

Deciphering the role of mycobacterial secreted proteins through structural biology

Citation for published version (APA):

Gijsbers Alejandro, A. (2022). *Deciphering the role of mycobacterial secreted proteins through structural biology*. [Doctoral Thesis, Maastricht University]. ProefschriftMaken. <https://doi.org/10.26481/dis.20220125ag>

Document status and date:

Published: 01/01/2022

DOI:

[10.26481/dis.20220125ag](https://doi.org/10.26481/dis.20220125ag)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

**Deciphering the role of
mycobacterial secreted proteins
through structural biology**

© copyright Abril Gijsbers, 2022

Printing: ProefschriftMaken || www.proefschriftmaken.nl

ISBN 978-94-6423-666-8

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior permission of the author or the copyright-owning journals for previous published chapters.

Deciphering the role of mycobacterial secreted proteins through structural biology

DISSERTATION

to obtain the degree of Doctor at the Maastricht University,
on the authority of the Rector Magnificus,
Prof.dr. Rianne M. Letschert
in accordance with the decision of the Board of Deans,
to be defended in public
on Tuesday, 25th of January 2022, at 16.00 hours

by

Abril Gijsbers Alejandre

Born in Oaxaca, Mexico, on May 4th, 1990

Promotor:

Prof. dr. P.J. Peters

Copromotors:

Dr. R.B.G. Ravelli

Dr. C. López-Iglesias

Assessment Committee:

Prof. dr. C.P.M. Reutelingsperger (chairman)

Dr. MD. MSc. B.C. de Jong, Institute of Tropical Medicine

Dr. G.A.F. Nicolaes

Prof. dr. T.K. Sixma, Netherlands Cancer Institute

Prof. dr. M.A.G.G. Vooijs

This research was part of the M4I research program and received financial support from the Dutch Province of Limburg under the LINK program, the Netherlands Organization for Scientific Research (NWO) in the framework of the Fund New Chemical Innovations (project numbers 731.016.407) and in the framework of the National Roadmap NEMI (project number 184.034.014). This research was also part from the European Union's Horizon 2020 Research and Innovation Programme (grant agreement 766970 Q-SORT) and the Marie Skłodowska-Curie Actions (call FP7-PEOPLE-2013-ITN).

We have to remember that what we observe is not nature in itself, but nature exposed to our method of questioning

—Werner Heisenberg

Table of contents

CHAPTER 1	General introduction	9
CHAPTER 2	Cryo-EM structures from sub-nL volumes using pin-printing and jet vitrification	25
CHAPTER 3	<i>Mycobacterium tuberculosis</i> ferritin: a suitable workhorse protein for cryo-EM development	45
CHAPTER 4	Priming mycobacterial ESX-secreted protein B to form a channel-like structure	63
CHAPTER 5	Structural analysis of the partially disordered protein EspK from <i>Mycobacterium tuberculosis</i>	95
CHAPTER 6	Structure of EspB–EspK complex: the non-identical twin of the PE–PPE–EspG secretion mechanism	115
CHAPTER 7	Impact paragraph	141
CHAPTER 8	Discussion and summary	147
	General discussion	149
	Summary (English)	155
	Samenvatting (Dutch)	157
	Resumen (Spanish)	159
CHAPTER 9	Appendices	163
	References	165
	List of publications	189
	Acknowledgments	191
	Curriculum vitae	193

CHAPTER 1

1

General introduction

Abril Gijbers

I. Using structural biology to understand a biological process

Nature is built mainly from a couple of chemical elements, which interact to form molecules, and then organised to form cells, the basic unit of life. From those molecules, proteins are the most structurally and functionally sophisticated of all. They are biomolecules composed of a series of amino acids organised in a specific sequence covalently linked by a peptide bond, which defines the structure and function of the protein. Participation in every biological process makes proteins essential for the correct maintaining and performance of all living organisms.

It was initially assumed that all proteins should acquire a structure to perform their function. Thus, obtaining the protein structure would be an open window towards understanding the role of proteins in the cell. The structure-function paradigm ruled for many years, assuming that its 3D arrangement determined the function of a protein, and the loss of structure would lead to a malfunction of the protein and consequently to a disease (1,2). This paradigm was based on the catalytic activity of enzymes and their lock-and-key model, where a specific substrate (key) fits perfectly into a rigid active site (lock). This idea was further supported by the development of X-ray crystallography for protein structure determination (3,4). The discovery of highly flexible and intrinsically disordered proteins changed this paradigm, where proteins do not need to be rigid or have a structure to be functional (1). Furthermore, it has been now demonstrated that the protein structure acquired by crystallography is a snapshot of the functional cycle of a protein and that proteins are not rigid structures; on the contrary, they are flexible and dynamic molecules. Bioinformatics studies suggest that the vast majority of the proteins have long disordered regions, some of which are primarily unstructured (5). Although the paradigm changed, the power of structural visualisation has not. Structure (or lack of structure) determination remains essential to characterise the function of a protein, and contains extremely valuable information to carry out other kind of studies.

A bit more than 60 years ago has passed since the first protein structure was described by John C. Kendrew and collaborators (4) using X-ray crystallography. As X-rays interact with electrons, scientists realised that exposing crystals to X-rays generated a particular diffraction pattern. From this information an electron density map can be calculated in which a 3D model of the molecule can be built. The development of macromolecular crystallography (MX) has solved 158,485 out of 179,954 protein structures [numbers taken from the Protein Data Bank (PDB) (6) on November 16th, 2021], which corresponds to 88% of the structures known. However, it comes with the prerequisite that macromolecules form a crystal, a rigid structure out of a naturally flexible molecule. Even if the crystal is formed, flexible regions of a protein adopting multiple conformations within the crystal leads to poorly defined densities for those regions, making them hard to be resolved. That is why protein flexibility is considered one of the “Achilles’ heel” of this technique (7). Scientists use strategies to remove flexible regions or limit the movement of a protein to allow crystal formation; however, flexibility

is the natural state of a protein, and doing so would reduce our view of its mechanism of action. Other techniques have been developed to obtain structural information at (near) atomic resolution and to complement the information at the macromolecule level to characterise the functional cycle of a protein. A complementary technique is nuclear magnetic resonance (NMR), which has the advantage that it uses protein in solution to solve the 3D structure of proteins and their dynamics. The disadvantage is that the data analysis becomes convoluted or gets lost when the protein size exceeds 40 kDa (8). That could be acceptable considering that the average protein size is 35-53 kDa (9), depending on the organism. However, most proteins carry their function in complex with itself and/or with other proteins making the functional unit much larger. Cryo-electron microscopy (cryo-EM) has revolutionised structural biology in recent years, resolving structures that could not be obtained before (10). It has the advantage that no crystals are needed, it can target large dynamic complexes and resolve different conformations of the molecule, and that it is significantly less sensitive to structural heterogeneity and impurities than the other two methods (11).

In essence, cryo-electron microscopy (cryo-EM) images specimens under cryogenic conditions with a transmission electron microscope. Depending on the method and sample, it can be divided into sub-disciplines: cryo-electron tomography (cryo-ET), single-particle analysis (SPA) cryo-EM, cryo-scanning transmission electron microscopy (cryo-STEM), and electron crystallography. As the name hints, the radiation source is electrons that are accelerated down in a high-vacuum column towards the sample (12). The sample then scatters the electrons, and the detector records a phase-contrast image. From those sub-disciplines, SPA cryo-EM is the most common technique to determine high-resolution structure of proteins. Significant developments like direct electron detectors (13), motion correction (14,15), and improvement in the **imaging** process algorithms (16,17) have taken cryo-EM from a “blobology era” to a continuous stream of near-atomic resolution structures (18). Further advances like better energy filters, aberration correctors, monochromators, and new generation detectors enhance the signal-to-noise ratio, making it possible to reach atomic resolution (19,20) or decrease the protein size-limit for cryo-EM (21). This technique is moving forward in gigantic steps, attracting more and more people worldwide and promising to solve most structural problems. However, intrinsic issues of the method and the sample are still to be alleviated to reach that point. Sample preparation is considered one of the main bottlenecks (22). One significant part is protein purification; however, this is true for all the *in vitro* techniques. Issues specific to SPA sample preparation can be listed from every step of the workflow (18). Some of them include the interaction of the protein with the air-water interface (preferential orientation, complex instability, or protein denaturation) (23,24), artefacts of the support film (deformation during image acquisition, charging, protein interaction) (25,26), low reproducibility caused by the current method/equipment, and the need of good experience to carry out the embedding of the sample in vitreous ice (vitrification) without ice contamination or grid deformation (18). Enormous efforts to

overcome these limitations or extend cryo-EM capabilities have been developed in the last years. Still, no single method/instrument seems to alleviate all the problems (27).

With the technological improvements in cryo-EM, structural biology is now in a new era, where protein, DNA, RNA, and complexes thereof that could not be crystallised or solved by NMR are now solved by cryo-EM. Good examples are membrane complexes like secretion systems (28), the tripartite multidrug efflux systems (29), or the transient receptor potential (TRP) channels that were this year's Nobel Prize subject (30,31). Furthermore, cryo-EM has the potential to classify and solve the structure of multiple functional states. By vitrifying after chemical or physical stimuli, one could obtain structural information of the different steps that form the functional cycle of the protein or protein complexes to explain a biological process (32).

2. Tuberculosis: a silent pandemic

2.1 Today's facts and numbers

Tuberculosis (TB) is one of the leading causes of death worldwide. Before the COVID-19 pandemic, TB was the main cause of death from a single infectious agent (33). Human tuberculosis results from an infection by bacillus *Mycobacterium tuberculosis*, spreading when people expel droplets that contain bacteria into the air by either coughing, sneezing, spitting, or simply breathing. The World Health Organization (WHO) estimates that one-quarter of the world's population is infected with latent or active *M. tuberculosis* (33). Although there is some doubt about this estimation (34), some facts cannot be questioned: approximately 10 million people fell ill in 2020 (**Figure 1**), from which 1.5 million people died (33), which is more than 4,000 people a day. To put these numbers in perspective: COVID-19 has turned our lives upside down since the beginning of 2020, a new disease that we did not know how to treat, using unspecific drugs and experimental methods has a mortality rate of 2%. On the other hand, tuberculosis has been known for thousands of years and is considered curable, yet 15% of patients die. So why is tuberculosis still a global health issue? There is not simple answer to that; TB is a very complex problem that should be seen from a scientific, medical, social, and economic perspective.

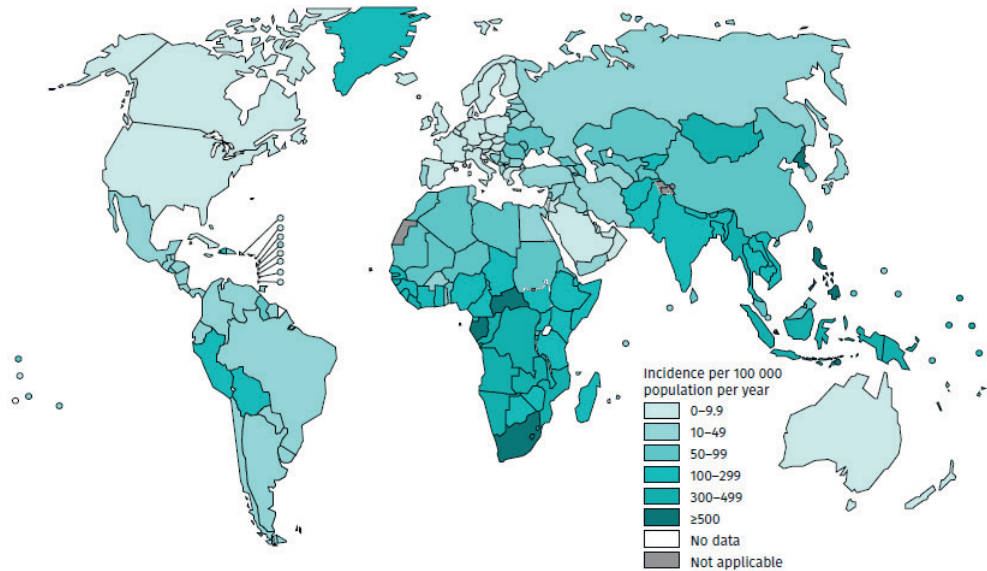


Figure 1. Estimated TB incidence rates from 2020. Figure taken from the global tuberculosis report 2021 (33).

TB affects primarily low and middle-income countries (**Figure 1**), facing fragile health and economic systems that cannot keep the fight by themselves. For this reason, the WHO set a strategy to end TB, and the United Nations (UN) has included TB as a target within its Sustainable Development Goals, calling all UN members to solve this global problem together. This thesis lies within the scope of the scientific challenge we have to accomplish WHO's strategy to end TB. Nonetheless, it is crucial to acknowledge the other factors if we want to tackle this problem successfully.

TB is considered a disease of poverty, causing vulnerability, stigma, and discrimination (35). The social impact of being ill with TB is a barrier in detecting and treating patients, making the task of TB eradication much harder. WHO, in coordination with national programmes accommodates money to offer treatment for free. However, TB patients and families still face a financial crisis from direct or indirect costs (33,36). Income loss is one of the most critical factors contributing to the further impoverishment of patients and households. Informal unemployment is a common practice in these countries, not including social benefits or support from the employer. This situation could drive the patient to start working before is advised, putting other people at risk of infection or delaying their full recovery. National programmes must consider some financial compensation in their budget to safeguard patient's integrity and the people who depend on them.

In the 1940s, the first antibiotic that *M. tuberculosis* was sensitive to was discovered, streptomycin. Unfortunately, soon after, bacteria acquired resistance. Other drugs were discovered in the 1950s and 60s, offering the application of combination therapy, reducing the opportunity for the bacteria to acquire resistance to one of the drugs, and shortening the therapeutic regimen from 18-24 months to 6 months (37). Today, the first-line of drugs anti-TB is composed of isoniazid, rifampin, ethambutol, and pyrazinamide, which have saved 66 million people in the last 20 years (33). However, treatment's length, together with strong side-effects, like hepatotoxicity, exanthema, and arthralgia, results in a lack of patient's adherence to up to 23% (38,39). This issue has enhanced the development of multidrug-resistant (MDR), and extensive drug-resistant (XDR) strains, depending on how many drugs the strain is resistant to, in some cases, to all. As far as we know, exchange of genomic material is limited in *M. tuberculosis* (40); thus, any resistance-conferring mutation probably occurs spontaneously. It is quite common to find a subpopulation of drug-resistant strains in a patient with an active infection (41). If there is no proper completion of the therapy (by any cause), it creates a drug-selection pressure that will allow the prevalence of such strains. For this reason, drug-resistant tuberculosis is considered a consequence of men's actions (41). The TB treatment is far from optimal, highlighting the value of the medical care in the hospitals and afterward with surveillance to fulfil the therapy regimen.

Undoubtedly, the number of TB incidence and deaths have decreased over time; however, extrapolation of the current trend suggests it could take another 40 years to eradicate TB and the investment of US\$ 300 billion (calculated from current funding) (33). Apart from the lives that will be lost during this time, it is estimated that diseases like TB contribute to 100 million people passing the poverty line each year (42), failing to accomplish the first Sustainable Development Goal of the UN: no poverty. For this reason, WHO and UN set the goal to eradicate TB by 2030; unfortunately, the surge of MDR-TB threatens any control of the situation, because it complicates the treatment at every level. Patients with drug resistance will be prescribed a combination of second and third-line TB drugs for 9-11 months or more (43). Compared to the standard treatment, this medicine is more expensive (in some cases paid by the patient), has more potent side effects, is less effective, and is not always available (37,44). This is a severe problem because, suppose no additional measures are taken, and there is a 40% increment of drug-resistant cases. In that case, it is estimated that 75 million additional people will die in the next 30 years compared to the UN mortality projections. The cost will not be US\$ 300 billion but US\$ 16.7 trillion, which will negatively affect the global gross domestic product (GDP). (45). The world has seen a step back in the eradication of TB before. In the 1990s, with the emergence of HIV, the number of TB cases and deaths changed their trend worldwide. Nowadays, co-infection is common, and has a tragic synergy: HIV increases the risk of having active TB infection, and TB is the number one cause of death among HIV-positive patients. The interaction of antitubercular and antiretroviral therapy has been described (46), which brings another variable to the table for physicians' consideration when a regimen is designed.

Everything was challenging enough, and the severe acute respiratory syndrome coronavirus 2 (SARS-CoV2) came to complicated everything even more. Since early 2020, the world's focus turned to COVID-19, the disease caused by this virus, affecting the work done on many communicable diseases. The long-term consequences are yet to be seen; however, the WHO recently reported years of regression in eradicating TB due to measurements taken to control the virus. Likewise, the limited access to diagnosis and treatment led to death increment for the first time in more than a decade (33). Lockdowns implemented in most cities around the world caused a dramatic decrease in TB diagnosis (47). As time passed and experience was gained, governments implemented strategies to deal with this problem, like the COVID-TB double test in India (48). Unfortunately, that still resulted in a decrease of diagnosis by 25%. Missing diagnosis leads to inadequate treatment, which in turn contributes to the severity and mortality rate of the disease and the risk of transmission and development of drug resistance (49,50). Co-infection of these two biological agents has been studied from the early stages of the COVID-19 pandemic (51), and preliminary results could indicate a similar synergy (52) as seen for HIV. Resource are other factors that will have a significant weight on the future perspectives of TB. Medical personnel involved in controlling of disease outbreaks have valuable expertise needed to control COVID-19, so many were recruited, leaving other diseases unattended (53). As emphasised before, due to the high level of non-adherence for TB patients, surveillance is a critical factor in disease control, and the lack of personnel could jeopardise the strategy to end TB. Do we need to talk about money? Lucica Ditiu, Executive Director of the Stop TB Partnership, spoke aloud in early 2020 about the unfairness of the monetary-resource difference that COVID-19 had compared to other diseases (54). A disease of no more than 2 years old now has more than ten vaccines available and more to come, while TB, the oldest disease on record, has a 100 years-old vaccine that is ineffective in adults. The development of new technologies that could lead to early diagnosis and therapies that improve patients' quality of life and that helps to eradicate this ancient enemy once and for all is essential to achieve the Sustainable Development Goals. However, gaps in the knowledge of the infection process are stopping us from moving forward.

2.2 Walking alongside *Mycobacterium tuberculosis* throughout time

It is now 2021; six centuries have passed since the end of the Middle Ages, characterised by strong religious power and obscurantism and the beginning of an era of scientific discoveries and developments. One of them is the discovery of *Mycobacterium tuberculosis* as the causative agent of TB by Robert Koch 141 years ago. From this, other discoveries followed, like the first vaccine against *M. tuberculosis*, by Albert Calmette and Camille Guérin 40 years later (100 years ago), an attenuated strain of *M. bovis* named Bacillus Calmette–Guérin (BCG). More details about this strain will be discussed below. Our knowledge on the pathogenesis of *M. tuberculosis* has improved since then, and the development of a therapeutic regimen has changed the disease from fatal to curable; unfortunately, this has not been enough to solve the health crisis. It is not surprising, considering that *M. tuberculosis* has challenged our immune system for thousands of

years, evolving to be one of the most efficient pathogens in human history. In order to design new therapies that could eradicate TB, a proper understanding of the pathogenesis is essential, including the evolution that the microorganism has gone through to prevent an unfavourable future.

Human TB is caused by *M. tuberculosis* and other members of a closely related cluster (>99.95% sequence similarity), known as the *Mycobacterium tuberculosis* complex (MTBC) (55). The human-associated MTBC are obligated human pathogens with no animal or environmental reservoir, making it plausible that human evolution and behaviour could influence genetic changes of MTBC over time. Different studies agree that MTBC originated in Africa (56,57) when an environmental microorganism acquired the ability to multiply within macrophages and spread from host to host. This adaptation could have occurred thanks to changes in human behaviour, like controlling fire and social gatherings (58). Initially, it was assumed that human-associated MTBC came from animals when domestication started; however, phylogenetic studies have proven that it was the other way around (59). There is a consensus of where MTBC came from, but not when this happened. Scientists have created models to calculate the age, reaching a wide range, from millions of years (57) to less than 6,000 years old (60). The later estimation seems not compatible with evidence of MTBC DNA and lipids found in humans and a bison remains estimated to be 11,000-9,000 years old (61,62) and 17,000 years old (63), respectively. Based on the close relationship between MTBC and its host, Comas *et al.* (2013) used a model based on human evolution comparison, suggesting a close association for 70,000 years. Many of these studies use mutation rates from contemporary or young ancestors as molecular clocks; however, we ignore if this can be extrapolated to the original MTBC or the effect of latency in the evolution of those microorganisms (64). Regardless of when it all started, it has been accepted that once the MTBC ancestor became a human pathogen, it began to diverge into different human-associated lineages and those transmitted and adapted to domestic and wild animals (55,65) (**Figure 2**).

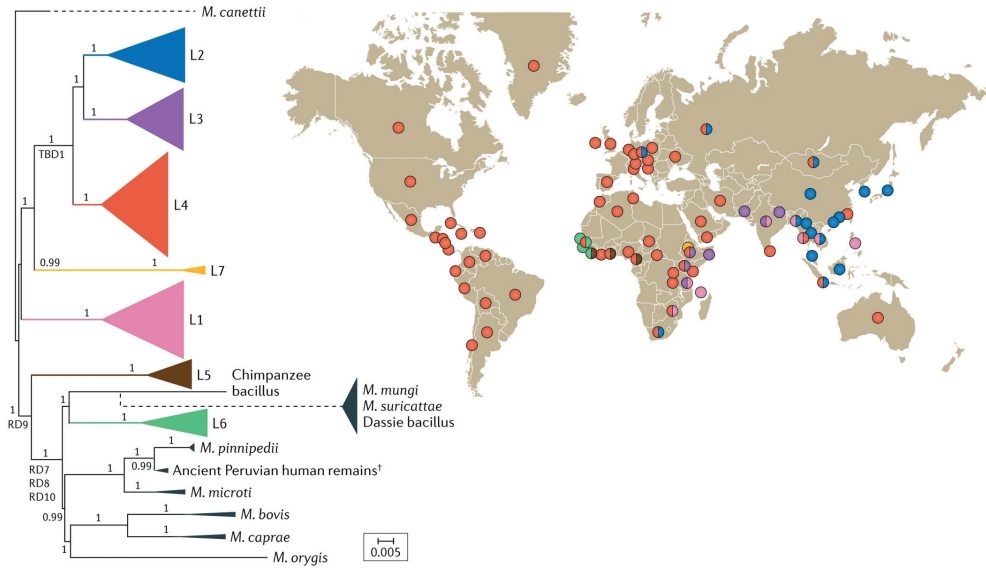


Figure 2. Geographic distribution of human-associated MTBC. Phylogeny of human-associated MTBC lineages (in colour) and animal-associated lineages (grey) rooted with *M. canettii*, and their global distribution. Characteristic deletion of the different lineages is indicated [M. tuberculosis-specific deletion 1 (TBD1) and other regions of difference (RD7-10)]. Scale bar represents the number of nucleotide substitutions—Figure taken from (64). Lineages L2 and L4 are present in most parts of the world due to globalisation. This map only shows the lineages that are most represented in those areas.

Human-associated lineages have a specific geographic distribution, some widely spread and others restricted to local areas. Africa is the only continent that harbours all lineages, supporting the idea that MTBC originated from Africa. Over time, MTBC started a local adaptation to its host behaviour. It is expected that in early times, when the human population was limited, MTBC would stay in latency for decades, allowing the human survival until reproductive age, and thus, the survival of the bacteria (66). That changed with domestication of plants and animals and urban developments, generating a human population explosion and with them of MTBC (56,67). With the first settlements and population growth, the latency phase would decrease, and the bacteria could become more virulent without compromising its own survival. This hypothesis is reflected in the virulence difference between lineages, like the hyper-virulent lineage 2 (also known as Beijing) that coincide with high-density cities in East Asia (67). Clinical data indicate that immune response, pathological signs, response to the BCG vaccine, drug-resistance development, and other features depend on the genetic signature of MTBC lineages (68,69). These genomic variations and phenotype differences increase the difficulty of understanding the infection process to develop new therapies and vaccines that efficiently deal with all the lineages (70). A deeper understanding will pave the way to successful therapy.

2.3 *Mycobacterium tuberculosis* pathogenesis

TB starts with inhaling of a few bacteria within a droplet that go deeply through the airway, reaching the alveoli. In this compartment, *M. tuberculosis* recruits permissive macrophages that lack microbicidal mechanisms by coating its surface with phthiocerol dimycocoserate (PDIM). The bacteria are then engulfed by the macrophages, allowing the cross through the epithelial barrier. (71). This step seems to be important for an effective infection and would not be possible within non-sterile parts of the respiratory system like the pharynx that contains commensal microbes, which are recognised by macrophages that produce microbicidal reactive nitrogen species (71). What comes next is a bit ambiguous because, despite the enormous progress describing different steps of the life cycle of *M. tuberculosis*, it has not always been easy to put the pieces together. One of the reasons is that we do not have an animal model that can fully reproduce each characteristic of human TB (72). Considering what has been described above, that this pathogen has adapted to humans for millennia, studies in zebrafish or mice will probably not reflect what happens in our body, making the extrapolation difficult or even erroneous. Another possible explanation could be that multiple outcomes are happening simultaneously: the host will successfully clear some bacteria, and others will enter in a latent state or continue with an active infection. It has been acknowledged that *M. tuberculosis* has a complex interaction with its host. By using a counterintuitive strategy, the pathogen embraces the immune system's actions to survive and flourish within this environment. For that, this pathogen has an arsenal of biomolecules to block or circumvent the immune system response, expected from a millennia-adapted microorganism.

If we simplify the life cycle of *M. tuberculosis* to what is commonly accepted: alveolar macrophages take *M. tuberculosis* and migrate to the lung's interstitium, where other immune cells are called up to form a granuloma in an attempt to contain the infection. Macrophages and other immune cells recognise foreign organisms, then internalise them via phagocytosis and digest to eliminate such threats or present antigens to develop adaptive immunity. This event triggers a cascade of actions by the pathogen, starting with the inhibition of the phagolysosome maturation (73), rupture of phagosome membrane and translocation to the cytosol of the macrophage where the bacteria proliferates (74), and finally the promotion of necrosis over apoptosis for bacterial content release (75) (**Figure 3**). Most of these events are associated with the type VII secretion system (T7SS), and for that reason, T7SS is considered to be one of the major virulence factors and a possible target for new treatment (76).

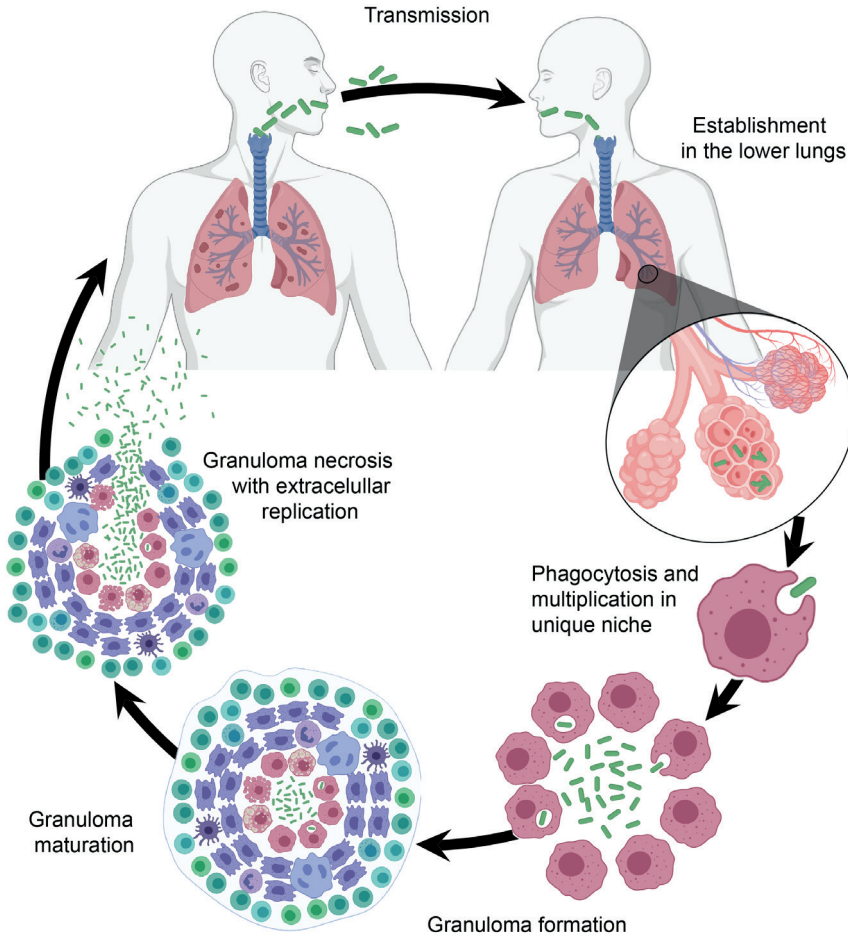


Figure 3. The life cycle of *Mycobacterium tuberculosis*. The infection starts with the inhalation of droplets containing a few bacteria that travels to the host’s alveoli. Alveolar macrophages engulf the pathogen and recruit other immune cells to form a granuloma. However, *M. tuberculosis* neutralises macrophages’ microbicidal actions by blocking the phagolysosome’s maturation and lysing the phagosome membrane, which leads to the replication of the bacteria in the cytosol of the host cell followed by necrosis for bacterial content release and dissemination —created with BioRender.com.

Genomic comparison identified regions of difference (RDs) between mycobacterial species (**Figure 2**); RD1 immediately attracted scientific attention due to its absence in many non-pathogenic species (77). Further studies concluded that the lack of this genomic region in the *M. bovis* BCG strain, used as TB-vaccine, was the main contributor to its attenuation (78). The RD1 contains part of a locus that codes for the ESX-1 secretion system, the first T7SS discovered in mycobacteria. Pathogenic mycobacteria have up to five different T7SS (ESX-1 to -5) paralogue protein complexes with specific functions unable to complement each other. ESX-1, -3, and -5 have been implicated with *M. tuberculosis* virulence. ESX-1 has been involved in the phagosomal

escape, cellular inflammation, host cell death, and dissemination of the bacteria to neighbouring cells (74,79-82). Similarly, ESX-5 participates in nutrient uptake (83) and host cell death (80). In contrast, ESX-3 has been linked to iron homeostasis (84), an inorganic molecule used in vital processes and depleted during infection. Finally, ESX-2 and ESX-4 are the least characterised, where ESX-4 is involved in DNA transfer in other species (85) and is seen as the ancestor of the five ESX-systems (86). All five systems are built with the same core proteins embedded in the inner membrane of mycobacteria, suggesting that the architecture of the protein complexes is overall the same; however, they differ in the protein substrates secreted. Due to their external location and possibility of manipulating the environment, ESX-secreted proteins are the subject of intense work. Despite the progress in recent years, missing information, like the structure and function of ESX-1 and other RD1 elements, the existence of the outer-membrane machinery and its components, hamper the complete understanding of ESX-1 function and its role in *M. tuberculosis* pathogenesis. My thesis aims to fill in some of these gaps in our knowledge.

3. Scope of the thesis

In general, this thesis focuses on contributions to the field of cryo-EM that can potentially improve the way protein structures are solved, followed by the structural characterisation of proteins relevant for the virulence of *M. tuberculosis* to better understand their function.

Chapter 2 describes the development of a new machine that improves sample preparation at different levels. The VitroJet uses pin printing to deposit nano-litres of sample on the EM-grid (carrier surface), which substantially reduces sample waste. The deposited protein solution is vitrified by jets of cryogen followed by submersion into a cryogen bath. This new system allows the use of premounted autogrids, making the workflow fully automated and user-friendly.

The lack of a robust standard protein limits the growth of the cryo-EM field, as standard proteins are required to test new hardware, like the VitroJet, and commission, maintain, and quality-control high-end microscopes. Ferritin has become a popular option; due to the high solubility and favourable dimensions, combined with its symmetry, this protein allows swift characterisation of the performance of the microscope using a minimal number of micrographs. Nevertheless, ferritin from different species used nowadays lacks stability or has a complicated purification protocol, leading to insufficient protein.

Chapter 3 describes a simple purification protocol with a high-yield of mycobacterial ferritin BfrB. As a proof of concept, we solved the structure to a resolution of 2.12 Å. Such structure displays the characteristic cage shape with a unique C-terminal extension that loops into the cage region of the shell and provides extra stability to the protein. Together with the description of the BfrB structure, we shared with the

scientific community a highly needed expression and purification protocol that could serve the growing cryo-EM community to characterise and push the limits of their electron microscopes and workflows.

Chapter 3 is a transition from cryo-EM improvement to its usage in the structural characterisation of biological molecules. In **chapter 4**, we use cryo-EM to gain insight into the functioning of the ESX-1 secretion system of *M. tuberculosis*. Here, we describe the factors and events that prime the substrate EspB to oligomerise. We show that the quaternary structure of EspB is conserved across slow-growing mycobacterial species but not in the fast-growing *M. smegmatis*. EspB assembles into a channel with dimensions and characteristics suitable for the transit of ESX-1 substrates, evidenced by the presence of another EspB trapped within. The structure and assembly of EspB suggest a possible function as a structural element of ESX-1.

Early studies suggested that the functioning of EspB was closely related to the substrate EspK; hence, characterisation of the latter could give insight into the mechanism of action of the former. **Chapter 5** describes a structural characterisation of the ESX-1 substrate EspK. This study was limited to low resolution due to the flexible nature of the protein. By small-angle X-ray scattering, we show that EspK is a long molecule with a maximal dimension of 228 Å. It consists of two independent domains at each end of the protein connected by a flexible unstructured region driving the protein to coexist as an ensemble of conformations. The C-terminal domain consists of a mixture of α -helices and β -strands, while the N-terminal portion is mainly helical with an elongated shape. Sequence conservation suggests that this architecture is preserved amongst the different mycobacteria species, proposing specific roles for the N- and C-terminal domains assisted by a central flexible linker. Previous studies had shown an interaction between the C-terminal domain and EspB; thus, this domain was further characterised by X-ray crystallography to gain insight into its function. In **chapter 6**, we describe the first high-resolution structure available of EspK. This structure bound to EspB shows that the interaction occurs in the helical tip of EspB, similar to its homologues PE-PPEs and their chaperone EspG. This interaction disrupts EspB oligomerisation, proposing that this interaction happens in the cytoplasm, ensuring EspB to be in a secretion-competent state. However, conservation of the interaction in *M. smegmatis*, where no EspB oligomer is present, suggests that the primary function of EspK is still to be determined. In analogy to PE-PPE and their chaperone EspG, the structure presented in this study supports the idea that EspK facilitates EspB secretion, being this a different secretion mechanism in ESX-1.

To finalise the thesis, I talk about the impact of this research in developing cryo-EM to improve protein structure determination to understare biological mechanisms, specifically, the pathogenesis of *M. tuberculosis* and how its elucidation could alleviate TB's health problem (**chapter 7**).

CHAPTER 2

2

Cryo-EM structures from sub-nL volumes using pin-printing and jet vitrification

Raimond B. G. Ravelli, Frank J. T. Nijpels, Rene J. M. Henderiks, Giulia Weissenberger, Sanne Thewessem, **Abril Gijsbers**, Bart W. A. M. M. Beulen, Carmen López-Iglesias and Peter J. Peters

Abstract

The increasing demand for cryo-electron microscopy (cryo-EM) reveals drawbacks in current sample preparation protocols, such as sample waste and lack of reproducibility. Here, we present several technical developments that provide efficient sample preparation for cryo-EM studies. Pin printing substantially reduces sample waste by depositing only a sub-nanolitre volume of sample on the carrier surface. Sample evaporation is mitigated by dew point control feedback loops. The deposited sample is vitrified by jets of cryogen followed by submersion into a cryogen bath. Because the cryogen jets cool the sample from the centre, premounted autogrids can be used and loaded directly into automated cryo-EMs. We integrated these steps into a single device, named VitroJet. The device's performance was validated by resolving four standard proteins (apoferritin, GroEL, worm hemoglobin, beta-galactosidase) to $\sim 3\text{\AA}$ resolution using a 200-kV electron microscope. The VitroJet offers a promising solution for improved automated sample preparation in cryo-EM studies.

I. Introduction

Within just a few years, single-particle cryo-electron microscopy (cryo-EM) has become a powerful mainstream technique to resolve high-resolution 3D structures of macromolecules. Technological breakthroughs in microscopes, detectors, and processing have contributed to the resolution revolution (87). However, numerous advancements are yet to be made (88,89). Reproducible sample preparation has emerged as one of the main bottlenecks (90). The sample preparation technique for cryo-EM, pioneered by Dubochet and others (91-94), allows particles to be preserved in a thin vitreous layer in their native state. The current leading commercial sample preparation devices still rely on the work of these pioneers. Sample preparation begins with an EM grid that is plasma-treated in a separate instrument to increase hydrophilicity. As this effect is only temporary, the wetting properties depend on the time between glow discharge and sample deposition, temperature, humidity, and cleanliness of the environment (26,90,95). Next, a few microliters of sample are manually pipetted onto the grid, and >99.99% of it is blotted away by filter paper, after which an aqueous film is allowed to thin by evaporation. During this dynamic process, air-water interfaces will form (91,96,97), which can be detrimental to the structure of interest as proteins tend to adsorb to such interfaces and (partly) denature (96). The resulting film of sample on a perforated carrier (25,98,99) has a generally concave shape (96,100) due to drying and draining, with the centre being the thinnest. The grid is held by tweezers and plunged into a bath of cryogen to vitrify the sample (101), so that it can be observed in the vacuum of the microscope. This vitrification should be fast enough to prevent the formation of ice crystals. Grids are stored under cryogenic conditions until analysis. For microscope loading and unloading, the fragile grid needs to be assembled into either a sturdy cartridge or the tip of a holder: both processes are cumbersome and often harmful to the fragile grid.

Many steps of the above protocol require skilled operators, careful handling, clean tools, clean cryogens, and controlled environments. Some steps are intrinsically difficult to control accurately, such as positioning of the filter papers, blotting force, humidity, and flatness of the filter papers, as well as the position and shape of the tweezers. As a result, reproducibility is lacking. The required training and skills to obtain reliable grid quality is a significant entry barrier for new scientists. The increasing demand and evolution of cryo-EM call for improved, scalable methods. We have developed a method which allows for better control and, ultimately, minimal operator intervention. Our method consists of 1) an integrated glow-discharge module to control and minimize the time between plasma cleaning and sample deposition; 2) pin-printing for sample application, which requires only sub-nanolitre sample volumes and eliminates sample blotting; and 3) jet vitrification, which allows for the handling of autogrids. We integrated these features into a single setup, termed the VitroJet, and used it to prepare four standard proteins to obtain high-resolution single-particle reconstructions.

2. Methods

2.1 Protein purification

Human apoferritin overexpressed in *Escherichia coli* was kindly provided via Evgeniya Pechnikova (Thermo-Fisher) by Dr. Fei Sun (Institute of Biophysics, Chinese Academy of Sciences). Chaperonin-60 from *E. coli* (GroEL) was ordered from Sigma (C7688), dissolved in a buffer comprised 50 mM Tris (pH 8.0), 100 mM KCl, 10 mM MgCl₂, 2 mM DTT, and 80 mM trehalose, and used without further purification at a concentration of 10 mg/mL. Approximately 400- μ L blood from earthworm *Lumbricus terrestris* (WHBG) was extracted from the seventh segment of the body (102) and injected into a Superose[®] 6 Increase 10/300 GL size-exclusion chromatography column (SEC) with an elution buffer of 20 mM Tris (pH 8.0), 150 mM NaCl, and 10 mM CaCl₂. After purification, WHBG was concentrated to 6 mg/mL. Beta-galactosidase (β -gal) from *E. coli*, ordered from Sigma (G5635), was further purified by SEC with a Superdex[®] 200 Increase 10/300 GL column and eluted with 25 mM Tris (pH 8.0), 50 mM NaCl, 2 mM MgCl₂, and 1 mM TCEP. The ligand phenylethyl β -D-thiogalactopyranoside (PETG) was purchased from Sigma-Aldrich (catalog #P1692) and prepared as described (103). We used a final β -gal protein concentration of 4 mg/mL with 7.5 mM PETG.

2.2 Sample preparation

Quantifoil R1.2/1.3 Au300 and UltraAufoil R1.2/1.3 Au300 grids (Quantifoil Micro Tools, Jena, Germany) were used as sample carriers. Grids were pre-clipped before entering the VitroJet. Sample deposition occurred in a climate chamber at room temperature with a humidity of 93%, where pin and grid were cooled toward dew-point temperature. A sub-nanolitre volume of sample was pin-printed onto the grids with a writing speed of 0.3 mm/s. Samples were vitrified within 80 ms after sample deposition by two pressurized jets of liquid ethane for 50 ms at 99 K and 1 bar.

2.3 Single-particle cryo-EM

Micrographs were collected on a 200-kV Thermo Fisher Tecnai Arctica microscope equipped with a Falcon 3 detector. For each standard protein, micrographs were collected with a calibrated pixel size of 0.935 Å. For apoferritin and GroEL, total integrated electron flux of $\sim 40 e^-/\text{Å}^2$ in counting mode at a defocus range of 0.7–1.5- μ m underfocus was used. For WHBG and β -gal, a total integrated electron flux of $\sim 43 e^-/\text{Å}^2$ in counting mode, at 0.6–1.4- μ m underfocus for WHBG and 0.5–1.3- μ m underfocus for β -gal, was used. We recorded 383 movies over 48 s for apoferritin (5 squares, 44–76 holes/square), 1,284 movies over 69 s for GroEL (8 squares, 67–201 holes/square), 1,115 movies over 77 s for WHBG (9 squares, 49–222 holes/square), and 718 movies over 78 s for β -gal (5 squares, 48–224 holes/square).

2.4 Image processing

The images were processed in RELION (104), where the frames of the movies were aligned and averaged using a Bayesian approach as described (105). The contrast transfer function (CTF) parameters were calculated with Gctf (106). Afterward, a subset of micrographs was used to pick ± 500 particles manually for initial 2D classification. These 2D classes were used for an iterative, automated particle-picking procedure where both the references and the autopicking parameters were improved using a subset of the micrographs. The complete data set was autopicked, particles were extracted and subjected to an iterative 2D classification scheme to reject bad particles. After 2D classification, 3D auto-refine was performed using starting models based from the PDB (4W1I, 2YNJ, 1X9F, 6CVM). The resulting reconstructions were low pass filtered and projected for another iteration of picking, 2D classification, and 3D auto-refine. Local CTF refinement, and local symmetry in the case of WHGB, resulted in the final maps. Given resolutions were estimated based on established, Fourier Shell Correlation (107) standards (108), as directed by the program RELION.

2.5 Model refinement

The PDB starting models mentioned above (4W1I, 2YNJ, 1X9F, 6CVM) were superimposed on the sharpened cryo-EM maps. Models were refined iteratively through rounds of manually adjustment in Coot (109), real space refinement in Phenix (110), and structure validation using MolProbity (111).

3. Results

3.1 Sample carriers

All results below were obtained with pre-mounted autogrids (112,113). Autogrids were initially developed for increased robustness in order to allow automated handling of grids in cryo-transmission electron microscopes (TEMs), such as the Titan Krios, Glacios, and Arctica (Thermo Fisher Scientific). Traditional vitrification devices are not able to vitrify pre-mounted (assembled at room temperature) autogrids. Therefore, autogrids are normally manually assembled under cryogenic conditions by clamping an EM grid into a sturdy cartridge by means of a flexible C-clip spring (so called post-mounting). The jet vitrification procedure described below overcomes this limitation.

3.2 Glow discharge module

Traditional vitrification devices use external glow discharge modules. We characterized this procedure by determining contact angles for typical glow discharge settings used in our lab. An ELMO Glow Discharge System (Cordouan Technologies) was operated at 7 mA, 0.35 mbar, and 30 s glow discharge time. Contact angles (measured using a Krüss drop shape analyzer, model DSA25) of $23 \pm 7^\circ$ ($n=6$) were obtained for minimal transfer time (1.5 min) between the glow discharge system and the drop shape analyser.

Within the VitroJet, the glow discharge module is an integral part of the device. Each grid is glow discharged separately, and transferred within 10 s from the glow discharge unit into the process chamber. We evaluated different glow discharge settings (0.3–20 mA; 0.05–0.2 mbar), using different grids (Quantifoil R2/2 Cu300, UltraAuFoil R2/2 Au300), different treatment times (5–40 s), and different durations between the glow discharge treatment and the contact angle measurement (1–17 min). Contact angles of $17 \pm 3^\circ$ ($n=5$) could be obtained from grids glow-discharged in the VitroJet. In general, longer glow discharge treatment gave lower contact angles. Not surprisingly (114), longer delays (17 min) between glow discharge and contact angle measurements gave larger contact angles (35°). The glow-discharged grids did not show any visible damage. The single particle data described below were obtained using grids that were glow discharged for 30 s at 0.5 mA and 0.1 mbar.

3.3 Process chamber

In order to minimize sample evaporation, current vitrification devices such as the VitroJet, Leica EM GP, and the Gatan Cryoplunge apply the sample inside a humidified chamber. Passmore (90) recommended to keep the relative humidity (RH) surrounding the specimen support at 100% to prevent changes in the solute concentration prior to freezing. However, it is difficult to achieve a reliable 100% RH. Once the air within the chamber is fully saturated with water vapour, condensation will occur. Water droplets on the humidity sensor and/or the grid compromise the reproducibility of the experiment: humidity sensors only work reliably under non-condensing conditions.

We first studied the theoretical and reported evaporation rates of thin layers of water on grids. Different evaporation velocities of thin (suitable for TEM) water layers have been reported, depending on temperature and RH: 5 nm/s at 4 °C and 90% RH (90) and 40–50 nm/s at 20 °C and 40% RH (97). We modelled the evaporation of a thin layer of water on a grid (Supplementary Note 1 in (115)). Parameters within this model include partial water pressure in air, saturated vapour pressure, velocity of the air flow over the sample, thermal conductivity of the sample carrier, and shape of the deposited water layer. For a given temperature and RH, the dew-point temperature of the grid can be calculated (**Figure 4A**). At <100% RH, the dew-point temperature of the grid will be lower than the temperature in the chamber. Our model predicts that the layer thickness decreases by 70 nm/s for every °C of dew-point error, for a thin line of pure water deposited onto the grid at 93% RH in the chamber (**Figure 4B; Supplementary Note 1** in (115)).

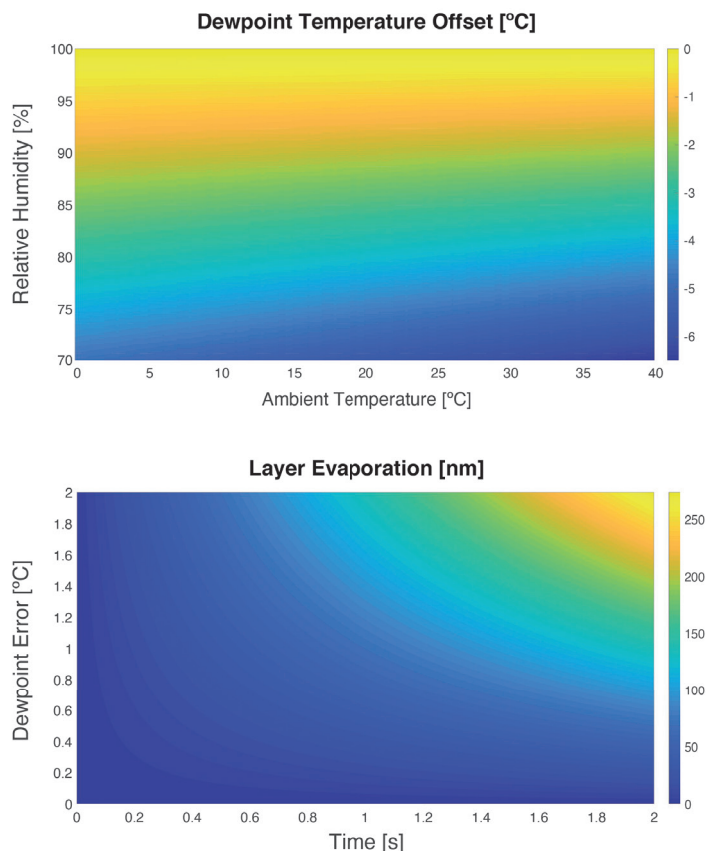


Figure 4. Dew-point temperature calculations and predictions. (A) Dew-point temperature offset (colour scale, in degrees) as a function of the relative humidity and ambient temperature. *E.g.*, when working at 20 °C and 94% relative humidity, the dew-point temperature would be 1 °C (orange) below the ambient temperature. (B) Predicted decrease in thickness of a thin sample layer due to evaporation (colour scale, in nm/s) with respect to the time and the dew-point error, which is the difference between sample temperature and its dew point. *E.g.*, for a temperature of 1 °C above dew point, 70 nm of the layer is estimated to evaporate within 1 s.

In order to prevent sample evaporation, the sample itself thus needs to be maintained at dew point. Based on the relationship shown in **Figure 4A**, we built a fast feedback loop to maintain the grid at dew-point temperature. The feedback loop, implemented in LabView, monitors the RH and temperature of the process chamber and calculates the dew-point temperature at a frequency of 100 Hz. Within the VitroJet, this chamber is maintained at 93% RH. The grid is Peltier-cooled through the autogrid cartridge. The optimal grid temperature was offset from its calculated dew point to account for heat exchange between the grid and its environment, resulting in an elevated temperature of the centre of the grid compared with the sensor positioned just below the grid. Even if the edge of the grid would be held perfectly at the dew-point temperature, the estimated temperature of the centre can be a few tenths of a degree higher (**Supplementary Note 1**

in (115)). Our theoretical model predicts that, if such temperature difference would not be accounted for, the layer thickness could still decrease by 14 nm/s. An incorporated camera enables visual inspection of evaporation and/or condensation on the grid and can be used to fine-tune the dew-point feedback loop (**Supplementary Movies 1 and 2** in (115)).

3.4 Pin printing

To improve sample deposition on EM grids, several developments have been presented, including droplet-based methods (116-120), *e.g.*, on nanowire grids (121), and using capillaries followed by sample thinning by controlled water evaporation (122). We sought a way to obtain sample thickness layers suitable for cryo-EM without the need for blotting, nanowire grids, and/or extra water evaporation steps. In our pin-printing method, a sub-nL volume of sample is deposited onto the grid using a pin. In order to determine the relative positions of the pin and the grid, we position them subsequently in the focal plane of the camera. Using these calibrated positions, the pin can be moved to a defined distance of the grid.

A stock sample volume of 0.5 μL is introduced into the process chamber by a positive displacement pipette. A metal pin is dipped into the sample to collect a sub-nanoliter volume. The pin is cooled down to dew point to prevent evaporation of the tiny droplet at its tip. The pin is moved to a predefined distance (typically 10 μm) from the carrier surface such that the sample forms a capillary bridge between the pin and grid. Once this bridge is formed, the pin is moved along the surface of the carrier. Capillary forces ensure that the liquid bridge follows the pin. If this parallel movement is sufficiently fast, a thin film will be deposited due to the viscous shearing on the liquid. The film thickness h is determined by the relative speed of the pin u moving over the carrier surface, the stand-off distance δ between pin and carrier, the viscosity μ , surface tension σ , and surface properties of the pin and the carrier (**Figure 5A**). For the data shown here, we used a velocity of 0.3 mm/s (**Supplementary Movie 3** in (115)). Moving at a higher speed should result in thicker sample layers and vice versa (114) whereas using a larger stand-off distance should result in a thicker layer. The pin diameter itself will influence the range of stand-off distances one can use and the area one can write. In summary, the pin-printing technique opens up a variety of possibilities to optimize sample layer thickness and counteract variations in fluid properties.

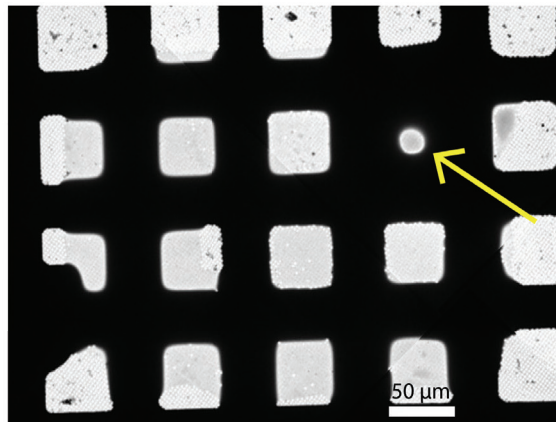
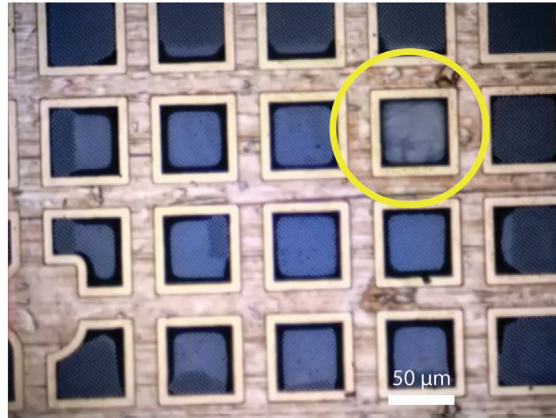
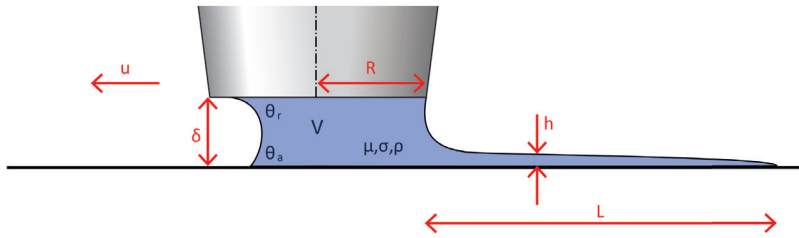


Figure 5. Pin-printing parameters and resulting sample deposition. (A) Schematic overview of relevant parameters for pin printing. The deposited layer in height (h) and length (L) is result of the sample properties (viscosity μ , surface tension σ , density ρ), contact angles (advancing at the grid θ_a , receding at the pin θ_r), and writing parameters (standoff distance δ , pin radius R , pin velocity u). (B) Photograph of the grid by the integrated camera shows the sample applied by pin printing. The pin is indicated by a yellow circle. (C) Low magnification cryo-EM atlas of the deposited and vitrified sample. One thicker square (arrow) is the result of retraction of the pin from the grid. Scale bars represent 50 μm .

While printing could be performed in any pattern, we found that printing in a partly overlapping spiral pattern (**Figure 5B and C**) resulted in a more uniform thickness compared with, for example, in one straight line across the carrier. Finally, the pin retracts from the surface and the remaining volume in the liquid bridge breaks up, leaving a thicker droplet at the end of the deposited line. Because only a sub-nanolitre volume is deposited per printing, numerous grids can be printed from the same stock volume of 0.5 μL .

Three factors could alter the film layer thickness between deposition and vitrification. These include liquid flow, evaporation, and condensation. Liquid flow is characterized by the redistribution time scale T_r for a surface-tension-driven thin-film flow (123). $T_r = \mu d^4 / \sigma b^3$, and is primarily dependent on d , the distance the fluid has to travel, and the thickness b of the sample layer. Redistribution occurs on two timescales: the first is very fast for thick layers of sample adhering to the grid bars, whereas the second is a much slower redistribution of liquid within the thin layer over the grid square surface. When the pin moves over a grid square, wicking by the grid bars takes place within milliseconds, resulting in an unusually thick layer just around the edges of the grid square which cannot be prevented in the deposition process. The second liquid redistribution is slower. For example, for a 50 nm thin layer of water (b), and $\mu \approx 10^{-3}$ Pas being the dynamic viscosity of the liquid, and $\sigma \approx 0.07/\text{m}$ being the surface tension of the liquid, the redistribution time scale would be >10 min.

Figure 5C shows a low magnification EM overview of the pin-printed area, whereas **Supplementary Figure 1** shows the ice distribution within a grid. Both figures show that the ice layer is somewhat thicker at the start of pin printing (middle of the atlas) compared with the end (outside of the squared spiral). Before the start of deposition, the pin approaches the grid and squeezes the droplet, leaving an excess of sample in the centre. This increased thickness is maintained due to time scale of liquid redistribution. Alteration of the deposited sample layer could still occur by evaporation or condensation. Both effects are mitigated by a tight control of the dew-point feedback loop described above. Overall, the sample layer thickness obtained directly after pin printing will be very similar to the vitrified one.

3.5 Vitrification module

After sample deposition, traditional vitrification devices rapidly plunge the bare grid into a bath of cryogenic liquid to vitrify the sample. Plunge vitrification starts cooling down from the bottom of the grid upward, and boiling of the coolant can form a gaseous insulating layer at the surface of the grid (101). This process can compromise the cooling time of the sample, which is estimated to be 10^{-4} s in the most favourable case of thin-layer vitrification (94). These current devices are not able to vitrify pre-mounted autogrids due to the extra thermal mass of the cartridge. The sturdy rim of autogrids would hit the cryogenic liquid first, whereas the area of interest (the sample on the middle of the grid) is cooled later at speeds too slow to prevent ice-crystal nucleation.

Inspired by cryofixation of thick tissues for room-temperature ultrastructural studies (124), we devised an alternative way to vitrify samples in pre-mounted autogrids: jet vitrification. Two streams of cryogenic liquid hit the sample and its carrier in the centre (**Supplementary Movie 4** in (115)), cooling it down to temperatures below 130 K in <1 ms. As jet vitrification cools the autogrid from the centre outward, the rim of the cartridge is cooled down last. The jets continue to spray for 50 ms to precool the cartridge; hereafter, the autogrid is submerged into a bath of liquid ethane to fully cool down the cartridge rim and gripper. The grid is subsequently slowly moved out of the ethane bath to allow excess ethane to flow off and prevent ethane solidification. Following vitrification, the gripper transfers the autogrid to a spring-loaded storage container in liquid nitrogen to connect directly to the microscope-loading workflow.

The preparation for jet vitrification is similar as for the Vitrobot. A cryochamber is precooled by liquid nitrogen before introducing the vitrification cryogen. At the start of the process ~13 mL of ethane or other cryogen (compared with the 6–7 mL used standard for a Vitrobot) is condensed in a precooled bath surrounded by a liquid nitrogen chamber. As the jetting requires <100 ms in total with ethane returning to the cryogen bath, ethane is recycled for the grids. Following the preparation of 8 grids, the ethane is typically refilled to maintain uniform vitrification conditions.

We conducted experiments with different cooling media at different temperatures, including liquid nitrogen (77 K), ethane/propane (37%/63% (v/v) at 79 and 93 K), and ethane (99 K). Of these different media, liquid nitrogen and liquid ethane/propane cooled the grid to the lowest final temperature, but the highest cooling rates were obtained with liquid ethane (**Supplementary Figure 2**). These cooling rates were measured with 25 μm constantan wires (Bare thermocouple wire, Omega, Norwalk) woven into a copper grid mesh to form a thermocouple: such wires are expected to cool much slower than the very thin sample. The centre of the grid showed significantly higher cooling rates with jetting compared with plunging, both for autogrids as well as regular grids (**Supplementary Figure 2**).

3.6 Integration

The above steps (glow discharge, process chamber, pin-printing, and jet vitrification) were implemented into the VitroJet (*Figure 6A*). A supply cassette provides up to 12 pre-clipped autogrids. A gripper picks up each autogrid individually and transports it sequentially through each of the different steps for sample preparation (**Figure 6B**; **Supplementary Movie 5** in (115)). The gripper is dried by nitrogen gas within the glow discharge unit before picking up the next carrier.

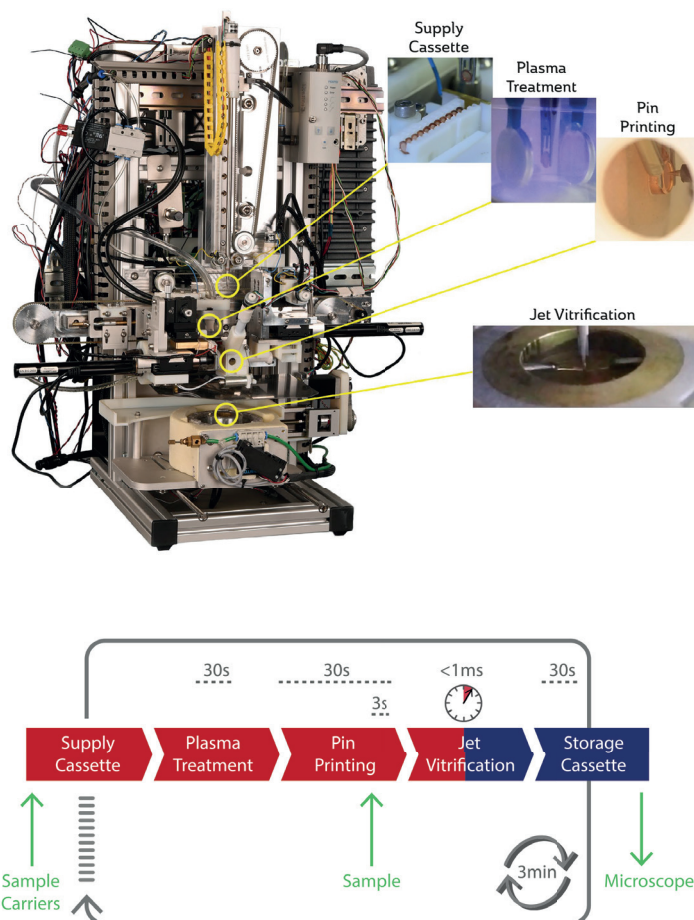


Figure 6. The VitroJet workflow. (A) The device and (B) the workflow. Sample carrier (up to 12) are introduced in a supply cassette and sample (up to 0.5 μL) in a pipette tip (yellow arrows). The VitroJet automatically processes the cassette by sequentially passing sample carriers through the plasma treatment, sample deposition by pin printing, jet vitrification and finally storing the grid in a cryogenic cassette ready to enter the electron microscope (green arrow). The supply cassette is at room temperature (red), the storage cassette at cryogenic temperature (blue). An animation showing all subsequent steps is shown in **Supplementary Movie 5** (115).

The cycle time of the workflow is ~3 min. Most of this time is taken up by the glow discharge module: evacuation of this chamber takes ~20 s (twice per cycle), while the glow discharge itself takes 30 s. For pin printing, calibration of the sample carrier and pin position within the process chamber requires <30 s, and the sample deposition a few seconds. Transfer of the grid between the process chamber and the vitrification chamber is completed within 80 ms. Vitrification of the sample lasts <1 ms, whereas another 300 ms is used to deeply cool down the entire autogrid assembly. Grid removal from the ethane bath and transfer requires <30 s.

Before each use, the VitroJet undergoes several automatic preparatory tasks, which are completed in ~15 min and includes cooling down the vitrification unit, filling the liquid ethane reservoir, and equilibrating the process module at the right humidity. The temperature-controlled liquid ethane reservoir is cooled by a liquid nitrogen bath.

3.7 Structure determination by single-particle analysis (SPA)

To validate the VitroJet, we prepared samples of several standard proteins (apoferritin, GroEL, worm hemoglobin, beta-galactosidase) and performed high-resolution SPA (see Methods). Each sample was pin-printed on ~16 squares of 300-mesh grids with perforated foils (R1.2/R1.3) and jet-vitrified using liquid ethane. Atlas overviews collected within the cryo-EM show excellent correlation with the visualization of the deposition within the VitroJet just prior to vitrification (**Figure 5B and C**). Cryo-EM data were collected on a 200-kV FEI Arctica microscope. From the squares that were pin-printed, most holes could be selected for data collection. Holes close to the grid bar were skipped to avoid thicker ice due to the wicking of the grid bars. Micrographs were recorded between 0.5 and 2.0 μm underfocus, showing good contrast (**Figure 7**).

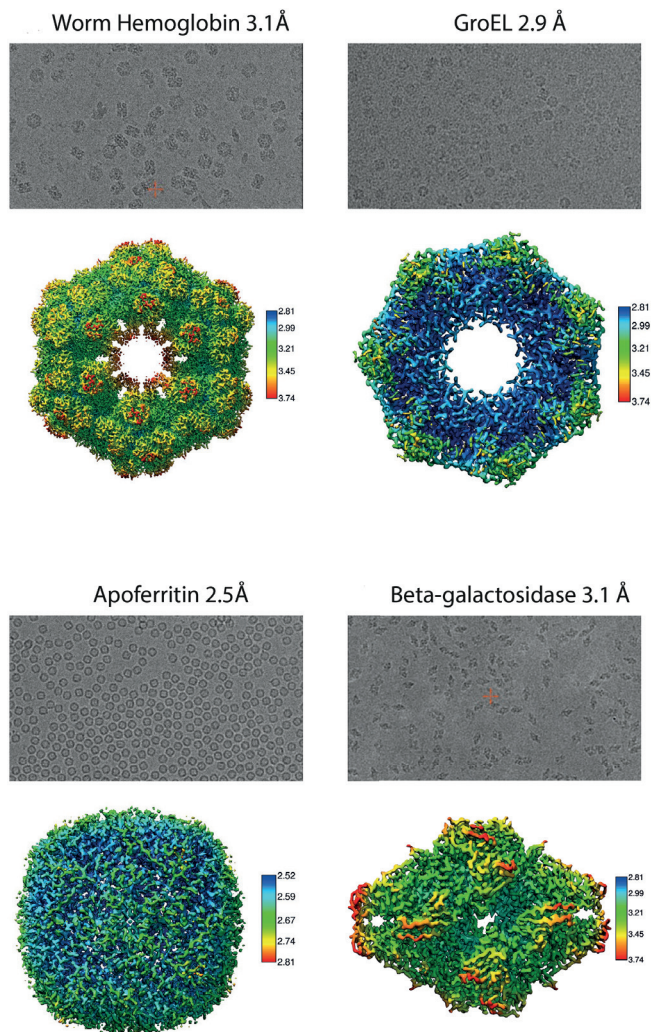


Figure 7. Sample preparation of four standard proteins by the VitroJet, and validated by cryo-EM analysis. Micrographs (top) and reconstructions (bottom) of worm haemoglobin, GroEL, apoferritin, and beta-galactosidase. The reconstructions are coloured according to local resolution as calculated using RELION.

The data, processed with RELION (16), yielded good 2D classes. After iterations both with particle picking and extraction, (local) contrast transfer function estimation, and 2D and 3D classifications, 3D reconstructions of $\sim 3\text{-\AA}$ resolution were obtained for each of these four proteins (**Figure 7**). Apoferritin gave a 3D reconstruction of 2.49-\AA resolution using 47,209 particles (O symmetry). For GroEL, we obtained a 3D reconstruction of 2.94-\AA resolution using 9809 particles (D7 symmetry). 3D reconstructions of worm hemoglobin and beta-galactosidase had respective resolutions of 3.11-\AA resolution (using 10,488 particles, D6 symmetry) and 3.11-\AA resolution (using 15,252 particles, C7 symmetry). Models refined against the four reconstructions were in accordance with earlier published models demonstrating the success of the VitroJet in preparing suitable samples for cryo-EM studies.

4. Discussion

Traditional cryo-EM sample preparation methods require multiple manual steps and depend on ill-controlled parameters such as blotting force, grid positioning, and time between glow discharge and sample application. Here, we present a workflow that minimizes operator dependency and provides control over relevant parameters. Prior to starting the sample preparation cycle, parameters regarding glow discharge, dew point, pin printing and jetting can be set. After initiation, the process is executed in an automated fashion. In an era where automation has made so much impact on microscope alignment, data collection, processing, and model building, automated control over sample preparation is a mandatory next step.

Samples applied to pre-mounted autogrids are difficult to blot and even more difficult to vitrify using the existing leading commercial devices or even the next-generation blotless commercial device Chameleon (125). We overcame both problems by using pin printing, which does not require blotting, and jet vitrification, which yields superior cooling rates starting from the centre of the grid where the sample is located. In addition to circumventing the problems associated with blotting, pin printing requires minute volumes of sample, which enables the study of macromolecules for which only micrograms can be obtained.

Proteins tend to absorb to the air–water interface where they can denature (23,96,126). It seems intuitive that reducing the time the sample is exposed to such an interface would help to prevent protein denaturation (100). However, using the Stokes–Einstein equation, it has been calculated that even for a minimal residence time of ~ 1 ms, particle–surface interactions will occur dozens of times before the water is frozen (127,128). Unfortunately, all existing devices have residence times varying between 11 ms (120) and several seconds and therefore cannot prevent particle–surface interactions. For the VitroJet, the residence time varies between seconds for the first written square down to 80 ms for the last square written prior to vitrification. A promising complementary

approach would be to control the surfaces present in the specimen support (127,129), as demonstrated for yeast fatty acid synthase applied to a substrate of hydrophilised graphene (96). We tested the device with different types of grids (continuous carbon, graphene oxide, different hole-sizes, UltraAuFoil) and found the pin printing procedure to be compatible with a multitude of (modified) grid supports.

Jet vitrification was originally demonstrated 40 years ago, on thick tissue spanned over a holder (124). The method was used as a prelude to freeze substitution, resin embedding and sectioning, for ultrastructural studies performed at room temperature. Sample evaporation was not a concern for the bulky tissues used. Here, we adapted jet vitrification to obtain thin layers of macromolecular samples. One might expect that sample evaporation problems would be insurmountable and that the jets would blow away the thin liquid layers of macromolecules prior to vitrification, resulting in empty holes. However, we have demonstrated the opposite. We believe that the sample already vitrifies before the liquid cryogen hits the sample (130).

The setup of the VitroJet is modular, making it easier to incorporate future developments to further advance the cryo-EM field. The pin printing applies sample to a part of the available area of the grid. Future generations of the VitroJet could be equipped with multiple pins moving simultaneously over different parts of the grids. Such schemes would enable higher-throughput screening as well as time-resolved studies (131,132), *e.g.*, combined with laser excitation. The process chamber provides the ability to control condensation as well as evaporation for an extended duration of time on a specific layer thickness, which could also offer benefits for soft condensed matter studies. While the VitroJet described here was developed for SPA, we aim to develop a branch of the VitroJet dedicated to the preparation and vitrification of cellular samples. Vitrification of cells is inherently more difficult than that of purified macromolecular samples. For example, it was stated that the centre of HeLa cells clearly undergoes incomplete vitrification (133). Preliminary results indicate that jet vitrification will help to reduce this problem, which would be a true asset for in situ structural biology. The VitroJet offers much-needed innovations in sample preparation, which will accelerate and perhaps even revolutionize future cryo-EM studies.

Data availability

The source data underlying **Supplementary Figure 1B** and **Supplementary Figure 2** are provided as a Source data file in the online version of the manuscript. Other data are available from the corresponding authors upon reasonable request.

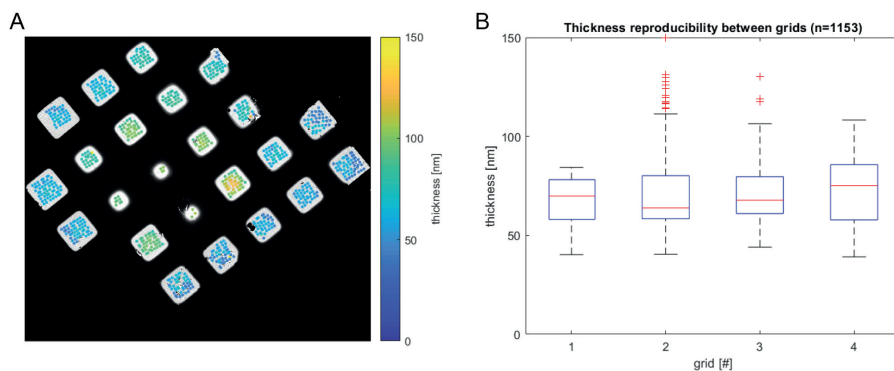
Author contribution

P. J. P., C. L-I, and R. B. G. R. designed and directed the project; F. J. T. N. designed and constructed the machine with input from all authors and B. W. A. M. M. B. as project leader; S. T. did preliminary tests; R. J. M. H. and G. W. performed the experiments; A. G. prepared the samples; R. B. G. R., R. J. M. H., and G. W. analyses the data and wrote the paper with input from all authors.

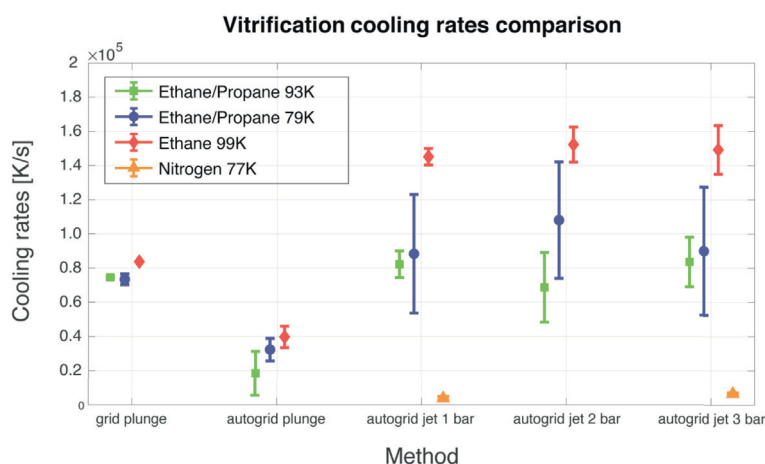
Acknowledgments and funding

We thank Dr. Fei Sun (**Institute** of Biophysics, Chinese Academy of Sciences) for providing apoferritin sample, Pascal Huysmans and Paul Kwant (IDEE, Maastricht) for engineering input, Peter Frederik for helpful discussions and Hang Nguyen for critical reading of the paper. Hans Duimel and Hirotooshi Furusho provided technical support from the UM Microscopy CORE Lab, Giancarlo Tria helped with initial experiments, Paul van Schayck with the IT infrastructure, and Roger Jeurissen with the theoretical framework. This research received funding from the Netherlands Organization for Scientific Research (NWO) in the framework of the Fund New Chemical Innovations, numbers 731.014.109 and 731.016.407, as well as from the Province of Limburg, the Netherlands.

Supplementary material



Supplementary Figure 1. Ice thickness distribution within a grid and the reproducibility of four grids prepared during the same session. Thickness measurements were determined using intensity comparisons as described by (134). (A) Ice thickness comparison over the pin printed area of a grid. Thickness ranges from 0 (dark blue) to 150 nm (yellow) over various holes measured on the grid-squares analysed. Deposition followed a square spiral pattern starting at the centre of the atlas image and spiralling outward. (B) Four grids with the same deposition parameters were prepared in sequential order. The box and whiskers plot describes the distribution of ice thickness for every grid has been made. The ends box represent the lower and upper quartile with the median shown as a red line in the box. The whiskers and dashed lines represent the range of the data. Outlier points are shown as red crosses. Source data are provided as a Source Data file in the online version of the manuscript.



Supplementary Figure 2. Cooling rates of sample carriers by the different vitrification methods. Measurement of average rate (K/s) at which the thermocouple is cooled down from 273 K to 123 K. The standard deviations are plotted as error bars. The thermocouple consists of constantan wire woven in a copper grid, which adds extra thermal mass to the grid and therefore influences the measurement. As the biological samples on the thin films are located away from the grid bars, this measurement is used as surrogate marker for the different cooling methods. Ethane jet vitrification results in significant higher cooling rates compared to ethane plunge vitrification. Similar cooling rates were found regardless of the jet vitrification pressure (1–3 bar). For all methods, liquid ethane (red diamonds) showed the highest cooling rates compared to their mixtures with propane (green and blue symbols) or to liquid nitrogen alone (orange). Source data are provided as a Source Data file in the online version of the manuscript.

3

CHAPTER 3

Mycobacterium tuberculosis ferritin: a suitable workhorse protein for cryo-EM development

Abril Gijssbers*, Yue Zhang*, Ye Gao, Peter J. Peters and Raimond B. G. Ravelli

*These authors contributed equally

Abstract

The use of cryo-EM continues to expand worldwide and calls for good-quality standard proteins with simple protocols for their production. Here, a straight forward expression and purification protocol is presented that provides an apoferritin, bacterioferritin B (BfrB), from *Mycobacterium tuberculosis* with high yield and purity. A 2.12 Å resolution cryo-EM structure of BfrB is reported, showing the typical cage-like oligomer constituting of 24 monomers related by 432 symmetry. However, it also contains a unique C-terminal extension (164–181), which loops into the cage region of the shell and provides extra stability to the protein. Part of this region was ambiguous in previous crystal structures but could be built within the cryo-EM map. These findings and this protocol could serve the growing cryo-EM community in characterising and pushing the limits of their electron microscopes and workflows.

Keywords

Cryo-EM, ferritin, *Mycobacterium tuberculosis*, single-particle analysis, expression and purification protocols.

I. Introduction

Single particle (SPA) cryogenic electron microscopy (cryo-EM) has become an indispensable tool for structural biology. The combination of direct electron detectors, motion correction (14,15), high-end electron microscopes and advanced imaging process algorithms, allowed for the “Resolution Revolution” (87) from whereon an increasing number of SPA structures with resolution below 2 Å could have been determined (135-138). Hardware improvements such as a monochromator, spherical aberration (C_s) corrector, energy filter and new generation of direct electron detector improved the signal-to-noise ratio even further and has brought SPA to real atomic resolution (19,20). The success of SPA attracted many new scientists into the field of cryo-EM and led to an impressive growth of its community as well as the number of instruments that can deliver high-resolution cryo-EM structures. Ferritin is often used as a benchmark to commission and validate these machines, as it is relatively straightforward to get good-resolution datasets of its 24-mers. In fact, as of March 2021, 11 out of 25 SPA structures deposited in the Protein Data Bank (PDB) with resolution better than 2 Å are ferritins, from different organisms. Ferritin has thus become a gold standard for cryo-electron microscopists to evaluate their setups and for pushing the limits of sample preparation (25,115,117,139), imaging (19,20) and data processing (17,21,135).

Ferritins are protein complexes involved in iron homeostasis and DNA repair by storing it in cage-like structures. This protein complex is formed by 12 or 24 subunits with tetrahedral 23-symmetric or octahedral 432-symmetric arrangements, depending on which sub-family it belongs. There are three sub-families of ferritin: iron-containing ferritin and heme-containing bacterioferritin involved in the iron storage, and DNA-binding protein in starved cells (Dps) that uses iron to protect DNA from free radical-mediated damage (140,141). Iron is a trace element vital for all living organisms, and although it is the second most common metal on Earth, it is not always bioavailable: for this reason organisms compete for the iron in the environment (142). The acquisition of iron is considered a key step in the development of any pathogen, and so, host cells have developed mechanisms to sequester the metal from infecting bacteria in an attempt to defend themselves (143). One of such pathogens is *Mycobacterium tuberculosis*, the causative agent of human tuberculosis.

In this study, a simple high-yield purification protocol is presented for bacterioferritin B (BrfB) of *M. tuberculosis*, together with its high-resolution cryo-EM structure. Previously, only lower resolution crystal structures had been reported. Our structure shows extra residues at both termini. The C-terminal tail of BrfB, being a target peptide for encapsulation (144) and playing a role in ferroxidase activity and iron release in addition to providing stability to the protein (145), loops into the cage region of the complex, prior to folding back via the three-fold channel to the B-pore. Two conformations of the C-terminal residue His175 were observed near the B-pore, suggesting a role of this residue in iron exchange. BrfB could serve the structural community in testing their

expanding fleet of equipment as well as aid to a better understanding of iron storage proteins, which are essential for the survival and progression of important human pathogens like *M. tuberculosis*.

2. Materials and methods

2.1 Expression and purification of BfrB

Codon-optimised gene of *M. tuberculosis* BfrB was cloned into a modified pRSET backbone (Eurofins Genomics) using *Nde*I and *Hind*III restriction sites to over-express in C41(DE3) *Escherichia coli*. Expression and purification protocol was adapted from Khare et al. (145) and Parida et al. (146). In summary, primary culture of lysogeny broth (LB) medium supplemented with 100 µg/mL ampicillin was prepared from a single colony at 37 °C and 200 rpm overnight, to then inoculate 500 mL fresh LB medium with 1:1000 dilution of the primary culture in the same conditions until reaching an optical density (OD₆₀₀) of 0.6. Protein over-expression was induced by adding isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM, for 3 h without changing the culture conditions. Cells were harvested by centrifugation and stored at -20 °C until further use.

Bacterial cells were resuspended in 20 mM Tris-HCl pH 8.0, 300 mM NaCl supplemented with 2 units/mL of benzonase (Sigma-Aldrich), EDTA-free protease inhibitor (Sigma-Aldrich) and 2 mM β-mercaptoethanol (BME). The cell suspension was lysed by sonication. Cellular debris was removed by centrifugation at 30,000 × g for 30 min. Supernatant was subjected to a saturation of 20% ammonium sulfate and incubated at 5 °C for 1 h with constant shaking before centrifugation at 15,000 × g for 20 min. The pellet was resuspended in 20 mM Tris-HCl pH 8.0 and 150 mM NaCl, and centrifuged at 10,000 × g for 5 min to remove precipitants. Sample was further purified by size exclusion chromatography with a Superdex® 200 Increase 10/30 column (GE Healthcare). Fractions were collected and protein purity was evaluated by denaturing polyacrylamide gel electrophoresis. Fractions containing BfrB were pooled together and stored at -80 °C until further use. The final yield was >100 mg for 1 L of culture.

2.2 Cryo-electron microscopy sample preparation, data acquisition and image processing

Purified BfrB was first used at a concentration of 11 mg/mL (calculated with Pierce™ BCA Protein Assay Kit). A volume of 2.5 µL was applied on glow discharged UltrAuFoil Au300 R1.2/1.3 grids (Quantifoil). Excess liquid was removed by blotting for 3 s (blot force 5) using filter paper followed by plunge freezing in liquid ethane using a FEI Vitrobot Mark IV operated under 100% humidity at 4 °C. Cryo-EM single particle data were collected on a Titan Krios at 300 kV with a BioQuantum K3 Imaging Filter with a 20 eV post-column energy filter. The detector was utilised in super-resolution counting mode at a nominal magnification of 130,000×. **Table 1** shows the statistics

of the data set. Data were processed using the RELION pipeline (16). Movie stacks were corrected for drift (5×5 patches) and dose-weighted using MotionCor2 (147). The local contrast transfer function (CTF) parameters were determined for the drift-corrected micrographs using Gctf (106). A first set of 2D-references were generated from manually picked particles in RELION (16) and these were then used for subsequent automatic particle picking. **Table 1** lists the number of particles in the final data set after particles picking, 2D-classification and 3D classification. The latter was with O symmetry. Beamtilt parameters, anisotropic magnification and local CTF parameters were refined and particles were polished (135). The resolution of the final full map, listed in **Table 1** and shown in **Figure 8**, was 2.12 Å using the gold-standard FSC=0.143 criterion (108). A B-factor plot according to Rosenthal and Henderson (148) was calculated using random subsets of the data with variable number of particles (**Supplementary Figure 3**).

Table 1 also includes the statistics of a data set collected in-house on a Tecnai Arctica microscope, operating at 200 kV, with Falcon 3 camera operated in electron counting mode (no energy filter). For this data set, we used a highly concentrated sample of 80 mg/mL, which gave beautiful micrographs with densely packed monolayers of BfrB in the middle of the holes of the UltrAuFoil Au300 R1.2/1.3 grids (**Figure 8A**). This dataset contained only a third of the number of micrographs compared to the K3 dataset, however, since it was collected on a more concentrated sample with a larger pixel size (0.651 versus 0.935 Å), similar number of particles were obtained for both datasets. The resolution of the 200 kV Falcon 3 data set was 2.39 Å.

2.3 Structure determination and refinement

We used the PDB model 3QD8 (145) as a starting model in Coot (149) for manual docking and building. The final model was refined against a sharpened cryo-EM map obtained by LocSpiral (150). The model was refined iteratively through rounds of manually adjustment in Coot (109), real space refinement in Phenix (110) and structure validation using MolProbity (111).

Chapter 3

Table 1. Data collection statistics.

Sample	BfrB <i>M. tuberculosis</i>	
Concentration (mg/mL)	11	80
Grid type	Quantifoil UltraAuFoil 300 mesh R1.2/1.3	Quantifoil UltraAuFoil 300 mesh R1.2/1.3
Plunge freezer	Vitrobot	Vitrobot
Microscope	Krios	Tecnai Arctica
Voltage (kV)	300	200
Energy filter (eV)	20	None
Camera	K3	Falcon 3
Detector Mode	Super-resolution counting	Electron Counting
Nominal magnification (k \times)	130	110
Physical pixel size (\AA)	0.6514	0.935
Exposure time (s)	1.3	46.33
Fluence ($e^- \text{\AA}^{-2}$)	40	41
Focus range (microns)	-0.4, -0.6, -0.7, -0.8, -0.9, -1.0, -1.2, -1.4	-0.75, -1.0, -1.25
Micrographs	2518	875
Number of fractions	50	50
Particles	163568	186025
Symmetry imposed	O	O
Average resolution (\AA)	2.13	2.39
FSC threshold	0.143	0.143
Map sharpening B factor (\AA^2)	-68	-109
Refinement		
Initial model used (PDB entry)	3qd8	
Model resolution against LocSpiral (\AA)	1.89	
FSC threshold	0.5	
Model resolution against RELION map (\AA)	2.18	
FSC threshold	0.5	
Model composition monomer		
Atoms	1471	
Hydrogen atoms	0	
Protein residues	177	
Waters	58	
B factors (\AA^2)		
Protein	28.33	
Water	30.79	
R.m.s. deviations		
Bond lengths (\AA)	0.017	
Bond angles ($^\circ$)	1.324	
Correlation coefficients		
Mask	0.88	
Box	0.87	
Validation		
<i>MolProbity</i> score	1.20	
Clashscore	3.49	
Poor rotamers (%)	0	

Sample	BfrB <i>M. tuberculosis</i>
Ramachandran plot	
Favored (%)	97.71
Allowed (%)	2.29
Disallowed (%)	0

3. Results and discussion

3.I Mycobacterial ferritin structure

We studied the structure of BfrB by cryo-EM and obtained a 2.12 Å-resolution density map based on gold-standard FSC (108) using 2,518 micrographs with 163,568 particles. The high-quality data allowed observation of secondary-structure features during 2D classification (**Figure 8**). The clear density for side chains and holes in aromatic rings illustrates the quality of the EM map (**Figure 9**). Our final model contains 1,392 water molecules, whereas the BfrB model built from the 3 Å X-ray crystallography map had 360 water molecules (145) and the 1.15 Å resolution cryo-EM structure of human apoferritin had 4,622 water molecules (20).

Previous models of *M. tuberculosis* BfrB have been determined by X-ray crystallography (PDB entries 3QD8 at 3 Å resolution, 3OJ5 at 2.85 Å resolution, and 3UNO at 2.5 Å resolution (145)(TB Structural Genomics Consortium, unpublished work)). We built our model in an enhanced cryo-EM map, a *LocSpiral* map, calculated using algorithms based on spiral phase transformation (**Figure 9**) (150). **Supplementary Figure 5A** shows a comparison between the RELION postprocess map and the *LocSpiral* sharpened map, revealing some extra features in the latter. To confirm these, we also calculated the *LocOccupancy* maps, which estimate the density occupancy (**Supplementary Figure 5B**) (150). The resolution of both maps was also estimated by comparing the FSC between the refined model and the map (110) at a cut-off value of 0.5 (**Supplementary Figure 4**): this is 1.89 and 2.18 Å when compared with the *LocSpiral* and RELION post-processing maps, respectively (**Table 1**).

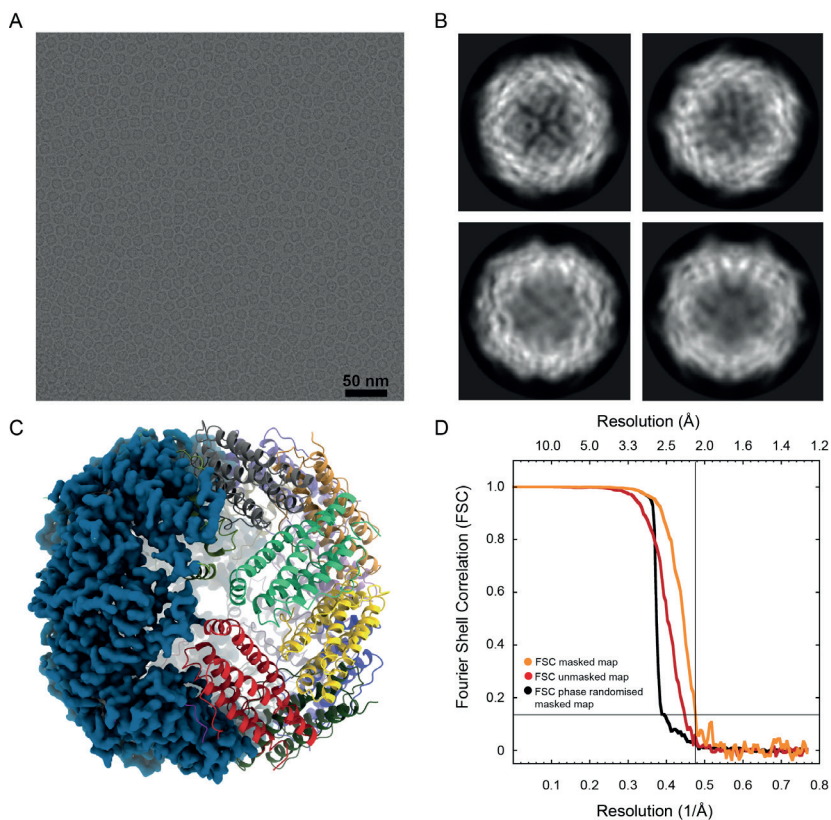


Figure 8. Single-particle analysis of BfrB. (A) Micrograph of highly concentrated (80 mg/mL) BfrB sample in vitreous ice, collected on a Falcon3 at 200 kV. (B) 2D class averages, the size of the shown box is 150 Å. (C) 3D reconstruction from 163,568 particles at 2.12 Å resolution, collected on a K3 at 300 kV. (D) Gold-standard Fourier shell correlation (FSC) before (red line) and after (orange line) masking, and the phase randomized FSC (black line).

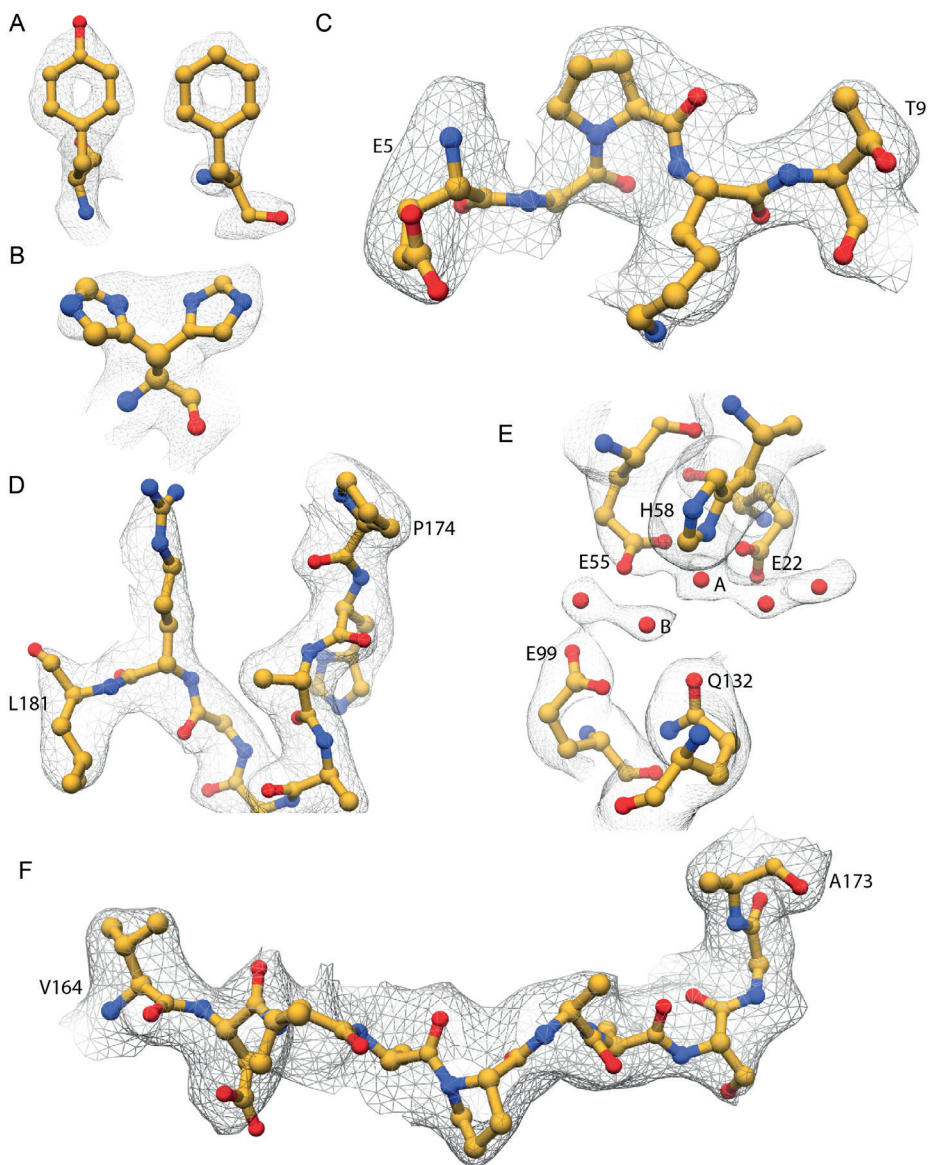


Figure 9. Representative regions of the density. (A) Post-processing map density for Phe23 and Tyr35. Local resolution scaled map of (B) His175, (C) Glu5-Thr9, (D) Crigid (Pro174-Leu181). (E) A string of density blobs near the ferroxidase sites, (F) Cflex (Val164-Ala173). The EM density is shown in gray mesh, the residue atoms are represented in ball-stick model.

At the N-terminus, we could add residues Glu5–Thr9, for which clear density was seen in the enhanced EM map (**Figure 9C**), but which were absent in our starting model (PDB entry 3QD8). The C-terminal region 164–181 consists of a flexible part (C_{flex} , 164–173) and a rigid part (C_{rigid} , 174–181), as described by Khare *et al.* (145). C_{flex} is ill-defined within the X-ray maps. The C-terminal region has been shown to be important for protein stabilisation and iron uptake (145), and thus confident assignment of these residues could aid our understanding of BfrB function. The extension of the C-terminal end in *M. tuberculosis* is unusual for ferritin even compared to heme-containing BfrA from the same organism. This extension has been shown to play an essential role in its function: BfrB exhibits a 3.5-fold reduction in oxidation rate of Fe(II) and a 20% reduction in the Fe(III) release rate upon removal of the C-terminal end (145). We found the extension of the C-terminal end to be located within the interior of the BfrB cage, which is remarkable as Mt-enc has been reported to encapsulate BfrB via this C-terminal extension (144). Our EM map displays well defined density for C_{rigid} (**Figure 9D**), residues 174–181, including a double conformation of residue His175 (**Figure 9B**). Similar to as in *S. coelicolor* bacterioferritin, Fe(II) enters BfrB from the B-pore and is converted to Fe(III) at the ferroxidase centres (145,151,152). We hypothesise that the double conformation of His175 might be relevant for the iron exchange of the protein, as this residue is located at the interior of the B-pore and in the vicinity of a cavity (**Figure 10** and **Supplementary Figure 6**). This cavity leads towards the ferroxidase sites A (Glu22, Glu55 and His58) and B (Glu55, Glu99 and Glu135) where the ferrous iron is oxidised by molecular oxygen (145,146). Previously deposited X-ray diffraction electron-density maps did not place iron ions at these sites. The 2.5 Å resolution map for PDB entry 3UNO only showed good density for the B site and weak density for the A site: water molecules were placed at both positions. The highly conserved Gln132 might favour iron binding at site B. The 3.0 Å resolution map for PDB entry 3QD8 showed some unmodelled density at the B site and no density for the A site. Finally, the 2.85 Å resolution map for PDB entry 3OJ5 did not show clear density for either the A or the B site. In our EM map, a string consisting of 5 density blobs could be seen (**Figure 9E**). We placed water molecules in here, as we lack experimental evidence for these water molecules being ions. Both the A and B site have clear densities, where the A site is coordinated by Glu22 (2.55 Å), Glu55 (2.78 and 2.82 Å), His58 (2.51 Å) and neighbouring water molecules (2.44 and 2.65 Å). The B site is coordinated by Glu55 (2.77 Å), Glu99 (3.03 Å), Gln132 (2.55 Å) and another water molecule (2.55 Å). Glu135 has a different conformation compared to the X-ray map and does not coordinate to the A or B site. The coordination distances are longer than those most commonly found for iron ions according to *MetalPDB* (153): 2.031–2.236 Å for Fe-N and 2.077–2.414 Å for Fe-O. Another string of multiple water-like densities was found near Asp37, Pro42, Lys46 and Ser50.

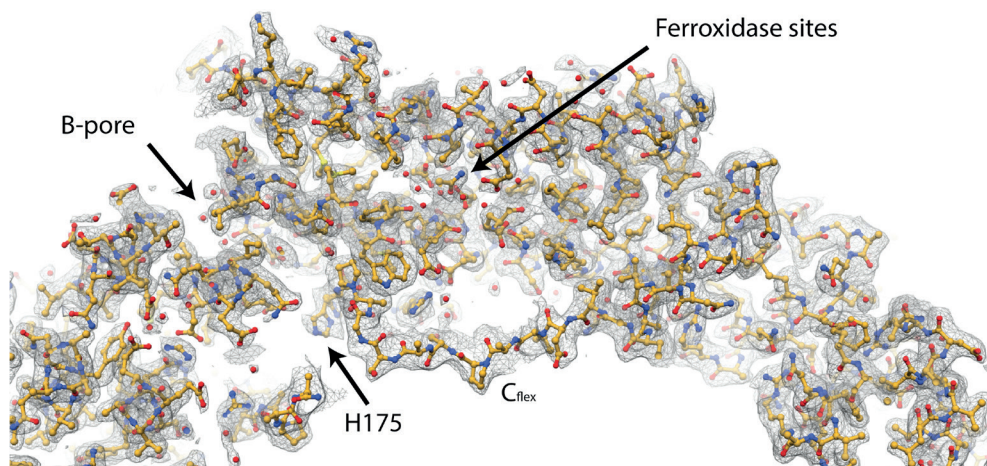


Figure 10. Density map and model for C_{flex}, which extends into the interior of the cage and is located above the cavity between the B-pore and the ferroxidase centres.

Our map reveals density for C_{flex}, which extends into the interior of the cage and is located above the cavity between the B-pore and the ferroxidase centres (**Figure 10**). The density for C_{flex} is less defined than other areas of the molecule (**Figure 9F** and **Supplementary Figure 5**), which would reflect flexibility and a possible functional role of this (154). *LocSpiral* (150) helped to further improve this part of the density.

3.2 Characteristics suitable for a good standard protein for cryo-EM

Ferritin has become more and more popular to serve as a standard protein for testing and training purposes in microscope laboratories. The good solubility and stability of most ferritins and their good intermediate size (12–13 nm diameter), combined with a 432-point group symmetry, allow swift characterisation of the performance of the microscope using a minimal number of micrographs. To this date, the most commonly used samples are the commercially available horse spleen ferritin and self-produced mouse ferritin (155) and human ferritin (20). However, samples and protocols are not always optimal. The most affordable sample is horse spleen ferritin, but unfortunately it seldom provides the sample quality needed to push the resolution, possibly due to the presence of broken particles that introduces sample heterogeneity. Khare *et al.* (145) compared the stability of ferritin subunits from different organisms and reported that horse ferritin is one of the least stable, together with mouse ferritin. The highest resolution achieved by cryo-EM is on human ferritin (20); however, the procedure described is an extensive protocol with more than ten steps that include the precipitation of nucleic acids, two ammonium sulfate precipitations, two sucrose gradients, two 24 h dialyses and ion-exchange chromatography. Although the final yield was not reported, it is expected that some protein will be lost with every step of the purification, resulting in high purity but

low yield. The length and the use of so many different steps is inconvenient if we aim for a protocol that can be reproduced worldwide, as it will require infrastructure and expertise that might not be available everywhere. Previously, we have tried expression and purification protocols of constructs from different organisms, leading to yields that were, in our hands, not sufficient for the extensive and routine use of the proteins with classical grid-preparation techniques such as the Vitrobot.

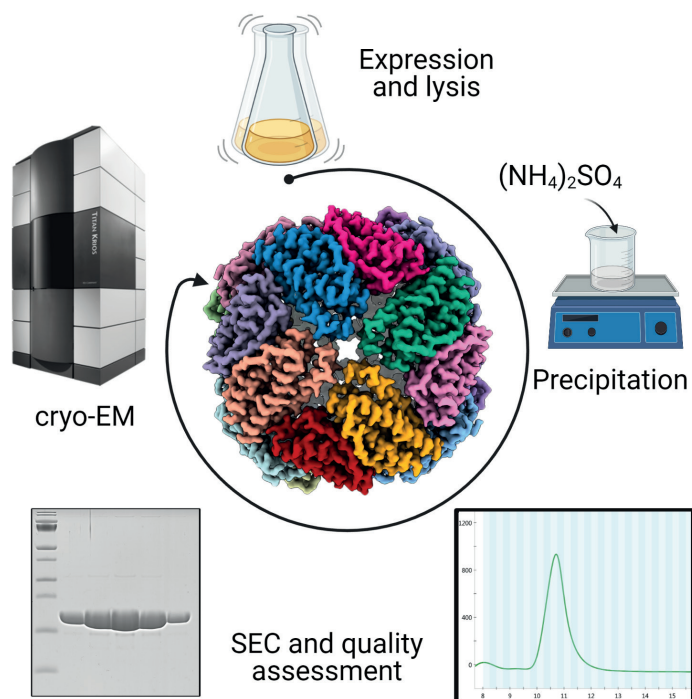


Figure 11. Simple purification workflow of BfrB protein. After protein expression, *E. coli* cells are lysed by sonication and cellular debris removed by centrifugation. Protein sample is precipitated by ammonium sulfate, and resuspended to be further purified by size exclusion chromatography. High yield of highly pure protein is obtained, suitable for cryo-EM studies.

In here, we have provided an easy and cheap method to purify the mycobacterial BfrB recombinantly expressed in *E. coli* C41 strain. The expression and purification protocol was adapted from Khare *et al.* (145) to increase the yield from ~17.5 mg to >100 mg of protein per litre of culture. Given the good expression, we did not design a tag for the BfrB plasmid, thereby overcoming the need for a tag-removal step. One ammonium sulfate precipitation step followed by size-exclusion chromatography was sufficient to obtain samples of high levels of purity, as confirmed by SDS-PAGE gel (Figure 11 and Supplementary Figure 7) as well as cryo-EM (Figure 8). Protein

concentration up to 115 mg/mL (calculated from the theoretical extinction coefficient) was achieved without any obvious precipitation, and no monomers were observed in the chromatogram, suggesting high solubility and stability of purified BfrB oligomer. The 2.12 Å resolution structure presented here was obtained from one 11 mg/mL aliquot stored at -80 °C, indicating that the protein is resistant to at least one freeze-thaw cycle. We prepared grids up to a concentration of 115 mg/mL, and observed a dense packing of non-overlapping BfrB particles at 80 mg/mL (**Figure 8A**), which is remarkably high but reproducible for our local grid-preparation system. The B-factor plot according to Rosenthal and Henderson (148) shows that we obtained resolutions better than 4, 3 and 2.5 Å for 100, 1000 and 10000 particles, respectively (**Supplementary Figure 3**). Although the fitted B-factor of 75 Å² of our reported data set is too high to expect record-breaking resolutions, it is a perfectly adequate sample to obtain 2.5 Å resolution with modest microscopes and settings (**Table 1**). Whereas the rigid part of BfrB can serve in characterising the performance of the microscope at regular times, the flexible C-terminus could also help to train users in some of the more advanced image data-processing SPA steps such as particle subtraction and focused classification. Plasmids are available upon request.

Over the last decade, there has been a fast development in every aspect of cryo-EM. Recent publications have proven that this technique has the potential to visualise macromolecules at atomic level (19,20). However, every protein has different physicochemical properties that makes it unique and that could challenge its structural elucidation by cryo-EM. For this reason, further improvements in cryo-EM are still necessary and thus a protein that serve as 'workhorse' to test them. BfrB could be one such protein.

Data availability

The refined model has been deposited in the Protein Data Bank as 7O6E and maps within the Electron Microscopy Data Bank as EMD-12738.

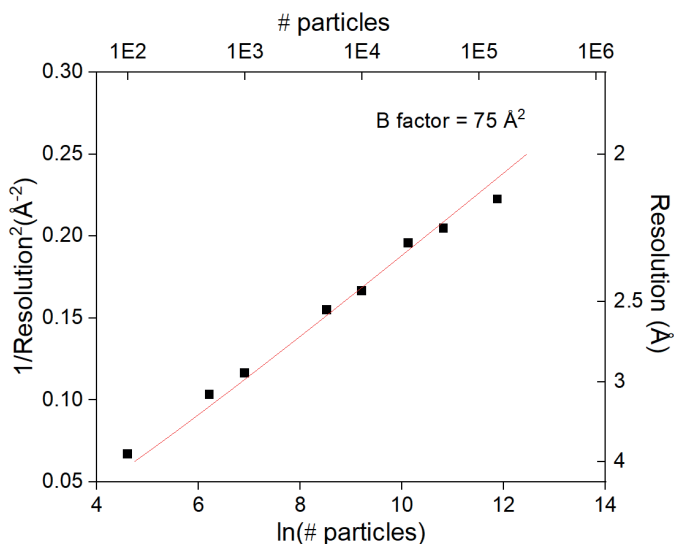
Author contribution

Abril Gijbers: Methodology, Investigation, Writing - Original Draft, Writing - Review & Editing. **Yue Zhang:** Formal analysis, Investigation, Writing - Original Draft, Writing - Review & Editing. **Ye Gao:** Methodology. **Peter J. Peters:** Supervised, Funding acquisition. **Raimond B.G. Ravelli:** Conceptualisation, Methodology, Formal analysis, Investigation, Funding acquisition, Supervision, Writing - Original Draft, Writing - Review & Editing.

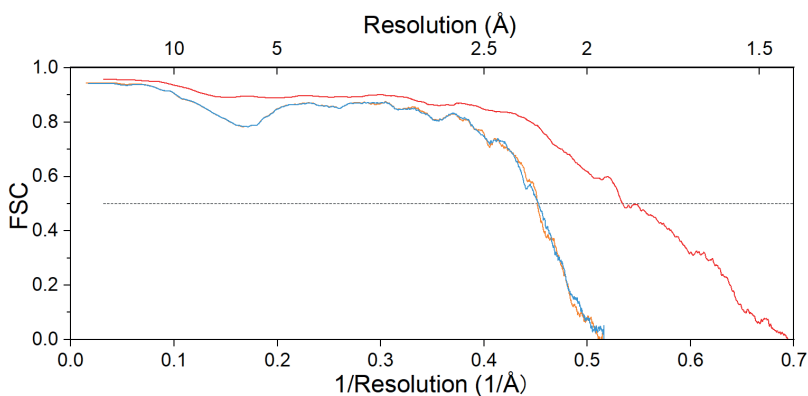
Acknowledgments and funding

We thank Paul van Schayck and Jan-Erik Thede (UM) for IT support, the Microscopy CORE Lab (UM) for support of the infrastructure, and Stephanie Siregar and Isabella Menart (both UM) for different purifications. This work benefited from access to the Netherlands Centre for Electron Nanoscopy (NeCEN) with assistance from Ludovic Renault and Wen Yang. This research received funding from the Netherlands Organisation for Scientific Research (NWO) in the framework of the National Roadmap NEMI project number 184.034.014, from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 766970 Q-SORT, as well as from the Province of Limburg, the Netherlands.

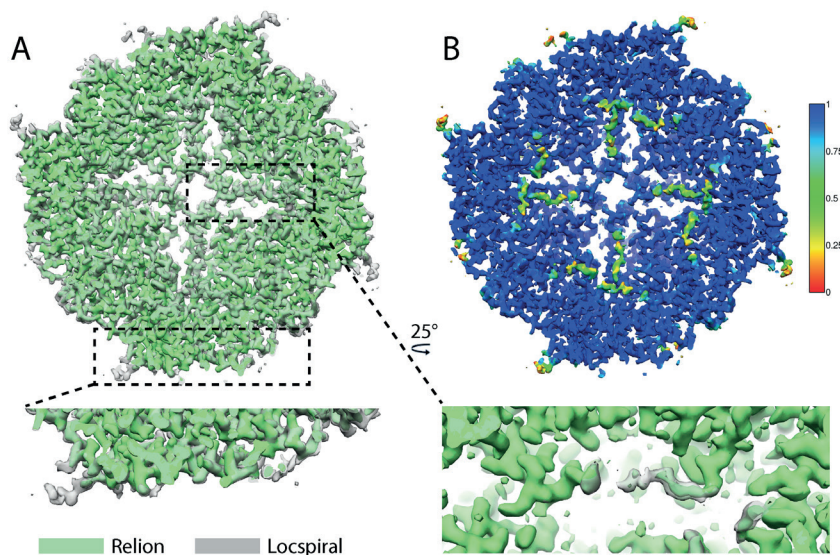
Supplementary material



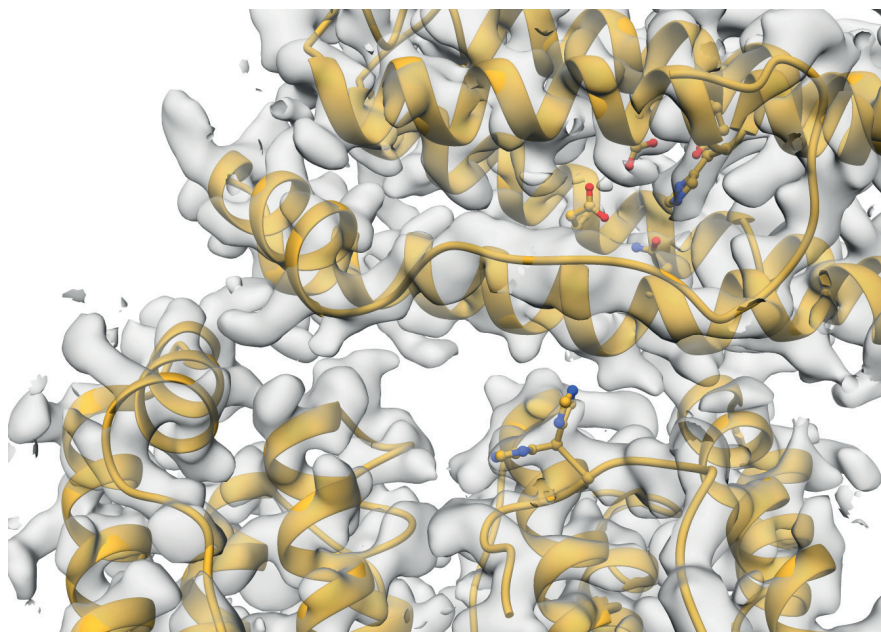
Supplementary Figure 3. B-factor plot according to Rosenthal and Henderson (148), where $1/d^2$, with d being the resolution of each refinement, is plotted against the natural logarithm of the number of particles in each subset. The fitted B-factor is 75 \AA^2 .



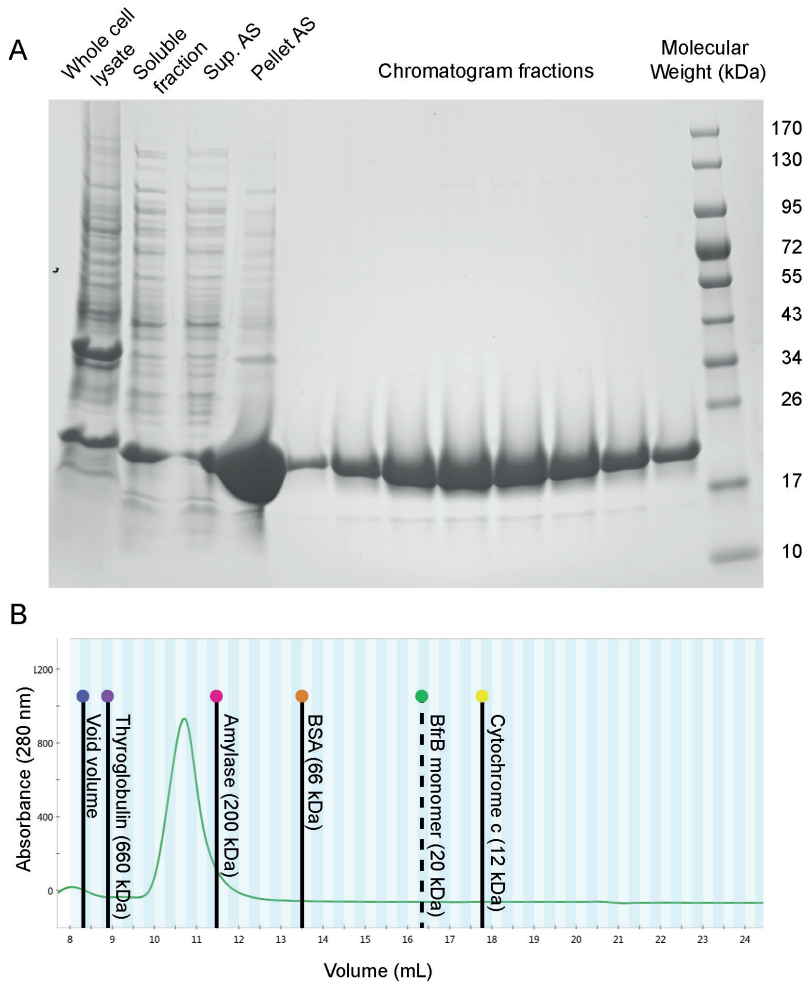
Supplementary Figure 4. Fourier Shell Correlation (FSC) between map and model. The blue and orange lines are the FSC curves between model and both half maps. The red curve reports a resolution for the final map, according to the FSC (model) = 0.5 criterium, of 1.89 \AA .



Supplementary Figure 5. (A) The superimposed sharpened maps obtained by *LocSpiral* (grey colour) and RELION postprocessing (green colour). The figures below show zoomed view of the region indicated with the dashed rectangles. **(B) The occupancy map of Mtb BfrB model obtained by LocOccupancy, calculated between 8 and 2.1 Å (156).** The occupancy ranges between [0, 1], the occupancy of residues at N-term and Cflex have occupancy lower than 1.



Supplementary Figure 6. Cryo-EM map and Mtb BfrB model with the B-pore channel shown in the middle with a double conformation of His175. His175 as well as the residues at the ferroxidase sites are shown in ball-stick model.



Supplementary Figure 7. Purification of BfrB purification. (A) Denaturing gel electrophoresis of the different purification steps. Lysis of bacterial suspension (Whole cell lysate), removal of cellular debris by centrifugation (Soluble fraction), protein precipitation by ammonium sulfate (Supernatant AS and Pellet AS). Fractions of BfrB oligomer peak obtained from (B) size exclusion chromatography. Reference proteins used for column calibration positioned in their corresponding elution volume.

CHAPTER 4



Priming mycobacterial ESX-secreted protein B to form a channel-like structure

Abril Gijbers, Vanesa Vinciauskaite, Axel Siroy, Ye Gao, Giancarlo Tria, Anjusha Mathew, Nuria Sánchez-Puig, Carmen López-Iglesias, Peter J. Peters and Raimond B. G. Ravelli

Abstract

ESX-1 is a major virulence factor of *Mycobacterium tuberculosis*, a secretion machinery directly involved in the survival of the microorganism from the immune system defence. It disrupts the phagosome membrane of the host cell through a contact-dependent mechanism. Recently, the structure of the inner-membrane core complex of the homologous ESX-3 and ESX-5 was resolved; however, the elements involved in the secretion through the outer membrane or those acting on the host cell membrane are unknown. Protein substrates might form this missing element. Here, we describe the oligomerisation process of the ESX-1 substrate EspB, which occurs upon cleavage of its C-terminal region and is favoured by an acidic environment. Cryo-electron microscopy data shows that quaternary structure of EspB is conserved across slow growing species, but not in the fast growing *M. smegmatis*. EspB assembles into a channel with dimensions and characteristics suitable for the transit of ESX-1 substrates, as shown by the presence of another EspB trapped within. Our results provide insight into the structure and assembly of EspB, and suggests a possible function as a structural element of ESX-1.

Keywords

Cryo-EM, EspB, ESX-1, mycobacteria, preferential orientation

I. Introduction

Tuberculosis (TB) is an infectious disease caused by the bacillus *Mycobacterium tuberculosis*. In 2018, 10 million people developed the disease from which 0.5 million were caused by multidrug-resistant strains. Even though TB is curable, 1.5 million people succumb to it every year (157). The current treatment is long and with serious side effects, often driving the patient to terminate the therapy before its conclusion (39). This has contributed to an increase in the number of patients suffering from multidrug- and extensively drug-resistant TB. While treatment is available for some of these resistant strains, the regimen is usually longer, more expensive and sometimes more toxic. For this reason, research on mycobacterial pathogenesis is vital to find a proper target in order to develop more effective therapeutics and vaccines.

The high incidence of TB relates to the ability of *M. tuberculosis* to evade the host immune system (158). This ability is related to multiple factors, one of which is a complex cell envelope of low permeability that plays a crucial role in drug resistance and in survival under harsh conditions (159). Likewise, pathogenic mycobacteria secrete virulence factors that manipulate the environment and compromise the host immune response. Mycobacteria have up to five specialised secretion machineries that carry out this process, named ESX-1 to -5 (together known as the type VII secretion system or T7SS). The core components of the inner-membrane part of the T7SS have been identified (160). Nevertheless, it remains unknown whether the translocation of substrates through the inner and outer membranes is functionally coupled or not, and if it deploys a specific outer-membrane complex to do so (161). Proteins from the PE/PPE family, characterised by Pro-Glu and Pro-Pro-Glu motifs and secreted by T7SS, are often associated with the outer most layer of the mycobacterial cell envelope, and have been suggested to play a role in the membrane channel formation (162-164). Recently, the intake of nutrients by *M. tuberculosis* was shown to be dependent on PE/PPE proteins, suggesting that these form small molecule-selective porins that allow the bacterium to take up nutrients over an otherwise impermeable barrier (164).

ESX-1 to -5 are paralogue protein complexes with specialised functions and substrates, unable to complement each other (76,165). ESX-1 is essential for the virulence of *M. tuberculosis*. It has been implicated in phagosomal escape, cellular inflammation, host cell death, and dissemination of the bacteria to neighbouring cells (74,79-82). Our knowledge about the structure of the machinery as well as the mechanism of secretion and regulation remains limited. ESX-3 is involved in iron homeostasis (84), and only recently the **molecular** architecture of its inner-membrane core has been determined (166,167). The complex consists of a dimer of protomers, made of four proteins: the ESX-conserved component (Ecc)-B, C, D ($\times 2$), and E. Despite the resolution achieved in both studies, there was no obvious channel through which the proteins substrates can traverse. Rosenberg and collaborators have described that one of the elements of the secretion system (EccC) forms dimers upon substrate binding, which then forms

higher-order oligomers (168). This is in agreement with observations that ESX-5, involved in nutrient uptake (83) and host cell death (80), forms a hexamer (169-171). A recent structure of an ESX-5 trimer of dimers shows that it is stabilised by a mycosin protease (MycP) positioned in the periplasm on top of EccB₅ (170). ESX-2 and ESX-4 are the least characterised, where ESX-4 is involved in DNA transfer (85) and is seen as the ancestor of the five ESX-systems (86). Located in different positions in the genome, the *esx* loci contain the genes that code for the Ecc proteins, MycP, a heterodimer of EsxA/B-like proteins, and one or more PE-PPE pairs. With high sequence similarity and conservation between paralogues (167,172), one could expect the inner-membrane core of the different systems to share a similar architecture. So what makes each one of them unique? Experimental data suggest that the answer lies with the substrates (173). The *esx-1* locus encodes for more than ten unique proteins that are known to be secreted (174), termed the ESX-1 secretion-associated proteins (Esp) (175). Amongst those is EspC, a protein present in pathogenic mycobacteria that was described to form filamentous structures *in vitro* and to localise on the bacterial surface *in vivo* (173). Due to the similarities between EspC and the needle protein of the type III secretion system, Lou *et al.* hypothesised that ESX-1 could be an injectosome system with EspC as its needle. This is of particular importance because, compared to the other secretion systems, ESX-1 function has been described to take place through a contact-dependent mechanism (176), which makes the discovery of an outer-membrane complex essential for understanding the system. Other proteins, like EspE that has been localised on the cell wall (174,177,178), are of interest as possible elements of such outer-membrane complex. The protein EspB has been the focus of attention due to its ability to oligomerise upon secretion (179,180), making it a strong candidate as a structural component of the machinery (181).

EspB belongs to the PE/PPE family, but unlike other family members that form heterodimers in mycobacteria, EspB consists of a single poly-peptide chain fusing the PE and PPE domains (179). EspB is a 48-kDa protein that matures during secretion: Its largely un-structured C-terminal region is cleaved in the periplasm by the protease MycP¹, leaving a mature 38-kDa isoform (182-184). The purpose of this maturation is not yet clear but it was shown that inactivation of MycP¹, and thus cleavage of EspB, deregulates the secretion of proteins by ESX-1 (183). Chen *et al.* observed specific binding of EspB to phosphatidylserine and phosphatidic acid after cleavage (185), suggesting that the C-terminal processing of EspB is important for its function, possibly involving lipid binding. The crystal structure of the monomeric N-terminal region of EspB from *M. tuberculosis* and *M. smegmatis* has been determined: It forms a four-helix bundle with high structural homology between species (179,180). During the preparation of this work for publication, the structure of an EspB oligomer from *M. tuberculosis* was published by Piton *et al.*, showing features of a pore-like transport protein (181).

EspB is the only member of the PE/PPE family described to date to form higher-order oligomers. In this work, we studied the oligomerisation ability and structures of EspB

from *M. tuberculosis*, *M. marinum*, *M. haemophilum* and *M. smegmatis*. We show that truncation of EspB at the MycP₁ cleavage site and an acidic environment promote the oligomerisation of EspB from the three slow-growing species but not from fast-growing *M. smegmatis*. Oligomerisation is mediated by intermolecular hydrogen bonds and amide bridges between residues highly conserved in the slow-growing species, but absent in *M. smegmatis*. The structures of oligomeric EspB consist of two domains: an N-terminal region that forms a cylinder-like structure with a tunnel large enough to accommodate a folded PE-PPE pair, and a partly hydrophobic C-terminal region that interacts with hydrophobic surfaces. The oligomer has similar inner-pore dimensions as was described for the pore within the periplasmic region of ESX-5 (170). Visualisation of a trapped EspB monomer within the channel supports the idea that it could transit secreted proteins through its tunnel. Overall, in this work we describe factors that prime the oligomerisation of EspB, and provide insight into its potential role in the ESX-1 machinery.

2. Materials and methods

2.1 Cloning, expression and purification of EspB constructs

Different constructs used in this study are listed in **Table 2**. DNA fragments were PCR-amplified with KOD Hot Start Master Mix (Novagen®) from genomic DNA of *M. tuberculosis* H37Rv, *M. marinum* or *M. smegmatis* [BEI Resources, National Institute of Allergy and Infectious Diseases (NIAID)], and cloned in a modified pRSET backbone (Invitrogen™) using *Nsi*I and *Hind*III restriction sites. Constructs included an N-terminal 6×His-tag followed by a tobacco etch virus (TEV) protease cleavage site. EspB mutants and construct EspB₂₋₃₅₈ and EspB₂₋₂₈₇ were generated using KOD-Plus- Mutagenesis kit (Toyobo Co., Ltd.) from the plasmid encoding the full-length protein. All plasmids were sequenced to verify absence of inadvertent mutations. *M. haemophilum* and PE25-PPE41 construct were synthesised and codon optimised for expression in *Escherichia coli* (Eurofins Genomics).

For the non-codon optimised constructs, proteins were expressed in Rosetta™ 2(DE3) *E. coli* cells in Overnight Express™ Instant LB Medium (EMD Millipore) supplemented with 100 µg/mL of carbenicillin and 25 µg/mL of chloramphenicol for 50 h at 25 °C. In the case of codon optimisation, the protein was expressed in C41 (DE3) *E. coli* cells in the same conditions with the respective antibiotic. Prior to protein purification, cells were resuspended in buffer containing 20 mM Tris-HCl (pH 8.0), 300 mM NaCl, 40 mM imidazole, 1 mM PMSF, and 25 U/mL benzonase, and were lysed using an EmulsiFlex-C3 homogenizer (Avestin). Proteins were purified with HisPur™ Ni-NTA Resin (Thermo Fisher Scientific) equilibrated in the lysis buffer and eluted in the same buffer supplemented with 400mM imidazole. The 6×His-tag was cleaved using TEV protease followed by a second Ni-NTA purification to remove the free 6×His-tag, uncleaved protein and the His-tagged protease (186). In case higher purity was needed,

proteins were purified on a size-exclusion Superdex® 200 Increase 10/300 GL column (GE Healthcare) in buffer containing 20 mM Tris-HCl (pH 8.0), 300 mM NaCl. Protein was stored at -80 °C until further use.

2.2 Analytical size exclusion chromatography (SEC)

Samples were dialysed overnight in the corresponding buffers and different concentrations of protein were loaded onto a size-exclusion Superdex® 200 Increase 3.2/300 column (GE Healthcare Life Science) at a flow rate of 50 µL/min. Basic buffer comprised 20 mM Tris-HCl (pH 8.0), 150 mM NaCl, while the acidic buffer was 20 mM acetate buffer (pH 5.5), 150 mM NaCl.

2.3 Native mass spectrometry

Native mass spectrometry was used to obtain the high resolution mass information of the samples. *M. tuberculosis* EspB₂₋₃₄₈ (5 mg/mL) was buffer exchanged with 100 mM NH₄CH₃CO₂ (at pH 5.5 and 8.5) using 3-kDa molecular weight cut-off dialysis membrane overnight followed by an extra hour buffer exchange with a fresh NH₄CH₃CO₂ solution at 4 °C. The buffer exchange of fragments produced by limited proteolysis (2 mg/mL) was performed using SEC on a Superdex® 200 Increase 3.2/300 column (GE Healthcare Life Science) with 100 mM NH₄CH₃CO₂ at pH 6.8. CH₃COOH and NH₄OH were used to adjust the pH of NH₄CH₃CO₂ solution. The mass spectrometry measurements were performed in positive ion mode on an ultra-high mass range (UHMR) Q-Exactive Orbitrap mass spectrometer (Thermo Fisher Scientific) with a static nano-electrospray ionization (nESI) source. In-house pulled, gold-coated borosilicate capillaries (187) were used for the sample introduction to the mass spectrometer, and a voltage of 1.2 kV was applied. Mass spectral resolution was set at 4375 to 8750 (at m/z=200) and an injection time of 100–200 ms was used. For each spectrum, 10 scans were combined, containing 5 to 10 microscans. The inlet capillary temperature was kept at 320 °C. Parameters such as in-source trapping, transfer m/z, detector m/z, trapping gas pressure and mass range were optimized for each analyte separately. All mass spectra were analysed using Thermo Scientific Xcalibur software and spectral deconvolutions were performed with the UniDec software (188).

2.4 Cryo-EM sample preparation, data acquisition and image processing

Samples, in 20 mM acetate buffer (pH 5.5), 150 mM NaCl, were diluted to the respective concentrations (**Table 2**). A volume of 2.5 µL of each sample was applied on glow-discharged UltrAuFoil Au300 R1.2/1.3 grids (Quantifoil), and excess liquid was removed by blotting for 3 s (blot force 5) using filter paper followed by plunge freezing in liquid ethane using a FEI Vitrobot Mark IV at 100% humidity and 4 °C. For PE25-PPE41, an acetate buffer (pH 6.5), 150 mM NaCl was used, due to precipitation of the protein at lower pH.

Cryo-EM single particle analysis (SPA) data were collected using untilted and tilted schemes (189). For EspB₂₋₂₈₇ from *M. tuberculosis* and EspB₂₋₂₈₆ from *M. marinum*,

untilted images were recorded on a Titan Krios (Thermo Fischer Scientific) at 300 kV with a K3 detector operated in super-resolution counting mode. Tilted SPA data were collected for EspB₂₋₃₄₈ from *M. tuberculosis* on a 200-kV Tecnai Arctica TEM using SerialEM (190), using a Falcon III detector in counting mode. **Table 3** shows all specifications and statistics for the data sets. Individual micrographs of EspB₂₋₂₈₇ from *M. haemophilum*, EspB₂₋₄₀₇ from *M. smegmatis* as well as PE25-PPE41 from *M. tuberculosis* were collected on the 200-kV Arctica.

Data were processed using the RELION-3 pipeline (135). Movie stacks were corrected for drift (5×5 patches) and dose-weighted using MotionCor2 (147). The local contrast transfer function (CTF) parameters were determined for the drift-corrected micrographs using Gctf (106). The EspB₂₋₃₄₈ data set was collected at two angles of the stage: 0 degrees and 40 degrees. For each tilt angle, a first set of 2D references were generated from manually picked particles in RELION (16) and these were used for subsequent automatic particle picking. **Table 3** lists the number of particles in the final data set after particle picking, 2D and 3D classification. The 3D classification was run without imposing symmetry and used to select the heptameric particles. Local CTF parameters were iteratively refined (135), which was particularly important for the tilted data set, beam-tilt parameters were estimated, and particles were polished. Particle subtraction followed by focused classification was used to characterise densities other than those described by the refined model described below. Due to extreme preferred orientation of the datasets of EspB₂₋₃₄₈ heptamers, automatic masking and automatic B-factor estimation in post-processing were hampered by missing wedge artefacts. For this data set, parameters were manually optimised by visual inspection of the resulting maps. Density within the heptameric pore was obtained by a combination of 2D and 3D classification. The initial density map of a loaded complex was generated by symmetry expansion of a C7 3D-refined particle list, followed by 3D classification in C1 without further image alignment. Later iterations employed 45,671 unique particles and 3D refinement in C1 while imposing local symmetry for the heptamer. The resulting 5.3 Å map was used to identify a total of eight EspB monomers (heptamer plus one in the middle), and local symmetry averaged. The final resolution of the heptamer maps, listed in **Table 3**, varied between 2.3 and 3.4 Å, using the gold-standard FSC=0.143 criterion (108).

2.5 Structure determination and refinement

The PDB model 4XXX (179) was used as a starting model in Coot (149) for manual docking and building into the tilted-scheme SPA data set of EspB₂₋₃₄₈ of *M. tuberculosis*. The final model was refined against the high-resolution sharpened map of EspB₂₋₂₈₇ of *M. tuberculosis*. This model was later used as reference for *M. marinum* model. Models were refined iteratively through rounds of manual adjustment in Coot (191), real space refinement in Phenix (192) and structure validation using MolProbity (193).

2.6 Cryo-electron tomography

EspB₂₋₃₄₈ from *M. tuberculosis*, in 20 mM acetate buffer (pH 5.5), 150 mM NaCl, was diluted to 0.5 mg/mL. A volume of 2.5 μ L was applied on glow-discharged UltrAuFoil Au200 R2/2 grids (Quantifoil), and excess liquid was removed by blotting for 3 s (blot force 5) using filter paper followed by plunge freezing in liquid ethane using a FEI Vitrobot Mark IV at 100% humidity at 4 °C.

Tomography data was acquired with a Titan Krios (Thermo Fisher Scientific) equipped with a K3 direct electron detector (Gatan) in electron counting mode and pixel size of 1.37 Å. Movies were acquired in SerialEM (190) using a stage tilt scheme of -60° to 60° in increments of 3° through a total fluence of 80 e⁻/Å² and a defocus target range of -3 to -6 μ m. Recorded movie fractions were corrected for beam-induced sample motion and processed into dose weighted stacks with MotionCor2 (147). Tilt series were aligned, 4 × binned and reconstructed with IMOD (194) using patch tracking and SIRT, respectively.

2.7 Limited proteolysis and Edman sequencing

Samples were incubated with trypsin for different length of time at a molar ratio of 1:6 (enzyme:substrate) following the Proti-Ace™ Kit (Hampton Research) recommendation. The reaction was stopped by adding SDS-PAGE loading buffer (63 mM Tris-HCl, 2% SDS, 10% glycerol, 0.1% bromophenol blue) and samples were resolved on a 12% polyacrylamide gel. Bands were transferred from the SDS-PAGE gel to a PVDF membrane and stained with 0.1% Coomassie Brilliant Blue R-250, 40% methanol, and 10% acetic acid until bands were visible. The membrane was then washed with water and dried, and EspB cleavage products were cut out. The first ten amino acids were determined by Edman sequencing at the Plateforme Protéomique PISSARO IRIB at the Université de Rouen.

2.8 Circular dichroism (CD) spectroscopy

The CD spectra of 5 μ M EspB₂₇₉₋₄₆₀ were recorded either in 50 mM phosphate (pH 8.0), 50 mM NaCl or 10 mM acetate (pH 5.5), 50 mM NaCl at 25 °C in the far-UV region using a Jasco J-1500 CD spectropolarimeter (JASCO Analytical Instruments) on a 0.1 cm path-length cell. Spectra correspond to the average of five repetitive scans acquired every 1 nm with 5-s average time per point and 1-nm band pass. Temperature was regulated with a Peltier temperature-controlled cell holder. Data were corrected by subtracting the CD signal of the buffer over the same wavelength region. The effect of 2,2,2-trifluoroethanol (TFE) was recorded using the aforementioned phosphate buffer. Temperature dependence of ellipticity was followed at 222 nm from 20 to 80 °C, a bandwidth of 1 nm, a time response of 16 s and a temperature gradient of 1 °C/min. The data was fitted to a Boltzmann distribution describing a standard two-state transition model to obtain the melting temperature.

3. Results

3.1 Oligomerisation is favoured by an acidic pH and maturation of EspB

Previously, it has been described that oligomerisation of EspB occurs after secretion (179). In the infection context, this secretion would lead the protein to the phagosomal lumen of a macrophage, an organelle known to have pH acidification as a functional mechanism. To evaluate the putative role of pH in the oligomerisation process, the mature form of *M. tuberculosis* EspB (residues 2–358) was incubated at different pH values and analysed by size exclusion chromatography (SEC). Results showed that the equilibrium is favoured towards an oligomer form at pH 5.5 compared to pH 8.0 (**Figure 12A**), as observed by a higher oligomer/monomer ratio at any protein concentration (**Figure 12B**). Native mass spectrometry experiments confirmed this behaviour and could identify different oligomeric states of EspB, with the heptamer being the most predominant. Intermediate states were observed (dimer to pentamer) and even higher oligomeric states (octamer) but in lower abundance compared to the heptamer (**Figure 12C and D**).

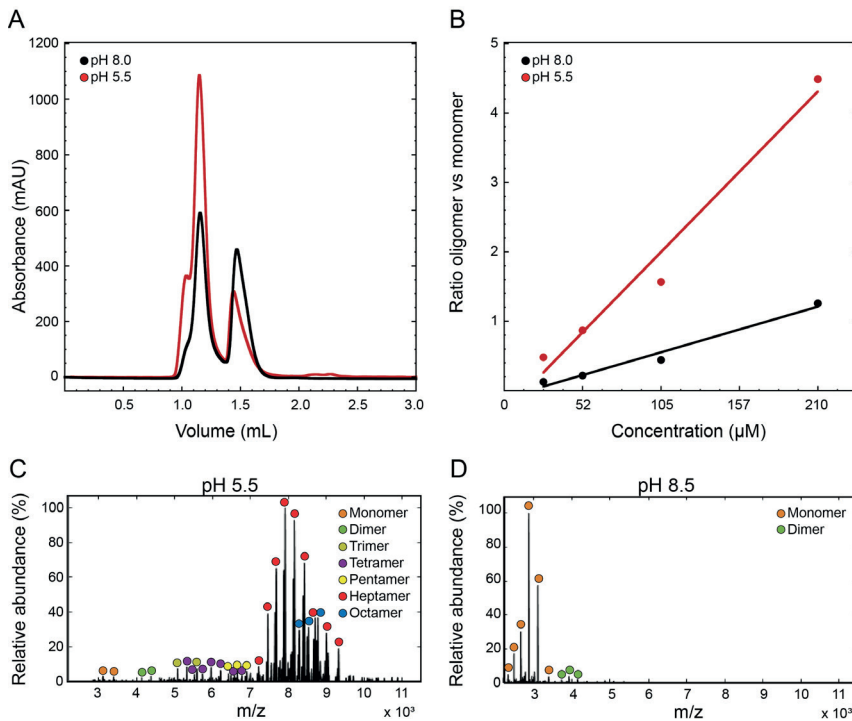


Figure 12. Oligomerisation of EspB is promoted by an acidic environment. (A) Size exclusion chromatography profiles of *M. tuberculosis* EspB₂₋₃₅₈ at 210 μM in 20 mM acetate buffer (pH 5.5), 150 mM NaCl and 20 mM Tris (pH 8.0), 150 mM NaCl. Void volume corresponds to 0.8 mL elution volume. (B) Oligomer/monomer ratios at different protein concentrations in conditions from panel (A). The absorbance values of the oligomer were taken at 1.14 mL while the monomer values were at 1.48 mL. (C–D) Presence of the different oligomer species from *M. tuberculosis* EspB₂₋₃₄₈ at pH 5.5 and pH 8.5 obtained by native mass spectrometry.

Because EspB undergoes proteolytic processing of its C-terminus during secretion, we investigated the effect of this cleavage at pH 5.5 on the quaternary structure of different *M. tuberculosis* EspB constructs that vary in their C-terminus length (**Figure 13**). With the exception of EspB₇₋₂₇₈ that did not oligomerise, we observed that oligomerisation was favoured for all other constructs at pH 5.5 (**Figure 13B**). The full-length EspB₂₋₄₆₀ (**Figure 13B**, blue trace) presented the lowest amounts of complex formation compared to the other constructs tested, while the highest was observed for the mature isoform, EspB₂₋₃₅₈ (**Figure 13B**, orange trace). Shorter versions showed oligomerisation, though not to the extent as EspB₂₋₃₅₈. These results suggest that MycP₁ cleaves EspB to allow oligomerisation, and that the remaining residues of the unstructured C-terminal region could aid complex stabilisation. Melting temperatures (T_m) calculated by circular dichroism (CD) highlighted EspB₇₋₂₇₈ as the least stable construct of all those tested, showing a 14 °C difference or more (**Figure 13A**), which could explain the negative impact on the oligomerisation.

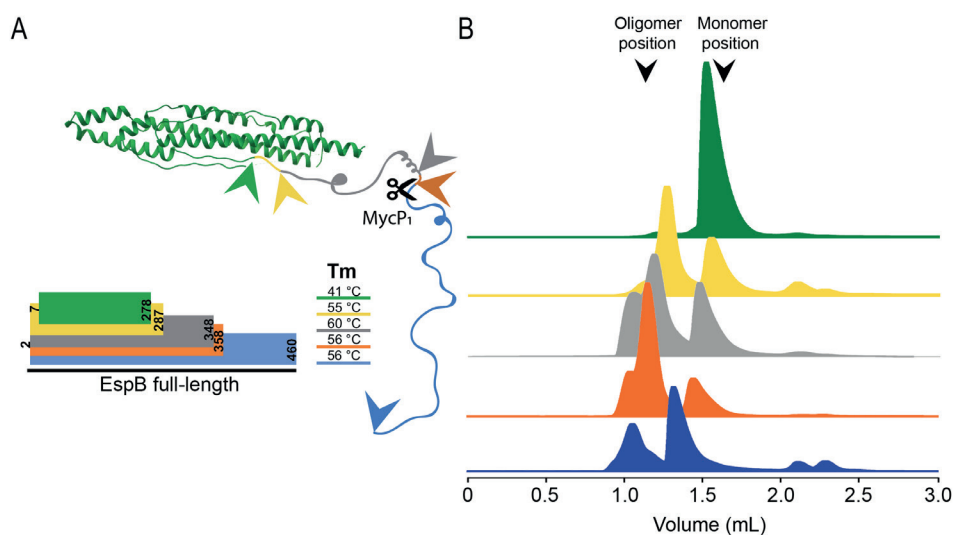


Figure 13. Impact of EspB C-terminus processing on oligomerisation. (A) Scheme of the different constructs used in this work, where EspB₂₋₄₆₀ is in blue, EspB₂₋₃₅₈ in orange (MycP₁ cleavage site), EspB₂₋₃₄₈ in grey, EspB₂₋₂₈₇ in yellow and EspB₇₋₂₇₈ in green. Structural model from PDB ID 4XXX, while the C-terminal region is a representation of an unfolded protein. Arrows represent the end of each construct. Melting temperatures were obtained following the dependency of the circular dichroism signal as a function of temperature. (B) Size exclusion chromatograms of each construct corresponding to the colours in panel (A), resulting from 50 μ L sample injection at 220 μ M eluted in 20 mM acetate buffer (pH 5.5), 150 mM NaCl. Void volume corresponds to 0.8 mL elution volume.

3.2 EspB from *M. smegmatis* is unable to oligomerise

To determine whether the oligomerisation ability is conserved across species, we performed cryo-EM analysis on different orthologues of the mature EspB. We selected *M. marinum* due to its wide use as model microorganism to study TB, *M. haemophilum* as an evolutionary distant specie from *M. tuberculosis*, and *M. smegmatis* representing a fast-growing *Mycobacterium* and a common organism used in research. Proteins from slow-growing species *M. tuberculosis*, *M. marinum* and *M. haemophilum* were able to oligomerise into ring-like structures while the fast-growing *M. smegmatis* did not, as seen by the lack of visible particles (**Figure 14A**); the structured region of an EspB monomer (30 kDa) has a signal-to-noise ratio too low to be visualised within these micrographs (21,195). Interestingly, comparison of the tertiary structure from the slow-growing species studied here with the published crystallographic model of EspB from *M. smegmatis* did not show substantial differences (RMSD C α 's 0.979–1.128 Å), apart from an extended α -helix 2 (**Figure 14B**), absent in our oligomeric structures. To determine whether the differences in oligomerisation ability between *M. smegmatis* and the slow-growing species were due to variations in their primary structure, we performed sequence alignment of multiple EspB orthologues. The species able to oligomerise have high sequence identity, together with other slow-growing mycobacteria species, whereas *M. smegmatis* has a low sequence identity, similar to other fast-growing species (**Figure 14C** and **Supplementary Figure 8**).

Because EspB belongs to the PE/PPE family, we included in the analysis a PE-PPE pair with a structure already published (196). PE25-PPE41 did not oligomerise (**Figure 14**), despite sharing a similar tertiary structure (RMSD C α 's 1.134 Å). With a low percentage identity (21.3%), it confirms the importance of specific amino acids sequence for the conservation of the quaternary structure.

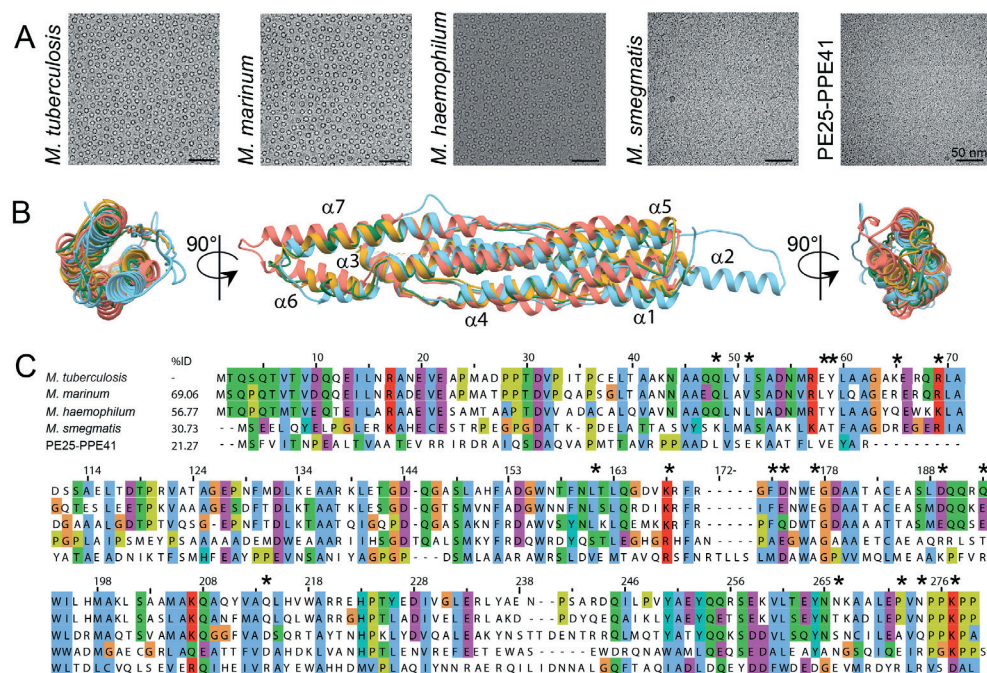


Figure 14. Oligomerisation differences between EspB orthologues despite sharing similar tertiary structure. (A) Evaluation of the oligomerisation of EspB orthologues and PE25-PPE41 by cryo-electron microscopy. Scale bars represent 50 nm. (B) Different views of structural alignment of EspB_{Mtb} (yellow – this work), EspB_{Mmar} (green – this work), EspB_{Msmeg} (light blue – PDB ID 4WJ1), and PE25-PPE41 (orange – PDB ID 4W4K). (C) Multi-alignment of amino acid sequences of different species from the *Mycobacterium* genus, as well as the protein pair PE25-PPE41. Numbering and sequence identity is based on the sequence of *M. tuberculosis*. Asterisks denote residues making inter-subunit contact in EspB from *M. tuberculosis* and *M. marinum*. Alignment was generated using ClustalW server, and figure was created using software Jalview (197). The colour scheme of ClustalX is used (198).

3.3 High-resolution cryo-EM structures of EspB oligomers

Next, we aimed to solve the high-resolution structure of EspB oligomers by cryo-EM. Initial experiments were performed with EspB₂₋₄₆₀ and EspB₂₋₃₄₈ from *M. tuberculosis*, which displayed an extreme preferential orientation where only “top views” could be seen (**Figure 15A**). Cryo-electron tomography revealed these molecules to be attached to the air-water interface (**Supplementary Figure 9**). Different oligomers were found: hexamers, heptamers, rings with an extra density in the middle, and octamers, with the heptameric ensemble being the predominant one (**Figure 15A**), in agreement with the results obtained in solution (**Figure 12C**). Preliminary 3D reconstructions could be obtained from data collected at one or more tilt angles (**Figure 15B**). Data processing resulted in 3–4 Å resolution maps from which the first heptamer models were built. Removal of C-terminal to residue 287 led to a different distribution of particles on the cryo-EM grids, now with random orientations (**Figure 15C**), denoting an interaction between this region and the air-water interface on the EM grid (23).

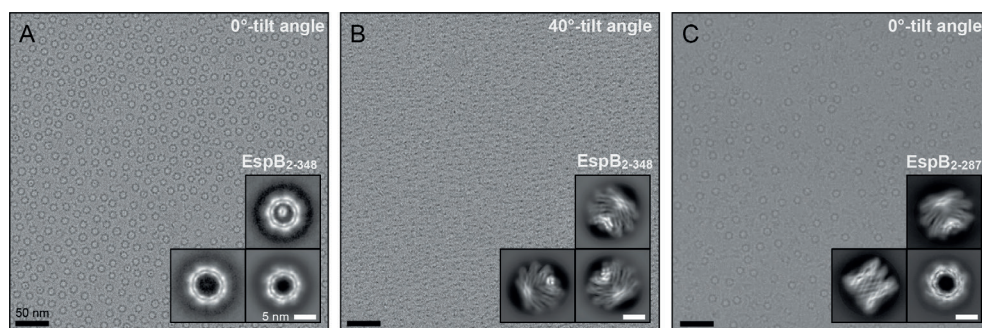


Figure 15. Loss of the EspB preferential orientation by removal of its C-terminal residues. (A–B) Representative micrograph of EspB₂₋₃₄₈ with preferential orientation at 0°-tilt angle or 40°-tilt angle. (C) Representative micrograph of EspB₂₋₂₈₇ with random orientation taken at 0°-tilt angle. Insets correspond to the respective 2D classes. Scale bars in A–C represent 50 nm; scale bars in insets represent 5 nm.

Experiments were repeated for constructs EspB₂₋₂₈₇ from *M. tuberculosis*, and the equivalent construct from *M. marinum* (at 0°-stage tilt), leading to high-resolution EM maps of 2.3 Å and 2.5 Å average resolution, respectively (**Figure 16A and B**, and **Supplementary Figure 10A–C**). We observed high structural conservation between the two structures. Both displayed a four-helix bundle, like the EsxA-B complex and PE25-PPE41 complex, with the WxG and YxxxD located on one end of the elongated molecule, referred to the top hereafter, making an H-bond interaction between the nitrogen of W176 with the oxygen of Y81, as was observed in the crystal structure (**Figure 16C**) (179). The helical tip is located on the opposite end, referred to as the bottom, for both EspB and PE25-PPE41 (179,180). The C-terminal region starts near the top end of the elongated molecule. The overall structure shows seven copies tilted 32° with respect to the symmetry-axes forming a cylinder-like oligomer with a width and a height of 90 Å (**Figure 16A and B**).

The single-particle analysis (SPA) map from *M. tuberculosis* EspB revealed several interaction pairs between monomers (**Table 4**), involving residues highly conserved in slow-growing mycobacteria (**Figure 16C** and **Supplementary Figure 8**). Mutation in one of these residues (Q48A) in the *M. tuberculosis* orthologue resulted in the disruption of the oligomer (**Supplementary Figure 10D**, yellow trace) as evidenced by the absence of a high molecular weight peak in the size exclusion chromatogram. Q48 forms an amide bridge with Q164 (**Figure 16C**), an uncommon interaction seen within the structures present in the Protein Data bank (PDB) (199). However, they are stronger interactions than typical hydrogen bonds and less affected by pH changes than salt bridges, another strong interaction (200). According to the theoretical pKa of amino acids, Histidines would be the only residue that would change its protonation state in the pH interval 5.5–8.0; however, no Histidines were found in the oligomerisation

interface. Alterations of the pKa can result from the influence of neighbouring residue in a 3D structure. Prediction by PROPKA did not show a significant change in other ionisable groups, suggesting that these residues mainly play a role in the pH-dependent overall charge distribution of the monomers (**Figure 17A and B**).

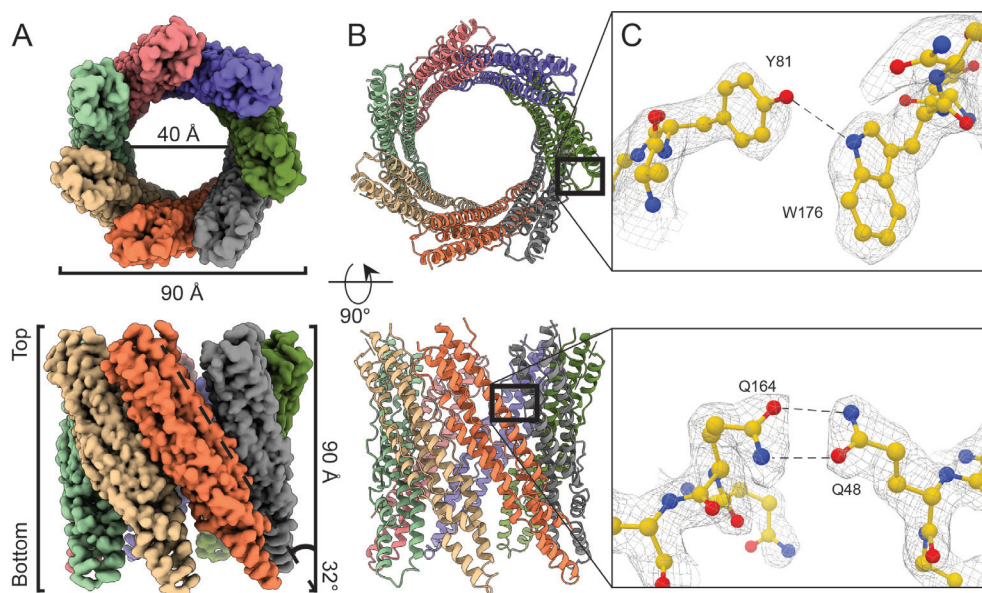


Figure 16. Cryo-EM reconstruction of EspB₂₋₂₈₇ heptamer complex. (A–B) Density map and structural model made with ChimeraX (201), showing each monomer in different colours. For (A) and (B), the upper panels show the top views and the bottom panels show the side views. Representation of the smallest inner-diameter calculated by HOLE (202). (C) Model and densities of intramolecular interaction at W176-Y81 and intermolecular interaction at Q48-Q164. Colours follow the conventional colouring code for chemical elements.

The high-resolution EspB oligomer maps did not reveal a continuous density for the PE-PPE linker. The proposed location of the linker within the crystal structure (179) overlaps with the oligomerisation interface, and would need to adopt a different position upon oligomerisation. Particle subtraction followed by focused classification showed partial densities for the linker at the periphery of the structure. To confirm that the PE-PPE linker adopts a different conformation upon oligomerisation, we created a double mutation, N55C (in the core of the monomer) and T119C (in the PE-PPE linker), that would lock the linker in its crystal position, thus preventing it from adopting a different conformation needed for the oligomerisation. This double mutant abolished oligomerisation of EspB, suggesting that such intramolecular disulfide bond was formed (**Supplementary Figure 10D**, red trace).

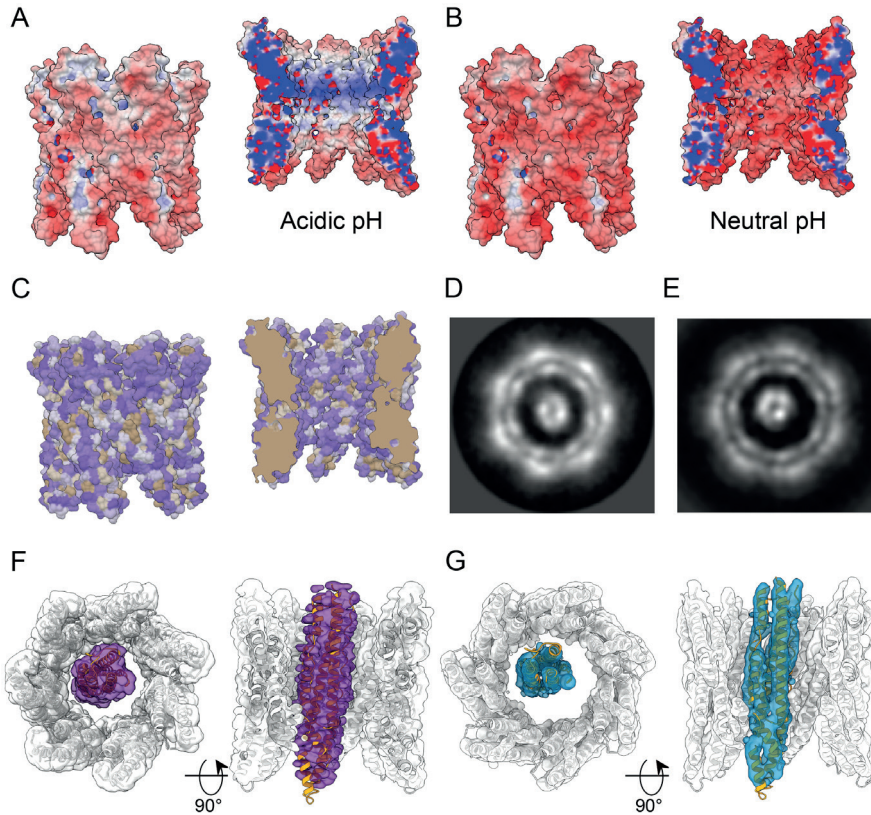


Figure 17. Evaluation of EspB pore. (A–B) Electrostatic potential of EspB oligomer at pH 5.4 and (B) neutral pH 6.7 (-10 kT/e – red, 10 kT/e – blue). The protonation state was assigned by PROPKA (203) and electrostatic calculations were generated by APBS (204) and PDBPQR (205). (C) Surface representation of amino acid hydrophobicity according to the Kyte-Doolittle scale (polar residues – purple, non-polar residues – gold). (D) High-resolution 2D class of EspB heptamer with extra density in the middle. (E) Projection of the 3D map obtained for the 7+1 EspB oligomer. (F) C1 3D map of 7+1 EspB oligomer with local symmetry applied to the heptamer ring. (G) C1 3D map of the 7+1 EspB oligomer with 8-fold local symmetry applied and models fitted to the map.

3.4 EspB, a possible transport channel for T7SS proteins

The EspB cylinder-like structure has an internal pore diameter of 40 \AA (Figure 16A), large enough to accommodate folded T7SS substrates such as EsxA/EsxB (diameter 35 \AA), PE25-PPE41 (diameter 27 \AA) or an EspB monomer itself (diameter 28 \AA). Analysis of the degree of hydrophobicity in the structure showed that the internal surface of the oligomer is mainly hydrophilic (Figure 17C), allowing other hydrophilic molecules to pass.

During the cryo-EM data processing, additional densities were consistently found within the EspB heptamer of all the different constructs, including constructs that lack the C-terminal region. Figure 4.6D shows a high-resolution 2D class of EspB₂₋₃₄₈ with

a well-defined density inside the channel from a subset of the data collected at 0°-stage tilt. This 2D class was found in ~7% of the particles recorded. The 2D classes obtained at 40°-stage tilt could not be unambiguously manually assigned to specific oligomerisation forms. Instead, 3D classification in RELION (16) was used to identify one class with solely C7 symmetry and one class with an extra density within the heptameric channel. Local symmetry averaging of the heptamer model while processing the overall map in C1 map revealed an extra density spanning the entire channel, in which an EspB monomer model can be fitted (**Figure 17E–G**).

3.5 Integrity of the PE-PPE linker is not essential for the oligomerisation of EspB

To determine if the PE-PPE linker absent in our model was essential for oligomerisation, we performed limited proteolysis analysis on the *M. tuberculosis* constructs. Incubation with trypsin fully digested EspB₇₋₂₇₈, perhaps due to a lower stability (**Figure 13A**), but resulted in two major fragments for the constructs EspB₂₋₄₆₀ and EspB₂₋₃₄₈, as shown by SDS-PAGE (**Supplementary Figure 11A**). N-terminal sequencing and mass spectrometry analysis revealed that the larger fragment corresponded to a section of the protein comprising residues V122 to R343 (corresponding to the PPE domain), while the smaller fragment included the N-terminal end of the protein sequence, with a few residues from the affinity tag, up to residue R121 (PE domain and linker) (**Supplementary Figure 11B–D**). Despite being split within the PE-PPE linker into two fragments, EspB₂₋₃₄₈ behaved in SEC as a single entity with the capacity to form oligomers (**Supplementary Figure 11E and F**), confirming that the integrity of this region is not necessary for the complex to form. It is noteworthy that trypsin did not cut before R343, even though there is a cleavage site in the allegedly unfolded C-terminal region, raising the question whether this region is truly fully unstructured.

3.6 Properties of the EspB C-terminal region

The function of the C-terminal region has puzzled the scientific community for a long time, partly because EspB is the only substrate known to date of the MycP₁ protease. Here, we described its processing as an important factor for the oligomerisation of the N-terminal region (1-287); however, the cleavage leaves ~70 residues for no-obvious reason. To gain insight in the properties of the C-terminal region that could hint for its function, and based on the preferential orientation-effect seen in cryo-EM, we performed a hydrophobicity analysis of this region on the different EspB orthologues. Analysis evidenced the presence of hydrophobic patches in the slow-growing species, and absent in EspB from *M. smegmatis* (**Supplementary Figure 12A**). Some of these patches are present in all the constructs with preferential orientation, leading to speculate that residues in the 297–324 region interact with the air-water interface of the cryo-EM grid (23).

To understand whether this effect is related to a structural change or a particular characteristic in the C-terminal region of the protein, we expressed a construct corresponding to residues 279–460 and carried out circular dichroism (CD) studies on it. Far UV CD spectra analysis of this region showed a negative band around 198 nm (**Supplementary Figure 12B**), characteristic of random coil structures. This result is in line with the high fraction (54%) of “disorder-promoting” residues within this region (lysine, glutamine, serine, glutamic acid, proline, and glycine: amino acids commonly found in intrinsically disordered protein regions). Interestingly, its proline content is 2.5 times higher than that observed for proteins in the PBD (206,207). Analysis of the CD difference spectra obtained at two pH [$\Delta\theta$ (pH 5.5 – pH 8.0)] revealed a positive signal close to 220 nm and a negative signal near 200 nm (**Supplementary Figure 12B inset**), showing that this region is able to adopt extended left-handed helical conformations [poly-L-proline type II or PPII (208)]. CD analysis of the C-terminal region in the presence of different concentrations of 2,2,2-trifluoroethanol (TFE) showed that this region has an intrinsic ability to attain helicity based on the decrease in the ellipticity signal at 222 nm (**Supplementary Figure 12C**) (209). The lack of a single isodichroic point at 200 nm suggests that the conformational changes elicited by TFE do not comply with a two-state model; most probably, the transition is accompanied by an intermediate, *e.g.*, the presence of more than one α -helix.

4. Discussion

In the present study, we describe different factors that facilitate oligomerisation of EspB: an acidic environment, the truncation of its C-terminal region, a flexible PE-PPE linker and the residues involved in the interaction. Our findings are in agreement with previous observations that EspB oligomerises upon secretion (179). Based on these results, the C-terminus of the full-length protein could prevent premature oligomerisation in the cytosol of mycobacteria, possibly through steric hindrance. However, this region is also likely to have other functions. Deletion of EspB C-terminus does not affect its own secretion (182,210) but rather the secretion of EsxA/EsxB, possibly by loss of interaction with the last residues of EspB (182). The sequence of the C-terminal end is highly conserved (**Supplementary Figure 8**), which makes it possible that this region interacts with other molecules in the cytoplasm of the bacterium.

The ability of EspB to oligomerise seems to be conserved across slow-growing species, in contrast to fast-growing mycobacteria, like *M. smegmatis*. This microorganism is a non-pathogenic species that uses ESX-1 system for horizontal DNA transfer (211). The exact mechanism of this transfer is unknown; however, evidence suggests that ESX-1 is not the DNA conduit but rather secretes proteins that act like pheromones, which in turn induces the expression of *esx-4* genes resulting in mating-pair interactions (85). The ESX-1 substrate EsxA was shown to undergo a structural change that allows membrane

insertion in *M. tuberculosis* when exposed to an acidic environment, however, this effect does not occur in its *M. smegmatis* orthologue (212,213). Taking the aforementioned antecedents and the oligomerisation differences between EspB proteins observed in this work, it is plausible that the mechanism of action of ESX-1 is distinct between these two species. High degree of sequence conservation in slow-growing species compared to the fast-growing ones suggests that EspB gained the ability to oligomerise during evolution. Slow-growth is usually a characteristic of pathogenic species; however, pathogenesis is a complex process involving more elements than the ESX-1 secretion system (76). Follow-up experiments are needed to fully understand the role of EspB in mycobacterial pathogenesis.

EspB interacts with the lipids phosphatidylserine and phosphatidic acid (185). It was suggested that EspB could transport phosphatidic acid (181), but the complex's interior is mainly hydrophilic, making this scenario less plausible. Despite the presence of lipids in the crystallisation setup, Korotkova *et al.* (2015) could not find lipids within the crystal structure of EspB₇₋₂₇₈, which lacks the C-terminus. Our results show that the C-terminus of EspB contributes to the protein's preferred orientation on an EM grid caused by interaction with the hydrophobic air-water interface (23), analogous to what could happen on a lipid membrane. With a PPII helix at the end of the channel followed by hydrophobic patches at the C-terminus, we hypothesise that this secondary structure interacts with the head group of the lipids, as it has been described for other PPII helices (214), allowing the hydrophobic residues to insert into bilayer membranes. Based on the chemical properties of the channel and supported by the evidence of an extra EspB monomer observed within the oligomer, we propose that EspB could be a structural element of ESX-1, allowing other substrates to transit through the channel.

The combined data presented in our work leads us to hypothesise three models of the role of the EspB oligomer. EspB within the cytosol is likely to be monomeric (179), either free or chaperoned by EspK (210), similar to the role of EspG and PE/PPE pairs (215,216). The binding of a chaperone to the helical tip of EspB would place the WxG and YxxxD bipartite secretion signal exposed on the top of EspB, ready to interact with the T7SS machinery. Upon exiting ESX-1 inner-membrane pore, the pre-protein EspB will be cleaved within the periplasm by MycP₁. Analogous to ESX-5 (170), we expect MycP₁ to cap the central periplasmic dome-like chamber formed by EccB₁, and to have its proteolytic site facing towards its central pore. The cleavage of the C-terminal region at A358 (184) will remove the most hydrophilic part of the C-terminus, leaving a hydrophobic tail (**Supplementary Figure 12A**).

From here, we propose three possible pathways for the oligomerisation of EspB. In one scenario, after processing the C-terminus, EspB binds the outer membrane of mycobacteria, decreasing its critical concentration to form an oligomer (**Figure 18 Model 1**). As suggested above, it can be assumed that EspB monomers would transit

through the inner-membrane pore with the top first, where the C-terminus, as well as the WxG and YxxxD motifs, are located. In this position, the monomers would already be properly oriented to form an oligomer on the outer membrane inner leaflet just like EspB₂₋₃₅₈ attaches to the air-water interface of a cryo-EM grid (**Figure 15A**). The inner pore of the EspB heptamer has similar dimensions compared to that proposed by Beckham *et al.* for the ESX-5 hexameric structure (161), albeit they later published a higher resolution structure with a more constricted pore in a close state (217). The space between the inner and outer-membrane has been reported to be 20–24 nm wide (174,218,219), which could accommodate the 9-nm long EspB heptamer. It was postulated (181) that the positively charged interior of the EspB channel could play a role in the transfer of negatively charged substrates such as DNA or phospholipids. However, in analogy to the negative lumen of a bacteriophage tail that is used to transfer DNA (220), we propose that the positively charged interior space of the EspB oligomer would channel substrates of the same charge, as negatively charged substrates would most likely bind and get trapped. Since the heptameric structure presented here lacks any trans-membrane domains and is highly soluble, it is unlikely to be embedded within the outer membrane. However, it could be anchored by its C-terminus forming part of a larger machinery that completes the ESX-1 core complex. EspB is well known to be secreted to the culture medium of mycobacteria (221); thus, in this model, EspB will help in its own secretion by forming a channel through which additional substrates could travel. This proposal is based on the fact that EspB is cleaved right after crossing the inner membrane of the bacterium. This cleavage is one of the factors that prime oligomerisation.

In a second scenario, when EspB is secreted outside the bacterium, it could interact with either the phagosomal membrane or the external face of the outer membrane (**Figure 18 Model 2**). The aforementioned hypothesis of how EspB is secreted (C-terminus and WxG/YxxxD motifs first) would favour interaction with the phagosomal membrane; however, there is also some evidence EspB being extracted from the outer-most layer of the bacterium (174). From different experiments (**Figure 12**), it is expected that for soluble oligomers, a high amount of protein would be needed before the complex is formed. Interaction with the secretion machinery or a membrane would increase the local concentration, potentially reducing the amount of protein required and making the system more efficient.

In the third more speculative model, EspB would undergo a conformational change, as observed for some pore-forming proteins such as the amphitropic gasdermins (222). Upon proteolysis, a pre-pore ring could assemble prior to membrane insertion (223). Recently, it was suggested that the PE/PPE family of proteins could form small molecule-selective channels analogous to outer-membrane porins, allowing *M. tuberculosis* to take up nutrients through its almost impermeable cell wall (164). Despite the evidence, it remains a mystery how such soluble heterodimers would insert into a membrane. We

hypothesise that, analogous to the heterodimer EsxA/EsxB where EsxA alone can insert into a membrane in acidic conditions (212), the amphiphilic helices of either PE or PPE alone might insert into the membrane. EspB is fundamentally different from PE/PPE pairs in the sense that its PE and PPE parts are fused into a single protein, joined by one long flexible linker able to adopt multiple conformations (181). Unlike the EsxA-EsxB heterodimer, where EsxA would act independently from EsxB upon membrane insertion, the PE moiety of EspB would still be linked to its PPE counterpart even if the latter would insert itself into a membrane. We speculate that such a linker could allow EspB to form tubular-like structures while exchanging PE and PPE domains between different molecules (**Figure 18 Model 3**). As described for EspC (173) and occasionally also found in our data (**Supplementary Figure 13**), such higher-order oligomers could be a component of the secretion apparatus.

Our hypothesis that EspB acts as a scaffold or structural component of the secretion apparatus is supported by earlier findings. As ESX-1 works through a contact-dependent mechanism and not by secretion of toxins (176), it is possible that the cytotoxic effects on macrophages observed by Chen *et al.* (2013) for EspB was the result of an increment in the machinery activity. Most of the work described here favours model 1 or 2. More evidence needs to be gathered to verify or falsify any of the models (**Table 5**). Techniques like *in situ* cryo-electron tomography of infected immune cells (224) could be used to provide more relevant insight.

In summary, this study reveals factors that prime the oligomerisation of EspB and presents evidence that supports the hypothesis that EspB is a structural element of ESX-1 secretion system, possibly acting on a lipid membrane. ESX-1 is essential for the virulence of infectious mycobacteria such as *M. tuberculosis*. However, after decades of arduous research, our understanding on the structure and the mechanism of action of this system remains limited. Here we provide a structural and possibly functional understanding of an ESX-1 element that could be used to guide structure-based drug and vaccine design in order to tackle the global health threat that tuberculosis is.

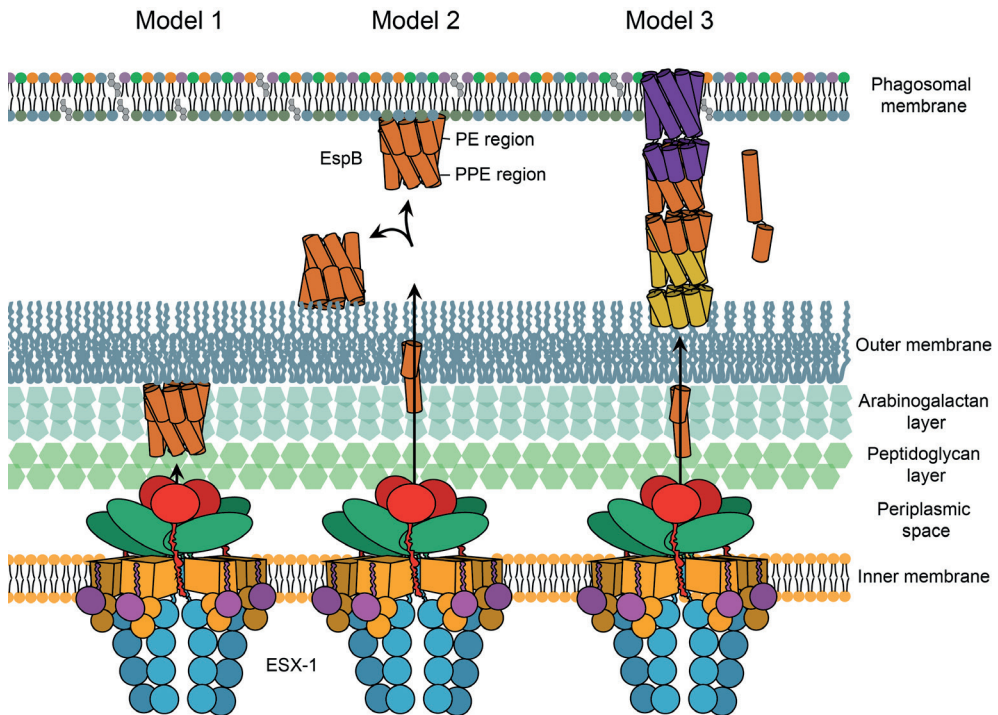


Figure 18. Putative pathways for the oligomerisation of EspB. In Model 1, EspB is cleaved in its C-terminus by the protease MycP₁ in the periplasm of mycobacteria leaving hydrophobic residues to insert into the outer membrane; an increase in the local concentration on the membrane leads to oligomerisation of EspB. In model 2, secretion of EspB across the double membrane after MyP₁ cleavage allows the protein to bind to either the phagosomal membrane or the external part of the outer membrane. In model 3, after cleavage in the periplasm and secretion to the exterior of the bacterium, EspB goes through a conformational change dissociating the PE and PPE domains and exposing hydrophobic residues that would allow the insertion into the membrane; while the PPE gets embedded into the membrane in an oligomeric form, the respective PE is able to interact with the PPE of a second molecule forming a tubular structure. Different colours are used for each heptamer-subunit. Regardless of what oligomerisation pathway EspB follows, oligomerised EspB is hypothesised to form part of the larger machinery that completes the inner-membrane complex of ESX-1.

Data availability

The final global and local B-factor sharpened maps as well as the half-maps and masks have been deposited in EMPIAR for *M. tuberculosis* and *M. marinum* under the access number EMD-13154 and EMD-13153, respectively. Likewise, refined models were deposited in the Protein Data Bank with the access code 7P13 and 7P0Z, respectively. Deposition of heptamer model containing a monomer inside (7 + 1) has been linked to the high-resolution heptamer model access code.

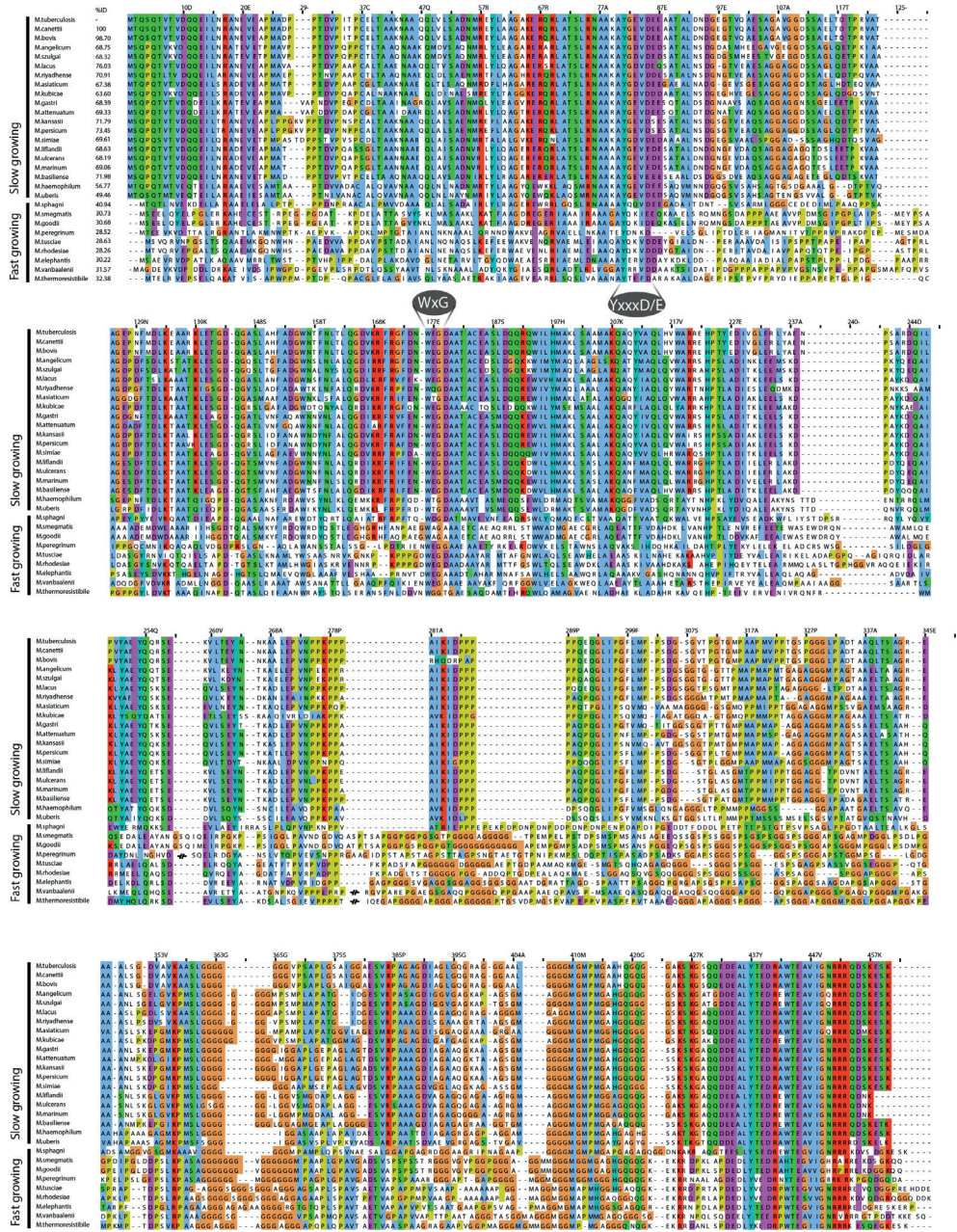
Author contribution

Abril Gijbers: Conceptualisation, Methodology, Formal analysis, Investigation, Visualisation, Writing–Original Draft, Writing–Review & Editing. **Vanesa Vinciauskaite:** Methodology, Investigation. **Axel Siroy:** Methodology, Investigation, Writing–Review & Editing. **Ye Gao:** Methodology. **Giancarlo Tria:** Methodology, Formal analysis. **Anjusha Mathew:** Methodology, Formal analysis. **Nuria Sánchez-Puig:** Methodology, Formal analysis, Writing–Review & Editing. **Carmen López-Iglesias:** Supervision, Writing–Review & Editing. **Peter J. Peters:** Conceptualisation, Supervision, Funding acquisition, Writing–Review & Editing. **Raimond B. G. Ravelli:** Conceptualisation, Formal analysis, Investigation, Supervision, Funding acquisition, Writing–Original Draft, Writing–Review & Editing.

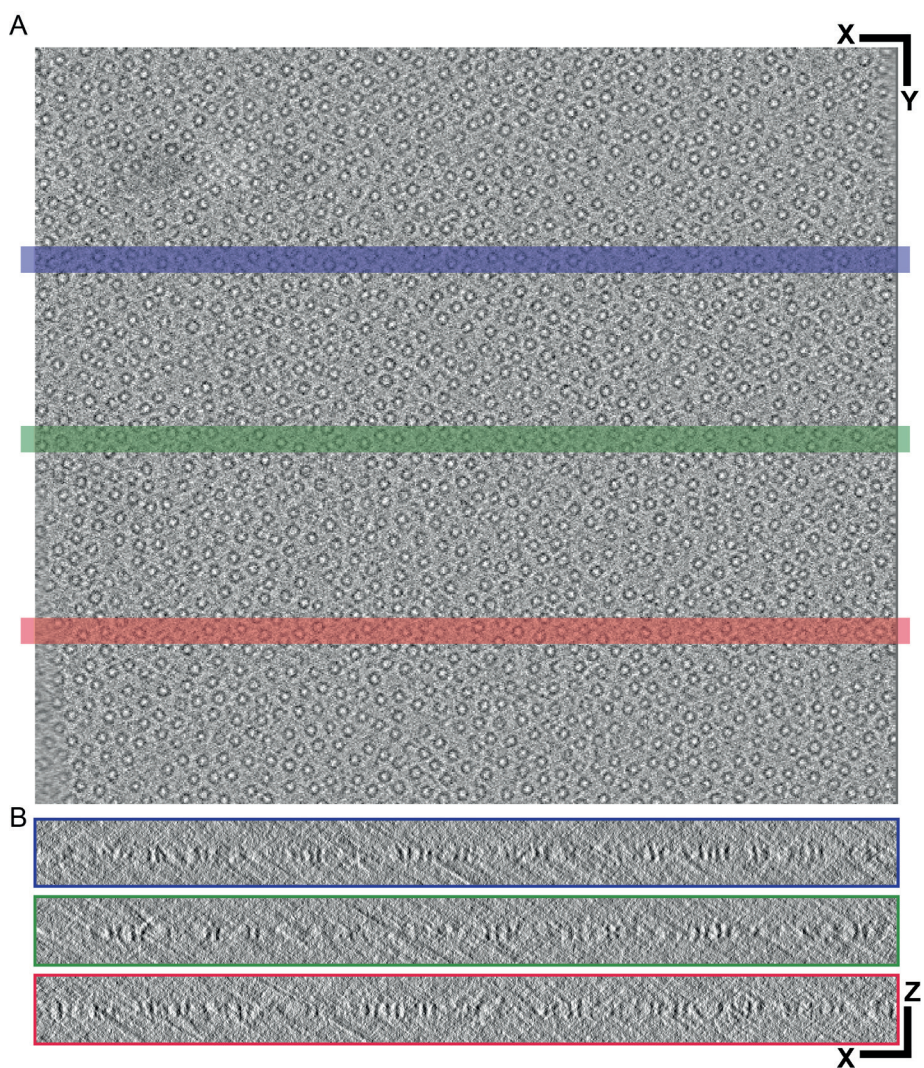
Acknowledgments and funding

We thank Paul van Schayck (UM) for SerialEM and IT support; the Microscopy CORE Lab (UM) for technical support; Yue Zhang (UM) for model refinement support; Chris Lewis (UM) for tomogram reconstruction support; Florence Pojer and Stewart Cole (Global Health Institute, Lausanne, Switzerland) for initial sample aliquots and preliminary studies; Ron Heeren and Shane Ellis (UM) for native mass spectrometry support; and Hang Nguyen (UM) for critical reading of the manuscript. Nuria Sánchez-Puig acknowledges the support from the PASPA-DGAPA programme from UNAM and CONACYT 283909. This research received funding from the Netherlands Organisation for Scientific Research (NWO) in the framework of the Fund New Chemical Innovations, numbers 731.016.407 and 184.034.014, from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 766970 Q-SORT. This research is also part of the M4I research programme supported by the Dutch Province of Limburg through the LINK programme.

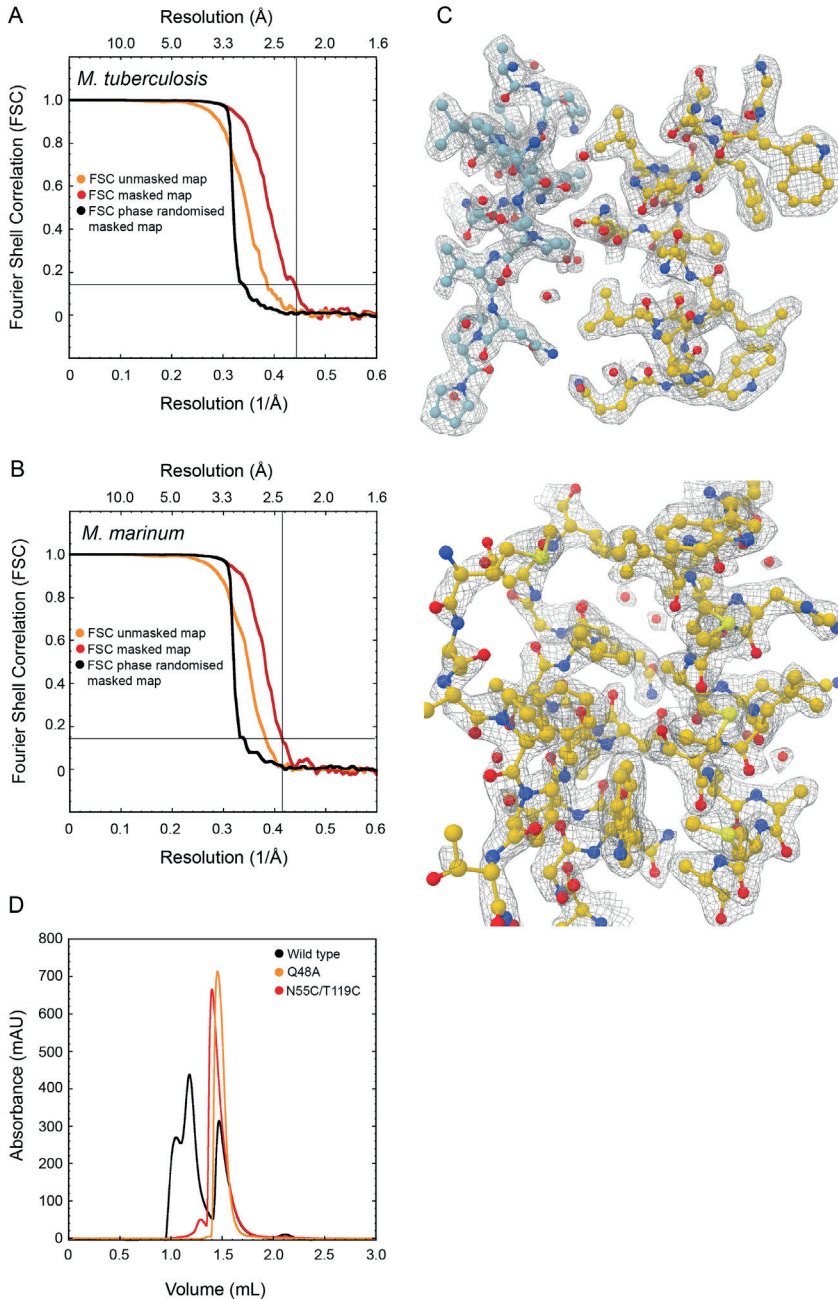
Supplementary material



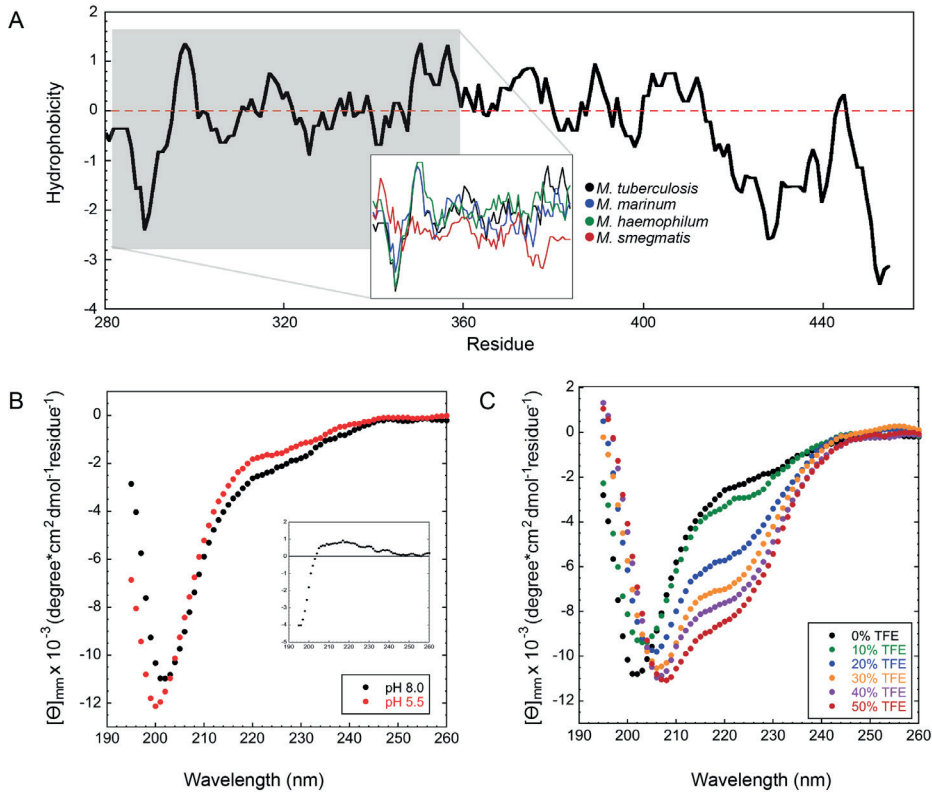
Supplementary Figure 8. Sequence alignment of EspB from different mycobacterial species. Numbering and sequence identity are based on the sequence of *M. tuberculosis*. Alignment was generated using ClustalW server, and figure was created using Jalview software (197). The colour scheme of ClustalX is used (198). Large inserts were removed, denoted by **///** symbol, to avoid empty spaces in the alignment: *M. peregrinum* 12 residues, *M. varbaanensis* 17 residues, and *M. thermoresistibile* 28 residues.



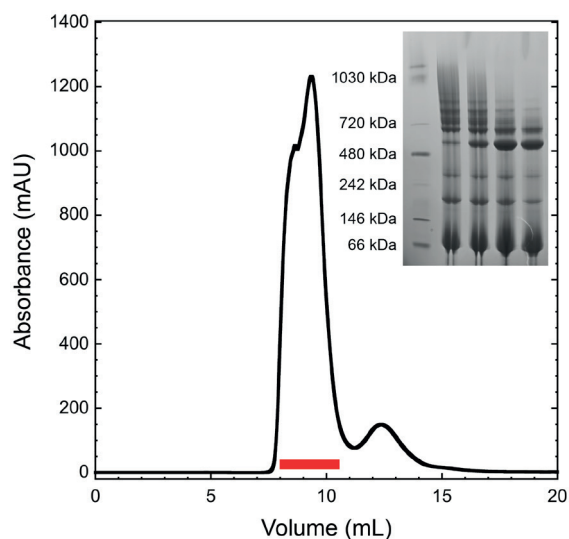
Supplementary Figure 9. EspB preferential orientation caused by an interaction to the air-water interface. (A–B) Tomogram slice of EspB₂₋₃₄₈ with 27 nm thickness in X,Y and X,Z orientation, respectively.



Supplementary Figure 10. Cryo-EM analysis of EspB structure. Gold-standard Fourier shell correlation (FSC) plot of (A) EspB₂₋₂₈₇ from *M. tuberculosis* and (B) EspB₂₋₂₈₆ from *M. marinum*. (C) Quality of cryo-EM-derived density map. Selected regions showing the fit of the derived atomic model to the cryo-EM density map (black mesh) (D) Size exclusion chromatograms of EspB₂₋₃₄₈ from *M. tuberculosis* and mutants that affect oligomerisation.



Supplementary Figure 12. Characterisation of the C-terminal region of EspB. (A) Kyte-Doolittle hydrophobicity plot of residues 280–460 of EspB from *M. tuberculosis*. Inset shows the degree of hydrophobicity of residues 280–360 from different species. Window size of 9 was used as parameter. (B, C) Far UV circular dichroism spectra of *M. tuberculosis* EspB_{279–460} at different pH and TFE concentrations. Inset in (B) shows the spectrum-difference between pH 5.5 and pH 8.0.



Supplementary Figure 13. Higher-order oligomer formation. Size exclusion chromatography profiles of EspB₂₋₃₄₈ from *M. tuberculosis* (20 mg/mL) injected onto a Superdex® 200 Increase 10/300 GL. Inset corresponds to a Blue-Native PAGE of the SEC fractions highlighted in red.

Table 2. Constructs and concentrations used in this study.

Plasmid name	Species	Gene product	Concentration used for cryo-EM experiments
pAG01	<i>M. tuberculosis</i>	6 × His-EspB 2-460	0.5 mg/mL
pAG02	<i>M. tuberculosis</i>	6 × His-EspB 2-358	0.5 mg/mL
pAG03	<i>M. tuberculosis</i>	6 × His-EspB 2-348	0.5 mg/mL
pAG04	<i>M. tuberculosis</i>	6 × His-EspB 2-287	6 mg/mL
pAG05	<i>M. tuberculosis</i>	6 × His-EspB 7-278	-
pAG06	<i>M. tuberculosis</i>	6 × His-MBP-EspB 279-460	-
pAG07	<i>M. tuberculosis</i>	6 × His-EspB 2-348 Q48A	-
pAG08	<i>M. tuberculosis</i>	6 × His-EspB 2-348 N55C/T119C	-
pAG09	<i>M. marinum</i>	6 × His-EspB 2-355	0.5 mg/mL
pAG10	<i>M. marinum</i>	6 × His-EspB 2-286	8.7 mg/mL
pAG11	<i>M. haemophilum</i>	6 × His-EspB 2-287	1 mg/mL
pAG12	<i>M. smegmatis</i>	6 × His-EspB 2-407	10 mg/mL
pAG13	<i>M. tuberculosis</i>	PE25 / 6 × His-PPE41	5 mg/mL

Table 3. Statistics of cryo-EM data collection, reconstruction and structure refinement.

	EspB 2-348 <i>M. tuberculosis</i>	EspB 2-287 <i>M. tuberculosis</i>	EspB 2-286 <i>M. marinum</i>
Grid type	Quantifoil UltraAuFoil Au200 mesh R2/2	Quantifoil UltraAuFoil Au300 mesh R1.2/1.3	Quantifoil UltraAuFoil Au300 mesh R1.2/1.3
Microscope	TFS Tecnai Arctica	TFS Krios	TFS Krios
Camera	Falcon III	K3 electron counting	K3 electron counting
Automated Data Acquisition Software	SerialEM	EPU	EPU
Nominal magnification (k \times)	110	105	105
Physical pixel size (\AA)	0.935	0.834	0.834
Exposure time (s)	43	1.8	1.8
Fluence ($e^- \text{\AA}^{-2}$)	40	40	40
Micrographs	1,457	2,334	2,421
#fractions	50	40	40
Particles	914,683	484,786	435,505
Symmetry imposed	C7	C7	C7
Average resolution (\AA)	3.4	2.29	2.43
FSC threshold	0.143	0.143	0.143
Map sharpening <i>B</i> factor (\AA^2)	-180	-80	-91
Refinement			
Initial model used (PDB entry)	7P13	4XXX	7P13
Model resolution (\AA)		2.2	2.2
FSC threshold		0.5	0.5
Model composition			
Atoms		13,727	12,873
Protein residues		1,743	1,652
<i>B</i> factors (\AA^2)		41.58	46.74
R.m.s. deviations			
Bond lengths (\AA)		0.016	0.015
Bond angles ($^\circ$)		1.057	1.084
Correlation coefficients			
Mask		0.87	0.86
Box		0.82	0.83
Validation			
MolProbity score		1.49	1.10
Clashscore		5.24	3.11
Rotamers outliers (%)		0.00	0.00
Ramachandran plot			
Favoured (%)		96.68	98.28
Allowed (%)		3.32	1.72
Outliers (%)		0.00	0.00

Table 4. Inter-subunit contacts in *M. tuberculosis* EspB and their respective residues in *M. smegmatis*.

<i>M. tuberculosis</i>	<i>M. smegmatis</i>
Q48 ↔ Q164	K45 / E168
L51 ↔ L161	A48 / S165
E58 ↔ R192	A55 / S197
E58 ↔ K168	A55 / R172
Y59 ↔ R192	T56 / S197
K65 ↔ F173 (backbone)	R62 / A178
K65 ↔ D174 (backbone)	R62 / E179
R69 ↔ E177	R66 / A182
N266 ↔ Q214	N268 / D219
P272 ↔ Q164	E274 / E168*
N274 ↔ Q193	R276 / T198
K277 ↔ D189	K279 / R194*

*Repulsive effect

Table 5. Current knowledge portraying each model.

Model	Supporting information	Information that could oppose the model
1	EspB is cleaved in the periplasm C-terminal region interacts with hydrophobic surfaces. C-terminal end is localised in the same end as the secretion motifs, suggesting secretion occurs with this region going first facilitating the interaction with membrane layer.	EspB is found in the culture filtrate. However, EspB could also secrete itself and what is found in the culture filtrate is not the total amount being secreted. Interaction with specific mycobacterial lipids are missing.
2	EspB is found in the culture filtrate C-terminal region interacts with hydrophobic surfaces. Phagosomal lumen is acidic.	Interaction with specific lipids are missing.
3	High-order oligomers shown by SEC and native-PAGE PE and PPE proteins are secreted together and after secretion they dissociate. PE-PPE linker could keep them in proximity. PE and PPE region interact through a hydrophobic interface that could interact with lipids when exposed. EsxE/EsxF oligomerise and form membrane pores	No condition was found to stabilise such higher-order oligomer. EspB is highly soluble and has not been found embedded into the membrane.

CHAPTER 5

5

Structural analysis of the partially disordered protein EspK from *Mycobacterium tuberculosis*

Abril Gijbers, Nuria Sánchez-Puig, Ye Gao, Peter J. Peters, Raimond B. G. Ravelli and Dritan Siliqi

Abstract

For centuries, tuberculosis has been a worldwide burden for human health and gaps in our understanding of its pathogenesis has hampered the development of new treatments. ESX-1 is a complex machinery responsible for the secretion of virulence factors that manipulate the host response. Despite the importance of these secreted proteins for pathogenicity, only a few of them have been structurally and functionally characterised. Here, we describe a structural study of EspK, a 74 kDa protein known to be essential for the secretion of other substrates and the cytolytic effects of ESX-1. SAXS data show that EspK is a long molecule with a maximal dimension of 228 Å. It consists of two independent folded regions at each end of the protein connected by a flexible unstructured region driving the protein to coexist as an ensemble of conformations. Limited proteolysis identified a 26 kDa globular domain at the C-terminus of the protein consisting of a mixture of α -helices and β -strands, as shown by CD and SAXS. In contrast, the N-terminal portion is mainly helical with an elongated shape. Sequence conservation suggests that this architecture is preserved amongst the different mycobacteria species, proposing specific roles for the N- and C-terminal domains assisted by the middle flexible linker.

Keywords

Disordered region, EspK, ESX-1 and SAXS.

I. Introduction

Mycobacterium tuberculosis is the causative agent of human tuberculosis (TB) and can be considered one of the most efficient pathogens in history, as it has threatened our health since the beginning of the *Homo sapiens* existence (61). Nowadays, it is estimated that one-quarter of the world's population has latent TB, from which 10 million people fell ill in 2018. Every year, 1.5 million people succumb from TB, placing it together with COVID-19 (225), as the leading cause of death from a single infectious agent (157). Although TB chemotherapy is considered a triumph of anti-infective research (37), changing the disease from fatal to curable, it is far from optimal. The toxicity of the drugs and the length of the treatment have contributed to the rise of drug-resistant strains that threaten the global health security (37). The need for new medicine that can cure or prevent TB is unquestionable, but gaps on the knowledge of mycobacteria pathogenesis hampers its development. Additionally, *M. tuberculosis* has diverged into several phylogenetic lineages with different virulence degrees hindering our further understanding of the molecular mechanisms of the pathogenesis. Nevertheless, there are essential events that contribute to the infectivity success of all pathogenic strains. Macrophages, as part of the innate immune system and first line of defence against pathogens, internalise the bacteria in a process called phagocytosis to degrade the microorganism (226). However, *M. tuberculosis* and other pathogenic species evade this fate by blocking the maturation of the phagosome and disrupting the phagosomal membrane to translocate into the cytosol of the host cell (74). This event is essential for the survival of mycobacteria, as it has been shown that non-pathogenic species are unable to translocate, leading to the lysis of the bacteria. This ability has been directly linked to the presence of the ESX-1 secretion system (81).

Mycobacteria have five different secretion systems (ESX-1 to -5) that facilitate the transport of virulence factors through a complex and almost impermeable cell wall (175). These are paralogue protein complexes with specific functions that are unable to complement each other (76). Despite their importance in the pathogenesis and survival of mycobacteria, the structure and mechanism of action remain poorly characterised. Only recently, the architecture of the inner-membrane complex of ESX-3 (166,167) and ESX-5 (170,171) was determined showing a protomer unit composed of the ESX-conserved components (EccB, EccC, EccD ($\times 2$) and EccE), which further oligomerise into dimers and hexamers. Each locus is composed of genes that code for the Ecc proteins and other proteins involved in their own expression and secretion regulation. These are (a) one pair of Esx proteins who provided the name for the secretion systems, (b) two or more proteins belonging to the PE/PPE family, (c) a mycosin protease (MycP), and (d) one or more ESX secretion-associated proteins (Esp) (175). Due to the high sequence similarity and conservation between the paralogue systems (167), it is hypothesised that the substrates are directly linked to the specific function of each secretion system. Compared to its paralogues, ESX-1 contains a large number of Esp that are essential for the mycobacteria virulence (227-230). One such protein is EspK, encoded by the *espK*

gene (also known as Rv3879c) which is missing in the attenuated strain *M. bovis* bacille Calmette-Guérin (BCG) used worldwide as vaccine against *M. tuberculosis* (77,231). The disruption of *espK* has been linked with the loss of cytolytic/cytotoxic effects in mycobacterial strains (228), as well as the decreased expression and secretion of other Esp (210,232). Physical interaction with the EspB protein has led to the hypothesis that EspK acts as chaperon of the former, but this is yet to be proven (210).

Here, we present a structural study of the EspK protein done by limited proteolysis, small-angle X-ray scattering and circular dichroism, which revealed the presence of two well-defined domains connected by an unstructured, low complexity linker. The N-terminal region consists of an elongated shape with a predicted two-helix bundle structure characteristic of the ESX-1 substrates, while the C-terminal end comprises a globular domain composed of a mixture of α -helices and β -sheets. Sequence conservation suggests that this architecture is preserved amongst the different mycobacteria species and that specific roles for the N- and C-terminal domains are assisted by the flexible linker.

2. Materials and methods

2.1 Multiple sequence alignment and secondary structure prediction

Multiple protein sequence alignment consisting of 16 representative sequences of the *Mycobacterium* genus was performed using ClustalW (198). Sequence alignments were visualised with the program Jalview 2.4 (<http://www.jalview.org/>) (197). Secondary structure prediction of EspK full-length was calculated using the Jpred 4 server (233).

2.2 Cloning, expression and protein purification of EspK constructs

The coding sequence of the full-length EspK protein was amplified from genomic DNA of *M. tuberculosis* H37Rv (BEI Resources, National Institute of Allergy and Infectious Diseases) by PCR. The DNA fragment was cloned in the pQLinkH vector (234) using the restriction sites *Nsi*I and *Hind*III. The sequence corresponding to the C-terminal region of EspK (residues 484–729) was cloned in the aforementioned vector by inverse polymerase chain reaction (235). The constructs encode a 6 \times His tag followed by a TEV protease recognition site and the corresponding EspK protein. Both proteins were expressed in Rosetta™ 2(DE3) *Escherichia coli* cells in Overnight Express™ Instant LB Medium (EMD Millipore) supplemented with 100 μ g/mL of carbenicillin and 25 μ g/mL of chloramphenicol for 50 h at 25 °C. Cell pellet was resuspended in 20 mM Tris-HCl (pH 8.0), 300 mM NaCl, 40 mM imidazole supplemented with 1 mM PMSF and 25 U/mL Benzonase® Nuclease (Merck), and lysis was performed with a C3 homogenizer (Emulsiflex, Avestin, ATA Scientific Pty Ltd.). The lysate was clarified by centrifugation at 100,000 \times g for 40 min and 4 °C. Soluble fraction was purified through a Ni²⁺ ion affinity chromatography using a 5 mL Ni-NTA Superflow column (Qiagen) and eluted with the same buffer containing 250 mM imidazole. The eluted protein was digested

with TEV protease to remove the histidine tag while dialysing overnight at 4 °C against the low imidazole buffer. The protein sample was further purified through a second Ni²⁺ ion affinity chromatography. Aliquots containing the protein of interest were pooled together and stored at –80 °C in 50 mM Tris-HCl (pH 8.0), 300 mM NaCl until further use. Sample purity was assessed by SDS-PAGE.

2.3 Limited proteolysis and N-terminal sequencing

A full-length EspK sample (4 μM) was incubated with trypsin for 30 min at different molar ratios following the Proti-Ace™ Kit (Hampton Research) instructions. Reactions were stopped by adding SDS-PAGE loading buffer and samples were resolved on a 12% polyacrylamide gel. Bands were transferred from the SDS-PAGE gel to a PVDF membrane and stained with 0.1% (w/v) Coomassie Brilliant Blue R-250, 40% methanol, and 10% acetic acid for 5 min before cutting out the section containing the cleavage product. The first ten amino acids were determined by Edman sequencing at the Plateforme Protéomique PISSARO IRIB at the Université de Rouen, France.

2.4 Circular dichroism spectroscopy (CD)

Circular dichroism scan measurements were obtained at 25 °C using a JASCO J-1500 spectropolarimeter equipped with a Peltier temperature controller (Jasco Inc., Easton, MD, United States). CD spectra were recorded using a 1 mm cuvette and a protein concentration of 1 μM for the full-length construct and 3 μM for the EspK C-terminal region in the far-UV (260–190 nm). Spectra were acquired in 50 mM phosphate buffer (pH 8.0), 50 mM NaCl, 1 mM dithioerythritol, at a 1-nm increase per step, an averaging time of 5 s, and a spectral resolution of 1 nm. Each spectrum corresponds to the average of five repetitive scans and was corrected by subtracting the CD signal of the buffer. Temperature dependence of ellipticity was followed by monitoring the signal at 222 nm from 20 to 80 °C, with a bandwidth of 1 nm, a response time of 16 s, a sample interval of 0.2 °C and a scan rate of 1 °C/min.

2.5 Small-angle X-ray scattering (SAXS) experiments

Structural characterisation of the EspK proteins was performed by SAXS coupled to an online size exclusion chromatography (Agilent 1200 HPLC, Santa Clara, CA, United States) equilibrated with 20 mM Tris-HCl (pH 8.0), 300 mM NaCl. Experiments for the C-terminal region of EspK were collected in the bioSAXS beamline B21 at Diamond Light Source, Harwell, United Kingdom. Protein sample consisting of 50 μL at a concentration of 13 mg/mL was run over a Shodex KW-403 size exclusion column at a flow rate of 0.08 mL/min. The eluted protein was directed through a 1.6-mm diameter quartz capillary cell held in vacuum. Data acquisition consisted of 580 frames (with 3-s exposure time) using a PILATUS 2M detector at a calibrated distance of 4.014 m from the sample. Images were corrected for variations in beam current, normalized for time exposure, and processed into one-dimensional scattering curves using GDA and the DAWN software (Diamond Light Source, UK). Data for the full-length EspK were collected in the bioSAXS beamline P12-EMBL at DESY Light Source, Hamburg,

Germany. Sample consisting of 50 μL at a concentration of 3.6 mg/mL was run over a Superdex[®] 200 Increase 3.2/300 (GE Health Life Science) size exclusion column attached to a FPLC–Malvern TDA system at a flow rate of 0.1 mL/min. The elution output was directed through a quartz capillary cell (50 μm thick wall and a 1.7 mm path length) held in vacuum. Data acquisition consisted of 900 frames (with 1-s exposure time) using a PILATUS 2M detector at the distance of 3.0 m from the sample. Images were corrected for variations in beam current, normalized for time exposure, and processed into one-dimensional scattering curves using integrated software at the beamline (236). Background was manually subtracted using the program CHROMIXS (237). SAXS parameters are listed in **Table 6**. Low-resolution structures were constructed by *ab initio* modelling using the program GASBOR (238) and DAMMIF (239) by aligning, averaging, and filtering ten independently calculated dummy residue and atom models by using DAMAVER program (240). Models for the different regions of EspK were predicted by the I-TASSER web server (241) considering residues 1–180 (N-terminus), 181–483 (flexible linker) and 484–729 (C-terminus). This division was based on the secondary structure prediction and limited proteolysis results. Then, multi-domain modelling was performed using the CORAL programme (242). As the middle part of EspK is predicted to be highly disordered, 40 residues at each end of this region were removed from the I-TASSER model and added by CORAL as linkers between the N-terminal and C-terminal regions.

Table 6. SAXS data collection and experimental parameters for the *M. tuberculosis* EspK proteins.

Data collection parameters		
Beamline	B21, Diamond Light Source, Harwell (UK)	P12, DESY/EMBL, Hamburg (Germany)
Detector	Pilatus 2M	Pilatus 2M
Beam size	0.2 × 0.2 mm	0.2 × 0.12 mm
Energy	12.4 kV	20.0 kV
Sample to detector distance (mm)	4,014	3,000
q range (\AA^{-1})	0.0038 – 0.42	0.0038 – 0.42
Exposure time (s)	3	1
Number of frames	580	900
Temperature (K)	293	293
Mode	SEC online	SEC online
Structural parameters	EspK C-terminus	EspK full-length
Concentration range (mg/mL)	13	3.6
q Interval for Fourier inversion (\AA^{-1})	0.011 – 0.341	0.008 – 0.110
R_g [from P(r)] (\AA)	22.33 ± 0.28	53.89 ± 0.46
R_g [from Guiner approximation] (\AA)	21.70 ± 1.22	53.53 ± 1.25
sR_g limits [from Guiner approximation]	0.19 – 1.30	0.60 – 1.28
Dmax (\AA)	83	228
Porod volume estimate (nm^3)	52	179
GASBOR excluded volume (nm^3)	42	130

Data collection parameters		
Molecular Mass (kDa)		
From Porod ($\times 0.53$)	28	94.5
From excluded volume ($\times 0.5$)	21	65
From sequence	30.3	78.3
Modelling		
Ambiguity score	0.0 (potentially unique)	1.74 (might be ambiguous)
Resolution (FSC) (Å)	25 ± 3	40 ± 3
SASBDB code	SASDKR4	SASDKQ4
Software employed		
Primary data reduction	DAWN pipeline (Diamond Light Source, UK)	Local pipeline (DESY Light Source, Germany)
Data processing	ScÅtter v3.1q, ATLAS	
<i>Ab initio</i> modelling	DAMMIF, GASBOR	
Validation and averaging	DAMAVR	
Computation of model intensities	CRY SOL	

$q = 4\pi\sin(\theta)/\lambda$, where 2θ is the scattering angle and λ is the wavelength of the incident X-ray beam

3. Results and discussion

3.1 Sequence conservation of EspK highlights discrete regions

Limited information exists on the structural features of EspK. Sequence analysis of the *M. tuberculosis* protein displays a region between residues 182 and 437 with an unusual large content of alanine and proline. Proline is an atypical amino acid due to its cyclic side chain that restricts the backbone conformation and is unable to act as hydrogen donor: for this reason, proline is known as the “helix-breaker” (243). Based on this observation, EspK could be divided in three discrete regions: an N-terminal domain containing a WxG motif commonly found in ESX-1 proteins (244), an A/P-rich middle region, and a C-terminal domain. Comparison of the amino acid composition present in these regions with a subset of proteins taken from the protein data bank (PDB) (207) representing structured proteins, showed a similar distribution of amino acids for the N- and C-terminal ends. In contrast, the middle region lacks residues that promote order and has 5-fold more prolines than average structured proteins (**Figure 19**). It is noteworthy that this high proline content is also unusual for disordered proteins, as it has a 3-fold enrichment compared to the average content of intrinsically unstructured proteins (207). Secondary structure prediction of EspK suggested that the N- and C-terminal ends are folded regions connected by a large unstructured linker (**Figure 20**), which is in agreement with the amino acid analysis showing an average content of order promoting residues. To determine if this organisation is peculiar for the *M. tuberculosis* EspK or whether it is a general characteristic of this protein, we carried out a sequence alignment of different species from the genus *Mycobacterium*. Analysis revealed that the N- and C-terminal domains have a high sequence conservation amongst the species, while the middle region is variable in length and sequence. Despite the sequence variation in the middle region, it retains a characteristic high alanine and proline content

that would preserve its physicochemical properties (Figure 20 and Supplementary Figure 14).

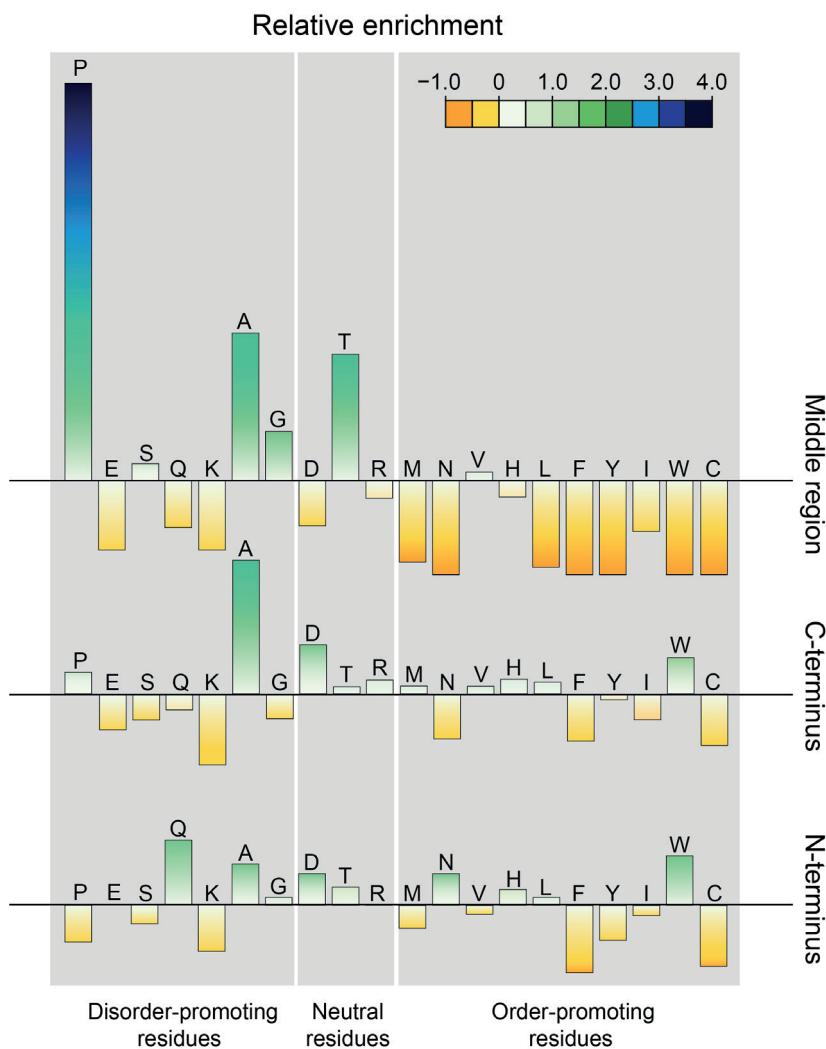


Figure 19. Relative amino acid enrichment in the different regions of EspK compared to proteins deposited in the protein data bank (207). Enrichment calculated as $(AA_{\text{EspK}} - AA_{\text{PDB}}) / AA_{\text{PDB}}$, where AA_{EspK} is the content of an amino acid in EspK, and AA_{PDB} is the corresponding amino acid content of a subset of structured proteins. Amino acids are distributed according to their (dis)order-promoting potential.

3.2 EspK contains a trypsin-resistant domain

To confirm the existence of the two well-defined domains, EspK full-length from *M. tuberculosis* H37Rv was recombinantly expressed in *E. coli* and further purified. Interestingly, EspK migrates anomalously in SDS-PAGE, appearing at a higher molecular weight than the expected 74 kDa (**Figure 21**). This has been observed in high-proline content molecules whose difference in migration is directly proportional to the percentage of prolines in the amino acid sequence (245). The limited proteolysis of EspK at different molar ratios of protease displayed the presence of a digestion-resistant fragment, with an apparent molecular weight of 26 kDa (**Figure 21**). The N-terminal sequencing showed that this fragment starts at Gly484 and based on the molecular weight calculated from the amino acid sequence (26.4 kDa), it extends all the way to the C-terminus of the protein. This result agrees with the secondary structure prediction and the conservation of the respective region, suggesting that it corresponds to a folded domain of EspK. Based on the amino acid sequence, the N-terminal domain would represent an 18-kDa fragment; however, no fragments were found below the 26-kDa molecular weight marker (**Figure 21**). This suggests that the N-terminal domain consists of either unstructured regions or contains exposed accessible digestion sites that destabilise the structure.

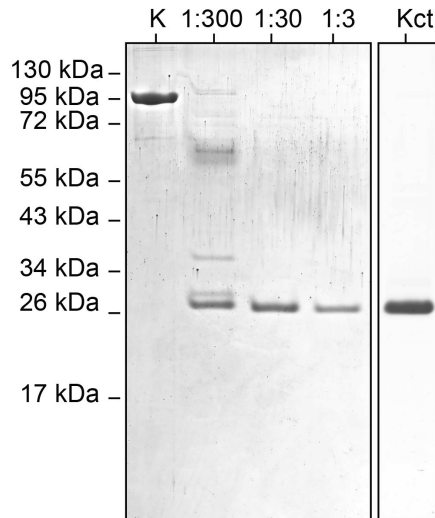


Figure 21. Identification of a trypsin-resistant C-terminal region on EspK by limited proteolysis. First lane (K) – full-length EspK, lanes 2–4: EspK digestion using different molar ratios of trypsin, lane 5: trypsin-resistant EspK fragment recombinantly expressed.

3.3 Characterisation of secondary structure by circular dichroism

The secondary structure content of EspK and its C-terminal region was evaluated by circular dichroism (**Figure 22A**). Closer inspection of the far-UV CD spectrum of the C-terminal fragment shows that it consists of the typical signature of a mixed secondary structure containing both α -helices and β -sheets with two negative minima at 210 and 220 nm and a positive maximum at 197 nm. These shifted negative minima result from the combined contribution of the characteristic negative bands for α -helices present at 208 and 222 nm and those of β -sheets present between 210 and 225 nm. In addition, the presence of the intense positive signal between 195 and 200 nm is characteristic of proteins containing β -sheets. Considering that the full-length construct comprises that of the C-terminal region, this CD spectrum also resembles a mixture of β -sheets and α -helices with a well-defined negative minimum at 208 nm instead of that at 210 nm, suggesting a larger content of α -helices and a plateau between 215 and 220 nm resulting from the β -sheets contribution. The positive maxima displaced toward the lower wavelength (193 nm) also indicates a larger content α -helices contributed most likely from the N-terminal region of EspK. The overall intensity of the full-length EspK spectra is smaller (absolute value) compared to that of the C-terminal region, implying that there is less secondary structure content per residue in the full-length protein and thus a larger content of unstructured regions. The thermal stability of EspK was determined by monitoring the ellipticity at 222 nm as a function of temperature (**Figure 22B**). The thermal denaturation for the EspK full-length and C-terminal region corresponded to an irreversible process as the refolded spectra did not overlay with the corresponding one before the heat treatment (data not shown). Data for both constructs described a single broad transition comprising approximately 45 °C from the onset of the denaturation process until reaching the denatured state. Despite the lack of clearly identifiable intermediate transitions, this long gradual denaturation process suggests the presence of intermediate states with similar secondary structure content, which cannot be evidenced by this technique. The apparent melting temperatures corresponded to 45 and 52 °C for the full-length EspK and the C-terminal region, respectively. The decrease in the melting point for the full-length protein compared to that of the C-terminus suggests that these two regions behave independently, as an interaction between them would increase the stability of the protein and thus the melting temperature of the full-length protein.

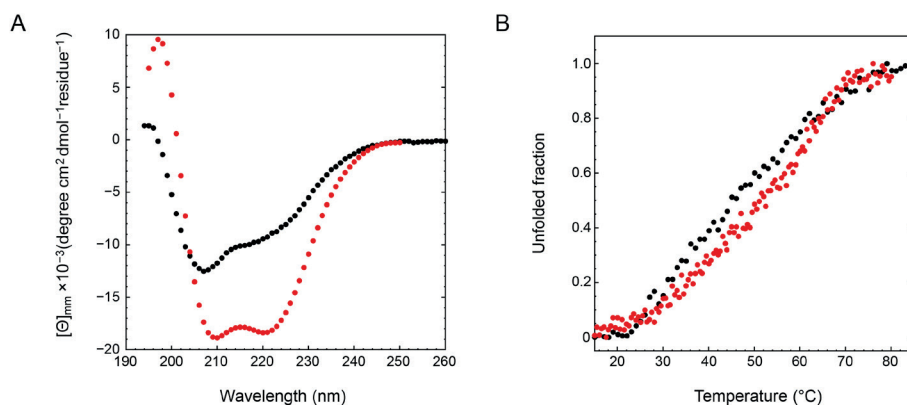


Figure 22. Circular dichroism of the EspK full-length protein (black) and its C-terminal region (red). (A) Far-UV spectra. (B) Thermal denaturation.

3.4 Structural analysis by small-angle X-ray scattering

We tried to gain insight into the tertiary structure of EspK by performing SAXS experiments. All parameters are listed in **Table 6**. The one-dimensional SAXS experimental curves were used to judge the quality of the data and obtain basic structural information related to the size and shape of the EspK protein and its C-terminal domain (**Figure 23A**). SAXS curve analysis confirmed that the proteins were monomeric, as the calculated molecular weight from the Porod plot corresponds to the expected value calculated from the amino acid sequence (**Table 6**). The full-length EspK and its C-terminal domain have a radius of gyration (R_g) of 53.53 Å and 21.70 Å, respectively, which are calculated from the slope of the Guinier plot (246), with a maximum dimension (D_{max}) of 228 Å and 83 Å obtained from the Pair-Distance Distribution Function ($P(r)$) (**Figure 23B**). Based on these two parameters, it is expected that the full-length EspK represents an elongated molecule, while the C-terminal domain comprises a globular shape.

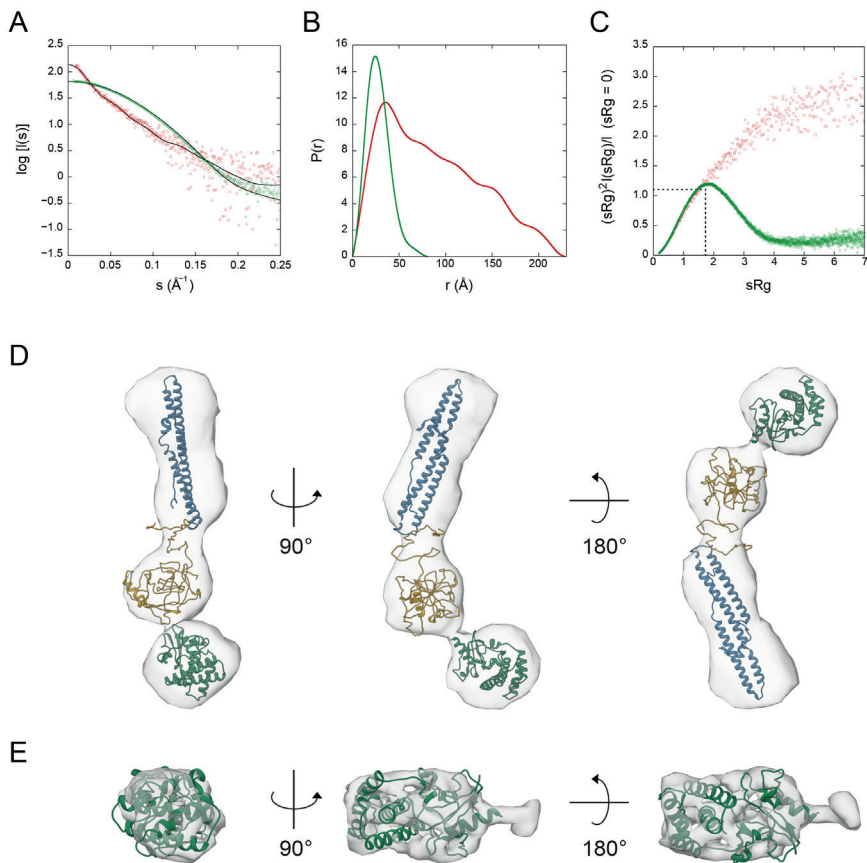


Figure 23. SAXS data for the EspK full-length protein (red) and its C-terminal domain (green). (A) Fit of the calculated SAXS scattering curves compared to the experimental scattering signal. (B) Pair distribution function plot. (C) Dimensionless Kratky plot. The intersection of the dotted black trace corresponds to the value for the reference protein bovine serum albumin (BSA). (D) Ab initio molecular envelope of the full-length EspK showing the fit of the models corresponding to the N-terminal (residues 1–180 in blue), the flexible linker (residues 181–438 in yellow), and C-terminal region (residues 484–729 in green). (E) Ab initio molecular envelope of the C-terminal region of EspK showing the fit of the corresponding model.

Analysis of the Porod exponent, a quantitative measurement of the increase of compactness of a protein (247), confirmed the flexible nature of the full-length protein in comparison with its C-terminal domain with values of 2.4 and 4.0, respectively. This observation was in agreement with the corresponding dimensionless Kratky and $P(r)$ distribution plots, where the C-terminal domain behaved as a globular and compact protein similar to the bovine serum albumin (BSA) used as a standard protein, compared to the highly flexible and elongated full-length EspK that seems to attain

multiple conformations (**Figure 23B–C**). To inquire on the tertiary structure of the proteins, we obtained the SAXS *ab initio* models of the EspK full-length and C-terminal domain using DAMMIF and GASBOR software programs, respectively, and compared them with their corresponding I-TASSER predicted models (**Figure 23D–E**). The ambiguity of the obtained reconstructions, as estimated by the program AMBIMETER (248) were 1.74 for the full-length EspK and 0.0 for the C-terminal region (**Table 6**). These values indicate that the 3D reconstruction for the full-length protein might be ambiguous, whereas that of the C-terminal is not. The model obtained for the full-length EspK represents an average of different conformations as suggested by the Pair-Distance Distribution Function ($P(r)$), resulting in such ambiguity. The resolution of the models, as determined by Fourier Shell Correlation (FSC) (249) is 40 Å and 25 Å for the EspK full-length and C-terminal domain, respectively. The I-TASSER prediction for the EspK full-length protein resulted in an unstructured model. As previous data pointed to the C-terminal end being folded, we performed independent predictions for the three regions of the protein (**Supplementary Figure 15**). In agreement with the secondary structure prediction and circular dichroism, the model for the N-terminal region consisted of only α -helices, and seven out of the best ten templates used to build it corresponded to Pro-Pro-Glu (PPE) proteins. These proteins are characterised by its Pro-Pro-Glu (PPE) motif and are known to be secreted together with their PE protein pair by ESX-1 and its paralogues (76,250). Despite the resemblance, the EspK N-terminal domain does not contain the PPE motif; instead, it only contains the WxG motif needed for the secretion of the protein. The top five models predicted for the middle region by I-TASSER resulted in a disordered region with different spatial distributions but no secondary structure content. In the case of the C-terminal domain, all predicted models displays a globular and compact protein composed of a mixture of α -helices and β -strands as also suggested by the circular dichroism results. For this domain, the correlation of $c^2 = 11.4$ between the experimental SAXS curves and the calculated one from the *ab initio* model (**Supplementary Figure 15C**) suggests a good level of confidence in the correctness of the model, which was further confirmed by calculating the normalised spatial discrepancy (251) between the SAXS and I-TASSER model (NSD = 2.48). For EspK full-length, the low-resolution envelope (DAMMIF) obtained is well described by the I-TASSER models from the three regions (**Figure 23D**). These models were subsequently used to build a multi-domain model with the CORAL program, which describes the experimental EspK full-length SAXS data confidently ($c^2 = 1.39$) (**Figure 24**).

The protein with the closest structural similarity found in the protein data bank which was used to build the I-TASSER model of EspK C-terminus corresponded to residues 184–410 of the Rv3899c protein from *M. tuberculosis* (252). This is a protein of unknown function found in the bacteria culture filtrates (253) and infected guinea pig lungs (254). It is noteworthy that the corresponding gene is located next to the *esx-2* locus, which is a paralogue of the ESX-1 secretion system to which EspK belongs. Limited information exists on ESX-2 but its exclusive presence in slow-growing mycobacteria (255), a group

of mycobacteria characterised to be pathogenic, implies a possible involvement in this process. Until now, from all five paralogues, ESX-1 is the only secretion system that contains multiple Esp-proteins. Based on the similarity found by I-TASSER between EspK and Rv3899c, and its location in the genome, it might be possible for Rv3899c to be secreted by ESX-2.

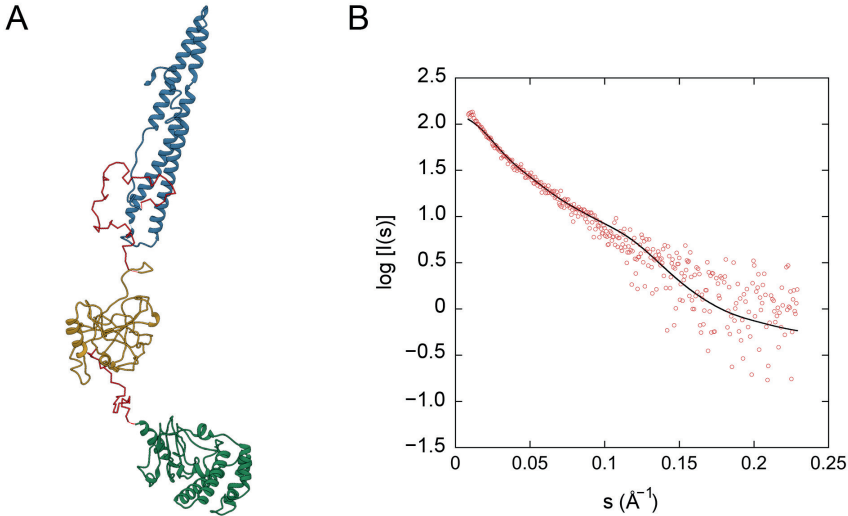


Figure 24. Multi-domain modelling by CORAL of the full-length EspK. (A) Multi-domain model of the N-terminal domain (blue), the middle region (yellow) and the C-terminal domain (green) connected by linkers (red). (B) Fit of the calculated SAXS scattering curves from the multi-domain model (black line) compared to the experimental scattering signal (red).

To this day, no function has been described for EspK except for a hypothetical role as a chaperone of EspB based on their interaction and that with the ESX-1 core protein EccCb₁ (210). Intrinsic disordered regions provide proteins with a unique ability to interact with several unrelated binding partners. With this in mind and based on the thermal denaturation results that suggest no interaction between the EspK domains, it is plausible that EspK assists other protein substrates such as EspB to be secreted, *e.g.*, one domain could interact with the ESX-machinery, while the other one interacts with a substrate. EspB is homologous to the PPE-PE proteins, which are chaperoned by EspG (196). The current model of EspK does not share structural similarity with EspG (256), implying that the system could make use of different secretion mechanism with specific chaperones. High-resolution structures are needed to test this hypothesis, leading to a better understanding of the action mechanism of ESX-1 secretory system.

5. Conclusion

Using a combined approach of SAXS, CD, and limited proteolysis, together with structure predictions, we show that EspK is formed by two discrete independent domains connected by a partially disordered region, with an unusual large content of proline that confers a high degree of flexibility to the protein. This architecture is conserved along the *Mycobacterium* genus, suggesting a specific function for the N- and C-terminal domain assisted by the flexible linker. To confirm the mechanism of action of such domains, high-resolution studies are needed in the presence of the binding partners.

Data availability

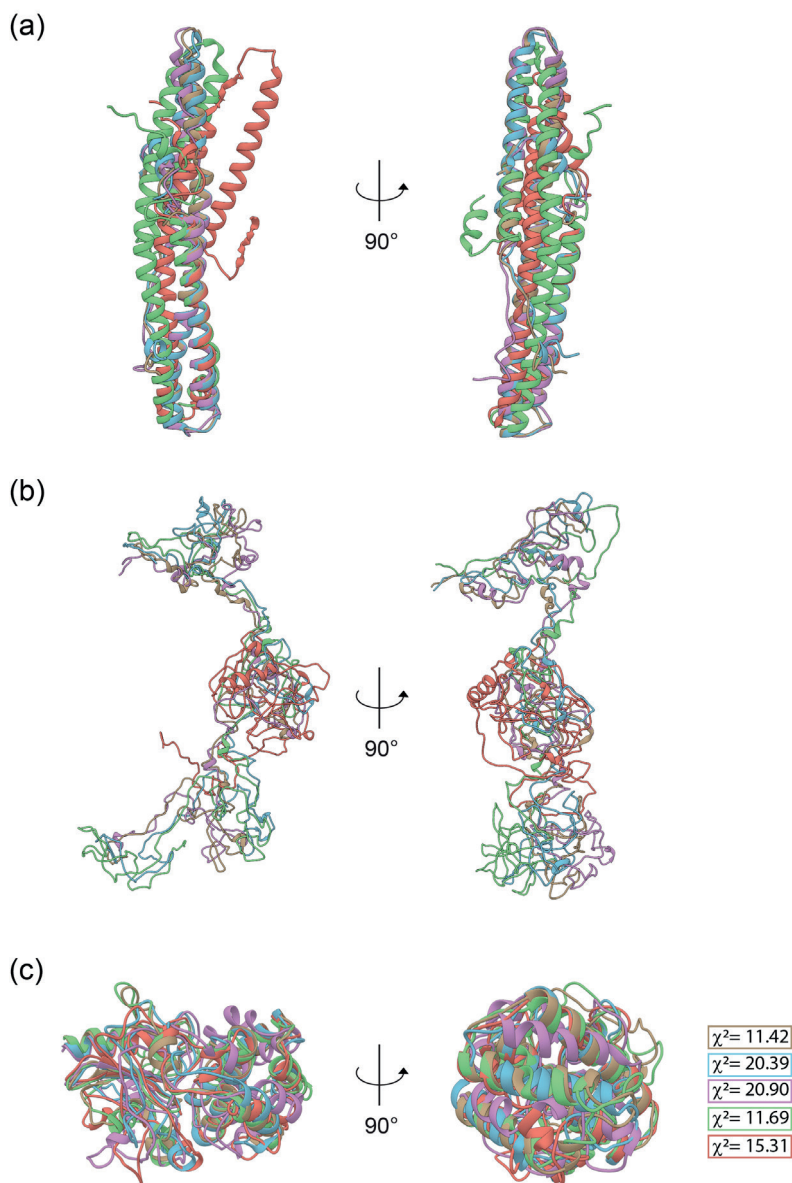
The SAXS data and models for the EspK full-length protein and its C-terminal domain have been deposited in the Small Angle Scattering Biological Data Bank (SASBDB (257)) with the access code SASDKQ4 and SASDKR4, respectively (www.sasbdb.org).

Author contribution

Abril Gijsbers: Conceptualisation, Methodology, Formal analysis, Funding acquisition, Writing–Original Draft, Writing–Review & Editing. **Nuria Sánchez-Puig:** Conceptualisation, Methodology, Formal analysis, Writing–Original Draft, Writing–Review & Editing. **Ye Gao:** Methodology. **Peter J. Peters:** Conceptualisation, Supervision, Funding acquisition, Writing–Review & Editing. **Raimond B. G. Ravelli:** Conceptualisation, Supervision, Funding acquisition, Writing–Review & Editing. **Dritan Siliqi:** (in bold) Conceptualisation, Methodology, Formal analysis, Funding acquisition, Writing–Original Draft, Writing–Review & Editing

Acknowledgments and funding

We thank Alexey Kikhney from DESY Light Source, and Nathan Cowieson and Nikul Khunti from Diamond Light Source for their assistance in the preparation of the SAXS experiments. This research is part of the M4I research programme supported by the Dutch Province of Limburg through the LINK programme. AG and RBGR acknowledge funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 766970 Q-SORT (H2020-FETOPEN-1-2016-2017), and NSP acknowledges the support from the PASPA-DGAPA program from UNAM and CONACYT 283909. The SAXS experiments were performed by proposals MX21741-1 SM21035-161 and SAXS-805 iNEXT 6260 on the beamlines B21, Diamond Light Source and P12-EMBL at DESY Light Source, respectively.



Supplementary Figure 15. Tertiary structure prediction of the different regions of EspK by I-TASSER. Top five models predicted for the N-terminal domain (A), the middle region (B) and the C-terminal domain (C). Each model is represented by different colour. χ^2 represents the correlation between the experimental SAXS data and that calculated for the different I-TASSER models of EspK C-terminal domain.

CHAPTER 6



Structure of EspB–EspK complex: the non-identical twin of the PE–PPE–EspG secretion mechanism

Abril Gijbers, Mathias Eymery, Isabella Menart, Andrew McCarthy, Dritan Siliqi, Vanesa Vinciauskaite, Ye Gao, Peter J. Peters and Raimond B. G. Ravelli

Abstract

Pathogenic species from the *Mycobacterium* genus are responsible for a number of health conditions in human and animals that threaten the health security and economy worldwide. Mycobacteria have up to five specialised secretion systems (ESX-1 to -5) in order to transport virulence factors across their complex cell envelope that allows the manipulation of their environment. In pathogenic species, these virulence factors influence the immune system's response and are responsible for membrane disruption and cell death. While structural details of these secretion systems have been recently described, there are gaps in the secretion mechanisms of most of the substrates, hampering a proper understanding of the systems. Here, we describe the structure of EspB bound to its chaperone EspK, which revealed a new fold. EspB interacts with EspK through its helical tip, similar to its homologues PE-PPEs and their chaperone EspG. This interaction disrupts EspB oligomerisation ensuring the protein to be in a secretion-competent. However, conservation of the interaction in *M. smegmatis*, where no EspB oligomer is present, suggests that the main function of EspK is to still to be characterised. Analogue to PE-PPE and their chaperone EspG, the structure presented in this study suggests a different secretion mechanism in ESX-1, where EspK facilitates the secretion of EspB.

Keywords

Chaperone, EspB, EspK, ESX-1, mycobacteria

I. Introduction

Human tuberculosis (TB) is a widespread infectious disease, caused by the bacillus *Mycobacterium tuberculosis*. In 2019, 10 million cases and 1.4 million deaths have been reported due to TB, making it the leading cause of death from a single infectious agent and a colossal burden on global human health (157). The World Health Organization foresees that the COVID-19 pandemic has had a major impact on the ongoing TB crisis, decreasing TB detection and treatment by as much as 25-50% in several high-burden countries (157). Additionally, both respiratory diseases seem to negatively impact each other, with COVID-19 case fatality rates likely being increased among TB patients (258), whilst TB seems to increase susceptibility and severity of COVID-19 disease (52). Furthermore, COVID-19 is prevalent in high TB burden countries, deteriorating socioeconomic conditions that further worsen TB epidemics (157). Moreover, increasing occurrence of drug-resistant *M. tuberculosis* strains impedes treatment of TB with currently available drugs (259). Thus, the global TB crisis calls for development of new drugs that efficiently target *M. tuberculosis*, which requires identification of treatment targets in the *M. tuberculosis* infection cycle.

Several important aspects of *M. tuberculosis* infection have been elucidated. A unique aspect of *M. tuberculosis* infection is the ability to escape lytic degradation in the phagolysosomes of alveolar macrophages (74,82), which phagocytose tubercle bacilli upon infection (226). The lack of this ability in non-pathogenic mycobacteria leads to lytic degradation (81), indicating that this event is crucial for intracellular survival and persistent infection by pathogenic mycobacteria, including *M. tuberculosis* (74). The translocation of tubercle bacilli occurs through the permeabilisation of the host phagosome, which has been directly linked to the ESAT-6 secretion systems (ESX), specifically ESX-1 (81,260).

ESX secretion systems are prominent effectors of mycobacterial virulence, and participate in substrate secretion across the mycobacterial envelope (260). Five paralogous ESX systems (ESX-1 to -5), unable to complement each other's function, have been identified across mycobacteria (76). While ESX-1 has been implicated in substrate transport across the inner membrane (76), and directly linked to phagosome permeabilisation by a contact-dependent mechanism (81,176), the details of its structure and function remain elusive. Since the five ESX systems show high similarity in genetic organisation and sequence conservation (261), it could be expected that they are unified by a similar architecture. Hence, the recent high-resolution structures of ESX-5 (170) and ESX-3 (166,167) might provide clues into the architecture and function of ESX-1.

The core secretion machinery, formed by ESX-conserved components (Ecc) EccB, EccC, EccD, and EccE (76,175), and stabilised by the subtilisin-like serine protease mycosin (MycP) (172), secretes several classes of substrates (262). These include proteins of the WxG100 superfamily composed of 100 amino acids and a highly conserved WxG

motif (263); proline-glutamate (PE)-rich and proline-proline-glutamate (PPE)-rich proteins secreted in pairs (264); and alanine-rich proteins (265), the transport of which is facilitated by a conserved YxxxD/E motif (266). It is currently known that ESX-1 is located in the inner membrane, but it is expected to extend to the outer membrane, although this remains to be determined (262). Several ESX-1 substrates have been suggested to form the missing components of the extended ESX-1 machinery (173,267), but additional studies are needed in order to have a more complete knowledge of ESX-1 secretion.

One of these substrates, the ESX-1 secretion-associated protein B (EspB), has been postulated as a structural component of the extended ESX-1 machinery (267). Deletion of *espB* attenuates *M. tuberculosis in vivo* growth and survival, and interferes with secretion of ESX-1 proteins *EsxA/B* (210,228), which emphasises that EspB is an essential component of ESX-1. EspB belongs to the PE/PPE family, sharing structural homology (179,180), however, different from the other protein members that form heterodimers, EspB is a single polypeptide that contains the bipartite secretion signal sequence, formed by WxG and YxxxD/E motifs (180). The latter, contained within the “export arm”, is hypothesised to be involved in the interaction to EccCb₁, targeting EspB for secretion (180). Once secreted across the inner mycobacterial membrane, EspB oligomerises into a cylinder-like heptamer (179,180), favoured by the removal of its C-terminal region and the acidic pH environment of the phagosomal lumen (267). Oligomerisation of EspB is driven by the PE/PPE N-terminal region (180), and mediated by the ESX-1 core complex MycP₁ protease, which cleaves the full-length 48-kDa precursor EspB monomer at its C-terminal domain (183), leaving a mature 38-kDa isoform to be secreted (210). The resulting EspB heptamer has been proposed to form channel-like structures, enabling other ESX-1 secreted substrates to pass through, beyond the inner mycobacterial envelope (267). Improving our understanding of EspB secretion and function, as well as the regulatory mechanism behind its secretion can thus reveal how EspB completes the ESX-1 core secretion machinery and participates in substrate secretion.

To date, it is recognised that the secretion signals are not enough to target the substrate to its specific secretion machinery (266), out of the five ESX systems, and that chaperones are needed to perform this role (165,196). Although it remains to be characterised, is expected that EspB secretion, like other substrates, relies on a chaperone to deliver it to the secretion machinery. As EspB secretion does not depend on EspG (210), unlike its structurally similar PE/PPE pairs, it is likely that EspB utilizes a different secretion pathway. EspB interacts with the EspK protein (210), which is encoded by a gene in the RD1 locus downstream of EspB (260). It has been suggested that the EspK C-terminal domain interacts with the EspB N-terminus in the cytosol, which is then delivered to EccCb₁ for secretion (210). Structural analysis of EspK at low resolution combined with structure predictions identified the presence of N- and C-terminal domains with independent behaviour, connected by an unstructured low-complexity (268). However,

the detailed structure of EspK and its role have yet to be determined, especially its interaction with EspB, and possibly other ESX-1 proteins. Thus, a high-resolution structure of EspK and characterisation of its interaction with EspB are necessary to consolidate EspK's role in ESX-1-mediated secretion in mycobacteria.

Here, we present the crystal structure of the EspK C-terminal domain bound to EspB from *M. tuberculosis*, obtained through macromolecular crystallisation, and show a new fold within the structure. We report the structure of the complex between *M. tuberculosis* EspB and EspK proteins, showing that EspK binds to the EspB helical tip, analogous to PE–PPE–EspG interactions. One of the regions of EspK binding to EspB is the PE/PPE loop, which is known to undergo conformational changes that allow for EspB oligomerisation, post-secretion. Thus, EspK binding to EspB prevents EspB oligomerisation not only by steric hindrance, but also by binding the PE/PPE loop, whose movement is required to allow for EspB oligomerisation. Electron microscopy in cryogenic conditions and size exclusion chromatography experiments confirmed that fewer EspB oligomers were found when EspK was present, suggesting that EspK acts as a chaperone by preventing a premature association, leaving EspB in a secretion-competent state. However, as the interaction is conserved in *M. smegmatis*, where EspB does not oligomerise, it is likely that EspK carries out another function that is still to be characterised. In analogous to what we see in the PE–PPE–EspG secretion mechanism, and based on our results, we propose that EspB is secreted through a new secretion route for the ESX-1 system, mediated by EspK, where the chaperone keeps EspB in a secretion-competent state and, most likely, directs the protein to the machinery.

2. Materials and methods

2.1 Protein expression and purification

Different constructs used in this study are listed in **Table 8**. EspK constructs from *M. smegmatis* and *M. marinum* were synthesised and codon optimised for expression in *Escherichia coli* (Eurofins). They were designed based on the *M. tuberculosis* construct. EspB constructs from all organisms, as well as EspK from *M. tuberculosis* were expressed in Rosetta™ 2(DE3) *E. coli* and purified by conventional chromatographic techniques as described in (267,268). For the codon-optimised constructs, proteins were expressed in C41(DE3) *E. coli* cells following the aforementioned procedure.

2.2 Analytical size exclusion chromatography

Samples were dialysed overnight in 20 mM Tris-HCl (pH 8.0), 150 mM NaCl and different concentrations of protein were loaded onto a size-exclusion Superdex® 200 Increase 3.2/300 column (GE Healthcare Life Science) and eluted with the same buffer at a flow rate of 50 µL/min.

2.3 Small-angle X-ray scattering (SAXS) experiments

SAXS experiments on the EspB constructs (**Table 8**) were originally collected in the bioSAXS beamline P12-EMBL at DESY Light Source (Hamburg, Germany) (269). Sample injections of 50 μL were run on a Superdex[®] 200 Increase 5/150 GL size exclusion column attached to a FPLC–Malvern TDA system at a flow rate of 0.1 mL/min. The elution output was directed through a quartz capillary cell (50- μm thick wall and a 1.7-mm path length) held in vacuum. Data acquisition consisted of 900 frames (with 1 s exposure time) using a PILATUS 2M detector at the distance of 3.0 m from the sample. Images were corrected for variations in beam current, normalized for time exposure, and processed into one-dimensional scattering curves using integrated software at the beamline (236). Background subtraction was done with the program CHROMIXS (237).

SAXS data on EspB and EspK were collected in the bioSAXS beamline B21 at Diamond Light Source (Harwell, UK) (270). Experiments were performed coupled to a Shodex KW-403 SEC column equilibrated with 20 mM Tris-HCl (pH 8.0), 300 mM NaCl. Injections of 50 μL were made. EspB–EspK mixtures were used in 1:5 molar ratio. The eluted protein was directed through a 1.6-mm diameter quartz capillary cell held in vacuum. Data acquisition consisted of 580 frames (with 3-s exposure time) using a PILATUS 2M detector at a calibrated distance of 4.014 m from the sample. Images were corrected for variations in beam current, normalized for time exposure, and processed into one-dimensional scattering curves using GDA and the DAWN software (Diamond Light Source, UK). SAXS parameters are listed in **Table 9**. Missing flexible regions were added to the crystal models using the CORAL program to improve fitting (242).

2.4 Structure determination by macromolecular crystallography

2.4.1 Crystallisation conditions

EspB (residues 2–348) or EspB (residues 2–300) were incubated together with EspK (residues 484–729) from *M. tuberculosis* for 30 min at 4 °C. The EspB–EspK protein complex was then isolated by gel-filtration using a Superdex[®] 200 Increase 10/300 GL column, and concentrated to 8 mg/mL in 50 mM Tris-HCl (pH 8.0) and 300 mM NaCl. Protein concentration of the complex mixture was determined by 280 nm absorbance using an extinction coefficient of 63 495 $\text{M}^{-1} \text{cm}^{-1}$ theoretically calculated with ExPASy.

Initial crystallisation conditions were identified at the High Throughput Crystallisation Facility, EMBL Grenoble, France. The best condition for the EspB (residues 2–348) and EspK (residues 484–729) complex was obtained from the JCSG-Plus[™] (Molecular Dimensions) condition F10 (1.1 M sodium malonate, 0.1 M HEPES, pH 7.0, and 0.5% (v/v) jeffamine ED-2003), with a high solvent content (>70 %) and anisotropic diffraction to ~ 4.6 Å observed. Manual optimisation and screening of various additives, with PEG3350 in particular identified, improved the crystal quality. Later, because the C-terminal region of EspB (residues 2–348) was observed to be disordered in initial

maps a shorter EspB (residues 2–300) construct was designed to further improve the diffraction quality. This EspB (residues 2–300) and EspK (residues 484–729) complex crystallised in similar conditions and produced the best diffracting crystals with reduced anisotropy. The best diffracting crystal conditions were produced by mixing 1 μL of the concentrated EspB–EspK complex (8 mg/mL) with 1 μL of a reservoir solution in a 24-well plate (0.2–1.2 M sodium malonate, 0.1 M HEPES (pH 7.8), and 8–14% PEG3350). These crystals were cryoprotected by supplementing the mother liquor with 25% glycerol, mounted onto a nylon loop and flash frozen in liquid nitrogen until data collection.

For mercury atoms (Hg-SAD), crystals were grown in a slightly different condition (0.2–1.2 M sodium malonate, 0.1 M HEPES (pH 7.8)) and soaked in a solution corresponding to the mother liquor, supplemented with 25% glycerol and 5 mM p-Chloromercuribenzoic acid (Hampton HR2-446-08), previously dissolved in DMSO as a 200 mM stock solution. Crystals were mounted onto a nylon loop and flash frozen in liquid nitrogen until data collection.

2.4.2 Data collection, analysis and structure refinement

The highest resolution native and S-SAD data were collected on ID30B at the European Synchrotron Radiation Facility (Grenoble, France) (271). Hg-derivative data were collected on EMBL beamline P13 at the PETRA III storage ring (DESY, Hamburg, Germany) (272). All data were integrated using the XDS suite (273). The diffraction data were highly anisotropic, with diffraction limits of ~ 2.3 Å, 2.9 Å, and 2.8 Å along the best direction for native, S-SAD and Hg derivative, respectively, but only ~ 3.1 Å, 3.6 Å, and 4 Å in the weakly diffracting direction. Therefore, all data were processed using STARANISO, as implemented in autoPROC (274), which applies non-elliptical anisotropic limits based on a locally averaged mean $I/\sigma(I)$ cut-off, performs a Bayesian estimation of structure amplitudes, and applies an anisotropic correction to the data.

The EspB–EspK complex structure was solved by molecular replacement with Phaser (275), using the EspB structure (PDB ID: 4XXX) as a search model (179). The experimental phases were further improved, and a preliminary model for EspK was generated using the Phenix AutoBuild module (276). The model was completed by cycles of manual building in Coot (109) and refinement rounds in BUSTER. MolProbity (111) was used for model validation.

2.5 Cryo-electron microscopy

Proteins were dialysed against fresh 20 mM Tris (pH 8.0), 150 mM NaCl prior to the experiment. EspB (15 μM) was mixed with EspK in a molar ratio of 1:1 and 1:3. Mixtures and individual proteins, as controls, were vitrified by applying 2.5 μL of sample on a glow-discharged UltrAuFoil Au300 R1.2/1.3 grids (Quantifoil), excess liquid removed by blotting for 3 s (blot force 5) using filter paper and plunge freezing in liquid ethane using a FEI Vitrobot Mark IV at 100% humidity at 4 °C. Cryo-EM data were collected

on an Arctica Tecnai (Thermo Fischer Scientific) at 200 kV with a Falcon III detector in counting mode. As only qualitative assessment was aimed, data were processed using the RELION-3 software (135) until 2D classification. Movie stacks were corrected for drift (5×5 patches) and dose-weighted using MotionCor2 (147). The local contrast transfer function (CTF) parameters were determined for the drift-corrected micrographs using Gctf (106).

2.6 Homology modelling

The C-terminal domain of EspK from *M. smegmatis* was modelled with AlphaFold (277), using a neural network, and by homology to crystal structure EspK from *M. tuberculosis* with Modeller (278) and Rosetta (279) software.

3. Results

3.1 Structure of EspK C-terminal domain suggests a new fold

We have shown that EspK is built into two domains connected by a low-complexity linker, where the C-terminal domain is trypsin-resistant (268), a region demonstrated to interact with EspB (210). In order to gain insight into the relationship between these proteins, we performed macromolecular crystallisation on the C-terminal domain of EspK (residues 484–729, here after termed EspK) from *M. tuberculosis*. The structure of this domain was solved, displaying seven α -helices surrounding one β -sheet formed by four antiparallel strands (**Figure 25**). Secondary structure content is consistent with CD results observed in earlier studies (268). A Dali search did not reveal a close model, suggesting this domain adopts a novel fold.

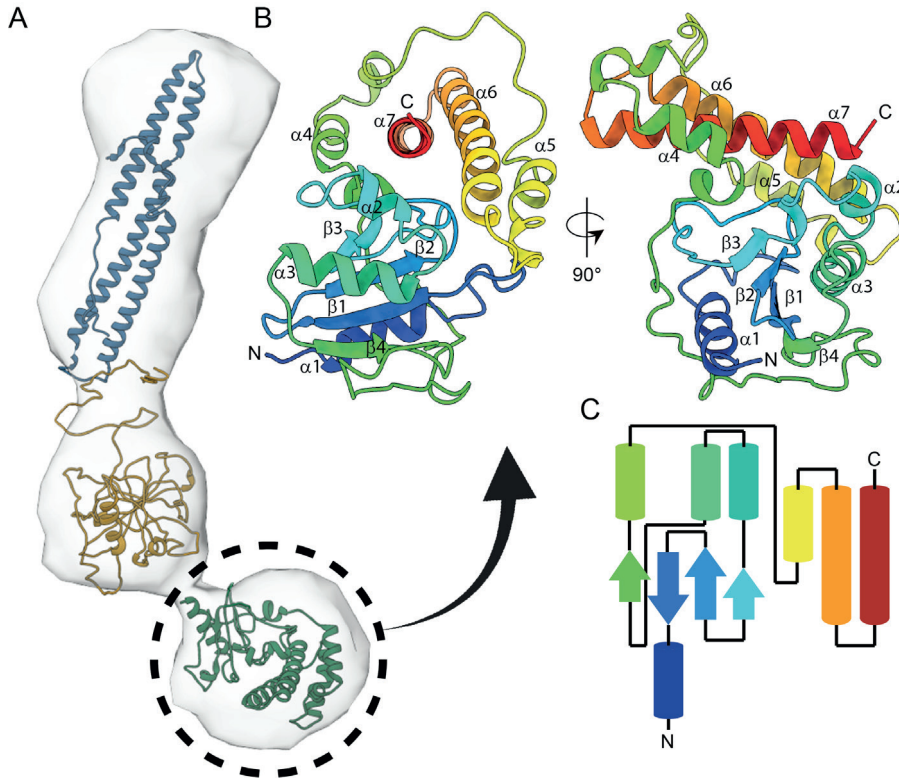


Figure 25. Structure of EspK₄₈₄₋₇₂₉ reveals a new fold. (A) Predicted model of EspK fitted in SAXS envelope from (268), with the dashed circle denoting the C-terminal domain (residues 484–729). (B) Ribbon representation of the C-terminal domain, where secondary elements are labelled $\alpha 1$ – $\alpha 7$ and $\beta 1$ – $\beta 4$. (C) Topology diagram displays overall distribution of secondary structure.

3.2 EspB and EspK form a hetero-dimer in solution

The interaction between EspB and EspK was first reported by McLaughlin *et al.* (2007) with a two-hybrid system and confirmed by pull-down assays. As those constructs differ from ours, we aimed to verify that the interaction was intact. SEC profiles of the proteins alone and together show that EspK interacts with the N-terminal region of EspB (residues 7–278, **Supplementary Figure 16**) from *M. tuberculosis*. Testing three other EspB constructs with longer C-terminal regions (EspB₇₋₂₇₈, EspB₂₋₃₄₈ and EspB₂₋₄₆₀; **Table 8**) showed similar binding (data not shown). To gain insight how the complex is formed, we first performed SAXS on the EspB constructs, all parameters are listed in **Table 9**. As expected, the EspB constructs showed differences in parameters like maximum dimension (Dmax) (**Supplementary Figure 17A**), radius of gyration (Rg), and the Porod exponent (level of compactness): longer C-terminal regions showed more flexible proteins, as evidenced by the change in the shape of the dimensionless Kratky plot, from a Gaussian distribution (EspB₇₋₂₇₈) to a plateau (EspB₂₋₄₆₀) (**Supplementary**

Figure 17B). SAXS experiments in the presence of EspK did not produce a significant change in these parameters (**Supplementary Figure 17C**), except for the estimated molecular weight, which confirmed the complex to have a stoichiometry of 1:1 (**Table 9**). The lack of change in the flexibility or maximal dimensions of EspB in the presence or absence of EspK suggests that the C-terminal region of EspB does not participate in the interaction with EspK and is only the N-terminal region where the contact points are.

3.3 EspB interacts with EspK through the helical tip

To understand how these two proteins interact, we solved the crystal structure of the complex. An initial crystal of low resolution was obtained with EspK and EspB full-length, from which a new EspB construct was created with a reduced flexible tail (residues 2-300) but retaining important contact points in the crystal packing. Improvement of the crystal diffraction led to a final resolution of 2.3 Å (**Table 7**). In the model, EspB structure covers from Val9 to Pro297, only missing Glu87–Ser112 from the PE-PPE loop, while EspK structure includes Gly484 to Ala726. The EspB structure, in complex with EspK, did not differ from the previous crystal structure (179) (0.676–0.781 Å backbone root-mean-square deviation), suggesting that no conformational change in EspB occurs upon complex formation. EspK binds to the helical tip of EspB (**Figure 26A**), similar to what has been described for PE-PPEs and their chaperone EspG (196,216,280,281). The interaction occurs between $\alpha 6-7$ and the PE-PPE loop of EspB (**Figure 26B**) and $\alpha 5-6$ of EspK (**Figure 26C**), where we can find hydrogen bonds, salt bridges, and hydrophobic interactions (**Figure 26D**). EspK appears to fill a cavity formed in EspB (**Figure 26E**). Overall, the EspB–EspK complex is stabilised by both electrostatic and hydrophobic interactions, as well as the shape formed by the contact points. Comparison of the calculated SAXS curve of this crystal structure into the experimental scatter signal further confirms this complex (**Supplementary Figure 17D**).

To further understand the EspB–EspK interaction, we performed SEC on EspK mutants in 1:1 mixtures with EspB. Shortening the amino acid side-chain of EspK with K663A substitution or making an F659D substitution in EspK to change the hydrophobicity of the residue abolished the interaction, suggestion these residues are essential for the complex formation (data not shown).

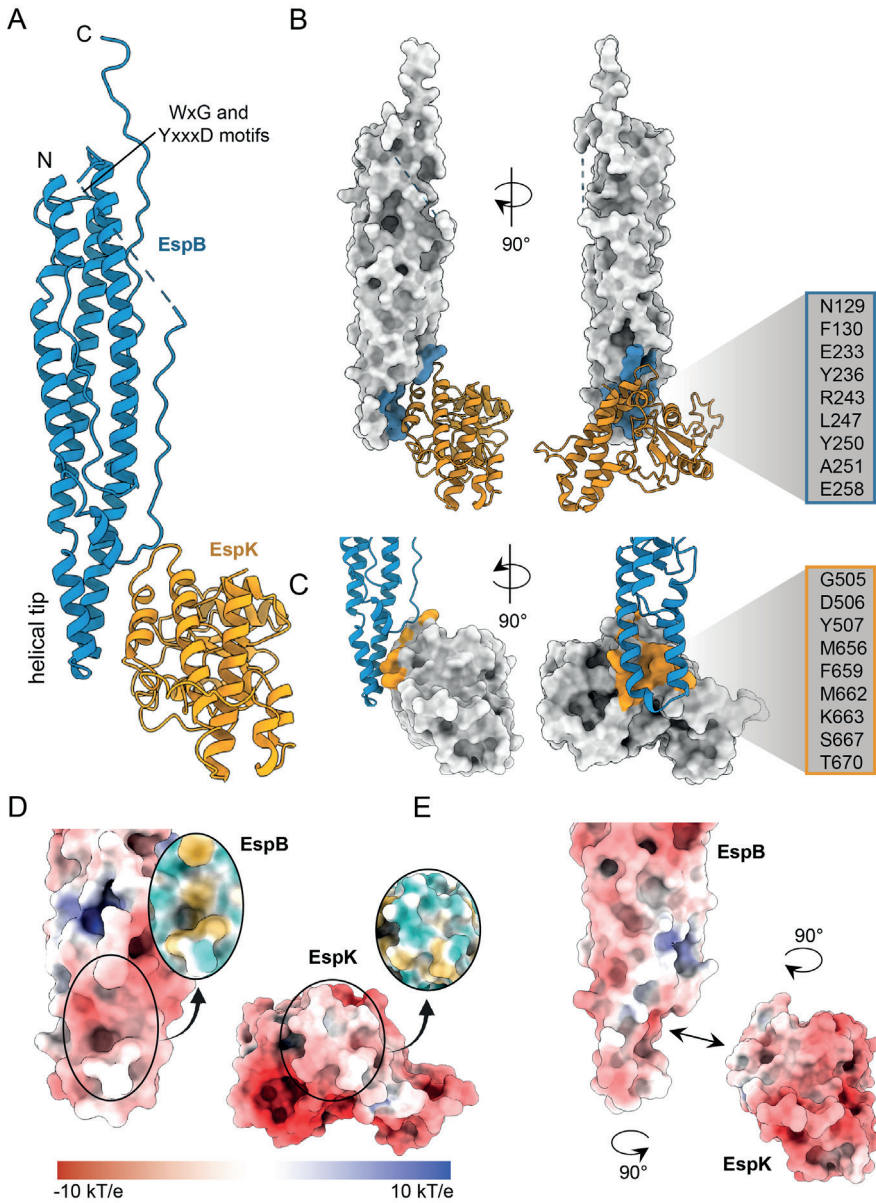


Figure 26. Structure of EspB–EspK complex. (A) Overview of protein complex highlighting EspB elements (blue). (B) Surface representation of EspB and the residues that make contact with EspK. (C) Surface representation of EspK and the residues that make contact with EspB. (D–E) Electrostatic potential of EspB and EspK highlighting the binding pocket in two different views. In D, the circled insets display the protein surface coloured by amino acid hydrophobicity according to the Kyte–Doolittle scale: polar residues in turquoise and non-polar residues in gold.

Table 7. Crystallographic data collection and refinement statistics of *M. tuberculosis* EspB–EspK.

Crystal	Native	S-SAD	Hg-derivative
Wavelength (Å)	0.976	2.066	1.008
Resolution range (Å)	85.7-2.3 (2.6-2.3)	86.2-2.9 (3.3-2.9)	86.5-2.8 (3.2-2.8)
Space group	P6 ₁ 22	P6 ₁ 22	P6 ₁ 22
Unit cell (Å)	101.6, 101.6, 377.0 90, 90, 120	99.5, 99.5, 375.8 90, 90, 120	99.9, 99.9, 376.93 90, 90, 120
Unique reflections	28129 (1406)	16171 (810)	13712 (687)
Multiplicity	6.2 (8.5)	19.5 (18.2)*	5.8 (6.0)*
Completeness (%)			
Spherical	53.2 (9.0)	65.5 (11.7)	47.7 (7.1)
Ellipsoidal	94.9 (88.4)	94.9 (74.4)	93.7 (77.9)
Mean <I/sI>	11.7 (1.9)	10.0 (2.0)	9.4 (1.9)
R _{min} (%)	3.4 (38.0)	10.5 (69.9)	9.6 (62.1)

3.4 Interaction with EspK disrupts oligomerisation of EspB

EspK binds to the helical tip of EspB, including a region of the PE-PPE loop. In earlier studies, we have demonstrated that this loop moves away from its “crystal” position for EspB to oligomerise (267). Alignment of EspB structure in complex with EspK and in the heptamer form suggests that EspK makes a steric hindrance in the oligomerisation interface. To prove that EspK binding is not compatible with the oligomerisation of EspB, we tested the ability of EspB to form oligomers in the presence of EspK, evaluated by SEC and cryo-EM. The maximum absorbance of the peak corresponding to EspB oligomer was reduced when EspK was present compared to EspB alone, at all molar ratios tested (**Figure 27A**). To confirm this result, we took advantage of the preferential orientation that EspB has in cryo-EM (267), making it easy to distinguish the oligomer, and the respective controls of EspK alone, showing that no higher-order structures are formed (**Figure 27B–D**). Mixture of EspB and EspK at 1:1 and 1:3 molar ratios disrupted the EspB oligomers in a concentration-dependent manner, as can be seen not just by the reduction of ring-like particles in the micrograph but also by the type of broken particles present in the mixture (**Figure 27E–F**). The same result obtained by two different techniques supports the idea that the presence of EspK reduces EspB oligomerisation.

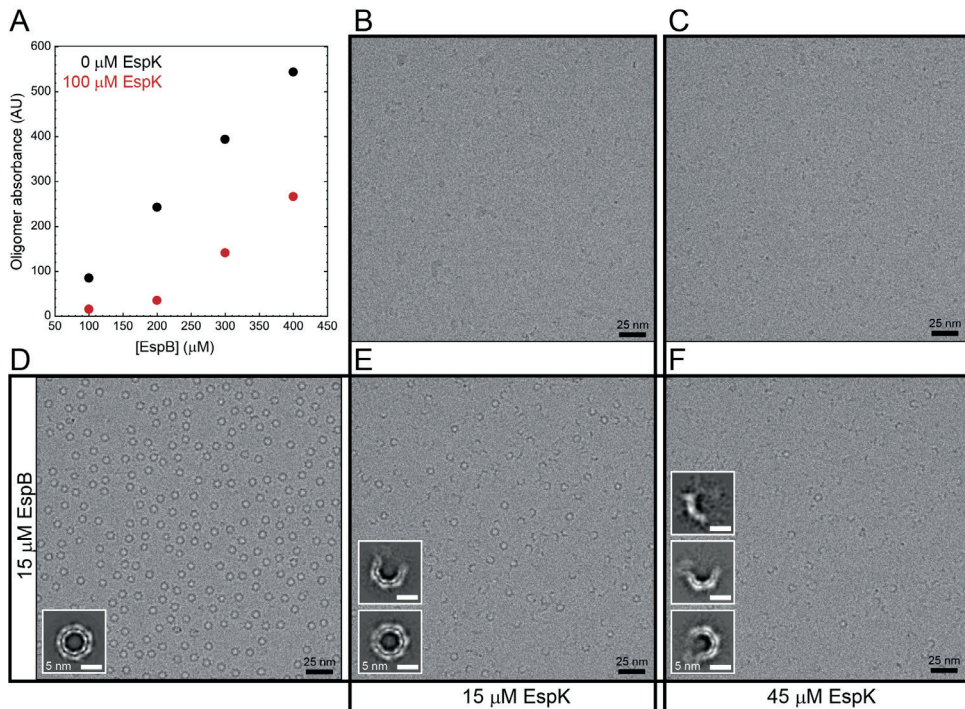


Figure 27. Disruption of EspB oligomer in the presence of EspK. (A) Maximum absorbance of EspB oligomer at different concentrations, in the absence or presence of EspK. Electron micrographs of EspK at (B) 15 μM and (C) 45 μM; (D) EspB at 15 μM; (E) mixture 1:1 molar ratio, and a (F) mixture 1:3 molar ratio. Insets correspond to the respective 2D classes.

3.5 EspB–EspK interaction is conserved in other mycobacterial species

The oligomerisation of EspB is a conserved ability for slow-growing mycobacterial species (267). To understand if EspK function is solely to regulate the oligomerisation of EspB, we assessed the interaction of this pair from two other mycobacterial species: *M. marinum*, another slow-growing mycobacteria widely used as a model microorganism to study TB and probed to form an EspB oligomer, and *M. smegmatis*, a fast-growing mycobacteria in which EspB does not oligomerise. SEC profiles showed that the interaction is conserved in both species (**Supplementary Figure 18**), suggesting that EspK plays another role outside regulating EspB oligomerisation.

The C-terminal domain of EspK from *M. tuberculosis* has a high percentage identity with its orthologue from *M. marinum* (EspK: 80.4%). Thus, we tested if EspB and EspK proteins from *M. tuberculosis* could interact with their counterparts from *M. marinum*. As observed by a peak with an earlier elution volume in the SEC profiles, proteins from these evolutionary close species are able to interact (**Supplementary Figure 19A–B**). However, none of the EspB and EspK proteins from *M. tuberculosis* and *M. marinum*

interacted with EspB and EspK from *M. smegmatis* (**Supplementary Figure 19C–F**). This result may be due to the low sequence identity of EspK (53.1–53.3%) and EspB (28.5–30.7%) from *M. smegmatis* compared to those from *M. tuberculosis* and *M. marinum*. To understand the molecular reasoning of these results, we predicted the EspK structure from *M. smegmatis* with AlphaFold (277), a neural network-based software, and Modeller (278) and RosettaCM (279), a homology-based software, using the EspK crystal structure from *M. tuberculosis*. EspK from *M. smegmatis* is predicted to have the same fold as its orthologue (**Figure 28**), regardless of the method used. Considering EspB also shares a similar structure between these three species (180,267), this result does not explain why there is no EspB–EspK interspecies interaction for *M. smegmatis*. Taking a closer look into the structures of *M. smegmatis* and *M. tuberculosis*, we can see that important contact points are different, including $\alpha 6$ in EspB and residues that change from apolar to polar in EspK (**Figure 29**). It is noteworthy that when aligning the *M. smegmatis* models into the *M. tuberculosis* complex, residues with the same charge would be facing each other, repelling the proteins. Considering that SEC experiments proved that the interaction between EspB and EspK from *M. smegmatis* occurs, it is possible that such complex would have a different arrangement.

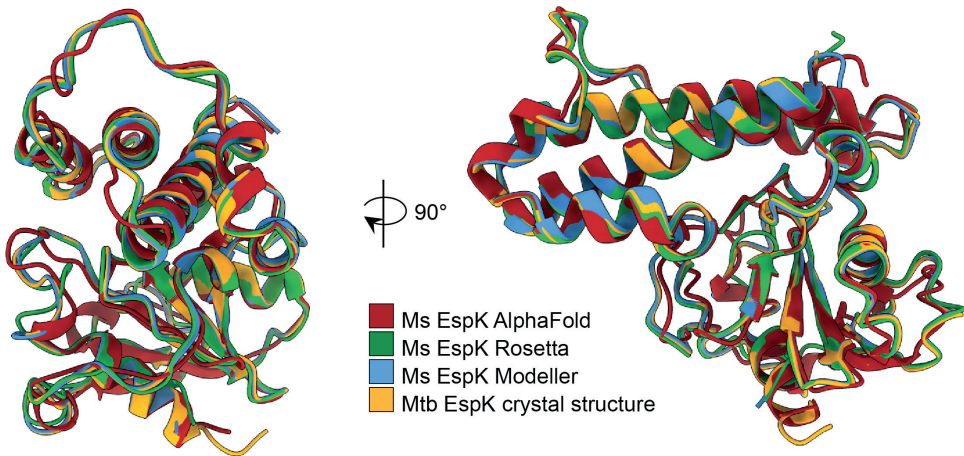


Figure 28. EspK model from *M. smegmatis* is similar to the *M. tuberculosis* orthologue. Comparison of EspK from *M. tuberculosis* and EspK model predictions from *M. smegmatis* obtained by AlphaFold (277), a neural network-based software, and Modeller (278) and RosettaCM (279), a homology-based software using the crystal structure of EspK from *M. tuberculosis*.

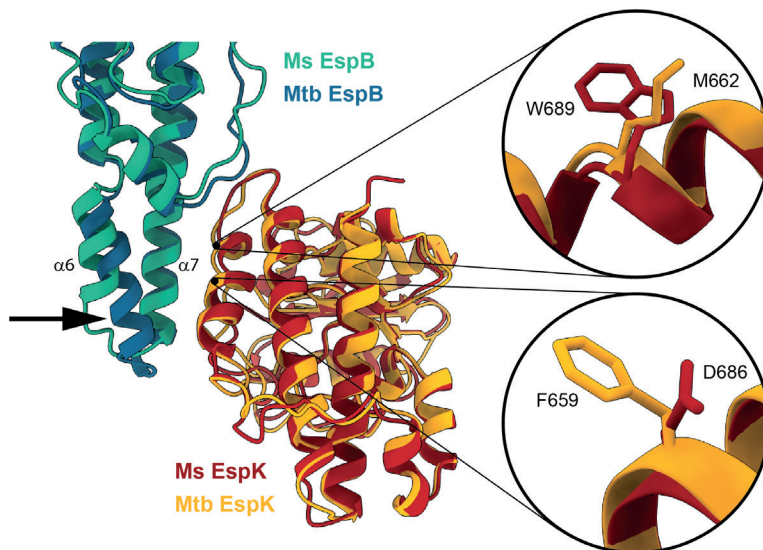


Figure 29. Structural differences between EspB and EspK from *M. tuberculosis* and *M. smegmatis* in the complex context. EspB from *M. smegmatis* (green) (PDB ID: 4WJ1) aligned to EspB crystal structure from *M. tuberculosis* (blue) obtained in this work. EspK from *M. smegmatis* (red) predicted by AlphaFold aligned to EspK crystal structure from *M. tuberculosis* (yellow).

4. Discussion

Despite the arduous efforts to investigate the functioning of ESX-1, important information like the secretion process of protein substrates is not fully understood and for some proteins is completely missing. One of these proteins is EspB. In early studies, we have described the quaternary structure of EspB and the possible role that could play as an element of the putative outer-membrane complex of ESX-1 (267). Knowing the importance of this protein in the proper functioning of the secretion system (228), we sought to gain information in the secretion mechanism of EspB by studying the interaction between EspB and its putative chaperone, EspK.

EspB belongs to the PE/PPE family, a group of proteins that shares structural homology and that comprises nearly 10% of the protein repertoire of pathogenic mycobacteria (282). These family of proteins is secreted by the T7SS with the assistance of the chaperone EspG; while all the system secretes only one PE–PPE complex, ESX-5 secretes 95% of them (283). Despite the resemblance, the secretion of EspB does not depend on EspG₁ (210). McLaughlin *et al.* (2007) described the interaction between EspB and the C-terminal end of EspK and the lack of EspB secretion in an *espK* mutant strain led them to hypothesised EspK as possible chaperone of EspB. To understand the relationship between these proteins, we solved the structure of EspK C-terminal domain (268) bound to EspB. To our surprise, EspK structure resulted in a novel fold with a

series of helices surrounding a β -sheet of four antiparallel strands. The structure of the complex revealed that EspK binds to the helical tip of EspB, a feature shared with its PE–PPE homologues and their chaperone EspG (196,216,280). The interaction at this position leaves the WxG and YxxxD motifs in EspB accessible for the interaction with ESX-1 secretion system supporting the current models of EspB mechanism of action (267). Different from the PE–PPE–EspG complex where the helical tip of the PPE seems to be inserted in a deep groove formed by EspG, EspK appears to fill a cavity formed in EspB.

To date, the structure of EspG₁, EspG₃ and EspG₅ have been solved, and although, they present the lowest sequence identity of all the paralogues in the ESX-systems (13–23%) their tertiary structure is conserved (256). However, they show enough structural differences to present secretion-specificity (**Figure 30**, (196)). Variances in two key regions alter the shape of the PPE binding surface that may translate in binding specificity: the b2-3 loop changes in length and structure and the α 2 is longer for EspG₃, which would clash with the PPE protein in the EspG₅ orientation (196). EspK does not share structural homology to EspG, despite binding in the same region.

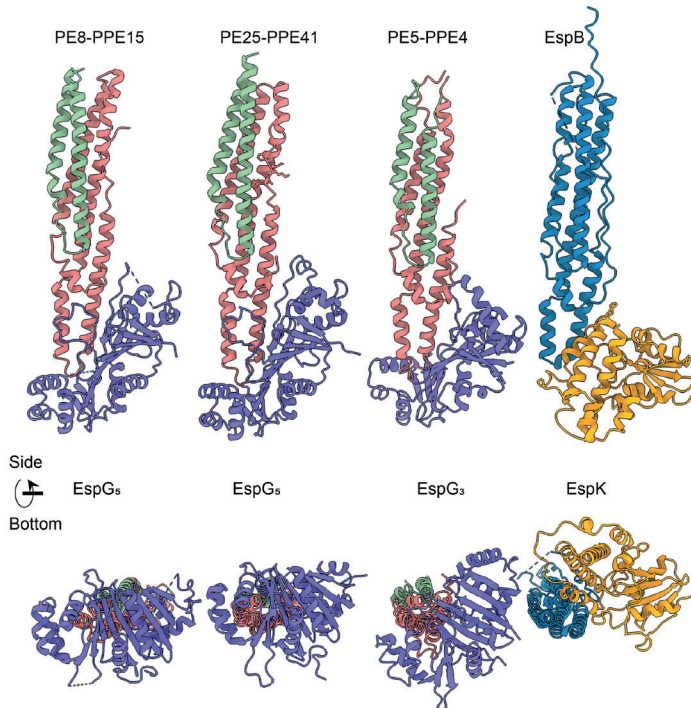


Figure 30. EspG and EspK bind to the helical tip of their respective binding partner. Structure comparison of the PE-PPE-EspG ternary complexes available in the literature (PDB ID: 5XFS, 4KXR, 6UUJ) and EspB-EspK complex. Where PE's are coloured in green, PPE's are in pink, EspG's are in purple, EspB in blue and EspK in yellow.

Chaperones involved in secretion systems play multiple functions to keep proteins in a secretion-competent state: stabilisation of insoluble proteins, preventing premature folding or self-oligomerisation, directing substrates to the secretion machinery, and regulating the hierarchy in the secretion process (196,284–286). It has been shown that EspG interaction keeps PE–PPEs soluble (196,250) and some experimental evidence suggests that this ternary complex is delivered to the machinery by the ESX-core component A (EccA) (287). EspB is a highly soluble protein and does not require a chaperone for the same reason as its homologues. Instead, EspK interferes with the oligomerisation of EspB, which supports the idea that the interaction occurs in the cytoplasm of the bacterium. Despite what could seem obvious, the location of the interaction has been never tested. McLaughlin *et al.* (2007) suggested it was in the cytoplasm because they could not find EspK secreted into the medium, however, since then, multiple studies have shown the presence of EspK in the capsule or culture filtrate (174,178,232). As EspB oligomerises upon secretion, it is expected that any control on this event would happen before its secretion. We hypothesised that EspK binds to EspB, not to maintain it soluble but to prevent a premature oligomerisation that could stop the secretion. We have shown that one of the factors that favours the oligomerisation of EspB is the cleavage of its C-terminal region that occurs during secretion, as well as the presence of an acidic environment (267). Nevertheless, oligomerisation of EspB full-length could happen in the no-so basic environment of mycobacterial cytoplasm. So blocking the oligomerisation of EspB by EspK could ensure EspB to go through the secretion machinery, making the process more efficient.

EspB oligomerisation seems to be a conserved characteristic of slow-growing mycobacteria, and missing in fast-growing species, suggesting that it was an advantage gained during evolution (267). However, we have shown that the interaction between EspB and EspK also occurs in *M. smegmatis*, a species belonging to the fast-growing group and that lacks an EspB oligomer. This could implicate that the appearance of EspK–EspB interaction could have happened in a shared ancestor, earlier than the oligomerisation, making it plausible that blocking oligomerisation is secondary to EspK main function. Furthermore, the lack of interaction between EspB and EspK proteins from the phylogenetic distant relative, *M. smegmatis*, with EspB or EspK from *M. marinum* and *M. tuberculosis*, suggests that the proteins have evolved together maintaining the interaction over time.

It has been suggested that EspK direct EspB to the machinery by connecting EspB to EccCb₁ (210). McLaughlin *et al.* (2007) described a model where the N-terminus of EspK binds to EccCb₁ and the C-terminus binds to EspB, in *M. tuberculosis* and *M. marinum*. The latter have been confirmed in this paper. However, they only included in the model an interaction with EccCb₁ through the N-terminus of EspK even though they showed interaction also with its C-terminus (see supplementary figure 3B of (210)). In the same study, they could not see interaction between EspB and EccCb₁, while Solomonson *et al.* (2015) saw a weak interaction in *M. smegmatis*. The secretion

of substrates via the T7SS is defined by a bipartite signal (YxxxD/E motif and WxG motif), and although the integrity of those motifs have been proved essential (266), no interaction of this motif has been described so far with the machinery or any other element (168,288,289). On the other hand, EccCb₁ recognises hydrophobic residues seven positions downstream of the YxxxD/E motif in the EsxB substrate (168,289). Sequence alignment on multiple ESX-secreted protein, including EspB, suggests that this is a conserved characteristic, and the name “export arm” was assigned to this region (180). It is noteworthy that in EspB from *M. smegmatis* the export arm is an extension of the $\alpha 2$ (180), while in *M. marinum* and *M. tuberculosis* it is unstructured (179,267), which might explain the difference in the aforementioned information. Taking together all the available information and considering the advantages and disadvantages of the experiments used to obtain it, is possible that the data is not mutually exclusive, so we suggest a putative model of the secretion mechanism of EspB: EspK recruits EspB through its C-terminal domain maintaining the protein monomeric, while its N-terminal domain contact the machinery (EccCb₁). The long and flexible linker that connects such domains (268) works as a rope pulling the C-terminal domain–EspB complex towards the machinery. Interaction of EspB with EccCb₁ allows recognition of EspB export arm and secretion occurs to form an oligomeric structure that allows the transit of proteins through (maybe) the outer membrane (**Figure 31**).

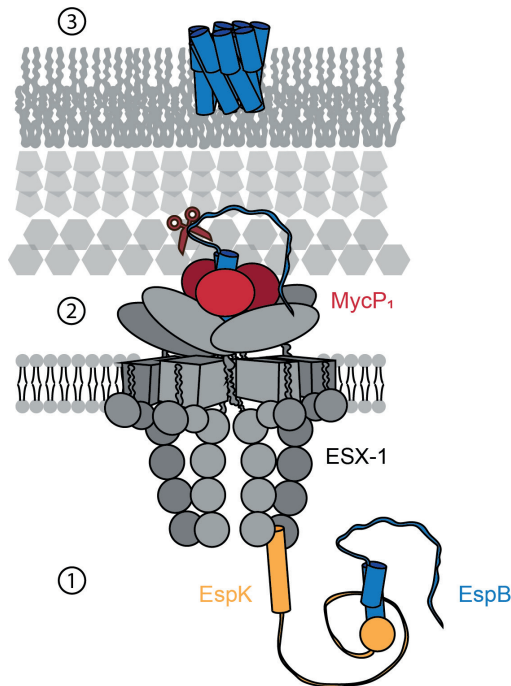


Figure 31. Scheme the secretion pathways of EspB. (1) EspB interacts with the C-terminal domain of EspK, which pulls it towards ESX-1 to be recognised by the machinery. (2) EspB is secreted through the channel and cleaved by MycP₁ at its C-terminal region. (3) EspB meets the conditions to oligomerise and be functionally active.

In conclusion, we have characterised the structure of EspB in complex with its chaperon EspK. Although the direction of EspB to the machinery as main function of EspK remains to be fully described, EspK keeps EspB in a secretion-competent state in slow-growing mycobacteria by avoiding a premature oligomerisation. Despite the enormous efforts to understand the function of the T7SS seeking a way to block it and finally eradicate TB, it has been hampered by the complexity of the structure and mechanism of secretion/action. The structure of EspB–EspK provides a possible new secretion pathway, helping to untangle little by little the secretion mechanism of ESX-1. Likewise, the structure of EspK represent an opportunity to design therapeutic compounds that could mimic the dissociation effect in EspB oligomer and thus diminishing virulence in mycobacteria.

Data availability

The crystallographic and SAXS data will be deposited in the Protein Data Bank. The SAXS data has been deposited in the Small Angle Scattering Biological Data Bank (SASBDB (257)) under the accession numbers SASDMD7 (EspB₂₋₄₀₆), SASDME7 (EspB₂₋₃₄₈), SASDMF7 (EspB₇₋₂₇₈), SASDMG7 (EspK₄₈₄₋₇₂₉ + EspB₂₋₄₀₆), SASDMH7 (EspK₄₈₄₋₇₂₉ + EspB₂₋₃₄₈), and SASDMJ7 (EspK₄₈₄₋₇₂₉ + EspB₇₋₂₇₈).

Author contribution

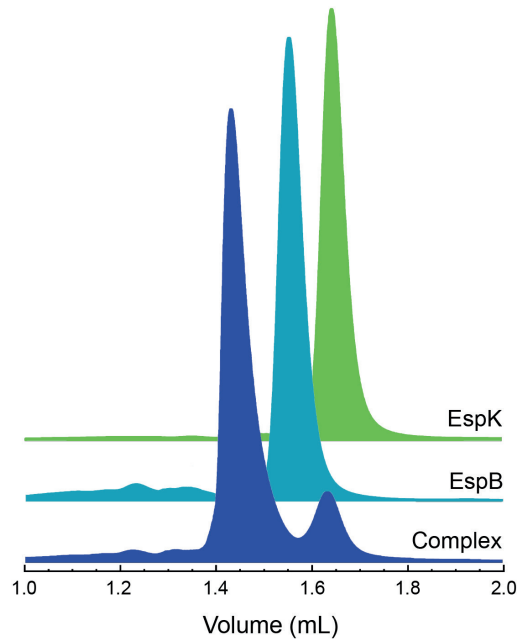
Abril Gijsbers: Conceptualisation, Methodology, Formal analysis, Investigation, Visualisation, Writing - Original Draft, Writing - Review & Editing. **Mathias Eymery:** Methodology, Formal analysis. **Isabella Menart:** Methodology, Investigation, Writing - Original Draft. **Andrew McCarthy:** Methodology, Formal analysis, Funding acquisition, Writing - Review & Editing. **Dritan Siliqi:** Methodology, Formal analysis, Funding acquisition, Writing - Review & Editing. **Vanesa Vinciauskaite:** Methodology, Investigation. **Ye Gao:** Methodology. **Peter J. Peters:** Conceptualisation, Supervision, Funding acquisition, Writing - Review & Editing. **Raimond B. G. Ravelli:** Conceptualisation, Formal analysis, Supervision, Funding acquisition, Writing - Review & Editing.

Acknowledgments and funding

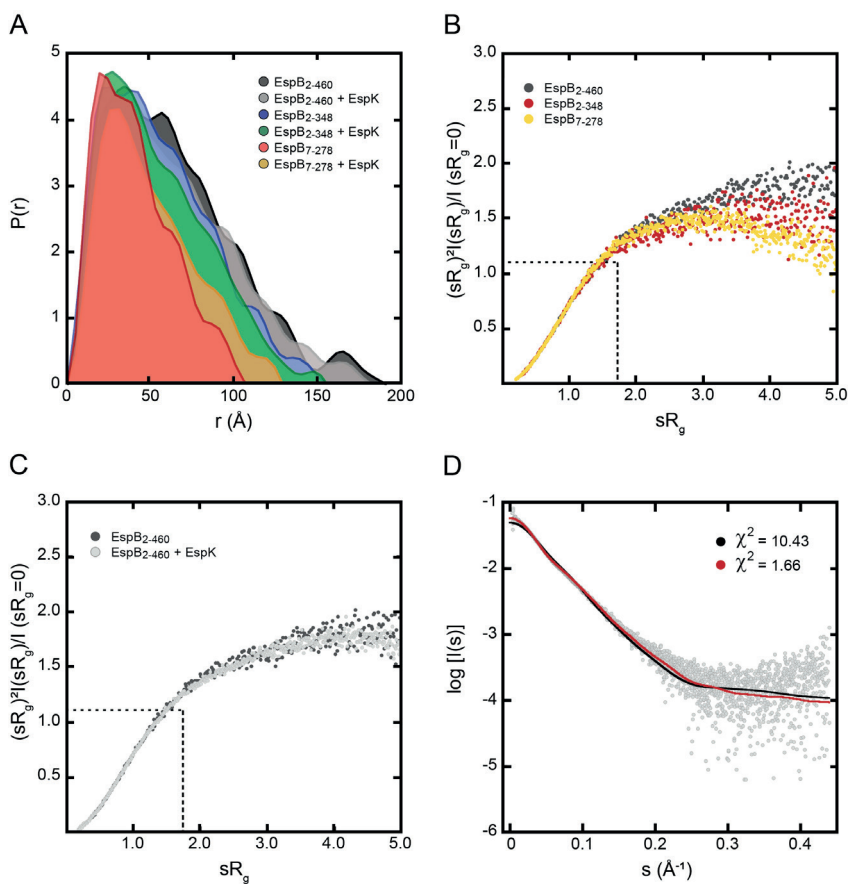
We thank the Microscopy CORE Lab (UM) for their technical and scientific support; and Hang Nguyen (UM) for critical reading of the manuscript. This research received funding from the Netherlands Organisation for Scientific Research (NWO) in the framework of the Fund New Chemical Innovations, numbers 731.016.407 and 184.034.014, from the European Union's Horizon 2020 Research and Innovation

Programme under Grant Agreement No 766970 Q-SORT. This research is also part of the M4I research programme supported by the Dutch Province of Limburg through the LINK programme. The SAXS experiments were funded from proposals MX21741-1 and MX21741-2 on the beamline B21, Diamond Light Source, respectively. We thank Nathan Cowieson and Nikul Khunti from Diamond Light Source for their assistance in the preparation of the SAXS experiments.

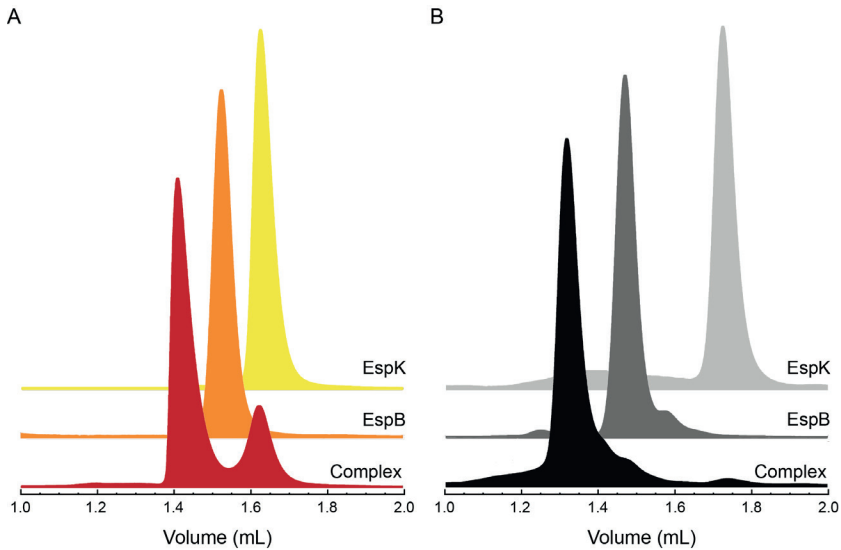
Supplementary material



Supplementary Figure 16. EspB and EspK form a complex. Size exclusion chromatography profiles of *M. tuberculosis* EspK₄₈₄₋₇₂₉, EspB₇₋₂₇₈, and a mixture of 1:1 molar ratio (labelled Complex). Experiments were performed at 100 μ M in 20 mM Tris-HCl (pH 8.0), 150 mM NaCl. Void volume correspond to 0.8 mL elution volume.



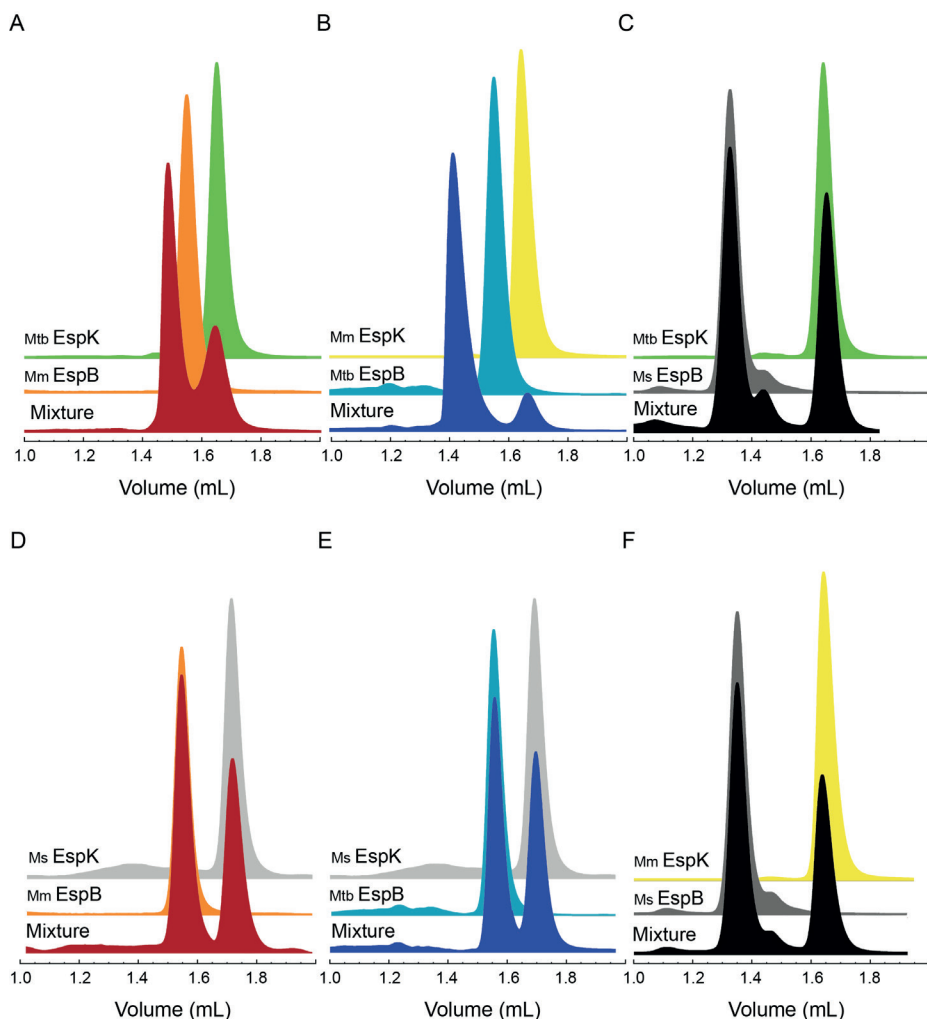
Supplementary Figure 17. Low resolution SAXS data suggests N-terminal region of EspB solely binds EspK. (A) Pair distribution function plot and dimensionless Kratky plot of EspB constructs in the (B) absence or (C) presence of EspK. In B and C, the intersection of the dotted black trace corresponds to the value for the reference protein bovine serum albumin (BSA). (D) Fit of the calculated SAXS scattering curves of the crystal structure (EspB₂₋₃₀₀) compared to the experimental scattering signal (black line). As experimental SAXS data was done on EspB₂₋₃₄₈, 48 residues were added to the crystal structure by the program CORAL (242) to make them comparable (red line).



Supplementary Figure 18. EspB and EspK from other species also form a complex. (A) Size exclusion chromatography profiles of (A) *M. marinum* EspK₅₃₂₋₇₇₆, EspB₂₋₂₈₆, and a mixture of 1:1 molar ratio (labelled Complex) and of (B) *M. smegmatis* EspK₅₁₅₋₇₅₅, EspB₂₋₄₀₇, and a mixture of 1:1 molar ratio (labelled Complex) Experiments were performed at 100 μ M in 20 mM Tris-HCl (pH 8.0), 150 mM NaCl. Void volume correspond to 0.8 mL elution volume.

Table 8. Constructs from this study and the experiments in which they were used.

Specie	Construct	Experiment
<i>M. tuberculosis</i>	EspB 2-460	SAXS and MX
<i>M. tuberculosis</i>	EspB 2-348	SAXS and cryo-EM
<i>M. tuberculosis</i>	EspB 2-300	MX
<i>M. tuberculosis</i>	EspB 7-278	SEC and SAXS
<i>M. tuberculosis</i>	EspK 484-729	SEC, MX, SAXS and cryo-EM
<i>M. marinum</i>	EspB 2-286	SEC
<i>M. marinum</i>	EspK 532-776	SEC
<i>M. smegmatis</i>	EspB 2-407	SEC
<i>M. smegmatis</i>	EspK 515-755	SEC



Supplementary Figure 19. No EspB and EspK inter-species interaction is formed with *M. smegmatis*. Size exclusion chromatography profiles of EspK, EspB, and a mixture of 1:1 molar ratio. (A) *M. tuberculosis* EspK₄₈₄₋₇₂₉ and *M. marinum* EspB₂₋₂₈₆. (B) *M. marinum* EspK₅₃₂₋₇₇₆ and *M. tuberculosis* EspB₇₋₂₇₈. (C) *M. tuberculosis* EspK₄₈₄₋₇₂₉ and *M. smegmatis* EspB₂₋₄₀₇. (D) *M. smegmatis* EspK₅₁₅₋₇₅₅ and *M. marinum* EspB₂₋₂₈₆. (E) *M. smegmatis* EspK₅₁₅₋₇₅₅ and *M. tuberculosis* EspB₇₋₂₇₈. (F) *M. marinum* EspK₅₃₂₋₇₇₆ and *M. smegmatis* EspB₂₋₄₀₇. Experiments were performed at 100 μ M in 20 mM Tris-HCl (pH 8.0), 150 mM NaCl. Void volume correspond to 0.8 mL elution volume.

Table 9. SAXS data collection and experimental parameters for *M. tuberculosis* EspB constructs alone and in complex of EspK.

Data collection parameters	P12, DESY (Hamburg, Germany)		B21, Diamond Light Source (Harwell, UK)			
	Pilatus 6M (3000 mm)		Pilatus 2M (4014 mm)			
Detector (distance to sample)	0.2 × 0.3 mm		0.2 × 0.2 mm			
Beam size	10.0 kV		12.4 kV			
Energy	0.0038 – 0.42		0.0038 – 0.42			
q range (\AA^{-1})	1		3			
Exposure time (s)	900		620			
Number of frames	293		293			
Temperature (K)	SEC online		SEC online			
Mode	SEC online		SEC online			
Structural parameters	EspB 2-460	EspB 2-348	EspB 7-278	EspK–EspB 2-460	EspK–EspB 2-348	EspK–EspB 7-278
Injection concentration (mg/mL)	6.1	5.6	6.7	4.3/1.5	8.6/2.4	8.6/1.9
q Interval for Fourier inversion (\AA^{-1})	0.009–0.159	0.009–0.200	0.011–0.251	0.008–0.216	0.011–0.184	0.016–0.195
R_g [from P(r)] (\AA)	54.70 ± 0.26	42.60 ± 0.12	34.11 ± 0.21	54.90 ± 0.42	45.24 ± 0.42	42.94 ± 0.12
R_g [from Guiner approximation] (\AA)	50.10 ± 0.04	40.00 ± 0.04	32.60 ± 0.17	49.60 ± 0.19	43.24 ± 0.12	40.77 ± 0.13
sR_g limits [from Guiner approximation]	0.45–1.28	0.40–1.30	0.27–1.25	0.42–1.30	0.40–1.30	0.67–1.30
D_{max} (\AA)	191	157	108	183	157	130
Porod volume estimate (nm^3)	119	75	56	128	102	113
Molecular Mass (kDa)						
Credibility interval (probability)	52–59 (92%)	41–46 (91%)	34–39 (92%)	71–78 (94%)	61–69 (94%)	53–59 (92%)
From sequence	48	37	30	75	64	57
Software employed						
Primary data reduction	DAWN pipeline (Diamond Light Source, UK)					
Data processing	ScÅtter v3.1 q, ATSAS					
Modelling	CORAL					
Computation of model intensities	CRY SOL					

7

CHAPTER 7

Impact paragraph

Abril Gijbers

I. Economical and societal impact

TB is the oldest disease on record, and still, too many patients succumb from it every year due to ineffective treatment. The development of new drugs has been challenging due to the complex infection process, the pathogen protecting itself in the interior of the host cells, and the variability of disease severity between patients linked to the different lineages (290). Nevertheless, the world would have a great benefit from an efficient therapy to prevent and treat TB. 1) Annual investment (US\$ 5 billion) in health care and medical units dedicated to TB, including personnel during and after hospitalisation, can be used to fulfil other medical or social needs. 2) TB fractures fragile economies by disabling a considerable number of people to work, compromising small business and family financials. Reducing patients with active infection could aid the country's growth by contributing to the workforce and saving families from crossing the impoverishment threshold. 3) The latter has more profound consequences; life quality would profit from poverty reduction, which could mitigate the stigma around the disease helping in social re-integration. Only a few pharmaceutical companies have shown interest in engaging this challenge (291). The reason is the little revenue expected from treating a disease with an available treatment combined with a market dominated by middle and low-income countries, leaving the task in the hands of academia with relative plodding progress. However, it is of most importance to have a joined front, like the one seen for COVID-19, if we want to win this long war against TB. Otherwise, MDR and XDR-TB threat to turn **this silent pandemic into a thunderous one**.

The new vaccines that target the Spike protein of the SARS-CoV2 are a few examples of how structural biology has changed how new therapeutics are developed, speeding the process by making a structure-based drug and vaccine design (292,293). However, this has been possible after decades of hardware and software development. Nevertheless, every protein represents a new challenge, limiting our capacity to determine protein structures. New developments or tools, like the VitroJet with a robust standard protein (BfrB) to test it, will potentially take the field of structural biology, in particularly cryo-EM, to characterise any protein or protein complex, and give a solution to many medical or biological problems, improving people's lives. In this thesis, I described the structural characterisation of three mycobacterial proteins necessary for the pathogenesis of *M. tuberculosis*, which could be target for treating TB. BfrB is a ferritin protein that works as a vault, storing iron for vital cellular processes. During infection, the host cell limits the concentration of iron to make the bacterium "starve". Disruption of BfrB function would compromise the survival of *M. tuberculosis* and the progression of the infection in this hostile environment, as preliminary data has shown (294). On the other hand, EspB and EspK belong to the ESX-1 secretion system, a protein complex responsible for the translocation of *M. tuberculosis* from the phagosome to the host cell's cytoplasm (74). This event is a characteristic of all pathogenic species and does not depend on genomic variations (lineages). Likewise, mutations in *espB* and *espK* genes are known to result in the attenuation of the pathogen. As secreted proteins

accessible to the environment, they could be good targets for structure-based drug and vaccine design.

2. Scientific impact

Mycobacterium tuberculosis has walked through history with us, killing more humans than any other pathogen. Despite the arduous work to understand the different mechanisms that *M. tuberculosis* puts in action to evade the immune system, we are far from succeeding. The subjects of this thesis are mycobacterial proteins that play a role in the survival of the bacteria during infection, particularly proteins belonging to the T7SS. Which role and how the proteins perform it were the questions that drove this research. The involvement of the T7SSs in different steps of pathogenesis has made them considered one of the most important virulence factors, attracting many scientists and making this a very competitive field. However, this has not been enough to speed the pace of its characterisation. The T7SS or the ESX-systems are protein complexes embedded in the membrane of mycobacteria, which has a double complication. 1) Lipids composition could be critical for the proper structure of the complex. Mycobacteria have an unusual membrane composition (295), which might explain why the successful purification of the machinery has only been reported from mycobacterial species. Nevertheless, that is not a simple answer to our problems. Mycobacteria is not a common organism for recombinant protein expression; hence there are limited molecular-biology tools compared to other model organisms and expertise. That is why recombinant work in heterologous systems like *E. coli* is still a common practice. 2) The ESX-1 system has six proteins that form the inner-membrane, 17 proteins that are secreted or regulate secretion, and other multiple proteins that regulate the system at the expression level. To understand how the machinery works, we need to identify all the elements involved and understand their function, which might implicate working in settings with multiple proteins simultaneously. Moreover, to put another level of difficulty, as expressed before, TB is caused by different mycobacterial lineages that have polymorphisms across their genome, changing the behaviour of proteins.

The elucidation of ESX-3 (166,167) and ESX-5 (170,171) structures are milestones for the scientific community, giving valuable insight into the secretion mechanism for all five systems (296). Unfortunately, they are not enough to understand the whole process, for example, because they cannot explain how proteins will be secreted through the outer membrane. The proteins that compose the outer-membrane complex are unknown, information critical for the proper understanding of the system; secreted proteins have to traverse the physical barrier, and ESX-1 cytolytic activity is carried out through physical contact (176). The subject's relevance highlights the importance of this work, especially on the protein EspB, which has the correct dimensions and physicochemical properties to allow the transit of proteins through it, suggesting its participation in the outer-membrane machinery. Our work could serve as a thread to untangle the biggest mystery of the T7SS and provide the basis for a vaccine like the spike protein of SARS CoV2.



CHAPTER 8

Discussion and summary

Abril Gijsbers

General discussion

The future of structural biology and protein structure determination

Protein structures have been solved for many decades, giving invaluable information on biological processes and their perturbation (*i.e.*, the effect of a mutation in a disease or a drug's mechanism of action (297)). However, the number of structures solved is minimum compared to the number of proteins in nature. It has been 40 years since the first genome sequence was reported, the human mitochondrion (298). From there, the determination of the sequence of a protein has no longer been the limiting factor. Only this year, the complete sequence of the human genome has been reported (299), after 20 years of the first draft (300). If the amino acid sequence of a protein determines its structure, why is it so difficult to predict the structure of proteins? Cyrus Levinthal described that protein folding is not randomly acquired; otherwise, there would be an astronomical number of possibilities, taking much longer for a small protein to fold than the universe's age (301). Even with the computational power existing nowadays, it is not trivial to predict the structure of a protein from its sequence. Structure prediction programmes need data available in the PDB to solve structures by homology or pairwise evolutionary correlation; however, results are far from accurate, especially when homologue structures are not available. The company DeepMind has developed in recent years a programme called AlphaFold, which predicts protein structures based on a neural network that learns evolutionary, physical, and geometric constraints from experimental data (277). Based on their results in the 14th Critical Assessment of protein (CAS14), AlphaFold promises to predict protein structures with high accuracy even for those with no structural homologues. This year, together with the European Bioinformatics Institute (EMBL-EBI), they have made a structural database of more than 20 proteomes, including the human (302). Does this mean that structural biology is becoming dispensable? Hardly. First, these predictions depend on more experimental data to learn and improve. Second, the proteome of an organism is not as simple as it has been depicted. After translation, proteins still go through changes that are not yet considered: chemical and physical post-translational modifications, conformational changes, and interaction with other proteins at different stages of the functional cycle. Though, the latter is currently in development. One good example is EspB. If you look at this protein in the AlphaFold server, it is shown as a full-length protein in a monomeric state. It ignores that EspB interacts with EspK, oligomerises in a heptamer-manner after secretion, or cleaved by MycP₁. Nevertheless, structure prediction by AlphaFold or any other programme, could be extremely valuable. It could temporarily serve as a reference for biochemical and cellular techniques (*e.g.*, to decide where to place a fluorescent probe) while using it to reduce the time needed to solve protein structures experimentally, either from the sample preparation or model building.

The challenge of proteome complexity is not only faced by bioinformatics. Experimental scientists realised that the structure determination of a single protein is not enough to understand its function. It is estimated that there are 42 million proteins in a eukaryote

single-cell organism like yeast (303) and a hundred times more in a human cell (304). Characterisation of proteins *in vitro* is simplistic (not easy), but for many techniques it is the only way. Unfortunately, we most likely ignore all the critical elements relevant to the protein's proper (natural) functioning, either organic or inorganic molecules from the environment. With this in mind and the fact that proteins do not work in isolation, new methods are being developed to study proteins *in situ*, like cryo-ET and sub-tomogram averaging. This technique's principle is similar to the SPA cryo-EM, with the main difference that particles are acquired and further processed from a tomogram rather than a 2D image. The advantage of such a technique is that information can be obtained from the natural environment, circumventing the bottleneck that protein purification represents for all the *in vitro* techniques. However, it comes with a heavy cost: sample preparation becomes a complex process and identifying proteins at a long distance (low magnification) within the crowd. Complementary techniques like fluorescence microscopy are being used to alleviate this problem, but this only works to a certain point as the localisation accuracy that can be obtained by light microscopy is limited by its wavelength (resolution of 0.2 μm). Currently, sub-tomogram averaging can only be used for specific cases: mega Dalton-size protein complexes within the cell or smaller (~ 300 kDa) proteins protruding from the virus capsid, with a very distinct morphology and in high abundance (305). Although SPA cryo-EM and tomography are the future of structural biology for many people, they might not achieve to solve all kind of protein structures, arguing for a **multimodal** approach to complement or corroborate our results if the time and money allow it, to have a close-to-complete view of the biological process where the protein is involved.

Democratising the future of structural biology

This year, the PDB celebrates 50 years since it was established as an open-accessed archive to share protein structures (306). The knowledge generated has been invaluable, and proof of that is the multiple servers and software based on its content (277,307,308). Unquestionably, a joined effort to continue feeding the structural reservoir is for everyone's benefit; unfortunately, financial resources limit the participation of research groups from middle and low-income countries in this global task. Cryo-EM technology is excessively expensive, for many countries, completely unaffordable. Installing a high-end microscope can cost more than €6 M with an annual running cost of more than €200,000 and needs unique infrastructure. In countries like Mexico, where average awarded grants are for €20,000, it has to be a national effort to keep a single microscope running.

International facilities would partly solve the problem, as illustrated by the access to high-quality X-ray experiments (MX and SAXS) in synchrotrons. However, the rest of the problem comes before acquiring the final data set. Sample preparation for cryo-EM experiments requires an iterative evaluation to produce the perfect sample: good sample distribution and not caught on the support film, random orientation, minimum sample overlapping, no denaturation at the air-water interface, and no aggregation. To

alleviate these problems, one needs continuous access to a cryo-EM microscope good enough to evaluate the sample. High-end microscopes are not necessary; in fact, recent studies have demonstrated the potential use of 100-kV microscopes for cost-effective sample evaluation or low-resolution structure determination (309). This possibility has vast advantages as low-voltage microscopes are already available worldwide or could be afforded. The downside is that there is no high-efficiency detector optimised for 100-kV microscopes so far (310,311). Richard Henderson (awarded Nobel Prize in Chemistry) and Chris Russo, from the MRC Laboratory of Molecular Biology, are strong advocates to build such equipment, pressuring manufacturers like Thermo Fisher Scientific (TFS) to make it happen. The launch of the 100-kV Tundra microscope by TFS could potentially have the effect that we hope for, reaching most countries' financial possibilities.

***M. tuberculosis* pathogenesis: challenges ahead and how structural biology could aid**

Despite arduous work to elucidate the pathogenesis of *M. tuberculosis*, we are not close to understand in detail all the steps taken by the bacterium to evade the immune system. Many difficulties contribute to the pace at which we progress, being the subject of intense work: culture conditions, molecular biology tools (312-314), biosafety issues, host model (315), and others intrinsic to the physiology of the genus.

There is no doubt that the pathogenesis of *M. tuberculosis* has been a challenge, and the diversity of the lineages complicates it even more. Defence mechanisms against the immune system and further manipulation are strongly related to the components of T7SSs. While the manipulation of the immune response can vary between lineages (316), one important event for the survival and progression of the illness is the translocation from the phagosome to the cytosol of macrophages. For this action, pathogenic mycobacteria require a functional ESX-1 (81). It has been shown that the machinery secretes proteins to the exterior and mediates membrane lysis through a contact-dependent mechanism, which means it should traverse the double membrane. This idea seems logical; nevertheless, the elements that could build the outer-membrane complex have not been identified. *Mycobacterium* genus is characterised by a complex cell envelope (**Figure 32**). First, there is a cytoplasmic membrane (inner membrane) made by glycerophospholipids, phosphatidylinositol mannosides (PIMs), and derivatives like lipomannans (LMs) and lipoarabinomannans (LAMs) characteristic of actinomycetes. Then, the periplasmic space contains covalently bound layers of a honeycomb structure built from peptidoglycan molecules and an arabinogalactan polymer. They, in turn, are bound to the mycolic acids present in the mycomembrane (outer membrane) (218). The latter is an impermeable layer, working as a barrier that enables the microorganism to survive under harsh environments, including exposure to antibiotics (almost 1,000 times less permeable to β -lactam antibiotics than Gram-negative cell wall (317)). Finally, the outermost layer is the capsule, a loose structure of polysaccharides and proteins implicated in the pathogenesis of *M. tuberculosis* (318). Its existence has not always been apparent; the presence of detergents such as Tween or tyloxapol, commonly used

in culture media to reduce the clumps, compromise the structure of the layer (174). Electron microscopy studies have shown that the cell envelope could be 70 nm thick, including the capsule (174). Based on structure determination of ESX-1 paralogues, EccB and MycP core components protrude 10 nm from the inner membrane, leaving ~55 nm to be occupied by the missing piece. The identification of the latter has been a challenging task as extraction from native sources has failed (167,172), suggesting that the protein complexes might not have a tight interaction. Other secretion systems have been extracted intact with a single detergent (28,319); however, the different nature of the two membranes could require unique conditions to keep the pieces together.

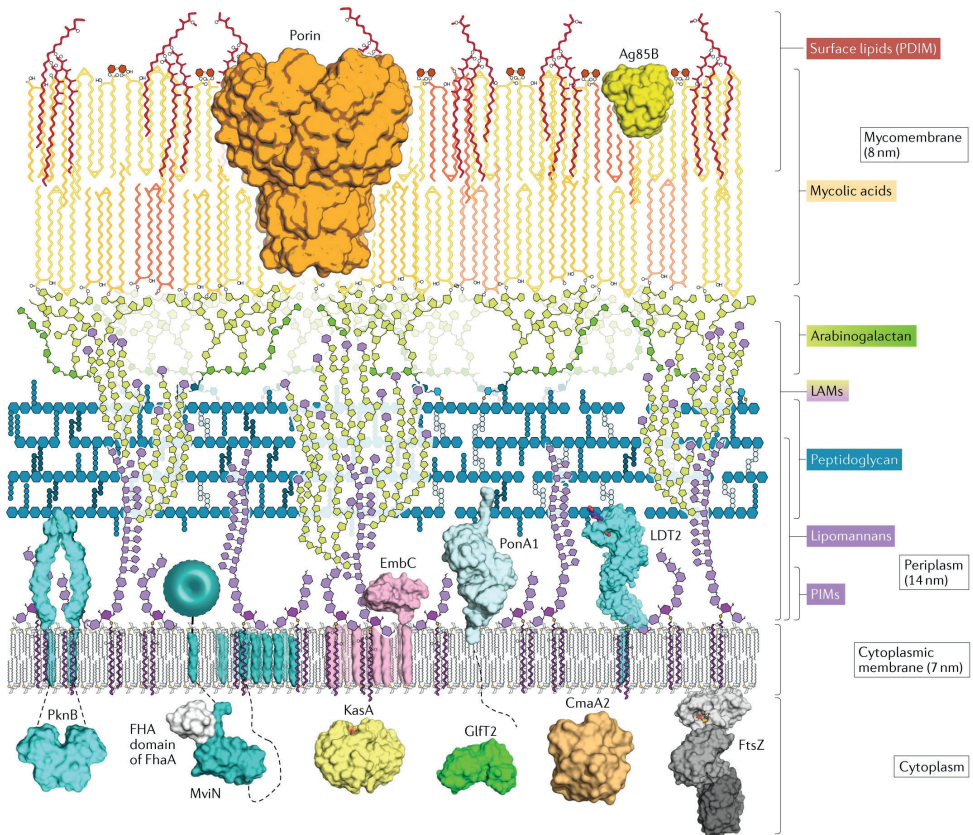


Figure 32. Model of *M. tuberculosis* cell envelope. Representation of the molecular composition of each layer. Structural information of proteins included in the figure are available in the PDB with the following entry codes: PknB (2KUI and 1MRU), Ag85B (1F0P), GlfT2 (4FIY), KasA (4C6X), FtsZ (4KWE), PonA1 (5CXW and 5HLD), EmbC (3PTY); MviN–FhaA forkhead- associated (FHA) domain complex (3OUN), CmaA2 (1KPI), LDT2 (3TUR and 5DUJ) and MspA porin (1UUN). Porin MspA is from *M. smegmatis* and is included only as an example of outer-membrane protein. No porin have been described for *M. tuberculosis*. Figure reproduced from (218).

In situ cryo-electron tomography has the advantage that no purification of the target is needed. Current projects at M4I-Maastricht University include the visualisation of T7SSs in and out of the infection context by this technique. Unfortunately, it has been unexpectedly challenging as the mycobacterial cell envelope is more electron-dense than other microorganisms (**Figure 33A**), and no evident needles stick out of the bacteria (**Figure 33B**). Likewise, the cytoplasmic elements (EccCa,b) are known to exhibit high flexibility, making it equally difficult to distinguish the machinery from the cell's interior. Additional Mega Dalton-tag or tags with heavy elements (gold) are being developed to resolve this issue.

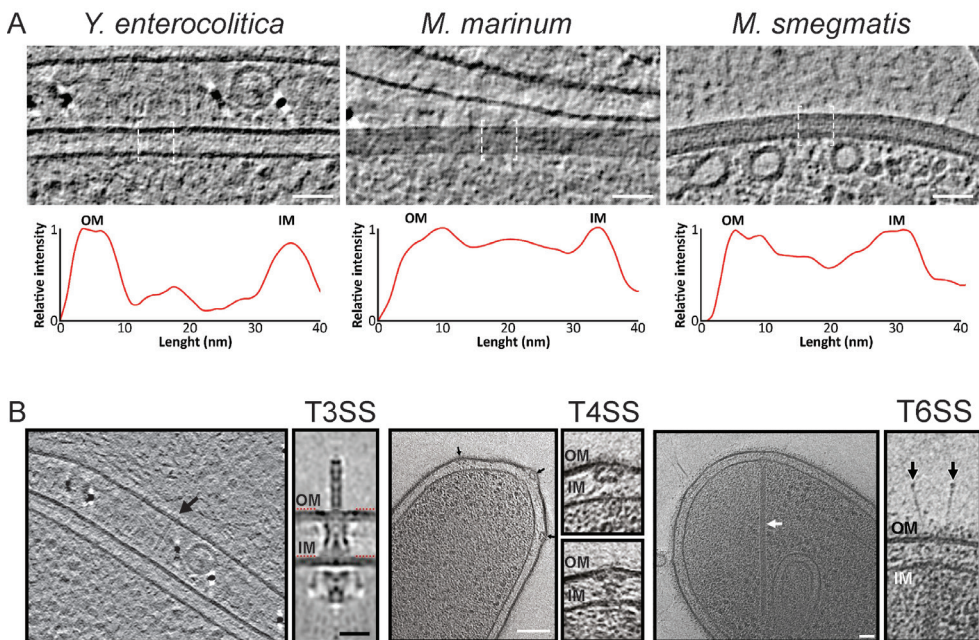


Figure 33. Mycobacterial membrane electron-density challenges the identification of T7SS. (A) Relative intensity of the different microorganism membranes showed in the micrographs depicts the difficulty to identify protein complexes within mycobacterial membranes. (B) Localization of the T3SS (*Yersinia enterocolitica*) (224), T4SS (*Legionella pneumophila*) (320), and T6SS (*Myxococcus xanthus*) (321) in the bacteria using electron microscopy techniques. Panel (A) was courtesy of Casper Berger (M4I-Maastricht University).

Tuberculosis and other health threats

We live in a world that is considered our property. No rules are imposed to avoid overpopulation and usage of limited natural resources, pollution, and extinction of flora and fauna to feed and spoil those people. Some would say that, of course, there are rules, but are there? When Japan promotes whale hunting, Canadian miners suck the natural resources of Latin-American countries while polluting the land, or European

countries criticise Brazil for destroying the Amazon forest for livestock when they are the ones buying the meat. Humans are selfish; they only look after their comfort (money), regardless of whether that comprises our planet's future. Climate change is an evident and threatening consequence of our actions but not the only one. There were predictions of a pandemic from a respiratory disease expressed for many years. They were so accurate that now that happened, a population sector believed that it was a conspiracy. Fortunately, a lot of people took action and started to change their habits. Although it has been difficult for most of us, I want to make a clear reminder that this is not the first disease and is not going to be the last. Globalisation has had a dramatic effect on the population increment and the movement of goods and people worldwide, for touristic reasons or migration caused by war, economic problems, and (soon) climatic change. In the reality that we live in, it is insufficient to have equipped and reliable health systems in rich countries when the rest of the world is unprotected. There are no borders for microbes, and the COVID-19 pandemic is proof of that. So the bad news is that the history will repeat; the question is where and how soon? I have always been uncomfortable with the amount of money that the war receives or projects intended to find life beyond our planet or build spacecraft to investigate water in Mars when the Earth is neglected. Bill Gates gave a Ted talk in 2015 where he suggested having an "army" of health workers with the same system as the military: paid people that train every day while waiting for their mission to save lives in their own country or elsewhere. He has a point: this pandemic exposed how unprepared we are to deal with wars against microscopic enemies. We need governments to invest in science and their medical system. *M. tuberculosis* is an opportunistic microorganism waiting in the dark for a compromised immune system to attack. The actual consequences of COVID-19 on TB are still to be seen, but this will continue to worsen with coming outbreaks until we develop effective therapies that could prevent and treat the disease.

Summary (English)

Proteins are the most sophisticated biomolecules in nature, and their involvement in every biological process makes them essential for the proper maintenance and performance of life. Structure determination of these elements represents invaluable information to understand their function, correct a medical problem, or use proteins for biotechnological purposes. Technological developments have made the structure determination a routine activity; however, characteristics intrinsic to proteins challenge current methods. In the last decade, electron microscopy in cryogenic conditions has gone through a resolution revolution, now comparable to other high-resolution techniques. Only limited by protein size (> 50 kDa), cryo-EM has the potential to determine the structure at (near-) atomic resolution while giving mechanistic information on the protein function. These advantages have attracted many users, explaining the exponential growth over the years. Unfortunately, this trend will slow down due to the different bottlenecks or difficulties found in the current workflow, which are the subject of intense work.

This thesis describes the development of a new instrument that could replace the semi-manual preparation of samples embedded in vitreous ice (vitrification) for cryo-EM to a fully automatic process (**chapter 2**). This new machine replaces the sample application via pipetting and excess removal with filter paper to a pin-printing method, reducing sample waste and artefacts caused by the filter paper. The incorporation of cryogen jets allows the use of premounted autogrids (metal-edged grids for sturdy handling) ready to be inserted into the microscope, avoiding manual handling of the grids post-vitrification that could lead to ice contamination or grid deformation.

During the development of the VitroJet, we were confronted with the lack of a standard protein available in large amounts to characterise the instrument. Other techniques have easily accessed proteins for training or maintenance purposes, like the lysozyme for MX or BSA for SAXS. In cryo-EM, ferritin is the protein of choice; its symmetry, solubility, and size allow structure determination with a few micrographs enabling prompt characterisation. Unfortunately, the commercially available ferritin (from horse spleen) lacks stability, while stable human ferritin has a complicated purification protocol that will limit its access to laboratories with specific biochemical expertise. With this in mind, we established a high-yielded and straightforward purification protocol for mycobacterial ferritin BfrB (**chapter 3**). As a proof of concept, Bachelor students with limited biochemical expertise carried out the protocol leading to >50 mg of protein with good purity. Structure determination led to a high-resolution model where conformation changes of amino acids and flexible regions were visible, valuable information that could aid to understand how the function is carried out. This chapter had a double intention: to provide the cryo-EM community a standard protein that could serve as a workhorse for instrument maintenance and development and structural characterisation of a mycobacterial protein essential for the pathogenesis of *M. tuberculosis*, which can serve as a therapeutic target.

In the following chapters, I used structural techniques, particularly cryo-EM, to describe virulence factors of *M. tuberculosis* and understand their role in the functioning of the type VII secretion system and possibly in the pathogenesis of this microorganism. One big mystery that surrounds T7SS is the elements that compose the outer-membrane portion, critical information for the proper understanding of the system. EspB protein has been proposed to be one of these elements. **Chapter 4** describes the conditions that trigger the oligomerisation of EspB and the characteristics that support the hypothesis that this protein is part of the missing components. To further understand the function of EspB and its regulation mechanisms, we studied its binding partner, EspK, another virulence factor of *M. tuberculosis*. Low-resolution studies by SAXS, complemented with CD and sequence alignments, described the presence of two structured domains linked by an unstructured region, characteristic of all the EspK proteins from the *Mycobacterium* genus (**chapter 5**). Previous studies described the interaction between the C-terminal domain and EspB; for this reason, it was further characterised at high resolution by X-ray crystallography in complex with EspB (**chapter 6**). Such structure resembled the PE-PPE bound to their chaperone EspG, suggesting a possible role of EspK as a chaperone of EspB. This hypothesis was supported by the fact that EspK disrupted the EspB oligomer, ensuring the latter to be in a secretion-competent state.

Overall, this work intended to demonstrate the power of structural visualisation applied in a specific health problem, tuberculosis, to close the knowledge gap that will lead to efficient therapy one day. Likewise, it is imperative to continue the development of structural techniques to expand our understanding of the mechanisms underlying nature.

Samenvatting (Dutch)

Eiwitten zijn de meest geavanceerde biomoleculen in de natuur, en hun betrokkenheid bij elk biologisch proces maakt ze essentieel voor het juiste onderhoud en de goede ontwikkeling van het leven. Het beschrijven van de structuur van deze elementen omvat onschatbare informatie om hun functie te begrijpen, een medisch probleem te corrigeren, of eiwitten voor biotechnologische doeleinden te gebruiken. Dankzij technologische ontwikkelingen is de structuurbepaling een routine-activiteit geworden; niettemin dagen intrinsieke kenmerken van eiwitten de huidige methoden uit. In het afgelopen decennium heeft elektronenmicroscopie in cryogene omstandigheden een resolutie-revolutie doorgemaakt, nu vergelijkbaar met andere hoge resolutie technieken. Alleen beperkt door de eiwitgrootte (> 50 kDa), heeft cryo-EM het potentieel om de structuur te bepalen met (bijna-) atomaire resolutie en tegelijkertijd mechanistische informatie te geven over de eiwitfunctie. Deze voordelen hebben veel gebruikers aangetrokken, wat de exponentiële groei door de jaren heen verklaart. Helaas zal deze trend langzamer gaan vanwege de verschillende knelpunten of moeilijkheden in de huidige *workflow*, wat het onderwerp is van intensief werk.

Dit proefschrift beschrijft de ontwikkeling van een nieuw instrument dat de semi-manuele voorbereiding van monsters ingebed in glasachtig ijs (vitrificatie) voor cryo-EM kan vervangen door een volautomatisch proces (**hoofdstuk 2**). Deze nieuwe machine vervangt de monstertoepassing via pipetteren en overtollige verwijdering met filterpapier naar een pin-printmethode, waardoor monsterverspilling en artefacten die door het filterpapier worden veroorzaakt, worden verminderd. De integratie van cryogene jets maakt gebruik van voorgemonteerde autogrids (roosters met metalen randen voor stevige hantering) mogelijk die gelijk in de microscoop kunnen worden geplaatst. Hierdoor wordt handmatige hantering van de roosters na verglazing vermeden, wat kan leiden tot ijsverontreiniging of roostervorming.

Tijdens de ontwikkeling van de VitroJet werden we geconfronteerd met het ontbreken van een standaard eiwit dat in grote hoeveelheden beschikbaar was om het instrument te karakteriseren. Andere technieken hebben gemakkelijk toegang tot eiwitten voor trainings- of onderhoudsdoeleinden, zoals het lysozyme voor MX of BSA voor SAXS. Bij cryo-EM is ferritine het eiwit bij uitstek; zijn symmetrie, oplosbaarheid en grootte maken zijn structuurbepaling mogelijk met een paar microfoto's, waardoor snelle karakterisering mogelijk is. Helaas mist het commercieel verkrijgbare ferritine (uit paardenmilt) stabiliteit, terwijl stabiel menselijk ferritine een ingewikkeld zuiveringsprotocol heeft dat de toegang tot laboratoria met specifieke biochemische expertise beperkt. Met dit in gedachten hebben we een hoogrenderend en eenvoudig zuiveringsprotocol opgesteld voor mycobacteriële ferritine BfrB (**hoofdstuk 3**). Als *proof of concept* voerden bachelor studenten met beperkte biochemische expertise het protocol uit dat leidde tot >50 mg eiwit met een goede zuiverheid. De structuurbepaling van BfrB leidde tot een model met hoge resolutie waarin conformatieveranderingen van aminozuren en flexibele regio's

zichtbaar waren; waardevolle informatie die zou kunnen helpen begrijpen hoe de functie wordt uitgevoerd. Dit hoofdstuk had een dubbele bedoeling: de cryo-EM-gemeenschap voorzien van een standaard eiwit dat zou kunnen dienen als een werkpaard voor het onderhoud en de ontwikkeling van instrumenten en de structurele karakterisering van een mycobacterieel eiwit dat essentieel voor de pathogenese van *M. tuberculosis* is, en kan dienen als een therapeutisch doelwit.

In de volgende hoofdstukken heb ik structurele technieken gebruikt, met name cryo-EM, om virulentiefactoren van *M. tuberculosis* te beschrijven en hun rol in het functioneren van het type VII-secretiesysteem en mogelijk in de pathogenese van dit micro-organisme te begrijpen. Een groot mysterie dat de T7SS omringt, zijn de elementen die het buitenste membraangedeelte vormen, essentiële informatie voor een goed begrip van het systeem. Er is voorgesteld dat het EspB-eiwit een van deze elementen is. Het EspB-eiwit is voorgesteld als één van deze elementen.. Hoofdstuk 4 beschrijft de omstandigheden die de oligomerisatie van EspB veroorzaken en de kenmerken die de hypothese ondersteunen dat dit eiwit deel uitmaakt van de ontbrekende componenten. Om de functie van EspB en zijn regulatiemechanismen beter te begrijpen, hebben we zijn bindende partner, EspK, een andere virulentiefactor van *M. tuberculosis*, bestudeerd. Lage-resolutie studies door SAXS, aangevuld met CD en sequentie-uitlijning, beschreven de aanwezigheid van twee gestructureerde domeinen die verbonden zijn door een ongestructureerde regio, kenmerkend voor het *Mycobacterium* genus (**hoofdstuk 5**). Eerdere studies beschreven de interactie tussen het C-terminale domein en EspB; om deze reden werd het verder gekarakteriseerd met hoge resolutie door röntgenkristallografie in een complex met EspB (**hoofdstuk 6**). Een dergelijke structuur leek op de PE-PPE gebonden aan hun chaperonne EspG, wat een mogelijke rol van EspK als chaperonne van EspB suggereert. Deze hypothese werd ondersteund door het feit dat EspK het EspB-oligomeer verstoort, waardoor de laatste in een secretiecompetente toestand verkeerde.

Over het algemeen was dit werk bedoeld om de kracht van structurele visualisatie te demonstreren toegepast bij een specifiek gezondheidsprobleem, tuberculose, om de kenniskloof te dichten die op een dag tot efficiënte therapie zal leiden. Tevens is het absoluut noodzakelijk om de ontwikkeling van structurele technieken voort te zetten om ons begrip van de onderliggende mechanismen van de natuur te vergroten.

Resumen (Spanish)

Las proteínas son la biomolécula más sofisticada en la naturaleza, y su participación en cada uno de los procesos biológicos las hacen esenciales para un adecuado funcionamiento y mantenimiento de la vida. La determinación estructural de estos elementos representa información invaluable para entender su función, corregir un problema médico, o utilizar la proteína para fines biotecnológicos. Los avances tecnológicos han hecho de la determinación estructural una actividad rutinaria; sin embargo, características intrínsecas de las proteínas desafían los métodos actuales. En la última década, la microscopía electrónica en condiciones criogénicas ha experimentado una *revolución de la resolución*, ahora comparable a otras técnicas de alta resolución. Sólo limitado por el tamaño de la proteína (> 50 kDa), crio-EM tiene el potencial de determinar la estructura a una resolución (casi) atómica al mismo tiempo que proporciona información sobre el mecanismo de acción de la proteína. Estas ventajas han atraído a muchos usuarios, lo que explica el crecimiento exponencial a lo largo de los años. Lamentablemente, esta tendencia se desacelerará debido a los diferentes cuellos de botella o dificultades que se tienen en el flujo de trabajo actual, los cuales son objeto de un intenso trabajo.

Esta tesis describe el desarrollo de un nuevo instrumento que podría reemplazar la preparación semi-manual de muestras embebidas en hielo vítreo (vitrificación) para crio-EM a un proceso completamente automático (**capítulo 2**). Esta nueva máquina reemplaza la aplicación de la muestra mediante pipeteo y eliminación del exceso con papel de filtro por un método de impresión con alfileres (del inglés, pin-printing), lo que reduce el desperdicio de muestra y los artefactos causados por el papel de filtro. La incorporación de chorros criogénicos permite el uso de autorejillas (rejillas con borde metálico para un manejo robusto) premontadas listas para ser insertadas en el microscopio, evitando la manipulación manual de las rejillas post-vitrificación que podría provocar contaminación por hielo o la deformación de la rejilla.

Durante el desarrollo del VitroJet, nos enfrentamos a la falta de una proteína estándar disponible en grandes cantidades para caracterizar el instrumento. Otras técnicas tienen proteínas de fácil acceso con fines de entrenamiento del personal o mantenimiento de equipos, como la lisozima para MX o BSA para SAXS. En crio-EM, la ferritina es la proteína de elección; su simetría, solubilidad y tamaño permiten la determinación de la estructura con unas pocas micrografías que permiten una rápida caracterización. Desafortunadamente, la ferritina disponible comercialmente (de bazo de caballo) carece de estabilidad, mientras que la ferritina humana es estable, pero tiene un protocolo de purificación complicado que limitará su acceso a laboratorios con experiencia bioquímica específica. Con esto en mente, establecimos un protocolo de purificación sencillo y de alto rendimiento para la ferritina micobacteriana BfrB (**capítulo 3**). Como prueba de este concepto, estudiantes de licenciatura con poca experiencia bioquímica llevaron a cabo el protocolo obteniendo > 50 mg de proteína con buena pureza. La determinación de la estructura condujo a un modelo de alta resolución donde se observaron cambios

conformacionales de aminoácidos y regiones flexible de la proteína, información valiosa que podría ayudar a comprender cómo se lleva a cabo la función. Este capítulo tenía una doble intención: proporcionar a la comunidad crio-EM una proteína estándar que pudiera servir como caballo de batalla para el mantenimiento y desarrollo de instrumentos y la caracterización estructural de una proteína micobacteriana esencial para la patogénesis de *M. tuberculosis*, la cual puede fungir como blanco terapéutico.

En los siguientes capítulos, utilicé técnicas estructurales, particularmente crio-EM, para describir factores de virulencia de *M. tuberculosis* y comprender su papel en el funcionamiento del sistema de secreción de tipo VII y posiblemente en la patogénesis de este microorganismo. Un gran misterio que envuelve a esta bio-máquina son los elementos que componen la parte de la membrana exterior, información crítica para la correcta comprensión del sistema. Se ha propuesto que la proteína EspB sea uno de estos elementos. El **capítulo 4** describe las condiciones que desencadenan la oligomerización de EspB y las características que apoyan la hipótesis de que esta proteína es uno de los componentes faltantes. Para comprender mejor la función de EspB y sus mecanismos de regulación, estudiamos su ligando, EspK, otro factor de virulencia de *M. tuberculosis*. Estudios de SAXS de baja resolución, complementados con CD y alineamientos de secuencia, describieron la presencia de dos dominios estructurados unidos por una región no estructurada, característico de todas las proteínas EspK del género *Mycobacterium* (**capítulo 5**). Estudios previos describieron la interacción entre el dominio C-terminal y la proteína EspB; por este motivo, se continuó con la caracterización a alta resolución mediante cristalografía de rayos X en complejo con EspB. Dicha estructura presentó un parecido a la proteína PE-PPE unido a su chaperona EspG, lo que sugiere un posible papel de EspK como chaperona de EspB. Esta hipótesis fue apoyada por el hecho de que EspK disoció el oligómero de EspB, asegurando que este último estuviera en un estado de secreción competente.

En general, este trabajo pretendía demostrar el poder de la visualización estructural aplicada en un problema de salud específico, la tuberculosis, para cerrar la brecha de conocimiento que conducirá a una terapia eficiente algún día. Asimismo, es imperativo continuar el desarrollo de técnicas estructurales para ampliar nuestra comprensión de los mecanismos que subyacen a la naturaleza.

CHAPTER 9



Appendices

References

1. Wright, P. E., and Dyson, H. J. (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* **293**, 321-331
2. Uversky, V. N. (2019) Protein intrinsic disorder and structure-function continuum. *Prog Mol Biol Transl Sci* **166**, 1-17
3. Perutz, M. F. (1963) X-ray analysis of hemoglobin. *Science* **140**, 863-869
4. Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., and Phillips, D. C. (1958) A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181**, 662-666
5. Dunker, A. K., Obradovic, Z., Romero, P., Garner, E. C., and Brown, C. J. (2000) Intrinsic protein disorder in complete genomes. *Genome Inform Ser Workshop Genome Inform* **11**, 161-171
6. Berman, H., Henrick, K., and Nakamura, H. (2003) Announcing the worldwide Protein Data Bank. *Nat Struct Biol* **10**, 980
7. Rupp, B. (2009) *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*, CRC Press
8. Yu, H. (1999) Extending the size limit of protein nuclear magnetic resonance. *Proc Natl Acad Sci U S A* **96**, 332-334
9. Netzer, W. J., and Hartl, F. U. (1998) Protein folding in the cytosol: chaperonin-dependent and -independent mechanisms. *Trends Biochem Sci* **23**, 68-73
10. Callaway, E. (2020) Revolutionary cryo-EM is taking over structural biology. *Nature* **578**, 201
11. Jahagirdar, D., Jha, V., Basu, K., Gomez-Blanco, J., Vargas, J., and Ortega, J. (2020) Alternative conformations and motions adopted by 30S ribosomal subunits visualized by cryo-electron microscopy. *RNA* **26**, 2017-2030
12. Milne, J. L., Borgnia, M. J., Bartesaghi, A., Tran, E. E., Earl, L. A., Schauder, D. M., Lengyel, J., Pierson, J., Patwardhan, A., and Subramaniam, S. (2013) Cryo-electron microscopy--a primer for the non-microscopist. *FEBS J* **280**, 28-45
13. Faruqi, A. R., and McMullan, G. (2011) Electronic detectors for electron microscopy. *Q Rev Biophys* **44**, 357-390
14. Li, X., Mooney, P., Zheng, S., Booth, C. R., Braunjfeld, M. B., Gubbens, S., Agard, D. A., and Cheng, Y. (2013) Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat Methods* **10**, 584-590
15. Scheres, S. H. (2014) Beam-induced motion correction for sub-megadalton cryo-EM particles. *Elife* **3**, e03665
16. Scheres, S. H. (2012) RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180**, 519-530
17. Punjani, A., Rubinstein, J. L., Fleet, D. J., and Brubaker, M. A. (2017) cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296
18. Danev, R., Yanagisawa, H., and Kikkawa, M. (2019) Cryo-Electron Microscopy Methodology: Current Aspects and Future Directions. *Trends Biochem Sci* **44**, 837-848

19. Nakane, T., Kotecha, A., Sente, A., McMullan, G., Masiulis, S., Brown, P. M. G. E., Grigoras, I. T., Malinauskaite, L., Malinauskas, T., Miehl, J., Uchanski, T., Yu, L. B., Karia, D., Pechnikova, E. V., de Jong, E., Keizer, J., Bischoff, M., McCormack, J., Tiemeijer, P., Hardwick, S. W., Chirgadze, D. Y., Murshudov, G., Aricescu, A. R., and Scheres, S. H. W. (2020) Single-particle cryo-EM at atomic resolution. *Nature* **587**, 152-+
20. Yip, K. M., Fischer, N., Paknia, E., Chari, A., and Stark, H. (2020) Atomic-resolution protein structure determination by cryo-EM. *Nature* **587**, 157-+
21. Zhang, Y., Tammara, R., Peters, P. J., and Ravelli, R. B. G. (2020) Could Egg White Lysozyme be Solved by Single Particle Cryo-EM? *J Chem Inf Model* **60**, 2605-2613
22. Lyumkis, D. (2019) Challenges and opportunities in cryo-EM single-particle analysis. *J Biol Chem* **294**, 5181-5197
23. Noble, A. J., Wei, H., Dandey, V. P., Zhang, Z., Tan, Y. Z., Potter, C. S., and Carragher, B. (2018) Reducing effects of particle adsorption to the air-water interface in cryo-EM. *Nat Methods* **15**, 793-795
24. Baldwin, P. R., and Lyumkis, D. (2020) Non-uniformity of projection distributions attenuates resolution in Cryo-EM. *Prog Biophys Mol Biol* **150**, 160-183
25. Russo, C. J., and Passmore, L. A. (2014) Electron microscopy: Ultrastable gold substrates for electron cryomicroscopy. *Science* **346**, 1377-1380
26. Russo, C. J., and Passmore, L. A. (2014) Controlling protein adsorption on graphene for cryo-EM using low-energy hydrogen plasmas. *Nat Methods* **11**, 649-652
27. Weissenberger, G., Henderikx, R. J. M., and Peters, P. J. (2021) Understanding the invisible hands of sample preparation for cryo-EM. *Nat Methods* **18**, 463-471
28. Hu, J., Worrall, L. J., Hong, C., Vuckovic, M., Atkinson, C. E., Caveney, N., Yu, Z., and Strynadka, N. C. J. (2018) Cryo-EM analysis of the T3S injectisome reveals the structure of the needle and open secretin. *Nat Commun* **9**, 3840
29. Daury, L., Orange, F., Taveau, J. C., Verchere, A., Monlezun, L., Gounou, C., Marreddy, R. K., Picard, M., Broutin, I., Pos, K. M., and Lambert, O. (2016) Tripartite assembly of RND multidrug efflux pumps. *Nat Commun* **7**, 10731
30. Cao, E., Liao, M., Cheng, Y., and Julius, D. (2013) TRPV1 structures in distinct conformations reveal activation mechanisms. *Nature* **504**, 113-118
31. Liao, M., Cao, E., Julius, D., and Cheng, Y. (2013) Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature* **504**, 107-112
32. Fischer, N., Konevega, A. L., Wintermeyer, W., Rodnina, M. V., and Stark, H. (2010) Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. *Nature* **466**, 329-333
33. World Health Organization. (2021) Global tuberculosis report 2021. Geneva: World Health Organization
34. Behr, M. A., Edelstein, P. H., and Ramakrishnan, L. (2019) Is Mycobacterium tuberculosis infection life long? *BMJ* **367**, l5770
35. Datiko, D. G., Jerene, D., and Suarez, P. (2020) Stigma matters in ending tuberculosis: Nationwide survey of stigma in Ethiopia. *BMC Public Health* **20**, 190

36. Tanimura, T., Jaramillo, E., Weil, D., Raviglione, M., and Lonroth, K. (2014) Financial burden for tuberculosis patients in low- and middle-income countries: a systematic review. *Eur Respir J* **43**, 1763-1775
37. Zumla, A. I., Gillespie, S. H., Hoelscher, M., Philips, P. P., Cole, S. T., Abubakar, I., McHugh, T. D., Schito, M., Maeurer, M., and Nunn, A. J. (2014) New antituberculosis drugs, regimens, and adjunct therapies: needs, advances, and future prospects. *Lancet Infect Dis* **14**, 327-340
38. Awofeso, N. (2008) Anti-tuberculosis medication side-effects constitute major factor for poor adherence to tuberculosis treatment. *Bull World Health Organ* **86**, B-D
39. Schaberg, T., Rebhan, K., and Lode, H. (1996) Risk factors for side-effects of isoniazid, rifampin and pyrazinamide in patients hospitalized for pulmonary tuberculosis. *Eur Respir J* **9**, 2026-2030
40. Madacki, J., Orgeur, M., Mas Fiol, G., Frigui, W., Ma, L., and Brosch, R. (2021) ESX-1-Independent Horizontal Gene Transfer by *Mycobacterium tuberculosis* Complex Strains. *mBio* **12**
41. Gandhi, N. R., Nunn, P., Dheda, K., Schaaf, H. S., Zignol, M., van Soolingen, D., Jensen, P., and Bayona, J. (2010) Multidrug-resistant and extensively drug-resistant tuberculosis: a threat to global control of tuberculosis. *Lancet* **375**, 1830-1843
42. World Health Organization. (2019) World Health Statistics 2019. Geneva: World Health Organization
43. World Health Organization. (2020) Module 4: treatment - drug-resistant tuberculosis treatment. in *WHO consolidated guidelines on tuberculosis*
44. Dheda, K., Gumbo, T., Maartens, G., Dooley, K. E., McNerney, R., Murray, M., Furin, J., Nardell, E. A., London, L., Lessem, E., Theron, G., van Helden, P., Niemann, S., Merker, M., Dowdy, D., Van Rie, A., Siu, G. K., Pasipanodya, J. G., Rodrigues, C., Clark, T. G., Sirgel, F. A., Esmail, A., Lin, H. H., Atre, S. R., Schaaf, H. S., Chang, K. C., Lange, C., Nahid, P., Udawadia, Z. F., Horsburgh, C. R., Jr., Churchyard, G. J., Menzies, D., Hesselning, A. C., Nuermberger, E., McIlleron, H., Fennelly, K. P., Goemaere, E., Jaramillo, E., Low, M., Jara, C. M., Padayatchi, N., and Warren, R. M. (2017) The epidemiology, pathogenesis, transmission, diagnosis, and management of multidrug-resistant, extensively drug-resistant, and incurable tuberculosis. *Lancet Respir Med*
45. All-Party Parliamentary Group. (2015) The Price of a Pandemic: Counting the Cost of MDR-TB. All-Party Parliamentary Group on Global tuberculosis
46. McIlleron, H., Meintjes, G., Burman, W. J., and Maartens, G. (2007) Complications of antiretroviral therapy in patients with tuberculosis: drug interactions, toxicity, and immune reconstitution inflammatory syndrome. *J Infect Dis* **196 Suppl 1**, S63-75
47. Stop TB Partnership Secretariat. (2020) The TB response is heavily impacted by the COVID-19 pandemic. http://stoptb.org/news/stories/2020/ns20_014.html
48. Ministry of Health and Family Welfare. (2021) India TB report 2021.
49. Alagna, R., Besozzi, G., Codecasa, L. R., Gori, A., Migliori, G. B., Raviglione, M., and Cirillo, D. M. (2020) Celebrating World Tuberculosis Day at the time of COVID-19. *Eur Respir J* **55**

50. Cilloni, L., Fu, H., Vesga, J. F., Dowdy, D., Pretorius, C., Ahmedov, S., Nair, S. A., Mosneaga, A., Masini, E., Sahu, S., and Arinaminpathy, N. (2020) The potential impact of the COVID-19 pandemic on the tuberculosis epidemic a modelling analysis. *EClinicalMedicine* **28**, 100603
51. Tadolini, M., Codecasa, L. R., Garcia-Garcia, J. M., Blanc, F. X., Borisov, S., Alffenaar, J. W., Andrejak, C., Bachez, P., Bart, P. A., Belilovski, E., Cardoso-Landivar, J., Centis, R., D'Ambrosio, L., Luiza De Souza-Galvao, M., Dominguez-Castellano, A., Dourmane, S., Frechet Jachym, M., Froissart, A., Giacomet, V., Goletti, D., Gard, S., Gualano, G., Izadifar, A., Le Du, D., Marin Royo, M., Mazza-Stalder, J., Motta, I., Ong, C. W. M., Palmieri, F., Riviere, F., Rodrigo, T., Silva, D. R., Sanchez-Montalva, A., Saporiti, M., Scarpellini, P., Schlemmer, F., Spanevello, A., Sumarokova, E., Taberner, E., Tambyah, P. A., Tiberi, S., Torre, A., Visca, D., Zabaleta Murguiondo, M., Sotgiu, G., and Migliori, G. B. (2020) Active tuberculosis, sequelae and COVID-19 co-infection: first cohort of 49 cases. *Eur Respir J* **56**
52. Chen, Y., Wang, Y., Fleming, J., Yu, Y., Gu, Y., Liu, C., Fan, L., Wang, X., Cheng, M., Bi, L., and Liu, Y. (2020) Active or latent tuberculosis increases susceptibility to COVID-19 and disease severity. *medRxiv*
53. Centers for Disease Control and Prevention. (2021) World Tuberculosis (TB) Day 2021. https://www.cdc.gov/tb/features/wtbd/2021WTBD_Feature.html
54. Ford, L. (2020) Millions predicted to develop tuberculosis as result of Covid-19 lockdown. *The Guardian*
55. Smith, N. H., Hewinson, R. G., Kremer, K., Brosch, R., and Gordon, S. V. (2009) Myths and misconceptions: the origin and evolution of Mycobacterium tuberculosis. *Nat Rev Microbiol* **7**, 537-544
56. Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K. E., Kato-Maeda, M., Parkhill, J., Malla, B., Berg, S., Thwaites, G., Yeboah-Manu, D., Bothamley, G., Mei, J., Wei, L., Bentley, S., Harris, S. R., Niemann, S., Diel, R., Aseffa, A., Gao, Q., Young, D., and Gagneux, S. (2013) Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. *Nat Genet* **45**, 1176-1182
57. Gutierrez, M. C., Brisse, S., Brosch, R., Fabre, M., Omais, B., Marmiesse, M., Supply, P., and Vincent, V. (2005) Ancient origin and gene mosaicism of the progenitor of Mycobacterium tuberculosis. *PLoS Pathog* **1**, e5
58. Chisholm, R. H., Trauer, J. M., Curnoe, D., and Tanaka, M. M. (2016) Controlled fire use in early humans might have triggered the evolutionary emergence of tuberculosis. *Proc Natl Acad Sci U S A* **113**, 9051-9056
59. Brosch, R., Gordon, S. V., Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K., Garnier, T., Gutierrez, C., Hewinson, G., Kremer, K., Parsons, L. M., Pym, A. S., Samper, S., van Soolingen, D., and Cole, S. T. (2002) A new evolutionary scenario for the Mycobacterium tuberculosis complex. *Proc Natl Acad Sci U S A* **99**, 3684-3689

60. Bos, K. I., Harkins, K. M., Herbig, A., Coscolla, M., Weber, N., Comas, I., Forrest, S. A., Bryant, J. M., Harris, S. R., Schuenemann, V. J., Campbell, T. J., Majander, K., Wilbur, A. K., Guichon, R. A., Wolfe Steadman, D. L., Cook, D. C., Niemann, S., Behr, M. A., Zumarraga, M., Bastida, R., Huson, D., Nieselt, K., Young, D., Parkhill, J., Buikstra, J. E., Gagneux, S., Stone, A. C., and Krause, J. (2014) Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* **514**, 494-497
61. Hershkovitz, I., Donoghue, H. D., Minnikin, D. E., Besra, G. S., Lee, O. Y., Gernaey, A. M., Galili, E., Eshed, V., Greenblatt, C. L., Lemma, E., Bar-Gal, G. K., and Spigelman, M. (2008) Detection and molecular characterization of 9,000-year-old Mycobacterium tuberculosis from a Neolithic settlement in the Eastern Mediterranean. *PLoS One* **3**, e3426
62. Baker, O., Lee, O. Y., Wu, H. H., Besra, G. S., Minnikin, D. E., Llewellyn, G., Williams, C. M., Maixner, F., O'Sullivan, N., Zink, A., Chamel, B., Khawam, R., Coqueugnot, E., Helmer, D., Le Mort, F., Perrin, P., Gourichon, L., Dutailly, B., Palfi, G., Coqueugnot, H., and Dutour, O. (2015) Human tuberculosis predates domestication in ancient Syria. *Tuberculosis (Edinb)* **95 Suppl 1**, S4-S12
63. Lee, O. Y., Wu, H. H., Donoghue, H. D., Spigelman, M., Greenblatt, C. L., Bull, I. D., Rothschild, B. M., Martin, L. D., Minnikin, D. E., and Besra, G. S. (2012) Mycobacterium tuberculosis complex lipid virulence factors preserved in the 17,000-year-old skeleton of an extinct bison, *Bison antiquus*. *PLoS One* **7**, e41923
64. Gagneux, S. (2018) Ecology and evolution of Mycobacterium tuberculosis. *Nat Rev Microbiol* **16**, 202-213
65. de Jong, B. C., Antonio, M., and Gagneux, S. (2010) Mycobacterium africanum--review of an important cause of human tuberculosis in West Africa. *PLoS Negl Trop Dis* **4**, e744
66. Blaser, M. J., and Kirschner, D. (2007) The equilibria that allow bacterial persistence in human hosts. *Nature* **449**, 843-849
67. Merker, M., Blin, C., Mona, S., Duforet-Frebourg, N., Lecher, S., Willery, E., Blum, M. G., Rusch-Gerdes, S., Mokrousov, I., Aleksic, E., Allix-Beguec, C., Antierens, A., Augustynowicz-Kopec, E., Ballif, M., Barletta, F., Beck, H. P., Barry, C. E., 3rd, Bonnet, M., Borroni, E., Campos-Herrero, I., Cirillo, D., Cox, H., Crowe, S., Crudu, V., Diel, R., Drobniewski, F., Fauville-Dufaux, M., Gagneux, S., Ghebremichael, S., Hanekom, M., Hoffner, S., Jiao, W. W., Kalon, S., Kohl, T. A., Kontsevaya, I., Lillebaek, T., Maeda, S., Nikolayevskyy, V., Rasmussen, M., Rastogi, N., Samper, S., Sanchez-Padilla, E., Savic, B., Shamputa, I. C., Shen, A., Sng, L. H., Stakenas, P., Toit, K., Varaine, F., Vukovic, D., Wahl, C., Warren, R., Supply, P., Niemann, S., and Wirth, T. (2015) Evolutionary history and global spread of the Mycobacterium tuberculosis Beijing lineage. *Nat Genet* **47**, 242-249
68. Ford, C. B., Shah, R. R., Maeda, M. K., Gagneux, S., Murray, M. B., Cohen, T., Johnston, J. C., Gardy, J., Lipsitch, M., and Fortune, S. M. (2013) Mycobacterium tuberculosis mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. *Nat Genet* **45**, 784-790
69. Hanekom, M., Gey van Pittius, N. C., McEvoy, C., Victor, T. C., Van Helden, P. D., and Warren, R. M. (2011) Mycobacterium tuberculosis Beijing genotype: a template for success. *Tuberculosis (Edinb)* **91**, 510-523

70. Schragar, L. K., Vekemens, J., Drager, N., Lewinsohn, D. M., and Olesen, O. F. (2020) The status of tuberculosis vaccine development. *Lancet Infect Dis* **20**, e28-e37
71. Cambier, C. J., Takaki, K. K., Larson, R. P., Hernandez, R. E., Tobin, D. M., Urdahl, K. B., Cosma, C. L., and Ramakrishnan, L. (2014) Mycobacteria manipulate macrophage recruitment through coordinated use of membrane lipids. *Nature* **505**, 218-222
72. Iakobachvili, N., and Peters, P. J. (2017) Humans in a Dish: The Potential of Organoids in Modeling Immunity and Infectious Diseases. *Front Microbiol* **8**, 2402
73. Seto, S., Matsumoto, S., Ohta, I., Tsujimura, K., and Koide, Y. (2009) Dissection of Rab7 localization on Mycobacterium tuberculosis phagosome. *Biochem Biophys Res Commun* **387**, 272-277
74. van der Wel, N., Hava, D., Houben, D., Fluitsma, D., van Zon, M., Pierson, J., Brenner, M., and Peters, P. J. (2007) M. tuberculosis and M. leprae translocate from the phagolysosome to the cytosol in myeloid cells. *Cell* **129**, 1287-1298
75. Keane, J., Remold, H. G., and Kornfeld, H. (2000) Virulent Mycobacterium tuberculosis strains evade apoptosis of infected alveolar macrophages. *J Immunol* **164**, 2016-2020
76. Abdallah, A. M., Gey van Pittius, N. C., Champion, P. A., Cox, J., Luirink, J., Vandenbroucke-Grauls, C. M., Appelmelk, B. J., and Bitter, W. (2007) Type VII secretion-mycobacteria show the way. *Nat Rev Microbiol* **5**, 883-891
77. Mahairas, G. G., Sabo, P. J., Hickey, M. J., Singh, D. C., and Stover, C. K. (1996) Molecular analysis of genetic differences between Mycobacterium bovis BCG and virulent M. bovis. *J Bacteriol* **178**, 1274-1282
78. Pym, A. S., Brodin, P., Brosch, R., Huerre, M., and Cole, S. T. (2002) Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines Mycobacterium bovis BCG and Mycobacterium microti. *Mol Microbiol* **46**, 709-717
79. Stanley, S. A., Johndrow, J. E., Manzanillo, P., and Cox, J. S. (2007) The Type I IFN response to infection with Mycobacterium tuberculosis requires ESX-1-mediated secretion and contributes to pathogenesis. *J Immunol* **178**, 3143-3152
80. Abdallah, A. M., Bestebroer, J., Savage, N. D., de Punder, K., van Zon, M., Wilson, L., Korbee, C. J., van der Sar, A. M., Ottenhoff, T. H., van der Wel, N. N., Bitter, W., and Peters, P. J. (2011) Mycobacterial secretion systems ESX-1 and ESX-5 play distinct roles in host cell death and inflammasome activation. *J Immunol* **187**, 4744-4753
81. Houben, D., Demangel, C., van Ingen, J., Perez, J., Baldeon, L., Abdallah, A. M., Caleechurn, L., Bottai, D., van Zon, M., de Punder, K., van der Laan, T., Kant, A., Bossers-de Vries, R., Willemsen, P., Bitter, W., van Soolingen, D., Brosch, R., van der Wel, N., and Peters, P. J. (2012) ESX-1-mediated translocation to the cytosol controls virulence of mycobacteria. *Cell Microbiol* **14**, 1287-1298
82. Simeone, R., Bobard, A., Lippmann, J., Bitter, W., Majlessi, L., Brosch, R., and Enninga, J. (2012) Phagosomal rupture by Mycobacterium tuberculosis results in toxicity and host cell death. *PLoS Pathog* **8**, e1002507

83. Ates, L. S., Ummels, R., Commandeur, S., van de Weerd, R., Sparrius, M., Weerdenburg, E., Alber, M., Kalscheuer, R., Piersma, S. R., Abdallah, A. M., Abd El Ghany, M., Abdel-Haleem, A. M., Pain, A., Jimenez, C. R., Bitter, W., and Houben, E. N. (2015) Essential Role of the ESX-5 Secretion System in Outer Membrane Permeability of Pathogenic Mycobacteria. *PLoS Genet* **11**, e1005190
84. Siegrist, M. S., Unnikrishnan, M., McConnell, M. J., Borowsky, M., Cheng, T. Y., Siddiqi, N., Fortune, S. M., Moody, D. B., and Rubin, E. J. (2009) Mycobacterial Esx-3 is required for mycobactin-mediated iron acquisition. *Proc Natl Acad Sci U S A* **106**, 18792-18797
85. Gray, T. A., Clark, R. R., Boucher, N., Lapierre, P., Smith, C., and Derbyshire, K. M. (2016) Intercellular communication and conjugation are mediated by ESX secretion systems in mycobacteria. *Science* **354**, 347-350
86. Gey van Pittius, N. C., Gamielien, J., Hide, W., Brown, G. D., Siezen, R. J., and Beyers, A. D. (2001) The ESAT-6 gene cluster of *Mycobacterium tuberculosis* and other high G+C Gram-positive bacteria. *Genome Biol* **2**, RESEARCH0044
87. Kuhlbrandt, W. (2014) Biochemistry. The resolution revolution. *Science* **343**, 1443-1444
88. Glaeser, R. M. (2016) How good can cryo-EM become? *Nature Methods* **13**, 28-32
89. Baker, M. (2018) Cryo-electron microscopy shapes up. *Nature* **561**, 565-567
90. Passmore, L. A., and Russo, C. J. (2016) Specimen Preparation for High-Resolution Cryo-EM. *Methods Enzymol* **579**, 51-86
91. Frederik, P. M., Stuart, M. C., and Verkleij, A. J. (1989) Intermediary structures during membrane fusion as observed by cryo-electron microscopy. *Biochim Biophys Acta* **979**, 275-278
92. Bellare, J. R., Davis, H. T., Scriven, L. E., and Talmon, Y. (1988) Controlled environment vitrification system: an improved sample preparation technique. *J Electron Microscop Tech* **10**, 87-111
93. Dubochet, J., Lepault, J., Freeman, R., Berriman, J. A., and Homo, J.-C. (1982) Electron microscopy of frozen water and aqueous solutions. *Journal of Microscopy* **128**, 219-237
94. Dubochet, J., Adrian, M., Chang, J. J., Homo, J. C., Lepault, J., McDowell, A. W., and Schultz, P. (1988) Cryo-electron microscopy of vitrified specimens. *Q Rev Biophys* **21**, 129-228
95. Grassucci, R. A., Taylor, D. J., and Frank, J. (2007) Preparation of macromolecular complexes for cryo-electron microscopy. *Nat Protoc* **2**, 3239-3246
96. D'Imprima, E., Floris, D., Joppe, M., Sanchez, R., Grininger, M., and Kuhlbrandt, W. (2019) Protein denaturation at the air-water interface and how to prevent it. *Elife* **8**
97. Frederik, P. M., and Hubert, D. H. (2005) Cryoelectron microscopy of liposomes. *Methods Enzymol* **391**, 431-448
98. Ermantraut, E., Wohlfart, K., and Tichelaar, W. (1998) Perforated support foils with pre-defined hole size, shape and arrangement. *Ultramicroscopy* **74**, 75-81
99. Quispe, J., Damiano, J., Mick, S. E., Nackashi, D. P., Fellmann, D., Ajero, T. G., Carragher, B., and Potter, C. S. (2007) An improved holey carbon film for cryo-electron microscopy. *Microsc Microanal* **13**, 365-371

100. Noble, A. J., Dandey, V. P., Wei, H., Brasch, J., Chase, J., Acharya, P., Tan, Y. Z., Zhang, Z., Kim, L. Y., Scapin, G., Rapp, M., Eng, E. T., Rice, W. J., Cheng, A., Negro, C. J., Shapiro, L., Kwong, P. D., Jeruzalmi, D., des Georges, A., Potter, C. S., and Carragher, B. (2018) Routine single particle CryoEM sample and grid characterization by tomography. *Elife* **7**
101. Kasas, S., Dumas, G., Dietler, G., Catsicas, S., and Adrian, M. (2003) Vitrification of cryoelectron microscopy specimens revealed by high-speed photographic imaging. *J Microsc* **211**, 48-53
102. Afanasyev, P., Seer-Linnemayr, C., Ravelli, R. B. G., Matadeen, R., De Carlo, S., Alewijnse, B., Portugal, R. V., Pannu, N. S., Schatz, M., and van Heel, M. (2017) Single-particle cryo-EM using alignment by classification (ABC): the structure of *Lumbricus terrestris* haemoglobin. *IUCrJ* **4**, 678-694
103. Bartesaghi, A., Merk, A., Banerjee, S., Matthies, D., Wu, X., Milne, J. L., and Subramaniam, S. (2015) 2.2 Å resolution cryo-EM structure of beta-galactosidase in complex with a cell-permeant inhibitor. *Science* **348**, 1147-1151
104. Scheres, S. H. (2012) A Bayesian view on cryo-EM structure determination. *J Mol Biol* **415**, 406-418
105. Zivanov, J., Nakane, T., and Scheres, S. H. W. (2019) A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* **6**, 5-17
106. Zhang, K. (2016) Gctf: Real-time CTF determination and correction. *J Struct Biol* **193**, 1-12
107. Van Heel, M. (1987) Similarity measures between images. *Ultramicroscopy* **21**, 95-100
108. Scheres, S. H., and Chen, S. (2012) Prevention of overfitting in cryo-EM structure determination. *Nat Methods* **9**, 853-854
109. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501
110. Afonine, P. V., Klaholz, B. P., Moriarty, N. W., Poon, B. K., Sobolev, O. V., Terwilliger, T. C., Adams, P. D., and Urzhumtsev, A. (2018) New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallogr D Struct Biol* **74**, 814-840
111. Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., Verma, V., Keedy, D. A., Hintze, B. J., Chen, V. B., Jain, S., Lewis, S. M., Arendall, W. B., 3rd, Snoeyink, J., Adams, P. D., Lovell, S. C., Richardson, J. S., and Richardson, D. C. (2018) MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci* **27**, 293-315
112. Zhang, J., Ji, G., Huang, X., Xu, W., and Sun, F. (2016) An improved cryo-FIB method for fabrication of frozen hydrated lamella. *J Struct Biol* **194**, 218-223
113. Medeiros, J. M., Bock, D., Weiss, G. L., Kooger, R., Wepf, R. A., and Pilhofer, M. (2018) Robust workflow and instrumentation for cryo-focused ion beam milling of samples for electron cryotomography. *Ultramicroscopy* **190**, 1-11
114. Houmard, M., Riassetto, D., Roussel, F., Bourgeois, A., Berthome, G., Joud, J. C., and Langlet, M. (2007) Morphology and natural wettability properties of sol-gel derived TiO₂-SiO₂ composite thin films. *Appl Surf Sci* **254**, 1405-1414

115. Ravelli, R. B. G., Nijpels, F. J. T., Henderikx, R. J. M., Weissenberger, G., Thewessem, S., Gijsbers, A., Beulen, B., Lopez-Iglesias, C., and Peters, P. J. (2020) Cryo-EM structures from sub-nl volumes using pin-printing and jet vitrification. *Nat Commun* **11**, 2563
116. Dandey, V. P., Wei, H., Zhang, Z., Tan, Y. Z., Acharya, P., Eng, E. T., Rice, W. J., Kahn, P. A., Potter, C. S., and Carragher, B. (2018) Spotiton: New features and applications. *J Struct Biol* **202**, 161-169
117. Jain, T., Sheehan, P., Crum, J., Carragher, B., and Potter, C. S. (2012) Spotiton: a prototype for an integrated inkjet dispense and vitrification system for cryo-TEM. *J Struct Biol* **179**, 68-75
118. Ashtiani, D., Venugopal, H., Belousoff, M., Spicer, B., Mak, J., Neild, A., and de Marco, A. (2018) Delivery of femtolitre droplets using surface acoustic wave based atomisation for cryo-EM grid preparation. *J Struct Biol* **203**, 94-101
119. Feng, X., Fu, Z., Kaledhonkar, S., Jia, Y., Shah, B., Jin, A., Liu, Z., Sun, M., Chen, B., Grassucci, R. A., Ren, Y., Jiang, H., Frank, J., and Lin, Q. (2017) A Fast and Effective Microfluidic Spraying-Plunging Method for High-Resolution Single-Particle Cryo-EM. *Structure* **25**, 663-670 e663
120. Kontziampasis, D., Klebl, D. P., Iadanza, M. G., Scarff, C. A., Kopf, F., Sobott, F., Monteiro, D. C. F., Trebbin, M., Muench, S. P., and White, H. D. (2019) A cryo-EM grid preparation device for time-resolved structural studies. *IUCr* **6**, 1024-1031
121. Wei, H., Dandey, V. P., Zhang, Z., Raczkowski, A., Rice, W. J., Carragher, B., and Potter, C. S. (2018) Optimizing “self-wicking” nanowire grids. *J Struct Biol* **202**, 170-174
122. Arnold, S. A., Albiez, S., Bieri, A., Syntychaki, A., Adaixo, R., McLeod, R. A., Goldie, K. N., Stahlberg, H., and Braun, T. (2017) Blotting-free and lossless cryo-electron microscopy grid preparation from nanoliter-sized protein samples and single-cell extracts. *J Struct Biol* **197**, 220-226
123. De Gennes, P. G., Brochard-Wyart, F., and Quere, D. (2004) Capillarity and Wetting Phenomena: Drops, Bubbles, Pearls, Waves. *Springer-Verlag*
124. Burstein, N. L., and Maurice, D. M. (1978) Cryofixation of tissue surfaces by a propane jet for electron microscopy. *Micron (1969)* **9**, 191-198
125. Darrow, M. C., Moore, J. P., Walker, R. J., Doering, K., Thaw, P., and King, R. S. (2019) chameleon: Next Generation Sample Preparation for CryoEM based on Spotiton. *Acta Crystallogr A* **75**, A424-A424
126. Taylor, K. A., and Glaeser, R. M. (2008) Retrospective on the early development of cryoelectron microscopy of macromolecules and a prospective on opportunities for the future. *J Struct Biol* **163**, 214-223
127. Naydenova, K., and Russo, C. J. (2017) Measuring the effects of particle orientation to improve the efficiency of electron cryomicroscopy. *Nat Commun* **8**, 629
128. Sun, F. (2018) Orienting the future of bio-macromolecular electron microscopy. *Chinese Phys B* **27**
129. Glaeser, R. M., and Han, B. G. (2017) Opinion: hazards faced by macromolecules when confined to thin aqueous films. *Biophys Rep* **3**, 1-7

130. Bouwhuis, W., van der Veen, R. C., Tran, T., Keij, D. L., Winkels, K. G., Peters, I. R., van der Meer, D., Sun, C., Snoeijer, J. H., and Lohse, D. (2012) Maximal air bubble entrainment at liquid-drop impact. *Phys Rev Lett* **109**, 264501
131. Berriman, J., and Unwin, N. (1994) Analysis of transient structures by cryo-microscopy combined with rapid mixing of spray droplets. *Ultramicroscopy* **56**, 241-252
132. Chen, B., Kaledhonkar, S., Sun, M., Shen, B., Lu, Z., Barnard, D., Lu, T. M., Gonzalez, R. L., Jr., and Frank, J. (2015) Structural dynamics of ribosome subunit association studied by mixing-spraying time-resolved cryogenic electron microscopy. *Structure* **23**, 1097-1105
133. Mahamid, J., Pfeffer, S., Schaffer, M., Villa, E., Danev, R., Cuellar, L. K., Forster, F., Hyman, A. A., Plitzko, J. M., and Baumeister, W. (2016) Visualizing the molecular sociology at the HeLa cell nuclear periphery. *Science* **351**, 969-972
134. Rice, W. J., Cheng, A., Noble, A. J., Eng, E. T., Kim, L. Y., Carragher, B., and Potter, C. S. (2018) Routine determination of ice thickness for cryo-EM grids. *J Struct Biol* **204**, 38-44
135. Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J., Lindahl, E., and Scheres, S. H. (2018) New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**
136. Bartesaghi, A., Aguerrebere, C., Falconieri, V., Banerjee, S., Earl, L. A., Zhu, X., Grigorieff, N., Milne, J. L. S., Sapiro, G., Wu, X., and Subramaniam, S. (2018) Atomic Resolution Cryo-EM Structure of beta-Galactosidase. *Structure* **26**, 848-856 e843
137. Hamaguchi, T., Maki-Yonekura, S., Naitow, H., Matsuura, Y., Ishikawa, T., and Yonekura, K. (2019) A new cryo-EM system for single particle analysis. *J Struct Biol* **207**, 40-48
138. Tan, Y. Z., Aiyer, S., Mietzsch, M., Hull, J. A., McKenna, R., Grieger, J., Samulski, R. J., Baker, T. S., Agbandje-McKenna, M., and Lyumkis, D. (2018) Sub-2 Å Ewald curvature corrected structure of an AAV2 capsid variant. *Nat Commun* **9**, 3628
139. Carragher, B., Cheng, Y., Frost, A., Glaeser, R. M., Lander, G. C., Nogales, E., and Wang, H. W. (2019) Current outcomes when optimizing 'standard' sample preparation for single-particle cryo-EM. *J Microsc* **276**, 39-45
140. Andrews, S. C. (2010) The Ferritin-like superfamily: Evolution of the biological iron storeman from a rubrerythrin-like ancestor. *Biochim Biophys Acta* **1800**, 691-705
141. Arosio, P., Ingrassia, R., and Cavadini, P. (2009) Ferritins: a family of molecules for iron storage, antioxidation and more. *Biochim Biophys Acta* **1790**, 589-599
142. Expert, D. (2012) Iron, an Element Essential to Life. in *Molecular Aspects of Iron Metabolism in Pathogenic and Symbiotic Plant-Microbe Associations* (Expert, D., and O'Brian, M. R. eds.), Springer. pp
143. Ratledge, C., and Dover, L. G. (2000) Iron metabolism in pathogenic bacteria. *Annu Rev Microbiol* **54**, 881-941
144. Contreras, H., Joens, M. S., McMath, L. M., Le, V. P., Tullius, M. V., Kimmey, J. M., Bionghi, N., Horwitz, M. A., Fitzpatrick, J. A., and Goulding, C. W. (2014) Characterization of a Mycobacterium tuberculosis nanocompartment and its potential cargo proteins. *J Biol Chem* **289**, 18279-18289
145. Khare, G., Gupta, V., Nangpal, P., Gupta, R. K., Sauter, N. K., and Tyagi, A. K. (2011) Ferritin structure from Mycobacterium tuberculosis: comparative study with homologues identifies extended C-terminus involved in ferroxidase activity. *PLoS One* **6**, e18570

146. Parida, A., Mohanty, A., Kansara, B. T., and Behera, R. K. (2020) Impact of Phosphate on Iron Mineralization and Mobilization in Nonheme Bacterioferritin B from *Mycobacterium tuberculosis*. *Inorg Chem* **59**, 629-641
147. Zheng, S. Q., Palovcak, E., Armache, J. P., Verba, K. A., Cheng, Y., and Agard, D. A. (2017) MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331-332
148. Rosenthal, P. B., and Henderson, R. (2003) Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J Mol Biol* **333**, 721-745
149. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132
150. Kaur, S., Gomez-Blanco, J., Khalifa, A. A. Z., Adinarayanan, S., Sanchez-Garcia, R., Wrapp, D., McLellan, J. S., Bui, K. H., and Vargas, J. (2021) Local computational methods to improve the interpretability and analysis of cryo-EM maps. *Nat Commun* **12**, 1240
151. Rui, H., Rivera, M., and Im, W. (2012) Protein dynamics and ion traffic in bacterioferritin. *Biochemistry* **51**, 9900-9910
152. Jobichen, C., Chong, T. Y., Rattinam, R., Basak, S., Srinivasan, M., Pandey, K. P., Ngoc, T. B., Shi, J., Angayarkanni, J., and Sivaraman, J. (2021) Cryo-EM Structure of Bacterioferritin Nanocages Provides Insight into the Bio-mineralization of Ferritins. *bioRxiv*, 2021.2002.2004.429857
153. Putignano, V., Rosato, A., Banci, L., and Andreini, C. (2018) MetalPDB in 2018: a database of metal sites in biological macromolecular structures. *Nucleic Acids Res* **46**, D459-D464
154. Teilum, K., Olsen, J. G., and Kragelund, B. B. (2009) Functional aspects of protein flexibility. *Cell Mol Life Sci* **66**, 2231-2247
155. Wu, M., Lander, G. C., and Herzik, M. A., Jr. (2020) Sub-2 Angstrom resolution structure determination using single-particle cryo-EM at 200keV. *J Struct Biol X* **4**, 100020
156. Kaur, S., Gomez-Blanco, J., Khalifa, A. A. Z., Adinarayanan, S., Sanchez-Garcia, R., Wrapp, D., McLellan, J. S., Bui, K. H., and Vargas, J. (2021) Local computational methods to improve the interpretability and analysis of cryo-EM maps. *Nature Communications* **12**, 1240
157. World Health Organization. (2020) Global tuberculosis report 2020. Geneva: World Health Organization
158. Ferluga, J., Yasmin, H., Al-Ahdal, M. N., Bhakta, S., and Kishore, U. (2020) Natural and trained innate immunity against *Mycobacterium tuberculosis*. *Immunobiology* **225**, 151951
159. Brennan, P. J., and Nikaido, H. (1995) The envelope of mycobacteria. *Annu Rev Biochem* **64**, 29-63
160. Pym, A. S., Brodin, P., Majlessi, L., Brosch, R., Demangel, C., Williams, A., Griffiths, K. E., Marchal, G., Leclerc, C., and Cole, S. T. (2003) Recombinant BCG exporting ESAT-6 confers enhanced protection against tuberculosis. *Nat Med* **9**, 533-539

161. Beckham, K. S., Ciccarelli, L., Bunduc, C. M., Mertens, H. D., Ummels, R., Lugmayr, W., Mayr, J., Rettel, M., Savitski, M. M., Svergun, D. I., Bitter, W., Wilmanns, M., Marlovits, T. C., Parret, A. H., and Houben, E. N. (2017) Structure of the mycobacterial ESX-5 type VII secretion system membrane complex by single-particle analysis. *Nat Microbiol* **2**, 17047
162. Cascioferro, A., Delogu, G., Colone, M., Sali, M., Stringaro, A., Arancia, G., Fadda, G., Palu, G., and Manganeli, R. (2007) PE is a functional domain responsible for protein translocation and localization on mycobacterial cell wall. *Mol Microbiol* **66**, 1536-1547
163. Burggraaf, M. J., Ates, L. S., Speer, A., van der Kuij, K., Kuijl, C., and Bitter, W. (2019) Optimization of secretion and surface localization of heterologous OVA protein in mycobacteria by using LipY as a carrier. *Microb Cell Fact* **18**, 44
164. Wang, Q., Boshoff, H. I. M., Harrison, J. R., Ray, P. C., Green, S. R., Wyatt, P. G., and Barry, C. E., 3rd. (2020) PE/PPE proteins mediate nutrient transport across the outer membrane of *Mycobacterium tuberculosis*. *Science* **367**, 1147-1151
165. Phan, T. H., Ummels, R., Bitter, W., and Houben, E. N. (2017) Identification of a substrate domain that determines system specificity in mycobacterial type VII secretion systems. *Sci Rep* **7**, 42704
166. Famelis, N., Rivera-Calzada, A., Degliesposti, G., Wingender, M., Mietrach, N., Skehel, J. M., Fernandez-Leiro, R., Bottcher, B., Schlosser, A., Llorca, O., and Geibel, S. (2019) Architecture of the mycobacterial type VII secretion system. *Nature*
167. Poweleit, N., Czudnochowski, N., Nakagawa, R., Trinidad, D. D., Murphy, K. C., Sasseti, C. M., and Rosenberg, O. S. (2019) The structure of the endogenous ESX-3 secretion system. *Elife* **8**
168. Rosenberg, O. S., Dovala, D., Li, X., Connolly, L., Bendebury, A., Finer-Moore, J., Holton, J., Cheng, Y., Stroud, R. M., and Cox, J. S. (2015) Substrates Control Multimerization and Activation of the Multi-Domain ATPase Motor of Type VII Secretion. *Cell* **161**, 501-512
169. Houben, E. N., Bestebroer, J., Ummels, R., Wilson, L., Piersma, S. R., Jimenez, C. R., Ottenhoff, T. H., Luirink, J., and Bitter, W. (2012) Composition of the type VII secretion system membrane complex. *Mol Microbiol* **86**, 472-484
170. Bunduc, C. M., Fahrenkamp, D., Wald, J., Ummels, R., Bitter, W., Houben, E. N. G., and Marlovits, T. C. (2021) Structure and dynamics of a mycobacterial type VII secretion system. *Nature* **593**, 445-448
171. Beckham, K. S. H., Ritter, C., Chojnowski, G., Ziemianowicz, D. S., Mullapudi, E., Rettel, M., Savitski, M. M., Mortensen, S. A., Kosinski, J., and Wilmanns, M. (2021) Structure of the mycobacterial ESX-5 type VII secretion system pore complex. *Sci Adv* **7**
172. van Winden, V. J., Ummels, R., Piersma, S. R., Jimenez, C. R., Korotkov, K. V., Bitter, W., and Houben, E. N. (2016) Mycosins Are Required for the Stabilization of the ESX-1 and ESX-5 Type VII Secretion Membrane Complexes. *mBio* **7**
173. Lou, Y., Rybniker, J., Sala, C., and Cole, S. T. (2017) EspC forms a filamentous structure in the cell envelope of *Mycobacterium tuberculosis* and impacts ESX-1 secretion. *Mol Microbiol* **103**, 26-38

174. Sani, M., Houben, E. N., Geurtsen, J., Pierson, J., de Punder, K., van Zon, M., Wever, B., Piersma, S. R., Jimenez, C. R., Daffe, M., Appelmelk, B. J., Bitter, W., van der Wel, N., and Peters, P. J. (2010) Direct visualization by cryo-EM of the mycobacterial capsular layer: a labile structure containing ESX-1-secreted proteins. *PLoS Pathog* **6**, e1000794
175. Bitter, W., Houben, E. N., Bottai, D., Brodin, P., Brown, E. J., Cox, J. S., Derbyshire, K., Fortune, S. M., Gao, L. Y., Liu, J., Gey van Pittius, N. C., Pym, A. S., Rubin, E. J., Sherman, D. R., Cole, S. T., and Brosch, R. (2009) Systematic genetic nomenclature for type VII secretion systems. *PLoS Pathog* **5**, e1000507
176. Conrad, W. H., Osman, M. M., Shanahan, J. K., Chu, F., Takaki, K. K., Cameron, J., Hopkinson-Woolley, D., Brosch, R., and Ramakrishnan, L. (2017) Mycobacterial ESX-1 secretion system mediates host cell lysis through bacterium contact-dependent gross membrane disruptions. *Proc Natl Acad Sci U S A* **114**, 1371-1376
177. Carlsson, F., Joshi, S. A., Rangell, L., and Brown, E. J. (2009) Polar localization of virulence-related Esx-1 secretion in mycobacteria. *PLoS Pathog* **5**, e1000285
178. Phan, T. H., van Leeuwen, L. M., Kuijl, C., Ummels, R., van Stempvoort, G., Rubio-Canalejas, A., Piersma, S. R., Jimenez, C. R., van der Sar, A. M., Houben, E. N. G., and Bitter, W. (2018) EspH is a hypervirulence factor for *Mycobacterium marinum* and essential for the secretion of the ESX-1 substrates EspE and EspF. *PLoS Pathog* **14**, e1007247
179. Korotkova, N., Piton, J., Wagner, J. M., Boy-Rottger, S., Japaridze, A., Evans, T. J., Cole, S. T., Pojer, F., and Korotkov, K. V. (2015) Structure of EspB, a secreted substrate of the ESX-1 secretion system of *Mycobacterium tuberculosis*. *J Struct Biol* **191**, 236-244
180. Solomonson, M., Setiaputra, D., Makepeace, K. A. T., Lameignere, E., Petrotchenko, E. V., Conrady, D. G., Bergeron, J. R., Vuckovic, M., DiMaio, F., Borchers, C. H., Yip, C. K., and Strynadka, N. C. J. (2015) Structure of EspB from the ESX-1 type VII secretion system and insights into its export mechanism. *Structure* **23**, 571-583
181. Piton, J., Pojer, F., Wakatsuki, S., Gati, C., and Cole, S. T. (2020) High resolution CryoEM structure of the ring-shaped virulence factor EspB from *Mycobacterium tuberculosis*. *J Struct Biol: X* **4**
182. Xu, J., Laine, O., Masciocchi, M., Manoranjan, J., Smith, J., Du, S. J., Edwards, N., Zhu, X., Fenselau, C., and Gao, L. Y. (2007) A unique *Mycobacterium* ESX-1 protein co-secreted with CFP-10/ESAT-6 and is necessary for inhibiting phagosome maturation. *Mol Microbiol* **66**, 787-800
183. Ohol, Y. M., Goetz, D. H., Chan, K., Shiloh, M. U., Craik, C. S., and Cox, J. S. (2010) *Mycobacterium tuberculosis* MycP1 protease plays a dual role in regulation of ESX-1 secretion and virulence. *Cell Host Microbe* **7**, 210-220
184. Solomonson, M., Huesgen, P. F., Wasney, G. A., Watanabe, N., Gruninger, R. J., Prehna, G., Overall, C. M., and Strynadka, N. C. (2013) Structure of the mycosin-1 protease from the mycobacterial ESX-1 protein type VII secretion system. *J Biol Chem* **288**, 17782-17790
185. Chen, J. M., Zhang, M., Rybniker, J., Boy-Rottger, S., Dhar, N., Pojer, F., and Cole, S. T. (2013) *Mycobacterium tuberculosis* EspB binds phospholipids and mediates EsxA-independent virulence. *Mol Microbiol* **89**, 1154-1166

186. Kapust, R. B., Tozser, J., Fox, J. D., Anderson, D. E., Cherry, S., Copeland, T. D., and Waugh, D. S. (2001) Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng* **14**, 993-1000
187. Mathew, A., Buijs, R., Eijkel, G. B., Giskes, F., Dyachenko, A., van der Horst, J., Byelov, D., Spaanderman, D. J., Heck, A. J. R., Porta Siegel, T., Ellis, S. R., and Heeren, R. M. A. (2021) Ion Imaging of Native Protein Complexes Using Orthogonal Time-of-Flight Mass Spectrometry and a Timepix Detector. *J Am Soc Mass Spectrom* **32**, 569-580
188. Marty, M. T., Baldwin, A. J., Marklund, E. G., Hochberg, G. K., Benesch, J. L., and Robinson, C. V. (2015) Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. *Anal Chem* **87**, 4370-4376
189. Tan, Y. Z., Baldwin, P. R., Davis, J. H., Williamson, J. R., Potter, C. S., Carragher, B., and Lyumkis, D. (2017) Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat Methods* **14**, 793-796
190. Mastronarde, D. N. (2005) Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol* **152**, 36-51
191. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallographica Section D* **66**, 486-501
192. Afonine, P. V., Klaholz, B. P., Moriarty, N. W., Poon, B. K., Sobolev, O. V., Terwilliger, T. C., Adams, P. D., and Urzhumtsev, A. (2018) New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallographica Section D* **74**, 814-840
193. Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., Verma, V., Keedy, D. A., Hintze, B. J., Chen, V. B., Jain, S., Lewis, S. M., Arendall III, W. B., Snoeyink, J., Adams, P. D., Lovell, S. C., Richardson, J. S., and Richardson, D. C. (2018) MolProbity: More and better reference data for improved all-atom structure validation. *Protein Science* **27**, 293-315
194. Mastronarde, D. N., and Held, S. R. (2017) Automated tilt series alignment and tomographic reconstruction in IMOD. *J Struct Biol* **197**, 102-113
195. Henderson, R. (1995) The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Q Rev Biophys* **28**, 171-193
196. Ekiert, D. C., and Cox, J. S. (2014) Structure of a PE-PPE-EspG complex from Mycobacterium tuberculosis reveals molecular specificity of ESX protein secretion. *Proc Natl Acad Sci U S A* **111**, 14758-14763
197. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191
198. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948

199. Joosten, R. P., Salzemann, J., Bloch, V., Stockinger, H., Berglund, A. C., Blanchet, C., Bongcam-Rudloff, E., Combet, C., Da Costa, A. L., Deleage, G., Diarena, M., Fabbretti, R., Fettahi, G., Flegel, V., Gisel, A., Kasam, V., Kervinen, T., Korpelainen, E., Mattila, K., Pagni, M., Reichstadt, M., Breton, V., Tickle, I. J., and Vriend, G. (2009) PDB_REDO: automated re-refinement of X-ray structure models in the PDB. *J Appl Crystallogr* **42**, 376-384
200. Xie, N. Z., Du, Q. S., Li, J. X., and Huang, R. B. (2015) Exploring Strong Interactions in Proteins with Quantum Chemistry and Examples of Their Applications in Drug Design. *PLoS One* **10**, e0137113
201. Goddard, T. D., Huang, C. C., Meng, E. C., Pettersen, E. F., Couch, G. S., Morris, J. H., and Ferrin, T. E. (2018) UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* **27**, 14-25
202. Smart, O. S., Goodfellow, J. M., and Wallace, B. A. (1993) The pore dimensions of gramicidin A. *Biophys J* **65**, 2455-2460
203. Olsson, M. H., Sondergaard, C. R., Rostkowski, M., and Jensen, J. H. (2011) PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J Chem Theory Comput* **7**, 525-537
204. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **98**, 10037-10041
205. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* **32**, W665-667
206. Uversky, V. N. (2013) The alphabet of intrinsic disorder: II. Various roles of glutamic acid in ordered and intrinsically disordered proteins. *Intrinsically Disord Proteins* **1**, e24684
207. Theillet, F. X., Kalmar, L., Tompa, P., Han, K. H., Selenko, P., Dunker, A. K., Daughdrill, G. W., and Uversky, V. N. (2013) The alphabet of intrinsic disorder: I. Act like a Pro: On the abundance and roles of proline residues in intrinsically disordered proteins. *Intrinsically Disord Proteins* **1**, e24360
208. Rucker, A. L., and Creamer, T. P. (2002) Polyproline II helical structure in protein unfolded states: lysine peptides revisited. *Protein Sci* **11**, 980-985
209. Luo, P., and Baldwin, R. L. (1997) Mechanism of helix induction by trifluoroethanol: a framework for extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water. *Biochemistry* **36**, 8413-8421
210. McLaughlin, B., Chon, J. S., MacGurn, J. A., Carlsson, F., Cheng, T. L., Cox, J. S., and Brown, E. J. (2007) A mycobacterium ESX-1-secreted virulence factor with unique requirements for export. *PLoS Pathog* **3**, e105
211. Flint, J. L., Kowalski, J. C., Karnati, P. K., and Derbyshire, K. M. (2004) The RD1 virulence locus of *Mycobacterium tuberculosis* regulates DNA transfer in *Mycobacterium smegmatis*. *Proc Natl Acad Sci U S A* **101**, 12598-12603

212. De Leon, J., Jiang, G., Ma, Y., Rubin, E., Fortune, S., and Sun, J. (2012) Mycobacterium tuberculosis ESAT-6 exhibits a unique membrane-interacting activity that is not found in its ortholog from non-pathogenic Mycobacterium smegmatis. *J Biol Chem* **287**, 44184-44191
213. Ma, Y., Keil, V., and Sun, J. (2015) Characterization of Mycobacterium tuberculosis EsxA membrane insertion: roles of N- and C-terminal flexible arms and central helix-turn-helix motif. *J Biol Chem* **290**, 7314-7322
214. Franz, J., Lelle, M., Peneva, K., Bonn, M., and Weidner, T. (2016) SAP(E) - A cell-penetrating polyproline helix at lipid interfaces. *Biochim Biophys Acta* **1858**, 2028-2034
215. Williamson, Z. A., Chaton, C. T., Ciocca, W. A., Korotkova, N., and Korotkov, K. V. (2020) PE5-PPE4-EspG3 heterotrimer structure from mycobacterial ESX-3 secretion system gives insight into cognate substrate recognition by ESX systems. *J Biol Chem*
216. Korotkova, N., Freire, D., Phan, T. H., Ummels, R., Creekmore, C. C., Evans, T. J., Wilmanns, M., Bitter, W., Parret, A. H., Houben, E. N., and Korotkov, K. V. (2014) Structure of the Mycobacterium tuberculosis type VII secretion system chaperone EspG5 in complex with PE25-PPE41 dimer. *Mol Microbiol* **94**, 367-382
217. Beckham, K. S., Ritter, C., Chojnowski, G., Mullapudi, E., Rettel, M., Savitski, M. M., Mortensen, S. A., Kosinski, J., and Wilmanns, M. (2020) Structure of the mycobacterial ESX-5 Type VII Secretion System hexameric pore complex. *bioRxiv*
218. Dulberger, C. L., Rubin, E. J., and Boutte, C. C. (2020) The mycobacterial cell envelope - a moving target. *Nat Rev Microbiol* **18**, 47-59
219. Zuber, B., Chami, M., Houssin, C., Dubochet, J., Griffiths, G., and Daffe, M. (2008) Direct visualization of the outer membrane of mycobacteria and corynebacteria in their native state. *J Bacteriol* **190**, 5672-5680
220. Zinke, M., Sachowsky, K. A. A., Oster, C., Zinn-Justin, S., Ravelli, R., Schroder, G. F., Habeck, M., and Lange, A. (2020) Architecture of the flexible tail tube of bacteriophage SPP1. *Nat Commun* **11**, 5759
221. Lodes, M. J., Dillon, D. C., Mohamath, R., Day, C. H., Benson, D. R., Reynolds, L. D., McNeill, P., Sampaio, D. P., Skeiky, Y. A., Badaro, R., Persing, D. H., Reed, S. G., and Houghton, R. L. (2001) Serological expression cloning and immunological evaluation of MTB48, a novel Mycobacterium tuberculosis antigen. *J Clin Microbiol* **39**, 2485-2493
222. Liu, X., and Lieberman, J. (2020) Knocking 'em Dead: Pore-Forming Proteins in Immune Defense. *Annu Rev Immunol* **38**, 455-485
223. Ruan, J., Xia, S., Liu, X., Lieberman, J., and Wu, H. (2018) Cryo-EM structure of the gasdermin A3 membrane pore. *Nature* **557**, 62-67
224. Berger, C., Ravelli, R. B. G., Lopez-Iglesias, C., Kudryashev, M., Diepold, A., and Peters, P. J. (2021) Structure of the Yersinia injectisome in intracellular host cell phagosomes revealed by cryo FIB electron tomography. *J Struct Biol* **213**, 107701
225. Dong, E., Du, H., and Gardner, L. (2020) An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* **20**, 533-534
226. Hirayama, D., Iida, T., and Nakase, H. (2017) The Phagocytic Function of Macrophage-Enforcing Innate Immunity and Tissue Homeostasis. *Int J Mol Sci* **19**

227. Sassetti, C. M., and Rubin, E. J. (2003) Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci U S A* **100**, 12989-12994
228. Gao, L. Y., Guo, S., McLaughlin, B., Morisaki, H., Engel, J. N., and Brown, E. J. (2004) A mycobacterial virulence gene cluster extending RD1 is required for cytolysis, bacterial spreading and ESAT-6 secretion. *Mol Microbiol* **53**, 1677-1693
229. Bottai, D., Majlessi, L., Simeone, R., Frigui, W., Laurent, C., Lenormand, P., Chen, J., Rosenkrands, I., Huerre, M., Leclerc, C., Cole, S. T., and Brosch, R. (2011) ESAT-6 secretion-independent impact of ESX-1 genes *espF* and *espG1* on virulence of *Mycobacterium tuberculosis*. *J Infect Dis* **203**, 1155-1164
230. Singh, P. K., Saxena, R., Tiwari, S., Singh, D. K., Singh, S. K., Kumari, R., and Srivastava, K. K. (2015) RD-1 encoded EspJ protein gets phosphorylated prior to affect the growth and intracellular survival of mycobacteria. *Sci Rep* **5**, 12717
231. Behr, M. A., Wilson, M. A., Gill, W. P., Salamon, H., Schoolnik, G. K., Rane, S., and Small, P. M. (1999) Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* **284**, 1520-1523
232. Champion, M. M., Williams, E. A., Pinapati, R. S., and Champion, P. A. (2014) Correlation of phenotypic profiles using targeted proteomics identifies mycobacterial *esx-1* substrates. *J Proteome Res* **13**, 5151-5164
233. Drozdetskiy, A., Cole, C., Procter, J., and Barton, G. J. (2015) JPred4: a protein secondary structure prediction server. *Nucleic Acids Res* **43**, W389-394
234. Scheich, C., Kummel, D., Soumailakakis, D., Heinemann, U., and Bussow, K. (2007) Vectors for co-expression of an unrestricted number of proteins. *Nucleic Acids Res* **35**, e43
235. Green, M. R., and Sambrook, J. (2019) Inverse Polymerase Chain Reaction (PCR). *Cold Spring Harb Protoc* **2019**
236. Franke, D., Kikhney, A. G., and Svergun, D. I. (2012) Automated acquisition and analysis of small angle X-ray scattering data. *Nucl Instrum Meth A* **689**, 52-59
237. Panjkovich, A., and Svergun, D. I. (2018) CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. *Bioinformatics* **34**, 1944-1946
238. Svergun, D. I., Petoukhov, M. V., and Koch, M. H. (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* **80**, 2946-2953
239. Franke, D., and Svergun, D. I. (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J Appl Crystallogr* **42**, 342-346
240. Volkov, V. V., and Svergun, D. I. (2003) Uniqueness of ab initio shape determination in small-angle scattering. *Journal of Applied Crystallography* **36**, 860-864
241. Zhang, Y. (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40
242. Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D., Konarev, P. V., and Svergun, D. I. (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *J Appl Crystallogr* **45**, 342-350
243. Williamson, M. P. (1994) The structure and function of proline-rich regions in proteins. *Biochem J* **297 (Pt 2)**, 249-260

244. Poulsen, C., Panjikar, S., Holton, S. J., Wilmanns, M., and Song, Y. H. (2014) WXG100 protein superfamily consists of three subfamilies and exhibits an alpha-helical C-terminal conserved residue pattern. *PLoS One* **9**, e89313
245. Breibeck, J., and Skerra, A. (2018) The polypeptide biophysics of proline/alanine-rich sequences (PAS): Recombinant biopolymers with PEG-like properties. *Biopolymers* **109**
246. Guinier, A. (1939) La diffraction des rayons X aux très petits angles : application à l'étude de phénomènes ultramicroscopiques. *Ann. Phys.* **11**, 161-237
247. Rambo, R. P., and Tainer, J. A. (2011) Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers* **95**, 559-571
248. Petoukhov, M. V., and Svergun, D. I. (2015) Ambiguity assessment of small-angle scattering curves from monodisperse systems. *Acta Crystallogr D Biol Crystallogr* **71**, 1051-1058
249. Tuukkanen, A. T., Kleywegt, G. J., and Svergun, D. I. (2016) Resolution of ab initio shapes determined from small-angle scattering. *IUCrJ* **3**, 440-447
250. Strong, M., Sawaya, M. R., Wang, S., Phillips, M., Cascio, D., and Eisenberg, D. (2006) Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A* **103**, 8060-8065
251. Kozin, M. B., and Svergun, D. I. (2001) Automated matching of high- and low-resolution structural models. *Journal of Applied Crystallography* **34**, 33-41
252. Liu, Y., Gao, Y., Li, D., Fleming, J., Li, H., and Bi, L. (2016) Crystal structure of Rv3899c184-410, a hypothetical protein from *Mycobacterium tuberculosis*. *Acta Crystallographica Section F* **72**, 642-645
253. Malen, H., Berven, F. S., Fladmark, K. E., and Wiker, H. G. (2007) Comprehensive analysis of exported proteins from *Mycobacterium tuberculosis* H37Rv. *Proteomics* **7**, 1702-1718
254. Kruh, N. A., Troudt, J., Izzo, A., Prenni, J., and Dobos, K. M. (2010) Portrait of a pathogen: the *Mycobacterium tuberculosis* proteome in vivo. *PLoS One* **5**, e13938
255. Newton-Foot, M., Warren, R. M., Sampson, S. L., van Helden, P. D., and Gey van Pittius, N. C. (2016) The plasmid-mediated evolution of the mycobacterial ESX (Type VII) secretion systems. *BMC Evol Biol* **16**, 62
256. Tuukkanen, A. T., Freire, D., Chan, S., Arbing, M. A., Reed, R. W., Evans, T. J., Zenkeviciute, G., Kim, J., Kahng, S., Sawaya, M. R., Chaton, C. T., Wilmanns, M., Eisenberg, D., Parret, A. H. A., and Korotkov, K. V. (2019) Structural Variability of EspG Chaperones from Mycobacterial ESX-1, ESX-3, and ESX-5 Type VII Secretion Systems. *J Mol Biol* **431**, 289-307
257. Kikhney, A. G., Borges, C. R., Molodenskiy, D. S., Jeffries, C. M., and Svergun, D. I. (2020) SASBDB: Towards an automatically curated and validated repository for biological scattering data. *Protein Sci* **29**, 66-75
258. Gao, Y., Liu, M., Chen, Y., Shi, S., Geng, J., and Tian, J. (2021) Association between tuberculosis and COVID-19 severity and mortality: A rapid systematic review and meta-analysis. *J Med Virol* **93**, 194-196
259. Mariandyshev, A., and Eliseev, P. (2017) Drug-resistant tuberculosis threatens WHO's End-TB strategy. *Lancet Infect Dis* **17**, 674-675

260. Groschel, M. I., Sayes, F., Simeone, R., Majlessi, L., and Brosch, R. (2016) ESX secretion systems: mycobacterial evolution to counter host immunity. *Nat Rev Microbiol* **14**, 677-691
261. Simeone, R., Bottai, D., Frigui, W., Majlessi, L., and Brosch, R. (2015) ESX/type VII secretion systems of mycobacteria: Insights into evolution, pathogenicity and protection. *Tuberculosis (Edinb)* **95 Suppl 1**, S150-154
262. Houben, E. N., Korotkov, K. V., and Bitter, W. (2014) Take five - Type VII secretion systems of Mycobacteria. *Biochim Biophys Acta* **1843**, 1707-1716
263. Pallen, M. J. (2002) The ESAT-6/WXG100 superfamily -- and a new Gram-positive secretion system? *Trends Microbiol* **10**, 209-212
264. Gey van Pittius, N. C., Sampson, S. L., Lee, H., Kim, Y., van Helden, P. D., and Warren, R. M. (2006) Evolution and expansion of the Mycobacterium tuberculosis PE and PPE multigene families and their association with the duplication of the ESAT-6 (esx) gene cluster regions. *BMC Evol Biol* **6**, 95
265. Chirakos, A. E., Balaram, A., Conrad, W., and Champion, P. A. (2020) Modeling Tubercular ESX-1 Secretion Using Mycobacterium marinum. *Microbiol Mol Biol Rev* **84**
266. Daleke, M. H., Ummels, R., Bawono, P., Heringa, J., Vandenbroucke-Grauls, C. M., Luijck, J., and Bitter, W. (2012) General secretion signal for the mycobacterial type VII secretion pathway. *Proc Natl Acad Sci U S A* **109**, 11342-11347
267. Gijbsbers, A., Vinciauskaite, V., Siroy, A., Gao, Y., Tria, G., Mathew, A., Sánchez-Puig, N., López-Iglesias, C., Peters, P. J., and Ravelli, R. B. G. (2021) Priming mycobacterial ESX-secreted protein B to form a channel-like structure. *Curr Res Struct Biol* **3**, 153-164
268. Gijbsbers, A., Sanchez-Puig, N., Gao, Y., Peters, P. J., Ravelli, R. B. G., and Siliqi, D. (2021) Structural Analysis of the Partially Disordered Protein EspK from Mycobacterium Tuberculosis. *Crystals* **11**
269. Blanchet, C. E., Spilotros, A., Schwemmer, F., Graewert, M. A., Kikhney, A., Jeffries, C. M., Franke, D., Mark, D., Zengerle, R., Cipriani, F., Fiedler, S., Roessle, M., and Svergun, D. I. (2015) Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). *J Appl Crystallogr* **48**, 431-443
270. Cowieson, N. P., Edwards-Gayle, C. J. C., Inoue, K., Khunti, N. S., Douth, J., Williams, E., Daniels, S., Preece, G., Krumpa, N. A., Sutter, J. P., Tully, M. D., Terrill, N. J., and Rambo, R. P. (2020) Beamline B21: high-throughput small-angle X-ray scattering at Diamond Light Source. *J Synchrotron Radiat* **27**, 1438-1446
271. McCarthy, A. A., Barrett, R., Beteva, A., Caserotto, H., Dobias, F., Felisaz, F., Giraud, T., Guijarro, M., Janocha, R., Khadrouche, A., Lentini, M., Leonard, G. A., Marrero, M. L., Malbet-Monaco, S., McSweeney, S., Nurizzo, D., Papp, G., Rossi, C., Sinoir, J., Sorez, C., Surr, J., Svensson, O., Zander, U., Cipriani, F., Theveneau, P., and Mueller-Dieckmann, C. (2018) ID30B-a versatile beamline for macromolecular crystallography experiments at the ESRF. *J Synchrotron Radiat* **25**, 1249-1260

272. Cianci, M., Bourenkov, G., Pompidor, G., Karpics, I., Kallio, J., Bento, I., Roessle, M., Cipriani, F., Fiedler, S., and Schneider, T. R. (2017) P13, the EMBL macromolecular crystallography beamline at the low-emittance PETRA III ring for high- and low-energy phasing with variable beam focusing. *J Synchrotron Radiat* **24**, 323-332
273. Kabsch, W. (2010) Xds. *Acta Crystallographica Section D-Biological Crystallography* **66**, 125-132
274. Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., Womack, T., and Bricogne, G. (2011) Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr D* **67**, 293-302
275. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674
276. Lieschner, D., Afonine, P. V., Baker, M. L., Bunkoczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L. W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J., and Adams, P. D. (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**, 861-877
277. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., and Hassabis, D. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*
278. Sali, A., and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234**, 779-815
279. Song, Y., DiMaio, F., Wang, R. Y., Kim, D., Miles, C., Brunette, T., Thompson, J., and Baker, D. (2013) High-resolution comparative modeling with RosettaCM. *Structure* **21**, 1735-1742
280. Chen, X., Cheng, H. F., Zhou, J., Chan, C. Y., Lau, K. F., Tsui, S. K., and Au, S. W. (2017) Structural basis of the PE-PPE protein interaction in Mycobacterium tuberculosis. *J Biol Chem* **292**, 16880-16890
281. Williamson, Z. A., Chaton, C. T., Ciocca, W. A., Korotkova, N., and Korotkov, K. V. (2020) PE5-PPE4-EspG3 heterotrimer structure from mycobacterial ESX-3 secretion system gives insight into cognate substrate recognition by ESX systems. *J Biol Chem* **295**, 12706-12715
282. Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E., 3rd, Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin,

- N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M. A., Rajandream, M. A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J. E., Taylor, K., Whitehead, S., and Barrell, B. G. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537-544
283. Abdallah, A. M., Verboom, T., Weerdenburg, E. M., Gey van Pittius, N. C., Mahasha, P. W., Jimenez, C., Parra, M., Cadieux, N., Brennan, M. J., Appelmelk, B. J., and Bitter, W. (2009) PPE and PE_PGRS proteins of *Mycobacterium marinum* are transported via the type VII secretion system ESX-5. *Mol Microbiol* **73**, 329-340
284. Minamino, T. (2014) Protein export through the bacterial flagellar type III export pathway. *Biochim Biophys Acta* **1843**, 1642-1648
285. Fenton, W. A., and Horwich, A. L. (1997) GroEL-mediated protein folding. *Protein Sci* **6**, 743-760
286. Izore, T., Job, V., and Dessen, A. (2011) Biogenesis, regulation, and targeting of the type III secretion system. *Structure* **19**, 603-612
287. Crosskey, T. D., Beckham, K. S. H., and Wilmanns, M. (2020) The ATPases of the mycobacterial type VII secretion system: Structural and mechanistic insights into secretion. *Prog Biophys Mol Biol* **152**, 25-34
288. Champion, P. A., Stanley, S. A., Champion, M. M., Brown, E. J., and Cox, J. S. (2006) C-terminal signal sequence promotes virulence factor secretion in *Mycobacterium tuberculosis*. *Science* **313**, 1632-1636
289. Wang, S., Zhou, K., Yang, X., Zhang, B., Zhao, Y., Xiao, Y., Yang, X., Yang, H., Guddat, L. W., Li, J., and Rao, Z. (2020) Structural insights into substrate recognition by the type VII secretion system. *Protein Cell* **11**, 124-137
290. Ernest, J. P., Strydom, N., Wang, Q., Zhang, N., Nuermberger, E., Dartois, V., and Savic, R. M. (2021) Development of New Tuberculosis Drugs: Translation to Regimen Composition for Drug-Sensitive and Multidrug-Resistant Tuberculosis. *Annu Rev Pharmacol Toxicol* **61**, 495-516
291. O'Brien, R. J., and Nunn, P. P. (2001) The need for new drugs against tuberculosis. Obstacles, opportunities, and next steps. *Am J Respir Crit Care Med* **163**, 1055-1058
292. Dagotto, G., Yu, J., and Barouch, D. H. (2020) Approaches and Challenges in SARS-CoV-2 Vaccine Development. *Cell Host Microbe* **28**, 364-370
293. Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., and Velesler, D. (2020) Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **183**, 1735
294. Choi, S., Choi, H. G., Shin, K. W., Back, Y. W., Park, H. S., Lee, J. H., and Kim, H. J. (2018) *Mycobacterium tuberculosis* Protein Rv3841 Activates Dendritic Cells and Contributes to a T Helper 1 Immune Response. *J Immunol Res* **2018**, 3525302
295. Bansal-Mutalik, R., and Nikaido, H. (2014) Mycobacterial outer membrane is a lipid bilayer and the inner membrane is unusually rich in diacyl phosphatidylinositol dimannosides. *Proc Natl Acad Sci U S A* **111**, 4958-4963
296. Rivera-Calzada, A., Famelis, N., Llorca, O., and Geibel, S. (2021) Type VII secretion systems: structure, functions and transport models. *Nat Rev Microbiol* **19**, 567-584

297. Congreve, M., Murray, C. W., and Blundell, T. L. (2005) Structural biology and drug discovery. *Drug Discov Today* **10**, 895-907
298. Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J., Staden, R., and Young, I. G. (1981) Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457-465
299. Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A. V., Mikheenko, A., Vollger, M. R., N., A., Uralsky, L., Gershman, A., Aganezov, S., Hoyt, S. J., Diekhans, M., Logsdon, G. A., Alonge, M., Antonarakis, S. E., Borchers, M., Bouffard, G. G., Brooks, S. Y., Caldas, G. V., Cheng, H., Chin, C. S., Chow, W., G., d. L. L., Dishuck, P. C., Durbin, R., Dvorkina, T., Fiddes, I. T., Formenti, G., Fulton, R. S., Functamman, A., Garrison, E., Grady, P. G. S., Graves-Lindsay, T. A., Hall, I. M., Hansen, N. F., Hartley, G. A., Haukness, M., Howe, K., Hunkapiller, M. W., Jain, C., Jain, M., Jarvis, E. D., Kerpedjiev, P., Kirsche, M., Kolmogorov, M., Korlach, J., Kremitzki, M., Li, H., Maduro, V. V., Marschall, T., McCartney, A. M., McDaniel, J., Miller, D. E., Mullikin, J. C., Myers, E. W., Olson, N. D., Paten, B., Peluso, P., Pevzner, P. A., Porubsky, D., Potapova, T., Rogaev, E. I., Rosenfeld, J. A., Salzberg, S. L., Schneider, V. A., Sedlazeck, F. J., Shafin, K., Shew, C. J., Shumate, A., Sims, Y., Simt, A. F. A., Soto, D. C., Sović, I., Storer, J. M., Streets, A., Sullivan, B. A., Thibaud-Nissen, F., Torrance, J., Wagner, J., Walenz, B. P., Wenger, A., Wood, J. M. D., Xiao, C., Yan, S. M., Young, A. C., Zarate, S., Surti, U., McCoy, R. C., Dennis, M. Y., Alexandrov, I. A., Gerton, J. L., O'Neill, R. J., Timp, W., Zook, J. M., Schatz, M. C., Eichler, E. E., Miga, K. H., and Phillippy, A. M. (2021) The complete sequence of a human genome. *BioRxiv*
300. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M. L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J.,

- Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y. H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigo, R., Campbell, M. J., Sjolander, K. V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yoosheph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y. H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., and Zhu, X. (2001) The sequence of the human genome. *Science* **291**, 1304-1351
301. Levinthal, C. (1969) How to Fold Graciously. in *Mössbauer Spectroscopy in Biological Systems Proceedings* (DeBrunner, J. T. P., and Munck, E. eds.), University of Illinois Press, Monticello, Illinois
302. Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Zidek, A., Bridgland, A., Cowie, A., Meyer, C., Laydon, A., Velankar, S., Kleywegt, G. J., Bateman, A., Evans, R., Pritzel, A., Figurnov, M., Ronneberger, O., Bates, R., Kohl, S. A. A., Potapenko, A., Ballard, A. J., Romera-Paredes, B., Nikolov, S., Jain, R., Clancy, E., Reiman, D., Petersen, S., Senior, A. W., Kavukcuoglu, K., Birney, E., Kohli, P., Jumper, J., and Hassabis, D. (2021) Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590-596
303. Ho, B., Baryshnikova, A., and Brown, G. W. (2018) Unification of Protein Abundance Datasets Yields a Quantitative *Saccharomyces cerevisiae* Proteome. *Cell Syst* **6**, 192-205 e193
304. Milo, R. (2013) What is the total number of protein molecules per cell volume? A call to rethink some published values. *Bioessays* **35**, 1050-1055
305. Briggs, J. A. (2013) Structural biology in situ--the potential of subtomogram averaging. *Curr Opin Struct Biol* **23**, 261-267
306. Bonvin, A. (2021) 50 years of PBD: a catalyst in structural biology. *Nat Methods* **18**, 448-449
307. Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015) The I-TASSER Suite: protein structure and function prediction. *Nat Methods* **12**, 7-8
308. Holm, L. (2020) Using Dali for Protein Structure Comparison. *Methods Mol Biol* **2112**, 29-42

309. Naydenova, K., McMullan, G., Peet, M. J., Lee, Y., Edwards, P. C., Chen, S., Leahy, E., Scotcher, S., Henderson, R., and Russo, C. J. (2019) CryoEM at 100 keV: a demonstration and prospects. *IUCr* **6**, 1086-1098
310. Peet, M. J., Henderson, R., and Russo, C. J. (2019) The energy dependence of contrast and damage in electron cryomicroscopy of biological molecules. *Ultramicroscopy* **203**, 125-131
311. van Schayck, J. P., van Genderen, E., Maddox, E., Roussel, L., Boulanger, H., Frojdh, E., Abrahams, J. P., Peters, P. J., and Ravelli, R. B. G. (2020) Sub-pixel electron detection using a convolutional neural network. *Ultramicroscopy* **218**, 113091
312. Beckham, K. S. H., Staack, S., Wilmanns, M., and Parret, A. H. A. (2020) The pMy vector series: A versatile cloning platform for the recombinant production of mycobacterial proteins in *Mycobacterium smegmatis*. *Protein Sci* **29**, 2528-2537
313. Arnold, F. M., Hohl, M., Remm, S., Koliwer-Brandl, H., Adenau, S., Chusri, S., Sander, P., Hilbi, H., and Seeger, M. A. (2018) A uniform cloning platform for mycobacterial genetics and protein production. *Sci Rep* **8**, 9539
314. Noens, E. E., Williams, C., Anandhakrishnan, M., Poulsen, C., Ehebauer, M. T., and Wilmanns, M. (2011) Improved mycobacterial protein production using a *Mycobacterium smegmatis* groEL1DeltaC expression strain. *BMC Biotechnol* **11**, 27
315. Iakobachvili, N., Leon-Icaza, S. A., Knoops, K., Sachs, N., Mazeres, S., Simeone, R., Peixoto, A., Murriss-Espin, M., Mazieres, J., Lopez-Iglesias, C., Ravelli, R., Neyrolles, O., Meunier, E., Lugo-Villarino, G., Clevers, H., Cougoule, C., and Peters, P. J. (2020) Mycobacteria-host interactions in human bronchiolar airway organoids. *BioRxiv*
316. Carmona, J., Cruz, A., Moreira-Teixeira, L., Sousa, C., Sousa, J., Osorio, N. S., Saraiva, A. L., Svenson, S., Kallenius, G., Pedrosa, J., Rodrigues, F., Castro, A. G., and Saraiva, M. (2013) Mycobacterium tuberculosis Strains Are Differentially Recognized by TLRs with an Impact on the Immune Response. *PLoS One* **8**, e67277
317. Jarlier, V., and Nikaido, H. (1990) Permeability barrier to hydrophilic solutes in *Mycobacterium chelonae*. *J Bacteriol* **172**, 1418-1423
318. Daffe, M., and Etienne, G. (1999) The capsule of *Mycobacterium tuberculosis* and its implications for pathogenicity. *Tuber Lung Dis* **79**, 153-169
319. Lunelli, M., Kamprad, A., Burger, J., Mielke, T., Spahn, C. M. T., and Kolbe, M. (2020) Cryo-EM structure of the *Shigella* type III needle complex. *PLoS Pathog* **16**, e1008263
320. Ghosal, D., Chang, Y. W., Jeong, K. C., Vogel, J. P., and Jensen, G. J. (2017) In situ structure of the *Legionella* Dot/Icm type IV secretion system by electron cryotomography. *EMBO Rep* **18**, 726-732
321. Chang, Y. W., Rettberg, L. A., Ortega, D. R., and Jensen, G. J. (2017) In vivo structures of an intact type VI secretion system revealed by electron cryotomography. *EMBO Rep* **18**, 1090-1099

List of publications

This thesis is based on the following publications

Cryo-EM structures from sub-nl volumes using pin-printing and jet vitrification

Ravelli, R. B. G., Nijpels, F. J. T., Henderikx, R. J. M., Weissenberger, G., Thewessem, S., **Gijsbers, A.**, Beulen, B., Lopez-Iglesias, C. and Peters, P. J. (2020). *Nature Communications* 11(1), 2563

***Mycobacterium tuberculosis* ferritin: a suitable workhorse protein for cryo-EM development**

Gijsbers, A. Zhang, Y., Gao, Y., Peters, P. J. and Ravelli, R. B. G. (2021). *Acta Crystallographica Section D: Structural Biology* 77(8), 1077-1083

Priming mycobacterial ESX-secreted protein B to form a channel-like structure

Gijsbers, A., Vinciauskaite, V., Siroy, A., Gao, Y., Tria, G., Mathew, A., Sánchez-Puig, N., López-Iglesias, C., Peters, P. J. and Ravelli, R. B. G. (2021). *Current Research in Structural Biology* 3, 153-164

Structural Analysis of the Partially Disordered Protein EspK from Mycobacterium Tuberculosis

Gijsbers, A., Sanchez-Puig, N., Gao, Y., Peters, P. J., Ravelli, R. B. G. and Siliqi, D. (2021). *Crystals* 11(1), 18

Structure of EspB–EspK complex: the non-identical twin of the PE–PPE–EspG secretion mechanism

Gijsbers, A., Eymery, M., Menart, I., McCarthy, A., Siliqi, D., Vinciauskaite, V., Gao, Y., Peters, P. J. and Ravelli, R. B. G. (Manuscript in preparation)

Other work published during the Ph.D.

Mutations in EFL1, an SBDS partner, are associated with infantile pancytopenia, exocrine pancreatic insufficiency and skeletal anomalies in a Shwachman-Diamond like syndrome

Stepensky, P., Chacón-Flores, M., Kim, K. H., Abuzaitoun, O., Bautista-Santos, A., Simanovsky, N., Siliqi, D., Altamura, D., Méndez-Godoy, A., **Gijsbers, A.**, Naser Eddin, A., Dor, T., Charrow, J., Sánchez-Puig, N. and Elpeleg, O. (2017). *Journal of Medical Genetics* 54 (8), 558-566

Interaction of the GTPase Elongation Factor Like-1 with the Shwachman-Diamond Syndrome Protein and Its Missense Mutations.

Gijsbers, A., Montagut, D.C., Mendez-Godoy, A., Altamura, D., Saviano, M., Siliqi, D. and Sánchez-Puig, N. (2018). *International Journal of Molecular Science* 19 (12), 4012

Exploring the role of elongation Factor-Like 1 (EFL1) in Shwachman-Diamond syndrome through molecular dynamics

Delre, P., Alberga, D., **Gijsbers, A.**, Sánchez-Puig, N., Nicolotti, O., Saviano, M., Siliqi, D. and Mangiatordi, G. F. (2019). *Journal of Biomolecular Structure and Dynamics* 38 (17), 5219-5229

Symmetry and planar chirality of a protein measured on an angular basis in a transmission electron microscope

Tavabi, A. H., Rosi, P., Ravelli, R. B. G., **Gijsbers, A.**, Rotunno, E., Guner, T., Zhang, Y., Rocaglia, A., Belsito, L., Pozzi, G., Tibeau, D., Gazzadi, G., Ghosh, M., Frabboni, S., Peters, P. J., Karimi, E., Tiemeijer, P., Dunin-Borkowski, R. E. and Grillo, V. (2021) preprint arXiv

Nature of polymorphisms in ESX-1 genes reveal evidence for converging evolution across MTBc Lineages

Vujkovic, A., Mulders, W., Vargas, R., Siroy, A., **Gijsbers, A.**, Keysers, J., Meehan, C. J., Farhat, M., Peters, P. J., de Jong, B. C. and Tzfadia, O. (Manuscript in preparation)

The presence of lipoprotein particles distorts the analysis and characterization of single extracellular vesicles

Lozano-Andrés, E., Enciso-Martínez, A., **Gijsbers, A.**, Libregts, S. F. W. M., Pinheiro, C., van Niel, G., Hendrix, A., Peters, P. J., Otto, C., Arkesteijn, G. J. A. and Wauben, M. H. M. (Manuscript in preparation)

Acknowledgments

This chapter of my life has come to an end. How do I feel? Many have asked me. Not a simple answer comes to my mind. There are many different feelings, some of them contradictory. I can be sure that I am proud of myself, what I have done, and what I have become. However, I could not have done it without the support of my family, friends, and colleagues. For that, I want to express my sincere gratitude.

Nuria, love, more than four years we spent apart, yet there was not a moment that I felt far from you. From the day we met, you knew that I had to leave the country to pursue my dreams, and the only thing I got from you was understanding and support. From your own experience, you helped me keep going by listening and giving me all the advice I needed. In the last months of this journey, you left home and compromised essential things to be by my side when I needed you the most. I will never thank you enough, but I can try by being there for you and giving you unconditional love.

Mamá y Johanna, hace seis años nuestras vidas cambiaron radicalmente por la pérdida de papá, un evento que nos marcó profundamente. Sin embargo, lo supimos superar juntas, uniéndonos como nunca. A ustedes les debo mi perseverancia, mi crecimiento, mi fuerza para sobrellevar las cosas. Sin su amor incondicional no hubiera podido seguir adelante. Por eso y mucho más les dedico este trabajo.

I was raised in Mexico, so when I say thank you to my family, I mean my whole family: from grandparents to nieces and nephews and beyond, my Mexican family, my Dutch family, my “chosen family” (friends). Thank you because you gave me the strength and support I needed! But especially to my Opa, a man with the biggest heart I have ever known. Opa, ik hou van je!!

To Peter, Raimond and Carmen, thank you all for believing and supporting me throughout this important period in my life. Thank you for allowing me to contribute something to the lab and the team. For the long and interesting talks, for the lovely moments.

In my time in Maastricht, I had the chance to meet incredible people that filled my life with joy and unforgettable moments, and that now I call friends. I started writing personal acknowledgements until I realised that no space would be enough to express my gratitude. With you, I ate, drank, travelled, and laughed so much. You taught me insignificant things and things that changed my life. Work-related and non-work-related. To that and much more, thank you: Navya, Yue, Rong, Hang, Nino, Jo, Klara, Frank, Giulia, Rene, Paul, Axel, Ye, Vanesa, Chris, Ruben, Kevin, Casper, Johana, Julia, Isabella, Dritan and Giancarlo.

Anita, you were there even before I knew I needed your help. You saved me countless times with bureaucratic stuff and even my housing. Thank you so much!

And last but not least, I want to thank all those who dared to call me supervisor, without whom I would not be the scientist I am today. I discovered the art and beauty of teaching with you, possibly my favourite part of the Ph.D. I wish you all the best in your future career.

Curriculum vitae

Abril Gijbbers Alejandre was born on May 4, 1990, in Oaxaca, Mexico, where she lived until 2008 that she moved to Mexico City to do her bachelor. In 2013, she obtained a diploma in Pharmaceutical and Biological Chemistry from the National Autonomous University of Mexico (UNAM). The subject of her dissertation was “Characterisation of the enzymic activity of the ribosomal GTPases EFTUD1/Efl1” from which Abril obtained a *magna cum laude* distinction. She continued her graduate studies in the same institution. In 2015, she received a Master’s degree in Biochemistry with *summa cum laude* distinction based on her thesis “Characterisation of the interaction between EFL1 protein and the protein involved in the Shwachman-Bodian-Diamond Syndrome (SBDS)”.



In 2016, Abril moved to the Netherlands and joined the Nanoscopy group at Maastricht MultiModal Molecular Imaging (M4I) Institute. The aim of the group was to solve the structure of the type VII secretion system from *Mycobacterium tuberculosis*. Within this subject, she designed and performed her project with the supervision of Prof. Peter J. Peters and Raimond B. G. Ravelli, which was initially dedicated to characterise protein substrates of the aforementioned secretion system. However, Abril was an active participant in multiple projects within and outside the Nanoscopy group, leading to other publications.

In 2021, she moved back to Mexico, where she will continue in academia doing a postdoctoral internship in Systems Biology in the Genomic Science Centre, UNAM.

