Technische Universität München
TUM School of Natural Sciences

HPLC-MS-based proteomics of mycobacteria: Bottom-up and intact mass approaches for characterizing antibiotic resistance profiles

Anja Sophie Dollinger

Vollständiger Abdruck der von der TUM School of Natural Sciences der

Technischen Universität München zur Erlangung des akademischen Grades einer

Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: Prof. Dr. Nicole Strittmatter

Prüfende der Dissertation:

1. apl. Prof. Dr. Christoph Haisch
2. Prof. Dr. Michael Groll

Die Dissertation wurde am 03.07.2025 bei der Technischen Universität München eingereicht und durch die TUM School of Natural Sciences am 11.09.2025 angenommen.

Acknowledgements

I would like to thank the Bavarian State Ministry for Science and the Arts through the research network bayresq.net for supporting this work financially.

To my lab colleagues **Gabi, Nici,** and **Sarah**: thank you for your hands-on help, the solidarity, and for being part of the everyday chaos, in the best possible way. Special thanks to my group (**AG Haisch**) for the scientific input, the collaborative troubleshooting, and your moral support every time my mass spectrometer decided to be dramatic. I am so grateful to **Alex**, who, during his brief time in the lab, helped me get started with the practical and conceptual basics that proved essential moving forward. My sincere thanks to **Dr. Woisch** for his willingness to share his knowledge so generously and for his kind support throughout my time in the lab. I would also like to thank **Luiz** and his lab team in Oregon for the warm welcome during my stay, for everything I was able to learn and try out in the lab, and for the scientific advice that continued well beyond my time there. Heartfelt thanks to **Lynette**, whose kindness and hospitality made my visit truly memorable.

A special thank-you to **Barbara**, who joined as an intern but contributed far beyond what was expected - your contribution is truly appreciated. I'm especially grateful to my supervisor, **Christoph**, for always being available, for the constant support, and for creating space for questions, ideas, and just the right dose of perfectionism. A heartfelt thank-you also goes to **Ines** for being a steady collaborator from start to finish, for the support, the shared work, and for getting through it all together over the past three and a half years (including some R tutorials along the way). Thanks also to **Jessica** and **Anna** – your guidance and encouragement meant a lot, both professionally and personally. And of course, to **Tom** – thank you for being the kind of person who helps even when you really don't have time. The world (and every lab) needs more people like you.

A big thank-you to my office mate and true partner-in-crime, **Susanne**, for all the laughs, the after-lunch walks, the deep (and not-so-deep) conversations, and the mental support that kept me grounded throughout. And to **Lilly**, who joined us later and quickly became an essential part of our little office world – thank you for your encouragement and for sharing in the everyday chaos and fun. I'm also grateful to my dear neighbor **Valentina** for the many coffee breaks and for reminding me that lunch is, in fact, not optional.

Above all, I want to thank my partner **Thomas** for always listening, always cheering me on, and never questioning the necessity of emergency chocolate. And finally, to **my family**: your unwavering belief in me has meant everything. I couldn't have done this without you.

Publications

The following article resulted from a side project conducted during my doctoral studies.

Dietrich, S., **Dollinger, A**., Wieser, A., & Haisch, C. (2025). Optimization of MALDI Matrices and Their Preparation for the MALDI-TOF MS Analysis of Oligonucleotides. *Rapid Communications in Mass Spectrometry*, *39*(15), e10061.

"Science is not only a discipline of reason but, also, one of romance and passion."

Stephen Hawking

Abstract

Tuberculosis (TB), with its drug-resistant strains, is the world's most deadly disease from a single infectious agent. This study gives new insights into antibiotic reaction mechanisms on the proteoform level of mycobacteria, by employing protein analysis either as intact mass analysis with untargeted labeling or with tandem-mass-tag (TMT) proteomics.

In contrast to the more commonly applied bottom-up proteomics, this study employed intact mass analysis, allowing the assessment of full-length protein species and their mass variants. High-performance liquid chromatography coupled to mass spectrometry (HPLC-MS) analysis of *Mycobacterium smegmatis (M. smegmatis)* wild-type resulted in the detection of 229 proteins. The investigation of dormant *M. smegmatis* proteins revealed 34 proteins, with 24 proteins matching the intact mass of the wild-type proteins. By modifying the untargeted labeling approach from Haisch *et al*. by using HPLC-MS instead of matrix-assisted laser desorption−ionization time-of-flight mass spectrometry (MALDI-TOF MS) readout, the proteome coverage was extended to hundreds of proteins. The dynamics of six intact proteins were analyzed in more detail over time, showing distinct response times for each protein. The effect of the first-line drug rifampicin on the protein *de novo* synthesis could be detected within about half the regular doubling cycle (1.5 h) for *M. smegmatis*. 15 of the intact proteins were identified by bottom-up proteomics. TMT analysis across multiple time points in wild-type and dormant bacterial states resulted in distinct patterns of up- and downregulated proteins in response to the antibiotics rifampicin and BTZ-043. Many of the proteins have not been fully characterized yet. Network analysis revealed several hub proteins, suggesting potential regulatory functions under both conditions.

This work, especially with the analysis of intact protein masses, offers unique insights into comprehensive proteomic profiles that may not be accessible through conventional bottom-up proteomics. The TMT analysis revealed several protein-protein interactions involving proteins that may represent novel drug targets. The results contribute to a deeper understanding of bacterial adaptation mechanisms and offer starting points for future therapeutic strategies.

Zusammenfassung

Tuberkulose (TB) ist mit ihren arzneimittelresistenten Stämmen die tödlichste Krankheit der Welt, die durch einen einzigen Infektionserreger verursacht wird. Diese Studie bietet neue Einblicke in die Mechanismen der Antibiotikareaktion auf Proteoform-Ebene von Mykobakterien, indem die Proteinanalyse entweder als intakte Massenanalyse mit ungerichteter Markierung oder als Tandem-Atommassenmarkierung (Tandem Mass Tag; TMT) eingesetzt wird.

Im Gegensatz zu der häufiger angewandten Bottom-up-Proteomik wurde in dieser Arbeit die Analyse intakter Proteine durchgeführt, wodurch sich vollständige Proteine und ihre Massenvarianten erfassen lassen. Die Analyse von Wildtyp-*Mycobacterium smegmatis (M. smegmatis)* mit Hochleistungsflüssigchromatographie gekoppelt mit Massenspektrometrie (engl.: high-performance liquid chromatography coupled to mass spectrometry; HPLC-MS) ergab 229 intakte Proteine. Die Untersuchung der dormanten *M. Smegmatis* Proteine zeigte 34 intakte Proteine, wovon 24 Proteine mit der intakten Masse der Wildtyp-Proteine übereinstimmten. Durch die Verwendung von HPLC-MS anstelle von Matrix-unterstützter Laser-Desorption/Ionisation Flugzeit Massenspektrometrie (engl.: matrix-assisted laser desorption−ionization time-of-flight mass spectrometry; MALDI-TOF) wie in der von Haisch *et al*. beschriebenen nicht-selektiven Proteinmarkierung, wurde die Proteomabdeckung um Hunderte von Proteinen erweitert. Sechs distinkte intakte Proteine wurden hinsichtlich ihrer zeitlichen Dynamiken näher analysiert, wobei jedes Protein eine charakteristische Reaktionszeit aufweist. Zusätzlich konnten die Effekte des Erstlinien-Antibiotikums Rifampicin auf die *de-novo*-Proteinsynthese innerhalb des halben Verdopplungszyklus (1,5 Stunden) von *M. Smegmatis* beobachtet werden. 15 intakte Proteine wurden durch Bottom-Up Untersuchungen eindeutig identifiziert. Die über mehrere Zeitpunkte hinweg durchgeführte TMT-Analyse von Wildtyp- und dormanten Mykobakterien ergab unterschiedliche Muster an hoch- und herunterregulierten Proteinen als Reaktion auf die Antibiotika Rifampicin und BTZ-043. Viele der Proteine sind bisher noch nicht vollständig charakterisiert. Eine Netzwerk-Analyse zeigte mehrere Schlüssel-Proteine, was auf mögliche regulatorische Funktionen hinweist.

Die vorliegende Arbeit, insbesondere die Untersuchung der intakten Proteinmassen, ermöglicht weiterführende Einblicke in umfassende Proteom-Profile, welche durch die konventionelle Bottom-Up Proteomik möglicherweise nicht oder nur unzureichend zugänglich wären. Die TMT-Analyse identifizierte mehrere Proteininteraktionen, die funktionell relevante Zielproteine beinhalten und potenzielle Angriffspunkte für Wirkstoffe darstellen. Die Ergebnisse tragen zu einem tieferen Verständnis der bakteriellen Anpassungsmechanismen bei und bieten Ansatzpunkte für künftige therapeutische Strategien.

Contents

1 Introduction

It is referred to as the "silent pandemic". Each year, millions of people die due to antimicrobial resistance (AMR) and its consequences. AMR occurs when antimicrobial medicines are no longer effective against bacteria, viruses, fungi, and parasites. This poses a significant threat to modern medicine, which is facing an antibiotics pipeline and access crisis.[1] Drug-resistant tuberculosis (TB) is an example of this threat. Once mainly considered manageable, TB has reemerged as a significant challenge. After the outbreak of the COVID pandemic, TB infections have been rising again since 2020, and in 2023, TB returned to being the world's most deadly disease from a single infectious agent. Approximately 1.25 million people died that year due to TB, with over 10 million infections worldwide.[2] One reason for this is the emergence of drug-resistant strains. Multi-drug-resistant TB (MDR-TB) and extensively drug-resistant TB (XDR-TB) are global health emergencies requiring complex, lengthy, and often side-prone treatments with limited success.[2] For this, it is necessary to better understand the metabolic bacterial reactions to the drugs and drug combinations to develop new antibiotics and improve current treatment regimens. So far, this has been investigated in *in vitro* experiments like Green Fluorescent Protein (GFP) reporter assays, resazurin assays, or enumerating the colony-forming unit (CFU). Yet, these studies do not consider the progressive, dynamic change within the bacterial cell, meaning that an important antibiotic effect on mycobacterial metabolism may be underestimated or overlooked. One possibility to address these challenges is to research the bacterial proteome, which has gained growing importance with increasing method development. Proteomic studies can enable new insights into the current bacterial cell state and aim to provide further information on their persistence and antibiotic tolerance mechanisms. New screening methods with untargeted labeling enable the observation of antimicrobial effects on the *de-novo* protein synthesis of mycobacteria.[3] However, approaches like tandem-mass-tag (TMT) labeling also enable multiplexing and relative, high-throughput protein quantification across multiple conditions.[4] It allows for simultaneous analysis of (drug-induced) proteomic changes, providing mechanistic insights and potential biomarker discovery, making it a valuable method for antibiotic research and tuberculosis studies.

2 Fundamentals

2.1 Mycobacteria

The genus *Mycobacterium* comprises 195 species. The species exhibit phylogenetic distinctions and can be distinguished into *fast-growing* and *slow-growing* mycobacteria.[5, 6] Most mycobacteria are fast-growing (doubling time ~2-6 h) and non-pathogenic. As they can proliferate outside a host, they can be found in the environment, like soil or water systems. The three significant pathogens for humans are *Mycobacterium leprae*, *Mycobacterium ulcerans*, and the *Mycobacterium tuberculosis* complex (MTBC), classified as slow-growers (doubling time > 10 h).[5, 7] The MTBC encompasses a group of closely related (> 99 % nucleotide sequence identity) bacilli responsible for tuberculosis (TB) in animals and humans.[5] Within the MTBC, *Mycobacterium tuberculosis* (*M. tuberculosis*) is the primary causative agent of human TB.[5] It was first discovered in 1882 by Robert Koch, making it the first microorganism to be identified as the pathogen behind a disease.[8]

M. tuberculosis are rod-shaped, acid-fast, aerobic bacteria incapable of active movement. The bacteria show characteristics of gram-negative and gram-positive features, but have been classified as gram-positive based on molecular taxonomy.[9] One unique characteristic of mycobacteria is their cell wall, which is schematically represented in **Figure 1**. The cell wall can be divided into three main compartments: the plasma membrane, the inner layer (cell wall core), and the outer layer. The plasma membrane surrounds the intracellular environment as a phospholipid bilayer and resembles other bacterial membranes. Next to the plasma membrane, cross-linked peptidoglycan (PG) molecules, similar to those of gram-positive bacteria, build a thick layer (10 to 20 layers). The PGs are covalently linked to an outer layer of arabinogalactans, which are again bonded to mycolic acids. The mycolic acids represent the major compartment of the cell wall and determine its permeability. Mycolic acids are long-chain fatty acids consisting of about 60 to 90 carbon atoms, imparting a high degree of hydrophobicity. This hydrophobicity is essential as it protects the cell from hydrophilic antibiotics or oxidative damage. The cell wall complex is surrounded by the outer layer, which is interleaved with further mycolic acids and multiple diverse lipids and proteins that act as signaling and effector molecules.[9, 10]

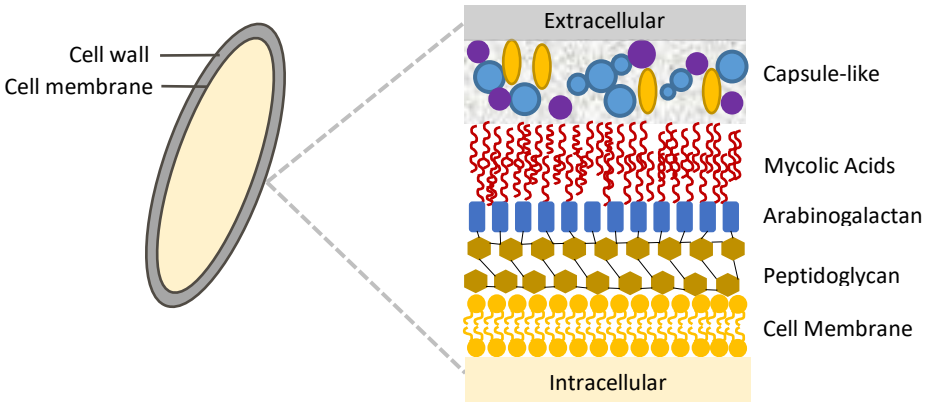


Figure 1: Scheme of the essential components of the bacterial cell wall, adapted from Hett and Rubin.[9]

In vitro experiments with *M. tuberculosis* prove to be complicated. *M. tuberculosis* requires bio-safety level (BSL) 3 standards, and the division rate is in the range of 20-24 h, making experiments in the laboratory time-consuming. Consequently, more easily handled surrogate organisms are often used as substitutes for *M. tuberculosis* in TB research.[11]

Mycobacterium smegmatis (M. smegmatis) is widely used as a model organism. Especially the *M. smegmatis* strain derivative MC^2 155, which was engineered in 1990, is the most commonly used strain due to its high genetic transformability.[12] Other factors contributing to the frequent use of *M. smegmatis* in research are the fast growth rate (2-6 h) and its non-pathogenic nature. Experiments with *M. smegmatis* only require BSL 2 (Germany) or BSL 1 (USA) standards. The genome size of *M. smegmatis* is larger than that of *M. tuberculosis*, comprising 6.8 Mb (NCBI RefSeq[13], 2019) compared to 4.4 Mb (NCBI RefSeq[14], 2013). Despite this difference in size, several essential genes are orthologous in both species. Over 2,800 of the about 4,000 protein-coding genes of *M. tuberculosis* are orthologous in *M. smegmatis*, with the core proteins functioning similarly in both species.[11] Also, the structural and physiological characteristics of the bacteria resemble each other. These similarities make *M. smegmatis* a suitable model organism for *M. tuberculosis*. They explain its wide usage for genetic and proteomic studies, optimizing extraction protocols, and drug development and target studies.[11, 15-17]

2.1.1 Antibiotic resistance and dormancy of *M. tuberculosis*

TB is generally preventable and curable. An overview of TB incidences and main events in TB-related research over time is represented in **Figure 2**. The first vaccine against TB (Bacille Calmette-Guérin) was discovered in 1921, and about 20 years later, the first anti-TB drugs were invented.[18, 19] With the World Health Organization's (WHO's) currently recommended regimen (4–6 months of anti-TB medication), approximately 85 % of TB cases can be successfully treated.[2] The WHO set the goal to reduce TB infections by over 50 % and a 20 % reduction in new cases by 2025, ultimately aiming for a 95 % reduction in deaths and a 90 % reduction in cases by 2035 ("End TB Strategy").[20] However, TB remains one of the most critical health concerns of the WHO. A significant challenge in TB treatment is antibiotic resistance and latency. Shortly after the first drug developments, the first drug-resistant TB strains occurred. In 2006, extensively drug-resistant (XDR) TB strains were officially defined, highlighting the growing challenge of drug resistance. Between 1995 and 2014, new TB infections rose from about 1 million to nearly 10 million due to multiple factors, including drug resistance and the HIV epidemic.[21, 22] As of 2023, approximately 400,000 people infected with TB were either resistant to the first-line drug rifampicin (RR-TB) or multidrug-resistant (MDR-TB), defined as resistance to isoniazid (H, first-line drug) and rifampicin (R).[23]

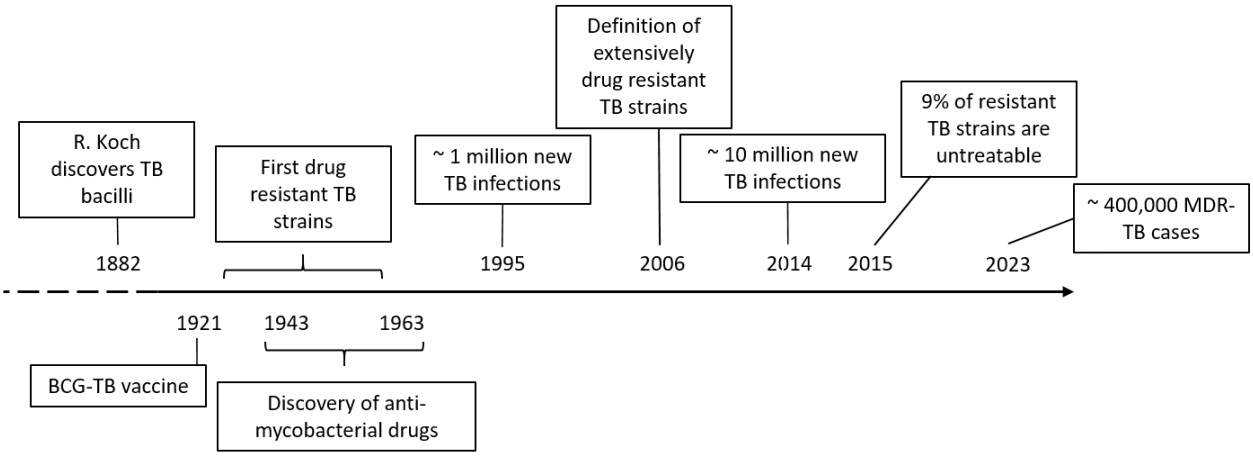


Figure 2: Schematic timeline of TB incidence and main events in TB-related research, adapted from Drapal et al.[21]

The antibiotic resistance of mycobacteria is attributed to multiple mechanisms and factors. One important role, as already mentioned before, plays the unique and complex mycobacterial cell envelope with its sugar, protein, and lipid layers. Those components are responsible for a lower

cell wall permeability, slowing the import of drugs into the cell and, therefore, limiting their interaction with intracellular targets.[24-26] The cell wall plays a vital role in resilience, but is not solely responsible for mycobacterial antibiotic resistance.[24]

The impermeable structure of the cell wall enables the bacteria to survive under extreme environmental stressors like hypoxia and nutrient depletion. Faced with such conditions, the bacteria adapt by entering a metabolically reduced state. Up to this point, there has been an extensive debate about the states of *metabolically inactive* or *low-activity* mycobacteria. The review by Chao and Rubin from 2010[27] summarizes and discusses these definitions, which are reiterated and further examined in the following section.

One term used for bacteria in a non-replicating, metabolically reduced but stable state is called *dormant*. In dormancy, the bacteria stop cell division, rendering most antibiotics that target growing cells ineffective. When conditions become more favorable, dormant bacteria return viable and can be cultured on nutrient agar to form colonies. Another metabolic state is *viable but not culturable* (VBNC). VBNC is primarily achieved upon long-term culturing and was observed in the murine Cornell model.[28, 29] VBNC bacteria are viable but cannot be cultured unless reactivated with exogenous factors such as immune suppression.[30-32] The phenotypic survival of bacterial cells under drug treatment is called persistence, and typically involves a small subpopulation of cells that enter a transient, drug-tolerant state. Persistent cells also show minimal metabolic activity, making them non-culturable. Latency describes the clinical condition in which no active disease emerges in patients. It is often assumed that bacteria of a latent infection are dormant, but this hypothesis has not been confirmed so far.[33] Furthermore, discussions exist about which model represents “true” dormancy or only a pre- or pseudo-dormant state. This study focuses not on differentiating between these states but on gaining deeper insights into the mechanisms underlying antibiotic tolerance and resistance and the progression toward it. The distinction between those states is complex, not fully understood, and yet to be clearly defined.[27, 34] They will collectively be referred to as dormant for simplification.

Several protocols are known to induce *in vitro* dormancy, each employing a different form of stress to the bacteria. The most common are hypoxia (Wayne model), nutrient depletion, hyper-acidity, high drug concentrations or prolonged antibiotic treatment, and various combinations

of these.[27] Dormant bacteria exhibit distinct morphological and physiological characteristics compared to wild-type or reactivated cells. This is shown in the scanning electron microscopy (SEM) images (s. **Figure 3**) presented by Neumann *et al*.[15] Dormant bacteria are smaller (1.76 ± 0.42 µm vs. 4.0 ± 1.1 µm) and tend to form clumped and 'pearl-string'-like colonies in contrast to the smooth and evenly shaped wild-type cells.[15] Reproduced *in vitro* dormancy could demonstrate that the bacteria can become resilient to the host defenses and that antibiotic therapies render less efficient.[35]

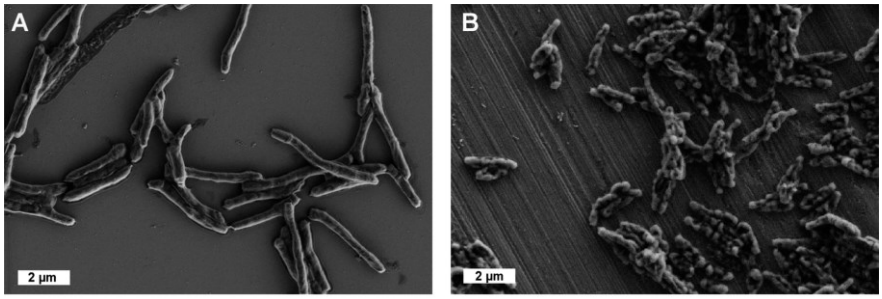


Figure 3: Scanning electron microscopy (SEM) images of *M. smegmatis* wild-type (A) and *M. smegmatis* in dormant state (B). Reprinted with permission from Neumann *et al*.[15] Copyright 2019 American Chemical Society.

2.1.2 Mycobacterial proteins and their role in antibiotic resistance

Since proteins are the main molecular effectors of bacterial cells, protein analysis has become indispensable for understanding cellular functions. Numerous proteomic studies have provided essential facts and insights into the mechanisms behind the antibiotic resistance of mycobacteria, investigating protein expressions, structures, and modifications.

Prior to protein analysis, proteins must be extracted from the mycobacterial cells. One key challenge in this context is the unique mycobacterial cell envelope, as mentioned before (s. chapter **2.1**). Standard lysis techniques, such as chemical lysis with commonly used buffers, are mostly ineffective due to the lipid shell of the bacteria. Therefore, harsh disruption methods have to be applied. Optimized lysis buffers containing detergents and chaotropic substances improve cell lysis and protein solubilization.[16, 36, 37] Additionally, treatment of the cells via bead beating, sonication, or French press can be applied to mechanically disrupt the mycobacterial cell wall.[16, 38, 39] Moreover, when HPLC-MS analysis is used, buffer-, salt-, and detergent-free conditions are required. Further steps for protein purification, such as precipitation or filtering, result in mycobacterial protein analysis being complex and time-consuming.[40-42]

A focus of proteomic investigations is the comparison of antibiotic-resistant and susceptible strains of mycobacteria. Differential protein expression experiments showed that specific proteins are overexpressed when treating *M. tb* with elevated drug concentration. With this, specific "escape" pathways of the bacteria were discovered, prolonging survival under antimicrobial exposure. Inactivation of some of these proteins in *M. tuberculosis* prevented the pathogen from escaping the rapid efficacy of antibiotics *in vitro*.[17] Investigating streptomycin-resistant strains, overexpression of specific proteins, and further *in silico* analysis showed significant interaction between those proteins and the used drug.[43] Similarly, experiments on amino-glycoside-resistant strains showed an overexpression of specific 'capsule proteins'. They were identified as being involved in intermediary metabolism, cell wall processes, and virulence.[44] A high-throughput proteogenomic study from de Souza *et al*. in 2024 identified uncharacterized proteins that contribute to drug tolerance. These newly discovered so-called microproteins, often not predicted by genomic analyses, suggest alternative pathways for antibiotic resistance and potential new drug targets.[45] Another study revealed that proteins can

alter the lipid composition of the cell wall by methylation. This promotes bacterial adaptation and survival under stress and acid conditions and increases antibiotic resistance.[46] One crucial function of membrane proteins is their role as efflux pumps, which actively transport substrates from the cytoplasm to the exterior of the cell. Bendre *et al.* summarized structural and mechanistic information on mycobacterial membrane proteins, emphasizing the importance of understanding their modes of action and potential as drug targets.[47] Further studies have also highlighted differences in the proteome of dormant and wild-type mycobacteria due to adaptive survival strategies to stress conditions.[48, 49] While ribosomal proteins remain mainly unchanged in expression, proteins from stress-response regulators are significantly upregulated when the bacteria are exposed to hypoxic conditions. For example, proteins induced from the transcription factor DosR regulon constitute approximately 20 % of the total cellular protein content, whereas, under normal growth conditions, they account for only about 1 %.[49] Modifications of proteins that are not directly encoded by genes initially, known as post-translational modifications (PTMs), also seem to play a crucial role in the case of antibiotic resistance. Sakatos *et al*. found phenotypically resistant subpopulations being regulated by the histone-like protein HupB. The protein exhibited PTMs on the lysine residues, such as acetylations and methylations. The authors demonstrated that those PTMs are crucial in forming antibiotic-resistant subpopulations, as their mutation inhibited subpopulation synthesis. These findings provided the first evidence that PTMs regulate cell states in prokaryotes [50]

To summarize, proteomic studies have demonstrated several connections between various adaptation/resistance mechanisms and proteins. Investigating the proteomic level of mycobacteria may contribute to developing novel drugs and treatment regimens. Despite the already-mentioned discoveries, the mycobacterial proteome remains not fully explored and requires further investigation to better understand the pathogen and its resistance.

2.2 Analytical methods for protein analysis

Within the last decades, the investigation of proteins has become of great interest to further understand complex biological systems. Due to the proteins' complex structure and dynamic nature, their analysis requires advanced analytical techniques. Particularly common for this are high-performance liquid chromatography and mass spectrometry.

2.2.1 High-performance liquid chromatography

In high-performance liquid chromatography (HPLC), the analyte mixture is injected into the chromatographic system, where it is pumped through the system with high pressure up to ~700 bar. A mobile phase guides the analytes through the system to a column packed with a stationary phase with which the analytes adsorb and interact depending on their chemical properties. Subsequently, the mobile phase, containing the analytes, is guided to a detector. The duration for each analyte to pass through the chromatographic system, including the interactions with the stationary phase, is called retention time (RT).

The stationary phase can be either polar, a so-called normal phase (NP), or an unpolar reversed phase (RP). RP chromatography is most frequently applied in bioanalytics. While reversed-phase liquid chromatography (RPLC) is a well-established technique in peptide separation, proteins exhibit greater structural and chemical diversity than smaller peptides, requiring additional considerations for optimizing separation conditions. RP columns often consist of silica particles modified with alkyl chains of various lengths (primarily C_4, C_8, or C_{18}). Compared to intact proteins, peptides are mostly separated on C_{18}-columns, as the hydrophobic stationary phase provides strong interaction with the peptide backbone. Pores with sizes of about 80-120 Å provide a high contact surface of the stationary phase and enable a high diffusion of the peptides into the porous material. For intact protein analysis, less hydrophobic stationary phases (C_4, C_8) are preferably employed as they offer higher protein recovery.[51, 52] Conventional small porous material is often less efficient at separating proteins than peptides. This is mainly due to a low mass transfer of the larger proteins into the small pores. This is usually improved with packing materials of larger pores (300-1,000 Å), leading to a higher diffusion coefficient of the proteins.[52]

For a stable and defined temperature for the separation, the column is placed in a programmable column oven. In protein analysis, separation temperatures are typically higher than room temperature (25-90 °C).[53-56] With elevated temperatures, the viscosity of the mobile phase is reduced, leading to lower backpressure, higher diffusion coefficients, and increased sorption kinetics. The results are enhanced separation efficiency and higher reproducibility.[52-57] However, excessively high temperatures can affect protein stability and column longevity.

For protein analysis, additives are often used in the mobile phase to reduce ionic interactions between the proteins and the functional groups of the stationary phase. A conventional additive is trifluoroacetic acid (TFA) in low concentration (e.g., 0.1 %), serving as an ion-pairing agent and improving protein solubilization. However, when LC is coupled to electrospray ionization (ESI-)MS, TFA can lead to ion suppression, comprising signal-to-noise ratios (S/N). A replacement for TFA is often formic acid (FA). While FA does not achieve the same chromatographic performance as TFA, it shows higher ionization efficiency.[52, 58]

Traditionally, LC is operated with flow rates of 0.1–2 mL/min through analytical and microbore columns, which exhibit an inner diameter (ID) of 1.5-4.6 mm. Reducing the ID, the required flow rate, and mobile phase volumes are decreased accordingly. If the elution time of the analyte remains the same, the concentration of the analyte is increased, and desolvation efficiency in the ESI is improved. As a result, signal intensity is enhanced, thereby improving S/N. Capillary columns (ID: 0.15-0.5 mm) and nano columns (ID: 10-150 µm) are commonly used for protein analysis. Capillary columns operate at 1-10 µL/min flow rates and exhibit higher sensitivity (5-30 times) than analysis with analytical columns. Besides improved ionization and high sensitivity, small-ID columns require less sample amount, which is especially beneficial in research fields, such as proteomics or clinical studies, where the sample amount is often limited. Therefore, analysis is frequently conducted in nano-flow mode (< 1.5 µL/min).[59-61] However, using capillary and especially nano columns requires specialized instruments due to their minimal flow rates, entailing several challenges, such as decreased robustness and more complex system handling.[59, 62]

Despite the technological developments in column design, RPLC of proteins is still more challenging than chromatography of peptides. This is mainly due to irreversible adsorption, carryover, insufficient mass transfer, and lower chromatographic performance.[63, 64]

Nevertheless, RPLC remains preferred due to its compatibility with volatile mobile phases and additives, allowing for mass spectrometry coupling.

2.2.2 Mass spectrometry

Mass spectrometry (MS) is an analytical tool to identify, characterize, and/or quantify molecules by measuring their mass-to-charge ratio (m/z). For this, the analytes are converted into the gas phase and ionized in an ion source. Afterward, a mass analyzer separates the ions according to their m/z. A detector then records the resulting ion current.

MS has become an essential tool in protein analysis due to its high sensitivity, accuracy, and throughput. To gain more information about the complex structure of proteins and peptides, such as the specific amino acid sequence, including mutations and PTMs, tandem mass spectrometry (MS/MS) is often applied. In MS/MS, a specific precursor ion is selected (MS1) and then fragmented (MS2). The corresponding product ions are characteristic for the particular protein or peptide. The type of fragment generated depends on the fragmentation method applied. Several fragmentation techniques were invented over time to improve sequence coverage for comprehensive protein analysis. The possible fragments and their nomenclature are depicted in **Figure 4**.

The most commonly used fragmentation method is collision-induced dissociation (CID) or collisionally activated dissociation (CAD). In CID, analyte ions are accelerated and collide with inert gas molecules like helium, argon, or nitrogen. This results in the cleavage of the weakest bonds: the amide N-C(O)-bonds along the peptide backbone. The fragments most likely for CID are b- and y-sequence ions. Generated product ions are primarily sufficient for protein identification. However, CID often provides limited sequence coverage and may cleave fragile PTM bonds. Due to this, CID works best for the fragmentation of small and low-charged peptides but is ineligible for detailed protein analysis.[65, 66] A fragmentation method similar to CID is the higher-energy collisional dissociation (HCD), referred to as “higher-energy C-trap dissociation” introduced by Olsen *et al*. in 2007.[67] The approach was initially invented to prevent low molecular weight cut-off of the fragment ions, which is a limitation when CID is applied in ion traps. HCD can be performed in hybrid-ion trap orbitrap mass spectrometers, where the fragmentation occurs outside the ion trap in a separate collision cell (HCD cell). HCD provides more comprehensive information about the analyte than ion-trap CID. The resulting

fragments are b- and y-, but also smaller internal fragments, providing higher sequence coverage.[65]

An alternative fragmentation technique is electron-capture dissociation, in which the multiply charged analytes interact with low-energy electrons (< 0.2 eV) emitted from a dispenser cathode. This leads to specific charge-reduced radical cations, quickly disintegrating along the N-Cα-bond into c- and z·-ions. ECD yields higher sequence coverage than CID and is, therefore, often used to analyze intact proteins and PTMs, as the fragmentation does not cleave labile PTM bonds. The requirement to efficiently trap low-energy electrons limits ECD to Fourier-transform ion cyclotron resonance mass spectrometers (FT-ICR-MS). This challenge was addressed by the invention of electron-transfer dissociation (ETD). ETD can be performed on a wide range of mass spectrometers, as the analyte cations receive the electrons by transfer from a radical anion (e.g., anthracene, fluoranthene). Like ECD, ETD provides c- and z·-ions while preserving labile PTM bonds intact and is thus used to analyze full-length proteins.[65]

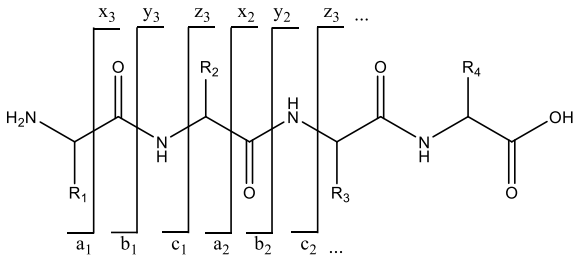


Figure 4: Illustration of the most common peptide fragmentations and their nomenclature. Demonstrated with a representative peptide including a-, b-, c- ions from the N-terminus and x-, y-, z- ions from the C-terminus. Residues are denoted as R.

2.3 Mass spectrometry-based proteomics techniques

Proteomics is the study of all proteins in a biological system, such as cells or living beings (proteome). As opposed to the genome, the proteome is dynamic and characterized by changes in multiple quantities and various protein forms, so-called *proteoforms*. The term *proteoform* was introduced by Smith and Kelleher to emphasize the complex naming of related proteins, considering all different modifications such as sequence variations, splice isoforms, and PTMs.[68]

Proteomic approaches can be separated into two major workflows, schematically represented in **Figure 5**. One method is the so-called bottom-up approach, in which the proteins are first enzymatically digested into peptides and then separated, for example, by liquid chromatography and analyzed via mass spectrometry. The other approach is top-down proteomics. In this workflow, the intact proteins are purified, separated (HPLC), and then analyzed via MS. After the first dimension of MS/MS analysis, the m/z signals of the proteins can be deconvoluted into the corresponding neutral mass of the protein (intact mass). The proteins of interest can be identified by fragmentation of the precursor ions via second MS (top-down). Both workflows, the bottom-up and the top-down approach, are discussed in more detail in the following two chapters.

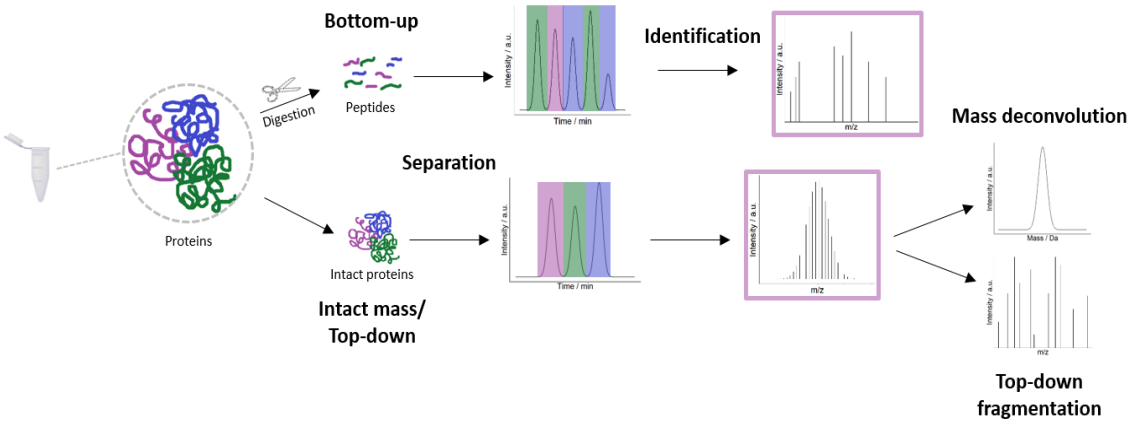


Figure 5: Schematic representation of bottom-up and intact mass/top-down proteomics principles.

2.3.1 Bottom-up proteomics

Bottom-up proteomics (BUP) is a popular method for identifying and characterizing proteins by analyzing their amino acid sequences. The term *bottom-up* implies that the information about a protein is reconstructed from its fragments. These fragments (peptides) are created by proteolysis (digestion) of the protein. For digestion, proteins are treated with endoproteases, like trypsin or Lys-C, cleaving the proteins into peptides. The peptides are separated via LC and fragmented with coupled MS/MS (shotgun proteomics). The resulting peptide patterns are matched with predicted database sequences or annotated peptide spectral libraries to identify and characterize the corresponding proteins.

BUP is a widely used approach in proteomics research and a well-developed and established method, enabling high-throughput analysis of multiple samples. As an advantage, digestion often only requires micro- to nanograms of proteins, allowing minimal sample consumption.[69-71] The chemical similarity of the peptides facilitates the analysis of especially complex samples, even if these originally contained (chemically) very different proteins. Above all, established bioinformatics tools and databases ease the identification of the corresponding proteins. One disadvantage is that digestion often requires multiple hours (4-16 h, 37 °C), making the sample preparation very time-consuming. Another limitation is that the peptide sequence coverage of proteins is usually incomplete, leading to information loss about the (native) protein structure. As the proteins are digested into peptides, potentially important regions may be cleaved and undetected in the MS. With this, the reconstruction of complete protein sequences is limited, and PTM localization or the differentiation between protein isoforms becomes challenging. Additionally, this can lead to false positives, assuming that proteins are intact, even when partially degraded. Also, proteins with similar sequences (homologs) might not be differentiated.[72, 73]

2.3.1.1 Data acquisition in bottom-up proteomics

To acquire data in bottom-up proteomics via MS/MS, the two common acquisition methods are data-dependent acquisition (DDA; Thermo), also referred to as information-dependent acquisition (IDA; SCIEX), and data-independent acquisition (DIA). Both approaches are schematically depicted in **Figure 6**. In DDA, in the first cycle of MS (survey scan), the most intense precursor ions (typically 15-20) are selected by predefined criteria. These are

subsequently fragmented and analyzed in the second stage of the MS/MS. The resulting data can be used to search existing databases. DDA stands out as a fast and simple setup, with straightforward data analysis and low demand for computational resources. One major limitation of DDA is its limited reproducibility, as the selection of the most intense precursors can vary between runs. Also, the loss of low-abundance ions further compromises the usage of this approach for quantification.[74]

DIA is an alternative to DDA, where all ions are fragmented rather than only the most intense ones. DIA is commonly applied with sequential window acquisition of all theoretical fragment ions (SWATH). Here, the scanning range is divided into a series of defined m/z-intervals (~25 Da), in which all ions are scanned with high speed (~100 ms) and subsequently fragmented.[74, 75] Compared to DDA, IDA is an unbiased, highly reproducible, and quantitative method that detects low-abundance analytes. The major drawback of DIA is that the data output requires advanced data analysis tools capable of handling large data sets, which involve bioinformatics tools and algorithms. Moreover, accurate data interpretation relies on spectral libraries, which may not always be available or comprehensive. Study-specific libraries must be generated in such cases, requiring more sample material, instrument time, and costs.[76]

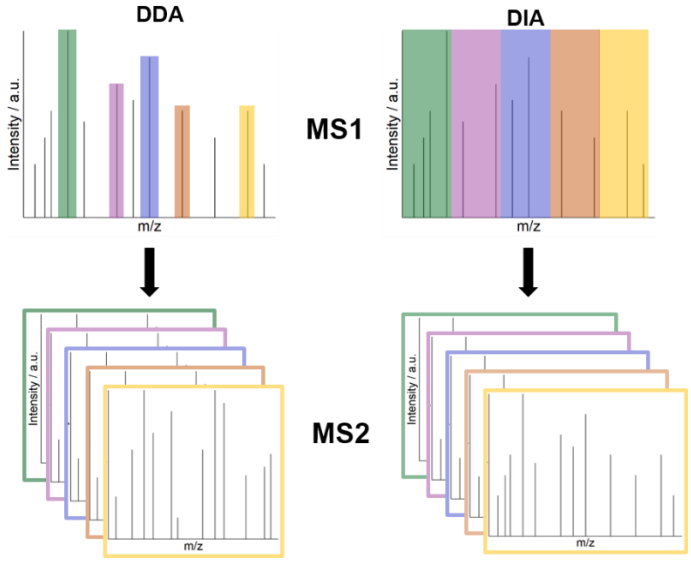


Figure 6: Schematic overview of a DDA-MS and DIA-MS approach. The figure is adapted from Krasny and Huang.[74]

2.3.1.2 Quantitative protein analysis by TMT

While both DIA and DDA have strengths and limitations in quantitative proteomics, alternative approaches, such as Tandem Mass Tag (TMT) labeling, offer a different method for quantification and the simultaneous analysis of various protein samples in one run (multiplexing). TMT labeling was first presented by Thompson *et al*. in 2003.[77] It revolutionized the field of proteomics, enabling fast and in-depth analysis of many up to 18 samples (TMTpro 18-plex) within one experiment.[4, 78] TMT labeling allows for studying complex biological systems, contributing to advancements in biomarker discovery[79-81], in the observation of protein dynamics[82, 83], post-translational modifications[84-86] and the investigation of drug target[87] and drug-protein interaction studies.[88]

The technique is based on mass tags assembled by three functional parts: an amine-reactive group, a mass normalizer, and a mass reporter, as seen in **Figure 7**. The chemical structure of each tag is identical, but the isotopic composition of the reporter and the mass normalizer group differs between the various tags due to the incorporation of stable isotopes. The mass normalizer balances the mass of the reporter group so that each TMT tag results in the same total mass (isobaric).[78]

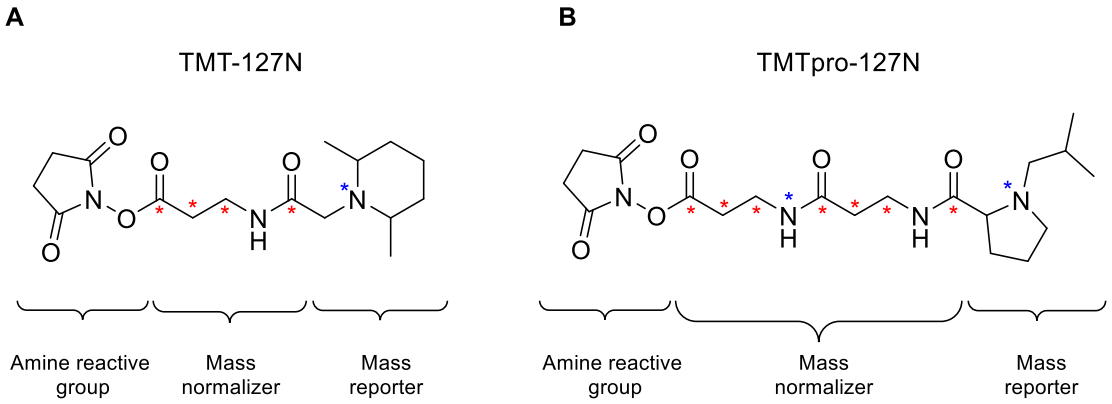


Figure 7: Chemical structure of a TMT tag (**A**) and TMTpro tag (**B**), both with a mass reporter group of 127 Da. The figure is adapted from Li *et al*.[89]

The principle of TMT tagging is as follows: The different TMT tags can be added to peptides from various samples, e.g., with different biological conditions. The reactive group of the tag covalently binds to the N-termini and ε-amino groups of lysines of the peptides, labeling them with the respective tag. After merging the different samples, LC-MS analysis is conducted. Peptides with identical amino sequences but different TMT tags coelute during LC separation

and show the same m/z value in the MS1 scan. The following fragmentation of those precursor ions generates specific reporter ions, enabling tag assignment and, with this, relative quantification across multiple samples. The mass of the reporter ions is also used for naming the tags (e.g., 127N or 128C), reflecting the specific isotopic composition of the reporter ions. 'C' denotes substitutions of ^{12}C with ^{13}C and 'N' indicates substitutions of ^{14}N with ^{15}N.[78]

Thompson *et al*. provided the first generation of TMT labeling with a set of two reagents.[29] This approach was improved by a more complex tag structure (s. **Figure 7 A)**, which enables the incorporation of five heavy isotopes (^{13}C or ^{15}N) in either the reporter or the balancer group. This provides six different TMT labels.[77, 81] Especially, the additional use of HCD in the MS3 on hybrid linear ion trap/orbitrap mass spectrometers provides higher fragmentation efficiency of the precursor ions, resulting in a more precise and reliable quantification than CID alone.[90-92] With these additional fragmentation techniques, it was possible to increase multiplexing by *neutron-encoding* and resolve the minimal mass difference (6.3 mDa) between heavy carbon (^{13}C) and heavy nitrogen (^{15}N) atoms, enabling the analysis of up to eleven different channels.[90, 93] In a 2019 study from Thompson *et al*., the linker region of the TMT tag was expanded, and the reporter ion was modified to a proline-based group (s. **Figure 7 B**).[94] This represents the latest generations of "TMTpro" labels, empowering the quantification of 16[85, 89, 94] (TMTpro 16-plex), or 18^4 (TMTpro 18-plex) different samples within one set.

2.3.2 Intact protein analysis/ top-down proteomics

In contrast to BUP, the proteins are not digested in top-down proteomics (TDP) but remain intact, mostly under denaturing conditions. This allows for complete sequence coverage of the protein, and further information about the proteoform, including PTMs, can be achieved. Due to this, TDP is often applied to better understand the protein functions and protein-protein interactions and to investigate specific PTMs potentially implicated in certain diseases and their development. Therefore, common approaches include biomarker discovery in clinical research, but single-cell analysis has also become more established in TDP.[66, 95-97] Compared to BUP, intact protein analysis, and TDP prove to be more challenging. Their limitation in sensitivity and high-throughput contributes to low proteome coverage. A common application for intact protein analysis is RPLC (s. chapter **2.2.1**) coupled to ESI-MS. RPLC serves to pre-separate the

proteins and reduce mass spectra complexity. Still, TD-spectra interpretation remains challenging and will be discussed in further detail in the next chapter.

2.3.2.1 Mass spectra interpretation in top-down proteomics

In contrast to Matrix-Assisted Laser Desorption/Ionization (MALDI), ESI leads on large molecules, such as proteins, to multiple charges. This is also due to the denatured and unfolded structure, which is more accessible for ionization. The corresponding spectra display a distribution of several charge states (charge state envelope). Each charge state comprises a series of signals from different isotopologues, as shown in **Figure 8.**

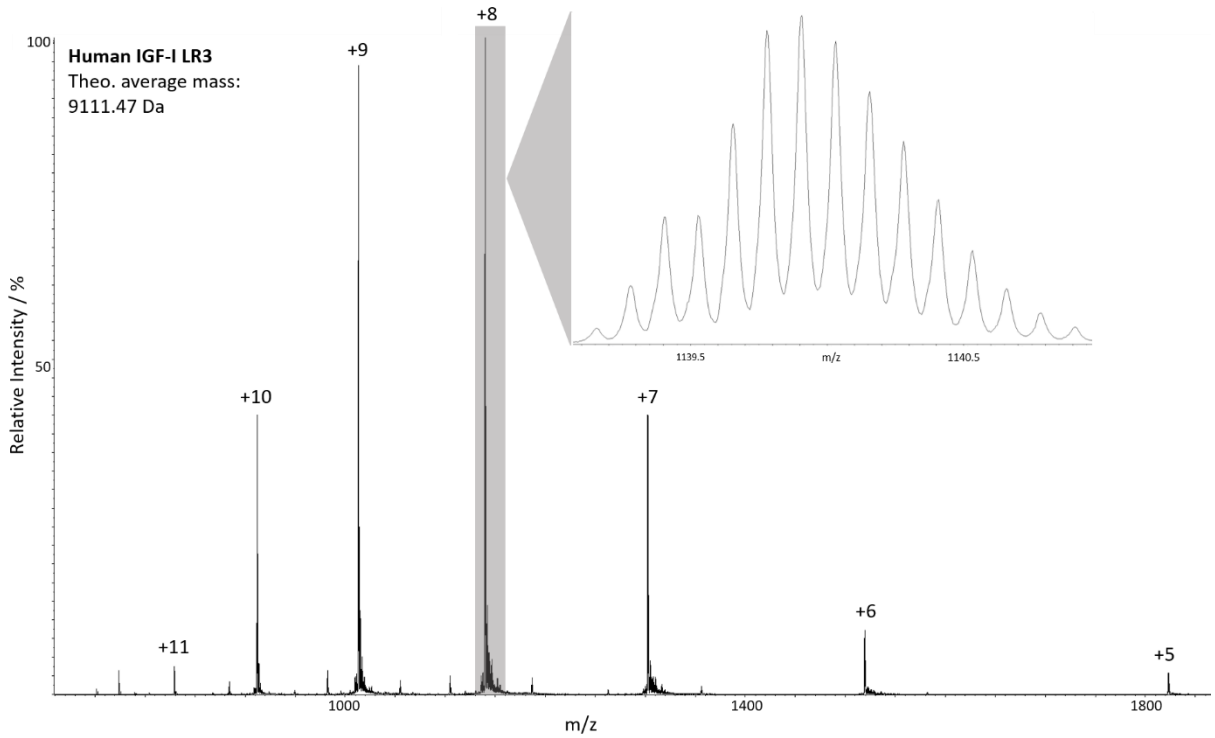


Figure 8: Mass spectrum of the protein Human IGF-I LR3 with its charge states indicated from +5 to +11.

Biomolecules mostly consist of hydrogen, carbon, nitrogen, oxygen, phosphorus, and sulfur. The isotopic distribution of biomolecules is mainly determined by the natural ratio of ^{12}C (98.89 %) to ^{13}C (~1.1 %). Molecules larger than 1000 Da show a significant contribution to the ^{13}C isotopes. For proteins heavier than approximately 15 kDa, the isotope distribution can be described by a Gaussian distribution, which can be seen in the mass spectrum of a 46-kDa protein in **Figure 9 B**. For determining the molecular weight of proteins, the differentiation between nominal, monoisotopic, average, and most abundant mass is, therefore, essential.[98-100]

The different mass types are schematically illustrated in the mass spectra of two different proteins in **Figure 9**.

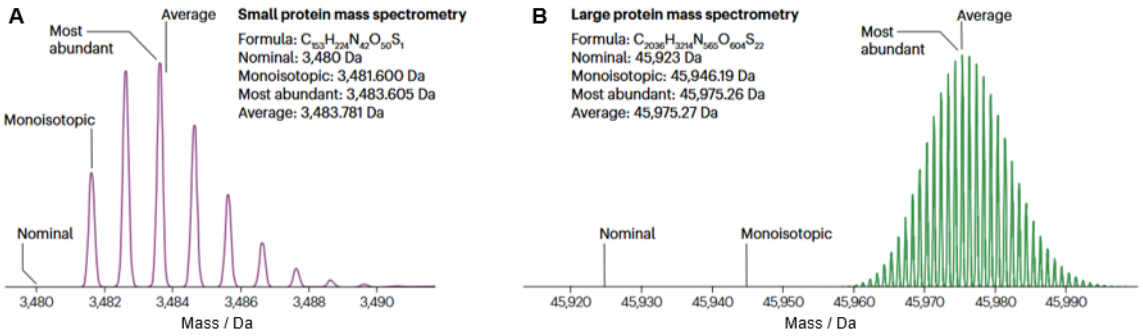


Figure 9: Mass spectra with the isotope distribution of a protein with a molecular weight of 3.5 kDa (A) and a protein with a mass of 46 kDa (B). Figure is reprinted and adapted with permission from Roberts *et al*.[66]

The nominal mass is defined by the integer mass of the most abundant isotope of the respective element without consideration of mass defects (e.g., H = 1 Da, C = 12 Da). The monoisotopic mass refers to the exact mass of the most abundant isotopes. The monoisotopic composition becomes less probable the larger a protein is. With a higher number of atoms, a larger number of combinations of atomic isotopes is possible. For small molecules, the monoisotopic mass is the most abundant, as the probability that all atoms are present in their most abundant isotopic form is high. Thus, the monoisotopic mass is displayed as a peak with high intensity (e.g., 3.5 kDa-protein, **Figure 9 A**) or the peak with the highest intensity (e.g., of peptides) in the mass spectra. In spectra of large proteins (e.g., 46 kDa protein, **Figure 9 B**), the most intense peak often no longer corresponds to the monoisotopic mass. Here, the most common isotopic distribution of the molecule, being displayed as the peak with the highest intensity in the isotopic cluster, reflects the most abundant mass. The peak of the monoisotopic mass of large proteins is of very low abundance or even non-detectable (s. **Figure 9 B**). Therefore, a protein's average or relative molecular mass (MR) is primarily used in databases and for the theoretical determination of the molecular weight. The sum of the average atomic masses calculates the MR.[98, 100-102]

This means that an individual protein is displayed by multiple signals. Not only multiple charge states but also PTMs (proteoform) and isotopic distribution make the direct interpretation of the mass spectra challenging. The average neutral mass has to be derived from the various m/z protein signals. This process is called mass deconvolution. Especially when isotopic resolution

is insufficient, but also due to data complexity and rising throughput in TDP, data deconvolution can be performed and supported by computational methods.

2.3.2.2 Data analysis in top-down proteomics

With the rapid advancements in TDP techniques, bioinformatic methods and software for protein analysis were also developed. By implementing peak assignment algorithms, probabilistic modeling, and machine learning techniques, the interpretation of intact and top-down protein spectra became easier and more accurate.[103, 104]

In the context of statistics and information theory, entropy, as introduced by Claude Shannon in 1948, is a measure of the uncertainty of a system under study.[105, 106] The entropy H is defined as:

$$H(\Omega) = -\sum_{i=1}^{n} P(\omega_i) \cdot log\ P(\omega_i) \qquad \text{(eq. 1)}$$

$\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ discrete sample space

$P(\omega_i)$ probability of outcome ω_i

This means that the maximum entropy represents the state with the highest uncertainty. The maximum entropy estimate is the least biased estimate, given the known constraints.[107] One of the first introduced deconvolution techniques for MS data is based on the maximum entropy algorithm. It was first applied in the early 1990s by Ferrige and Seddon as software (MaxEnt) for the analysis of protein spectra generated by ESI-MS.[108, 109] MaxEnt was the pioneer software with embedded maximum entropy deconvolution, and its technique is further presented here. MaxEnt generates multiple trial spectra and matches those with the experimental data. From the best fits, the spectra with the highest entropy are selected. The results are displayed as a probability distribution. The entropy indicates the distribution's width (error bars), reflecting the match's certainty.[108, 109]

MaxEnt approaches are robust and can improve S/N, enabling accurate molecular mass determination from complex spectra.[103, 109] Maximum entropy deconvolution is, for example, embedded in the MassHunter BioConfirm software (B.06.00 and newer) released by Agilent Inc. and the Bio Tool Kit plug-in for the PeakView software (V 2.1 and newer) from AB SCIEX.

Another algorithm based on probability theory is the Bayesian deconvolution. It was implemented by Marty *et al*. in 2015 in the open-source UniDec (Universal Deconvolution; https://unidec.chem.ox.ac.uk/) software as a deconvolution method for mass and ion mobility spectra.[110] The Bayesian data analysis can be divided into three parts: (1) an *a priori* assumption based on knowledge of the given data, (2) conditioning the model (from 1) to the observed data (posterior distribution), and (3) the evaluation of the fit.[111]

The Bayesian framework of Unidec is applied by iterating the above-mentioned three steps for separating the mass and charge components from an m/z spectrum. The most probable charge state is determined by using either the charge state distribution, mass distribution, or isotope distribution. The *a priori* assumption in UniDec is an equal probability for all charge states. In the first step, the most probable charge state distribution for each m/z data point in the spectrum (charge-vs-m/z-matrix) is refined. This is done by filtering the intensity of charge states depending on the intensity of neighboring charge states. In the second step, all charge assignments in the two-dimensional charge-vs-m/z-matrix are summed into a simulated one-dimensional m/z spectrum. The spectrum can be convolved with predefined peak shapes and widths. The simulated mass spectrum intensities are compared and adapted to the data in the third step. These three steps are repeated until the intensities of the simulated mass spectrum fit or align closely with the original data.[110, 112]

UniDec is widely used, especially in academic and industrial settings, mainly due to its open-source availability, user-friendly interface, and speed and flexibility. Novel pipeline developments enhance the further potential of UniDec for high-throughput data processing in, e.g., biotherapeutic and drug discovery research.[113]

2.4 Labeling techniques to study mycobacterial protein dynamics

After the genetic information of the deoxyribonucleic acid (DNA) has been transcribed into ribonucleic acid (RNA), it is translated into proteins. Consequently, proteins reflect all changes and factors that affect this process, including genetic and posttranscriptional modifications and environmental factors. The observation of changes in proteins and protein expression has always been of great interest to characterize and deepen the understanding of biological systems. One approach to observe changes in protein expression is stable isotope labeling by amino acids in a cell culture (SILAC), in which cells are cultured in two different media. One of those media contains amino acids with heavy isotopes. Lysine and arginine are most commonly used, labeled with six ^{13}C-atoms instead of ^{12}C. The bacterial cells incorporate the heavy amino acid, and newly synthesized proteins show a higher mass than the proteins of the unlabeled cell culture. For analysis, both cell cultures are merged, the proteins are digested, and labeled peptides are identified by their higher mass. Comparing the signal intensities of the unlabeled and labeled peptide homologs enables the quantification of the proteins and their comparison across different conditions.[114] SILAC has been widely used to answer various research questions. Some of them are analyzing proteomic alterations between biological samples, observing the dynamics of PTMs, biomarker discovery, and assessing proteome turnover on a proteome-wide scale.[114, 115] A modern method variant is pulsed SILAC (pSILAC), in which the cells are labeled with the heavy amino acids only for a distinct period. This allows a time-dependent analysis of protein dynamics and the possibility of analyzing protein synthesis and/or degradation for defined time frames.[116-118] One requirement for executing (p)SILAC is amino acid synthesis-deficient mutants. Mycobacteria, for comparison, can synthesize their own amino acids, and it is rather complex to induce auxotrophy in these bacteria.[119] So far, no SILAC experiment on mycobacteria has been described in the literature.

A new concept of using isotopic labeling to analyze *de novo* protein synthesis in mycobacteria was recently introduced by a study from Haisch *et al*.[3] In the study, fully supplemented medium with ^{13}C-labeled glycerol (G*) was used to grow mycobacteria. The bacteria can incorporate the heavy isotopes into newly synthesized proteins. The resulting higher mass and corresponding higher m/z-value of the proteins were observed by MALDI-TOF MS analysis.

After distinct incubation times of the bacteria (0.5 h–24 h), saturation of ^{13}C incorporation could be observed. To enable comparative and quantitative analysis of the protein signals over time, the center of gravity (COG) was used for data readout. For this, the COG values were normalized to the maximum shift (saturation point), expressed in percentage, and the maximum was calculated using the following, time-dependent, equation:

$$COG(t) = Y_0 + A \cdot e^{\frac{-t}{\tau}} \quad \text{(eq. 2)}$$

τ specific rise time

Y_0 maximum offset

A normalization factor of the maximum saturation to 100 %

This approach stands out as robust and highly reproducible, independently of the absolute intensity of the respective peaks. By comparing the differences between the COG values (ΔCOG), it is possible to monitor the protein turnover over time, exemplified by the protein integration host factor in **Figure 10**.

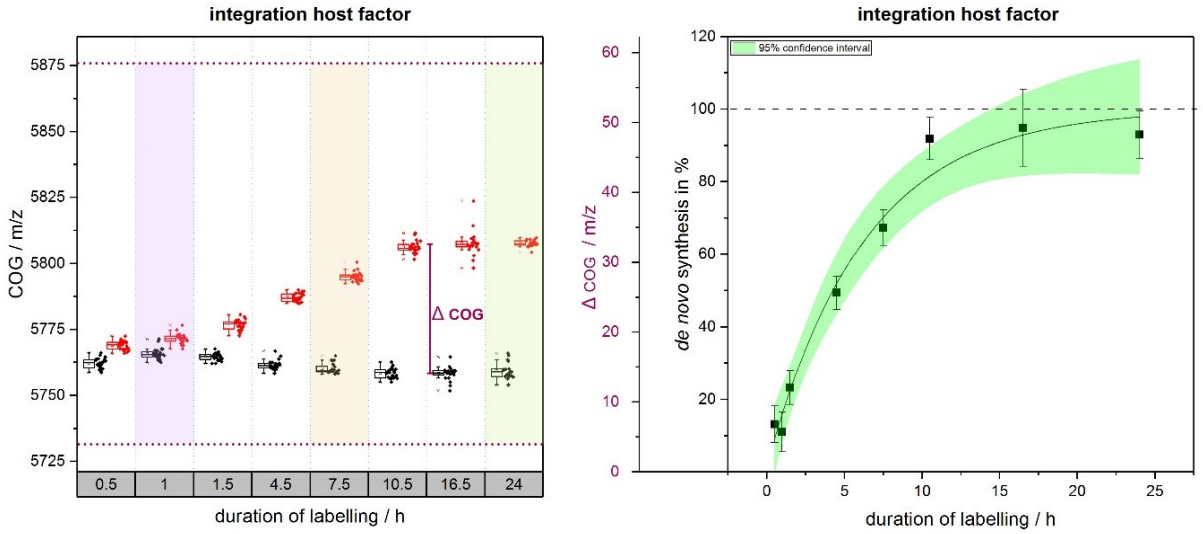


Figure 10: Results for individual assessment steps: *De novo* synthesis calculation by using the COG around the native protein mass (left) and normalized COG spectra (right). The figure is reprinted from Haisch *et al.*[3], under the terms of the CC BY 4.0 license.

The concept allows for tracking protein synthesis with high temporal resolution and observing changes within half the doubling cycle of the mycobacteria. It was possible to analyze the impact of four different drugs (rifampicin as a first-line drug, linezolid and bedaquiline as second-line drugs, and BTZ-043 as a drug candidate) on the *de novo* protein synthesis and

metabolism. In doing so, the authors have introduced a novel technique that allows rapid screening and testing of the initiation and scope of anti-mycobacterial drugs and drug candidates by using MALDI-TOF MS analysis.

Implementing MALDI-TOF MS aligns with the authors' initial idea of a fast and efficient screening tool. With minimal sample preparation and only small sample amounts, high-throughput measurement can be achieved, consistent with the use of an extraction protocol with a 96-well plate. One disadvantage of MALDI-TOF is its limitation as a one-dimensional analytical system with limited separation capacity. An example is the restricted differentiation of signals from molecules with similar molecular weights. Overlapping mass signatures may not be distinguished, leading to a loss of information. This is particularly important in complex biological systems with many similar molecules. As in the study, only seven proteins were analyzed as they were "easily accessible in the spectra and not overlaid by other proteins". This severely restricts the approach's potential, considering the mycobacterial cells' markedly higher protein amount.[13] The lack of multidimensional separation also constrains access to further information about the analytes, like hydrophobicity or polarity, or the construction of the molecules, such as isoforms or modifications.

Another challenge is the limited reproducibility of MALDI, which is often induced by matrix effects. Inhomogeneous matrix crystallization can lead to various ionization efficacies, resulting in signal variability. The consequence is a limited reproducibility of the energy deposition in the matrix and the energy transfer to the analyte ions. The ions experience different flight times in the TOF analyzer, leading to low mass accuracy. Further matrix effects are ion suppression and background noise, which result in limited sensitivity and reproducibility.[120] Also, as MALDI mainly generates singly charged ions, the m/z value directly reflects the molecular mass. With this, the analysis of molecules with high mass ($m/z > 25{,}000$) is often limited due to the restricted detection range of the mass analyzer and the lower detection efficiency of high-mass ions (e.g., with microchannel plate detectors).[120, 121]

2.5 Aim of the thesis

The aim of this thesis is to gain a deeper insight into the mechanisms of drug resistance in *M. tuberculosis*. Differences across various conditions, such as dormant and wild-type states of the model organism *M. Smegmatis*, are investigated. In addition, the effects of various drugs are examined. The onset of antibiotics on the mycobacterial metabolism is examined with varying concentrations over different time periods. Proteomic studies are performed via HPLC-MS. Before HPLC-MS analysis, protein extraction has to be optimized to achieve the highest possible yield. Additionally, various approaches are tested to render the protein samples suitable for HPLC-ESI-MS analysis.

The proteomic studies are divided into two major parts: the analysis of intact mycobacterial proteins and a quantitative, bottom-up approach by TMT labeling. In the first approach, intact proteome analysis is performed. So far, mainly bottom-up analysis has been examined to gain information about mycobacterial proteins. With the intact-mass approach, the presence of the proteins in their full-length constitution is ensured, and additional information about, e.g., PTMs can be yielded. Wild-type, as well as dormant mycobacteria induced by an acidification protocol, are examined. Additionally, the newly introduced stable isotopic labeling approach by Haisch *et al*.[3] is applied to observe the temporal protein dynamics of the mycobacteria. Building upon the group's prior experimental approach by the analysis via MALDI-TOF MS, the incorporation of heavy carbon isotopes into the proteins is observed via micro-flow RPLC-ESI-MS. Proteome coverage is improved by the second-dimensional separation (RPLC), and higher resolution is yielded by implementing a Q-TOF mass spectrometer. The mycobacterial proteomic changes are observed with treatment with the first-line drug rifampicin. Mycobacteria are treated either with a 1-fold or a 20-fold minimum inhibitory concentration (MIC) of rifampicin for five different time frames. Mass spectra evaluation is performed by manual deconvolution with the built-in deconvolution function of the PeakView software plug-in Bio Tool Kit. In addition, another approach for automatic mass spectra deconvolution over the entire chromatogram is introduced, which was developed in cooperation with a bioinformatics team of the Helmholtz Center Munich.

The second approach was examined at Oregon State University using a bottom-up proteomic approach with TMT-based quantification. Changes in mycobacterial proteins, when exposed to

elevated drug concentrations exceeding the respective MIC, are investigated over time. For this, wild-type, as well as dormant *M. smegmatis*, are treated either without drug treatment as a control or with 1-fold, 5-fold, or 20-fold MIC of the first-line drug rifampicin and the drug candidate BTZ-043. Drug treatment is administered over three distinct time periods (wild-type: 4.5 h, 10.5 h, or 24 h; dormant: 1 d, 4 d, 10 d). Mycobacterial peptides are labeled with TMT 16-plex, resulting in seven sets of 16 TMT labels. The analysis is performed with a nano UPLC and an Orbitrap MS.

3 Materials and Methods

3.1 Chemicals and materials

3.1.1 Chemicals and solvents

The following chemicals and solvents were obtained from commercial sources and used without further purification unless otherwise stated.

Table 1: List of chemicals and solvents.

Product	Company	Location
Acetone, LC-MS-Grade	Carl Roth	Karlsruhe, Germany
Acetonitrile for LC-MS	Supelco	Bellefonte, Pennsylvania
Acrylamide	Serva	Heidelberg, Germany
Brilliant Blue R250	Sigma-Aldrich	Taufkirchen, Germany
Bromphenol blue	Merck	Darmstadt, Germany
BTZ-043	HAPILA GmBH	Gera, Germany
Chloroform, p.a.	Merck	Darmstadt, Germany
Dimethyl sulfoxide, ≥ 99.9 %	Merck	Darmstadt, Germany
DL-Dithiothreitol, ≥ 99.0 %	Sigma-Aldrich	Taufkirchen, Germany
Dulbecco's Phosphate Buffered Saline	Sigma-Aldrich	Taufkirchen, Germany
Ethylenediamine tetraacetic acid disodium salt dihydrate, ≥ 99 %	Carl Roth	Karlsruhe, Germany
Formic Acid, LC-MS-Grade	Carl Roth	Karlsruhe, Germany
Glycerol-1,3-$^{13}C_2$	Sigma-Aldrich	Taufkirchen, Germany
Glycerol	Carl Roth	Karlsruhe, Germany

Product	**Company**	**Location**
Hydrochloric acid, 1 M	Sigma-Aldrich	Taufkirchen, Germany
Methanol, HPLC	Carl Roth	Karlsruhe, Germany
M7H9 Broth base	Sigma-Aldrich	Taufkirchen, Germany
Myoglobin, 95-100 %	Sigma-Aldrich	Taufkirchen, Germany
OADC Enrichment	Becton Dickinson and Company	Franklin Lakes, New Jersey
Phosphate-buffered saline (PBS)	Sigma-Aldrich	Taufkirchen, Germany
Pierce™ Intact Protein Mix Standard Mix	Thermo Fisher Scientific	Waltham, Massachusetts
Rifampicin	Sigma-Aldrich	Taufkirchen, Germany
SDS, ultra-pure	Carl Roth	Karlsruhe, Germany
N,N,N',N'-Tetramethylethylendiamine (TEMED), p.a.	Biomol GmbH	Hamburg, Germany
Trifluoroacetic acid for HPLC	Sigma-Aldrich	Taufkirchen, Germany
Thiourea, ≥ 99.0 %	Sigma-Aldrich	Taufkirchen, Germany
TRIS, ≥ 99.9 %	Carl Roth	Karlsruhe, Germany
TWEEN 80	Sigma-Aldrich	Taufkirchen, Germany
Tuning Kit	AB SCIEX	Framingham, Massachusetts
Urea, ≥ 99.6 %	Carl Roth	Karlsruhe, Germany
Water, Ultra LC-MS grade	Carl Roth	Karlsruhe, Germany

3.1.2 Other materials

Protein samples were handled in 1.5 mL and 2.0 mL Protein Lo-Bind Tubes (Eppendorf SE, Hamburg, Germany). Filtering of proteins was tested in Pall Nanosep centrifugal devices with an Omega membrane and a molecular weight cut-off of 3 kDa (Pall Corporation, Port Washington, New York). Protein fractionation was performed by using a T-piece (1/32", 0.25 mm drilling) from Vici Valco (Houston, Texas).

3.1.3 Instruments

- Bandelin Electronic Sonoplus HD 2070 Homogenisator (Bandelin electronic GmbH & Co. KG, Berlin, Germany)
- Thermomixer comfort (Eppendorf SE)
- Invitrogen Qubit 4 Fluorometer (Thermo Fisher Scientific, Waltham, Massachusetts)

The optical density of the bacterial culture at 600 nm (OD_{600}) was measured using an Ultrospec 3100 pro UV/Visible spectrophotometer (Amersham Biosciences, Little Chalfont, United Kingdom).

3.2 Experiments and methods for the study of mycobacteria

All experiments were conducted with the wild-type *M. smegmatis* strain MC^2 155 (ATCC 700084).

3.2.1 MIC determination

The MIC is the lowest concentration that completely inhibits bacterial growth. The determination of the MIC of the antimicrobial agents used in this study was performed in our laboratory as described previously, following the established broth dilution protocols and methodologies.[3, 122]

In short, mycobacterial cultures were cultivated under agitation (140 rpm) overnight at 37 °C with various drug concentrations. Subsequently, the samples were plated in serial dilution on Columbia agar (5 % sheep blood) and incubated at 37 °C for 3 days. To determine the bacterial

reduction factors, the CFU on the plates was counted. The log reduction of wild-type mycobacteria after treatment was calculated as follows:

$$Reduction\ factor = log_{10}(N_0) - log_{10}(N_t) \quad \text{(eq. 3)}$$

N_0 number of wild-type mycobacteria before treatment

N_t number of wild-type mycobacteria after treatment

Bacterial growth is assessed by measuring the optical density at λ = 600 nm (OD_{600}) of the bacterial solution with a spectrophotometer.

3.2.2 Growth of *M. smegmatis* under different conditions

Growth of *M. smegmatis*

Reference wild-type cultures were cultivated overnight (37 °C, 140 rpm) in M7H9 media supplemented with oleic albumin dextrose catalase (OADC) and 0.05 % Tween-80. Frozen glycerol stocks of *M. smegmatis* (-80 °C, Middlebrook 7H9, OADC, 0.05 % Tween-80, 20 % glycerol) were thawed and incubated overnight in a sterilized M7H9 medium, supplemented with OADC and 0.05 % Tween-80 (37 °C, 140 rpm, 15 h). This overnight culture was used for labeling and antibiotic treatment.

Labeling with 1,3-^{13}C labelled glycerol (G*)

The incubation of *M. smegmatis* cultures with 1,3-^{13}C-labeled glycerol was done as described previously.[3] In short, labeling experiments were performed using a starting OD_{600} of 0.1 in fresh cultivation medium (M7H9 medium supplemented with OADC) to which either unlabeled glycerol or glycerol-1,3-$^{13}C_2$ was added to a final concentration of 2.44 µL glycerol per mL medium. Labeling was performed over several different time intervals (1.5 h, 3 h, 4.5 h, 6 h, 7.5 h, 10.5 h, and 24 h).

Antibiotic treatment

Rifampicin solution was dissolved in dimethyl sulfoxide (DMSO) and vortexed thoroughly to ensure complete solubilization. Treatments were carried out at different concentrations corresponding to 1×, 5×, or 20× MIC with final concentrations of 842 ng/mL (1× MIC), 4.21 µg/mL (5× MIC), and 16.8 µg/mL (20× MIC), respectively. BTZ-043 was prepared from

a 5 µg/mL stock solution. Final treatment concentrations were adjusted to 4.73 ng/mL (1× MIC), 23.6 ng/mL (5× MIC), and 94.6 ng/mL (20× MIC). Bacterial cultures were treated with the respective antibiotics for defined time intervals, depending on the experimental setup. Cultures were incubated under agitation (37 °C, 140 rpm) in glass tubes positioned at a 45° angle. To ensure sufficient biomass for downstream extraction, at least two samples (total volume: 9 mL) were pooled. All experiments were performed in at least two biological replicates and analyzed in technical duplicates by HPLC-MS.

3.2.3 Growth of dormant and resuscitated mycobacteria

Dormancy was induced into *M. smegmatis* by acidification[123], as described by previous work from our group.[15] In short, *M. smegmatis* were cultivated (37 °C, 140 rpm) for 24 h in a nutrient broth medium containing 8.0 g of nutrient broth per liter. The next day, an aliquot of 1 mL of this culture was transferred to 100 mL of Sauton's medium, which contained per liter: 0.18 g K_2HPO_4; 0.166 g $MgSO_4 \cdot 7 \cdot H_2O$; 0.056 g NaH_2PO_4; 1.33 g L-asparagine; 20 mL glycerol; 0.017 g ferric ammonium citrate; 0.66 g citric acid; 0.833 g polysorbate 80. The pH was adjusted to 6.0 using 1 M HCl, and the final media was supplemented with 0.05 % Tween-80. Dormant bacteria were incubated for 20 days (37 °C, 140 rpm). Resuscitation of dormant *M. smegmatis* cultures was performed by centrifuging the dormant culture (10,000 g, 10 min, 7 °C), and the bacterial pellet was washed with PBS to remove residual media. Bacteria were resuspended in new M7H9 media (OADC, 0.05 % Tween-80) and incubated for 24 h (37 °C, 140 rpm).

3.2.4 Protein extraction

After incubation, the bacteria were separated from the medium by centrifugation (10,000 g, 10 min, 4 °C), transferred to 1.5 mL Eppendorf tubes, and washed twice with PBS buffer (10,000 g, 5 min, 7 °C). The pellets were resuspended in lysis buffer (~1:3 v/v, 25 mM TrisCl, 1 mM EDTA, 2.5 mM DTT, 7 M urea, 2 M thiourea, pH = 7.5). Subsequently, bacteria were lysed by beating in a 3:1 ratio of solution to beads (0.1 mm glass, PowerBead Tubes, Qiagen GmbH, Hilden) on the FastPrep-24™ 5G bead beating grinder and lysis system (M.P. Biomedicals, LLC, Santa Ana, California) with an installed *M. tuberculosis* program (2·45 s, 240 s break, 6 m/sec). Following the mechanical and chemical treatment of the bacteria, the

samples were centrifuged (15,000 g, 2 min, 7 °C). The supernatant containing the proteins was transferred into 1.5 mL Protein LoBind Eppendorf tubes, and proteins were precipitated with acetone in a 1:4 ratio (-20 °C) overnight. Afterward, the solutions were centrifuged (7,000 g, 5 min, 4 °C), the supernatant was removed, and the protein pellets were washed with ice-cold acetone. Resolubilization of the proteins was performed according to a protocol of Doucette *et al*.[42] For this, proteins were incubated with ice-cold 80 %-FA (2 min, -20 °C) and then diluted to the final concentration with LC-MS grade water and centrifuged to ensure clear solutions for follow-up measurements (15,000 g, 10 min, 7 °C). In cases where immediate analysis was not possible, samples were preserved at –80 °C until further use. Protein concentration was determined by using the Invitrogen Qubit 4 Fluorometer (Thermo Fisher Scientific) and the accompanying Invitrogen reagents according to the manufacturer's protocol.

3.2.5 Sample preparation for LC-MS analysis of intact proteins

For LC-MS analysis of the intact proteins, samples were thawed on ice, and solvents were evaporated under a vacuum (Concentrator plus, Eppendorf SE). Subsequently, proteins were re-dissolved in solvents matching the initial solvent composition of the HPLC gradient (s. chapter **3.3.2**; 90 % H_2O, 10 % acetonitrile (ACN), 0.1 % FA) and shaken for at least 30 minutes at 550 rpm at 7 °C to ensure complete solvation. Possible undissolved residues were separated by centrifugation at 13,000 g for 10 minutes at 7 °C.

3.2.6 Sample preparation for bottom-up analysis

Extracted proteins were dried under vacuum, and subsequently digested with the Preomics iST Kit (PreOmics GmbH, Planegg, Germany) according to the manufacturer's protocol.

3.2.7 Protein identification by Helmholtz Munich Core Facility

Protein identification of the fractions (s. chapter **4.3.4**) was carried out by the Core Facility Metabolomics and Proteomics (CF-MPC) at Helmholtz Munich – German Research Center for Environmental Health, under the direction of Dr. Stefanie Hauck and Dr. Christine von Törne.

For sample preparation, trichloroacetic acid was added to a final concentration of 25% (v/v), and samples were incubated for 10 minutes at 4 °C. Proteins were then pelleted by centrifugation at 14,000 g for 5 minutes at room temperature. The supernatants were discarded, and the pellets were washed three times with ice-cold acetone. Subsequently, pellets were resuspended in 50 mM ammonium bicarbonate buffer, sonicated, and incubated for 10 minutes at room temperature with gentle agitation. Reduction was performed using 20 mM dithiothreitol (DTT) for 30 minutes at 60 °C. After cooling to room temperature, alkylation was carried out by incubating the samples with 40 mM iodoacetamide (IAA) for 30 minutes in the dark. Excess IAA was quenched with 40 mM DTT. Proteins were digested sequentially with 0.1 µg Lys-C for 2 hours at room temperature, followed by 1 µg trypsin overnight at 37 °C. Peptide clean-up was performed as described by Kulak *et al*.[124]

Mass spectrometry data were acquired in DIA mode using a Q Exactive HF-X mass spectrometer (Thermo Fisher Scientific Inc.). Equal amounts of peptides were automatically loaded onto an RSLC system (Ultimate 3000, Thermo Fisher Scientific Inc.) coupled online to the mass spectrometer. Peptide enrichment was achieved using a nano-trap column (300 µm ID × 5 mm packed with Acclaim PepMap100 C_{18}, 5 µm, 100 Å; LC Packings, Sunnyvale, CA, USA), followed by separation via reversed-phase chromatography on an analytical column (Acquity UPLC M-Class HSS T3, 75 µm ID × 250 mm, 1.8 µm; Waters, Eschborn, Germany) maintained at 40 °C. Peptides were eluted at a flow rate of 250 nL/min using a linear gradient of 3 % to 40 % ACN in 0.1 % FA over 90 minutes. The DIA method included a full MS survey scan from m/z 300 to 1,500 at a resolution of 120,000, with an automatic gain control (AGC) target of 3×10^6 and a maximum injection time of 120 ms. Fragmentation was performed using higher-energy collisional dissociation (HCD) with a normalized collision energy of 28. Precursor ions were isolated in 37 variable windows covering the m/z range from 300 to 1,650 at a resolution of 30,000, with an AGC target of 3×10^6 and automatic injection time. All spectra were acquired in profile mode.

3.2.8 Digestion, TMT labeling, and analysis

Sample preparation (digestion and TMT labeling) was conducted in the laboratory of Dr. Luiz Bermudez (Department of Microbiology) at Oregon State University (OSU).

Digestion

Frozen protein pellets were thawed on ice and digested following the in-solution digestion protocol from the ProteaseMAX Surfactant Trypsin Enhancer protocol (Promega, Madison, Wisconsin).

TMT labeling

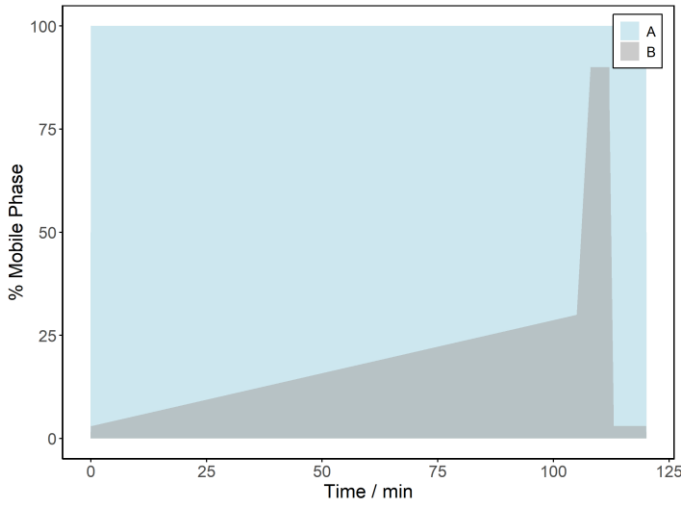
TMT labeling was performed with the TMTpro 16plex Isobaric Label Reagents and Kit from Thermo Fisher Scientific after the manufacturer's protocol.

Sample acquisition and data processing

All samples were analyzed in collaboration with the OSU - Mass Spectrometry Facility under the supervision of Dr. Claudia Maier. Stanislau Stanisheuski conducted the measurements.
Samples were analyzed using a Waters M-class UPLC coupled to a Thermo Orbitrap Fusion Lumos mass spectrometer. 1 µL of the sample was loaded onto a Symmetry C_{18} trapping column (100 Å, 180 µm × 20 mm, 5 µm; Waters, Milford, Massachusetts) for 3 min at 3 % eluent B with a flow rate of 5 µL/min. Separation was performed on an ethylene bridged hybrid (BEH) C_{18} analytical column (130 Å, 75 × 15 cm, 1.7 µm; Waters, Milford, Massachusetts) with a constant flow rate of 300 nL/min. Eluent A was H_2O with 0.1 % FA, and eluent B was ACN with 0.1 % FA. The active gradient is depicted in **Figure 11**. The electrospray voltage was set to 2300 V. The temperature of the ion transfer tube was 300 °C, and the S-lens RF level was set to 30 %. The precursor scan was performed from 400 to 1600 m/z at 120,000 resolving power with the automated gain control set to 400,000. The most intense precursor ions with charges 2 to 7, exceeding the intensity threshold of 25,000, were isolated for the MS2 analysis using a 0.7 Th isolation window. These ions were accumulated until AGC 100,000 was reached or at most 118 ms before being fragmented via HCD at a collision energy of 35 %. The cycle time was set to 3 s. The fragments were analyzed at 60,000 resolving power, with the first mass value in the scan equaling 100 Th. Dynamic exclusion was used for 60 s with a mass tolerance

of 10 ppm. Raw data were processed using Proteome Discoverer[125] (3.0) with the Chimerys search engine.

Further data acquisition was performed with the support of Cheng-Wei Liao and Dr. Ines Assum (Technical University of Munich; Helmholtz Center Munich) in the framework of the DynamicKit subproject of the bayresq.net project.



Time / min	% A	% B
0	97	3
105	70	30
108	10	90
112	10	90
113	97	3
120	97	3

Figure 11: HPLC gradient for TMT analysis with eluent A: H_2O, 0.1 % FA, and eluent B: ACN, 0.1 % FA.

3.2.9 SDS-PAGE

All materials referenced in this chapter are from Bio-Rad Laboratories (Hercules, California) unless otherwise indicated or listed before (s. chapter **3.1**). The exact composition of the separation and collection gel can be seen in **Table 2**. The protein quality was checked by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). 10 µg of protein extract (in 12 µL 25 mM TrisCl, 1 mM EDTA buffer) was mixed with 4 µL 4xLaemmli sample buffer and incubated for 5 min at 95 °C. Samples and protein standards (PageRule, prestained protein ladder, 10-180 kDa; Thermo Fisher Scientific) were loaded on a self-cast 12 % SDS-PAGE gel and run for 65 min at 200 V. Proteins were visualized by staining with coomassie blue (0.5 % (w/v) Coomassie Brilliant Blue R250, 50 % (v/v) methanol, 7 % (v/v) acetic acid) for 1 h and rinsed with ultrapure water for 30 min.

Table 2: SDS-gel composition for four gel preparations.

Component	**Separation gel (12 % acrylamide)**	**Collection gel**
Ultra-pure water	3.4 mL	3 mL
1.5 M Tris (pH = 8.8)	5.6 mL	-
0.5 M Tris (pH = 6.8)	-	1.5 mL
Acrylamide	6 mL	1.5 mL
SDS solution (10 %)	150 µL	50 µL
APS (10 %)	112.5 µL	35 µL
TEMED	9.45 µL	7 µL

3.3 HPLC-MS analysis

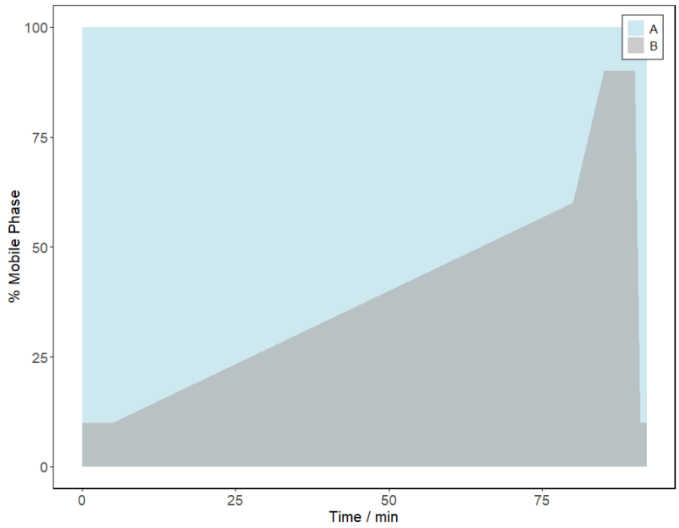
The entire HPLC-MS setup was supplied by AB Sciex. The protein samples were analyzed using liquid chromatography (Eksigent Ekspert NanoLC 425) coupled to a TripleTOF 5600+ with an Eksigent 25 Micron ESI Electrode.

3.3.1 Equilibration, tuning, and mass calibration of the system

Before each measurement series, the column oven was set to the specified temperature, and the HPLC column was equilibrated under the initial conditions for at least 10 column volumes until stable pressure was achieved. Tuning of the mass spectrometer was conducted as necessary, depending on the observed performance (resolution and sensitivity). For this, the tuning solution (AB Sciex) was injected via the integrated syringe pump system with a 1 mL syringe (Hamilton, Reno, Nevada), and the instrument optimization, as provided by the control software, was performed. An additional mass calibration was performed using the PPG positive Mix (AB Sciex) to achieve a more precise mass calibration of masses within the region of interest.

3.3.2 LC-MS parameters for intact mass analysis

A Triart Bio C_4 column (30 nm, S-1.9 µm, 150 x 0.3 mm ID, 1/32", YMC Europe GmbH, Dinslaken, Germany) was utilized to separate intact proteins. A C_4 capillary guard (YMC-Triart Bio C4, 1/32" capillary column, 30 nm, S-1.9 µm, 5x 0.3 mm ID) was implemented to prevent significant contamination of the chromatography column. The HPLC gradient for intact protein separation can be seen in **Figure 12**. The HPLC system was equilibrated by flushing the initial conditions for 3 minutes before and 2 minutes after each measurement.



Time / min	% A	% B
0	90	10
5	90	10
80	40	60
85	10	90
90	10	90
91	90	10
92	90	10

Figure 12: HPLC gradient for intact protein analysis with eluent A: H_2O, 0.1 % FA, and eluent B: ACN, 0.1 % FA.

Parameters for MS analysis are listed in **Table 3**. The measurements were controlled using the manufacturer's software Analyst TF 1.8, provided by AB Sciex.

Table 3: MS parameters for intact protein analysis.

Parameter	
Ion Source Gas 1	10 psi
Ion Source Gas 2	2 psi
Curtain Gas	25 psi
Temperature	300 °C
Ion Spray Voltage Floating	5000 V
Declustering Potential	150 V
Collision Energy	10 V
Accumulation Time	0.25 s

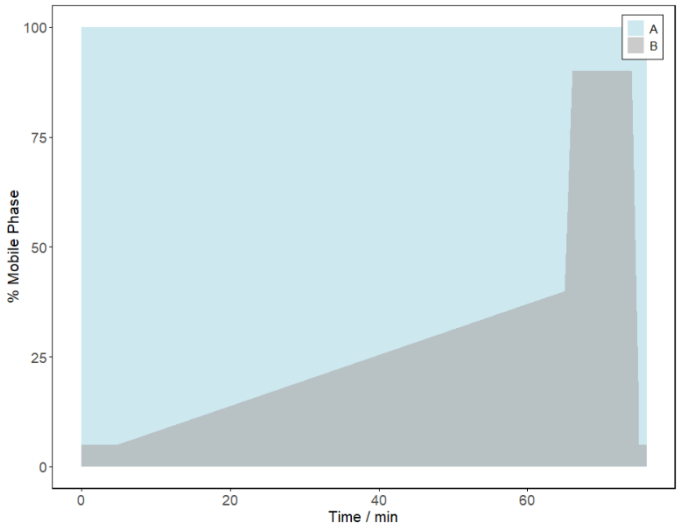
3.3.3 LC-MS parameters for bottom-up analysis

Measurements to analyze the digested proteins were conducted in positive mode as information-dependent acquisition (IDA). Precursor ions were included with charges from 2 to 5 and a minimum intensity threshold of 100 cps. The maximum number of candidate ions to monitor per cycle was set to 20. Previously targeted ions were excluded within a 30-second exclusion window, and dynamic background subtraction was allowed. Mass tolerance was 100 ppm. Further experimental parameters of the QTOF are listed in **Table 4**.

Table 4: IDA parameters for peptide analysis.

Parameter	MS1	MS2
Ion Source Gas 1	10 psi	10 psi
Ion Source Gas 2	15 psi	15 psi
Curtain Gas	25 psi	25 psi
Temperature	300 °C	300 °C
Ion Spray Voltage Floating	5500 V	5500 V
Declustering Potential	60 V	60 V
Collision Energy	10 V	35 V
Accumulation Time	50 ms	10 ms
Mass Range	350 - 1300 m/z	100 - 1500 m/z

The HPLC gradient for peptide analysis can be seen **Figure 13**. Before and after each measurement, the HPLC system was flushed with the initial flow rate conditions for 2 and 5 minutes, respectively.



Time / min	% A	% B
0	95	5
5	95	5
65	60	40
66	10	90
74	10	90
75	95	5
76	95	5

Figure 13: HPLC gradient for bottom-up analysis with eluent A: H_2O, 0.1 % FA, and eluent B: ACN, 0.1 % FA.

3.4 Data processing

3.4.1 Bottom-up data analysis

Bottom-up data analysis was carried out by Dr. Ines Assum. Network analysis with the TMT data was additionally supported by Cheng-Wei Liao. For all analyses, the *M. smegmatis* reference proteome (strain ATCC 700084 / mc²155, UP000000757) was used, obtained in FASTA format from the UniProt database.

Protein identification for intact mass assignments

Peptide data were analyzed with MaxQuant[126] (2.1.4.0). As group-specific parameters, trypsin and Lys-C were set for the digest type. Protein group output was analyzed using R[127] (4.2.0), and calculation of the MW of the proteins was done by using the Protein Molecular Weight Calculator from Altogen Labs.[128]

Data processing for the network analysis of the TMT data

Protein quantifications from Proteome Discoverer were further analyzed in R. A total of 2174 proteins were identified. For further analysis, only proteins with non-zero intensity in at least two technical replicates per three biological replicates under all nine experimental conditions were kept. All proteins that were missing more than 40 % in all analyzed samples were removed,

resulting in 669 remaining proteins. Missing values were preliminarily imputed with K-nearest-neighbor imputation (K = 20, DEP[129], 1.18.0 impute) for the following batch correction using ComBat_seq (SVA[130], 3.44) with a negative binomial regression. Quality control (QC) samples, the group conditions treatment, time, dose, and dormancy status were used to adjust for SET batch effects commonly known in TMT-labeling experiments. After batch correction, QC samples were removed from further analysis, and all data were $\log_2$ transformed. Missing values from the first step were re-imputed with K-nearest-neighbor imputation (K=20, DEP, 1.18.0, impute), and variance stabilization normalization (DEP, 1.18.0, normalize_vsn) was applied, followed by differential expression analysis between treated and untreated samples for each condition, i.e., timepoint, drug, and MIC using the test_diff function (DEP, 1.18.0). Significance was reached with an adjusted p-value of 0.05 (Benjamini-Hochberg) and an absolute $\log_2$ fold change of 1. All significant proteins across all conditions were merged, and the UniProt IDs annotated with the STRING database[131] (STRINGdb, 2.8.4) using all available connections (rbioapi, rba_string_interactions_network, 0.8.1) and associated biological processes (rbioapi, rba_string_annotation, 0.8.1) were grouped for simplified representation (s. **Table 5**). Network plots were generated with the packages (igraph[132], 2.0.3; ggraph[133], 2.2.1), using for each protein the condition with the maximum absolute fold change for visualization. All untreated dormant samples were compared to the untreated viable samples using the time point in hours as a covariate.

Table 5: Grouping of biological processes for the network analysis.

General metabolic processes	Metabolic process
	Cellular metabolic process
	Primary metabolic process
	Organic substance metabolic process
	Small-molecule metabolic process
Catabolic processes	Catabolic process
	Cellular catabolic process
	Small-molecule catabolic process
	Organic substance catabolic process
	Carbohydrate catabolic process
	Cellular carbohydrate catabolic process
	Alcohol catabolic process
	Polyol catabolic process
	Alditol catabolic process
	Glycerol catabolic process
	Aromatic compound catabolic process
	Hydrogen peroxide catabolic process
	Organic hydroxy compound catabolic process
Biosynthetic processes	Biosynthetic process
	Cellular biosynthetic process
	Small-molecule biosynthetic process
	Organic substance biosynthetic process
	Vitamin biosynthetic process
	Water-soluble vitamin biosynthetic process
	Cobalamin biosynthetic process
	Heterocycle biosynthetic process
	Aromatic compound biosynthetic process
	Tetrapyrrole biosynthetic process
	Cellular nitrogen compound biosynthetic process
	Organic cyclic compound biosynthetic process
	Organonitrogen compound biosynthetic process
Carbohydrate and alcohol metabolism	Carbohydrate metabolic process
	Cellular carbohydrate metabolic process
	Alcohol metabolic process
	Glycerol metabolic process
	Glycerol-3-phosphate metabolic process
	Alditol metabolic process
	Alditol phosphate metabolic process
	Polyol metabolic process
	Organic hydroxy compound metabolic process
	Carbon utilization
Acid and organic compound metabolism	Organic acid metabolic process
	Carboxylic acid metabolic process

	Oxoacid metabolic process
	Dicarboxylic acid metabolic process
	Chorismate metabolic process
Nitrogen and phosphorus metabolism	Nitrogen compound metabolic process
	Cellular nitrogen compound metabolic process
	Phosphorus metabolic process
	Phosphate-containing compound metabolic process
	Organophosphate metabolic process
	Phosphorylation
	Organonitrogen compound metabolic process
Cyclic and aromatic compound metabolism	Cellular aromatic compound metabolic process
	Aromatic compound biosynthetic process
	Aromatic compound catabolic process
	Heterocycle metabolic process
	Heterocycle biosynthetic process
	Organic cyclic compound metabolic process
	Organic cyclic compound biosynthetic process
	Tetrapyrrole metabolic process
	Tetrapyrrole biosynthetic process
Vitamin and coenzyme metabolism	Vitamin metabolic process
	Water-soluble vitamin metabolic process
	Vitamin biosynthetic process
	Water-soluble vitamin biosynthetic process
	Cobalamin metabolic process
	Cobalamin biosynthetic process
Stress response and detoxification	Response to stimulus
	Cellular response to stimulus
	Response to stress
	Response to oxidative stress
	Response to chemicals
	Cellular response to chemical stimulus
	Response to toxic substance
	Cellular response to toxic substance
	Detoxification
	Cellular detoxification
	Cellular oxidant detoxification
	Reactive oxygen species metabolic process
	Hydrogen peroxide metabolic process
	Hydrogen peroxide catabolic process
	Response to abiotic stimulus
	Response to hypoxia
	Response to decreased oxygen levels
	Response to oxygen levels

Energy and respiration processes	Generation of precursor metabolites and energy Energy derivation by oxidation of organic compounds Anaerobic respiration Electron transport chain Respiratory electron transport chain Cellular respiration
Transport and localization	Transport Transmembrane transport Localization Establishment of localization
Miscellaneous processes	Cellular process Carbohydrate derivative metabolic process

3.4.2 Intact protein analysis

LC-MS data of intact proteins were analyzed with PeakView software (2.1, AB Sciex). The built-in "Reconstruct Protein" function from Bio Tool Kit (1.0) was used for manual deconvolution of the protein signals. Predefined mass ranges for COG determination are listed in **Table 6**. A mass range for deconvolution was predetermined for each protein. For the calculation of the COG of the six proteins, deconvoluted mass ranges were individually adapted to each protein, respectively.

Table 6: Mass ranges for deconvolution and COG determination.

Protein mass / Da	Deconvolution range / Da	Mass Area Start / Da	Mass Area End /Da
7215	7000-7400	7210	7380
10974	10000-12000	10970	11100
12868	11000-13500	12800	13000
10642	10000-11000	10640	10735
9401	9000-10000	9400	9490
6877	6000-7500	6870	6970

COG analysis and curve fitting

Mass spectrometry data were processed in R (4.3.3) with the use of the packages minpack.lm[134], dplyr[135] and tidyr[136]. For each protein, a COG was calculated at each time point based on the intensity-weighted average of detected isotope peaks. Prior to this, a baseline of 5000 units was

subtracted from all area values, and resulting negative values were excluded. COGs were determined for each combination of condition, replicate, time point, and protein. Differential COG (ΔCOG) was calculated by subtracting the median COG of G from the median COG of G*. Uncertainty was estimated by non-parametric bootstrapping with replacement (n = 1000), resulting in a 95 % confidence interval. ΔCOG data from the untreated control ("no antibiotic") was fitted using a mono-exponential growth function as described by Haisch *et al*. ($COG(t) = Y_0 + A \times \exp(-t / \tau)$) and normalized to 100 % (maximum shift) to model the temporal incorporation of labeled carbon atoms. The uncertainty of the fitted model was also assessed by bootstrapping (n = 1000), using the empirical standard deviation of ΔCOG to simulate resampled input data. For antibiotic-treated conditions (1-fold MIC rifampicin, 20-fold MIC rifampicin), no fitting was performed; ΔCOG shifts were normalized based on the control fit. All visualizations were generated using ggplot2[137] and the package RColorBrewer.[138]

External data processing

External data processing was performed by Dr. Ines Assum as follows: Wiff/wiff.scan files were converted to mzML using ProteoWizard MSConvert[139] (3.0.22101-a968175) with centroiding (Peak Picking with vendor presets). MzML files were then processed in R 4.2.0 using the xcms package[140-142] (3.18.0), performing peak detection with the MatchedFilterParam routine. Resulting peaks were clustered by retention time with the dbscan clustering in the fpc package (2.2-10; radius 1, minimal points per cluster 3). Per cluster, the scan matching the median retention time of peaks closest was selected as the representative spectrum for this region of interest (ROI). Each such ROI spectrum was then deconvoluted in UniDec[110] (6.0.3), resulting in intact masses with relative quantifications (int_quant) and quality-of-fit scoring (DScore). Peak calling included in UniDec was used to determine the discrete masses that were present (mass range = 100 Da-500 kDa, mass bins = 1 Da, peak window = 1 Da, peak thresh = 0.1). Quantification thresholds between 0.1 and 0.3 were used throughout various analyses to prefilter masses based on intensity. To compare mass shifts resulting from isotopic labelling, data across samples were filtered to retain consistent retention time and intact mass ranges. For each sample, a single weighted intact mass was calculated by summing the products of individual intact masses and their corresponding intensities, then dividing by the total intensity.

4 Results and Discussion

4.1 Optimization of the LC-MS system for intact mass analysis

4.1.1 Optimization of the MS parameters

The five parameters Ion Source Gas 1 (GS1) and Ion Source Gas 2 (GS2), the Curtain Gas (CUR), Ion Spray Voltage Floating (ISFV), and the Temperature (TEM) affect various aspects of the ion source operation, like the ionization process, spray stability, and ion transfer. The parameters depend on the flow rate and the samples to be analyzed. They are described in more detail in the manufacturer's manual.[143] In short, GS1 and GS2 are responsible for the sample's nebulization/dispersion (GS1) and its drying (GS2) in the ESI. They depend, therefore, mainly on the used flow rate, the solvents, and the applied temperature. The CUR flushes the MS inlet with nitrogen gas. Refining the CUR value is essential, as too high CUR values would lead to a reduced ion transfer to the MS, whereas too low values may lead to the introduction of solvent vapors and contaminations into the system. The TEM is the temperature of the heating gas at the ion source, influencing the desolvation efficiency and the signal sensitivity. ISVF is the voltage applied to the ESI capillary, initiating the electrospray ionization and the affecting signal intensity. The installed ESI electrode is constructed for flow rates of 5-20 µL/min. A flow rate of 5 µL/min was chosen for the following experiments to analyze within the specified manufacturer's recommended parameter range, but also to yield high sensitivity with the minimal flow rate. Therefore, parameters were adapted to this specified flow rate. For optimizing the default MS parameters for intact protein analysis, a standard protein solution (0.1 µg/µL) was injected into the MS. Myoglobin was chosen as a reference protein, which, with a molecular weight of 17 kDa, falls within the size range of mycobacterial proteins. The parameters were ramped over time within the specified ranges, using steps of 10 psi for GS1, GS2, and CUR, 100 °C for TEM, and 500 V for ISVF. The default and optimized parameter values are listed in **Table 7**.

Table 7: Optimized and default MS parameters for intact protein analysis.

Parameter	Optimized	Default
Ion Source Gas 1	10 psi	18 psi
Ion Source Gas 2	15 psi	33 psi
Curtain Gas	25 psi	35 psi
Temperature	300 °C	180 °C
Ion Spray Voltage Floating	5000 V	3500 V

Individual parameter differences were not tested for statistical significance, as a combined effect of all parameters is expected and is addressed with further analysis. Therefore, a protein standard mix consisting of six different proteins was measured three times with the specific (default or optimized) parameters. The resulting signal intensities of the six different protein peaks can be seen in **Figure 14**. The comparison of the measurements with the different instrument settings shows a difference in signal intensity of two orders of magnitude, indicating a superior measurement performance with the new parameter settings. To appropriately visualize this difference, a logarithmic (log_{10}) scale was applied to the signal intensities in **Figure 14**.

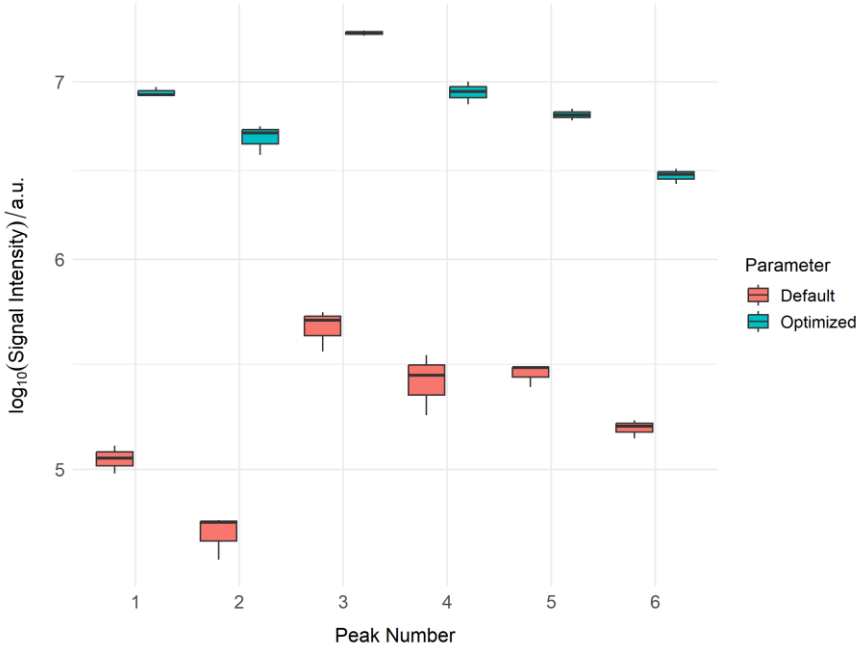


Figure 14: Boxplots showing the logarithmic signal intensities of six proteins (peak numbers 1-6) acquired using default and optimized MS parameters.

4.1.2 Evaluation of the HPLC conditions

Comparison of a C_4- and a C_8-RP column

For protein separation at the target flow rate of 5 µL/min, two RP columns with an ID of 0.3 mm were evaluated: a C_4 column with 300 Å pores and a C_8 column with 120 Å pores. Both columns are designed for biopharmaceutical separation. However, further information for the separation of proteins with a molecular weight (MW) range up to 150 kDa is only given for the C_4 column.[144] No such specification was found for the C_8 column.

The columns were tested for protein separation of the protein standard mixture, consisting of six different proteins ranging from 9 kDa to 67 kDa, which are listed in **Table 8**.

Table 8: Proteins of the Intact Protein Standard Mix (Thermo Fisher) with the corresponding theoretical (theo.) average and monoisotopic masses and their chromatographic elution order specific to the subsequent experiment.

Elution order	Protein name	Theo. average mass / Da	Theo. monoisotopic mass / Da
3	Human IGF-I LR3	9111.47	9105.34872
4	Human Thioredoxin	11865.52	11858.04393
1	*Streptococcus dysgalactiae* Protein G	21442.61	21429.75915
5	Bovine Carbonic Anhydrase II	28981.29	28963.6881
2	*Streptococcus* Protein AG	50459.74	50429.84641
6	*Escherichia coli* Exo Klenow	68001.15	67959.42515

The results can be visualized either as a total ion chromatogram (TIC)**,** which displays the total intensity over time (s. **Figure 15**, upper pane) or as a LC-MS contour plot, which provides more information by incorporating the mass-to-charge ratio (m/z) as a second dimension (s. **Figure 15**, middle and bottom pane). While the TIC offers a simplified overview of ion signal distribution, the contour plot allows for recognizing individual analytes based on their m/z and retention time. Proteins can typically be recognized based on their distinct m/z and RT characteristics. Since intact proteins are often detected as a series of charge states, they appear as a distribution of m/z values at a given retention time (here: $RT_{range} \approx 0.5$-3 min).

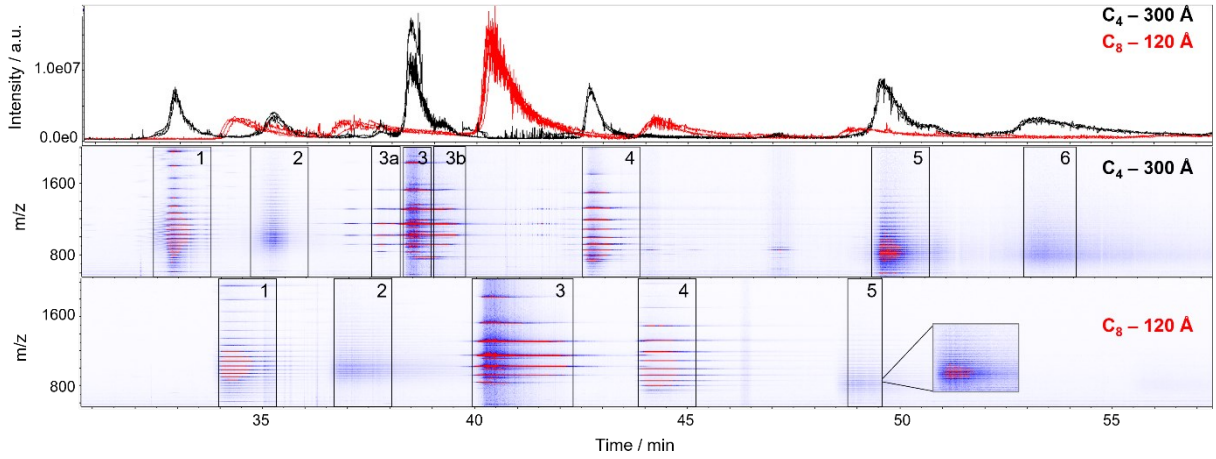


Figure 15: Upper pane: TICs of the protein standard separated on a C_4 (black) or C_8 (red) HPLC column. Measurements were conducted in triplicate. The middle and bottom panes show the respective LC-MS contour plots with protein signals highlighted in numbered boxes.

The protein mixture eluted over approximately 24 min on the C_4 column and over 16 min on the C_8 column, respectively, with the same elution order on both columns. However, only five of six proteins were separated with the C_8 column, with the biggest protein (*Escherichia coli (E. coli)* Exo Klenow, 68 kDa) absent in the chromatogram. This may be attributed to the pore size of the stationary phase. Traditional reversed-phase HPLC columns with pore sizes around 120 Å are unusual for analyzing large molecules, as these may not diffuse sufficiently into the small pores or may be completely excluded. Columns with larger pore sizes, such as 300 Å or bigger, are more appropriate for separating large proteins, as they allow the molecules better accessibility into the stationary phase, resulting in improved retention and peak shape. Additionally, it can be seen that the protein Human IGF-I LR3 (peak 3) was separated into three peaks (**Figure 15**, peak 3a, 3, and 3b), indicating the presence of possible isoforms or PTMs. The PTMs were confirmed by the corresponding deconvoluted mass spectra (16 Da, oxidation). In summary, the C_4 column shows a better performance for protein separation, especially for proteins covering a broad range of molecular weights. Moreover, the 300 Å-C_4 column provides more information about structural characteristics of intact proteins than the 120 Å-C_8 column.

Evaluation of different column temperatures on the protein separation efficiency

The influence of the column temperature was evaluated to optimize the efficiency of protein separation. For this, the protein standard mixture was analyzed in three technical replicates at each of the following temperatures: 30 °C, 40 °C, or 50 °C, which is the maximum temperature of the installed column oven. Due to the fact that the spectra exhibit signal fluctuations, likely

due to a short accumulation time (250 ms) and electrospray instability, Gaussian smoothing was applied to the spectra to facilitate peak analysis. The smoothing width was set to 15 points, ensuring that the original peak shapes remain largely unaffected (s. **Figure 16**, red graphs). Peak shoulders (for example, peak 3a, 3b) were not considered for peak comparison, as they result from modifications and were observed at each condition similarly.

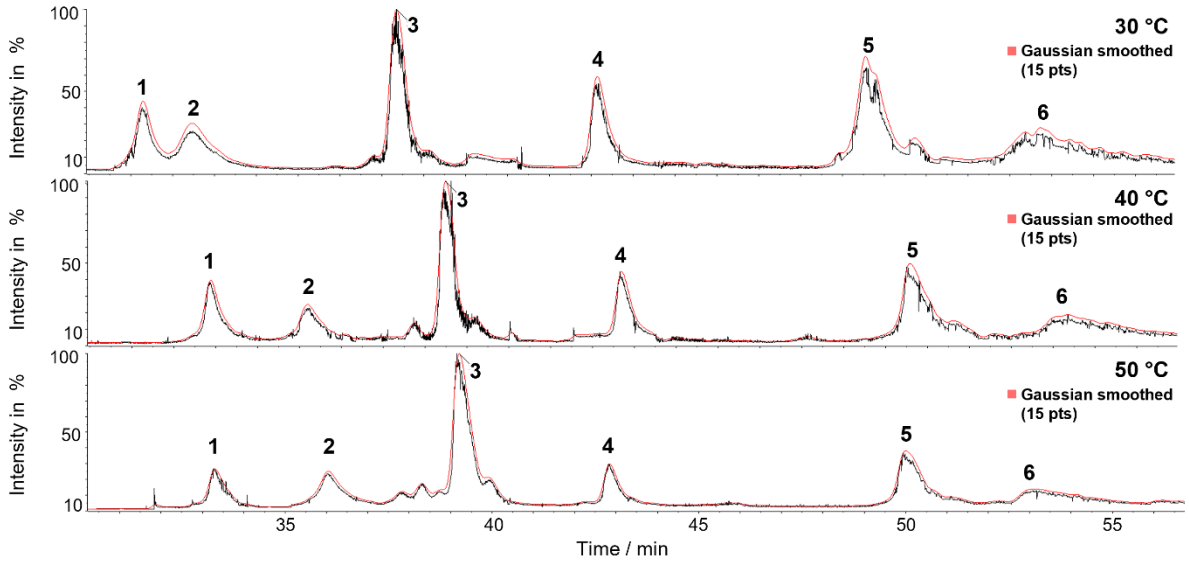


Figure 16: TICs of the protein standard analyzed at 30 °C, 40 °C, or 50 °C, respectively. The ion signals were Gaussian smoothed by 15 points and are indicated in red.

The six proteins eluted within approximately 26 minutes, while protein separation at 30 °C took about 1.5 min longer (from ~ 30.5-56 min) than at 40 °C (from ~ 32-56 min) or at 50 °C (from ~ 32.5-56 min). This may be due to stronger and longer adsorption of the proteins to the stationary phase at the lower temperature, while interaction of the proteins with the stationary phase may be weaker at higher temperatures (40 °C, 50 °C). Though protein elution at 30 °C required the longest retention time, peaks 1 and 2 overlap and could not be distinctly separated at this temperature. In contrast, peak separation is observed at 40 °C and 50 °C, respectively.

To evaluate the chromatographic performance, the six specific peaks were compared by their full width at half maximum (FWHM) and signal intensities (**Figure 17**). No significant difference between these parameters was found when operating at the different separation temperatures. As no major qualitative difference was observed between 40 °C and 50 °C, 40 °C was selected for further analysis in order to preserve the column material and reduce the risk of protein degradation.

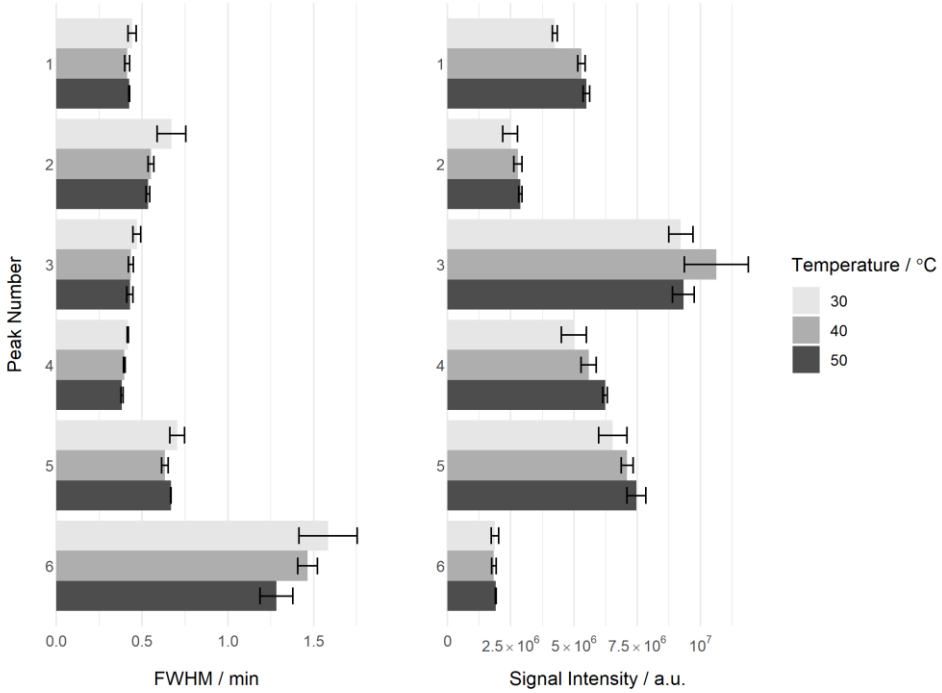


Figure 17: FWHM values and signal intensities of the six proteins (peak numbers 1-6) at either 30 °C, 40 °C, or 50 °C. Measurements were performed in triplicate.

4.1.3 Evaluation of the analytical performance

The protein standard mixture was analyzed six times daily (intra-day) on four different days (inter-day) to evaluate the performance of the LC-MS system's reliability and consistency and to assess the accuracy of the mass deconvolution (by PeakView software). Protein masses were deconvoluted with the Bio Tool Kit. To assess the mass accuracy of the analysis, the deviations of the deconvoluted masses from the theoretical average masses were calculated and are presented in the following **Figure 18**.

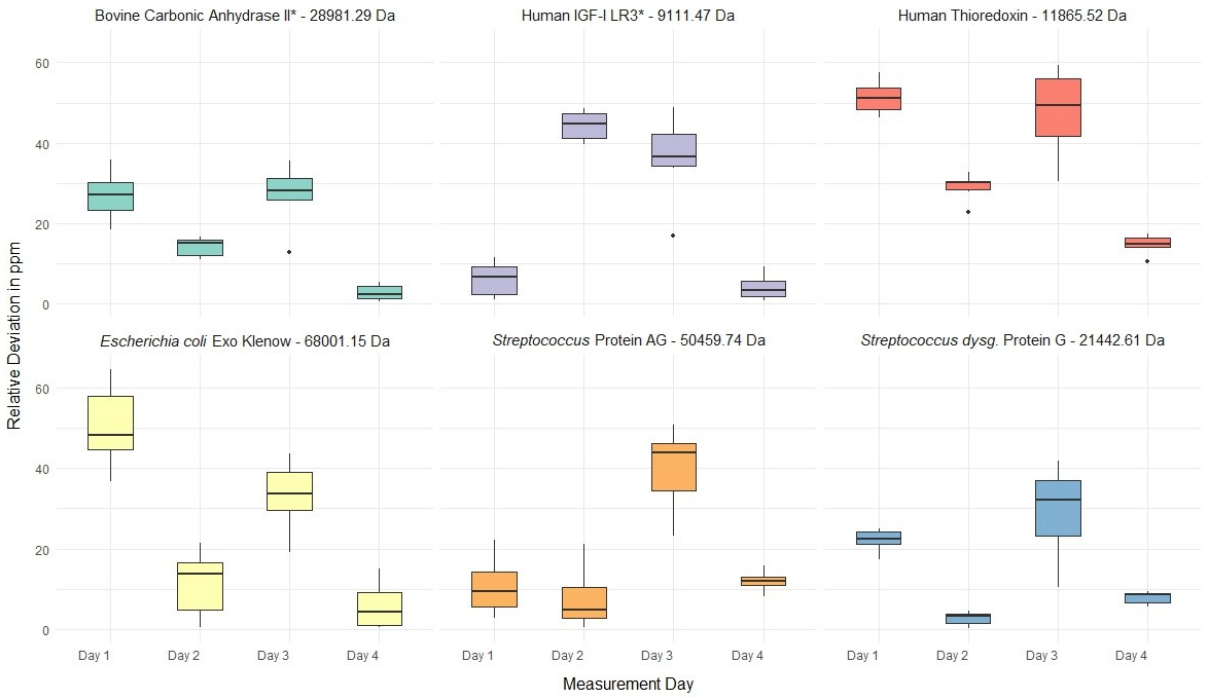


Figure 18: Relative mass deviations in parts-per-million (ppm) from the theoretical mass of six proteins (Pierce Intact Protein Standard Mix). Measurements were performed over four different days with six technical replicates per day.

The mass deviations range from 0.3 ppm (*Streptococcus dysgalactiae* Protein G, day 2) to a maximum of 65 ppm (*E. coli* Exo Klenow, day 1). The deconvoluted masses from *E. coli* Exo Klenow showed the highest standard deviation (SD) over all four measurement days (SD > 5.9 ppm), which may be attributed to the instrument's limited resolving power and its upper mass detection limit of 40 kDa, as mentioned before. ANOVA analysis was performed to test whether the mass deviations are significantly different within (intraday) and across (interday) the measurement days. For the intraday analysis, mass deviations between the six different proteins on each individual day were compared. They showed highly significant

($p < 0.001$, days 1, 2, and 4) and significant ($p < 0.05$, day 3) differences. The interday analysis was done by comparing relative deviations between the four different measurement days. The analysis showed highly significant ($p < 0.001$) variations, indicating day-to-day variability. For PTM identification, the limited mass accuracy could be critical. All mass deviations remained below a maximum of 65 ppm, which is a mass difference of a few Daltons (4.4 Da, *E. coli* Exo Klenow). PTMs often exhibit more than 4 Da, for example, formylation (28 Da), methylation (14 Da), and oxidation (16 Da), but, for example, deamidation (0.98 Da) would possibly not be identified. Although the results show statistically significant differences, the observed variations remain within an acceptable range for the intended application of observing relative mass shifts. If the mass accuracy is adequate for protein identification has to be further evaluated.

4.2 Optimization of the mycobacterial protein extraction and preparation for LC-MS analysis

Numerous methodologies were evaluated to investigate the influence of different lysis conditions on mycobacterial protein extraction and downstream LC-MS sample preparation. They are illustrated in **Figure 19**. For protein extraction, chemical and mechanical lysis approaches were explored. Following a protocol of Rabodoarivelo *et al.*[16], different lysis buffers combined with mechanical lysis methods were tested. Mechanical lysis was conducted either with probe sonication or with bead beating with varying bead-to-lysis buffer ratios. This protocol was compared to lysis approaches involving the use of 100 % TFA (after a protocol of Abele *et al.*[145]) or acetone precipitation (after a protocol of Neumann *et al.*[15]). For LC-MS analysis, protein purification has to be examined to remove the lysis buffer and possible impurities. For this, acetone and methanol-chloroform precipitation, and filtering with centrifugal filters of a molecular weight cut-off (MWCO) of 3000 Da were tested. Additionally, the protein resolubilization was optimized to enhance protein recovery.

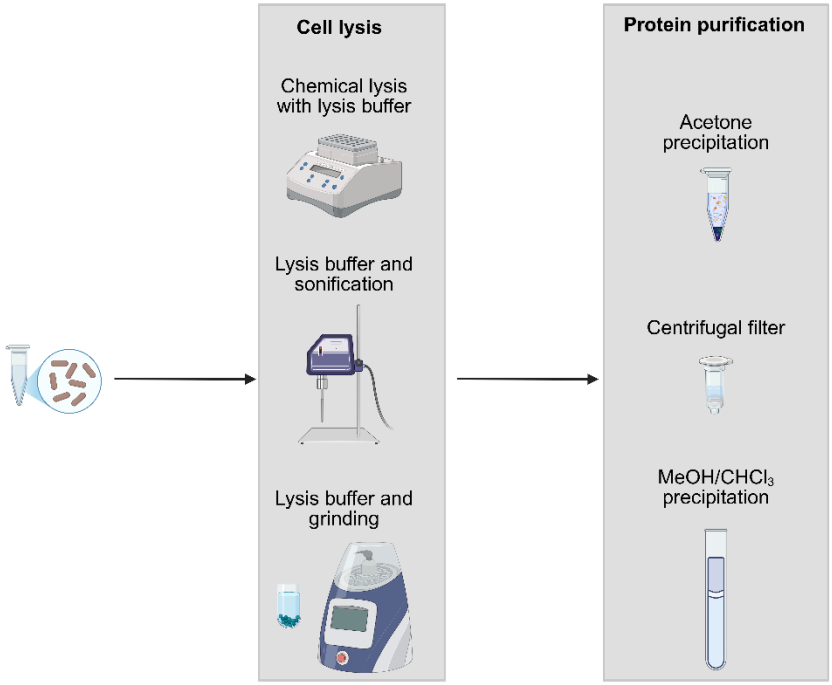


Figure 19: Schematic overview of the tested cell lysis methods and protein purification approaches. The figure was created with BioRender.com.

4.2.1 Testing of different protein extractions

To test the effect of different lysis conditions on the mycobacterial protein extraction, the methodologies utilized in the study by Rabodoarivelo *et al*.[16] were tested. Repetition of the experiments was performed to verify the robustness of the findings, ensuring that the results are not attributable to random chance or specific conditions.

In the study, four lysis buffers (Lb) were evaluated: an initial lysis buffer (Lbi; 50 mM NH_4HCO_3, pH = 7.4, 10 mM $MgCl_2$, 0.1 % NaN_3, 1 mM EGTA, protease inhibitor), Lbi supplemented with thiourea and urea (LbthUr), LbthUr with added dithiothreitol (DTT) (LbthUr-DTT), and LbthUr-DTT supplemented with sodium dodecyl sulfate (SDS). Additionally, various mechanical lysis methods were tested: sonication for 5 or 15 minutes, and/or bead beating, employed with zirconia beads with different bead-to-buffer volume ratios (1:1, 1:2, and 1:3).

Chaotropic agents like urea and thiourea disrupt hydrogen bonds within proteins and weaken hydrophobic interactions between them. This process promotes protein unfolding and denaturing, thereby enhancing their solubilization.[16] Similarly, DTT influences the denaturing and increases the solubility of proteins by reducing disulfide bonds. SDS works as a detergent, for example, by releasing membrane proteins from the membrane and solubilizing them.
Building upon the protocol, two modifications were introduced. First, a modified initial lysis buffer (LbA; 25 mM TrisCl, 1 mM EDTA, pH = 7.5) was employed. Second, DTT was added in lower concentrations (2.5 mM instead of 5 mM) and tested solely in combination with the initial lysis buffer (LbB). The composition of the lysis buffers can be seen in **Table 9**.

Table 9: Tested lysis buffers and their composition with minor modifications (printed in bold) based on a protocol of Rabodoarivelo *et al*.[16]

Lysis buffer	Composition
A	**25 mM TrisCl, 1 mM EDTA**
B	**25 mM TrisCl, 1 mM EDTA, 2.5 mM DTT**
C	25 mM TrisCl, 1 mM EDTA, 2 M thiourea, 7 M urea
D	25 mM TrisCl, 1 mM EDTA, **2.5 mM** DTT, 2 M thiourea, 7 M urea
E	25 mM TrisCl, 1 mM EDTA, 2.5 mM DTT, 2 M thiourea, 7 M urea, 1% SDS

Cells from an overnight *M. smegmatis* culture were centrifuged and subsequently washed with PBS. Aliquots from the cell-PBS suspension were used for lysis using the different lysis buffers A-E (s. **Table 9**). Lysis with the respective buffer was tested in triplicate using bead beating (1:4, beads:buffer, v/v), and protein concentration was measured using the Qubit protein assay. For quality control, extracted proteins were separated on a 12 % sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). The gel image and the resulting protein concentrations are presented in the following **Figure 20**.

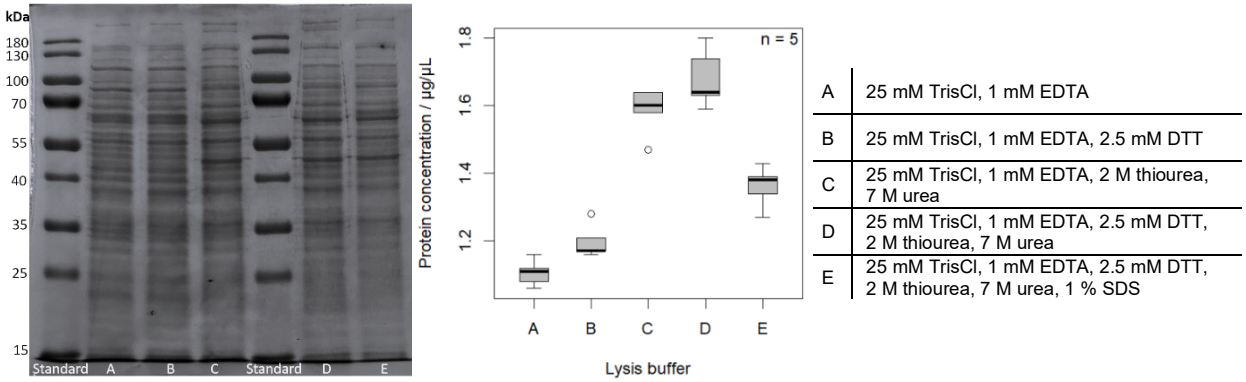


Figure 20: **Left**: SDS-PAGE of the proteins extracted with different lysis buffers (A-E). **Middle**: Boxplot of five technical replicates of the respective protein concentrations. Outliers are shown as individual points, falling outside the range defined by 1.5 times the interquartile range below the first or above the third quartile. **Right**: Table with the composition of lysis buffers A-E.

In the SDS-PAGE gel, no significant qualitative differences could be seen among the proteins from the different lysis buffers (s. **Figure 20**, left). The determined concentrations showed highly significant ($p < 0.001$) differences between the groups (ANOVA). The highest concentrations were yielded with the initial lysis buffer containing (thio-)urea (LbC, $c = 1.59 \pm 0.07$ µg/µL), as well as with the buffer containing (thio-)urea and DTT (LbD, $c = 1.68 \pm 0.09$ µg/µL). In comparison, in the study from Rabodoarivelo *et al.,* the buffer with

(thio-)urea and DTT (LbD) did not result in higher protein yields than the control buffer (LbA), which differs from the results of this experiment. The difference in concentrations between LbC and LbD showed no statistical significance, while being significantly higher than those of the other buffers A, B, and E (Tukey HSD). The lysis buffer containing (thio-)urea and DTT (LbD) was chosen for further experiments.

4.2.2 Testing of different mechanical lysis methods

Bead beating (BB) or sonication can be used as additional approaches to disrupt the cell wall mechanically. In contrast to the approach of Rabodoarivelo *et al.*, sonication was tested only for two runs of 45 s (5 cycles, 60 % amplitude) to have a better comparison to the bead beating approach (2·45 s). Between each cycle (sonication/BB), the samples were put on ice for 240 s to prevent protein degradation due to excessive heating. BB was tested with a proportion of beads:buffer equal to 1:2, 1:3, or 1:4. **Figure 21** shows the yielded protein concentrations.

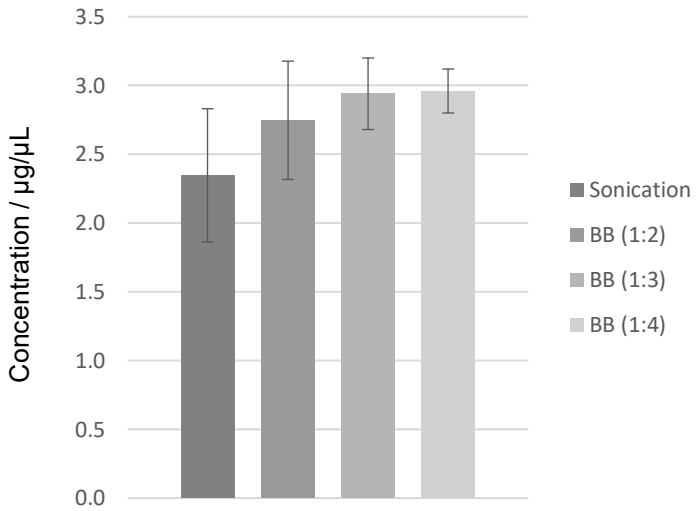


Figure 21: Protein concentrations resulting from *M. smegmatis* cells lysed with either probe sonication or bead beating with various bead:buffer ratios (1:2, 1:3, 1:4). Experiments were conducted in triplicate.

Sonication (2.35 ± 0.48 µg/µL) did not yield higher concentrations than BB (2.75 ± 0.43 µg/µL to 2.94 ± 0.26 µg/µL). Additionally, it is worth noting that the usage of probe sonication is considered critical due to the generation of aerosols that may pose a risk to the user in an open environment. To ensure the protocols remain suitable for potential future applications with *M. tuberculosis*, a closed-system sonication device would have been needed. However, due to the higher protein concentrations, BB is preferred as a further lysis method. The ratios of

beads:buffer did not differ significantly between the different approaches. BB with a ratio of beads:buffer of 1:3 was chosen as a further extraction method, as it yields a remaining sample volume that is optimal for downstream processing. Protein precipitation steps (e.g., acetone precipitation at a 1:4 ratio) can subsequently be performed in standard 1.5 mL microcentrifuge tubes.

4.2.3 Comparison with different lysis protocols

Based on the protocol of Neumann *et al.*[15], mycobacterial protein extraction was achieved by acetone precipitation for 5 min. After removal of acetone, extracted proteins were resolubilized with 50 % ACN, 35 % FA, and 15 % H_2O (v/v/v). Another extraction protocol was published by Abele *et al.*[145], presenting a unified workflow protocol for the in-depth analysis of bacterial proteomes, among others, mycobacteria. They lysed cells by incubation with 100 % trifluoroacetic acid (TFA) for 5 minutes at 55 °C and subsequently neutralized with 2 M TRIS. Both lysis methods were tested in comparison to the previously described lysis method (LbD and BB). The resulting protein concentrations are depicted in the following **Figure 22**. Lysis with LbD combined with bead beating resulted in the highest concentrations (1.48 ± 0.2 µg/µL), while both chemical lysis methods yielded lower concentrations (TFA: 1.11 ± 0.1 µg/µL; acetone precipitation: 1.03 ± 0.14 µg/µL). The values differ significantly ($p < 0.05$; ANOVA) from each other. Therefore, LbD combined with bead beating was chosen as the protein extraction method for *M. smegmatis* cells.

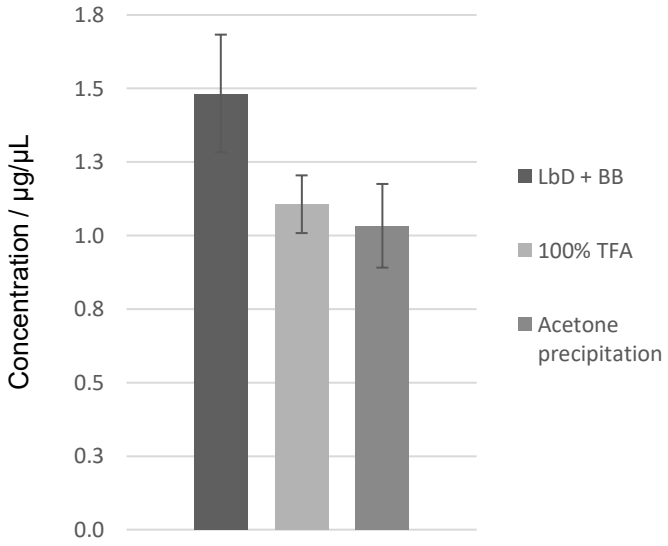


Figure 22: Protein concentrations yielded with lysis buffer D combined with bead beating (LbD + BB), 100 % TFA, or with acetone precipitation. Experiments were conducted in triplicate.

4.2.4 Protein purification

Donnelly *et al*. (2019) provided best practice guidelines for intact protein analysis by top-down mass spectrometry, recommending protein precipitation for moderately complex samples (5–100 proteins) and two-dimensional separation prior to MS analysis for highly complex samples (>100 proteins).[41] Although whole proteome analysis (with over 100 proteins) is the aim of this study, protein precipitation and molecular weight cut-off (MWCO)-based filtration (recommended for 1-5 proteins) were tested as sample preparation methods. The decision not to follow the 2D pre-separation was based on practical considerations regarding the compatibility of workflows and sample throughput.

Centrifugal filters with a MWCO of 3000 Da were chosen for filtration of mycobacterial lysates. While literature recommends using ultrafiltration devices with an MWCO not greater than half the molecular mass of target proteins to avoid sample loss, filters with smaller MWCOs were not commercially available for the intended setup. Therefore, the loss of proteins smaller than 6000 Da could not be excluded. In addition, the use of MWCO filters was deemed unsuitable for the intended workflow due to clogging observed during filtration of the bacterial lysates. In addition, the high cost per filter rendered MWCO-based filtration impractical for routine use in a high-throughput proteomics context. These considerations led to the exclusion of MWCO filters from further experiments.

To concentrate proteins and remove contaminants such as salts or detergents, protein precipitation was tested either with chloroform-methanol or with ice-cold acetone after a protocol from Donelly *et al*.[41] For the chloroform-methanol precipitation, a volume ratio of 1:1:4:3 of aqueous protein sample:chloroform:methanol:water was used. The supernatant was subsequently removed by aspiration, and the precipitated pellet was further washed with one more addition and removal of methanol. For acetone precipitation, an aqueous protein sample was incubated with a 1:4 ratio (v/v) of ice-cold acetone for 5 minutes at -20 °C. The supernatant was then discarded, and protein pellets were washed with ice-cold acetone, which was discarded afterward. Protein pellets were air-dried (room temperature) for a maximum of 5 min to prevent the formation of insoluble protein aggregates. Protein pellets from both precipitation methods were resolubilized for 15 min at -20 °C with 80 % FA (25 % of the desired end volume) and subsequently diluted with water. The resulting concentrations are displayed in **Figure 23**. As a result, acetone precipitation was chosen for further experiments as it yielded significantly (Welch's t-test; $p < 0.05$) higher protein concentrations than the methanol-chloroform precipitation.

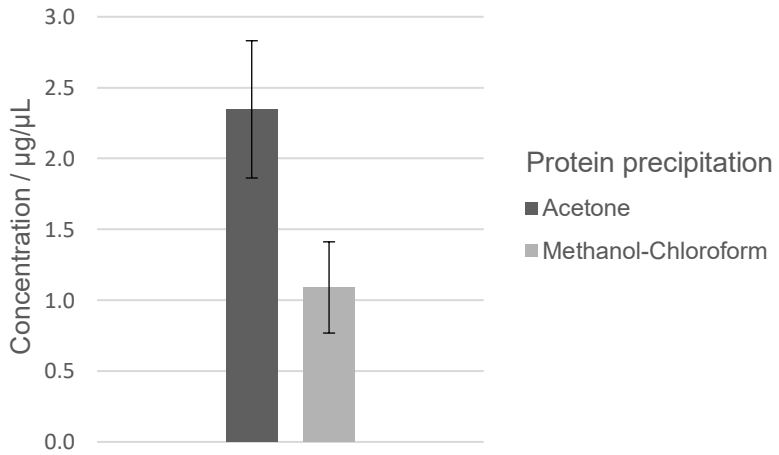


Figure 23: Protein concentrations yielded after acetone or methanol-chloroform precipitation (each in triplicate).

4.2.5 Precipitation duration

As commonly recommended in literature, a minimum acetone precipitation time of at least 1 h is required to yield sufficient protein amounts. This was tested and compared with different precipitation times in triplicate: 5 min, 30 min, 1 h, and overnight. The resulting concentrations can be seen in **Figure 24**. Protein concentrations after 5 min and 30 min of acetone precipitation gave similar values in a range of 1.48 ± 0.06 µg/µL (5 min) and 1.46 ± 0.026 µg/µL (30 min). Higher values were observed with 1 h (1.01 ± 0.38 µg/µL) and overnight precipitation (1.85 ± 0.12 µg/µL), but it is worth noting that concentrations yielded with precipitation for 1 h show a high standard deviation. The concentrations differ significantly from each other (ANOVA; $p < 0.05$), though Tukey's post hoc test showed no significant pairwise differences. However, acetone precipitation conducted either for 1 h or overnight showed the highest mean values, suggesting the most effective precipitation. To facilitate workflow, future precipitation was carried out overnight.

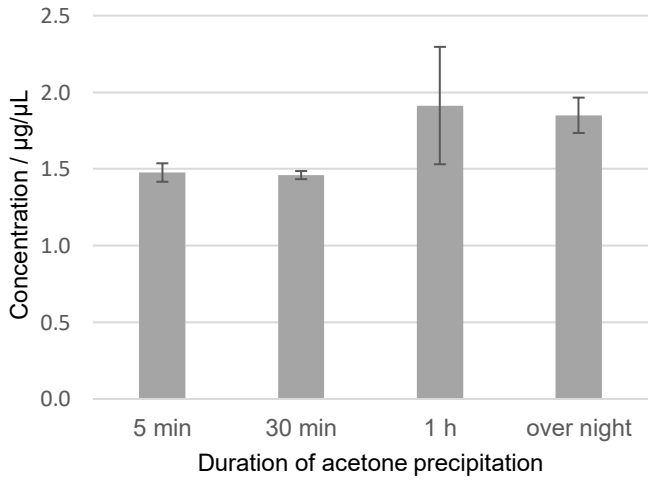


Figure 24: Yielded protein concentrations after different acetone precipitation times (5 min, 30 min, 1 h, and overnight). Experiments were performed in triplicate.

4.2.6 Testing of different resolubilization solvents

Precipitation of proteins is an important step for removing detergents and various contaminants. However, it also poses a risk of analyte loss by incomplete resolubilization of the precipitated proteins. Therefore, different solvents were tested to yield the highest protein concentration after precipitation. One approach follows a protocol of Doucette *et al*.[42], in which protein pellets are incubated using a small volume (25 % of the desired end volume) of ice-cold 80 % FA for 2 min at -20 °C and are then diluted to the desired final concentration with LC-MS grade water. Other protein resolubilization approaches follow Neumann *et al.* (v/v/v; 50 % ACN, 35 % FA, 15 % H_2O), with the initial solvent starting conditions of the LC-MS measurement (v/v/v; 90 % H_2O, 10 % ACN, 0.1 % FA). Each experiment was tested in quadruplicate, and the yielded protein concentrations can be seen in **Figure 25** (light grey bars). The protein samples from the resolubilization approaches with either 80 % FA or 50 % ACN, 35 % FA, and 15 % H_2O caused clogging of the LC-MS system. Therefore, an additional sample equilibration step was introduced. Solvents were evaporated using a centrifugal vacuum concentrator, and the remaining protein pellets were again resolubilized with the initial HPLC-solvents (90 % H_2O, 10 % ACN, 0.1 % FA). Proteins resolubilized with 90 % H_2O, 10 % ACN, and 0.1 % FA did not clog the system. Nevertheless, the samples were also vacuum-dried and subsequently resolubilized to assess the potential impact of the drying process on protein solubility. To evaluate protein recovery in the new solvent composition, protein concentrations were measured again and can be seen in **Figure 25** as dark grey bars. The solvent compositions for the first and second resolubilization are indicated in the legend.

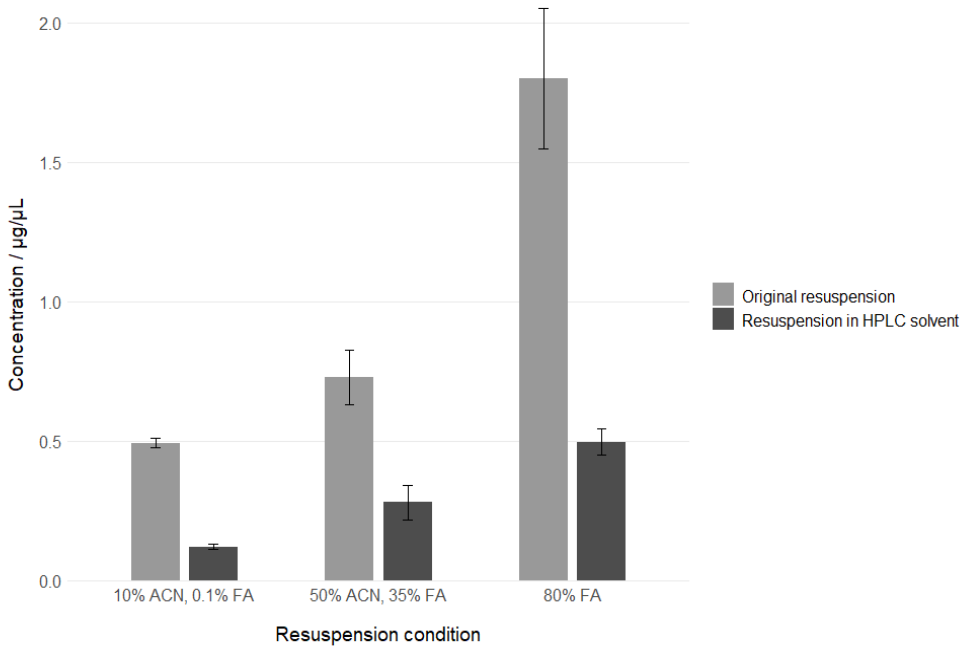


Figure 25: Protein concentrations yielded after resuspension in different solvent compositions. Grey bars: 10 % ACN with 0.1 % FA, 50 % ACN with 35 % FA, or 80 % FA (v/v). All compositions were adjusted to 100 % with water. Dark grey bars: Protein concentrations after vacuum drying and resolubilizing with the initial HPLC solvent (90 % H_2O, 10 % ACN, 0.1 % FA). All experiments were conducted in quadruplicate.

Protein resolubilization in 80 % FA yielded significantly higher concentrations than the other solvents (ANOVA; $p < 0.01$). Also, after vacuum drying and subsequent resolubilization in the initial LC-MS solvents, ANOVA indicated a significant difference in concentrations ($p < 0.001$). However, Tukey's post hoc test revealed no significant pairwise differences. Nonetheless, resolubilization in 80 % FA showed the highest mean concentration and was therefore chosen for the subsequent experiments.

4.3 Protein analysis of *M. smegmatis*

4.3.1 Intact mass analysis of *M. smegmatis* wild-type

To analyze the intact proteins of *M. smegmatis*, bacteria were incubated for 24 h, and proteins were extracted as described above. Two biological replicates were measured. In **Figure 26**, deconvoluted protein m/z signals are shown on the left, and the respective TIC and the LC-MS contour plot of one replicate are shown on the right. Only one biological replicate ("Biol. replicate 1") was fully analyzed, as the manual data processing involved is highly time-consuming. The second replicate showed similar signal patterns, supporting the reproducibility of the data.

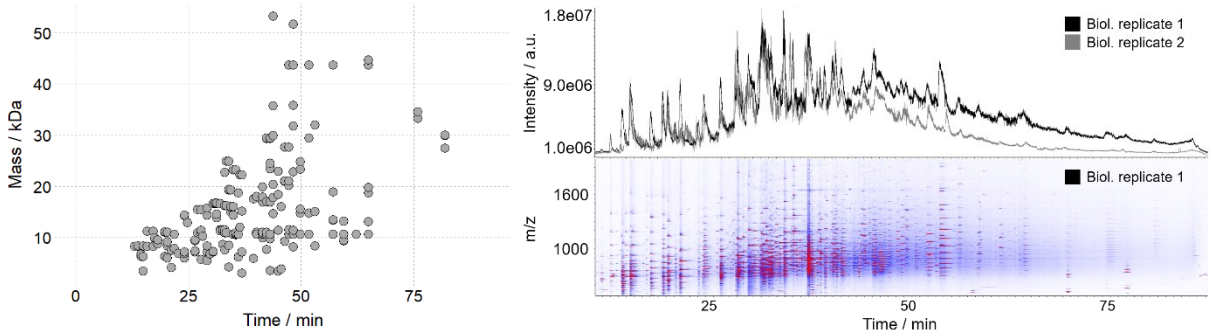


Figure 26: Left: Deconvoluted masses from biological replicate 1 in kDa over the retention time. **Right**: TICs (upper pane) of two biol. replicate *M. smegmatis*-protein extracts and a respective LC-MS contour plot of biol. replicate 1.

Every isotopic distribution of a protein showed co-eluting isotopic envelopes in the higher m/z range at the same retention time. They likely represent non-specific artifacts, such as in-source oxidation or adduct formation, as these features were also observed when analyzing the protein standard. They were not included in further analysis as they are assumed to be of limited biological relevance.

229 protein masses could be detected, whereas a possibly higher number of proteins is present, but could not be deconvoluted due to low S/N. The detected proteins are present in a mass range of 2-52 kDa, with a higher number in the lower mass range up to 20 kDa. The maximum amount of proteins (47) can be observed with a molecular weight of 10-12 kDa. The lower abundance of big proteins (> 20 kDa) can result from various factors, but it can be assumed that they originate from analytical limitations, as a higher MW of *M. smegmatis* could be observed in SDS-PAGE as shown before (s. **Figure 20**). For comparison, a molecular weight distribution

of intact *M. smegmatis* is visualized (s. **Figure 27**) based on protein entries retrieved from SwissProt. SwissProt is a manually curated and high-quality subset of the UniProt Knowledgebase (UniProtKB). Unlike the full UniProt database, SwissProt does not include automatically annotated and computationally predicted protein sequences and maintains sequence redundancy to a minimum.[146, 147] As a result, the total number of proteins from SwissProt is lower than the total number found in UniProt (for *M. smegmatis,* approximately one-tenth). However, SwissProt ensures higher data reliability, as the dataset consists only of reviewed and well-characterized (e.g., with protein function, PTMs, binding sites) proteins. In total, 593 proteins are annotated for *M. smegmatis* in SwissProt ("strain ATCC 700084 / mc(2)155), *Mycobacterium smegmatis* reference proteome UP000000757"). The theoretical molecular weights were calculated with the protein-molecular-weight calculator from Altogen Labs.[128]

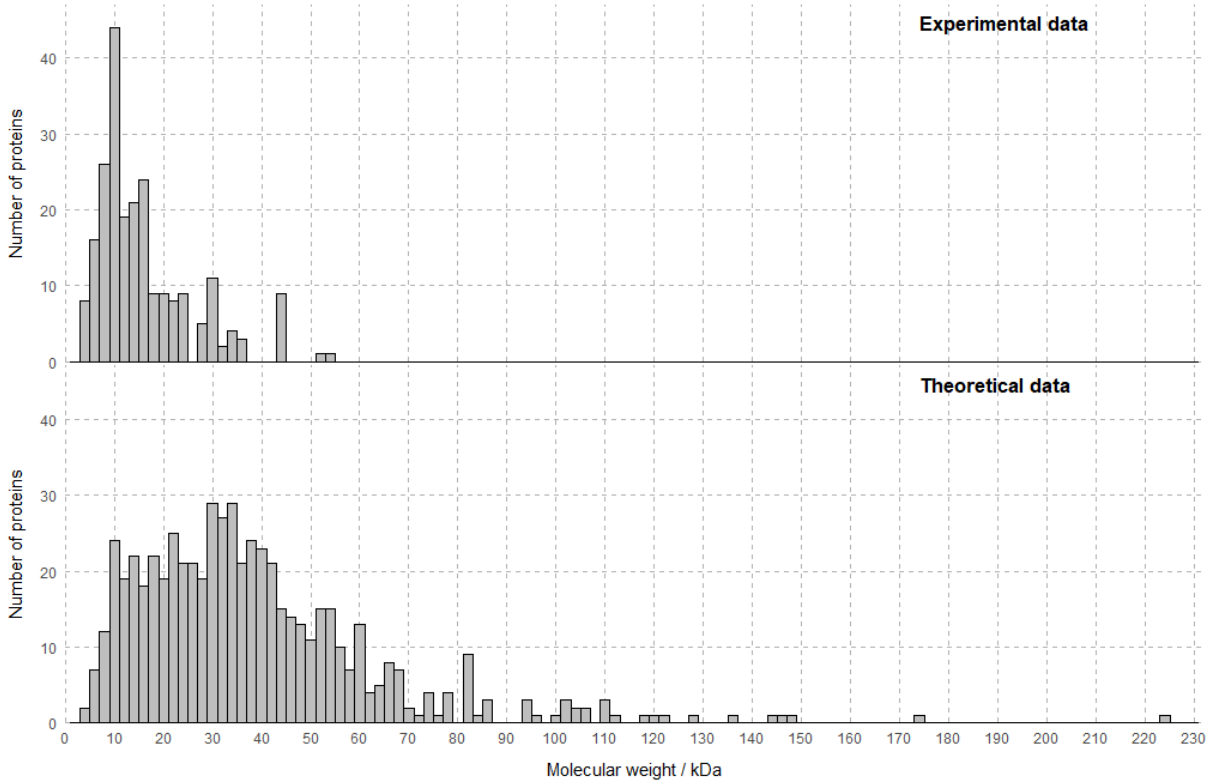


Figure 27: Molecular weight distribution of intact *M. smegmatis* proteins analyzed by HPLC-MS (experimental data) and annotated *M. smegmatis* proteins from the SwissProt reference proteome UP000000757 (theoretical data). The theoretical molecular weights were calculated with the protein-molecular-weight calculator from Altogen Labs.[128]

The distribution of the SwissProt proteins shows a strong right skew, with the majority of proteins in the lower molecular weight range (1-90 kDa), peaking around 30–40 kDa. A lower

protein frequency, with 26 proteins, is observed with molecular weights higher than 90 kDa, and 21 of these exceeding 100 kDa. This pattern suggests that *M. smegmatis* mostly expresses proteins up to 90 kDa. Comparing both *M. smegmatis* MW distributions, the experimental data do not match the SwissProt data, as the maximum protein size detected is 52 kDa. Additionally, only a subset of proteins was detected, indicating that not all possible proteins were identified in the analysis. Furthermore, it cannot be ruled out that some of the deconvoluted masses originate exclusively from proteoforms (e.g., the same protein with various PTMs), which, for example, are not classified as different proteins in SwissProt. It is possible that higher molecular masses (> 52 kDa) were not detected due to the instrument's inherent limitations. According to the manufacturer, the TripleTOF 5600 has an upper mass detection limit of 40 kDa.[148] While the detection of proteins bigger than 40 kDa (s. chapter **4.1.3**) demonstrates that this limit can be exceeded under certain conditions, it may still explain the low frequency and the absence of high-mass (> 40 kDa) ions. Moreover, proteins of higher MW often exhibit lower ionization efficiency and show low S/N. Especially when coeluting with low MW proteins, big proteins often remain undetected. Similar observations have been described in previous studies, demonstrating that single-dimensional RPLC coupled to MS only identifies a part of the intact proteoforms and is limited in the detection of larger (> 60 kDa) molecules.[149]

4.3.2 Intact mass analysis of dormant *M. smegmatis*

M. smegmatis were grown under acidification to generate dormant bacteria. Following the respective incubation period of 20 days, bacterial cultures were centrifuged, and an aliquot of the cells was incubated overnight in nutrient-rich medium to assess their resuscitation ability. Only cells that resumed growth under these conditions were classified as dormant and were included in subsequent analyses. This approach is in line with previously described methods for testing bacterial dormancy.[15]

Three biological replicates were analyzed via LC-MS. The LC-MS contour plots (s. **Figure 28**, right) show that the dormant cultures exhibit fewer protein signals than the wild-type bacteria (s. chapter **4.3.1**), though the same overall protein concentrations were injected. Protein masses were deconvoluted with Bio Tool Kit, and only proteins that were detected in at least two of the three biological replicates were considered for further analysis. The mean deconvoluted mass is shown in the left panel of **Figure 28**.

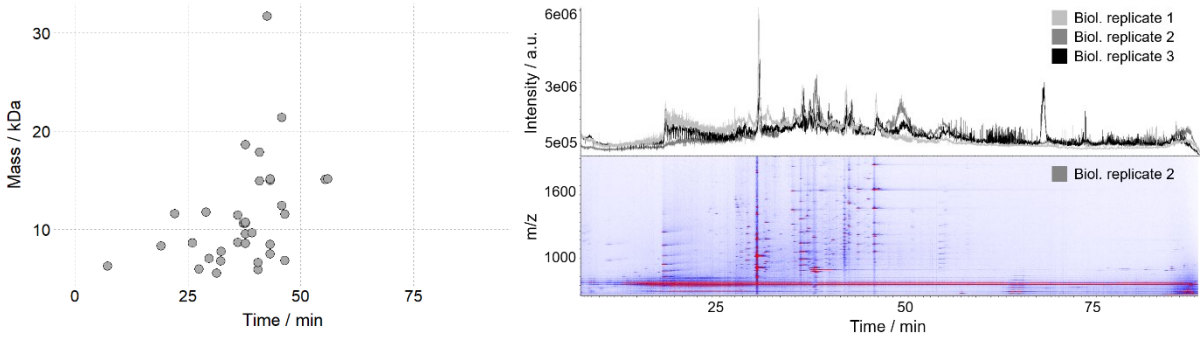


Figure 28: Averaged protein mass (of at least two biological replicates) in kDa over time (left pane) and the respective TIC of the three biological replicates and the contour plot of replicate 2 (right pane).

Biological replicate 1 contains fewer proteins (32) than the other two biological replicates (36 and 37, respectively). 34 proteins were detected in at least two out of the three samples analyzed, with molecular masses ranging from 5.5 kDa to 31.7 kDa. In **Figure 29**, the mass distributions from wild-type and dormant mycobacterial proteins are depicted as violin plots. In contrast to the proteins of the wild-type bacteria, the dormant bacteria showed fewer large (> 20 kDa) proteins. The median intact mass of dormant proteins is with 10123.9 Da lower than the wild-type median with 12133.6 Da.

Figure 29: Violin plots of the deconvoluted protein masses of wild-type (blue) and dormant (red) *M. smegmatis*.

When allowing a 65 ppm mass tolerance, as previously defined as mass accuracy (s. chapter **4.1.3**), 24 masses matched each other, while ten proteins were only detected in dormant bacteria (s. **Figure 30**). It cannot be ruled out that these ten proteins are also present in wild-type bacteria, but may have stayed undetected due to overlapping signals, resulting from an overall higher protein content in the wild-type condition. Alternatively, their detection in the

dormant condition could also be caused by an upregulation of, for example, stress response proteins that were reported previously by bottom-up analysis.[49]

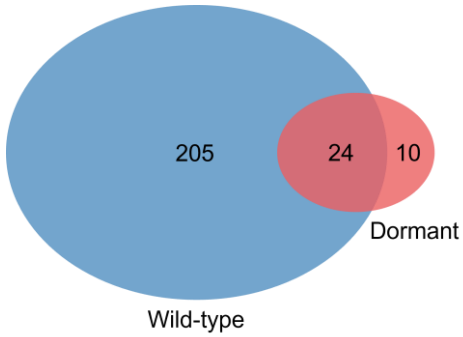


Figure 30: Venn diagram of deconvoluted protein masses from wild-type and dormant mycobacteria. 24 masses match each other with a mass tolerance of 65 ppm.

4.3.3 Dynamic measurement of intact proteins

In this section, parts of the experimental design and data analysis are adapted from the protocol of Haisch *et al.*[3] Non-targeted stable isotope labeling with glycerol-1,3-$^{13}C_2$ was performed to monitor the *de novo* protein synthesis in mycobacteria. In contrast to their use of MALDI-TOF MS, HPLC-MS was employed as the analytical readout method, allowing for a broader proteome coverage and higher resolution. For this, *M. smegmatis* cultures were incubated with ^{12}C-glycerol as a control and with ^{13}C-glycerol for labeling across multiple time points (1.5 h, 3 h, 4.5 h, 6 h, 7.5 h, 10.5 h, and 24 h). **Figure 31** shows a section of representative mass spectra of a protein (7215 Da, charge state seven) after four different incubation periods (1.5 h, 4.5 h, 7.5 h, and 24 h). After incubation of the bacteria with ^{12}C-glycerol (unlabeled, black), the protein peak shows its maximum at ~1032 m/z. After incubating with heavy ^{13}C-glycerol (labeled, red), the initial protein peak is less intense, while a peak at higher m/z is detected. Glycerol is incorporated into various metabolic pathways of the mycobacteria and remains likely a complete structural unit. Accordingly, the higher mass peak is probably not formed by a continuous mass shift over time but results from molecular fragments with discrete m/z values. For the sake of simplicity, however, the term *shift* will be further used in the following. While several mycobacterial proteins, showing a mass shift over time due to labeling with ^{13}C, were detectable by LC-MS, this study focuses on a more detailed analysis of six proteins, including the effects of varying incubation times and different MIC levels of the antibiotic rifampicin.

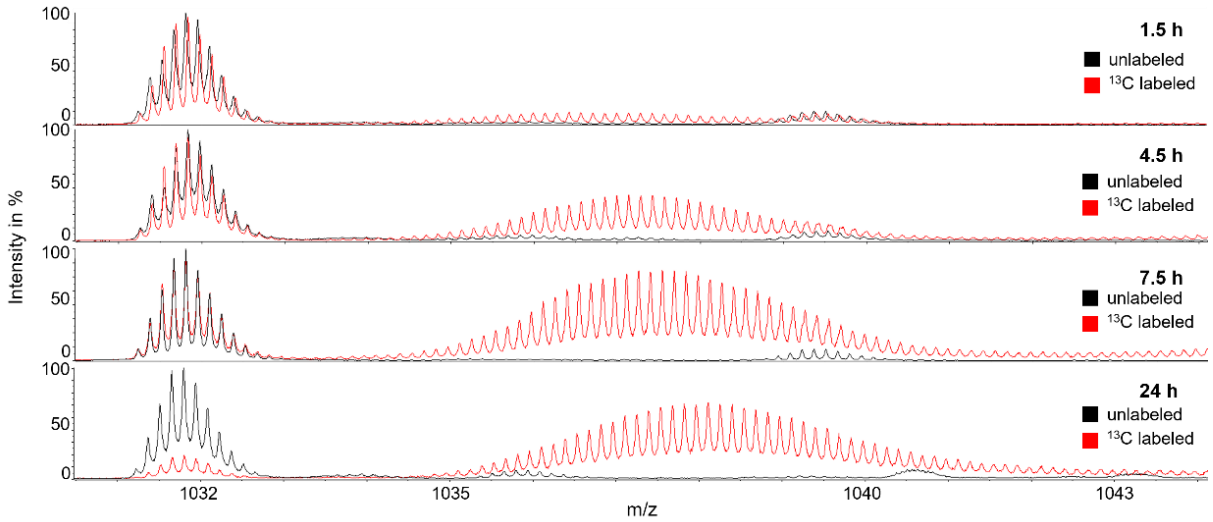


Figure 31: Mass spectra of a mycobacterial protein with a mass of 7215 Da (here: charge state seven) after different incubation times (1.5 h, 4.5 h, 7.5 h, and 24 h). Protein signals after incubation in medium containing ^{12}C-glycerol are indicated in black, and when incubated with ^{13}C-glycerol, in red.

Timeline with rifampicin

In alignment with the experimental design of the study by Haisch *et al*., the observation period with rifampicin treatment was limited to 7.5 h of incubation. This phase was selected because most of the dynamic changes in ^{13}C-incorporation and potential effects of antibiotic treatment are expected to occur during this time. At later time points, the incorporation typically approaches a steady state, limiting the ability to distinguish between different kinetics.

The COG was determined to observe the protein *de novo* synthesis over time. For this, the mass spectra of the respective proteins were deconvoluted manually over a defined mass window around the expected protein mass. The COG was then determined by analyzing this mass area, representing the intensity-weighted average of the detected deconvoluted mass peaks, accounting for possible PTMs and minor mass shifts (s. eq. 4).

$$COG = \frac{\sum_j m_j \cdot I_j}{\sum_j I_j} = \frac{\sum_j m_j \cdot A_j}{\sum_j A_j} \qquad \text{(eq. 4)}$$

m_j deconvoluted mass of peak j

I_j intensity of peak j

A_j area under the curve of peak j

As a limitation of the deconvolution algorithm used in PeakView, it should be noted that the deconvolution process appears to primarily utilize charge state detection to reconstruct neutral

masses. It remains unclear to what extent isotopic patterns are considered in the deconvolution algorithm itself. To reduce potential background noise and enhance the reliability of detected signals, we empirically established a threshold intensity value of 5000. By subtracting this threshold from the area of each deconvoluted mass, we aimed to filter out low-intensity signals that are less likely to represent meaningful features. Although this threshold is somewhat arbitrary, it provided a practical balance between sensitivity and noise suppression in our dataset. **Figure 32** presents the COG values of six representative proteins (6877 Da, 7215 Da, 9401 Da, 10642 Da, 10974 Da, 12868 Da) under different conditions, including unlabeled (G) and labeled (G*) cultures without antibiotic, with 1-fold MIC, and 20-fold MIC of rifampicin across multiple time points.

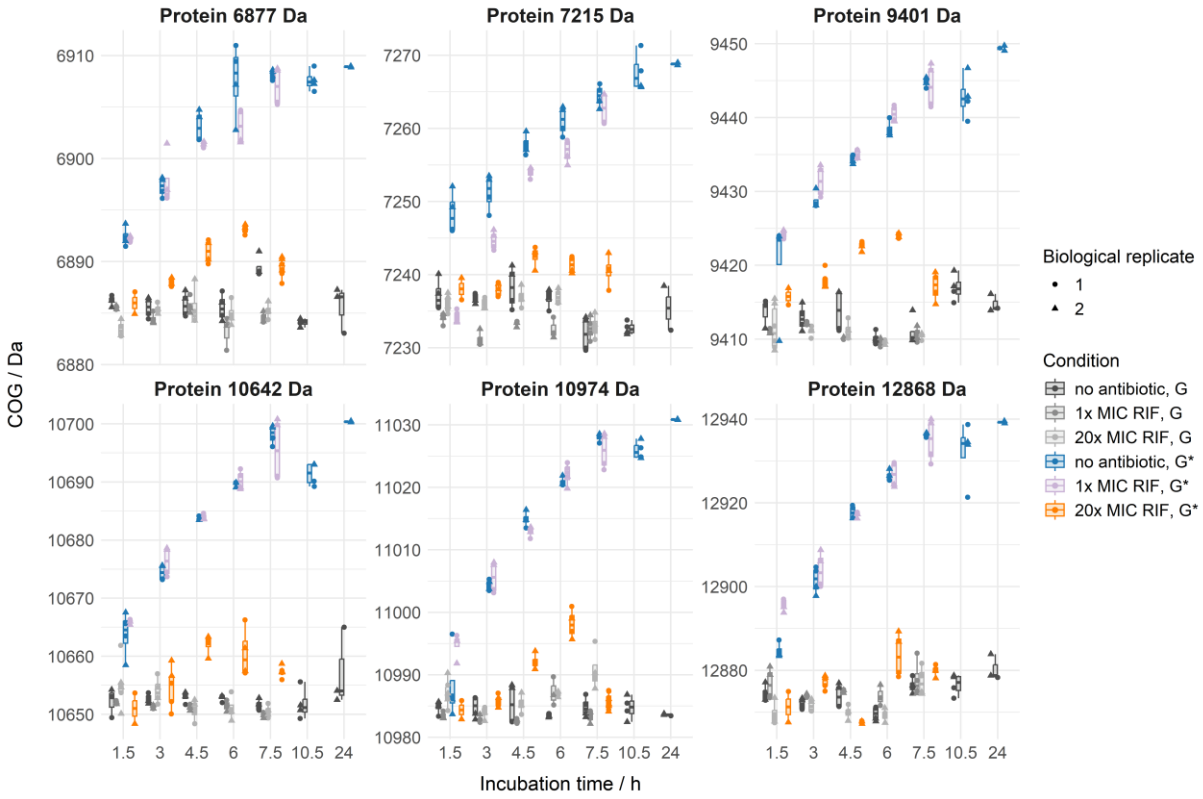


Figure 32: COG of six different *M. Smegmatis* proteins over different incubation periods with medium containing ^{12}C (G) or ^{13}C (G*), and with different treatment conditions (no antibiotic, 1-fold, and 20-fold MIC of rifampicin (RIF)).

The data illustrate the incorporation dynamics of heavy carbon into proteins across different time points and rifampicin exposures. In the absence of the antibiotic ("no antibiotic, G*"), a shift in COG values is observed over time, indicating ongoing *de novo* protein synthesis. Under exposure to 1-fold MIC of rifampicin, COG progression appears similar to the COG progression

of the control. At 20-fold MIC, the incorporation of labeled carbon seems to be suppressed or completely absent for some proteins, indicating inhibition of protein biosynthesis. To eliminate possible artifacts and to ease the dynamics interpretation, the differential COG (ΔCOG) was determined. As the data only provides information about four measurements (2 biological, 2 technical each), the median was chosen as a robust location estimator, as it's less affected by outliers and does not assume normally distributed data. With this, the ΔCOG was determined by subtracting the median COG value of the peaks from bacteria grown in G from those grown in G* medium. The measurement uncertainty was estimated by a non-parametric bootstrapping approach. G and G* values were independently resampled with replacement (n = 1000), and in each iteration, the difference of the medians was recalculated. This resulting distribution was used to determine the 95 % confidence interval (CI; 2.5-97.5 % quantile), which is indicated by error bars (s. **Figure 33**). Additionally, for the control group ("no antibiotic"), the temporal evolution of ΔCOG was fitted using the exponential decay equation (s. **2.4**) and normalized to 100 % (maximum shift). The measurement uncertainty of the fitted model was also assessed by bootstrapping (n = 1000), using the empirical standard deviation of ΔCOG to simulate resampled input data. The resulting 95 % CI is visualized as a light-green area in **Figure 33**.

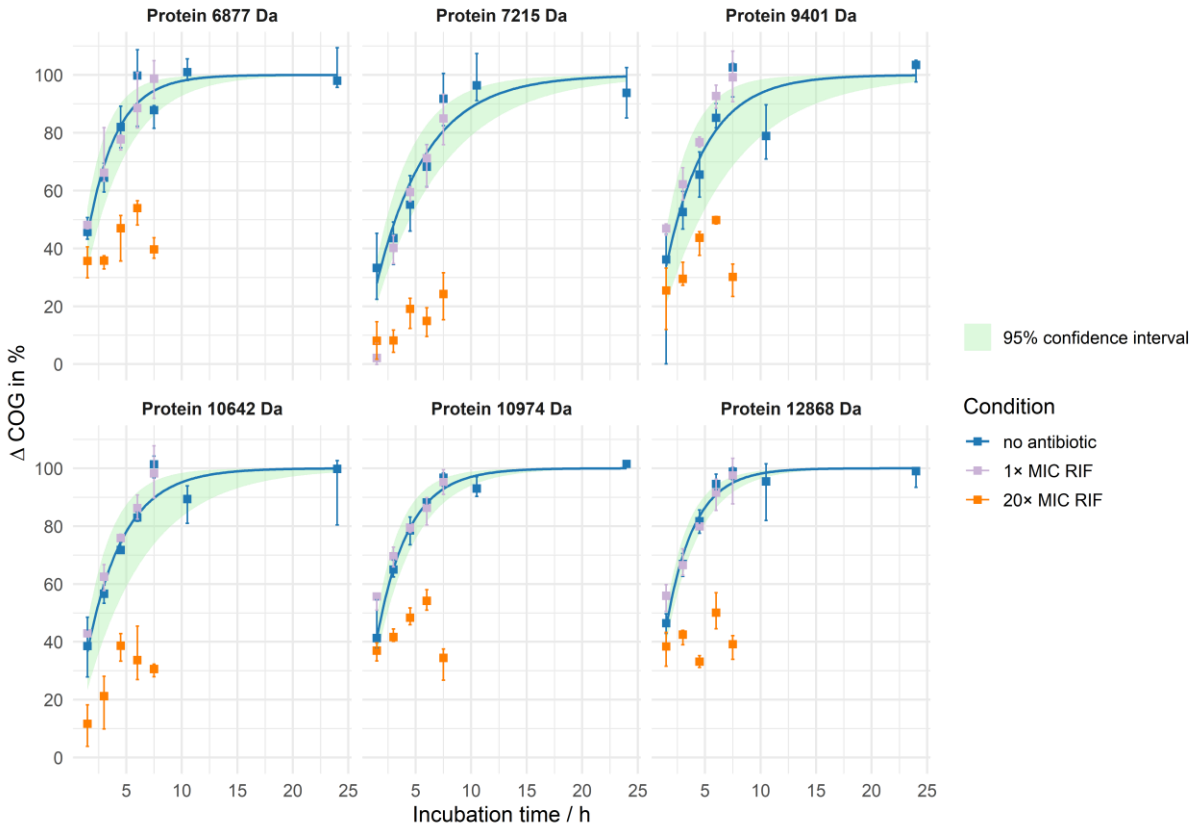


Figure 33: Protein-specific ΔCOG values over time, with the single-exponential decay function (eq. 2) being fitted to the control values ("no antibiotic"; blue line). ΔCOG values of proteins treated with a 1-fold or 20-fold MIC of rifampicin (RIF) are displayed relative to this fit.

The protein-specific parameters from the fitted model, such as specific rise time (τ) and the time to reach 50 % contribution, are summarized in **Table 10**. The data shows the different temporal behavior of each protein with τ ranging from 2.50 h (95 % CI: 2.11–3.00 h) to 4.56 h (95 % CI: 3.39-6.42 h). The time to reach 50 % contribution (t_{50}) ranges from 1.78 h to 3.16 h. These findings align with the observations reported by Haisch *et al.*, with t_{50} ranging from 1.50 ± 0.62 h to 5.65 ± 2.06 h. By analyzing the ΔCOG values depicted in **Figure 33**, proteins after treatment with 1-fold MIC of rifampicin show similar dynamics to the untreated control, lying within the 95 % CI. Only the protein with 7215 Da shows different kinetics. The first value after 1.5 h incubation is significantly lower than the control, suggesting suppressed protein synthesis at the early stage. All ΔCOG values of the mycobacterial proteins treated with a 20-fold MIC of rifampicin exhibit a different dynamic, to which the model could not be fitted. They show lower values than the control, indicating downregulated protein synthesis. A change in the ΔCOG shift after 6 h (protein 10642 Da) or 7.5 h (proteins 6877 Da, 9401 Da, 10975 Da,

12868 Da) could indicate an interruption of protein *de novo* synthesis or enhanced protein degradation.

Table 10: Specific rise time τ and time to reach 50 % (t_{50}) contribution (assuming exponential behavior) for the six analyzed proteins.

Protein / Da	τ / h	95 % CI for τ / h	t_{50} / h
6877	2.57	1.93 - 3.79	1.78
7215	4.56	3.39 - 6.42	3.16
9401	3.62	2.41 - 6.46	2.51
10642	3.29	2.41 - 5.59	2.28
10974	2.84	2.24 - 3.54	1.97
12868	2.50	2.11 - 3.00	1.73

Similar to the MALDI-TOF-based data in the original study, the COG dynamics in this study provide insight into translational regulation under antibiotic pressure. However, extending the approach by implementing an automated protein detection would be required to assess a broader protein coverage, obtaining more information about the mycobacterial proteome dynamics. Moreover, additional antibiotics should be tested to gain a better understanding of their metabolic activation pathways in mycobacteria. Protein identification of the observed proteins is essential to link the analyzed labeling patterns to specific biological functions. This will be addressed in the next section.

4.3.4 Protein identification via time-dependent fractionation

An assignment of peptides from a whole-proteome analysis (via digestion) to their corresponding intact proteins can be challenging, due to shared peptides among proteins and the presence of proteins with similar masses. Therefore, intact proteins were measured via LC-MS, and simultaneously, fractions were collected online over specific time intervals. The setup and a schematic representation of the time-dependent fraction collection are depicted in **Figure 34**. The proteins of the fractions were subsequently identified via bottom-up analysis. By this, sample complexity is reduced, narrowing down the number of identified proteins. Parallel intact mass analysis ensures the presence of the analytes in the respective fractions. Fractionation was performed by integrating a T-piece after the HPLC column. To ensure equal backpressure on

both outlets, tubing of identical length was installed, with an ESI capillary connected at each end. One end of the connected tubing directed the flow to the MS for intact mass analysis, while the other end was used for manual fraction collection.

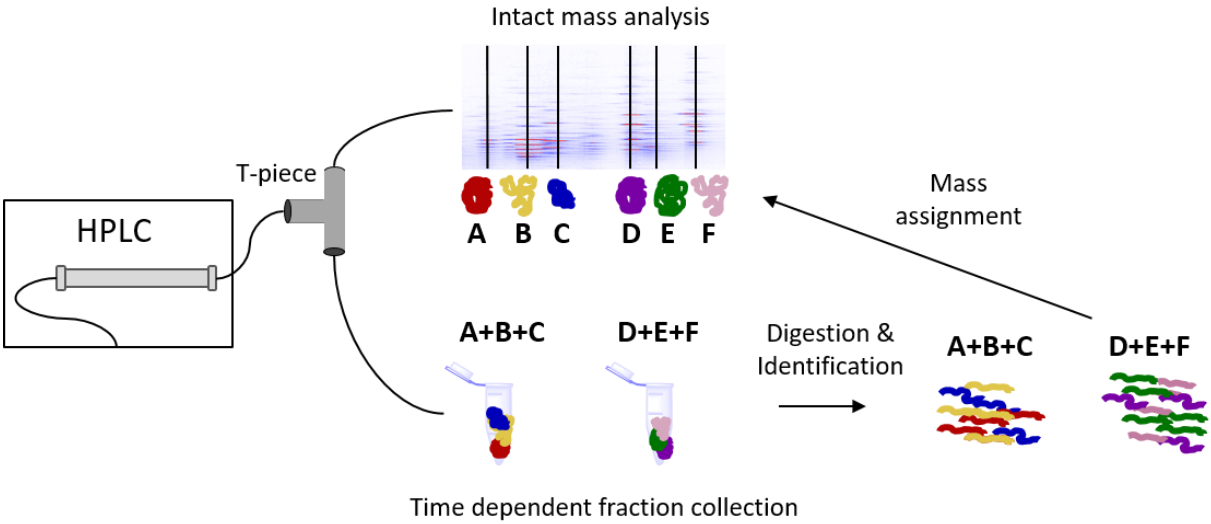
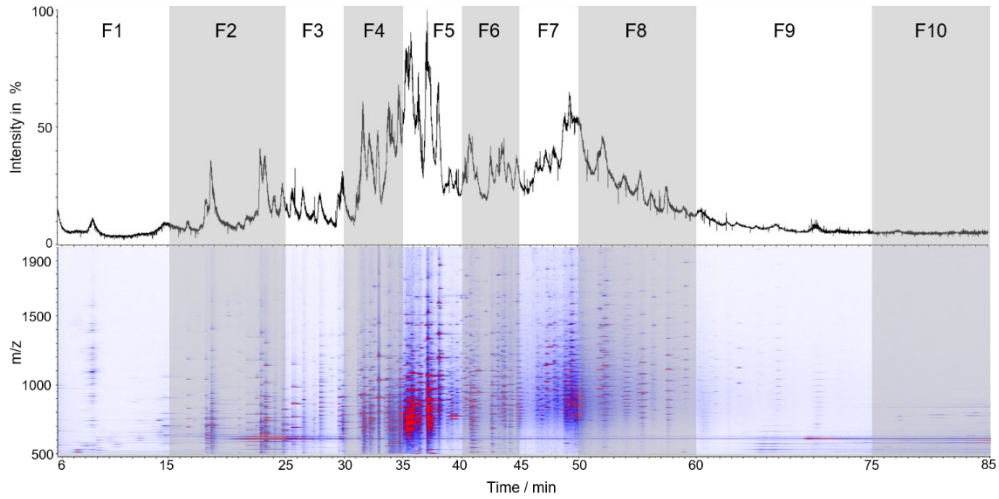


Figure 34: Scheme of the time-dependent fraction collection to identify the intact proteins.

For the fractionation of *M. smegmatis* protein extracts, two biological replicates were analyzed. Fractionation was started right after elution of the injection peak (6 min), and fractionation periods were adapted to the protein abundance eluting throughout the gradient: at less intense signal periods, samples were collected for 9-15 min, while at higher concentrations (25-50 min), fractions were collected over 5 min intervals (s. **Figure 35**). Fractions were pooled over four measurements to ensure sufficient protein amount for digestion.

Online intact mass analysis was compromised due to the reduced sample amounts and the decreased flow resulting from the T-piece splitting. Therefore, additional intact mass measurements were conducted with one port of the T-piece being capped to ensure the complete flow-through to the MS. These measurements yielded higher S/N and facilitated the intact mass analysis. However, a shift in retention time was observed due to the higher backpressure resulting from the capped T-piece. This time shift was considered with a time tolerance of ± 2 min when mass assignment of the time-dependent fractions was performed.



Fraction (F)	Time range / min
1	6–15
2	15–25
3	25–30
4	30–35
5	35–40
6	40-45
7	45–50
8	50–60
9	60–75
10	75–85

Figure 35: Time-dependent collection of fractions as illustrated by the TIC and contour plot of an *M. smegmatis* protein extract.

To ease protein detection, an automated deconvolution pipeline for intact mass analysis was applied, developed in cooperation with Dr. Ines Assum (s. chapter **3.4.2**, external data processing). For the mass assignment, protein masses were prefiltered (int. quant factor = 0.3), and the masses that matched the integer mass of the identified (digested) protein were considered for further analysis by manual evaluation with the Bio Tool Kit deconvolution. Overall, 15 proteins could be identified by mass assignments to the digested proteins of the respective fraction and are listed in **Table 11**. The table is split by a bold vertical line, with the left part showing the masses originating from the analysis following the automatic data pipeline, and the right part displaying the manually detected protein masses.

Table 11: Identified proteins by pipeline analysis with protein accession and theoretical, integer mass (left), and the manually detected protein masses with their mass difference to the theoretical mass (right).

Biological Replicate	Fraction	Protein Accession	Mass / Da	Deconvoluted Mass / Da	Mass Difference / Da	Mass Difference in ppm
1	F2	A0QSG0	11206	11205.4	-0.6	-54
2				11205.6	-0.4	-36
3				11205.8	-0.2	-18
1	F3	A0QV42	12868	12867.1	-0.9	-70
2				12867.5	-0.5	-39
3				12867.6	-0.4	-31
1	F4	A0QSD8	15742	15741.4	-0.6	-38
2				15741.9	-0.1	-6
3				15742.1	0.1	6
1	F5	A0R7F6	15923	15922.4	-0.6	-38
2				15923.0	0.0	0
3				15923.2	0.2	13
1		A0R441	16633	16636.2	3.2	192
2				16636.8	3.8	228
3				16637.0	4.0	240
1	F6	A0QV28	12116	12115.1	-0.9	-74
2				12115.8	-0.2	-17
3				12115.5	-0.5	-41
1		A0R561	17934	17933.4	-0.6	-34
2				17934.2	0.2	11
3				17934.1	0.1	6
1		A0QSH8	19871	19870.5	-0.5	-25
2				19871.4	0.4	20
3				19871.4	0.4	20
1	F7	A0QUY6	27649	27655.1	6.1	221
2				27656.2	7.2	260
3				27656.3	7.3	264
1		A0QSH8	19871	19870.5	-0.5	-25
2				19871.4	0.4	20
3				19871.4	0.4	20
1		A0R597	18349	18346.5	-2.5	-136
2				18347.3	-1.7	-93
3				18347.5	-1.5	-82
1		A0QTY6	11541	11562.5	21.5	1860
2				11562.8	21.8	1885
3				11562.8	21.8	1885
1	F8	A0QSD0	11433	11430.1	-2.9	-254
2				11430.3	-2.7	-236
3				11430.3	-2.7	-236
1		A0R1B1	22805	22803.6	-1.4	-61
2				22804.2	-0.8	-35
3				22804.5	-0.5	-22
1		A0QWV7	31942	31936.8	-5.2	-163
2				31938.2	-3.8	-119
3				31938.3	-3.7	-116

Matching protein masses were identified in the following fractions: one in F2, F3, and F4; two in F5; three in F6 and F8; and four in F7. Those matched masses show a maximal mass difference of ± 7.3 Da, while one protein (A0QTY6, F7) showed a mass difference of 21.5-21.8 Da. This mass difference could possibly be caused by a replacement of a proton by a sodium adduct (22 Da), which is typical for ESI-MS and has been reported in earlier studies.[150] The identified proteins were queried in the UniProt database. Information on cellular localization, molecular function, biological process, and review status (Swiss-Prot) is summarized in the following **Table 12**.

Table 12: Characterization of the identified proteins via LC-MS analysis. Information was obtained from the Uniprot database (May 2025).

Protein Accession	Protein name	Cellular localisation	Molecular Function	Biological process	Reviewed (SwissProt)
A0QSG0	Large ribosomal subunit protein uL24	50S ribosomal subunit	Ribonucleoprotein complex, ribosomal protein, rRNA binding, structural constituent of the ribosome	Translation	Reviewed
A0QV42	Large ribosomal subunit protein bL19	Cytosolic large ribosomal subunit	Structural constituent of the ribosome	Translation	Reviewed
A0QSD8	Large ribosomal protein uL16	Cytosolic large ribosomal subunit	tRNA binding, rRNA binding, structural constituent of ribosome	Translation	Reviewed
A0R7F6	50S ribosomal protein bL19	50S ribosomal subunit	Ribonucleoprotein complex, ribosomal protein, rRNA binding, structural constituent of ribosome	Translation	Reviewed
A0R441	Cold-shock DNA-binding domain protein	Unknown	DNA binding (cold-shock response)	Unknown	Unreviewed
A0QV28	Nitrogen regulatory protein P-II	Cytosol	ATP binding, enzyme regulator activity, regulation of nitrogen utilization,	Transcription, transcription regulation	Unreviewed
A0R561	RNA polymerase-binding transcription factor CarD	Cytoplasm	RNA polymerase-binding transcription factor CarD	rRNA transcription	Reviewed

Protein Accession	Protein name	Cellular localisation	Molecular Function	Biological process	Reviewed (SwissProt)
A0QSH8	Adenylate kinase	Cytoplasm	Adenylate kinase activity, ATP binding, AMP salvage	Unknown	Reviewed
A0QUY6	5-Carboxymethyl-2-hydroxymuconate delta-isomerase	Unknown	Acetylpyruvate hydrolase activity, metal ion binding	Unknown	Unreviewed
A0R597	Inorganic pyrophosphatase	Cytoplasm	Inorganic diphosphate phosphatase activity, magnesium ion binding, phosphate-containing compound metabolic process	Unknown	Unreviewed
A0QTY6	Stress-responsive A/B Barrel Domain superfamily protein	Unknown	Unknown	Unknown	Unreviewed
A0QSD0	Small ribosomal subunit protein uS10	Ribosome	Ribonucleoprotein complex, ribosome, tRNA binding, translation, structural constituent of ribosome	Translation	Reviewed
A0R1B1	Uncharacterized protein	Unknown	Unknown	Unknown	Unreviewed
A0QWV7	Nucleotide-binding protein	Unknown	ATP binding, GTP binding, displays ATPase and GTPase activities	Unknown	Reviewed

The proteins identified are either ribosomal and translation-related, non-ribosomal and functionally active (regulatory and enzymatic), or of unknown localization and function. Proteins that function as structural constituents of the ribosome, particularly those involved in rRNA binding and the 50S subunit, play a major role as targets in the antibiotic treatment of TB. One example is linezolid, which intervenes in the translational process.[151] Proteins that function in transcription regulation, or in energy-related pathways (ATP-/GTP-binding), play an important role as targets for the first-line drug like rifampicin, and the second-line drug bedaquiline. Rifampicin inhibits DNA-dependent RNA synthesis, while bedaquiline interacts

in the ATP synthesis.[152, 153] Additionally, the transcription factor CarD (A0R561) may be interesting as a potential drug target, as the protein was identified as an essential regulator in the persistence of *M. tuberculosis*.[154] Uncharacterized proteins of unknown function may also be promising drug targets, requiring further analysis. Moreover, an extended analysis of the identified proteins is needed to cover a wider spectrum of proteins. So far, only the protein with a mass of 12868 Da from chapter **4.3.3** could be identified as A0QV42. A0QV42 represents a structural constituent of the ribosome (large ribosomal subunit protein bL19) being interactive in the translational pathway. Given its similar dynamical behavior to the other five analyzed proteins, it can be assumed that all analyzed proteins are of ribosomal nature. This would need to be further validated. Incorporating greater mass deviations in mass assignments, e.g., by allowing additional masses from PTMs, may enhance the protein mass coverage.

4.3.5 TMT analysis of the mycobacterial proteome under antibiotic treatment

Taking advantage of TMT labeling, which allows for the simultaneous analysis of several different samples within one run, the effect of different conditions, such as antibiotics with different MICs, and various incubation times on wild-type and dormant *M. smegmatis* was investigated (s. **Table 13**). One group of samples (9 samples) was labeled but not included in the downstream analysis and is mentioned for transparency regarding the total number of tags used. Wild-type (WT) *M. smegmatis* were investigated either without antibiotics (control), with rifampicin, or with the drug candidate BTZ-043 in either 1-fold, 5-fold, or 20-fold MIC. Incubation times were chosen in alignment with the bacterial division rate: 4.5 h and 10.5 h within the exponential growth phase, and 24 h as a later-stage growth for observation in the potential stationary phase. Dormant mycobacteria were incubated either without antibiotics (control) or with a 5-fold MIC of the respective drug (rifampicin or BTZ-043). Incubation times were longer than for WT bacteria, as dormant bacteria feature reduced metabolic activity; therefore, a prolonged reaction time to the antibiotics is expected. **Figure 36** is a schematic representation of how internal quality controls were established. WT samples (4.5 h, 10.5 h, 24 h; each in triplicate) without antibiotic treatment were pooled for a WT control mix. Similarly, dormant control samples (1 d, 4 d, 10 d; triplicate each) were combined into a dormant control mix. The control mixes were then merged for one overall reference mix, which was added to each TMT set as an internal control. Seven full TMT sets (112 tags) were employed. The remaining unused (six) TMT tags were used as QC samples with the WT and dormant control mix to complete the sets.

Table 13: Prepared samples (biological triplicate each) for the TMT analysis.

M. smegmatis	Antibiotic	MIC	Incubation time	Quality control	
Wild-type (WT)	BTZ-043	1	4.5 h 10.5 h 24 h		
		5	4.5 h 10.5 h 24 h		
		20	4.5 h 10.5 h 24 h		
	Rifampicin	1	4.5 h 10.5 h 24 h		
		5	4.5 h 10.5 h 24 h		
		20	4.5 h 10.5 h 24 h		
	-	-	4.5 h 10.5 h 24 h	WT control mix	Reference mix
Dormant	-	-	1 day 4 days 10 days	Dormant control mix	
	BTZ-043	5	1 day 4 days 10 days		
	Rifampicin	5	1 day 4 days 10 days		

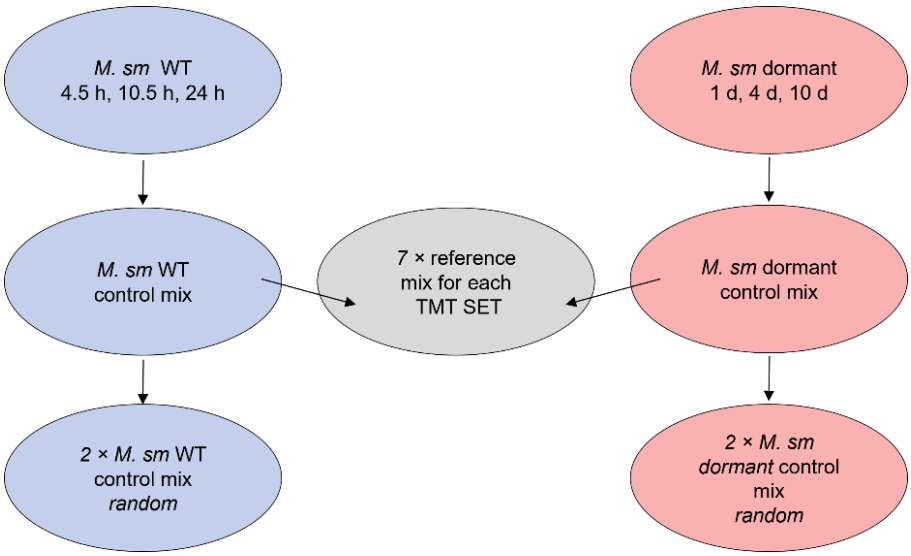


Figure 36: Schematic overview of generated internal quality controls for the TMT setup**.**

For the details of the data processing, see chapter **3.4.1**. Proteins identified across all experimental conditions and timepoints were compared to one another. Only proteins present in at least 60 % of all samples were included in the batch correction for the analysis. Additionally, they had to be detected in at least two of three biological replicates, and within those, in at least two technical replicates. Statistical significance was determined using a false discovery rate (FDR) threshold of 0.05, and proteins were considered differentially expressed with a $\log_2$ fold change (FC) of ± 1.

In **Table 14**, all proteins that were significantly up- or downregulated are listed. The analysis showed eleven significantly downregulated proteins and one upregulated (printed in bold) protein in WT bacteria. In the dormant samples, twelve proteins were significantly downregulated. Additional information, such as protein name, cellular localization, molecular function, biological processes, and whether the protein is reviewed by SwissProt, was obtained from the UniProt database (accessed May 2025).

Table 14: Significantly upregulated (printed in bold) or downregulated proteins from TMT analysis of wild-type and dormant bacteria.

Condition	Protein Accession	Protein name	Cellular localization	Molecular Function	Biological process	Reviewed (SwissProt)
Wild-type	A0QTS8	CsbD family protein	Unknown	Unknown	Unknown	Unreviewed
	A0QUM6	Hydrogenase-2, small subunit	Cell envelope	3 iron, 4 sulfur cluster binding, 4 iron, 4 sulfur cluster binding, electron transfer activity, ferredoxin hydrogenase activity, hydrogenase (acceptor) activity, metal ion binding, oxidoreductase	Anaerobic respiration	Unreviewed
	A0QUM7	Hydrogenase-2, large subunit	Cell envelope	Ferredoxin hydrogenase activity, hydrogenase (acceptor) activity, nickel cation binding, oxidoreductase	Unknown	Unreviewed
	A0QV17	Hemerythrin HHE cation binding region	Unknown	Unknown	Unknown	Unreviewed
	A0QZA1	Universal stress protein MSMEG_3950/ MSMEI_3859	Unknown	ATP binding	Unknown	Reviewed
	A0R5G1	**Acetyl-Coenzyme A synthetase**	Cytosol	Acetate-CoA ligase activity, AMP binding, ATP binding, metal ion binding, ligase	Acetyl-CoA, biosynthe-tic process from acetate	Reviewed
	Q3L887	Broad-specificity linear acyl-CoA dehydrogenase FadE5	Plasma membrane	Long-chain fatty acyl-CoA dehydrogenase activity, medium-chain fatty acyl-CoA dehydrogenase activity, short-chain fatty acyl-CoA dehydrogenase activity, oxidoreductase	Fatty acid metabolic process, fatty acid meta-bolism, lipid metabolism	Reviewed
	A0R2Q4	Fasciclin domain protein	Extracellular matrix, extracellular space	Cell adhesion molecule binding	Cell adhesion, extracel-lular matrix organiza-tion	Unreviewed
	A0QZ96	Universal stress protein family protein	Unknown	Unknown	Unknown	Unreviewed

Condition	Protein Accession	Protein name	Cellular localization	Molecular Function	Biological process	Reviewed (SwissProt)
	A0R692	DNA protection during starvation protein	cytoplasm nucleoid	DNA binding, ferric iron binding, oxidoreductase activity, acting on metal ions	Intracellular iron ion homeo-stasis	Reviewed
	A0QXT5	Ferritin-like domain-containing protein	Unknown	Unknown	Unknown	Unreviewed
	A0R4B1	Uncharacter-ized protein	Unknown	Unknown	Unknown	Unreviewed
Dormant	A0QU81	Secreted chorismate mutase	Extracellular region	Chorismate mutase activity, isomerase	Chorismate metabolic process, salicylic acid biosyn-thetic process	Reviewed
	A0R0A1	Glyoxalase/ bleomycin resistance protein/dioxy-genase	Unknown	Unknown	Unknown	Unreviewed
	A0QX24	ATPase, MoxR family protein	Unknown	ATP binding, ATP hydrolysis activity	Unknown	Unreviewed
	A0R248	Beta-carbonic anhydrase 1	Unknown	Carbonate dehydratase activity, zinc ion binding	Unknown	Unreviewed
	A0QSK7	Luciferase-like domain-containing protein	Cytosol	Oxidoreductase activity, acting on paired donors with incorporation or reduction of molecular oxygen	Unknown	Unreviewed
	A0QSN8	PduO protein	Unknown	Unknown	Unknown	Unreviewed
	A0QSP1	Glycerol dehydratase large subunit	Unknown	Cobalamin binding, glycerol dehydratase activity	Unknown	Unreviewed
	A0R729	Glycerol kinase	Cytosol	ATP binding, glycerol kinase activity	Glycerol catabolic process, glycerol-3-phosphate metabolic process	Unreviewed
	A0QZL1	Zinc-binding alcohol dehydrogenase	Unknown	NADPH binding, oxidoreductase activity, acting on NAD(P)H	Unknown	Unreviewed

Condition	Protein Accession	Protein name	Cellular localization	Molecular Function	Biological process	Reviewed (SwissProt)
	P0CH36	NADP-dependent alcohol dehydrogenase C 1	Unknown	Oxidoreductase, alcohol dehydrogenase (NADP+), zinc ion binding	Unknown	Reviewed
	A0QXX7	Catalase-peroxidase 2	Cytosol	Catalase activity, heme binding, metal ion binding; oxidoreductase, peroxidase	Cellular response to hydrogen peroxide, hydrogen peroxide catabolic process	Reviewed
	A0QV17	Hemerythrin HHE cation binding region	Unknown	Unknown	Unknown	Unreviewed

To ease data interpretation, proteins are shown in network plots in **Figure 37**. Here, each protein is displayed as a node (circle) with its size representing the $\log_2$(FC). The shape at the center of the node indicates the treatment condition: circles for bacteria treated with BTZ-043, triangles for rifampicin treatment, and diamond shapes when proteins are up-/downregulated in both treatments. The shape's color denotes the direction of regulation, with red for upregulation and blue for downregulation. Different node colors (pie charts) reflect functional annotation/pathway involvement of the proteins within the bacterium. Various segments represent multiple functional categories, and the lines indicate an interaction or connection between the proteins.

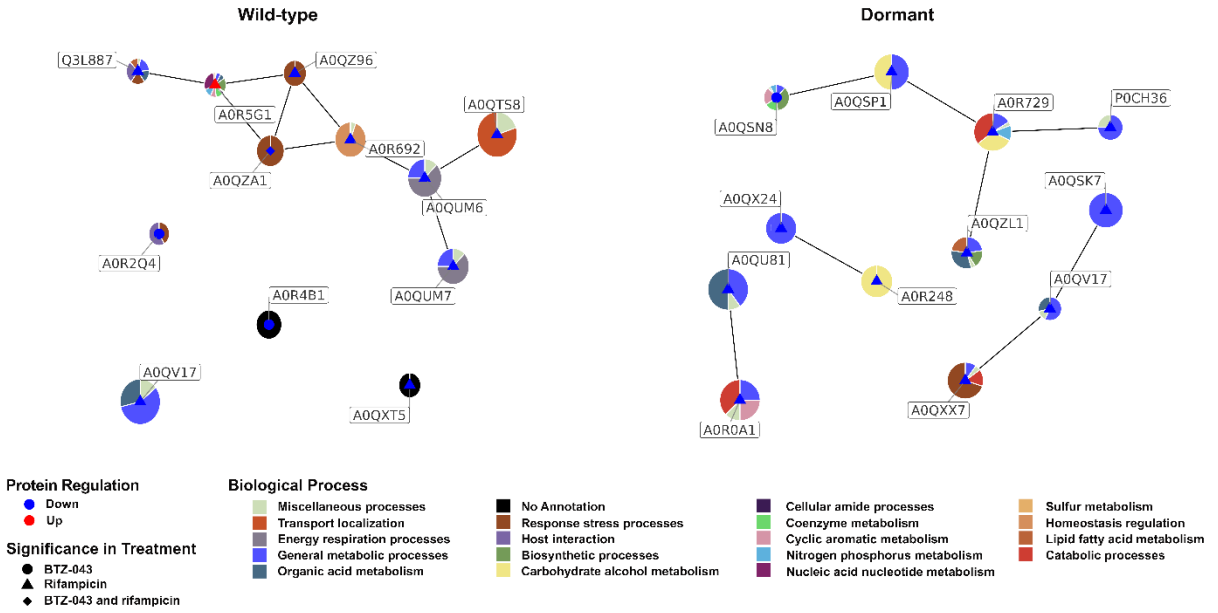


Figure 37: Network analysis of wild-type (left) and dormant (right) *M. smegmatis* proteins being significantly up- or downregulated when treated with rifampicin or BTZ-043.

The network analysis of differentially expressed proteins via TMT quantification reveals distinct patterns between WT and dormant *M. smegmatis*. In the WT condition, a subset of eight regulated proteins forms a connected interaction cluster, whereas four others remain unlinked. In the dormant state, proteins are divided into several smaller clusters (five, three, and two times two nodes). While these patterns do not indicate a specific mechanism, they may reflect differences in stress-related responses. Interestingly, protein A0QV17 was significantly downregulated in both the dormant and WT conditions. While it did not show any further interaction in the WT bacteria, it is connected with two other downregulated proteins (A0QXX7, A0QSK7) in the dormant state. A0QV17 is so far uncharacterized, but it seems to play a role in stress adaptation or metabolic regulation when treated with rifampicin. Due to its interaction with A0QXX7, which is a stress-associated protein, it suggests a transition into non-replicative or dormancy-related states. Besides this, proteins showing a high degree of connectivity to other proteins (hub proteins) could represent key control points in biological pathways. For example, the protein A0R729 in dormant bacteria represents a central point by being connected to three other proteins (P0CH36, A0QZL1, A0QSP1). A0R729 is so far not reviewed by SwissProt, but was described as a glycerol kinase with ATP binding function. Its downregulation, therefore, suggests an adaptive mechanism for energy conservation in the bacteria. Within this interaction network, all proteins except one are downregulated in response to rifampicin, while the

(uncharacterized) protein A0QSN8 is suppressed under BTZ-043 treatment. This pattern suggests that the two antibiotics affect distinct but potentially interacting pathways. Investigating potential synergistic effects between both compounds could reveal complementary mechanisms. In WT bacteria, the big interaction cluster of eight proteins could uncover important metabolic pathway regulations. Especially, the significant upregulation of only one protein (A0R5G1) within a network of otherwise exclusively downregulated proteins may indicate a compensatory activity of A0R5G1 for the metabolic pathways of the other proteins. With this, A0R5G1 may be a potential target for treatment approaches. Moreover, its connection to other, so far functionally unknown proteins (A0QZ96, A0QTS8) may lead to new, unrecognized or regulatory mechanisms. In general, many of the proteins found are not fully characterized yet. Further functional annotation and experimental validation would be required to understand whether these network structures correlate with biological processes relevant to dormancy, persistence, or antibiotic tolerance.

5 Summary and Outlook

5.1 Summary

This thesis contains two main parts. In the first part, intact mass analysis of *M. smegmatis* proteins was performed. For this, protein extraction and HPLC-MS settings were optimized. A lysis buffer containing 25 mM TrisCl, 1 mM EDTA, 2.5 mM DTT, 2 M thiourea, and 7 M urea, combined with bead beating (beads:buffer; 1:3), yielded the highest protein concentrations. For protein purification, acetone precipitation and resolubilization in 80 % FA after a protocol of Doucette *et al*.[42], with subsequent vacuum drying and resolubilization in the initial HPLC solvents (v/v/v; 90 % H_2O, 10 % ACN, 0.1 % FA) was chosen as the preferred approach to yield the highest protein concentrations. Intact mass analysis of *M. smegmatis* wild-type resulted in the detection of 229 proteins. The number is possibly restricted due to the instrument's limitation and only a one-dimensional RPLC setup prior to MS analysis. LC-MS analysis of intact, dormant *M. Smegmatis* proteins revealed 34 proteins, with 24 proteins matching the intact mass (± 65 ppm) of the wild-type proteins. By adapting the untargeted labeling approach from Haisch *et al*., but implementing HPLC-MS for protein analysis, it was possible to see a mass shift of labeled proteins over time. The dynamics of six distinct intact proteins were analyzed in more detail by observing the ΔCOG values of the protein masses over time. The values follow exponential behavior with specific rise times for each of the different proteins. The effect of rifampicin on protein *de novo* synthesis could be detected within about half the regular doubling cycle for *M. smegmatis* (1.5 h). To identify intact proteins, an online fractionation via T-piece splitting over specific time intervals was conducted, while intact mass analysis was performed. Fractions were investigated with bottom-up analysis, and the masses of the identified proteins were assigned to the intact masses. 15 proteins could be identified by mass assignments to the digested proteins of the respective fraction with maximum mass differences of 21.8 Da. The protein A0QV42 (12868 Da) represents a ribosomal protein and was analyzed by untargeted labeling.

In the second part, TMT analysis resulted in eleven significantly down-regulated and one protein being significantly up-regulated in WT bacteria when treated with rifampicin. In dormant bacteria, eleven proteins were significantly downregulated after treatment with rifampicin, and one, the (uncharacterized) protein A0QSN8, after BTZ-043 treatment. Many of the proteins have not been fully characterized yet. Network analysis of differentially expressed

proteins revealed distinct patterns between WT and dormant *M. smegmatis* proteins, with the protein A0QV17 being significantly downregulated in both conditions. In WT bacteria, the significant upregulation of only one protein (A0R5G1) within a network of otherwise exclusively downregulated proteins may indicate a compensatory activity for the metabolic pathways of the other proteins.

5.2 Outlook

Our analysis, especially with the application of intact mass analysis, offers unique insights into comprehensive proteomic profiles. Implementing automated protein detection for the identification of intact proteins would enable broader coverage of the mycobacterial proteome and provide deeper insights into protein dynamics. To improve mass assignment, it is also essential to account for the PTM masses. Employing top-down proteomics, fragmenting intact proteins at the MS2/3 level, could further support the identification of currently undetected proteins. Moreover, the inclusion of a further pre-separation step, e.g., by size exclusion chromatography, could possibly extend the number of detectable proteins. In addition, using a different mass spectrometer might be more suitable, as the TripleTOF 5600 has a limited mass range and is not optimized for intact protein analysis, potentially restricting detection of higher-mass proteoforms. With this, it may be possible to characterize numerous specific proteomic fingerprints as a reaction to different antibiotics, unravelling new mechanistic functions and potential targets. Future studies should apply the untargeted labeling approach combined with HPLC-MS readout to *M. tuberculosis* proteins to further investigate the pathogen's complex and dynamic proteome. Moreover, a wider range of antibiotics and concentrations should be investigated to assess more mechanisms of action, adaptive responses, and potential synergistic effects.

The proteins identified through TMT analysis may represent novel drug targets. Follow-up studies will be needed to determine whether these proteins are linked to biological processes such as dormancy, persistence, or antibiotic tolerance.

6 Abbreviations

ACN	Acetonitrile
AGC	Automatic gain control
AMR	Antimicrobial resistance
ANOVA	Analysis of variance
a. u.	Arbitrary unit
BB	Bead beating
BDQ	Bedaquiline
BEH	Ethylene bridged hybrid
BSL	Bio-safety level
BTZ	Benzothiazinone
BUP	Bottom-up proteomics
CAD	Collisionally activated dissociation
CFU	Colony-forming unit
CI	Confidence interval
CID	Collision-induced dissociation
COG	Center of gravity
CUR	Curtain gas
DDA	Data-dependent acquisition
DMSO	Dimethyl sulfoxide

DNA	Desoxyribonucleic acid
DTT	Dithiothreitol
ECD	Electron-capture dissociation
ESI	Electrospray ionization
ETD	Electron-transfer dissociation
FA	Formic acid
FC	Fold change
FT-ICR-MS	Fourier-transform ion cyclotron resonance mass spectrometer
FWHM	Full width at half maximum
GS1	Ion source gas 1
GS2	Ion source gas 2
h	Hours
HCD	Higher-energy collisional dissociation
HPLC	High-performance liquid chromatography
IAA	Iodoacetic acid
ID	Inner diameter
IDA	Information-dependent acquisition
ISVF	Ion spray voltage floating
LC	Liquid chromatography

MALDI-TOF	Matrix-assisted laser desorption−ionization time-of-flight
MDR-TB	Multi-drug-resistant tuberculosis
MIC	Minimum inhibitory concentration
min	Minutes
MR	Relative molecular mass
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
M. sm	*Mycobacterium smegmatis*
M. smegmatis	*Mycobacterium smegmatis*
M. tb	*Mycobacterium tuberculosis*
MTBC	*Mycobacterium tuberculosis* complex
M. tuberculosis	*Mycobacterium tuberculosis*
MW	Molecular weight
MWCO	Molecular weight cut-off
m/z	Mass-to-charge ratio
NP	Normal phase
OADC	Oleic albumin dextrose catalase
OD 600	Optical density at a wavelength of 600 nm
PBS	Phosphate-buffered saline

PG	Peptidoglycan
PTM	Post-translational modification
R	Resolution
RIF	Rifampicin
RNA	Ribonucleic acid
RP	Reversed phase
RPLC	Reversed-phase liquid chromatography
RT	Retention time
SD	Standard deviation
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate – polyacrylamide gel electrophoresis
SEM	Scanning electron microscopy
SILAC	Stable isotope labeling by/with amino acids in cell culture
SWATH	Sequential window acquisition of all theoretical fragment ions
TB	Tuberculosis
TDP	Top-down proteomics
TEM	Temperature
TEMED	*N,N,N′,N′*-Tetramethylethylendiamine
TFA	Trifluoroacetic acid

TIC	Total ion chromatogram
TMT	Tandem mass tag
T80	Tween-80
VBNC	Viable but not culturable
WHO	World Health Organization
WT	Wild-type
XDR-TB	Extensively drug-resistant tuberculosis

7 References

(1) Murray, C. J. L.; Ikuta, K. S.; Sharara, F.; Swetschinski, L.; Robles Aguilar, G.; Gray, A.; Han, C.; Bisignano, C.; Rao, P.; Wool, E.; et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet* **2022**, *399* (10325), 629-655. DOI: 10.1016/S0140-6736(21)02724-0 (accessed 26 February 2025).

(2) World Health Organization. *Global tuberculosis report 2024*; 2024.

(3) Haisch, C.; Neumann-Cip, A.-C.; Imhof, A.; Schmidt, A.; Forne, I.; Hoelscher, M.; Wieser, A. Dynamic Measurement of Protein Translation in Mycobacteria Using Nontargeted Stable Isotope Labeling in Combination with MALDI-TOF Mass Spectrometry-Based Readout. *Analytical Chemistry* **2025**. DOI: 10.1021/acs.analchem.4c03931.

(4) Li, J.; Cai, Z.; Bomgarden, R. D.; Pike, I.; Kuhn, K.; Rogers, J. C.; Roberts, T. M.; Gygi, S. P.; Paulo, J. A. TMTpro-18plex: the expanded and complete set of TMTpro reagents for sample multiplexing. *Journal of proteome research* **2021**, *20* (5), 2964-2972.

(5) Gagneux, S. Ecology and evolution of *Mycobacterium tuberculosis*. *Nature Reviews Microbiology* **2018**, *16* (4), 202-213.

(6) Armstrong, D. T.; Eisemann, E.; Parrish, N. A brief update on mycobacterial taxonomy, 2020 to 2022. *Journal of clinical microbiology* **2023**, *61* (4), e00331-00322.

(7) Qin, M.-H.; Madiraju, M. V.; Rajagopalan, M. Characterization of the functional replication origin of *Mycobacterium tuberculosis*. *Gene* **1999**, *233* (1-2), 121-130.

(8) Sakula, A. Robert Koch: centenary of the discovery of the tubercle bacillus, 1882. *Thorax* **1982**, *37* (4), 246-251.

(9) Hett, E. C.; Rubin, E. J. Bacterial growth and cell division: a mycobacterial perspective. *Microbiology and Molecular Biology Reviews* **2008**, *72* (1), 126-156.

(10) Jacobo-Delgado, Y.; Rodríguez-Carlos, A.; Serrano, C.; Rivas-Santiago, B. *Mycobacterium tuberculosis* cell-wall and antimicrobial peptides: a mission impossible? Front Immunol 14: 1194923. 2023.

(11) Sparks, I. L.; Derbyshire, K. M.; Jacobs Jr, W. R.; Morita, Y. S. Mycobacterium smegmatis: the vanguard of mycobacterial research. *Journal of bacteriology* **2023**, *205* (1), e00337-00322.

(12) Smith, I. *Mycobacterium tuberculosis* pathogenesis and molecular determinants of virulence. *Clinical microbiology reviews* **2003**, *16* (3), 463-496.

(13) NCBI RefSeq. Genes annotated on *Mycolicibacterium smegmatis* MC2 155 ASM430705v1 (GCF_004307055.1). 2019.

(14) NCBI RefSeq. Genes annotated on *Mycobacterium tuberculosis* H37Rv ASM19595v2 (GCF_000195955.2). 2013.

(15) Neumann, A. C.; Bauer, D.; Hoelscher, M.; Haisch, C.; Wieser, A. Identifying Dormant Growth State of Mycobacteria by Orthogonal Analytical Approaches on a Single Cell and Ensemble Basis. *Analytical Chemistry* **2019**, *91* (1), 881-887. DOI: 10.1021/acs.analchem.8b03646 From NLM Medline.

(16) Rabodoarivelo, M. S.; Aerts, M.; Vandamme, P.; Palomino, J. C.; Rasolofo, V.; Martin, A. Optimizing of a protein extraction method for *Mycobacterium tuberculosis* proteome analysis using mass spectrometry. *Journal of microbiological methods* **2016**, *131*, 144-147.

(17) Danelishvili, L.; Shulzhenko, N.; Chinison, J. J.; Babrak, L.; Hu, J.; Morgun, A.; Burrows, G.; Bermudez, L. E. *Mycobacterium tuberculosis* proteome response to antituberculosis compounds reveals metabolic "escape" pathways that prolong bacterial survival. *Antimicrobial Agents and Chemotherapy* **2017**, *61* (7), 10.1128/aac. 00430-00417.

(18) Luca, S.; Mihaescu, T. History of BCG vaccine. *Maedica* **2013**, *8* (1), 53-58.

(19) Lugosi, L. Theoretical and methodological aspects of BCG vaccine from the discovery of Calmette and Guerin to molecular biology. A review. *Tubercle and Lung Disease* **1992**, *73* (5), 252-261.

(20) Organization, W. H. The End TB Strategy (World Health Organization). **2014**.

(21) Drapal, M.; Fraser, P. D. Metabolite Profiling: A Tool for the Biochemical Characterisation of *Mycobacterium sp*. *Microorganisms* **2019**, *7* (5), 148.

(22) World Health Organization. *Global tuberculosis report 2015*; 2015.

(23) World Health Organization. *Global tuberculosis report 2023*; 2023.

(24) Jarlier, V.; Nikaido, H. Mycobacterial cell wall: structure and role in natural resistance to antibiotics. *FEMS microbiology letters* **1994**, *123* (1-2), 11-18.

(25) Bansal-Mutalik, R.; Nikaido, H. Mycobacterial outer membrane is a lipid bilayer and the inner membrane is unusually rich in diacyl phosphatidylinositol dimannosides. *Proceedings of the National Academy of Sciences* **2014**, *111* (13), 4958-4963.

(26) Fishbein, S. R.; Tomasi, F. G.; Wolf, I. D.; Dulberger, C. L.; Wang, A.; Keshishian, H.; Wallace, L.; Carr, S. A.; Ioerger, T. R.; Rego, E. H. The conserved translation factor LepA is required for optimal synthesis of a porin family in *Mycobacterium smegmatis*. *Journal of Bacteriology* **2021**, *203* (6), 10.1128/jb. 00604-00620.

(27) Chao, M. C.; Rubin, E. J. Letting sleeping dos lie: does dormancy play a role in tuberculosis? *Annual review of microbiology* **2010**, *64* (1), 293-311.

(28) McCune Jr, R. M.; Tompsett, R. Fate of *Mycobacterium tuberculosis* in mouse tissues as determined by the microbial enumeration technique: I. The persistence of drug-susceptible tubercle bacilli in the tissues despite prolonged antimicrobial therapy. *The Journal of experimental medicine* **1956**, *104* (5), 737-762.

(29) McCune Jr, R. M.; Tompsett, R.; McDermott, W. The fate of *Mycobacterium tuberculosis* in mouse tissues as determined by the microbial enumeration technique: II. The conversion of tuberculous infection to the latent state by the administration of pyrazinamide and a companion drug. *The Journal of experimental medicine* **1956**, *104* (5), 763-802.

(30) Xu, H.-S.; Roberts, N.; Singleton, F.; Attwell, R.; Grimes, D. J.; Colwell, R. Survival and viability of nonculturable *Escherichia coli* and *Vibrio cholerae* in the estuarine and marine environment. *Microbial ecology* **1982**, *8*, 313-323.

(31) Scanga, C. A.; Mohan, V.; Joseph, H.; Yu, K.; Chan, J.; Flynn, J. L. Reactivation of latent tuberculosis: variations on the Cornell murine model. *Infection and immunity* **1999**, *67* (9), 4531-4538.

(32) Li, L.; Mendis, N.; Trigui, H.; Oliver, J. D.; Faucher, S. P. The importance of the viable but non-culturable state in human bacterial pathogens. *Frontiers in microbiology* **2014**, *5*, 258.

(33) Lipworth, S.; Hammond, R.; Baron, V.; Hu, Y.; Coates, A.; Gillespie, S. Defining dormancy in mycobacterial disease. *Tuberculosis* **2016**, *99*, 131-142.

(34) Zhang, Y. Persistent and dormant tubercle bacilli and latent tuberculosis. *Front Biosci* **2004**, *9* (1-3), 1136-1156.

(35) Gengenbacher, M.; Kaufmann, S. H. *Mycobacterium tuberculosis*: success through dormancy. *FEMS microbiology reviews* **2012**, *36* (3), 514-532.

(36) Rabilloud, T. Use of thiourea to increase the solubility of membrane proteins in two‐dimensional electrophoresis. *Electrophoresis* **1998**, *19* (5), 758-760.

(37) Betts, J. C.; Dodson, P.; Quan, S.; Lewis, A. P.; Thomas, P. J.; Duncan, K.; McAdam, R. A. Comparison of the proteome of *Mycobacterium tuberculosis* strain H37Rv with clinical isolate CDC 1551. *Microbiology* **2000**, *146* (12), 3205-3216.

(38) Rezwan, M.; Lanéelle, M.-A.; Sander, P.; Daffé, M. Breaking down the wall: fractionation of mycobacteria. *Journal of microbiological methods* **2007**, *68* (1), 32-39.

(39) Odumeru, J.; Gao, A.; Chen, S.; Raymond, M.; Mutharia, L. Use of the bead beater for preparation of *Mycobacterium paratuberculosis* template DNA in milk. *Canadian Journal of Veterinary Research* **2001**, *65* (4), 201.

(40) Melby, J. A.; Roberts, D. S.; Larson, E. J.; Brown, K. A.; Bayne, E. F.; Jin, S.; Ge, Y. Novel strategies to address the challenges in top-down proteomics. *Journal of the American Society for Mass Spectrometry* **2021**, *32* (6), 1278-1294.

(41) Donnelly, D. P.; Rawlins, C. M.; DeHart, C. J.; Fornelli, L.; Schachner, L. F.; Lin, Z.; Lippens, J. L.; Aluri, K. C.; Sarin, R.; Chen, B. Best practices and benchmarks for intact protein analysis for top-down mass spectrometry. *Nature methods* **2019**, *16* (7), 587-594.

(42) Doucette, A. A.; Vieira, D. B.; Orton, D. J.; Wall, M. J. Resolubilization of precipitated intact membrane proteins with cold formic acid for analysis by mass spectrometry. *Journal of Proteome Research* **2014**, *13* (12), 6001-6012.

(43) Sharma, P.; Kumar, B.; Singhal, N.; Katoch, V. M.; Venkatesan, K.; Chauhan, D. S.; Bisht, D. Streptomycin induced protein expression analysis in *Mycobacterium tuberculosis* by two-dimensional gel electrophoresis & mass spectrometry. *Indian Journal of Medical Research* **2010**, *132* (4), 400-408.

(44) Sharma, D.; Gautam, S.; Srivastava, N.; Khan, A. M.; Bisht, D. Comparative Proteomic Analysis of Capsule Proteins in Aminoglycoside-Resistant and Sensitive *Mycobacterium tuberculosis* Clinical Isolates: Unraveling Potential Drug Targets. *The International Journal of Mycobacteriology* **2024**, *13* (2), 197-205.

(45) De Souza, E. V.; Dalberto, P. F.; Miranda, A. C.; Saghatelian, A.; Pinto, A. M.; Basso, L. A.; Machado, P.; Bizarro, C. V. Large-scale proteogenomics characterization of microproteins in *Mycobacterium tuberculosis*. *Scientific Reports* **2024**, *14* (1), 31186.

(46) Ali, S.; Ehtram, A.; Arora, N.; Manjunath, P.; Roy, D.; Ehtesham, N. Z.; Hasnain, S. E. The M. tuberculosis Rv1523 methyltransferase promotes drug resistance through methylation-mediated cell wall remodeling and modulates macrophages immune responses. *Frontiers in Cellular and Infection Microbiology* **2021**, *11*, 622487.

(47) Bendre, A. D.; Peters, P. J.; Kumar, J. Recent Insights into the Structure and Function of Mycobacterial Membrane Proteins Facilitated by Cryo-EM. *J Membr Biol* **2021**, *254* (3), 321-341. DOI: 10.1007/s00232-021-00179-w From NLM Medline.

(48) Gopinath, V.; Raghunandanan, S.; Gomez, R. L.; Jose, L.; Surendran, A.; Ramachandran, R.; Pushparajan, A. R.; Mundayoor, S.; Jaleel, A.; Kumar, R. A. Profiling the proteome of *Mycobacterium tuberculosis* during dormancy and reactivation. *Molecular & cellular proteomics* **2015**, *14* (8), 2160-2176.

(49) Schubert, O. T.; Ludwig, C.; Kogadeeva, M.; Zimmermann, M.; Rosenberger, G.; Gengenbacher, M.; Gillet, L. C.; Collins, B. C.; Röst, H. L.; Kaufmann, S. H. Absolute proteome composition and dynamics during dormancy and resuscitation of *Mycobacterium tuberculosis*. *Cell host & microbe* **2015**, *18* (1), 96-108.

(50) Sakatos, A.; Babunovic, G. H.; Chase, M. R.; Dills, A.; Leszyk, J.; Rosebrock, T.; Bryson, B.; Fortune, S. M. Posttranslational modification of a histone-like protein regulates phenotypic resistance to isoniazid in mycobacteria. *Science advances* **2018**, *4* (5), eaao1478.

(51) Catherman, A. D.; Skinner, O. S.; Kelleher, N. L. Top down proteomics: facts and perspectives. *Biochemical and biophysical research communications* **2014**, *445* (4), 683-693.

(52) Staub, A.; Guillarme, D.; Schappler, J.; Veuthey, J.-L.; Rudaz, S. Intact protein analysis in the biopharmaceutical field. *Journal of pharmaceutical and biomedical analysis* **2011**, *55* (4), 810-822.

(53) Fekete, S.; Murisier, A.; Beck, A.; Lawhorn, J.; Ritchie, H.; Boyes, B.; Guillarme, D. New wide-pore superficially porous stationary phases with low hydrophobicity applied for the analysis of monoclonal antibodies. *Journal of Chromatography A* **2021**, *1642*, 462050.

(54) Villacrés, C.; Mizero, B.; Spicer, V.; Viner, R.; Saba, J.; Patel, B.; Snovida, S.; Jensen, P.; Huhmer, A.; Krokhin, O. V. Toward an Ultimate Solution for Peptide Retention Time Prediction: The Effect of Column Temperature on Separation Selectivity. *Journal of Proteome Research* **2024**, *23* (4), 1488-1494.

(55) Fekete, S.; Horváth, K.; Guillarme, D. Influence of pressure and temperature on molar volume and retention properties of peptides in ultra-high pressure liquid chromatography. *Journal of Chromatography A* **2013**, *1311*, 65-71.

(56) Wang, Y.; Balgley, B. M.; Rudnick, P. A.; Lee, C. S. Effects of chromatography conditions on intact protein separations for top-down proteomics. *Journal of Chromatography A* **2005**, *1073* (1-2), 35-41.

(57) Greibrokk, T.; Andersen, T. High-temperature liquid chromatography. *Journal of Chromatography A* **2003**, *1000* (1-2), 743-755.

(58) García, M.; Hogenboom, A.; Zappey, H.; Irth, H. Effect of the mobile phase composition on the separation and detection of intact proteins by reversed-phase liquid chromatography–electrospray mass spectrometry. *Journal of Chromatography A* **2002**, *957* (2), 187-199.

(59) Maciel, E. V. S.; de Toffoli, A. L.; Sobieski, E.; Nazário, C. E. D.; Lancas, F. M. Miniaturized liquid chromatography focusing on analytical columns and mass spectrometry: A review. *Analytica chimica acta* **2020**, *1103*, 11-31.

(60) Boychenko, A.; Meding, S.; Decrop, W.; Ruehl, M.; Swart, R. Capillary-flow LC-MS: combining high sensitivity, robustness, and throughput. TN: 2017.

(61) Wilson, S. R.; Vehus, T.; Berg, H. S.; Lundanes, E. Nano-LC in proteomics: recent advances and approaches. *Bioanalysis* **2015**, *7* (14), 1799-1815.

(62) Carr, C. D. Recent developments in column technologies for the analysis of proteins and peptides. *LC GC Magazine-North America-Solutions for Separation Scientists* **2006**, *29*, 44-48.

(63) Chen, B.; Brown, K. A.; Lin, Z.; Ge, Y. Top-down proteomics: ready for prime time? *Analytical Chemistry* **2017**, *90* (1), 110-127.

(64) Bobály, B.; Sipkó, E.; Fekete, J. Challenges in liquid chromatographic characterization of proteins. *Journal of Chromatography B* **2016**, *1032*, 3-22.

(65) Jones, A. W.; Cooper, H. J. Dissociation techniques in mass spectrometry-based proteomics. *Analyst* **2011**, *136* (17), 3419-3429.

(66) Roberts, D. S.; Loo, J. A.; Tsybin, Y. O.; Liu, X.; Wu, S.; Chamot-Rooke, J.; Agar, J. N.; Paša-Tolić, L.; Smith, L. M.; Ge, Y. Top-down proteomics. *Nature Reviews Methods Primers* **2024**, *4* (1), 38.

(67) Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M. Higher-energy C-trap dissociation for peptide modification analysis. *Nature methods* **2007**, *4* (9), 709-712.

(68) Smith, L. M.; Kelleher, N. L. Proteoform: a single term describing protein complexity. *Nature methods* **2013**, *10* (3), 186-187.

(69) Hervey; Strader, M. B.; Hurst, G. B. Comparison of digestion protocols for microgram quantities of enriched protein samples. *Journal of proteome research* **2007**, *6* (8), 3054-3061.

(70) Hildonen, S.; Halvorsen, T. G.; Reubsaet, L. Why less is more when generating tryptic peptides in bottom‐up proteomics. *Proteomics* **2014**, *14* (17-18), 2031-2041.

(71) Strader, M. B.; Tabb, D. L.; Hervey, W. J.; Pan, C.; Hurst, G. B. Efficient and specific trypsin digestion of microgram to nanogram quantities of proteins in organic− aqueous solvent systems. *Analytical Chemistry* **2006**, *78* (1), 125-134.

(72) Dupree, E. J.; Jayathirtha, M.; Yorkey, H.; Mihasan, M.; Petre, B. A.; Darie, C. C. A critical review of bottom-up proteomics: the good, the bad, and the future of this field. *Proteomes* **2020**, *8* (3), 14.

(73) Ntai, I.; LeDuc, R. D.; Fellers, R. T.; Erdmann-Gilmore, P.; Davies, S. R.; Rumsey, J.; Early, B. P.; Thomas, P. M.; Li, S.; Compton, P. D. Integrated bottom-up and top-down proteomics of patient-derived breast tumor xenografts. *Molecular & Cellular Proteomics* **2016**, *15* (1), 45-56.

(74) Krasny, L.; Huang, P. Data-independent acquisition mass spectrometry (DIA-MS) for proteomic applications in oncology. Mol Omics 17: 29–42. 2021.

(75) Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Molecular & Cellular Proteomics* **2012**, *11* (6).

(76) Ludwig, C.; Gillet, L.; Rosenberger, G.; Amon, S.; Collins, B. C.; Aebersold, R. Data‐independent acquisition‐based SWATH‐MS for quantitative proteomics: a tutorial. *Molecular systems biology* **2018**, *14* (8), e8126.

(77) Thompson, A.; Schäfer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Hamon, C. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Analytical Chemistry* **2003**, *75* (8), 1895-1904.

(78) Rauniyar, N.; Yates III, J. R. Isobaric labeling-based relative quantification in shotgun proteomics. *Journal of proteome research* **2014**, *13* (12), 5293-5309.

(79) Weiner, S.; Sauer, M.; Visser, P. J.; Tijms, B. M.; Vorontsov, E.; Blennow, K.; Zetterberg, H.; Gobom, J. Optimized sample preparation and data analysis for TMT proteomic analysis of cerebrospinal fluid applied to the identification of Alzheimer's disease biomarkers. *Clinical Proteomics* **2022**, *19* (1), 13.

(80) Sinclair, J.; Timms, J. F. Quantitative profiling of serum samples using TMT protein labelling, fractionation and LC–MS/MS. *Methods* **2011**, *54* (4), 361-369.

(81) Dayon, L.; Hainard, A.; Licker, V.; Turck, N.; Kuhn, K.; Hochstrasser, D. F.; Burkhard, P. R.; Sanchez, J.-C. Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags. *Analytical Chemistry* **2008**, *80* (8), 2921-2931.

(82) Murphy, J. P.; Stepanova, E.; Everley, R. A.; Paulo, J. A.; Gygi, S. P. Comprehensive temporal protein dynamics during the diauxic shift in Saccharomyces cerevisiae. *Molecular & Cellular Proteomics* **2015**, *14* (9), 2454-2465.

(83) Perez-Perri, J. I.; Noerenberg, M.; Kamel, W.; Lenz, C. E.; Mohammed, S.; Hentze, M. W.; Castello, A. Global analysis of RNA-binding protein dynamics by comparative and enhanced RNA interactome capture. *Nature protocols* **2021**, *16* (1), 27-60.

(84) Schweppe, D. K.; Rusin, S. F.; Gygi, S. P.; Paulo, J. A. Optimized workflow for multiplexed phosphorylation analysis of TMT-labeled peptides using high-field asymmetric waveform ion mobility spectrometry. *Journal of proteome research* **2019**, *19* (1), 554-560.

(85) Yu, K.; Wang, Z.; Wu, Z.; Tan, H.; Mishra, A.; Peng, J. High-throughput profiling of proteome and posttranslational modifications by 16-Plex TMT labeling and mass spectrometry. In *Quantitative Methods in Proteomics*, Springer, 2021; pp 205-224.

(86) Romanick, S. S.; Ulrich, C.; Schlauch, K.; Hostler, A.; Payne, J.; Woolsey, R.; Quilici, D.; Feng, Y.; Ferguson, B. S. Obesity-mediated regulation of cardiac protein acetylation: parallel analysis of total and acetylated proteins via TMT-tagged mass spectrometry. *Bioscience reports* **2018**, *38* (5), BSR20180721.

(87) Phaneuf, C. G.; Aizikov, K.; Grinfeld, D.; Kreutzmann, A.; Mourad, D.; Lange, O.; Dai, D.; Zhang, B.; Belenky, A.; Makarov, A. A. Experimental strategies to improve drug-target identification in mass spectrometry-based thermal stability assays. *Communications Chemistry* **2023**, *6* (1), 64.

(88) Johnson, F. D.; Hughes, C. S.; Liu, A.; Lockwood, W. W.; Morin, G. B. Tandem mass tag-based thermal proteome profiling for the discovery of drug-protein interactions in cancer cells. *STAR protocols* **2023**, *4* (1), 102012.

(89) Li, J.; Van Vranken, J. G.; Pontano Vaites, L.; Schweppe, D. K.; Huttlin, E. L.; Etienne, C.; Nandhikonda, P.; Viner, R.; Robitaille, A. M.; Thompson, A. H. TMTpro reagents: a set of isobaric labeling mass tags enables simultaneous proteome-wide measurements across 16 samples. *Nature methods* **2020**, *17* (4), 399-404.

(90) McAlister, G. C.; Huttlin, E. L.; Haas, W.; Ting, L.; Jedrychowski, M. P.; Rogers, J. C.; Kuhn, K.; Pike, I.; Grothe, R. A.; Blethrow, J. D. Increasing the multiplexing capacity of TMTs using reporter ion isotopologues with isobaric masses. *Analytical chemistry* **2012**, *84* (17), 7469-7478.

(91) Pichler, P.; Köcher, T.; Holzmann, J.; Möhring, T.; Ammerer, G.; Mechtler, K. Improved precision of iTRAQ and TMT quantification by an axial extraction field in an Orbitrap HCD cell. *Analytical Chemistry* **2011**, *83* (4), 1469-1474.

(92) Dayon, L.; Pasquarello, C.; Hoogland, C.; Sanchez, J.-C.; Scherl, A. Combining low-and high-energy tandem mass spectra for optimized peptide quantification with isobaric tags. *Journal of proteomics* **2010**, *73* (4), 769-777.

(93) Paulo, J. A.; McAllister, F. E.; Everley, R. A.; Beausoleil, S. A.; Banks, A. S.; Gygi, S. P. Effects of MEK inhibitors GSK1120212 and PD0325901 in vivo using 10 ‐ plex quantitative proteomics and phosphoproteomics. *Proteomics* **2015**, *15* (2-3), 462-473.

(94) Thompson, A.; Wölmer, N.; Koncarevic, S.; Selzer, S.; Böhm, G.; Legner, H.; Schmid, P.; Kienle, S.; Penning, P.; Höhle, C. TMTpro: design, synthesis, and initial evaluation of a proline-based isobaric 16-plex tandem mass tag reagent set. *Analytical Chemistry* **2019**, *91* (24), 15941-15950.

(95) Brown, K. A.; Melby, J. A.; Roberts, D. S.; Ge, Y. Top-down proteomics: challenges, innovations, and applications in basic and clinical research. *Expert review of proteomics* **2020**, *17* (10), 719-733.

(96) Melby, J. A.; Brown, K. A.; Gregorich, Z. R.; Roberts, D. S.; Chapman, E. A.; Ehlers, L. E.; Gao, Z.; Larson, E. J.; Jin, Y.; Lopez, J. R. High sensitivity top–down proteomics captures single muscle cell heterogeneity in large proteoforms. *Proceedings of the National Academy of Sciences* **2023**, *120* (19), e2222081120.

(97) Su, P.; Hollas, M. A.; Rubakhin, S.; Butun, F. A.; Greer, J. B.; Early, B. P.; Fellers, R. T.; Caldwell, M. A.; Sweedler, J. V.; Kafader, J. O. Top-down proteomics of 10,000 single brain cells. *bioRxiv* **2023**, 2023.2005. 2031.543176.

(98) Soares, R.; Franco, C.; Pires, E.; Ventosa, M.; Palhinhas, R.; Koci, K.; de Almeida, A. M.; Coelho, A. V. Mass spectrometry and animal science: protein identification strategies and particularities of farm animal species. *Journal of proteomics* **2012**, *75* (14), 4190-4206.

(99) Trauger, S. A.; Webb, W.; Siuzdak, G. Peptide and protein analysis with mass spectrometry. *Journal of Spectroscopy* **2002**, *16* (1), 15-28.

(100) Rehm, H.; Letzel, T. In *Der Experimentator: Proteinbiochemie/Proteomics*, 2016; pp 267-326.

(101) Strupat, K. Molecular weight determination of peptides and proteins by ESI and MALDI. *Methods in enzymology* **2005**, *405*, 1-36.

(102) Yergey, J.; Heller, D.; Hansen, G.; Cotter, R. J.; Fenselau, C. Isotopic distributions in mass spectra of large molecules. *Analytical Chemistry* **1983**, *55* (2), 353-356.

(103) Xu, G.; Stupak, J.; Yang, L.; Hu, L.; Guo, B.; Li, J. Deconvolution in mass spectrometry based proteomics. *Rapid Communications in Mass Spectrometry* **2018**, *32* (10), 763-774.

(104) Matthiesen, R. Methods, algorithms and tools in computational proteomics: a practical point of view. *Proteomics* **2007**, *7* (16), 2815-2832.

(105) Shannon, C. E. A mathematical theory of communication. *The Bell system technical journal* **1948**, *27* (3), 379-423.

(106) Eshima, N. *Statistical data analysis and entropy*; Springer, 2020.

(107) Jaynes, E. T. Information theory and statistical mechanics. *Physical review* **1957**, *106* (4), 620.

(108) Ferrige, A.; Seddon, M.; Jarvis, S.; Skilling, J.; Aplin, R. Maximum entropy deconvolution in electrospray mass spectrometry. *Rapid communications in mass spectrometry* **1991**, *5* (8), 374-377.

(109) Ferrige, A.; Seddon, M.; Green, B.; Jarvis, S.; Skilling, J.; Staunton, J. Disentangling electrospray spectra with maximum entropy. *Rapid Communications in Mass Spectrometry* **1992**, *6* (11), 707-711.

(110) Marty, M. T.; Baldwin, A. J.; Marklund, E. G.; Hochberg, G. K.; Benesch, J. L.; Robinson, C. V. Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. *Analytical Chemistry* **2015**, *87* (8), 4370-4376.

(111) Gelman, A.; Carlin, J. B.; Stern, H. S.; Dunson, D. B.; Vehtari, A.; Rubin, D. B. *Bayesian Data Analysis*; CRC Press LLC, 2013.

(112) Kostelic, M. M.; Marty, M. T. Deconvolving native and intact protein mass spectra with UniDec. In *Proteoform Identification: Methods and Protocols*, Springer, 2022; pp 159-180.

(113) Phung, W.; Bakalarski, C. E.; Hinkle, T. B.; Sandoval, W.; Marty, M. T. UniDec processing pipeline for rapid analysis of biotherapeutic mass spectrometry data. *Analytical Chemistry* **2023**, *95* (30), 11491-11498.

(114) Chen, X.; Wei, S.; Ji, Y.; Guo, X.; Yang, F. Quantitative proteomics using SILAC: Principles, applications, and developments. *Proteomics* **2015**, *15* (18), 3175-3192.

(115) Dittmar, G.; Selbach, M. SILAC for biomarker discovery. *PROTEOMICS–Clinical Applications* **2015**, *9* (3-4), 301-306.

(116) Boisvert, F.-M.; Ahmad, Y.; Gierliński, M.; Charrière, F.; Lamont, D.; Scott, M.; Barton, G.; Lamond, A. I. A quantitative spatial proteomics analysis of proteome turnover in human cells. *Molecular & Cellular Proteomics* **2012**, *11* (3).

(117) Pratt, J. M.; Petty, J.; Riba-Garcia, I.; Robertson, D. H.; Gaskell, S. J.; Oliver, S. G.; Beynon, R. J. Dynamics of protein turnover, a missing dimension in proteomics. *Molecular & Cellular Proteomics* **2002**, *1* (8), 579-591.

(118) Huenten, S.; Kaller, M.; Drepper, F.; Oeljeklaus, S.; Bonfert, T.; Erhard, F.; Dueck, A.; Eichner, N.; Friedel, C. C.; Meister, G. p53-regulated networks of protein, mRNA, miRNA, and lncRNA expression revealed by integrated pulsed stable isotope labeling with amino acids in cell culture (pSILAC) and next generation sequencing (NGS) analyses. *Molecular & cellular proteomics* **2015**, *14* (10), 2609-2629.

(119) Deng, W.; Hong, B. Y.; de Hoog, C. L.; Stoynov, N.; Li, Y.; Foster, L. J.; Finlay, B. B. Quantitative proteomic analysis of type III secretome of enteropathogenic *Escherichia coli* reveals an expanded effector repertoire for attaching/effacing bacterial pathogens. *Molecular & Cellular Proteomics* **2012**, *11* (9), 692-709.

(120) Albrethsen, J. Reproducibility in protein profiling by MALDI-TOF mass spectrometry. *Clinical chemistry* **2007**, *53* (5), 852-858.

(121) Remoortere, A. v.; Van Zeijl, R. J.; Van den Oever, N.; Franck, J.; Longuespée, R.; Wisztorski, M.; Salzet, M.; Deelder, A. M.; Fournier, I.; McDonnell, L. A. MALDI imaging and profiling MS of higher mass proteins from tissue. *Journal of the American Society for Mass Spectrometry* **2010**, *21*, 1922-1929.

(122) Wiegand, I.; Hilpert, K.; Hancock, R. E. Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nature protocols* **2008**, *3* (2), 163-175.

(123) Kudykina, Y. K.; Shleeva, M.; Artsabanov, V. Y.; Suzina, N.; Kaprelyants, A. Generation of dormant forms by *Mycobacterium smegmatis* in the poststationary phase during gradual acidification of the medium. *Microbiology* **2011**, *80*, 638-649.

(124) Kulak, N. A.; Pichler, G.; Paron, I.; Nagaraj, N.; Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nature methods* **2014**, *11* (3), 319-324.

(125) Orsburn, B. C. Proteome discoverer—a community enhanced data processing suite for protein informatics. *Proteomes* **2021**, *9* (1), 15.

(126) Cox, J.; Mann, M. MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology* **2008**, *26* (12), 1367-1372.

(127) Team, R. D. C. R: A language and environment for statistical computing. *(No Title)* **2010**.

(128) Altogen Labs. *Protein molecular weight calculator*. n.d. https://altogenlabs.com/resources/protein-molecular-weight-calculator/ (accessed 4 April 2025).

(129) Zhang, X.; Smits, A. H.; Van Tilburg, G. B.; Ovaa, H.; Huber, W.; Vermeulen, M. Proteome-wide identification of ubiquitin interactions using UbIA-MS. *Nature protocols* **2018**, *13* (3), 530-550.

(130) Leek, J. T.; Johnson, W. E.; Parker, H. S.; Jaffe, A. E.; Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **2012**, *28* (6), 882-883.

(131) Szklarczyk, D.; Kirsch, R.; Koutrouli, M.; Nastou, K.; Mehryary, F.; Hachilif, R.; Gable, A. L.; Fang, T.; Doncheva, N. T.; Pyysalo, S. The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic acids research* **2023**, *51* (D1), D638-D646.

(132) Csardi, G.; Nepusz, T. The igraph software. *Complex syst* **2006**, *1695*, 1-9.

(133) Pedersen, T. L. ggraph: An implementation of grammar of graphics for graphs and networks. *R package version* **2020**, *2* (3), 1.

(134) Elzhov, T. V.; Mullen, K. M.; Spiess, A.-N.; Bolker, B.; Mullen, M. K. M.; Suggests, M. Package 'minpack. lm'. *Title R Interface Levenberg-Marquardt Nonlinear Least-Sq. Algorithm Found MINPACK Plus Support Bounds* **2016**.

(135) Wickham, H. dplyr: A grammar of data manipulation. *R package version 04.* **2015**, *3*, p156.

(136) Wickham, H.; Vaughan, D.; Girlich, M. tidyr: Tidy Messy Data. R package version 1.1. 3. *CRAN. R-project. org/package= tidyr* **2020**.

(137) Wickham, H. Toolbox. In *ggplot2: Elegant Graphics for Data Analysis*, Springer, 2016; pp 33-74.

(138) Neuwirth, E. RColorBrewer: colorbrewer palettes. **2014**.

(139) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J. A cross-platform toolkit for mass spectrometry and proteomics. *Nature biotechnology* **2012**, *30* (10), 918-920.

(140) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Analytical chemistry* **2006**, *78* (3), 779-787.

(141) Tautenhahn, R.; Böttcher, C.; Neumann, S. Highly sensitive feature detection for high resolution LC/MS. *BMC bioinformatics* **2008**, *9*, 1-16.

(142) Benton, H. P.; Want, E. J.; Ebbels, T. M. Correction of mass calibration gaps in liquid chromatography–mass spectrometry metabolomics data. *Bioinformatics* **2010**, *26* (19), 2488-2489.

(143) SCIEX. *TripleTOF 5600/5600+ System – Benutzerhandbuch*. SCIEX, 2022. https://sciex.com/content/dam/SCIEX/pdf/customer-docs/user-guide/5600-system-user-guide-de.pdf (accessed 1 April 2025).

(144) YMC. *YMC-Triart Bio C4*. https://www.ymc.co.jp/en/columns/ymc_triart_bio_c4/ (accessed 1 April 2025).

(145) Abele, M.; Doll, E.; Bayer, F. P.; Meng, C.; Lomp, N.; Neuhaus, K.; Scherer, S.; Kuster, B.; Ludwig, C. Unified workflow for the rapid and in-depth characterization of bacterial proteomes. *Molecular & Cellular Proteomics* **2023**, *22* (8).

(146) Gasteiger, E.; Jung, E.; Bairoch, A. SWISS-PROT: connecting biomolecular knowledge via a protein database. *Current issues in molecular biology* **2001**, *3* (3), 47-55.

(147) Bairoch, A.; Boeckmann, B. The SWISS-PROT protein sequence data bank. *Nucleic acids research* **1992**, *20* (Suppl), 2019.

(148) SCIEX. *TripleTOF™ 5600 System*. (n.d.). https://sciex.com/br/products/mass-spectrometers/qtof-systems/tripletof-systems/tripletof-5600-system (accessed March 11, 2025).

(149) Cai, W.; Tucholski, T.; Chen, B.; Alpert, A. J.; McIlwain, S.; Kohmoto, T.; Jin, S.; Ge, Y. Top-down proteomics of large proteins up to 223 kDa enabled by serial size exclusion chromatography strategy. *Analytical Chemistry* **2017**, *89* (10), 5467-5475.

(150) Thoben, C.; Hartner, N. T.; Hitzemann, M.; Raddatz, C.-R.; Eckermann, M.; Belder, D.; Zimmermann, S. Regarding the Influence of Additives and Additional Plasma-Induced Chemical Ionization on Adduct Formation in ESI/IMS/MS. *Journal of the American Society for Mass Spectrometry* **2023**, *34* (5), 857-868.

(151) Swaney, S. M.; Aoki, H.; Ganoza, M. C.; Shinabarger, D. L. The oxazolidinone linezolid inhibits initiation of protein synthesis in bacteria. *Antimicrobial agents and chemotherapy* **1998**, *42* (12), 3251-3255.

(152) Campbell, E. A.; Korzheva, N.; Mustaev, A.; Murakami, K.; Nair, S.; Goldfarb, A.; Darst, S. A. Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Cell* **2001**, *104* (6), 901-912.

(153) Hards, K.; Robson, J. R.; Berney, M.; Shaw, L.; Bald, D.; Koul, A.; Andries, K.; Cook, G. M. Bactericidal mode of action of bedaquiline. *Journal of Antimicrobial Chemotherapy* **2015**, *70* (7), 2028-2037.

(154) Stallings, C. L.; Stephanou, N. C.; Chu, L.; Hochschild, A.; Nickels, B. E.; Glickman, M. S. CarD is an essential regulator of rRNA transcription required for *Mycobacterium tuberculosis* persistence. *Cell* **2009**, *138* (1), 146-159.