by a dotted line. **B** shows the temperature-dependent ellipticity at 222 nm in the course of the heating step (left) and the cooling step (right) of the TC. The graphs show the fits for a one-step transition from folded to unfolded protein during the heating step (red) and vice versa during the cooling step (blue). The approximate melting temperature T_m is indicated as determined by the fit. The fits' residuals are shown in the respective colors above each graph. **C** shows the *van't Hoff* plot. The data points used for the linear fit (saturated color) correspond to the temperature span where the transition in the melting curve in **B** shows a linear behavior. **D** shows the free energy of folding ΔG for EsxU/T as a function of temperature. The standard deviation is indicated in shaded color.

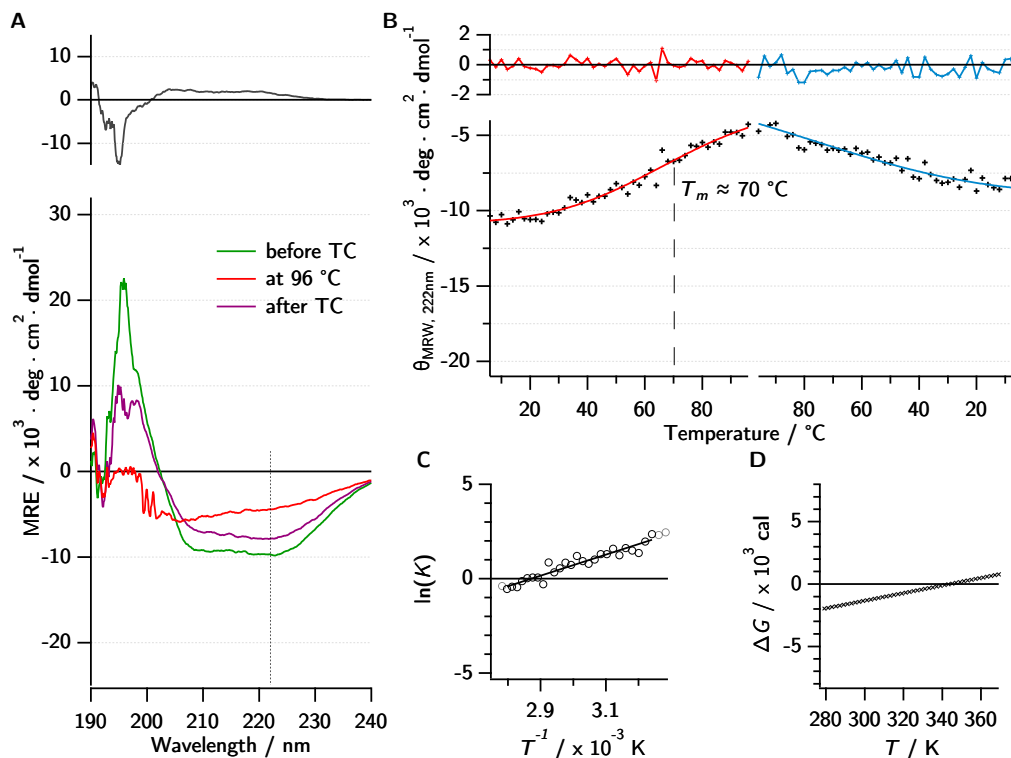


Figure 4.14: Temperature-dependent CD spectroscopy analysis of the EsxB-like monomer EsxU from ESX-4. Graph **A** shows the CD spectra before (green) and after (magenta) the temperature cycle (TC) consisting of a heating step (from 6 °C to 96 °C) and a cooling step (from 96 °C to 6 °C) and at 96 °C (red). The divergence of the curve after TC from the curve before TC is shown in black above the graph. The wavelength of 222 nm, which is used for temperature-dependent observation of ellipticity, is indicated by a dotted line. **B** shows the temperature-dependent ellipticity at 222 nm in the course of the heating step (left) and the cooling step (right) of the TC. The graphs show the fits for a one-step transition from folded to unfolded protein during the heating step (red) and vice versa during the cooling step (blue). The approximate melting temperature T_m is indicated as determined by the fit. The fits' residuals are shown in the respective colors above each graph. **C** shows the *van't Hoff* plot. The data points used for the linear fit (saturated color) correspond to the temperature span where the transition in the melting curve in **B** shows a linear behavior. **D** shows the free energy of folding ΔG for the EsxU monomer as a function of temperature. The standard deviation is indicated in shaded color.

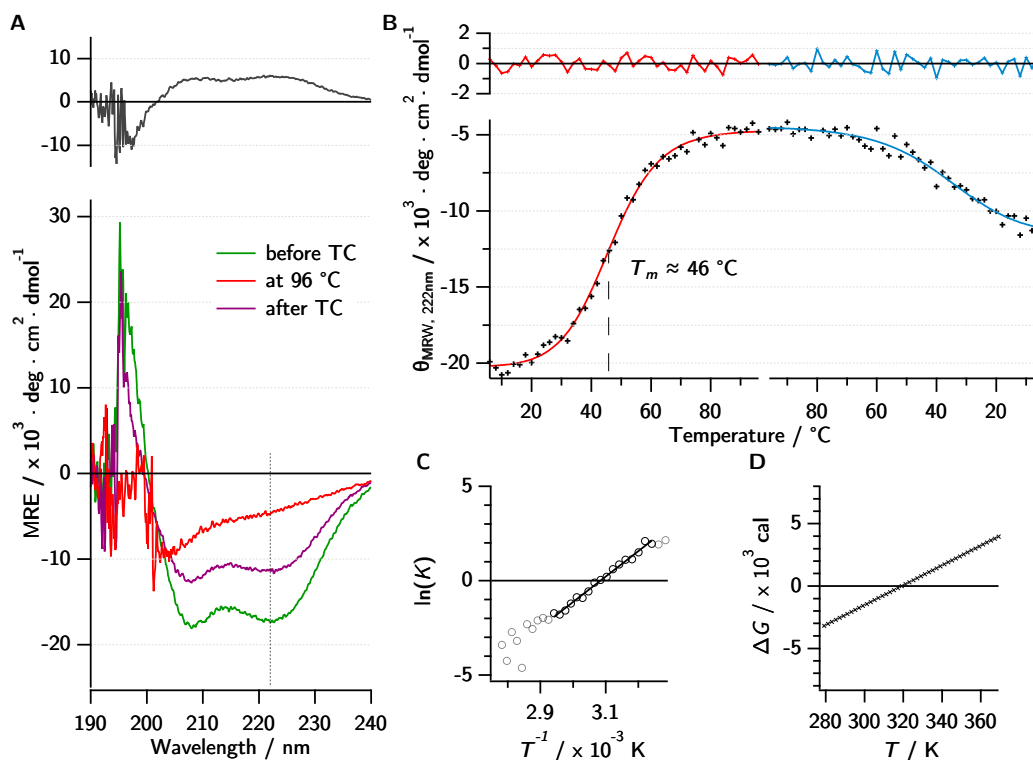


Figure 4.15: Temperature-dependent CD spectroscopy analysis of the EsxA-like monomer EsxT from ESX-4. Graph **A** shows the CD spectra before (green) and after (magenta) the temperature cycle (TC) consisting of a heating step (from 6 °C to 96 °C) and a cooling step (from 96 °C to 6 °C) and at 96 °C (red). The divergence of the curve after TC from the curve before TC is shown in black above the graph. The wavelength of 222 nm, which is used for temperature-dependent observation of ellipticity, is indicated by a dotted line. **B** shows the temperature-dependent ellipticity at 222 nm in the course of the heating step (left) and the cooling step (right) of the TC. The graphs show the fits for a one-step transition from folded to unfolded protein during the heating step (red) and vice versa during the cooling step (blue). The approximate melting temperature T_m is indicated as determined by the fit. The fits' residuals are shown in the respective colors above each graph. **C** shows the *van't Hoff* plot. The data points used for the linear fit (saturated color) correspond to the temperature span where the transition in the melting curve in **B** shows a linear behavior. **D** shows the free energy of folding ΔG for the EsxT monomer as a function of temperature. The standard deviation is indicated in shaded color.

Table 4.3: Thermodynamic parameters from thermal unfolding of EsxU/T and its monomers. The transition temperature T_m and the change in enthalpy ΔH_0 as determined by the 1T fit. The entropy ΔS_0 is calculated at T_m , where $\Delta G_0=0$. R indicates the recovery in % after TC, determined from the ellipticity at 222 nm.

Sample	T_m / K	ΔH_0 / kcal mol ⁻¹	ΔS_0 / cal K ⁻¹	R / %
EsxU	343.75 ± 6.76	10.47 ± 0.68	30.46 ± 2.59	79.8
EsxT	319.21 ± 0.38	25.37 ± 0.68	79.48 ± 2.22	54.9
EsxU/T	327.20 ± 0.85	28.00 ± 1.27	85.57 ± 4.10	67.9

To investigate the secondary structure content of the monomers and the complex, the CD spectra before and after the TC were analyzed with BeStSel and DichroWeb. The results

of secondary structure content analysis are shown in Figure 4.16. The two different analysis methods suggest the highest helical content to be found in either the complex (38.5 % according to BeStSel) or EsxT (55.3 % according to DichroWeb). According to BeStSel, the β -sheet content of EsxT is three times higher than indicated by DichroWeb (13.3 % instead of 4.4 %), while the percentage of turn structures is accordingly lower. The EsxB-like monomer EsxU shows the least helical content of the samples. According to both analysis algorithms, the decrease in helical content of either of the three samples is, for the most part, in favor of an increase in β -sheet.

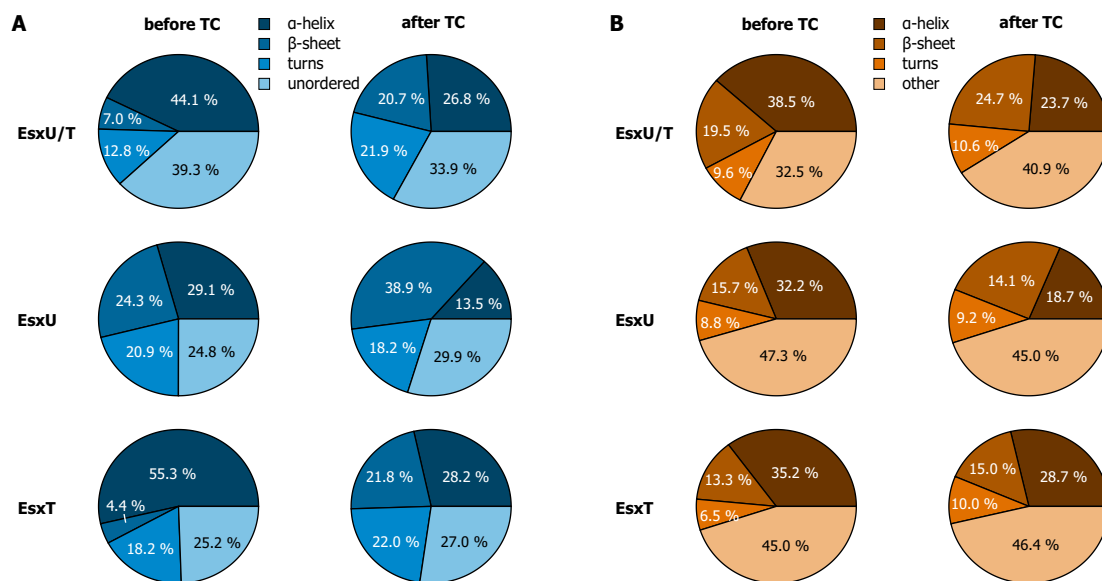


Figure 4.16: Secondary structure composition of EsxU/T and its monomers. The composition of the secondary structures of the EsxU/T complex and its monomers, as determined by deconvolution of the CD spectra before and after the temperature cycle (TC). Spectral deconvolution was done with **A** the web-service of DichroWeb [152] using the method Selcon3 and reference set 4 and **B** the web-services of BeStSel [156].

Taking together the curves of ΔG (part **D** in Figures 4.13-4.15, the change in entropy ΔS_0 and enthalpy ΔH_0 of unfolding (Table 4.3), it seems that the complex EsxU/T is most structured and stable, followed by the EsxA-like EsxT monomer. EsxU has the lowest change in entropy and enthalpy while unfolding, suggesting a less structured monomeric conformation. At the same time, it shows the lowest change in free energy, indicating less stability of the monomer than the other two samples. Interestingly, the recorded CD spectrum of EsxU before and after a complete TC resemble one another the best (compared to the other two samples), suggesting a high structural recovery after heat treatment. Together with the relatively high T_m , EsxU is the most thermostable of these three samples.

4.5 Single Molecule FRET (smFRET) Experiments

To establish the smFRET method with double-labeled EsxB for inter-loci complex formation studies, the FRET efficiency was analyzed over the period of several hours, both of the EsxB monomer and after the addition of the native complex partner EsxA. Likewise, double-labeled EsxB was measured with an increasing amount of the EsxA-like protein EsxT from ESX-4 to investigate inter-loci complex formation.

4.5.1 EsxA Titration to Study Binding Kinetics

The concentration of double-labeled EsxB was chosen in such a way that the correlation board of the setup showed an amplitude of between 7 and 10 for the auto-correlation curve (see section 3.4.1) and was kept consistent during every experiment block. The results presented here are grouped into four experiment blocks, according to the time range in which the experiments were conducted (April 2019, October 2019, January 2021 and March 2021). The respective data points in the graphs and values in the tables of each experiment block are indicated by color and the first letter of the block's month.

FRET measurements of the EsxB-AF488-AF647 monomer resulted in a histogram showing a centrally positioned peak with a maximum slightly below a FRET efficiency of 0.5 (see Figure 4.17, left). Even over the course of several hours, the monomer showed no significant changes in the resulting histogram. In total, the results of 23 measurements of the double-labeled EsxB monomer were analyzed and evaluated. Each histogram, resulting from a measurement of one hour, was fit with a Gauss-function in Igor Pro 9 in the FRET efficiency range between 0.2 and 0.8 as shown in Figure 4.17. The x-values of each curve's maximum are summarized in a box plot (Figure 4.18, left) and show a median of around 0.47. Considering the interquartile range (IQR), the median x-value of gauss-fits in 50 % of the monomer measurements is 0.466 ± 0.018 . According to equation (3.14), this corresponds to a distance of around 57 Å between the fluorophores.

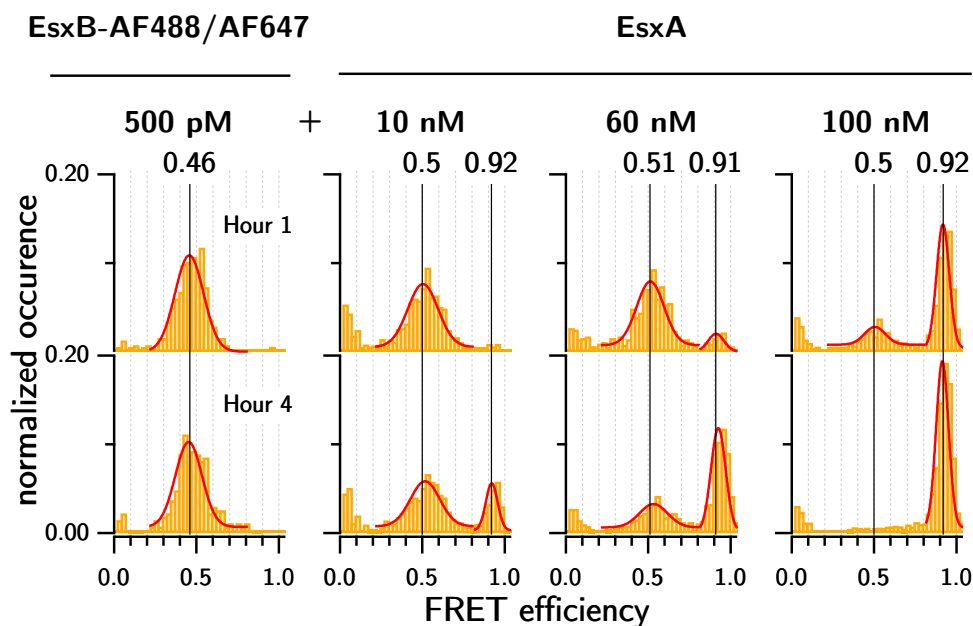


Figure 4.17: EsxA Titration. The first and fourth hour of each experiment is shown for selected titration steps to elucidate the change over time. The leftmost panel histograms show the FRET measurements of the monomer. To its right are the histograms of three titration steps, namely those with the addition of 10 nM, 60 nM, and 100 nM EsxA, respectively. The red curves are gauss-fits of the FRET histograms over the FRET efficiency range of 0.2 to 0.8 and 0.8 to 1.0, where applicable. The black vertical lines highlight the position of the first hour’s maximum.

When adding the native binding partner EsxA to the double-labeled EsxB, a peak at a FRET efficiency of over 0.9 appears in addition to the peak mentioned before. With increasing measurement time and/or with increasing EsxA concentration, the amplitude of this central peak progressively decreases while the amplitude of the peak at a high FRET efficiency increases. As shown in Figure 4.17, the central peak is still present after four hours with the addition of 10 nM EsxA, while the addition of the ten-fold amount of the binding partner results in the disappearance of this peak after the same time. A FRET efficiency of 0.92 corresponds to a distance of around 37 Å between the fluorophores.

Interestingly, the position of the central peak’s maximum is –in most cases– slightly shifted towards a higher FRET efficiency upon the addition of EsxA. This can be seen in the lower FRET histograms of the measurement of the double-labeled EsxB with 10 nM or 60 nM EsxA in Figure 4.17. It becomes even more evident when looking at Figure 4.18. Here, the data points in the right box plots represent the maxima of Gauss fits of the central peaks of over 50 measurements of EsxB in the presence of EsxA. Again, each point represents the result of a one-hour-measurement. However, the EsxA concentrations of the resulting data vary among each other and, therefore, can not be directly compared. Furthermore, some of the evaluated histograms result from dividing a single measurement of several hours into a number of consecutive one-hour experiments. A possible change in the histograms over time must also be taken into account before interpreting this observation. Therefore, the data points have been color-coded to indicate whether they stem from an experiment with a lower or higher EsxA concentration and whether they were obtained from a later time point

of a measurement. The points seem to be clearly divided into two distinct groups, termed “high” and “low”. While results from within the first two hours of an experiment can be seen in both populations, results from later time points are found exclusively in the “high” population, regardless of the EsxA concentration. To analyze the statistical significance of this result, the Dunnett’s test was performed. It is based on the student’s t-test and is utilized to compare the means of different experimental groups to a control group [162]. Hence, the data points of the “high” and “low” group from the measurements with EsxA were compared to the monomer group, which served as the control. As expected, the “low” group is not significantly different from the monomer group. However, the “high” group’s mean differs significantly, with a p -value of about $1.8 \cdot 10^{-6}$.

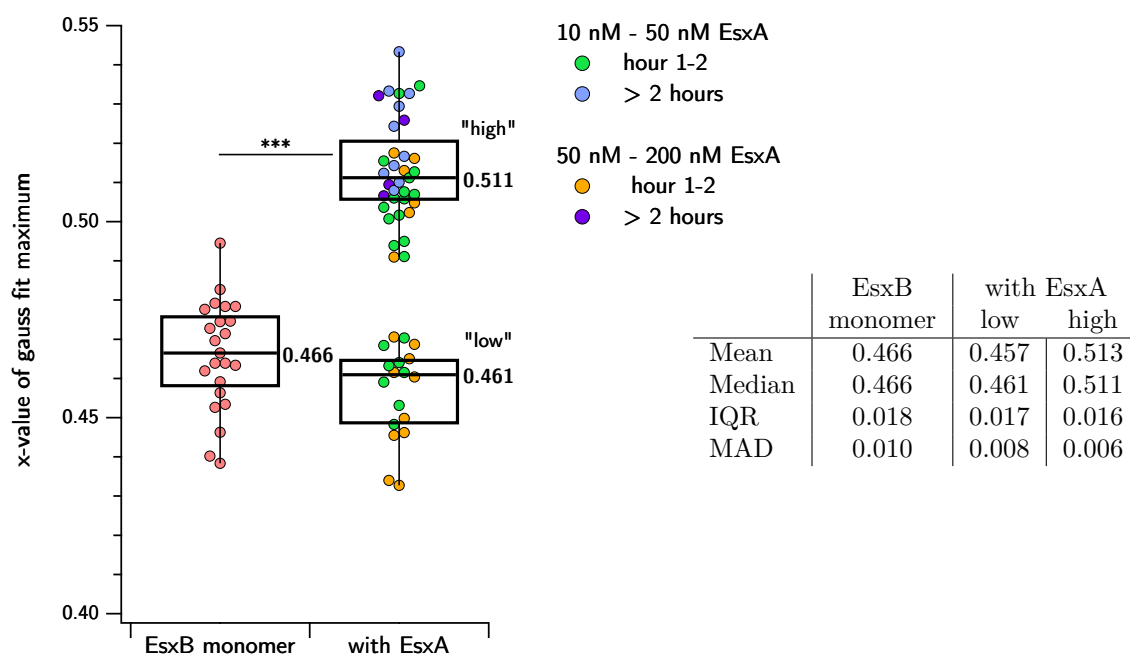


Figure 4.18: Change of center peak position in FRET histogram. The x-values of the Gauss fit’s maxima in FRET histograms are different when the EsxB monomer is measured alone (left) or with its binding partner EsxA (right). The median of each box plot is indicated next to its line. In the right plot, there are two distinct distributions of data points, indicated as “high” and “low”. Box plots for each distribution are shown, and medians are indicated next to each box. The median of the high group differs significantly from that of the monomer measurement on the left. The asterisks “***” refer to this significance with a very low p -value of only about $1.8 \cdot 10^{-6}$. The color code of the data points refers to the concentration of EsxA (either 10 nM to 50 nM or 50 nM to 200 nM) and to the hour of the experiment (being within the first two hours or later). The table on the far right gives statistical values of the three groups: EsxB monomer, the low and high population with EsxA. IQR = interquartile range, MAD = median absolute deviation.

As seen in Figure 4.17, the FRET histogram changes over time regarding the ratios of the peaks. This observable change is dependent on the EsxA concentration. Since the protein mixture is heated just prior to the measurement to induce complete unfolding of the proteins, the start point of the measurement can be viewed as the start of the binding process between

the two monomers ¹. A time-wise evaluation of multiple-hour measurements allows the analysis of the time-dependent changes in the histogram. For this, each histogram is divided into three regions according to the occurring peaks: low FRET efficiency (0 to 0.2), mid FRET efficiency (0.2 to 0.8), and high FRET efficiency (0.8 to 1). Assuming the peak at a high FRET efficiency of around 0.9 represents EsxB in complex with EsxA, the fraction of bound EsxB (F_b) can be calculated for every histogram and the increase over time can be visualized as in Figure 4.19. The data points for each EsxA concentration are fit with a single exponential function

$$F_b = A \cdot (1 - e^{-kx}) \quad (4.1)$$

where A is the normalized concentration of EsxB at equilibrium and k equals the observed pseudo-first-order rate constant k_{obs} . With an increase in EsxA concentration, saturation is reached faster (k_{obs} increases).

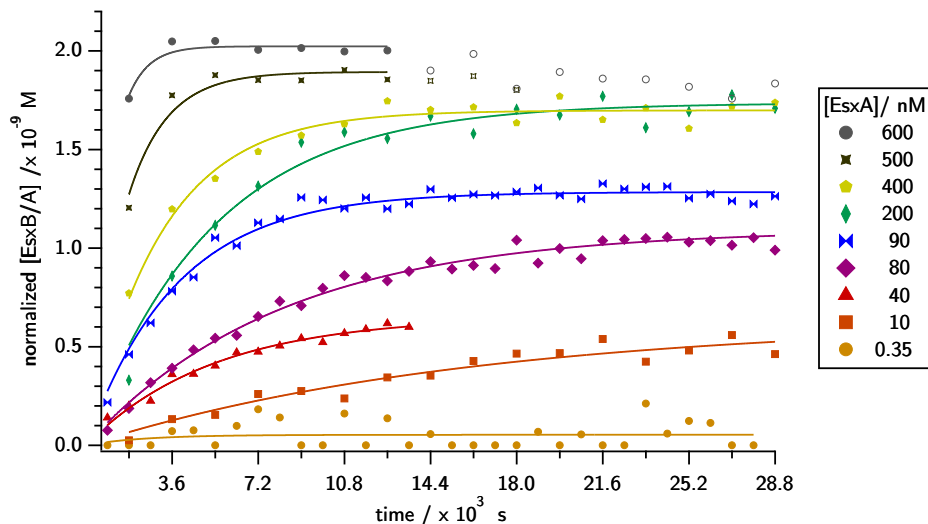


Figure 4.19: Time-dependent progression of EsxB/A complex formation in dependence of EsxA concentration. Each experiment was analyzed by dividing the data in consecutive periods of 15 or 30 minutes. From the resulting FRET histograms, the ratio of complexed EsxB, meaning the ratio of the peak between 0.8 and 1.0 FRET efficiency, was calculated. Data points are fit with a “pseudo-first-order” function. The curve’s respective EsxA concentration is indicated on the right. Data points of the measurements with 500 nM and 600 nM EsxA after four hours were excluded from fit (points not filled), as they showed a decreasing trend.

Once the data points have reached a plateau, an equilibrium is reached, meaning k_{off} equals k_{on} . The fit parameters A and k_{obs} of the EsxA-titration experiments are listed in Figures 4.20 and 4.21. Plotting A against the respective EsxA concentration reveals the concentration-dependent extent of complex formation. This is shown in Figure 4.20. As EsxA is present in large excess over EsxB, the amount of free, unbound EsxA can be considered to be only negligibly lower than that of the total amount of EsxA in the solution ($[\text{EsxA}]_{\text{free}} \approx [\text{EsxA}]_{\text{total}}$). By determining the half-saturation, the equilibrium dissociation constant K_d can be derived.

¹given that no significant amount of time has elapsed between heating of the protein mixture and the start point of the FRET measurement

The fit function to obtain the K_d value from the binding data is a sigmoidal function

$$F_b = A \cdot \left(\frac{x}{x + K_d} \right) + O \quad (4.2)$$

with A being the maximum concentration reached by complexed EsxB and O being a term for a possible y-axis offset. Performing a global fit of the data by linking the K_d value and holding O at the y-offset of $1.14 \cdot 10^{-11}$ M (which was determined for the EsxB-monomer without the addition of EsxA) results in a K_d of 12.9 nM with a large error of 322.9 nM.

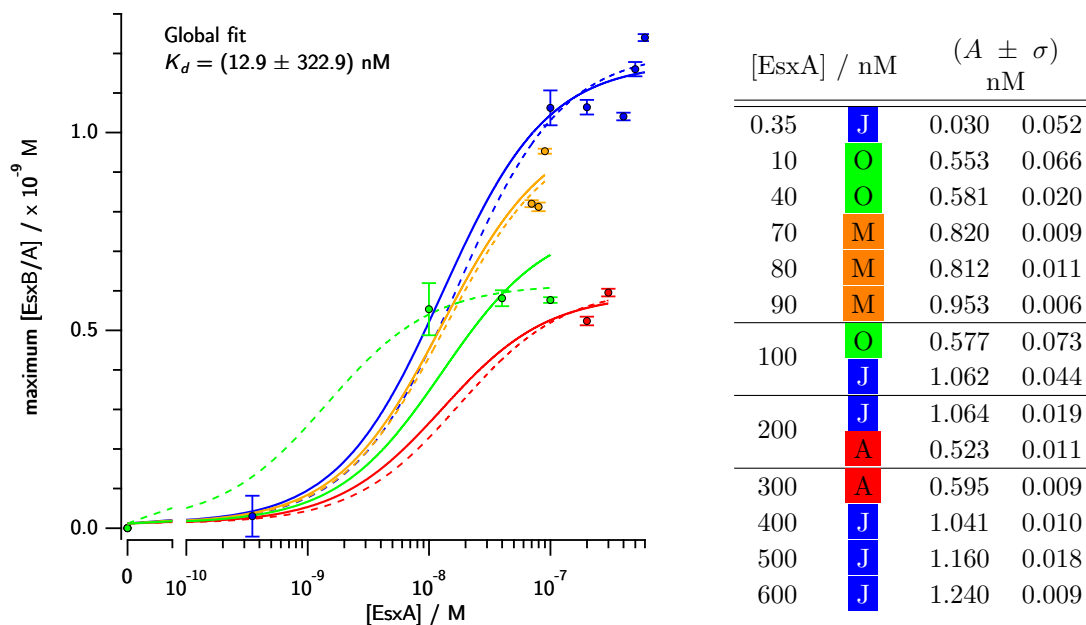


Figure 4.20: Maximum EsxB/A complex concentration measured with the addition of varying concentrations of binding partner EsxA. The values determined by the pseudo-first-order fit of titration experiments are listed in the table on the right. Color- and letter-coding in the graph and the table refer to the different experiment blocks. Error bars implicate the standard deviation. Individual sigmoidal fits are shown in dashed lines, and global sigmoidal fits linking the K_d value are shown in solid lines. The global fit results in a K_d value of (12.9 ± 322.9) nM.

Plotting k_{obs} against the respective EsxA concentration shows an increase of the rate constant with increasing concentration of binding partner. A one-step mechanism of association/dissociation would result in a linear plot of k_{obs} as a function of the binding partner's concentration [163]. Assuming this type of binding mechanism, the rates of association (k_{on}) and dissociation (k_{off}) can be determined with a line-fit of the k_{obs} data using

$$k_{\text{obs}} = k_{\text{on}} \cdot [\text{EsxA}] + k_{\text{off}} \quad (4.3)$$

Performing a global fit of the data by linking the individual rate constants results in a k_{on} of $(1259.33 \pm 344.38) \text{ s}^{-1} \text{ M}^{-1}$ and a k_{off} of $(9.67 \pm 9.46) \cdot 10^{-5} \text{ s}^{-1}$.

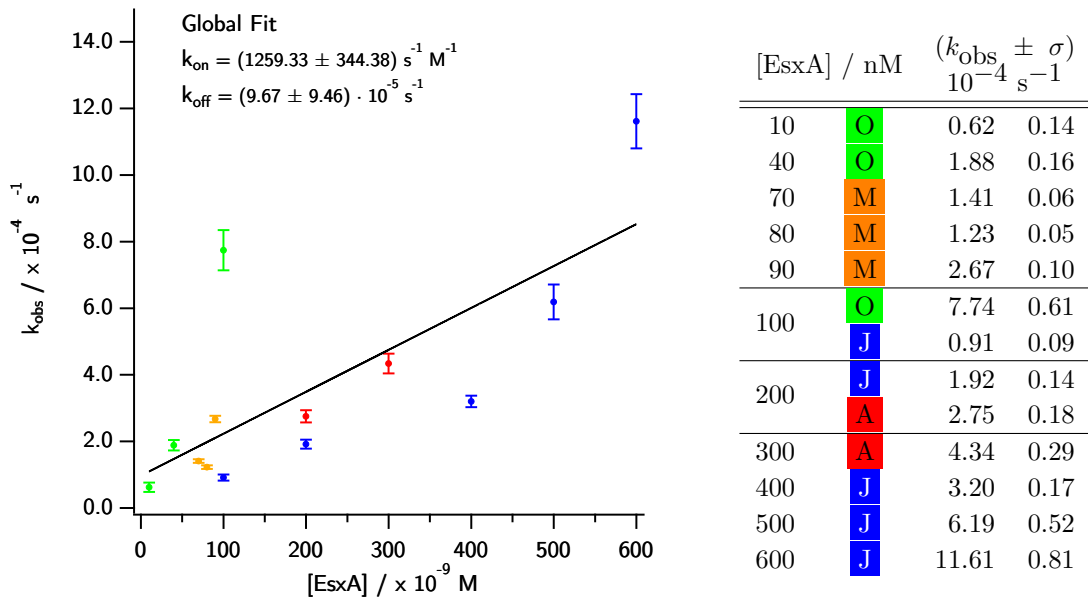


Figure 4.21: Experimentally observed pseudo-first-order rate constant as a function of the concentration of the binding partner EsxA. The values for k_{obs} determined by the pseudo-first-order fit of titration experiments are listed in the table on the right. Color- and letter-coding in the graph and the table refer to the different experiment blocks. Error bars implicate the standard deviation. The global line-fit is shown as a black solid line, the resulting values for k_{on} and k_{off} are indicated in the graph.

4.5.2 EsxT Titration to Study Inter-loci Complex Formation

In order to investigate the inter-loci dimer formation between EsxB and the EsxA-like protein EsxT from ESX-4, the FRET efficiency of double-labeled EsxB monomer was measured with an increasing amount of EsxT over the course of several hours. The results shown in Figure 4.22 show the FRET histograms of the first and sixth hour of different experiments. There is a clear increase of occurrences with high FRET efficiency with an increase in EsxT concentration and experiment time, comparable to that seen in the EsxA titration experiments (see Figure 4.17). However, it should be noted that the concentration of EsxT needed to detect a peak at a high FRET efficiency is four orders of magnitude higher than that of the natural binding partner EsxA.

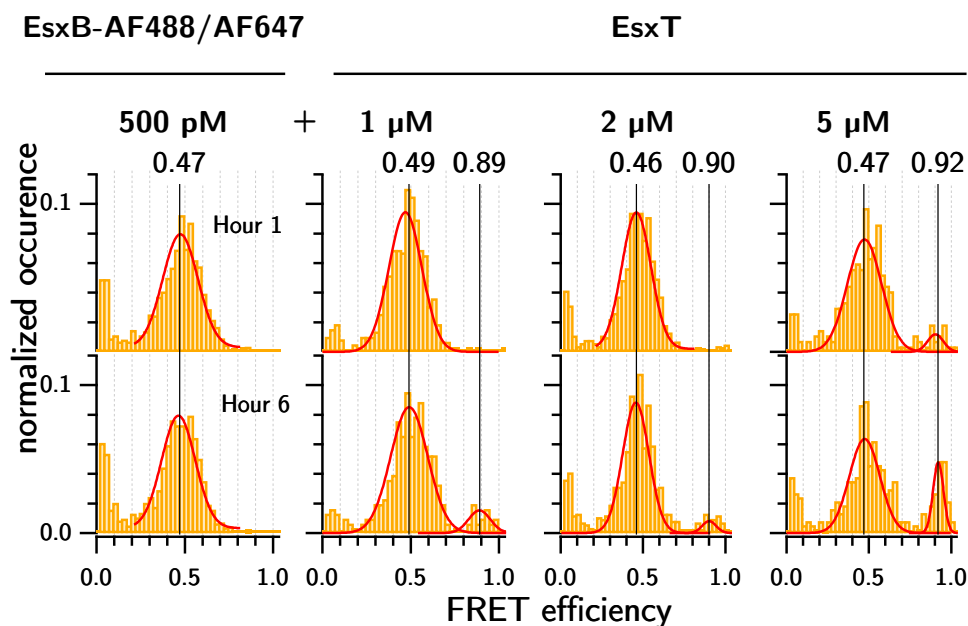


Figure 4.22: EsxT Titration. The first and sixth hour of each experiment is shown for selected titration steps to elucidate the change over time. The leftmost panel histograms show the FRET measurements of the monomer. To its right are the histograms of three titration steps, namely those with the addition of 1 μM , 2 μM , and 5 μM EsxT, respectively. The red curves are gauss-fits of the FRET histograms over the FRET efficiency range of 0.2 to 0.8 and 0.8 to 1.0, where applicable. The black vertical lines highlight the position of the first hour's maximum.

As done for the EsxA-Titration experiments (see Section 4.5.1), the multiple-hour measurements were evaluated in a time-wise fashion. This allows the graphic representation of the time-dependent behavior of the high-FRET-efficiency peak and the calculation of the observed rate (k_{obs}). Figure 4.23 shows the time-dependent progression of EsxB/T complex formation in dependence of EsxT concentration.

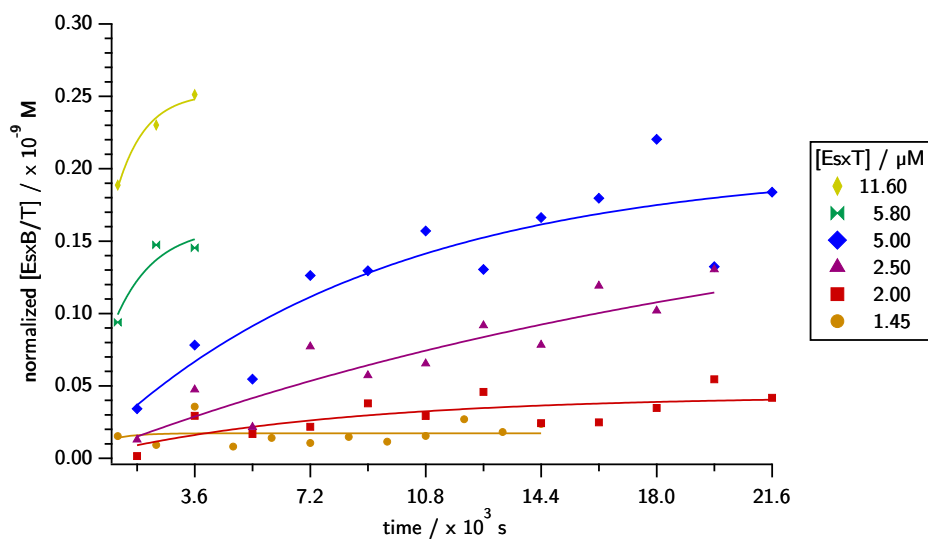


Figure 4.23: Time-dependent progression of EsxB/T complex formation in dependence of EsxT concentration. Each experiment was analyzed by dividing the data into consecutive periods of 15, 20, or 30 minutes. From the resulting FRET histograms, the ratio of complexed EsxB, meaning the ratio of the peak between 0.8 and 1.0 FRET efficiency, was calculated. Data points are fit with a “pseudo-first-order” function. The curve’s respective EsxT concentration is indicated on the right.

4.5.3 Combining Crosslinking with smFRET

The idea behind combining chemical crosslinking with FRET measurements is that some protein-protein interactions might only be temporary or so short-lived that they are not detected by smFRET experiments. Therefore, the chemical crosslinking reagents NHS and EDC were added to the protein mixture prior to FRET measurements (see section 3.4.4) in order to “freeze” the interacting proteins by covalently linking them. For this experiment, photons of the double-labeled EsxB monomer were recorded only in buffer or under crosslinking conditions for four hours each, once with and once without the addition of a potential binding partner. As binding partners, the native partner EsxA and EsxT from ESX-4 were tested.

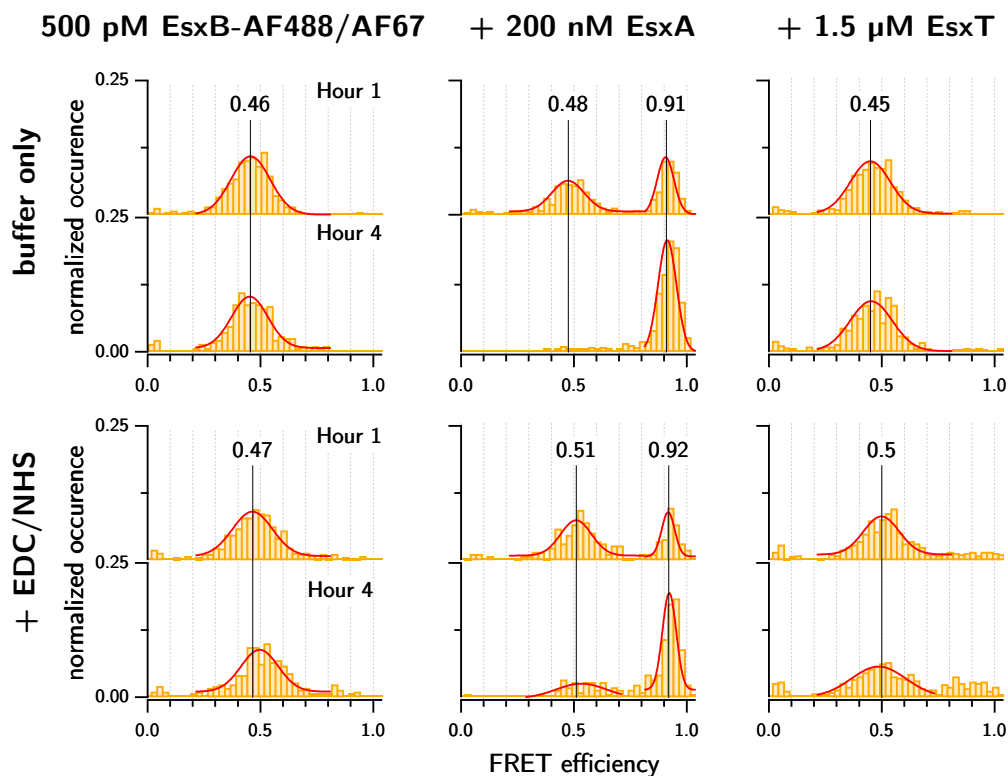


Figure 4.24: Combining smFRET and chemical crosslinking with EDC and NHS to detect protein-protein interaction. The first and fourth hour of each experiment is shown for each condition to elucidate the change over time. The top panel shows the control measurements performed with ordinary buffer, the bottom panel shows the results under crosslinking conditions (with addition of EDC and NHS). The left column shows the measurements of the EsxB monomer, the center column shows the results after the addition of 200 nM EsxA, and the right column shows the results after the addition of 1.5 μ M EsxT. The red curves are gauss-fits of the FRET histograms over the FRET efficiency range of 0.2 to 0.8 and 0.8 to 1.0, where applicable. The black vertical lines highlight the position of the first hour's maximum.

As Figure 4.24 shows, the addition of NHS and EDC to the EsxB monomer results in a slight shift of the center peak's maximum from 0.47 to around 0.5 after four hours (bottom left), which can not be seen in the control condition (top left). This behavior is rather comparable to the peak shift observed after the addition of EsxA (see Section 4.5.1 and Figures 4.17 and 4.18). An increase in occurrences of FRET efficiencies over 0.8 after four hours under crosslinking conditions is visible but nominal. When adding the native binding partner EsxA, virtually all detected occurrences in the fourth hour of the experiment are attributed to FRET efficiencies around 0.9 (center top). Interestingly, if NHS and EDC are added to the mixture prior to measurement, the trend is the same; however, a broad peak with a right-shifted center remains after four hours at FRET efficiencies of around 0.5. A slight but clear increase in occurrences at higher FRET efficiencies (0.8 - 1.0) can be seen when mixing EsxB with EsxT and crosslinking reagents in contrast to the buffer-only condition (right column). This suggests that combining chemical crosslinking and FRET could indeed be a valid method to detect transient protein-protein interactions. However, the possibility of

the cross-linked monomer having an increased affinity or binding capability with non-native binding partners needs to be considered.

4.6 Effect of pH on the EsxB/A Complex and its Monomers

Since the host defense mechanisms include acidifying vacuoles entrapping the pathogen, it was of great interest to analyze structural changes of the Esx proteins in the face of lower pH values. First, FRET measurements of the double-labeled monomer EsxB-AF488-AF647 were conducted under different pH conditions using sodium-phosphate-citrate buffers.

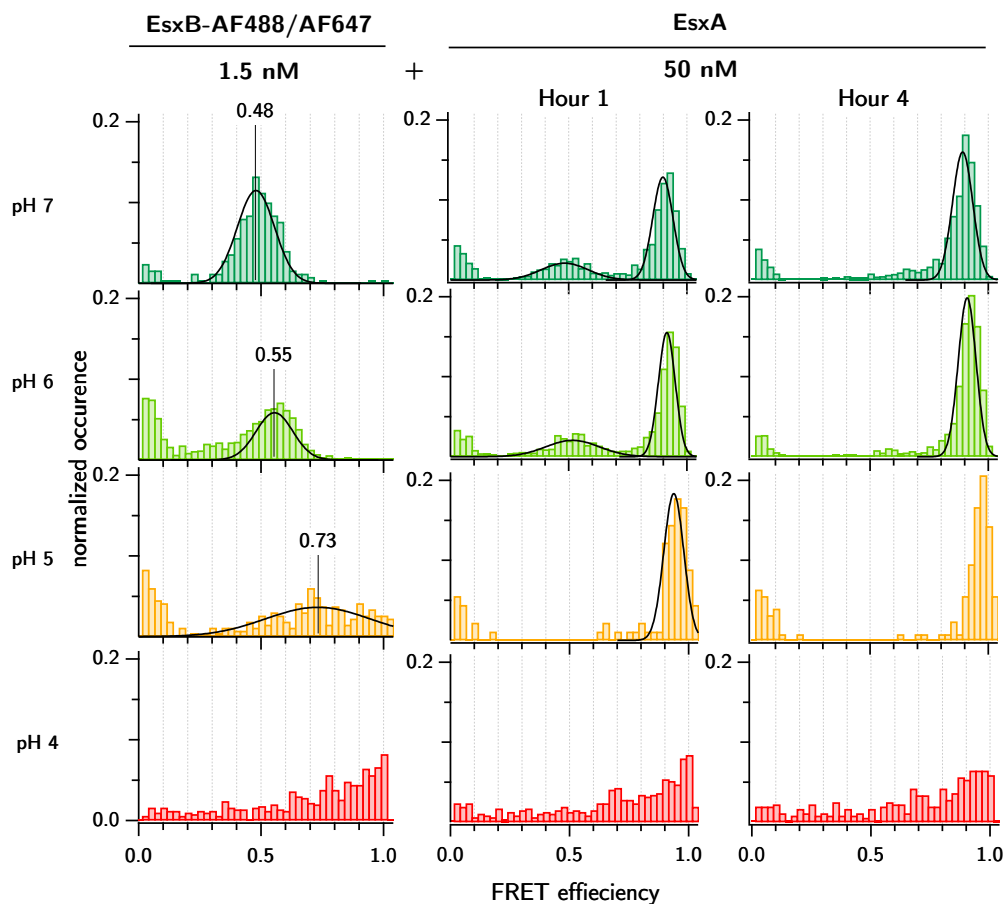


Figure 4.25: FRET histograms of the EsxB-AF64-AF488 monomer and with the addition of EsxA under different pH conditions. Left: FRET histograms showing the double-labeled EsxB-monomer in phosphate citrate buffers of different pH values. Each is the result of a one-hour measurement. For pH 7, 6 and 5 the data was fit with a Gauss-function (black line) and the resulting center of each peak is indicated above. Right: FRET histograms after the addition of 50 nM EsxA. The FRET histograms of the first and fourth hour of each experiment are shown for each indicated pH value. Gauss-fit of data is shown as a black line.

Interestingly, decreasing pH values showed a gradual shift towards higher FRET efficiencies for the monomer, as can be seen in Figure 4.25 on the left. Comparing the EsxB-monomer measurements and the mixture of EsxB and EsxA titration at the different pH values in Figure 4.25 suggests that complex formation between EsxB and EsxA does happen in a pH

range of 5 to 7. Even though the FRET histogram of the EsxB monomer is shifted towards higher FRET efficiencies at pH 4 (Figure 4.25 bottom left), a distinct peak can not be detected after the addition of EsxA.

The kinetics of complex formation were analyzed, as explained in the previous chapter, by plotting the occurrences of high FRET efficiencies (0.8-1.0) against the passed time in seconds and fitting the data with a pseudo-first-order function. This is presented in Figure 4.26. While the curves for pH 6 and 7 closely resemble each other and the data's fit reveals a similar value for k_{obs} ($(7.91 \pm 0.68) \cdot 10^{-4} \text{ s}^{-1}$ for pH 6 and $(8.67 \pm 0.70) \cdot 10^{-4} \text{ s}^{-1}$ for pH 7), the curve for pH 5 is visibly steeper, with the reaction rate for complex formation being around one order of magnitude larger than for the other two ($(27.39 \pm 2.09) \cdot 10^{-4} \text{ s}^{-1}$).

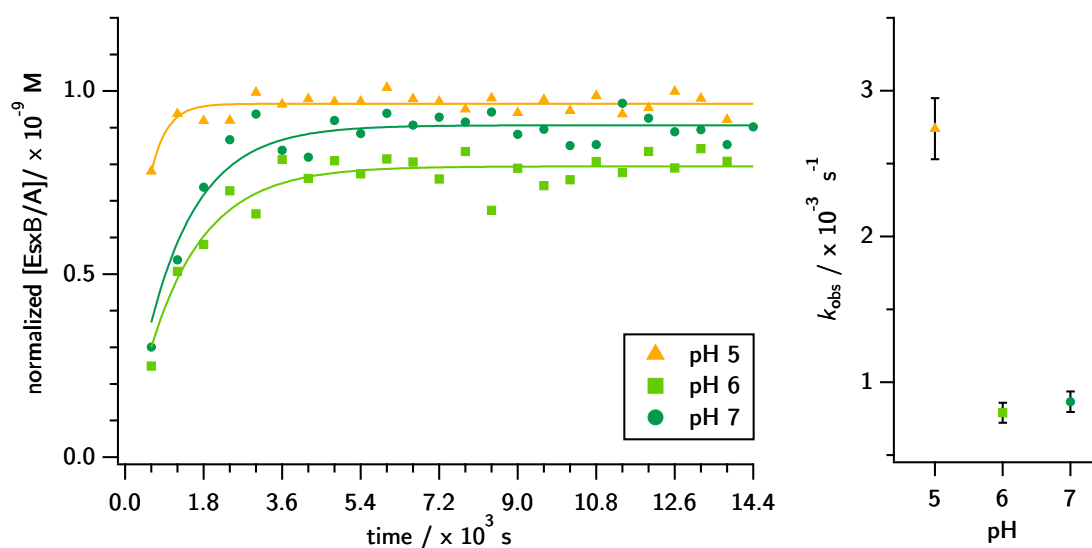


Figure 4.26: Time-dependent progression of EsxB/A complex formation in dependence of the pH value. Na-phosphate-citrate buffers of different pH values were used to measure the complex formation of EsxB/A. The respective pH values are indicated by colors and markers. Left: The resulting data was divided into consecutive periods of 10 minutes. From the resulting FRET histograms, the ratio of complexed EsxB, meaning the ratio of the peak between 0.8 and 1.0 FRET efficiency, was calculated. Right: The resulting values for k_{obs} from the pseudo-first-order fit are shown for each pH value.

Next, the structural changes of the EsxB/A complex and its monomers under different pH conditions were analyzed using CD spectroscopy. As represented in Figure 4.27, the EsxB/A complex was measured at pH values from 4 to 7, while the measurements of the individual monomers were performed at pH 4 and pH 7.

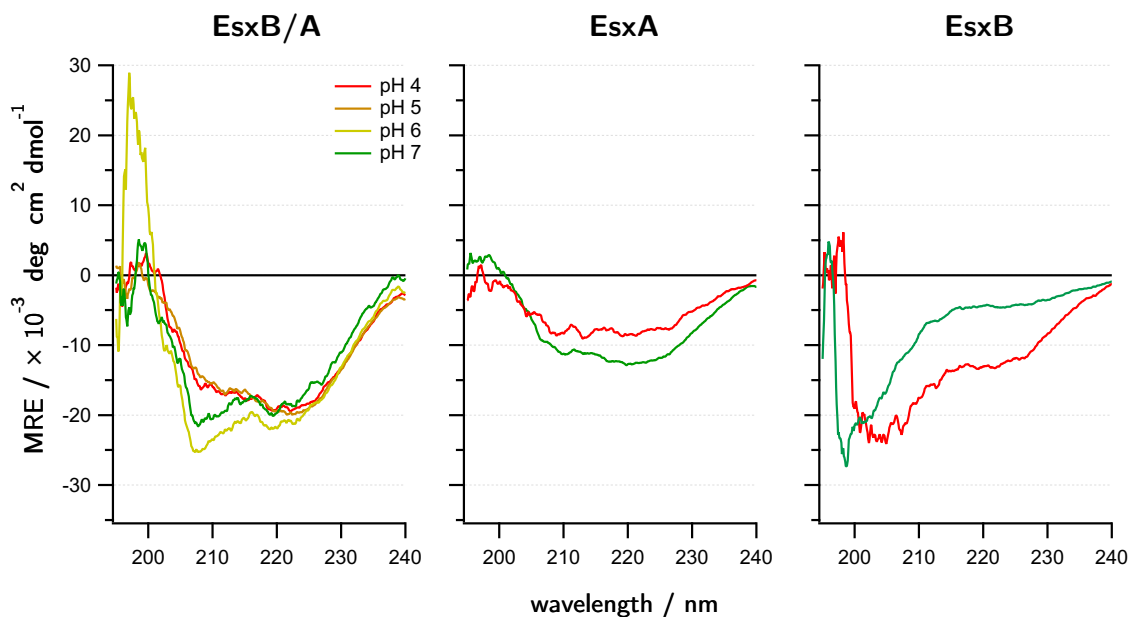


Figure 4.27: CD spectroscopy of EsxB/A and its monomers in different pH conditions. The results of CD spectroscopy of the complex EsxB/A in buffers of different pH values are shown on the left, and the results of the monomers EsxA and EsxB are shown in the center and on the right, respectively. The pH values for each experiment are indicated by color.

While the spectrum shape of the EsxB/A complex at pH 4 and pH 5 and at pH 6 and pH 7 are each congruent beyond a wavelength of 200 nm, the two groups clearly differ. Regarding the two minima at 209 nm and 222 nm (typical for helical structures) the minimum at 209 nm is much less prominent for the lower pH values. Regarding the measurements of the monomers, the spectra of EsxB show the most striking difference between pH 7 and pH 4.

4.7 Phylogenetic Assessment of EsxU and EsxF

The phylogenetic relationships of the Esx proteins from *M. tuberculosis* as published in 2001 by Pittius et al. [50] are shown in Figure 4.28 **A**. The pair EsxF/E of *M. tuberculosis* is phylogenetically the oldest Esx protein pair after EsxT/U from ESX4, which is the most ancient one. Figure 4.28 **B** shows the relationship between taxa of high- and low- G+C content Gram-positive bacteria where WXG100 proteins have been detected and verified. To evaluate the phylogenetic relationship of homodimer- and heterodimer-forming WXG100 proteins, sequence alignment of EsxU and EsxF from *M. tuberculosis* with a number of EsxB- and EsxA-like proteins known to form either homo- or heterodimers was performed with Clustal O [165] (see Figure A.2 **A**). The five most closely related EsxB-like proteins from *M. tuberculosis* (*mtbEsxU*, *mtbEsxF*, *mtbEsxB*, *mtbEsxG* and *mtbEsxS*, see Figure 4.28 **A**) along with the EsxB-like *codEsxB* from *Corynebacterium diphtheriae* make up the group of proteins known to form heterodimers with EsxA-like partners. The group of homodimer-forming proteins includes the EsxB-like protein *banEsxB* from *Bacillus anthracis* and the EsxA-like proteins *sagEsxA* from *Streptococcus agalactiae*, *sauEsxA* from *Staphylococcus aureus*, *limEsxA* from *Listeria monocytogenes*, *basYukE* from *Bacillus subtilis* and *getEsxA* from

Geobacillus thermodenitrificans. Sequence alignment of the “homodimer” group shows numerous highly conserved residues and besides the WYG-motif, the **I^K/R^HS/TPEELR**-motif² at the N-terminus is also highly conserved (see Figure A.2 **A**, bottom half). In contrast to this, the group of “heterodimers” shows less conserved features, with most of them being present in the other group as well (see Figure A.2 **A**, top half). The neighbor-joining phylogenetic tree of listed proteins can be seen in Figure 4.28 **C**, which also shows the result of pairwise alignment of heterodimers and homodimers, respectively, in a pairwise identity matrix (PIM) (the complete PIM showing the sequence identity between homo- and heterodimers is shown in Figure A.2 **B**). The greatest sequence identity is between *mtbEsxU* and *codEsxB* with 27.45 % and between *mtbEsxF* and *mtbEsxB* with 22.77 %. The sequence identity between *mtbEsxF* and *mtbEsxU* is the next highest value of 18.45 %. These values are in agreement with the phylogenetic tree in Figure 4.28 **A**.

²capital or lower case letter: strong or less strong conservation, bold letter: same residue in all compared sequences, fraction: either/or is conserved, “H”: hydrophobic residue

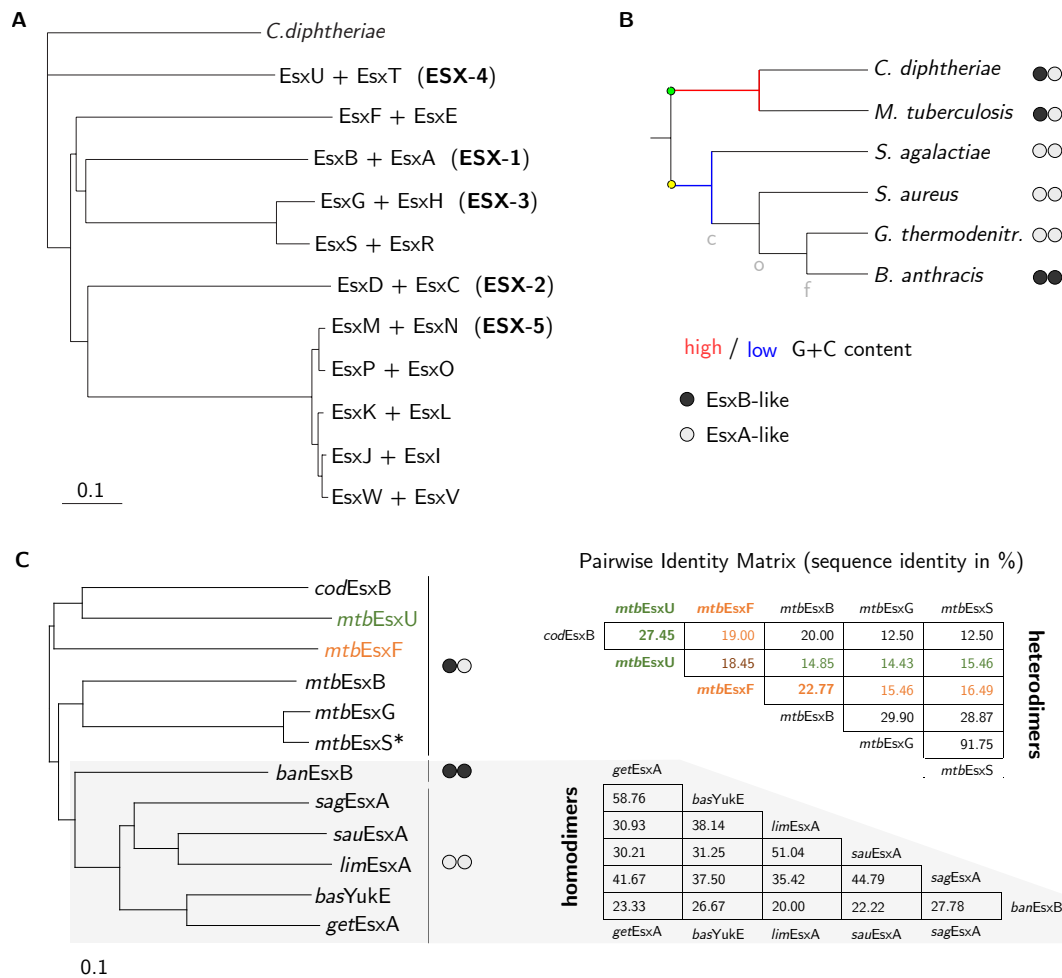


Figure 4.28: Phylogenetic trees showing the relationship between Esx proteins from *M. tuberculosis* and between selected Gram-positive bacteria. **A** Neighbor-joining phylogenetic tree using the EsxB/A-like protein orthologs of *Corynebacterium diphtheriae* as the outgroup. Reproduced after [50]. **B** Phylogenetic relationships of Gram-positive bacteria with a high (red) or low (blue) genomic G+C content of the phyla Actinomycetota (green node) and Bacillota (yellow node). Esx-dimers consisting of EsxB-like (black circle) and/or EsxA-like (grey circle) have been detected and structurally verified for all listed species as shown next to each name. Grey letters on tree are annotated for readers guidance and specify the position of nodes: *c* class, *o* order, *f* family. Tree created with iTOL [164]. **C** Neighbor-joining phylogenetic tree showing the relationship between selected heterodimer-forming EsxB-like proteins from *M. tuberculosis* (*mtb*) and *C. diphtheriae* (*cod*) with homodimer-forming Wxg100 proteins of other species (abbreviated as *ban*: *Bacillus anthracis*, *sag*: *Streptococcus agalactiae*, *sau*: *Staphylococcus aureus*, *lim*: *Listeria monocytogenes*, *bas*: *Bacillus subtilis*, *get*: *Geobacillus thermodenitrificans*. *EsxS forms heterodimers and -tetramers. The Pairwise Identity Matrix on the right shows the sequence identity of the heterodimers and homodimers in %.

Chapter 5

Discussion

This chapter aims to summarize the findings of this work and discuss them in light of the most recent research. The methodology of smFRET for kinetic studies of EsxB complex formation will be discussed first, followed by the results of the characterization of the orphan protein pair EsxF/E and the protein pair EsxU/T from ESX-4. In the latter part, special attention will be given to the implications of the homodimer formation of said protein pairs. This will be based on the possible interpretations of the smFRET results and the phylogenetic context of the proteins, compared to homo-dimer forming members of WXG100 from other bacteria.

5.1 Kinetic Binding Studies with smFRET

The smFRET method was used to establish a protocol for binding studies and the evaluation of kinetics. In order to evaluate binding affinities from titration experiments, it is important to consider a few experimental aspects. According to Jarmoskaite et al. [166], it is essential to adjust the incubation time in order to reach an equilibrium and to adjust the concentration of the constant complex partner in order to determine a reliable K_d value. If incubation times are too short and an equilibrium has not yet been reached, the calculated K_d most likely would underestimate the true binding strength. The time until equilibrium is reached greatly depends on the binding affinity. The lower k_{off} , the longer the incubation time needs to be to reach equilibrium. Further, if the concentration of the constant component is not well below the dissociation constant ($[\text{EsxB}] \ll K_d$), the resulting K_d value can differ from the true value by orders of magnitude. With the appropriate quadratic binding equation, the case where the trace component (EsxB) is similar to or in modest excess over K_d can also deliver reliable results. As the K_d value was not known prior to the experiments, and the concentration of the double-labeled EsxB was more or less dictated by the labeling efficiency and the experimental setup (adjusting the concentration to the demands of a proper fluorescence signal), there was little room for variance in this aspect.

5.1.1 EsxA Titration

One interesting finding was the apparent shift of the EsxB-“monomer peak” after the addition of the native binding partner EsxA towards a slightly higher FRET efficiency, especially at time points beyond two hours of incubation (shown in Figures 4.17 and 4.18). Even though the difference accounts for only about 2 Å, it was statistically significant. The fact that this shift was only observed in the presence of EsxA creates the impression that EsxB and

EsxB somehow interact before forming a complex. With regards to the two most commonly used models for protein-ligand binding, “conformational selection” and “induced fit”, neither can exclusively explain these observations. However, it is likely that real coupled folding and binding events occur in a combination of the two. Presumably, the largely unstructured EsxB undergoes small conformational changes upon encountering EsxA before forming a structured complex. A similar mechanism was shown for an intrinsically disordered domain of the transcription factor CREB (named pKID) for its association with the KIX domain of the CREB binding protein [167, 168]. Here, it was proposed that pKID forms a weak encounter complex with KIX. This encounter complex ensemble is stabilized by hydrophobic contacts and remains largely unstructured. The final complex with the helically structured pKID forms by evolving from a partly folded intermediate.

For kinetic studies, the increase of the structured EsxB/A-complex was observed over several hours upon the addition of different EsxA concentrations. In order to determine the fraction of bound EsxB at equilibrium, the data points of the time-dependent progression of bound EsxB (see Figure 4.19) are fit using the equation for pseudo-first-order kinetics (Equation 2.7). If done so with data points that do not reach a plateau, the determined values have large errors. Such values should not be used for further analysis. However, especially at low protein concentrations, an equilibrium was often not reached during the time of measurement. The fit for the eight-hour measurement after addition of 10 nM EsxA (see Figure 4.19) predicts an equilibrium at around 0.55 nM bound EsxB with a rate constant of approximately $0.6 \times 10^{-4} \text{ s}^{-1}$. If this rate constant were true, the incubation time to reach equilibrium would have been over 16 hours, given that $k_{\text{equil}} = k_{\text{obs}}$ and $t_{\text{equil}} = 5 \cdot t_{1/2} = 5 \cdot \frac{0.693}{k_{\text{equil}}}$. Concentrations lower than 10 nM of the binding partner EsxA would have needed even more time to reach equilibrium. However, measurements exceeding nine hours posed a problem, as the droplet of immersion water tends to evaporate during such long measurements.

Once the curve’s saturation was determined, the values were plotted against the concentration of EsxA (see Figure 4.20) in order to determine the binding constant. It was commonly observed that different FRET measurements with supposedly the same protein concentrations yielded quite different FRET histograms and showed diverging time-dependent behavior. This could be attributed to the fact that measurements were performed in different blocks, with months or even years time in between. The greatly varying storage time and multiple freeze-thaw cycles of the double-labeled EsxB surely influenced the condition of the protein. This posed a problem for the reproducibility and subsequent evaluation. Especially the data recorded in October 2019 (see “O” (green) in Figure 4.21) stands out from the other data points. Taking these points into the data set for global fitting evokes a very large error for the calculated K_d ($(12.9 \pm 322.9) \text{ nM}$). Excluding these points for evaluation seems advisable, however they are the only applicable data points for the lower concentration range of EsxA (10 nM - 70 nM). Evaluation including only the other data sets with a global fit gives a more exact K_d value of $(12.7 \pm 26.2) \text{ nM}$. Even though this result still has a relatively large standard deviation, the computed value falls well in the range of the reported K_d of between “below 11 nM” [169] and 50 nM for EsxB/EsxA [170].

A linear fit of the pseudo-first-order rates k_{obs} against the concentration of EsxA (see Figure 4.21) suggests a dissociation rate characteristic for a strong binding affinity. As mentioned above, the data collected over the course of many months has led to a bad comparability between the data sets and relatively high deviations. Nevertheless, a global

fit of all points yields reasonable rates of association and dissociation. With values of $k_{\text{on}} = (1259.33 \pm 344.38) \text{ s}^{-1} \text{ M}^{-1}$ and $k_{\text{off}} = (9.67 \pm 9.46) \times 10^{-4} \text{ s}^{-1}$, the dissociation constant K_d lies in the range of 2 nM to 120 nM ($K_d = (6.1 \pm 5.9) \times 10^{-8} \text{ M}^{-1}$). Considering the given standard deviations of the rates, K_d lies in the same concentration range as the values found in literature and determined using the graph in Figure 4.19 in this work. Previously, it was determined by de Jonge et al. [171] with surface plasmon resonance assays that at pH 6.5, the dissociation rate k_{off} of EsxB/A was $9 \times 10^{-4} \text{ s}^{-1}$, which deviates approximately one order of magnitude from the value determined in this work. It should be mentioned here that the method used in this work is based on simplified models. Determination of association and dissociation rates were performed under the assumption of a simplified one-step binding scheme “ $A + B \xrightleftharpoons[k_{\text{off}}]{k_{\text{on}}} AB$ ”, which does not account for any intermediate conformations. Realistically, binding mechanisms involve (numerous) conformational transitions and are commonly described with either the model of “conformational selection” or the mechanism of “induced fit”. According to Vogt et al. [172], measuring the binding dynamics at equilibrium does not allow us to distinguish between the two models.

Overall, the titration experiments performed using smFRET are suited for binding and kinetic studies, as performed in this work. Nevertheless, several considerations should be kept in mind for similar future studies with the given setup. The following points shall summarize the problems faced in this work during titration experiments with smFRET and suggest solutions:

- *Saturation is not reached:* Measurement times must be adjusted to equilibration times, or additional incubation must precede the measurements,
- *Variation between measurements:*
 - Experiments shall be recorded with the same protein batches within a limited time period of one or two months,
 - If different protein batches of the double-labeled binding partner are used, data sets need to be normalized according to the concentration determined by fluorescence correlation,
 - Titration steps with low concentrations of binding partner shall be repeated several times to minimize the errors introduced by preparation of low-volume samples of low concentrations,
 - The temperature during measurements shall be kept constant and be recorded.

pH-dependent Binding Studies

As pH changes belong to the key factors during phagosome maturation, smFRET measurements and CD spectroscopy were additionally performed at varying pH values. The early phagosome is characterized by a mild acidic environment with a pH of around 6.3, which then decreases to around 5.5 in a late-stage phagosome. In the final state, the phagolysosome can have pH values of as low as 4.5 [173]. It was reported that at pH 4, the EsxB/A complex dissociates, allowing the monomers to interact with membranes [171]. Lightbody et al. have performed chemical denaturation studies with EsxB/A and EsxH/G at different pH values. They showed that the resistance of both complexes to guanidine hydrochloride

remains constant within a pH range of 4.5 to 6.5, with EsxB/A even showing increased stability at pH 4.5 [174]. Thus, they concluded that the functional form of the proteins inside the phagosome is most likely the complex rather than the individual monomers. In conclusion, binding studies at low pH values have resulted in seemingly contradicting conclusions. In this work, the results of the FRET experiments suggest that complex formation between EsxB and EsxA occurs in a pH range between 5 and 7 (see Figure 4.25). The association rate is seemingly even higher at pH 5, compared to k_{obs} at pH 6 and 7 (see Figure 4.26). However, no complex formation was detected with this method at pH 4. Protein stability is influenced by the environmental ionic strength, and so is protein folding [175]. The different buffers vary considerably in their compositions of citric acid and sodium phosphate (see Table 3.5). It should be pointed out that FRET measurements performed in this work observed the complex formation of monomers in a solution of certain pH values (= association), while the two other studies mentioned before monitored a possible dissociation of the complex in the face of varying pH values. Looking at the results of CD spectroscopy, the spectra of EsxB/A at pH 5 and pH 4 are effectively the same (see Figure 4.27). Even though the minimum at 209 nM is less distinct in this pH range than for pH 6 and pH 7, the spectra resemble that of a mostly helical protein. Combining the results from FRET and CD measurements, it can be argued that while the complex formation of the individual monomers is possible only in a pH range of 7-5 and not at pH 4, the bound complex does not necessarily dissociate at pH 4. This suggests that the sodium-phosphate buffer with pH 4 used in these experiments seems to evoke conformational changes in the monomers, which prevent complex formation.

5.1.2 EsxT Titration

After the establishment of a titration protocol for kinetic studies, it was used to characterize the inter-loci complex formation between EsxB and EsxT, which was suggested by preceding chemical cross-linking studies (not shown). An apparent complex formation between EsxB and the EsxA-like protein from ESX-4 could be replicated with smFRET measurements (see Figure 4.22). However, reliable data analysis for computing a K_d or association and dissociation rates could not be performed. Fairly high EsxT concentrations were needed to evoke complex formation, and even after six hours of measurement, an equilibrium was often not reached (see Figure 4.23). Complex formation was observed only after the addition of EsxT in 1000-fold excess over EsxB. This raises the question of how realistic a complex formation under such conditions is *in vivo*.

As the binding equilibrium is strongly dependent on the binding affinity (K_d), detecting low-affine interactions requires a very high concentration of binding partners. This poses a problem, as the smFRET measurements are performed at very low protein concentrations. Therefore, the method of chemical crosslinking with EDC and sNHS was combined with smFRET (“Cross-FRET”) as an attempt to get around these restrictions and to uncover potentially short-lived protein-protein interactions. As presented in Figure 4.24, it was possible to visualize inter-loci complex formation with Cross-FRET after four hours (see bottom right), which would otherwise have gone undetected under native conditions (see top right). Interestingly, Cross-FRET measurements of only the double-labeled EsxB monomer revealed a slight shift of the peak’s maximum from 0.47 to around 0.50 (see Figure 4.24 bottom left), which is comparable to the “monomer-peak” shift observed after addition of EsxA, discussed

earlier. Further, a broad, right-shifted center peak remained four hours after the addition of 200 nM EsxA and crosslinking agents. In comparison to the complete disappearance of the center peak in the buffer measurement, this might indicate the presence of cross-linked EsxB monomers, which might no longer be able to interact with and/or bind to EsxA. If a portion of EsxB was intramolecularly linked in such a way that it is inhibited from protein-protein interaction – even with its native binding partner – this decreased pool of “native” EsxB would need to be considered and quantified for a correct interpretation of future results.

5.2 Characterization of EsxF/E and EsxU/T

The crystal structures of the “orphan” EsxF/E and the ESX-4-associated EsxU/T from *M. tuberculosis* (*mtb*) are yet to be determined, as well as the proteins’ concrete functions. Besides presenting in the form of a secreted, water-soluble heterodimeric complex, recent studies have suggested that EsxF and EsxE play an important role in mediating toxin secretion by interacting with membranes and forming hetero-oligomeric transmembrane pores [176]. The homologs from *Mycobacterium abscessus* (*mab*) were shown to form a heterodimeric complex in the typical four-helix bundle conformation (PDB: 4I0X) [7].

The ESX-4 cluster has been studied in different mycobacterial species, including *M. smegmatis* (*mse*), *M. abscessus* (*mab*) and *M. marinum* (*mma*). *mse*ESX-4 was shown to play an essential role in the conjugal DNA transfer [177], whereas *mab*ESX-4 is required for the survival of the bacteria inside amoebae and macrophages [178]. While *mab*EsxU/T could be detected as a secreted heterodimer in the culture supernatant [179], the *mma*EsxU/T complex is found in the cell lysate rather than in the supernatant [180]. Likewise, it has not been proven that *mse*EsxU/T and *mtb*EsxU/T are actually secreted into the environment by ESX-4 or stay associated with the bacterial cell wall [181]. It was stated in previous studies that *mab*EsxU/T is capable of interacting with artificial membranes and forming higher-order oligomeric structures in lipid bilayers, which is thought to follow a similar mechanism as the membrane-disruption of *mtb*EsxB/A [182]. A number of GxxxG and AxxxA motifs are present in *mtb*EsxF, *mtb*EsxU and *mtb*EsxT (see Figures A.2 **A** and A.3) as well as in *mtb*EsxE [176]. These motifs are common α -helical interaction motifs in membrane proteins, which further supports the theory of possible membrane interaction of these Esx proteins [183, 184].

In this work, *mtb*EsxU/T produced in *M. smegmatis* was found to be predominantly cell wall-associated. Only under slightly denaturing lysis conditions was it possible to get sufficient soluble EsxU/T complex (see Figure 4.4). However, the obtained protein complex was prone to precipitation upon storage at 4 °C and/or multiple freeze-thaw cycles (data not shown). The failure of *M. tuberculosis*, *M. smegmatis* and *M. marinum* to secrete EsxU/T was previously assigned to the missing C-terminal secretory signal motif YxxxD/E [181], which is commonly found in ESX-substrates and is present in over 80 % of *mtb*EsxB-like proteins. Together with the WXG-motif of the EsxA-like partner, the complex is thought to form a two-part type 7 recognition signal needed for secretion [69]. Alignment with Esx-homologues from other species has revealed the C-terminal consensus sequence HxxxD/ExxhxxxH, where *H* and *h* stand for highly and less conserved hydrophobic residues, respectively, and “D/E” for either aspartic or glutamic acid. This is found both in the bi-cistronically expressed EsxB-like and EsxA-like proteins associated with the Type 7 secretion system (T7SS) as well as in

the mono-cistronically expressed *sagEsxA*-like proteins associated with the secretion system of type 7b (T7bSS) [5]. Alignment of the four different EsxT and EsxU proteins (*mtb*, *mse*, *mab*, *mma*) shows the presence of the secretion signal in all EsxT homologs, while the motif is incomplete and partly shifted in the EsxU homologs. On the other hand, the EsxU homologs show an intact WxG motif, which is altered in the EsxT proteins (see Figure A.3).

Naturally occurring homodimers of mycobacterial WXG100 proteins have so far not been reported. There have been implications of homodimer formation of *mtbEsxB* and *mtbEsxA* using the method of the yeast two-hybrid system [185], but such homodimers have not been detected in culture supernatant or under soluble native conditions *in vitro* so far. However, SDS-resistant, high MW oligomers were previously observed during the purification process of *mtbEsxF/E*, which showed to increase after storage at 4 °C for two days [176]. These high MW oligomers are thought to be long filaments, which could be visualized by electron microscopy. Yet, to the best of my knowledge, it was not verified that these constructs actually consist of both monomers EsxF and EsxE, meaning it was not shown if they were *hetero*- or *homo*-oligomers.

As of yet, homodimers have been exclusively detected in other related Gram-positive organisms. Protein structures of homodimers verified by X-ray diffraction include *sagEsxA* (PDB: 3GWK) from *Streptococcus agalactiae* [5], *sauEsxA* (PDB: 2VRZ) from *Staphylococcus aureus* [9], *banEsxB* (PDB: 4J11) from *Bacillus anthracis* [186] and *getEsxA* (PDB: 3ZBH) from *Geobacillus thermodenitrificans* [187]. For other WXG100 proteins, homodimer-formation is suggested by empirical data, such as for *sauEsxB* from *S. aureus* [9] and YukE from *Bacillus subtilis* [188].

In Chapter 4 of this work, the biophysical and thermodynamic characterization of the two Esx-pairs EsxF/E and EsxU/T were presented. Further, several indications for homodimer-formations of EsxU and EsxF have been observed. The results from purification by SEC, chemical crosslinking studies, CD spectroscopy, and FRET experiments will be discussed in more detail and summarized in the following sections.

5.2.1 Biophysical and Structural Characterization of EsxF/E and EsxU/T

Besides characterizing the secondary structure of the complexes and the monomers, special focus was also directed on thermodynamic characterization. Both were analyzed with CD spectroscopy. Two commonly used programs were used for the evaluation of the CD spectra: DichroWeb and BeStSel web services. While the α -helical content determined with either method for spectral deconvolution largely agreed (deviations ranging mostly from <1 % to 4 %), the estimation of other structural components showed large differences. Since the secondary structure of WXG100 proteins is presumed to be mostly helical, other structural components are disregarded for data analysis. Interpretation of CD spectra with either of the two methods relies on reference data sets of proteins with known structures. One big issue with this is that the interpretation of the secondary structure of proteins in solution is based on conformational characteristics identified in crystal structures. Further, the choice of data sets is restricted by the low wavelength cutoff. In this work, reliable data could only be collected beyond 200 nm. Also, it should be mentioned at this point that PBS is not the ideal buffer for CD spectroscopy, as it has a lower wavelength limit of 200 nm because of the relatively high chloride content. Chloride ions absorb strongly below 200 nm [142], which can

distort the obtained protein spectrum, especially at low protein concentrations. However, the overall lower wavelength limit of the experimental setup (comprising the influences of the spectropolarimeter, the cuvettes, and the proteins) could not be decreased to below 200 nm, regardless of the buffer. Therefore, any effects of PBS on the spectra exceeding the blank subtraction are neglected. Considering all aforementioned points, it should be refrained from taking the fraction values for secondary structures for face value. Yet, the obtained values can be used to compare the secondary structure and temperature stability of monomers and complexes. To give the reader an overview of structural and thermodynamic characteristics of different *mtb*Esx proteins published in scientific literature and in this work, the parameters are summarized in Table 5.1.

Table 5.1: Thermodynamic parameters and structural characteristics of different mycobacterial Esx protein pairs (top) and monomers (bottom). Thermodynamic values determined with different methods: ^a*ITC*: isothermal titration calorimetry, ^b*CD*: circular dichroism spectroscopy, ^c*DSC*: differential scanning calorimetry. ΔG and ΔS are given for 298 K (25 °C) or if marked tm for T_m . Determination of secondary structure content from CD spectra done with ^K K2D server, ^{BD} BeStSel/Dichroweb server or ^{CP} CDPPro software. *n.d.*: no data available.

	EsxB/A	EsxF/E	EsxG/H	EsxK/L	EsxU/T	EsxV/W
$T_m / ^\circ\text{C}$	53.4 ^c	65.1 ^b	70.0 ^{b,1}	50.5 ^c	54.1 ^b	48.0 ^b
K_d / nM^{-1}	50 ^a	<i>n.d.</i>	11 ^a	97 ^a	<i>n.d.</i>	480 ^a
$\Delta H / \text{kcal mol}^{-1}$	-40.3 ^a	-38.3 ^b	-40.8 ^a	-36.6 ^a	-28.0 ^b	-33.5 ^a
$\Delta G / \text{kcal mol}^{-1}$	-9.95	0 tm	-10.81	-9.56	0 tm	-8.55
$\Delta S / \text{cal mol}^{-1} \text{K}^{-1}$	-101.00	-113.20 tm	-100.00	-154.91	-85.57 tm	-83.70
content α -helix	64 % ^K	51 % ^{BD}	60 % ^{CP,1}	<i>n.d.</i>	41 % ^{BD}	70 % ^K
Ref.	[170]	Fig 4.8, Tab 4.2	[174] ¹ , [189]	[190]	Fig 4.13, Tab 4.3	[189]

	$T_m / ^\circ\text{C}$	CD spectrum	Ref.		$T_m / ^\circ\text{C}$	CD spectrum	Ref.
EsxA	33.0 ^{b,1}	49 % α -helix ^{K,1}	[170] ¹ ,	EsxK	<i>n.d.</i>	random coil	[190]
EsxB	<i>n.d.</i>	random coil	[169]	EsxL	<i>n.d.</i>	random coil	
EsxE	53.6 ^b	40 % α -helix ^{BD}	Figs 4.9-	EsxU	70.6 ^b	31 % α -helix ^{BD}	Figs 4.14-
EsxF	44.9 ^b	30 % α -helix ^{BD}	4.11	EsxT	46.1 ^b	45 % α -helix ^{BD}	4.16
EsxG	<i>n.d.</i>	random coil	[174]	EsxV	<i>n.d.</i>	33 % α -helix ^K	[189]
EsxH	<i>n.d.</i>	30 % α -helix ^{CP}		EsxW	<i>n.d.</i>	34 % α -helix ^K	

Analysis of the CD spectra of EsxF/E implicates that over 50 % of the secondary structure is helical (see Figures 4.8 and 4.11). When measured individually, both monomers show a CD spectrum typical for proteins with significant helical structure (see Figures 4.9 **A** and 4.10 **A**). Spectral deconvolution suggests a slightly higher α -helical content for EsxE (~ 40 %) than for EsxF (~ 30 %) (see Figure 4.11). The CD spectra of EsxU/T and of both monomers also reveal curves typical for largely helical proteins (see Figures 4.13, 4.14 and 4.15 **A**). The analysis of CD spectra of the EsxU/T complex indicates a helical content of around 40 % and for the monomer EsxU around 30 %. The interpretations of the CD spectrum of the EsxT monomer deviated distinctly, ranging from 35 % (BeStSel) to 55 % (DichroWeb) (see Figure 4.16).

With some Esx pairs, the EsxA-like monomers are helically structured, and their EsxB-like partners show no structure (like it is the case for EsxH vs. EsxG and EsxA vs. EsxB). The monomers of other pairs are both unstructured (EsxK and EsxL) or both helical (EsxV and EsxW) (see Table 5.1). Despite presenting a CD spectrum typical for helical proteins, chemical denaturation studies indicate that EsxA and EsxH seemingly lack a stable tertiary structure under native conditions and probably resemble a molten globule state [169, 174]. Structural examinations of EsxA have revealed a small hydrophobic core within the loop region between the two helices, which is missing in EsxB. This region is thought to induce partial folding of EsxA and to stabilize the monomer [169, 191]. Many of the key residues involved in stabilizing the hydrophobic core within EsxA (L36, W43, Y51, V54 and Q55) are also present in EsxH, EsxT and EsxE, which might explain the helix-typical CD spectra of these monomers. However, the curves from thermal denaturation of the monomers EsxF, EsxE and EsxT presented in this work are classical cooperative unfolding curves (see Figures 4.9, 4.10 and 4.15 **B**) which is not the case for EsxA [170]. Further, the loss of structural features caused by increasing temperatures is visibly reversible upon cooling, which is another indication of a stable tertiary structure of the monomers. The denaturing curve of EsxU stands out, as it is rather shallow and does not show a pronounced sigmoidal form. Although the monomer shows little resistance to increasing temperature, as proteins with little structure would, the high structural recovery of almost 80 % upon cooling speaks for a stable tertiary structure. Together with an astonishingly high T_m of over 70 °C, the monomer EsxU proves to be highly thermostable and thermoresistant.

Regarding the results of thermodynamic analysis, it came as a surprise that EsxF and EsxE showed an extremely high structural recovery after thermal unfolding (>95 %), both as monomers and in complex (see Table 4.2). Interestingly, a similar high reversibility of thermal unfolding has been observed for the monomer *mtb*EsxA and the complex *mtb*EsxB/A [170], as well as for the *mtb*EsxV/W complex [189]. The homodimer *sag*EsxA showed a similar high refolding capacity after heat-denaturation with about 95 % structural recovery [5]. In comparison to a reported T_m of around 53 °C of EsxB/A, the “orphan” EsxF/E complex is significantly more thermostable, showing a T_m of around 65 °C. Only the complex EsxG/H was reported to have an even higher T_m of 70 °C [174].

The data of thermal denaturation of EsxF/E is best represented by a fit assuming two transitions (see Figure 4.8 **B**). In the denaturation process of a complex, T_{m1} would typically correspond to the dissociation of the complex into the monomers, and T_{m2} would correspond to the unfolding of the individual proteins. If this model were applied to the calculated terms of a two-step transition (2T) (see Table 4.2), this would mean that dissociation of EsxF/E would occur at a temperature of around 31 °C and the unfolding of the individual monomers would occur at around 67 °C ($T_{m1}=305$ K, $T_{m2}=340$ K). In turn, this would imply that the complex does not associate inside the human host with 37 °C. Thus, the first transition at T_{m1} probably marks conformational changes within the complex rather than its dissociation. With regards to the reported and computed values for unfolding enthalpy, it appears that the change in enthalpy upon thermal unfolding of the EsxF monomer ((31.83 ± 1.95) kcal mol⁻¹) is comparable to that of the complex EsxF/E and other complexes (see Table 5.1). Compared to this, the values for the EsxE monomer are roughly half (see Table 4.2).

The presented structural characteristics of EsxF, EsxE and EsxT could hint at a tertiary

structure in the monomeric state, possibly due to homodimer formation. Next, further evidence for this assumption will be evaluated.

5.2.2 Indications for Homodimers of EsxF and EsxU

The purification process by SEC of the complex His₆-EsxF/E and His₆-EsxU/T reproducibly resulted in the elution of two protein species: the heterodimeric complex and some protein form of apparent higher molecular weight, consisting only of the His-tagged EsxB-like partner (see Figures 4.3 and 4.4). As the calibration curve of the SEC column (Figure A.1) was performed with a set of standard globular proteins, the relationship between molecular weight and the elution volume V_e cannot be used to determine the definitive MW of a protein of unknown structure. However, comparison of V_e s allows for relative statements about the molecular weight of eluted protein species. Both heterodimeric complexes eluted earlier than expected for their calculated MW. This could be a result of flexible arms at the C-terminus (including the His₆-tag on EsxB-like proteins) and at the N-terminus of the proteins, which increases the effective hydrodynamic radius. Although larger by calculation, the His₆-EsxU/T complex eluted as a species of an apparent lower MW than that of His₆-EsxF/E, which suggests a more compact conformation of EsxU/T. Interestingly, Pandey et al. have reported the presence of only one peak in the SEC elution of EsxU/T [192]. However, the group expressed EsxT and EsxU as inclusion bodies and solubilized the proteins using 8 M urea before NiNTA purification. Reportedly, only about 20 % of proteins could be recovered during the subsequent refolding process by dialysis, while the rest succumbed to precipitation. Purification under denaturing conditions and high protein loss to denaturation might explain the apparent discrepancy between their observations and the ones presented in this work. With both “EsxB-like only species” eluting as fractions beyond the exclusion limit of the column, it cannot be conclusively determined if these are the result of homodimers or of some form of higher aggregates. When considering the results of the crosslinking experiments, it seems most probable that they are indeed homodimers, as the predominant form of cross-linked constructs fit an apparent MW of twice the EsxB-like proteins (see Figures 4.5 **D**, 4.6 and 4.7 **B**). However, the formation of higher aggregates cannot be definitely excluded, as fainter bands of apparent MW of over 30 kDa can also be seen in the respective figures. One striking difference between the apparent homodimer-formation of EsxU and EsxF in the crosslinking experiments is that EsxU-only complexes exclusively occurred if EsxT was not present (see Figure 4.7 **B**). The crosslinking of EsxU with EsxT resulted in only one complex-band, whereas the crosslinking of EsxF with EsxE showed two complex-bands. This could be an indication that the heterodimeric complex is strongly favored over the homodimer for EsxU/T. However, assuming that each partner of a pair is expressed equally, the question arises of how the EsxU-only fraction during SEC originated. One solution to this could be that EsxU (and EsxF) transition between a dimeric form and a monomeric form. It was previously shown that EsxL exhibits such a behavior. Under non-reducing conditions, the protein displays a monomer-dimer equilibrium, with around 30 % of EsxL occurring as monomers [190]. Thus, untagged EsxA-like partners are lost during the first purification step by IMAC, leading to an excess of EsxB-like monomers in the second purification step by SEC. It is also possible that the homodimer and heterodimer protein species show similar behavior in gel electrophoresis and thus appear as one band.

FRET Experiments Suggest a Highly Structured Conformation of EsxF

The initial plan was to perform titration studies with the double-labeled EsxF monomers in the same manner as it was previously done with EsxB-AF488-AF647. However, the monomer EsxF is much more (helically) structured than the EsxB monomer, leading to a FRET histogram with a broad peak at relatively high FRET efficiencies (E_T), spanning from 0.50 to 1.00 with a maximum at 0.93 (see Figure 4.12 **A**). This is not an ideal basis for binding studies, as the results are most accurate between an E_T of 0.2 and 0.8. The maximum transfer efficiency is similar to that of the EsxB-AF488-AF647/A complex ($E_T = 0.92$, see Figure 4.17), however the distance-distribution is larger. The broad peak resulting from the EsxF-Atto647N-AF488 monomer indicates a variety of V13C-T91C distances, ranging from 50 Å to 32 Å (and less). Such a heterogeneous E_T -distribution suggests a number of different protein conformations, whereas it can not be distinguished whether such different conformations result from differently structured monomers (like is the case with the EsxB-AF488-AF647 monomer) or from oligomers comprised of different “protein species”, such as complexes containing double-labeled and/or differently single-labeled proteins. Simultaneous shotgun labeling of a double-cysteine mutant protein (Cys1-Cys2) with two fluorophores (A and B) principally results in nine different “species”: unlabeled protein (0-0), three sorts of single labeled proteins for each fluorophore (A-0, 0-A, A-A, B-0, 0-B, B-B) and two double-labeled protein forms (A-B, B-A). Assuming a dimer formation, this would result in 81 possible different combinations, 50 of which would be double-labeled species (containing at least one fluorophore A and B each). A titration series with increasing concentrations of urea was performed to evaluate the stoichiometry and observe the denaturation process. Increasing denaturant induces complex dissociation and thus was used as a means to estimate the proportion of double-labeled monomers and complexes. These experiments were additionally conducted with CD spectroscopy, which yielded similar results (see Figure 4.12 **B**). The urea concentration at the midpoint of transition determined with FRET experiments was around 2.5 M and that determined with CD spectroscopy was around 2.3 M. In contrast to the thermal denaturation curves, which clearly show cooperative unfolding (see Figure 4.9 **B**), data displaying the chemical denaturation of EsxF by urea show a shallow curve and a relatively low resistance to chemical denaturation. However, chemical unfolding is reported to be much less cooperative than thermal unfolding [193], meaning this does not necessarily point to a lack of structure. Considering the relatively constant number of events in each histogram, it seems likely that the histograms show double-labeled EsxF-monomers instead of complexes. However, since the labeling efficiency was very low for both fluorophores (around 3.5 % each), most (double-) labeled monomers would interact with unlabeled EsxF if complex formation did occur. In order to limit the possible combinations of potential complexes, FRET experiments were subsequently performed after mixing separately single-labeled proteins. Using double-cysteine mutants (EsxF-VT) in the labeling process would result in 16 possible dimer-combinations, 9 of which would be double-labeled. Mixing separately labeled single-cysteine mutants (EsxF-V13C) would leave only one possible combination of a double-labeled dimer. Surprisingly, these experiments generated basically the same results: FRET histograms of double-labeled protein species showing a very low FRET efficiency (0.0 to 0.2), which corresponds to distances of 63 Å and more (see Figure 4.12 **C** and **D**). The FRET histogram of

the mixed separately labeled single-cysteine mutants EsxF-V13C is a very strong indication of the complex formation of EsxF.

Phylogenetic Context of Homo- and Heterodimers

Despite the presence of ESX clusters in other Gram-positive species, phylogenetic studies have shown that multiple copies of such loci are only found in mycobacteria [50]. Further, heterodimer formation of Esx proteins has been shown only for proteins from species of the Actinomycetota phylum, while Esx-homodimers were only accounted for in the Bacillota phylum (see Figure 4.28 **B**). The protein pair EsxU/T and the corresponding cluster ESX4 are the most ancient of their kind found in *M. tuberculosis* followed by the “orphan” pair EsxF/E [50] (see Figure 4.28 **A**). This evokes the speculation that these Esx-pairs are phylogenetically closer to homodimer-forming Esx proteins found in other species. Perhaps they have retained the ability to form homodimers besides forming heterodimers with their respective partners. However, alignment of the five oldest and most closely related EsxB-like proteins from *M. tuberculosis* (*mtbEsxU*, *mtbEsxF*, *mtbEsxB*, *mtbEsxG* and *mtbEsxS*) with the EsxB-like *codEsxB* and the homodimer-forming *banEsxB*, *sagEsxA*, *sauEsxA*, *limEsxA*, *basYukE* and *getEsxA*, revealed that, phylogenetically, *mtbEsxB*, *mtbEsxG* and *mtbEsxS* are more closely related to the homodimer-forming proteins, being grouped together in one of the two initial internal nodes of the phylogenetic tree (see Figure 4.28 **C**). On the other hand, the phylogenetic tree suggests that *mtbEsxU* and *mtbEsxF* show less difference to the group of monomer-forming proteins than to the group comprised of the other mycobacterial EsxB-like proteins. In summary, an exceptionally close phylogenetic relationship of EsxF and EsxU to homodimer-forming Esx proteins could not be shown, thus weakening the thesis of correlation between homodimer-formation and protein sequence or relatedness.

The two evaluated protein pairs EsxU/T and EsxF/E are phylogenetically the oldest Esx pairs of *M. tuberculosis*. EsxF and EsxU probably display a monomer-dimer equilibrium under native conditions and can form dimers with their respective partners, EsxE or EsxT, as well as homodimers. For the ESX4-associated EsxU/T pair, the heterodimeric form seems to be favored over the homodimer. Complexes or aggregates of EsxF with higher stoichiometry were observed in cross-linking studies too.

5.3 Conclusions and Outlook

Although WXG100 proteins of *M. tuberculosis* play an essential role in the survival of the bacterium and its pathogenesis, many of these proteins have been poorly characterized. The structural and thermodynamic analysis of EsxF/E and EsxT/U presented in this work, helps to come one step closer to a better understanding of their functions.

The smFRET method was successfully established at the institute to investigate the binding kinetics and possible dimer formation of different WXG100 proteins. Despite certain limitations, in particular the low fluorescence labeling efficiency and long measurement times, the method was able to provide valuable insights. Not only can it be used as a means to estimate the K_d value, but also the rates of association and dissociation. This was shown with the titration experiments performed with EsxB and EsxA. However, the studies with EsxT were of limited value due to the low stability of the monomers in solution. Hence, it was

not possible to stably express and label a double-cysteine mutant of EsxU, as it continuously precipitated during the labeling process. Since both experimental findings and literature references indicate that EsxU/T of the ESX-4 cluster is most probably membrane-bound, future smFRET experiments with lipid vesicles could present a possibility to analyze the kinetics of these proteins. Since a membrane-interaction was also predicted for EsxF/E in literature, it was also attempted to produce giant unilamellar vesicles (GUVs) in order to analyze the proteins within a lipid environment. However, no stable GUVs could be produced with the available options. Nevertheless, characterization and binding studies of proteins like EsxU/T and EsxF/E in membrane-like environments should be performed in the future, not only with smFRET but also with CD spectroscopy.

In addition, numerous results presented in this work indicate that EsxF, and possibly also EsxU, can form homodimers or even multimeric complexes. In contrast to the complex and the monomers EsxU and EsxT from ESX-4, EsxF and EsxE could be stably purified and stored (both as a complex and as monomers). A long-term investigation of EsxF monomers would be of particular interest, as the previous study from Tak et al. [176] has indicated the formation of filamentous structures after prolonged storage of EsxF/E. It is highly interesting to investigate if such filaments consist of both monomers or solely of EsxF. The high protein yield during purification seems promising for future crystallization experiments to investigate the structural properties of EsxF in more detail.

In addition, the studies on inter-loci complex formation between EsxB and EsxT indicate that proteins of different loci can also interact and possibly form stable complexes. This topic should be further investigated, as interactions or even complex formation between different WXG100 proteins could make a significant contribution to the versatility of the bacterium. A more precise characterization of these interactions could not only lead to a better understanding of the pathogenicity of *M. tuberculosis* and the disease progression of tuberculosis, but also open up new approaches for diagnostic or therapeutic options.

Bibliography

- [1] World Health Organization (Geneva), “Global Tuberculosis Report 2023,” 2023.
- [2] R. R. Nathavitharana and J. S. Friedland, “A tale of two global emergencies: tuberculosis control efforts can learn from the Ebola outbreak,” *European Respiratory Journal*, vol. 46, no. 2, pp. 293–296, 2015.
- [3] J. Furin, H. Cox, and M. Pai, “Tuberculosis,” *The Lancet*, vol. 393, no. 10181, pp. 1642–1656, 2019.
- [4] Y. Bao, L. Wang, and J. Sun, “A Small Protein but with Diverse Roles: A Review of EsxA in Mycobacterium–Host Interaction,” *Cells*, vol. 10, no. 1645, pp. 1–19, 2021.
- [5] C. Poulsen, S. Panjikar, 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*, vol. 9, no. 2, p. e89313, 2014.
- [6] D. Ilghari, K. L. Lightbody, V. Veverka, L. C. Waters, F. W. Muskett, P. S. Renshaw, and M. D. Carr, “Solution Structure of the *Mycobacterium tuberculosis* EsxG · EsxH Complex,” *Journal of Biological Chemistry*, vol. 286, no. 34, pp. 29993–30002, 2011.
- [7] M. A. Arbing, S. Chan, L. Harris, E. Kuo, T. T. Zhou, C. J. Ahn, L. Nguyen, Q. He, J. Lu, P. T. Menchavez, A. Shin, T. Holton, M. R. Sawaya, D. Cascio, and D. Eisenberg, “Heterologous Expression of Mycobacterial Esx Complexes in *Escherichia coli* for Structural Studies Is Facilitated by the Use of Maltose Binding Protein Fusions,” *PLOS ONE*, vol. 8, no. 11, p. e81753, 2013.
- [8] M. A. Arbing, M. Kaufmann, T. Phan, S. Chan, D. Cascio, and D. Eisenberg, “The crystal structure of the *Mycobacterium tuberculosis* Rv3019c-Rv3020c ESX complex reveals a domain-swapped heterotetramer,” *Protein Science*, vol. 19, no. 9, pp. 1692–1703, 2010.
- [9] R. Sundaramoorthy, P. K. Fyfe, and W. N. Hunter, “Structure of *Staphylococcus aureus* EsxA Suggests a Contribution to Virulence by Action as a Transport Chaperone and/or Adaptor Protein,” *Journal of Molecular Biology*, vol. 383, no. 3, pp. 603–614, 2008.
- [10] R. Pal, M. K. Bisht, and S. Mukhopadhyay, “Secretory proteins of *Mycobacterium tuberculosis* and their roles in modulation of host immune responses: focus on therapeutic targets,” *The FEBS Journal*, vol. 289, no. 14, pp. 4146–4171, 2022.
- [11] E. J. Peterson, A. A. Abidi, M. L. Arrieta-Ortiz, B. Aguilar, J. T. Yurkovich, A. Kaur, M. Pan, V. Srinivas, I. Shmulevich, and N. S. Baliga, “Intricate genetic programs controlling dormancy in *Mycobacterium tuberculosis*,” *Cell Reports*, vol. 31, no. 4, p. 107577, 2020.

- [12] W. Bitter, E. N. G. Houben, D. Bottai, P. Brodin, E. J. Brown, J. S. Cox, K. Derbyshire, S. M. Fortune, L.-Y. Gao, J. Liu, N. C. Gey 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*, vol. 5, no. 10, p. e1000507, 2009.
- [13] R. Koch, “Die Ätiologie der Tuberkulose,” *Berliner klinische Wochenschrift*, no. 15, pp. 221–230, 1882.
- [14] World Health Organization (Geneva), “Global Tuberculosis Report 2022,” 2022.
- [15] World Health Organization (Geneva), “Global Tuberculosis Report 2020,” 2020.
- [16] S. Y. Rodriguez-Takeuchi, M. E. Renjifo, and F. J. Medina, “Extrapulmonary Tuberculosis: Pathophysiology and Imaging Findings,” *RadioGraphics*, vol. 39, no. 7, pp. 2023–2037, 2019.
- [17] World Health Organization (Geneva), “BCG vaccines: WHO position paper,” *Weekly Epidemiological Record*, vol. 93, no. 8, pp. 73–96, 2018.
- [18] I. Abubakar, L. Pimpin, C. Ariti, R. Beynon, P. Mangtani, J. A. C. Sterne, P. E. Fine, P. G. Smith, M. Lipman, D. Elliman, J. M. Watson, L. N. Drumright, P. F. Whiting, E. Vynnycky, and L. C. Rodrigues, “Systematic review and meta-analysis of the current evidence on the duration of protection by bacillus Calmette–Guérin vaccination against tuberculosis,” *Health Technology Assessment*, vol. 17, no. 37, pp. 1–372, 2013.
- [19] World Health Organization (Geneva), “WHO consolidated guidelines on tuberculosis.,” in *WHO Operational Handbook on Tuberculosis* (W. H. O. (WHO), ed.), ch. Module 4: Treatment - Drug-susceptible Tuberculosis Treatment, 2022.
- [20] R. J. Wallace Jr., “Treatment of infections caused by rapidly growing mycobacteria in the era of the newer macrolides,” *Research in Microbiology*, vol. 147, pp. 30–35, 1996.
- [21] P. Supply and R. Brosch, “The Biology and Epidemiology of *Mycobacterium canettii*,” *Advances in Experimental Medicine and Biology*, pp. 27–41, 2017.
- [22] M. Orgeur and R. Brosch, “Evolution of virulence in the *Mycobacterium tuberculosis* Complex,” *Current Opinion in Microbiology*, vol. 41, pp. 68–75, 2018.
- [23] T. Fedrizzi, C. J. Meehan, A. Grottola, E. Giacobazzi, G. F. Serpini, S. Tagliazuchi, A. Fabio, C. Bettua, R. Bertorelli, V. D. Sanctis, F. Rumpianesi, M. Pecorari, O. Jousson, E. Tortoli, and N. Segata, “Genomic characterization of Nontuberculous Mycobacteria,” *Scientific Reports*, vol. 7, no. 45258, 2017.
- [24] E. Tortoli, “Chapter 1 - The Taxonomy of the Genus Mycobacterium,” in *Nontuberculous Mycobacteria (NTM)*, pp. 1–10, Elsevier, 2019.
- [25] P. Ehrlich, “Über die Färbung des Tuberkelbazillus,” *Deutsche Medizinische Wochenschriften*, no. 19, pp. 269–270, 1882.
- [26] F. Ziehl, “Zur Färbung des Tuberkelbacillus,” *Deutsche Medizinische Wochenschrift*, vol. 8, no. 33, p. 451, 1882.
- [27] F. Neelsen, “Ein casuistischer Beitrag zur Lehre von der Tuberkulose,” *Zentralblatt für die medizinischen Wissenschaften*, no. 28, pp. 497–501, 1883.

- [28] P. J. Bishop, "The History of the Ziehl-Neelsen Stain," *Tubercle*, vol. 51, no. 2, pp. 196–206, 1970.
- [29] S. Trifiro, A.-M. Bourgault, F. Lebel, and P. René, "Ghost Mycobacteria on Gram Stain," *Journal of Clinical Microbiology*, vol. 28, no. 1, pp. 146–147, 1990.
- [30] S. M. Batt, D. E. Minnikin, and G. S. Besra, "The thick waxy coat of mycobacteria, a protective layer against antibiotics and the host's immune system," *Biochemical Journal*, vol. 477, no. 10, pp. 1983–2006, 2020.
- [31] R. Bansal-Mutalik and H. Nikaido, "Mycobacterial outer membrane is a lipid bilayer and the inner membrane is unusually rich in diacyl phosphatidylinositol dimannosides," *PNAS*, vol. 111, no. 13, pp. 4958–4963, 2014.
- [32] D. E. Minnikin, O. Y. Lee, H. H. T. Wu, V. Nataraj, H. D. Donoghue, M. Ridell, M. Watanabe, L. Alderwick, A. Bhatt, and G. S. Besra, "Pathophysiological Implications of Cell Envelope Structure in *Mycobacterium tuberculosis* and Related Taxa," in *Tuberculosis - Expanding Knowledge* (W. Ribon, ed.), ch. 7, pp. 145–175, IntechOpen, 2015.
- [33] P. J. Brennan, "Structure, function, and biogenesis of the cell wall of *Mycobacterium tuberculosis*," *Tuberculosis*, vol. 83, no. 1-3, pp. 91–97, 2003.
- [34] K. C. Onwueme, C. J. Vos, J. Zurita, J. A. Ferreras, and L. E. N. Quadri, "The dimycocerosate ester polyketide virulence factors of mycobacteria," *Progress in Lipid Research*, vol. 44, no. 5, pp. 259–302, 2005.
- [35] S. Pitarque, G. Larrouy-Maumus, B. Payré, M. Jackson, G. Puzo, and J. Nigou, "The immunomodulatory lipoglycans, lipoarabinomannan and lipomannan, are exposed at the mycobacterial cell surface," *Tuberculosis*, vol. 88, no. 6, pp. 560–565, 2008.
- [36] M. Daffé and P. Draper, "The Envelope Layers of Mycobacteria with Reference to their Pathogenicity," *Advances in Microbial Physiology*, vol. 39, pp. 131–203, 1998.
- [37] M. Daffé and G. Etienne, "The capsule of *Mycobacterium tuberculosis* and its implications for pathogenicity," *Tubercle and Lung Disease*, vol. 79, no. 3, pp. 153–169, 1999.
- [38] C. Schaffer and P. Messner, "Review: Surface-layer glycoproteins: an example for the diversity of bacterial glycosylation with promising impacts on nanobiotechnology," *Glycobiology*, vol. 14, no. 8, pp. 31R–42R, 2004.
- [39] L. Brown, J. M. Wolf, R. Prados-Rosales, and A. Casadevall, "Through the wall: extracellular vesicles in Gram-positive bacteria, mycobacteria and fungi," *Nature Reviews Microbiology*, vol. 13, pp. 620–630, 2015.
- [40] V. Jarlier and H. Nikaido, "Permeability Barrier to Hydrophilic Solutes in *Mycobacterium chelonae*," *Journal of Bacteriology*, vol. 172, no. 3, pp. 1418–1423, 1990.
- [41] P. Brennan and H. Nikaido, "The Envelope of Mycobacteria," *Annual Review of Biochemistry*, vol. 64, pp. 29–63, 1995.
- [42] P. S. Raikwar, M. Mahapatra, V. Singh, and A. Bhardwaj, "Chapter 2: Lipid biosynthetic pathways as potential drug targets for emerging mycobacterial pathogens," in *Developments in Microbiology - Biology of Mycobacterial Lipids*, Academic Press, 2022.

- [43] A. Ortalo-Magné, M.-A. Dupont, A. Lemassu, A. B. Andersen, P. Gounon, and M. Daffé, “Molecular composition of the outermost capsular material of the tubercle bacillus,” *Microbiology*, vol. 141, no. 7, pp. 1609–1620, 1995.
- [44] R. Hunter, M. Olsen, C. Jagannath, and J. Actor, “Review: Multiple Roles of Cord Factor in the Pathogenesis of Primary, Secondary, and Cavitory Tuberculosis, Including a Revised Description of the Pathology of Secondary Disease,” *Annals of Clinical & Laboratory Science*, vol. 36, no. 4, pp. 371–386, 2006.
- [45] C. Pirson, R. Engel, G. J. Jones, T. Holder, O. Holst, and H. M. Vordermeier, “Highly Purified Mycobacterial Phosphatidylinositol Mannosides Drive Cell-Mediated Responses and Activate NKT Cells in Cattle,” *Clinical and Vaccine Immunology*, vol. 22, no. 2, pp. 178–184, 2015.
- [46] A. Tsirigotaki, J. De Geyter, N. Šoštaric´, A. Economou, and S. Karamanou, “Protein export through the bacterial Sec pathway,” *Nature Reviews Microbiology*, vol. 15, no. 1, pp. 21–36, 2016.
- [47] K. E. Chatzi, M. F. Sardis, A. Economou, and S. Karamanou, “SecA-mediated targeting and translocation of secretory proteins,” *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, vol. 1843, no. 8, pp. 1466–1474, 2014.
- [48] R. Patel, S. M. Smith, and C. Robinson, “Protein transport by the bacterial Tat pathway,” *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, vol. 1843, no. 8, pp. 1620–1628, 2014.
- [49] T. R. D. 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*, vol. 13, pp. 343–359, 2015.
- [50] N. Gey Van Pittius, J. Gamielien, W. Hide, G. Brown, R. Siezen, and A. Beyers, “The ESAT-6 gene cluster of *Mycobacterium tuberculosis* and other high G+C Gram-positive bacteria,” *Genome Biology*, vol. 2, no. 10, 2001.
- [51] A. M. Abdallah, N. C. G. van Pittius, P. A. D. Champion, J. Cox, J. Luirink, C. M. J. E. Vandenbroucke-Grauls, B. J. Appelmek, and W. Bitter, “Type VII Secretion - Mycobacteria show the way,” *Nature Reviews Microbiology*, vol. 5, pp. 883–891, 2007.
- [52] P. Andersen, A. B. Andersen, A. L. Sørensen, and S. Nagai, “Recall of long-lived immunity to *Mycobacterium tuberculosis* infection in mice,” *The Journal of Immunology*, vol. 154, no. 7, pp. 3359–3372, 1995.
- [53] S. T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry III., F. Tekaiia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M.-A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead, and B. G. Barrell, “Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence,” *Nature*, vol. 393, pp. 537–544, 1998.
- [54] K. Lewis, R. Liao, K. Guinn, M. Hickey, S. Smith, M. Behr, and D. Sherman, “Deletion of RD1 from *Mycobacterium tuberculosis* mimics Bacille Calmette-Guérin attenuation,” *The Journal of Infectious Diseases*, vol. 187, no. 1, pp. 117–123, 2003.

- [55] T. Hsu, S. Hingley-Wilson, B. Chen, M. Chen, A. Dai, P. Morin, C. Marks, J. Padiyar, C. Goulding, M. Gingery, D. Eisenberg, R. G. Russell, S. C. Derrick, F. Collins, S. L. Morris, C. H. King, and W. R. Jacobs, “The primary mechanism of attenuation of bacillus Calmette-Guérin is a loss of secreted lytic function required for invasion of lung interstitial tissue,” *PNAS*, vol. 100, no. 21, pp. 12420–12425, 2003.
- [56] M. Newton-Foot, R. M. Warren, S. L. Sampson, P. D. van Helden, and N. C. G. van Pittius, “The plasmid-mediated evolution of the mycobacterial ESX (Type VII) secretion systems,” *BMC Evolutionary Biology*, vol. 16, no. 62, 2016.
- [57] E. Dumas, E. C. Boritsch, M. Vandenberg, R. C. R. de la Vega, J.-M. Thiberge, V. Caro, J.-L. Gaillard, B. Heym, F. Girard-Misguich, R. Brosch, and G. Sapiel, “Mycobacterial Pan-Genome Analysis Suggests Important Role of Plasmids in the Radiation of Type VII Secretion Systems,” *Genome Biology and Evolution*, vol. 8, no. 2, pp. 387–402, 2016.
- [58] T. D. Mortimer, A. M. Weber, and C. S. Pepperell, “Evolutionary thrift: mycobacteria repurpose plasmid diversity during adaptation of type VII secretion systems,” *Genome Biology and Evolution*, vol. 9, no. 3, pp. 398–413, 2017.
- [59] A. Rivera-Calzada, N. Famelis, O. Llorca, and S. Geibel, “Type VII secretion systems: structure, functions and transport models,” *Nature Reviews Microbiology*, vol. 19, pp. 567–584, 2021.
- [60] 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*, vol. 53, no. 6, pp. 1677–1693, 2004.
- [61] 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*, vol. 187, no. 4, pp. 1238–1245, 2005.
- [62] K. S. H. Beckham, L. Ciccarelli, C. M. Bunduc, H. D. T. Mertens, R. Ummels, W. Lugmayr, J. Mayr, M. Rettel, M. M. Savitski, D. I. Svergun, W. Bitter, M. Wilmanns, T. C. Marlovits, A. H. A. Parret, and E. N. G. Houben, “Structure of the mycobacterial ESX-5 type VII secretion system membrane complex by single-particle analysis,” *Nature Microbiology*, vol. 2, 2017.
- [63] N. Famelis, A. Rivera-Calzada, G. Degliesposti, M. Wingender, N. Mietrach, J. M. Skelhel, R. Fernandez-Leiro, B. Böttcher, A. Schlosser, O. Llorca, and S. Geibel, “Architecture of the ESX-3/Type VII secretion system,” *Nature*, vol. 576, no. 7786, pp. 321–325, 2019.
- [64] N. Poweleit, N. Czudnochowski, R. Nakagawa, D. D. Trinidad, K. C. Murphy, C. M. Sassetti, and O. S. Rosenberg, “The structure of the endogenous ESX-3 secretion system,” *eLife*, vol. 8, 2019.
- [65] C. M. Bunduc, D. Fahrenkamp, J. Wald, R. Ummels, W. Bitter, E. N. G. Houben, and T. C. Marlovits, “Structure and dynamics of a mycobacterial type VII secretion system,” *Nature*, vol. 593, pp. 445–448, 2021.

- [66] T. L. Ramsdell, L. A. Huppert, T. A. Sysoeva, S. M. Fortune, and B. M. Burton, “Linked Domain Architectures Allow for Specialization of Function in the FtsK/SpoIIIE ATPases of ESX Secretion Systems,” *Journal of Molecular Biology*, vol. 427, no. 5, pp. 1119–1132, 2015.
- [67] Schrödinger, LLC, “The PyMOL Molecular Graphics System.” Version 2.5.4, August 2022.
- [68] M. J. Pallen, “The ESAT-6/WXG100 superfamily – and a new Gram-positive secretion system?,” *TRENDS in Microbiology*, vol. 10, no. 5, pp. 209–212, 2002.
- [69] M. H. Daleke, R. Ummels, P. Bawono, J. Heringa, C. M. J. E. Vandenbroucke-Grauls, J. Luirink, and W. Bitter, “General secretion signal for the mycobacterial type VII secretion pathway,” *PNAS*, vol. 109, no. 28, pp. 11342–11347, 2012.
- [70] S. Uplekar, B. Heym, V. Friocourt, J. Rougemont, and S. T. Cole, “Comparative Genomics of *esx* Genes from Clinical Isolates of *Mycobacterium tuberculosis* Provides Evidence for Gene Conversion and Epitope Variation,” *Infection and Immunity*, vol. 79, no. 10, pp. 4042–4049, 2011.
- [71] R. Prabhu and V. Singh, “The History of Tuberculosis: Past, Present, and Future,” *Advances in Microbiology*, vol. 9, no. 11, pp. 931–942, 2019.
- [72] 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*, vol. 2, no. 1, pp. 16076:1–23, 2016.
- [73] N. Reiling, S. Homolka, T. A. Kohl, C. Steinhäuser, K. Kolbe, S. Schütze, and J. Brandenburg, “Shaping the niche in macrophages: Genetic diversity of the *M. tuberculosis* complex and its consequences for the infected host,” *International Journal of Medical Microbiology*, vol. 308, no. 1, pp. 118–128, 2018.
- [74] G. R. Stewart, B. D. Robertson, and D. B. Young, “Tuberculosis: a problem with persistence,” *Nature Reviews Microbiology*, vol. 1, no. 2, pp. 97–105, 2003.
- [75] C. E. Barry 3rd, H. I. Boshoff, V. Dartois, T. Dick, S. Ehrt, J. Flynn, D. Schnappinger, R. J. Wilkinson, and D. Young, “The spectrum of latent tuberculosis: rethinking the biology and intervention strategies,” *Nature Reviews Microbiology*, vol. 7, pp. 845–855, 2009.
- [76] C. H. Liu, H. Liu, and B. Ge, “Innate immunity in tuberculosis: host defense vs pathogen evasion,” *Cellular & Molecular Immunology*, vol. 14, pp. 963–975, 2017.
- [77] S. Ehlers and U. E. Schaible, “The Granuloma in Tuberculosis: Dynamics of a Host-Pathogen Collusion,” *Frontiers in Immunology*, vol. 3, pp. 411:1–9, 2013.
- [78] J. Banchereau and R. Steinman, “Dendritic cells and the control of immunity,” *Nature*, vol. 392, pp. 245–252, 1998.
- [79] S. Hugues, A. Boissonnas, S. Amigorena, and L. Fetler, “The dynamics of dendritic cell - t cell interactions in priming and tolerance,” *Current Opinion in Immunology*, vol. 18, pp. 491–495, 2006.

- [80] L. Huang, E. V. Nazarova, S. Tan, Y. Liu, and D. G. Russell, "Growth of *Mycobacterium tuberculosis* in vivo segregates with host macrophage metabolism and ontogeny," *Journal of Experimental Medicine*, vol. 215, no. 4, pp. 1135–1152, 2018.
- [81] Y. Sibille and H. Reynolds, "Macrophages and polymorphonuclear neutrophils in lung defense and injury," *The American Review of Respiratory Disease*, vol. 141, no. 2, pp. 471–501, 1990.
- [82] R. Johnston, C. Godzik, and Z. Cohn, "Increased superoxide anion production by immunologically activated and chemically elicited macrophages," *The Journal of Experimental Medicine*, vol. 148, no. 1, pp. 115–127, 1978.
- [83] L. B. Adams, S. G. Franzblau, Z. Vavrin, J. B. Hibbs Jr., and J. L. Krahenbuhl, "L-Arginine-dependent macrophage effector functions inhibit activity of *Mycobacterium leprae*," *The Journal of Immunology*, vol. 147, no. 5, pp. 1642–1646, 1991.
- [84] L. Adams, M. C. Dinauer, D. E. Morgenstern, and J. L. Krahenbuhl, "Comparison of the roles of reactive oxygen and nitrogen intermediates in the host response to *Mycobacterium tuberculosis* using transgenic mice," *Tubercle and Lung Disease*, vol. 78, no. 5-6, pp. 237–246, 1997.
- [85] D. L. Granger, "Macrophage production of nitrogen oxides in host defence against microorganisms," *Research in Immunology*, vol. 147, no. 7, pp. 570–572, 1991.
- [86] J. S. Stamler, D. J. Singel, and J. Loscalzo, "Biochemistry of Nitric Oxide and Its Redox-Activated Forms," *Science*, vol. 258, no. 5090, pp. 1898–1902, 1992.
- [87] A. Cooper, J. Pearl, J. Brooks, S. Ehlers, and I. Orme, "Expression of the Nitric Oxide Synthase 2 Gene Is Not Essential for Early Control of *Mycobacterium tuberculosis* in the Murine Lung," *Infection and Immunity*, vol. 68, no. 12, pp. 6879–6882, 2000.
- [88] A. Koul, T. Herget, B. Klebl, and A. Ullrich, "Interplay between mycobacteria and host signalling pathways," *Nature Reviews Microbiology*, vol. 2, pp. 189–202, 2004.
- [89] A. O. Fels and Z. A. Cohn, "The alveolar macrophage," *Journal of Applied Physiology (1985)*, vol. 60, no. 2, pp. 353–369, 1986.
- [90] S. Sturgill-Koszycki, P. H. Schlesinger, P. Chakraborty, P. L. Haddix, H. L. Collins, A. K. Fok, R. D. Alien, S. L. Gluck, J. Heuser, and D. G. Russell, "Lack of Acidification in Mycobacterium Phagosomes Produced by Exclusion of the Vesicular Proton-ATPase," *Science*, vol. 263, no. 5147, pp. 678–681, 1994.
- [91] J. A. Armstrong and P. D. Hart, "Phagosome-lysosome interactions in cultured macrophages infected with virulent tubercle bacilli," *The Journal of Experimental Medicine*, vol. 142, no. 1, pp. 1–16, 1975.
- [92] C. M. Rosenberger and B. B. Finlay, "Phagocyte sabotage: disruption of macrophage signalling by bacterial pathogens," *Nature Reviews Molecular Cell Biology*, vol. 4, pp. 385–396, 2003.
- [93] 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. B. 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*, vol. 14, no. 8, pp. 1287–1298, 2012.

- [94] S. V. Jamwal, P. Mehrotra, A. Singh, Z. Siddiqui, A. Basu, and K. V. Rao, "Mycobacterial escape from macrophage phagosomes to the cytoplasm represents an alternate adaptation mechanism," *Scientific Reports*, vol. 6, 2016.
- [95] T. Dallenga, B. Corleis, and U. E. Schaible, "Infection of human neutrophils to study virulence properties of *Mycobacterium tuberculosis*," *Methods in Molecular Biology (MIMB)*, 2015.
- [96] T. Dallenga, U. Repnik, B. Corleis, J. Eich, R. Reimer, G. W. Griffiths, and U. E. Schaible, "*M. tuberculosis*-Induced Necrosis of Infected Neutrophils Promotes Bacterial Growth Following Phagocytosis by Macrophages," *Cell Host & Microbe*, vol. 22, no. 4, pp. 519–530, 2017.
- [97] A. M. Cooper, "Cell mediated immune responses in Tuberculosis," *Annual Review of Immunology*, vol. 27, pp. 393–422, 2009.
- [98] A. J. Wolf, B. Linas, G. J. T.-N. nez, E. Kincaid, T. Tamura, K. Takatsu, and J. D. Ernst, "*Mycobacterium tuberculosis* infects dendritic cells with high frequency and impairs their function in vivo," *The Journal of Immunology*, vol. 179, no. 4, pp. 2509–2519, 2007.
- [99] X. Jiao, R. Lo-Man, P. Guernonprez, L. Fiette, E. Dériaud, S. Burgaud, B. Gicquel, N. Winter, and C. Leclerc, "Dendritic Cells Are Host Cells for Mycobacteria In Vivo That Trigger Innate and Acquired Immunity," *The Journal of Immunology*, vol. 168, no. 3, pp. 1294–1301, 2002.
- [100] A. J. Pagán and L. Ramakrishnan, "Immunity and Immunopathology in the tuberculous Granuloma," *Cold Spring Harbor Perspectives in Medicine*, vol. 5, no. 9, 2015.
- [101] M. Allen, C. Bailey, I. Cahatol, L. Dodge, J. Yim, C. Kassissa, J. Luong, S. Kasko, S. Pandya, and V. Venketaraman, "Mechanisms of Control of *Mycobacterium tuberculosis* by NK Cells: Role of Glutathione," *Frontiers in Immunology*, vol. 6, 2015.
- [102] C.-C. Lu, T.-S. Wu, Y.-J. Hsu, C.-J. Chang, C.-S. Lin, J.-H. Chia, T.-L. Wu, T.-T. Huang, J. Martel, D. M. Ojcius, J. D. Young, and H.-C. Lai, "NK cells kill mycobacteria directly by releasing perforin and granulysin," *Journal of Leukocyte Biology*, vol. 96, pp. 1119–1129, 2014.
- [103] M. Oddo, T. Renno, A. Attinger, T. Bakker, H. R. MacDonald, and P. R. Meylan, "Fas ligand-induced apoptosis of infected human macrophages reduces the viability of intracellular *Mycobacterium tuberculosis*," *Journal of Immunology*, vol. 160, no. 11, pp. 5448–5454, 1998.
- [104] A. M. Dannenberg Jr., "Immunopathogenesis of Pulmonary Tuberculosis," *Hospital Practice*, vol. 28, pp. 51–58, 1993.
- [105] C. L. Cosma, D. R. Sherman, and L. Ramakrishnan, "The Secret Lives of the Pathogenic Mycobacteria," *Annual Review of Microbiology*, vol. 57, pp. 641–676, 2003.
- [106] P.-J. Cardona and J. Ruiz-Manzano, "On the nature of *Mycobacterium tuberculosis*-latent bacilli," *European Respiratory Journal*, vol. 24, pp. 1044–1051, Dec. 2004.
- [107] L. Skipper, "PROTEINS | Overview," in *Encyclopedia of Analytical Science*, pp. 344–352, Elsevier, 2005.

- [108] T. M. Gupte, M. Ritt, and S. Sivaramakrishnan, *ER/K-link—Leveraging a native protein linker to probe dynamic cellular interactions*, vol. 647. Elsevier, 2021.
- [109] J. L. Huffman and R. G. Brennan, “Prokaryotic transcription regulators: more than just the helix-turn-helix,” *Current opinion in Structural Biology*, vol. 12, no. 1, pp. 98–106, 2002.
- [110] A. N. Lupas and M. Gruber, “The structure of α -helical coiled coils,” *Advances in Protein Chemistry*, vol. 70, pp. 37–78, 2005.
- [111] J. Liu, Q. Zheng, Y. Deng, C.-S. Cheng, N. R. Kallenbach, and M. Lu, “A seven-helix coiled coil,” *PNAS*, vol. 103, no. 42, pp. 15457–15462, 2006.
- [112] H. C. Jubb, A. P. Pandurangan, M. A. Turner, B. Ochoa-Montaño, T. L. Blundell, and D. B. Ascher, “Mutations at protein-protein interfaces: Small changes over big surfaces have large impacts on human health,” *Progress in Biophysics and Molecular Biology*, vol. 128, pp. 3–13, 2017.
- [113] S. Majd, J. H. Power, and H. J. M. Grantham, “Neuronal response in Alzheimer’s and Parkinson’s disease: the effect of toxic proteins on intracellular pathways,” *BMC Neuroscience*, vol. 16, no. 69, 2015.
- [114] D. C. Fry and L. T. Vassilev, “Targeting protein–protein interactions for cancer therapy,” *Journal of Molecular Medicine*, vol. 83, pp. 955–963, 2005.
- [115] A. Bitran, W. M. Jacobs, X. Zhai, and E. Shakhnovich, “Cotranslational folding allows misfolding-prone proteins to circumvent deep kinetic traps,” *PNAS*, vol. 117, no. 3, pp. 1485–1495, 2020.
- [116] K. Ghosh, S. B. Ozkan, and K. A. Dill, “The ultimate speed limit to protein folding is conformational searching,” *Journal of the American Chemical Society*, vol. 129, no. 39, pp. 11920–11927, 2007.
- [117] T. Zou, N. Williams, S. B. Ozkan, and K. Ghosh, “Proteome folding kinetics is limited by protein half-life,” *PLOS ONE*, vol. 9, no. 11, 2014.
- [118] P. Atkins and J. de Paula, *Physical Chemistry*, ch. Chapter 22: The rates of chemical reactions, pp. 791–829. Oxford University Press, 8 ed., 2006.
- [119] P. L. Privalov, E. I. Tiktopulo, S. Y. Venyaminov, Y. V. Griko, G. I. Makhatadze, and N. N. Khechinashvili, “Heat capacity and conformation of proteins in the denatured state,” *Journal of Molecular Biology*, vol. 205, no. 4, pp. 737–750, 1989.
- [120] S. Yadav and F. Ahmad, “A new method for the determination of stability parameters of proteins from their heat-induced denaturation curves,” *Analytical Biochemistry*, vol. 283, no. 2, pp. 207–213, 2000.
- [121] A. Leitner, T. Walzthoeni, A. Kahraman, F. Herzog, O. Rinner, M. Beck, and R. Aebersold, “Probing Native Protein Structures by Chemical Cross-linking, Mass Spectrometry, and Bioinformatics,” *Molecular and Cellular Proteomics*, vol. 9, no. 8, pp. 1624–1649, 2010.
- [122] J. C. Sheehan and J. J. Hlavka, “The Cross-linking of Gelatin Using a Water-soluble Carbodiimide,” *Journal of the American Chemical Society*, vol. 79, pp. 4528–4529, 1957.

- [123] Z. Grabarek and J. Gergely, "Zero-Length Crosslinking Procedure with the Use of Active Esters," *Analytical Biochemistry*, vol. 185, no. 1, pp. 131–135, 1990.
- [124] *Bioconjugation and crosslinking technical handbook*, ch. 1: Chemical reactivity of crosslinking reagents, pp. 5–13. Thermo Fisher Scientific Inc., 2020.
- [125] *The Molecular Probes® Handbook: A Guide to Fluorescent Probes and Labeling Technologies*, ch. Chapter 2: Thiol-Reactive Probes, pp. 99–121. Thermo Fisher Scientific Inc., 11 ed., 2010.
- [126] E. Getz, M. Xiao, T. Chakrabarty, R. Cooke, and P. Selvin, "A Comparison between the Sulfhydryl Reductants Tris(2-carboxyethyl)phosphine and Dithiothreitol for Use in Protein Biochemistry," *Analytical Biochemistry*, vol. 273, pp. 73–80, 1999.
- [127] J. A. Burns, J. C. Butler, J. oran, and eorge M. Whitesides, "Selective Reduction of Disulfides by Tris(2-carboxyethyl)phosphine," *The Journal of Organic Chemistry*, vol. 56, no. 8, pp. 2648–2650, 1991.
- [128] T. Förster, "Energiewanderung und Fluoreszenz," *Die Naturwissenschaften*, vol. 33, no. 6, pp. 166–175, 1946.
- [129] T. Förster, "Zwischenmolekulare Energiewanderung und Fluoreszenz," *Annalen der Physik*, vol. 437, no. 1-2, pp. 55–75, 1948.
- [130] T. Förster, "Experimentelle und theoretische Untersuchung des zwischenmolekularen Übergangs von Elektronenanregungsenergie," *Zeitschrift für Naturforschung A*, vol. 4, no. 5, pp. 321–327, 1949.
- [131] A. Jabłoński, "Efficiency of Anti-Stokes Fluorescence in Dyes," *Nature*, vol. 131, pp. 839–840, June 1933.
- [132] P. F. Barbara, G. C. Walker, and T. P. Smith, "Vibrational modes and the dynamic solvent effect in electron and proton transfer," *Science*, vol. 256, no. 5059, pp. 975–981, 1992.
- [133] K. Nishiyama, Y. Watanabe, N. Yoshida, and F. Hirata, "Solvent dependence of stokes shift for organic solute–solvent systems: A comparative study by spectroscopy and reference interaction-site model–self-consistent-field theory," *The Journal of Chemical Physics*, vol. 139, no. 9, 2013.
- [134] G. G. Stokes, "On the Change of Refrangibility of Light," *Proceedings of The Royal Society of London*, vol. 6, pp. 195–200, 1852.
- [135] J. W. Lichtman and J.-A. Conchello, "Fluorescence microscopy," *Nature Methods*, vol. 2, no. 12, pp. 910–919, 2005.
- [136] D. A. McCarthy, *Cytometric analysis of cell phenotype and function*, ch. 3 - Fluorescence and fluorochromes, pp. 45–73. Cambridge: Cambridge University Press, 2001.
- [137] J. R. Lakowicz, *Principles of Fluorescence Spectroscopy*. Springer New York, NY, 3 ed., 2006.
- [138] R. M. Clegg, "Fluorescence resonance energy transfer," *Current Opinion in Biotechnology*, vol. 6, pp. 103–110, 1995.

- [139] P. Wu and L. Brand, “Resonance Energy Transfer: Methods and Applications,” *Analytical Biochemistry*, vol. 218, no. 1, pp. 1–13, 1994.
- [140] P. Schwille and E. Haustein, “Fluorescence correlation spectroscopy - an introduction to its concepts and applications,” *Spectroscopy*, vol. 94(22), 2001.
- [141] N. J. Greenfield, “Methods to Estimate the Conformation of Proteins and Polypeptides from Circular Dichroism Data,” *Analytical Biochemistry*, vol. 235, no. 1, pp. 1–10, 1996.
- [142] S. M. Kelly, T. J. Jess, and N. C. Price, “How to study proteins by circular dichroism,” *Biochimica et Biophysica Acta (BBA)*, vol. 1751, no. 2, pp. 119–139, 2005.
- [143] S. Prasad, I. Mandal, S. Singh, A. Paul, B. Mandal, R. Venkatramani, and R. Swaminathan, “Near UV-Visible electronic absorption originating from charged amino acids in a monomeric protein,” *Chemical Science*, vol. 8, no. 8, pp. 5416–5433, 2017.
- [144] K. Rosenheck and P. Doty, “The far ultraviolet absorption spectra of polypeptide and protein solutions and their dependence on conformation,” *PNAS*, vol. 47, no. 11, pp. 1775–1785, 1961.
- [145] N. J. Greenfield and G. D. Fasman, “Computed circular dichroism spectra for the evaluation of protein conformation,” *Biochemistry*, vol. 8, no. 10, pp. 4108–4116, 1969.
- [146] K. Park, A. Perczel, and G. D. Fasman, “Differentiation between transmembrane helices and peripheral helices by the deconvolution of circular dichroism spectra of membrane proteins,” *Protein Science*, vol. 1, no. 8, pp. 1032–49, 1992.
- [147] N. J. Greenfield, “Using circular dichroism spectra to estimate protein secondary structure,” *Nature Protocols*, vol. 1, no. 6, pp. 2876–2890, 2006.
- [148] C. Magaña Vergara, C. J. L. Kallenberg, M. Rogasch, C. G. Hübner, and Y.-H. Song, “A versatile vector for mycobacterial protein production with a functional minimized acetamidase regulon,” *Protein Science*, vol. 26, no. 11, pp. 2302–2311, 2017.
- [149] *The Molecular Probes® Handbook: A Guide to Fluorescent Probes and Labeling Technologies*, ch. Chapter 1: Fluorophores and Their Amine-Reactive Derivatives, pp. 10–97. Thermo Fisher Scientific Inc., 11 ed., 2010.
- [150] ATTO-TEC GmbH, “Product information: Thiol-reactive atto-label (maleimides).” <https://www.atto-tec.com/images/ATTO/Procedures/Mal.pdf>, 2023.
- [151] ATTO-TEC GmbH, “Product information: Atto 647n.” https://www.atto-tec.com/fileadmin/user_upload/Katalog_Flyer_Support/ATTO_647N.pdf, 2022.
- [152] A. Lobley, L. Whitmore, and B. A. Wallace, “Dichroweb: an interactive website for the analysis of protein secondary structure from circular dichroism spectra,” *Bioinformatics*, vol. 18, no. 1, pp. 211–212, 2002.
- [153] N. Sreerama and R. W. Woody, “A self-consistent method for the analysis of protein secondary structure from circular dichroism,” *Analytical Biochemistry*, vol. 209, no. 1, pp. 32–44, 1993.
- [154] N. Sreerama and R. W. Woody, “Estimation of protein secondary structure from circular dichroism spectra: Comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set,” *Analytical Biochemistry*, vol. 287, no. 2, pp. 252–260, 2000.

- [155] A. J. Miles and B. A. Wallace, “Synchrotron radiation circular dichroism spectroscopy of proteins and applications in structural and functional genomics,” *Chemical Society Reviews*, vol. 35, no. 1, pp. 39–51, 2006.
- [156] A. Micsonai, F. Wien, L. Kernya, Y.-H. Lee, Y. Goto, M. Réfrégiers, , and J. Kardos, “Accurate secondary structure prediction and foldrecognition for circular dichroism spectroscopy,” *PNAS*, vol. 112, no. 24, pp. E3095–E3103, 2015.
- [157] N. J. Greenfield, “Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions,” *Nature Protocols*, vol. 1, no. 6, pp. 2527–2535, 2006.
- [158] J. Seelig and H.-J. Schönfeld, “Thermal protein unfolding by differential scanning calorimetry and circular dichroism spectroscopy: Two-state model versus sequential unfolding,” *Quarterly Reviews of Biophysics*, vol. 49, no. e9, pp. 1–24, 2016.
- [159] T. J. Lambert, “FPbase: a community-editable fluorescent protein database,” *Nature Methods*, vol. 16, pp. 277–278, 2019.
- [160] A. Micsonai, Éva Moussong, F. Wien, E. Boros, H. Vadász, N. Murvai, Y.-H. Lee, T. Molnár, M. Réfrégiers, Y. Goto, Ágnes Tantos, and J. Kardos, “Bestsel: webserver for secondary structure and fold prediction for protein cd spectroscopy,” *Nucleic Acids Research*, vol. 50, no. W1, pp. W90–W98, 2022.
- [161] A. J. Miles, S. G. Ramalli, and B. A. Wallace, “Dichroweb, a website for calculating protein secondary structure from circular dichroism spectroscopic data,” *Protein Science*, vol. 31, no. 1, pp. 37–46, 2021.
- [162] C. W. Dunnett, “A multiple comparison procedure for comparing several treatments with a control,” *Journal of the American Statistical Association*, vol. 50, no. 272, pp. 1096–1121, 1955.
- [163] P. J. Tummino and R. A. Copeland, “Residence Time of Receptor-Ligand Complexes and Its Effect on Biological Function,” *Biochemistry*, vol. 47, no. 20, pp. 5481–5492, 2008.
- [164] I. Letunic and P. Bork, “Interactive tree of life (itol) v6: recent updates to the phylogenetic tree display and annotation tool,” *Nucleic Acids Research*, vol. 52, no. W1, pp. W78–W82, 2024.
- [165] F. Madeira, N. Madhusoodanan, J. Lee, A. Eusebi, A. Niewielska, A. R. N. Tivey, R. Lopez, and S. Butcher, “The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024,” *Nucleic Acids Research*, vol. 52, no. W1, pp. W521–W525, 2024.
- [166] I. Jarmoskaite, I. AlSadhan, P. P. Vaidyanathan, and D. Herschlag, “How to measure and evaluate binding affinities,” *eLife*, vol. 9, p. e57264, 2020.
- [167] P. E. Wright and H. J. Dyson, “Linking folding and binding,” *Current Opinion in Structural Biology*, vol. 19, no. 1, pp. 31–38, 2009.
- [168] K. Sugase, H. J. Dyson, and P. E. Wright, “Mechanism of coupled folding and binding of an intrinsically disordered protein,” *Nature*, vol. 447, pp. 1021–1025, 2007.

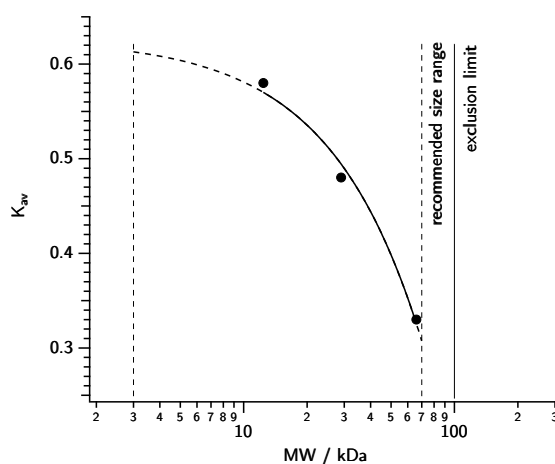
- [169] P. S. Renshaw, P. Panagiotidou, A. Whelan, S. V. Gordon, R. G. Hewinson, R. A. Williamson, and M. D. Carr, "Conclusive evidence that the major T-cell antigens of the *Mycobacterium tuberculosis* Complex ESAT-6 and CFP-10 form a tight, 1:1 complex and characterization of the structural properties of ESAT-6, CFP-10, and the ESAT-6 · CFP-10 complex," *Journal of Biological Chemistry*, vol. 277, no. 24, pp. 21598–21603, 2002.
- [170] A. K. Meher, N. C. Bal, K. V. R. Chary, and A. Arora, "*Mycobacterium tuberculosis* H37Rv ESAT-6–CFP-10 complex formation confers thermodynamic and biochemical stability," *The FEBS Journal*, vol. 273, no. 7, pp. 1445–1462, 2006.
- [171] M. I. D. Jonge, G. Pehau-Arnaudet, M. M. Fretz, F. Romain, D. Bottai, P. Brodin, N. Honoré, G. Marchal, W. Jiskoot, P. England, S. 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*, vol. 189, no. 16, pp. 6028–6034, 2007.
- [172] A. D. Vogt and E. D. Cera, "Conformational Selection or Induced Fit? A Critical Appraisal of the Kinetic Mechanism," *Biochemistry*, vol. 51, no. 30, pp. 5894–5902, 2012.
- [173] H.-J. Lee, Y. Woo, T.-W. Hahn, Y. M. Jung, and Y.-J. Jung, "Formation and maturation of the phagosome: A key mechanism in innate immunity against intracellular bacterial infection," *Microorganisms*, vol. 8, no. 9, p. 1298, 2020.
- [174] K. L. Lightbody, D. Ilghari, L. C. Waters, G. Carey, M. A. Bailey, R. A. Williamson, P. S. Renshaw, and M. D. Carr, "Molecular features governing the stability and specificity of functional complex formation by *Mycobacterium tuberculosis* CFP-10/ESAT-6 family proteins," *Journal of Biological Chemistry*, vol. 283, no. 25, pp. 17681–17690, 2008.
- [175] B. N. Dominy, D. Perl, F. X. Schmid, and C. L. B. III, "The effects of ionic strength on protein stability: The cold shock protein family," *Journal of Molecular Biology*, vol. 319, no. 2, pp. 541–554, 2002.
- [176] U. Tak, T. Dokland, and M. Niederweis, "Pore-forming Esx proteins mediate toxin secretion by *Mycobacterium tuberculosis*," *Nature Communications*, vol. 12, no. 349, 2021.
- [177] T. A. Gray, R. R. Clark, N. Boucher, P. Lapierre, C. Smith, and K. M. Derbyshire, "Intercellular communication and conjugation are mediated by ESX secretion systems in mycobacteria," *Science*, vol. 354, no. 6310, pp. 347–350, 2016.
- [178] 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," *PNAS*, vol. 115, no. 5, pp. E1002–E1011, 2018.
- [179] M. Lagune, V. Le Moigne, M. D. Johansen, F. Vásquez 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*, vol. 18, no. 8, p. e1010771, 2022.

- [180] 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*, vol. 25, no. 1, p. 103585, 2022.
- [181] N. Sankey, H. Merrick, P. Singh, J. Rogers, A. Reddi, S. D. Hartson, and A. Mitra, "Role of the *Mycobacterium tuberculosis* ESX-4 Secretion System in Heme Iron Utilization and Pore Formation by PPE Proteins," *mSphere*, vol. 8, no. 2, pp. e00573–22, 2023.
- [182] M. Lagune, C. Petit, F. V. Sotomayor, M. D. Johansen, K. S. H. 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*, vol. 167, no. 7, 2021.
- [183] G. Kleiger, R. Grothe, P. Mallick, and D. Eisenberg, "GXXXG and AXXXA: common alpha-helical interaction motifs in proteins, particularly in extremophiles," *Biochemistry*, vol. 41, no. 19, pp. 5990–5997, 2002.
- [184] M. G. Teese and D. Langosch, "Role of GxxxG Motifs in Transmembrane Domain Interactions," *Biochemistry*, vol. 54, no. 33, pp. 5125–5135, 2015.
- [185] J. Teutschbein, G. Schumann, U. Möllmann, S. Grabley, S. T. Cole, and T. Munder, "A protein linkage map of the ESAT-6 secretion system 1 (ESX-1) of *Mycobacterium tuberculosis*," *Microbiological Research*, vol. 164, no. 3, pp. 253–259, 2009.
- [186] Y. Fan, K. Tan, G. Chhor, R. Jedrzejczak, and A. Joachimiak, "The Crystal Structure of a Secreted Protein EsxB (Wild-Type, in P21 Space Group) from *Bacillus anthracis* Str. Sterne," *to be published*, 2013.
- [187] M. Zoltner, W. M. Ng, T. Palmer, and W. N. Hunter, "Crystal Structure of the Homodimeric Esat-6-Family Proteins EsxA and EsxB from *Geobacillus thermodenitrificans*," *to be published*, 2012.
- [188] T. A. Sysoeva, M. A. Zepeda-Rivera, L. A. Huppert, and B. M. Burton, "Dimer recognition and secretion by the esx secretion system in *Bacillus subtilis*," *PNAS*, vol. 111, no. 21, pp. 7653–7658, 2014.
- [189] A. Mahmood, S. Srivastava, S. Tripathi, M. A. Ansari, M. Owais, and A. Arora, "Molecular characterization of secretory proteins Rv3619c and Rv3620c from *Mycobacterium tuberculosis* H37Rv," *The FEBS Journal*, vol. 278, no. 2, pp. 341–353, 2011.
- [190] H. Pandey, S. Tripathi, K. Srivastava, D. K. Tripathi, M. Srivastava, S. Kant, K. K. Srivastava, and A. Arora, "Characterization of culture filtrate proteins Rv1197 and Rv1198 of ESAT-6 family from *Mycobacterium tuberculosis* H37Rv," *Biochimica et Biophysica Acta*, vol. 1861, no. 2, pp. 396–408, 2017.
- [191] P. Brodin, M. I. de Jonge, L. Majlessi, C. Leclerc, M. Nilges, S. T. Cole, and R. Brosch, "Functional analysis of early secreted antigenic target-6, the dominant T-cell antigen of *Mycobacterium tuberculosis*, reveals key residues involved in secretion, complex formation, virulence, and immunogenicity," *Journal of Biological Chemistry*, vol. 280, no. 40, pp. 33953–33959, 2005.
- [192] H. Pandey, F. Fatma, S. M. Yabaji, M. Kumari, S. Tripathi, K. Srivastava, D. K. Tripathi, S. Kant, K. K. Srivastava, and A. Arora, "Biophysical and immunological characterization of the ESX-4 system ESAT-6 family proteins Rv3444c and Rv3445c from *Mycobacterium tuberculosis* H37Rv," *Tuberculosis*, vol. 109, pp. 85–96, 2018.

- [193] J. Seelig and A. Seelig, “Chemical Protein Unfolding – A Simple Cooperative Model,” *The Journal of Physical Chemistry B*, vol. 127, no. 39, pp. 8285–8508, 2023.
- [194] A. Waterhouse, J. Procter, D. Martin, M. Clamp, and G. J. Barton, “Jalview version 2 - a multiple sequence alignment editor and analysis workbench,” *Bioinformatics*, vol. 9, no. 25, pp. 1189–1191, 2009.
- [195] K. Tamura, G. Stecher, and S. Kumar, “MEGA11: Molecular Evolutionary Genetics Analysis Version 11,” *Molecular Biology and Evolution*, vol. 38, no. 7, pp. 3022–3027, 2021.

Chapter A

Supplementary Figures



Column-specific data		Run-specific data			
CV	V_0	V_{delay}	sample volume	flow rate	buffer
120 ml	40 ml	1.3 ml	300 μl	1 ml/min	PBS

Protein (MW)	V_e / ml	corrected V_e / ml	V_e/V_0
Cytochrome c (12.4 kDa)	67.67	66.37	1.66
Carbonic Anhydrase (29.0 kDa)	79.52	78.22	1.96
BSA (66.0 kDa)	87.33	86.03	2.15

Figure A.1: Calibration curve of SEC column Superdex 16/600 75 μg . The chromatogram of SEC with a mix of standard proteins is shown on the left. The resulting calibration curve is shown on the right. The manufacturer gives the recommended range of protein size for separation as between 3 kDa and 70 kDa; the exclusion limit is noted as 100 kDa. The table below the graphs lists column-specific and run-specific data used for evaluation.

CV: column volume, V_0 : column void volume, V_{delay} : delay in sample application due to tubing from sample inlet to the column, (corrected) V_e : elution volume ($- V_{\text{delay}}$),

$$K_{\text{av}}: \frac{V_e - V_0}{CV - V_0}$$


```

      W-G                                     H---D--h---H
-----MNADPVL SYNFD AIEYSVRQE IHTT AARFN AALQELRSQIAPLQQL W TREA A AAYHA EQLKWHQAASALNEILDLGNVAVRHG A D D V A H A D R E A A G A W A R -----
-----MSQLSYDFGEIEYTVRQE IATHSRFNAALEELRAQIAPLQQY W TREA A AAYQVEQARWNOAAAALNEILFSLGNVAVRDGSDVEVAATDRS A A N A W G V -----
-----MSQIITYNHGEIDALVA-DVKG IINKFQAGLEELQHDIQPLVQQA E C Q E A T A Y Q E Y Q K A W H Q S A E D L N Q I L T Q L N A K V D Q G N Q P A Q H T D Q S A A G A W H H -----
-----MDPVL SYNFD AIEHTVRQE IHTTSMRFNAALEQLRSQIAPLQQL W SREA A A A Y Q T E Q F K W H Q A A A A L N Q I L V E L G N V A V R D G A D E V A G A D R R A A G G W A R -----
-----MSTPNTLNADFDLMRSVAGITDARNEEIR A M L Q A F I G R M S G V P P S V W G C L A A A R F Q D V V D R W M A E S T R L Y H V L H A I A D T I R H N E A A L R E A A G Q I H A R H I A A A G G D L
-----MSTPLGADFVMTTVAGQIDVYRNDDIRSMLQTFIGRMNTVPPTV W G C V A A T R F R D V V E R W M A E S L L H T T L Q R I A E T I R H N E R T L R E A A D G H A Q R L I G G V G E T L
-----MAVFQNDLALLDSTAKKIDGKYQEF T A M Q S Q L R - D R V A V G T S T W C Q Q A R H A F D E A M A R F D Q E M G D I Q K V L I G I H D T M E S N K R R I Q E M D E S Q T F -----
MSSPAGANPFTTSTASNTLSADFDLMRTVA A T T D A R N E E I R A L L Q A F I G R M N D V P A S V W G C V A A V R F K H A M D R W G A E S T R L Y H A L H T T I A E T I R H N E A L R E A G Q N H A H H I G A A G A D L
      W-G                                     H---D--h---H

```

Figure A.3: Alignment of mycobacterial EsxU and EsxT homologues. Multiple Sequence alignment performed with Clustal Omega [165], processing done with Jalview [194]. Species abbreviations are *mtb*: *M. tuberculosis*, *mse*: *M. smegmatis*, *mab*: *M. abscessus*, *mma*: *M. marinum*. WxG motif is annotated above and highlighted in the sequences in red, consensus secretory signal HxxxD(E)xxhxxxH is annotated above and highlighted in the sequences in yellow. *H*: highly conserved and *h*: less conserved hydrophobic residue. Light blue boxes in sequences of EsxU proteins show the shifted position of the “*h*” residue of the secretion signal. Grey boxes indicate AxxxA and GxxxG motifs.

Chapter B

List of Figures

2.1	Simplified cell wall structures of Gram-negative and Gram-positive bacteria and mycobacteria.	6
2.2	The ESX gene loci and the ESX secretion system model.	10
2.3	Schematic representation: cellular immune response and granuloma formation after infection with Mtb.	13
2.4	Models of coupled protein binding and folding.	17
2.5	Reaction scheme of EDC/sNHS-mediated crosslinking.	21
2.6	Sulfhydryl reductants and maleimide reaction.	22
2.7	Jabłoński diagram.	24
2.8	Schematic representation of fluctuations in fluorescent signal and creation of an auto-correlation function.	29
2.9	Polarized light.	31
2.10	The CD effect.	32
2.11	Wavelength-dependent absorption of peptides with different secondary structures and their distinct CD spectra.	33
3.1	Positioning of cysteine mutations in EsxF and EsxU according to EsxB-TE	37
3.2	Fluorescent dyes used in this work.	40
3.3	Schematic representation of the confocal FRET setup.	46
3.4	Overview of FRET data recording and analysis.	50
3.5	Characteristics of each FRET-pair using AF488 as a donor (D) and either AF647 or Atto647N as an acceptor (A).	50
4.1	Typical purification process of EsxB/A	52
4.2	Different purified Esx complexes and their respective monomers	52
4.3	Purification of the orphan complex EsxF/E	53
4.4	Purification of the complex EsxU/T from ESX4	54
4.5	Crosslinking experiments different Esx proteins	55
4.6	Crosslinking experiments with EsxF/E complex and monomers	56
4.7	Crosslinking experiments with EsxT/U complex and monomers	56
4.8	Temperature-dependent CD spectroscopy analysis of the EsxF/E complex.	58
4.9	Temperature-dependent CD spectroscopy analysis of the EsxF monomer.	59
4.10	Temperature-dependent CD spectroscopy analysis of the EsxE monomer.	60
4.11	Secondary structure composition of EsxF/E and its monomers.	61
4.12	FRET measurements with double- and single-labeled EsxF monomers.	63
4.13	Temperature-dependent CD spectroscopy analysis of the EsxU/T complex.	65
4.14	Temperature-dependent CD spectroscopy analysis of the EsxU monomer.	66

4.15	Temperature-dependent CD spectroscopy analysis of the EsxT monomer. . .	67
4.16	Secondary structure composition of EsxU/T and its monomers.	68
4.17	EsxA Titration.	70
4.18	Change of center peak position in FRET histogram.	71
4.19	Time-dependent progression of EsxB/A complex formation in dependence of EsxA concentration.	72
4.20	Maximum EsxB/A complex concentration measured with the addition of varying concentrations of binding partner EsxA.	73
4.21	Experimentally observed pseudo-first-order rate constant as a function of the binding partner EsxA.	74
4.22	EsxT Titration.	75
4.23	Time-dependent progression of EsxB/T complex formation in dependence of EsxT concentration.	76
4.24	Combining smFRET and chemical crosslinking with EDC and NHS to detect protein-protein interaction.	77
4.25	FRET histograms of the EsxB monomer and with the addition of EsxA under different pH conditions	78
4.26	Time-dependent progression of EsxB/A complex formation in dependence of the pH value.	79
4.27	CD spectroscopy of EsxB/A and its monomers in different pH conditions . .	80
4.28	Phylogenetic trees showing the relationship between Esx proteins from <i>M. tuberculosis</i> and between selected Gram-positive bacteria.	82
A.1	Calibration curve of SEC column Superdex 16/600 75 pg.	i
A.2	Alignment of different WXG100 proteins.	ii
A.3	Alignment of mycobacterial EsxU and EsxT homologues.	iii

Chapter C

List of Tables

2.1	ESX systems in different mycobacteria and their roles.	8
2.2	Esx proteins of <i>M. tuberculosis</i>	11
2.3	Typical features of CD spectra of proteins in different conformations.	33
3.1	Overview of expression vectors.	36
3.2	Primers used for QuikChange site-directed mutagenesis.	36
3.3	Parameters of different Esx-protein monomers and complexes.	39
3.4	Fluorophore characteristics.	41
3.5	Composition of phosphate-citrate buffers used for pH-dependent smFRET measurements.	45
4.1	List of purified Esx-complexes and yield	51
4.2	Thermodynamic parameters from thermal unfolding of EsxF/E and its monomers.	60
4.3	Thermodynamic parameters from thermal unfolding of EsxU/T and its monomers.	67
5.1	Thermodynamic parameters and structural characteristics of different mycobacterial Esx protein pairs and monomers.	89

Chapter D

Curriculum Vitae

Personal Data

Christina Jana Luise Kallenberg



Career

since 10/2024 Project Manager
National Reference Center for Mycobacteria, Research Center Borstel

University Studies and Doctorate

05/2017 - 10/2022 Scientific Assistant and Doctoral Student
Institute of Physics, University Lübeck

10/2013 - 03/2017 **Master Study: Infection Biology** (Grade: 1.6)
University of Lübeck

10/2010 - 09/2013 **Bachelor Study: Biology** (Grade: 2.3)
Friedrich-Alexander-University, Erlangen-Nürnberg

Education

2005 - 2010 Secondary School in Erlangen, Germany (“Abitur” Grade: 2.5)

2001 - 2005 Secondary School in Taipei, Taiwan

1998 - 2001 Elementary School in Bangkok, Thailand

1997 - 1998 Elementary School in Erlangen, Germany