Proteomics of Staphylococcus aureus

Understanding pathogenicity and identification of diagnostic biomarkers

Svetlicic, Ema

Publication date:
2024

Document Version
Publisher's PDF, also known as Version of record

Proteomics of *Staphylococcus aureus*: Understanding pathogenicity and identification of diagnostic biomarkers

PhD Thesis

Ema Svetličić

The Novo Nordisk Foundation Center for Biosustainability

Technical University of Denmark

July 2024

Proteomics of *Staphylococcus aureus*: Understanding pathogenicity and identification of diagnostic biomarkers

PhD Thesis

2024

Ph.D. thesis written by Ema Svetličić

Supervised by Prof. Ivan Mijakovic and Dr. Carsten Jers

Research conducted in the Bacterial Signal Transduction Group

The Novo Nordisk Foundation Center for Biosustainability

Technical University of Denmark

Assessment Committee:

Senior Researcher Alberto Santos Delgado

The Novo Nordisk Foundation Center for Biosustainability

Technical University of Denmark, Denmark

Professor Boris Macek

Interfaculty Institute for Cell Biology

University of Tuebingen, Germany

Professor Christophe Grangeasse

Molecular Microbiology and Structural Biochemistry

The National Centre for Scientific Research, Franc

Preface

This dissertation was prepared at The Novo Nordisk Center for Biosustainability (DTU Biosustain) at the Technical University of Denmark in partial fulfilment of the requirements for acquiring the PhD degree under supervision of Professor Ivan Mijakovic and Senior researcher Carsten Jers. The work was carried out from November 2021 to August 2024 at DTU in Bacterial Signal Transduction Group and during secondment at The Institute of Biomedicine, University of Gothenburg. The thesis was part of the project titled Pioneering Strategies Against Bacterial Infections - PEST-BIN, funded by the European Commission, Horizon 2020 Program as part of Marie Skłodowska-Curie Innovation Training Network (ITN).

Ema Svetličić

Copenhagen, 2024

Summary

Staphylococcus aureus is a major pathogen responsible for a range of serious infections, posing clinical challenges due to its ability to evade the host immune system, develop resistance to multiple antibiotics, and persist in various environments. This study employs proteomic approaches to elucidate the molecular mechanisms underlying *S. aureus* pathogenicity and to identify potential diagnostic biomarkers. Firstly, a benchmarking standard was utilized to systematically evaluate the impact of data acquisition, commonly used software packages and quantification strategies on protein identifications and quantitative accuracy. The results of this study helped us choose the most optimal proteomics workflow to analyze proteome dynamics of *S. aureus* during invasion of human cells. Our findings reveal distinct protein expression patterns associated with virulence factors, invasion mechanisms and bacterial survival strategies. Furthermore, by analyzing the surface proteome of several clinical *S. aureus* we identified several potential peptide biomarkers with high diagnostic potential, which could facilitate early and accurate detection of *S. aureus* infections. These insights not only enhance our understanding of the pathogen's virulence mechanisms but also pave the way for the development of targeted diagnostic and therapeutic strategies.

Resume:

Staphylococcus aureus er en væsentligt patogen, der er ansvarligt for en række alvorlige infektioner. Der forårsager kliniske udfordringer på grund af dens evne til at unddrage sig værtens immunsystem, udvikle resistens over for flere antibiotika og persistere i forskellige miljøer. Denne undersøgelse anvender proteomiske tilgange til at belyse de molekylære mekanismer, der ligger til grund for *S. aureus* patogenicitet, og til at identificere potentielle diagnostiske biomarkører. Indledningsvis blev en benchmarking-standard brugt til systematisk at evaluere indvirkningen af dataindsamling, almindeligt anvendte softwarepakker og kvantificeringsstrategier på proteinidentifikationer og kvantitativ nøjagtighed. Resultaterne af denne undersøgelse hjalp os med at vælge den mest optimale proteomiske arbejdsgang til at analysere proteomdynamikken i *S. aureus* under invasion af menneskelige celler. Vores resultater afslører distinkte proteinekspressionsmønstre forbundet med virulensfaktorer, invasionsmekanismer og bakterielle overlevelsesstrategier. Ved at analysere overfladeproteomet af flere kliniske *S. aureus* stammer identificerede vi desuden adskillige nye proteinbiomarkører med højt diagnostisk potentiale, som vil kunne lette tidlig og præcis påvisning af *S. aureus*-infektioner. Disse indsigter forbedrer ikke kun vores forståelse af denne patogene bakteries virulensmekanismer, men baner også vejen for udvikling af målrettede diagnostiske og terapeutiske strategier.

Acknowledgements

This PhD Thesis was made possible with the support of many individuals, to whom I would like to extend my heartfelt gratitude. Firstly, I would like to thank my supervisor Professor Ivan Mijakovic for giving me the opportunity to do this PhD. I want to thank him for his positivity, encouragement, guidance and for having great trust in me.

I also want to thank my co-supervisor Carsten Jers for the lengthy discussions, valuable suggestions, honesty, and for always being available to answer any questions. No less important, I am thankful for all the chats on the way to the coffee machine, lifts home, lessons in Danish and hangouts outside of work.

I also appreciate all the help and support given by my colleagues at Bacterial Signal Transduction Group. Special thanks to Mukil Madhusudanan for all the lab assistance and Mohammed Ghalib Enayathullah for all the discussions and for taking part in the infection experiments. Thanks also to Belay Tilahun Tadesse for giving me the opportunity to collaborate on his project.

Moreover, I want to thank Leonarda Acha Alarcon and Roger Karlsson from University of Gothenburg, for their guidance and assistance in experimental and data analysis work in this thesis. Special thanks to Leonarda, who made me feel so welcome during my secondment in Gothenburg, for her company in Tromsø and for becoming a dear friend.

I would like to thank all the collaborators on the PEST-BIN project, especially PhD students who made all the Summer Schools an unforgettable experience.

I am thankful to Marie Vestergaard Lukassen and Maike Wennekers Nielsen from DTU Proteomics core for performing mass spectrometry analyses and for their assistance with the sample preparation.

I want to thank my family and all my friends in Croatia. Hvala mojim roditeljima na podršci i ljubavi. Hvala Luki i Idi na svim zajedničkim trenutcima i redovnim posjetima u Kopenhagenu za koje se nadam da će se nastaviti. Hvala i svim mojim prijateljima na druženjima, zabavi i osloncu. Lastly, I want to thank Martin for his continous support, reassurance and understanding.

Publications

Included in the Thesis:

- Chapter 3: **Svetlicic E**, Enayathullah MG, Jers C, Mijakovic. Proteome dynamics of *Staphylococcus aureus* during co-cultivation with human alveolar epithelial cells (Manuscript in preparation)
- Chapter 4: **Svetlicic E**, Alarcon L.A, Karlsson R, Jers C, Mijakovic I. Mass spectrometry-based analysis of surface proteins in *Staphylococcus aureus* clinical strains: Identification of promising peptide targets for diagnostics. (Manuscript in preparation)

Not Included in the Thesis:

- Tadesse BT, **Svetlicic E**, Zhao S, Berhane N, Jers C, Solem C, Mijakovic I. Bad to the bone? - Genomic analysis of Enterococcus isolates from diverse environments reveals that most are safe and display potential as food fermentation microorganisms. Microbiol Res. 2024 Jun;283:127702. doi: 10.1016/j.micres.2024.127702. Epub 2024 Mar 22. PMID: 38552381.
- X. Chen, S. Pandit, L. Shi, V. Ravikumar, J. B. Køhler, **E. Svetlicic**, Z. Cao, A. Garg, D. Petranovic, I. Mijakovic, Graphene Oxide Attenuates Toxicity of Amyloid-β Aggregates in Yeast by Promoting Disassembly and Boosting Cellular Stress Response. Adv. Funct. Mater. 2023, 33, 2304053. https://doi.org/10.1002/adfm.202304053
- **Svetlicic E**, Jaén-Luchoro D, Klobucar RS, Jers C, Kazazic S, Franjevic D, Klobucar G, Shelton BG, Mijakovic I. Genomic characterization and assessment of pathogenic potential of *Legionella spp*. isolates from environmental monitoring. Front Microbiol. 2023 Jan 12;13:1091964. doi: 10.3389/fmicb.2022.1091964. PMID: 36713227; PMCID: PMC9879626.
- **Svetličić E**, Dončević L, Ozdanovac L, Janeš A, Tustonić T, Štajduhar A, Brkić AL, Čeprnja M, Cindrić M. Direct Identification of Urinary Tract Pathogens by MALDI-TOF/TOF Analysis and De Novo Peptide Sequencing. Molecules. 2022 Aug 25;27(17):5461. doi: 10.3390/molecules27175461. PMID: 36080229; PMCID: PMC9457756.

Table of Contents:

Chapter 1

Introduction

1 Introduction

1.1 Background and Significance

Staphylococcus aureus colonizes 20-30% of healthy individuals´ skin and mucosal surfaces such as nose, throat, vaginal wall, and gastrointestinal tract. Although it is regarded as part of the healthy human microbiota it is also a causative agent of serious life-threatening infections (1). *S. aureus* is a Gram-positive bacterium first described in the late 19th century by the Scottish surgeon Alexander Ogston who observed spherical bacteria isolated from the surgical abscess under the microscope. Due to their specific clustered growth, he noted the bacteria resembled bunches of grapes, hence the name Staphylococcus, from a Greek word *staphyle* (bunch of grapes). German physician Friedrich Julius Rosenbach further differentiated the species based on the color of the colonies: *S. aureus* (from the Latin word for gold, *aurum*) (2). The bacterium was also one of the first species to develop resistance to penicillin, only two years after the introduction of the antibiotic in the 1940s. Even when the semi-synthetic beta-lactam antibiotics have replaced penicillin, resistant strains appeared, with methicillin-resistant *S. aureus* (MRSA) described already in the early 1960s (3). For the most part of the 20th century, MRSA was predominantly associated with hospital acquired infections, however in the last decade the bacteria have spread from the hospitals and the infections caused by community associated MRSA are nowadays common (4). These infections can be mild and non-serious, however life threating infections such as bacteremia or septicemia, pneumonia and endocarditis are frequently caused by *S. aureus*. The recent The European Centre of Disease Prevention and Control (ECDC) report shows more than 150 000 infections and 7000 deaths in EU and EEA alone attributed to *S. aureus* (5). This represents a significant health burden marked by high mortality and morbidity. Therefore, It is essential to devise new strategies to combat these infections which can only be achieved through a comprehensive understanding of the complex relationship between *S. aureus* and its host (6).

1.2 *Staphylococcus aureus* pathogenicity

Despite being traditionally regarded as an extracellular bacterium, *S. aureus* has the capability to invade and persist within human cells which is hypothesized to be a mechanism of escaping the host immune system and antimicrobial treatment (7,8). To promote infection, *S. aureus* produces a variety of virulence and immune evasion factors of protein and non-protein origin which allow for its survival in the host environment. The expression of proteins implicated in virulence is controlled by regulators and varies depending on the conditions such

as pH, population density, nutrient availability, oxidative stress etc. (9). While many virulence and regulatory proteins have been identified and their function elucidated, several aspects of the pathogenicity such as the full extent of protein interactions in infection, the complete range of virulence factors, stress adaptation mechanisms and adaptive immune evasion remain incompletely understood (28).

1.2.1 Surface proteins

Cell wall anchored (CWA) proteins typically have multiple functions, as their direct contact with host cells subjects them to selective pressure. Moreover, functional redundancy of CWA proteins has been reported, for example different fibrinogen-binding proteins have been identified from *S. aureus*; Clumping factor A and B (ClfA and ClfB), Coagulase (Coa), Efb and Fibronectin binding protein A and B (FbpA and FbpB) (10,11). The CWA proteins are usually categorized based on structural motifs. One of the major categories is Microbial surface components recognizing adhesive matrix molecules (MSCRAMMs). Originally, this term was associated with proteins which recognize host extra cellular matrix (ECM) proteins such as fibrinogen and fibronectin. However, it is now evident that a variety of CWA proteins can bind to the extracellular matrix (ECM), and that some MSCRAMMs possess additional functions beyond merely promoting adhesion to proteins. Therefore, the term nowadays refers to structural characteristics and applies to proteins with two adjacent IgG-like folds and common mechanisms of ligand binding (12). These proteins include ClfA and ClfB, Serine–aspartate repeat-containing proteins CDE (SdrC, SdrD, SdrE), FbpA, FbpB and Collagen-binding protein (Cna). The functions of these proteins are adhesion to host cells, skin and nasal colonization, immune evasion and ECM adhesion, which reflects their importance in pathogenicity. The second major CWA category is the near iron transporter (NEAT) domain proteins which are responsible for capturing heme from hemoglobin, contributing to survival of bacteria during iron restricted conditions (12). These include iron-regulated surface (Isd) proteins which transfer heme to a membrane transporter which further transfers it to the cytoplasm where the free iron is released (13). There is also accumulating evidence that Isd proteins have several roles in the pathogenesis in addition to acquiring iron. For example IsdA promotes adhesion to the ECM and epithelial cells, while IsdH facilitates evasion of neutrophil uptake by accelerating degradation of the serum opsonin C3b (14). The third structural category comprises a single protein ubiquitous to *S. aureus* called Protein A (Spa) which binds to the Fc region of IgG, the A1 domain of von Willebrand factor and to the Fab region of IgM, thus impairing bacterial phagocytosis (15). G5-E Repeat Domain proteins include *S. aureus* surface

protein G (SasG) which promotes biofilm formation and adherence to human epithelial cells via its A domain (16).

1.2.2 Extracellular proteins

In addition to surface proteins, *S. aureus* secretes extracellular proteins (exotoxins) which target host cell membrane and exhibit distinct functions. The exotoxins of *S. aureus* are divided into cytotoxins, superantigens and cytotoxic enzymes based on their functions. Cytotoxins induce lysis of cells by targeting the cell membrane of the host, superantigens trigger cytokine production and T and B cell proliferation, while cytotoxic enzymes are involved in damage of the host cells (17). The most characterized cytotoxin is an alpha-toxin (alpha-hemolysin) which is a protein monomer encoded by the *hly* gene. Upon secretion, alpha-hemolysin interacts with an ADAM10 receptor on the host´s membrane which contains metalloprotease and disintegrant domains. The interaction causes cleavage of epithelial cadherin proteins (E-cadherins) leading to disruption of the epithelial barrier. *S. aureus* also produces a series of bi-component toxins that are structurally similar to alpha-toxin and belong to the beta-barrel pore-forming toxin family. These include the toxins which induce leukocyte damage, thus called leukocidins: Panton-Valentine leukocidin (PVL, consisting of the LukS and LukF proteins), the leukocidins LukDE and LukAB (also known as LukGH), and gamma-toxin (gamma-hemolysin, HlgA, HlgB, and HlgC)(18). Cytotoxins also include Phenol soluble modulins (PSMs) which are amphipathic peptides unique to staphylococci, with various roles in *S. aureus* pathogenesis including cell lysis, biofilm formation, and immune modulation (19,20). They are encoded in the core genome (genes present in the all or the majority of *S. aureus* strains), secreted by the ABC transporter Pmt, and can influence host immune responses through the activation of the formyl peptide receptor 2 (FPR2) which then induces a proinflammatory response (17). Exotoxins also include T cell superantigens which can trigger a massive immune response by activating T cells non-specifically. Unlike conventional antigens that are processed and presented to T cells in a specific manner, superantigens bind directly to the major histocompatibility complex (MHC) class II molecules on antigen-presenting cells and T cell receptors (TCR) on T cells. This interaction occurs outside the typical antigen-binding sites, leading to the simultaneous activation of a large number of T cells (21,22). The family of *S. aureus* superantigens consists of T cell superantigens (Sags) which include staphylococcal enterotoxins (SEs), staphylococcal enterotoxin-like (SE-l) SAgs, and toxic shock syndrome toxin-1 (TSST-1). SAgs are present great danger to the host organism due to its resistant to heat, low pH, and digestion (pepsin and trypsin) (23). Another member

of the S. aureus superantigens are B-cell superantigens which include only one surface Immunoglobulin binding protein Spa. By binding to Fcγ domain of Immunoglobulin it was shown that Spa prevents opsonization (24). Cytotoxic enzymes of *S. aureus* include β-toxin (also known as β-hemolysin) and Exfoliative toxins (ETs) are also known as epidermolytic toxins. The β-toxin encoding gene, *hlb*, is part of the core *S. aureus* genome, but its expression is often disrupted by the *hlb*-converting prophage, which carries immune evasion factors. β-toxin is a Mg^{2+}-dependent neutral sphingomyelinase that cleaves phospholipid sphingomyelin to produce ceramide and phosphocholine, exhibiting "hot-cold" hemolysin activity. Exfoliative toxins (ETs) ETA, ETB, ETC, and ETD, are serine proteases causing staphylococcal scalded skin syndrome (SSSS), predominantly affecting neonates, infants, and immunocompromised adults. Each ET is encoded on different mobile genetic elements and causes blistering and desquamation (skin peeling) (17).

1.2.3 Regulation of virulence

The regulation of the discussed virulence factors is an important aspect of *S. aureus* pathogenesis, and it is influenced by host and environment derived factors such as antibodies, cytokines, nutrient availability, pH, oxygen level, antibiotic pressure and bacterial density (25). The best characterized regulatory system is the accessory gene regulator (Agr) which is responsible for regulation of several virulence factors such as or repression of SpA, Coa, FnBP and activation of several exoproteins (e.g. alpha-toxin, beta-hemolysin, TSST-1, leukocidins) during the post-exponential phase (26). The *agr* locus consists of five genes (*agrA, agrC, agrD, agrB* and *hld*) which encodes a signaling circuit (Quorum sensing) that both produces and senses the autoinducer, a small peptide named the AIP (autoinducing peptide). The AIP is produced from the *agrD* gene and accumulates in the extracellular space and it activates the Agr system when it reaches critical concentration. The concentration of AIP is proportional to the cell density, and it serves as a communication between bacteria in a process called quorum sensing. The AIP is detected by a classic two-component signaling system (TCS) which consists of a histidine kinase serving as a membrane receptor (AgrC) and a response regulator (AgrA). TCSs controls many virulence determinants in bacterial pathogens through the same principle. A transmembrane sensor kinase receives a signal which induces autophosphorylation of the kinase. Subsequently, the sensor kinase phosphorylates and thereby activates its cognate response regulator that in turn leads to modulation of expression of its target genes (26). *S. aureus* encodes a total of 16 conserved pairs of TCSs that are involved in various cellular processes such as virulence regulation, antimicrobial resistance and metabolism (27). Other

well characterized TCSs in *S. aureus* include SaeRS and SrrAB, The sae operon consists of four genes (*saeP, saeQ, saeR, and saeS*), and the two promoters P1 and P3. SaeR and SaeS are the response regulator and HK of the system, respectively and are constitutively expressed from promotor P3. The P1 promoter is autoinduced by SaeRS and drives the expression of all four genes. The SaeRS was shown to positively regulate the expression of toxins (*hla, hlb, hld, lukGH* etc.), adhesins (*fnbA, fnbB, coa*), extracellular proteases (e.g *splA, splB, splC*,), superantigen-like proteins (SSLs) and many others virulence determinants (27,28). The SrrAB TCS consists of the cytoplasmic response regulator SrrA and the membrane-bound histidine kinase SrrB of the signal transduction system. The studies showed that SrrAB enables growth under hypoxia, oxidative stress and nitrosative stress by upregulating genes involved in fermentation (*pflAB*, *adhE*, *nrdDG*), H_2O_2 resistance (*kat, ahpC, dps*) and repair of nitrosative damage (*hmp, scdA*) (29,30). Moreover, SrrAB negatively regulates agr system, protein A (spa), and TSST-1 under low-oxygen condition by direct binding to their promotors (31). Other TCSs in *S. aureus* with implications in virulence include TCSs involved in response to antimicrobials (VraSR, GraXSR, BraRS), Cell Wall Metabolism, Autolysis (WalRK, ArlRS, LytSR), nitrate metabolism (NreCBA), respiration (AirRS), nutrient sensing (HssSR, PhoRP). Their role and current knowledge is detailed in an comprehensive review (27)

Apart from TCSs, *S. aureus* generates a family of DNA-binding proteins that control virulence factor expression. The most thoroughly researched among them is the staphylococcal accessory regulator nucleic acid-binding protein, SarA. It regulates virulence by upregulation of the Agr system which leads to the upregulation of toxins and repression of protein A (Spa). Evidence of direct posttranscriptional regulation have been found, where SarA binds to and consequently alters the mRNA levels of target genes such as *spa* (32). SarA binds to its own promoter, thereby auto upregulating its expression. Its activity is negatively regulated by the SarA-like protein SarR and positively regulated by the alternative sigma factor σ^B (33).

1.2.4 Immune evasion

In addition to producing multiple virulence factors, *S. aureus* has developed mechanisms to evade human immune system. Neutrophils play a crucial role in combating *S. aureus* infections. This was first observed in patients with neutrophil deficiencies, who exhibited extreme susceptibility to these infections (34,35). However, the majority of infections are present in individuals without a compromised immune system. This is because *S. aureus* targets neutrophils on several levels to evade the immune response: 1) the inhibition of neutrophil extravasation from the bloodstream into the tissues, neutrophil activation, and

chemotaxis, 2) inhibition of phagocytosis by aggregation, protective surface structures, and biofilm formation, 3) inhibition of opsonization, 4) inhibition of neutrophil killing mechanisms, and 5) direct elimination of neutrophils by cytolytic toxins or triggering of apoptosis (36,37). The fact that *S. aureus* has acquired a wide array of immune evasion factors, in contrast to other bacterial pathogens that manage with relatively few, can be understood by considering that *S. aureus* establishes life-long associations with its human hosts (38). Understanding the adaptive immune evasion mechanisms of *S. aureus* remains a significant challenge. This is partly due to the fact that many of the immune evasion proteins are specific to human hosts, complicating the study of *S. aureus* infections in animal models such as mice (39).

1.3 Proteomics in infectious microbiology

1.3.1 Proteomics in general

Proteomics field encompasses the identification and quantification of proteins and their interactions within a given type of cell or organism (40). Proteomics aims to provide a comprehensive understanding of how proteins contribute to various biological processes and how their expression changes between different conditions. Unlike genomics, which focuses on the study of genes, proteomics deals with the functional end-products of gene expression (i.e. proteins), giving insight into the actual biological activities occurring in cells. The proteomics field consists of a wide range of methodologies; however, the field has mostly been driven by the development of mass spectrometry (MS) based proteomics methods (41–43). MS-based proteomics can be broadly categorized into top-down, bottom-up, and targeted proteomics, each serving different applications and using different methodology. In top-down proteomics, intact proteins are analyzed for the purpose to detect protein degradation products, isoforms and post-translational modifications (PTMs), however the method is challenging due to protein solubility, lower sensitivity due to poor efficiency of protein ionization, difficulties in protein fragmentation etc. (44,45). In bottom-up and targeted proteomics proteins are digested into peptides by a specialized protease, usually trypsin, and analyzed by MS. The difference between the two approaches is that targeted approach can accurately detect and quantify a set of predefined peptides, therefore often being used when the protein/peptide of interest is already known, such as biomarker detection in clinical samples (46–48). Bottom-up proteomics, on the other hand aims to identify and quantify as many proteins as possible in the given sample, offering a broad overview on protein expression patterns and enabling comparison of global protein expression between different samples (49). The focus of this chapter is on the application of bottom-up proteomics in microbiology, primarily in host-

pathogen interactions and discovery of potential biomarkers of bacterial pathogens, thus the other two approaches will not be further discussed.

1.3.2 Proteomics workflow

The general workflow of bottom-up proteomics consists of 1) extraction of proteins from the studied sample 2) quantification of total protein content and normalization 3) protein digestion to peptides 4) purification and peptide desalting 5) offline fractionation of peptides (optional) 6) liquid chromatography (LC) separation of peptides and MS analysis 7) protein identification by database search. An efficient protein extraction should encompass maximized cell lysis, inactivation of pathogens for subsequent safe handling and solubilization of proteins. Lysis is usually achieved by lysis buffers containing chaotropic agents (urea, guanidinium hydrochloride) and detergents (sodium dodecyl-sulfate(SDS)) in combination with mechanical lysis (bead-beating) or ultrasonication (50,51). The protocol for mechanical lysis, including detergent concentration, power, and duration, typically needs to be tailored to different bacterial species. This is particularly important for species with thick cell walls, such as Gram-positive bacteria and spores (52–54). In cases of difficulty in cell lysis, protein amount limitation or other, certain enzymes targeting the cell wall can be used such as lysozyme and lysostaphin, however the incubation time and concentration should be carefully chosen to avoid further protease activity. (55,56). After isolating the proteins, the total protein concentration should be measured and ideally normalized to ensure that an equal amount of proteins is present in all samples. The most commonly used methods are the Bradford and bicinchoninic acid (BCA) assays which offer an accurate estimation of protein amount; however the lysis buffer components can interfere with the measurement, therefor the choice of the assay should depend on the buffer in which proteins are solubilized (57). High salt concentrations, detergents and chaotropic agents from the lysis buffer should be removed or minimized to ensure efficient protein digestion. This can be accomplished by protein precipitation, dilution or buffer exchange using centrifugal filters (58–60). Digestion is most commonly performed using trypsin or LysC (or combination) and incubating overnight. The resulting peptides need to be free from any salt, detergent or any other components which can cause poor ionization efficiency, signal suppression or distorted chromatography during analysis. Therefore, peptides are usually desalted prior to analysis, using reverse phase resins such as C18 matrix. The peptides are then separated by reversed-phase LC before they are ionized and introduced into a mass spectrometer.

The main components of any mass spectrometer are ion source, a mass analyzer and a detector. In bottom-up proteomics, the most common type of ionization is electrospray ionization (ESI) where liquid phase peptides are converted to gaseous peptide ions and are separated based on their mass to charge ratio (m/z) in a mass analyzer (61). Widely used mass analyzers include time-of-flight (TOF), Orbitraps, quadrupoles, and ion traps, each with distinct properties. These analyzers can either separate all analytes in a sample for comprehensive analysis or function as filters, directing only ions of specific *m/z* towards the detector (62). The detector captures separated ions from the mass analyzer, converts their impacts into electrical signals, and transforms these signals into digital data for the analysis. In typical bottom-up proteomics, two acquisition strategies have emerged: data dependent acquisition (DDA) and data independent acquisition (DIA). The DDA is a more traditional approach, where peptide ions are eluted, ionized and all ions within a wide m/z range (for example: 300-1000m/z) are detected in a survey scan resulting in an MS1 spectrum of precursors. The most abundant ions are then isolated through a narrow m/z window and subsequently fragmented resulting in an MS2 spectrum for each isolated precursor, from which the peptide sequence can be inferred. The peptides are identified by matching the experimentally derived spectra with a theoretical database containing precursor and fragments masses. While the spectra are relatively easily interpreted, DDA can suffer from relatively low reproducibility due to the stochastic nature of precursor selection leading to so called “missing values” even in technical replicates of the same sample. To address the shortcomings of DDA, a DIA approach was developed where precursors from the MS1 scan are co-isolated within a wider m/z range (usually 25 m/z) and then fragmented, resulting in a MS2 spectra containing fragments from multiple peptide precursors. The approach decreases variability and boosts peptide identifications; however, the spectra are more difficult to interpret and usually require spectral libraries generated by an extensive analysis of fractionated peptides by DDA (63,64). Recently, library free approaches have been developed, where the libraries were generated from protein sequences by machine learning, therefore accurately mimicking experimental libraries (65).

1.3.3 Quantification strategies

In bottom-up proteomics, quantification is usually not absolute but is based on the comparison of measured intensities between different samples. Besides acquisition method, bottom-up proteomics also has several approaches for peptide quantification, and these are label-based quantification and label free quantification (LFQ). Label-based quantification

enables each sample to have its specific identifier which allows for multiplexing of samples therefore reducing the instrument time and technical variability of the samples. Labeling can be performed on a protein level *in vivo* in a process called stable isotope labeling by/with amino acids in cell culture (SILAC). The principle of SILAC involves growing two populations of cells, one in medium that contains normal amino acids (light) and the other population of cells is fed medium containing amino acids (arginine/lysin) labeled with stable heavy isotopes. This enables accurate quantification of peptides on a MS1 level, however, only studies involving mammalian cell culture or model microbial organisms which can be made auxotrophic for arginine/lysine to be able to incorporate "heavy" amino acids during growth (66). Labeling can also be performed on a peptide level, with the most commonly used labels being tandem mass tags (TMT) which are based on isotopic mass tag variants. The chemical tag is composed of an amine-reactive group, a mass normalization group and a reporter ion group, Therefore, upon sample multiplexing, a labeled peptide co-elutes as a single peak with the same m/z value in MS1 scan regardless from which sample originates. The differentiation of samples is possible upon fragmentation, when the mass normalizer region gets cleaved off, resulting in reporter ion peaks of distinct m/z corresponding to the label. The relative quantification of different peptides is possible by comparing intensities of reporter ions in MS2 scan (67). Due to co-isolation of MS1 peaks sent for fragmentation, MS2 spectra sometimes contain fragments from more than one precursor, therefore skewing the quantification. New instruments with three mass analyzers (tribrids) have been introduced for that reason, enabling quantification in MS3 level, thus overcoming the interference from multiple co-isolated precursors (68). In addition to TMT labeling, commonly used labels are dimethyl labeling (69) and ITRAQ (70). Label-based approaches have been mostly associated with DDA analysis, however several recent studies have presented the multiplexed workflow using DIA in combination with labeling approaches (71,72).

1.3.4 Application to pathogenic bacteria

The potential of proteomics applied to pathogenic bacteria is wide. Instead of focusing on a single protein or a pathway, with modern proteomics it is possible to capture a global overview of the entire protein landscape, allowing for comprehensive analysis of protein expression, post-translational modifications, interactions, and dynamics within the bacterial cell under various conditions. Proteomics has been an important technique for studying bacterial adaptation to stress conditions such as hypoxia (73,74), pH changes (75,76), deprivation of nutrients (77,78) and oxidative stress (79). Studying the proteomics of bacteria

under stress conditions is crucial because it reveals how pathogens adapt to and survive in hostile environments encountered during host immune responses, such as phagocytosis. Proteomics applications also include differential proteome profiling between antibiotic tolerant/resistant MRSA strains (80,81) as well as studying factors involved in bacterial persistence (80). In several other studies, proteomics was utilized to study bacterial response to antibiotics (82–84), metal based nanoparticles (85–87) and graphene oxide (88). In addition, different MS-based proteomics techniques have been applied in profiling the relationship between bacterial pathogens and the host. In one such study, researchers monitored newly synthesized proteins of macrophages infected with *Salmonella enterica* across different cellular compartments and infection times, using click chemistry and SILAC-based proteomics (89). Other labeling approaches have also been used, such as in a study where intestinal organoids were infected with different strains of *Listeria monocytogenes*, and the remodeling of the host proteome was profiled using TMT labeling (90). LFQ DDA approaches have also been utilized in several *in vitro* (91–94) and *in vivo* (95–97) studies to profile proteome response of the host to the bacterial infection. Profiling bacterial response in host environment conditions is more challenging due to the prevalence of host proteins, therefore advanced methodologies have been utilized to profile both host and bacterial proteome response. Recently, a study reported an improvement in bacterial peptide detection from a mixture of *Salmonella*/HeLa cells using an optimized hybrid library generation workflow for DIA mass spectrometry. This workflow relied on data-dependent acquisition and *in silico*-predicted spectral libraries (98). Another study generated a global ion library using DDA for DIA analysis and successfully applied it to study response of *S. aureus* in human bronchial epithelial cells and in a murine pneumonia model (56). The mentioned studies analyzed the cellular proteome. Post-translational modifications in infectious bacteria regulate many biological processes (99), but due to the low percentage of modified amino acid sites compared to eukaryotes, the knowledge of distribution of PTMs in pathogenic bacteria is limited. In one of the most comprehensive studies of PTMs in pathogens, researchers have profiled phosphoproteome *S. aureus* in serine/threonine kinase and phosphatase mutants to better understand the impact of protein phosphorylation on the complex signaling networks involved in the bacterial pathogenicity (100).

In addition, diagnostics of pathogens is a major potential application of bottom-up proteomics, including identification of species-unique or associated peptides directly from the MS data in a process called “proteotyping “(101). Moreover, diagnostics is commonly performed using immunological methods where antibodies are raised against a specific protein

or peptide called a biomarker. Proteomics has an important role in this process by enabling the discovery and validation of novel biomarkers. For example, a study identified several host plasma proteins that could serve as biomarkers for increased susceptibility to *Mycobacterium tuberculosis* infection or predict the development of tuberculosis in different patients (102). Bacterial biomarkers were also identified in a study where proteomics was applied to identify biomarkers associated with invasive *Streptococcus* strains causing meningitis and sepsis (103). Furthermore, these biomarkers can pave a way towards new therapeutics and the development of vaccines against bacterial infections (104–106).

The further advances of MS based proteomics, including improvements in instrument resolution and sensitivity, development of DIA technologies and computational advances promise to facilitate studying proteomes of complex biological systems, paving the way for more accurate and detailed studies of host-pathogen interactions.

1.4 References

1. Tong SYC, Davis JS, Eichenberger E, Holland TL, Fowler VG. *Staphylococcus aureus* Infections: Epidemiology, Pathophysiology, Clinical Manifestations, and Management. Clin Microbiol Rev. 2015 Jul;28(3):603–61.

2. Licitra G. Etymologia: Staphylococcus. Emerg Infect Dis. 2013 Sep;19(9):1553.

3. Chambers HF, DeLeo FR. Waves of Resistance: *Staphylococcus aureus* in the Antibiotic Era. Nat Rev Microbiol. 2009 Sep;7(9):629–41.

4. Myles IA, Datta SK. *Staphylococcus aureus*: an introduction. Semin Immunopathol. 2012 Mar;34(2):181–4.

5. Cassini A, Högberg LD, Plachouras D, Quattrocchi A, Hoxha A, Simonsen GS, et al. Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. Lancet Infect Dis. 2019 Jan 1;19(1):56–66.

6. Raineri EJM, Altulea D, van Dijl JM. Staphylococcal trafficking and infection-from “nose to gut” and back. FEMS Microbiol Rev. 2022 Jan 18;46(1):fuab041.

7. Casadevall A, Fang FC. The intracellular pathogen concept. Mol Microbiol. 2020;113(3):541–5.

8. Flannagan RS, Heit B, Heinrichs DE. Intracellular replication of *Staphylococcus aureus* in mature phagolysosomes in macrophages precedes host cell death, and bacterial escape and dissemination. Cell Microbiol. 2016;18(4):514–35.

9. Liu GY. Molecular Pathogenesis of *Staphylococcus aureus* Infection. Pediatr Res. 2009 May;65(5 Pt 2):71R-77R.

10. Ko YP, Kang M, Ganesh VK, Ravirajan D, Li B, Höök M. Coagulase and Efb of *Staphylococcus aureus* Have a Common Fibrinogen Binding Motif. mBio. 2016 Jan 5;7(1):e01885-15.

11. da Costa TM, Viljoen A, Towell AM, Dufrêne YF, Geoghegan JA. Fibronectin binding protein B binds to loricrin and promotes corneocyte adhesion by *Staphylococcus aureus*. Nat Commun. 2022 May 6;13(1):2517.

12. Foster TJ. Surface Proteins of *Staphylococcus aureus*. Microbiol Spectr. 2019 Jul 5;7(4):10.1128/microbiolspec.gpp3-0046–2018.

13. Geoghegan JA, Foster TJ. Cell Wall-Anchored Surface Proteins of *Staphylococcus aureus*: Many Proteins, Multiple Functions. Curr Top Microbiol Immunol. 2017;409:95–120.

14. Visai L, Yanagisawa N, Josefsson E, Tarkowski A, Pezzali I, Rooijakkers SHM, et al. Immune evasion by *Staphylococcus aureus* conferred by iron-regulated surface determinant protein IsdH. Microbiology. 2009;155(3):667–79.

15. Cruz AR, Bentlage AEH, Blonk R, de Haas CJC, Aerts PC, Scheepmaker LM, et al. Toward Understanding How Staphylococcal Protein A Inhibits IgG-Mediated Phagocytosis. J Immunol. 2022 Sep 15;209(6):1146–55.

16. Geoghegan JA, Corrigan RM, Gruszka DT, Speziale P, O'Gara JP, Potts JR, et al. Role of Surface Protein SasG in Biofilm Formation by *Staphylococcus aureus*. J Bacteriol. 2010 Nov;192(21):5663–73.

17. Tam K, Torres VJ. *Staphylococcus aureus* Secreted Toxins and Extracellular Enzymes. Microbiol Spectr. 2019 Mar 15;7(2):10.1128/microbiolspec.gpp3-0039–2018.

18. Otto M. *Staphylococcus aureus* toxins. Curr Opin Microbiol. 2014 Feb;0:32–7.

19. Dastgheyb SS, Villaruz AE, Le KY, Tan VY, Duong AC, Chatterjee SS, et al. Role of Phenol-Soluble Modulins in Formation of *Staphylococcus aureus* Biofilms in Synovial Fluid. Infect Immun. 2015 Jul;83(7):2966–75.

20. Peschel A, Otto M. Phenol-soluble modulins and staphylococcal infection. Nat Rev Microbiol. 2013 Oct;11(10):667–73.

21. Xu SX, McCormick JK. Staphylococcal superantigens in colonization and disease. Front Cell Infect Microbiol. 2012 Apr 17;2:52.

22. Tuffs SW, Goncheva MI, Xu SX, Craig HC, Kasper KJ, Choi J, et al. Superantigens promote *Staphylococcus aureus* bloodstream infection by eliciting pathogenic interferon-gamma production. Proc Natl Acad Sci U S A. 2022 Feb 22;119(8):e2115987119.

23. Li SJ, Hu DL, Maina E k., Shinagawa K, Omoe K, Nakane A. Superantigenic activity of toxic shock syndrome toxin-1 is resistant to heating and digestive enzymes. J Appl Microbiol. 2011;110(3):729–36.

24. Chen X, Schneewind O, Missiakas D. Engineered human antibodies for the opsonization and killing of *Staphylococcus aureus*. Proc Natl Acad Sci U S A. 2022 Jan 25;119(4):e2114478119.

25. Jenul C, Horswill AR. Regulation of *Staphylococcus aureus* Virulence. Microbiol Spectr. 2019;7(2):10.1128/microbiolspec.GPP3-0031-2018

26. Bronner S, Monteil H, Prévost G. Regulation of virulence determinants in *Staphylococcus aureus*: complexity and applications. FEMS Microbiol Rev. 2004 May 1;28(2):183–200.

27. Haag AF, Bagnoli F. The Role of Two-Component Signal Transduction Systems in *Staphylococcus aureus* Virulence Regulation. Curr Top Microbiol Immunol. 2017;409:145-198

28. Liu Q, Yeo WS, Bae T. The SaeRS Two-Component System of *Staphylococcus aureus*. Genes. 2016 Oct 3;7(10):81.

29. Kinkel TL, Roux CM, Dunman PM, Fang FC. The *Staphylococcus aureus* SrrAB Two-Component System Promotes Resistance to Nitrosative Stress and Hypoxia. mBio. 2013 Nov 12;4(6):10.1128/mbio.00696-13.

30. Mashruwala AA, Boyd JM. The *Staphylococcus aureus* SrrAB Regulatory System Modulates Hydrogen Peroxide Resistance Factors, Which Imparts Protection to Aconitase during Aerobic Growth. PLoS ONE. 2017 Jan 18;12(1):e0170283.

31. Yarwood JM, McCormick JK, Schlievert PM. Identification of a novel two-component regulatory system that acts in global regulation of virulence factors of *Staphylococcus aureus*. J Bacteriol. 2001 Feb;183(4):1113–23.

32. Morrison JM, Anderson KL, Beenken KE, Smeltzer MS, Dunman PM. The staphylococcal accessory regulator, SarA, is an RNA-binding protein that modulates the mRNA turnover properties of late-exponential and stationary phase *Staphylococcus aureus* cells. Front Cell Infect Microbiol. 2012;2:26.

33. Cheung AL, Zhang G. Global regulation of virulence determinants in *Staphylococcus aureus* by the SarA protein family. Front Biosci J Virtual Libr. 2002 Aug 1;7:d1825-1842.

34. Lekstrom-Himes Julie A., Gallin John I. Immunodeficiency Diseases Caused by Defects in Phagocytes. N Engl J Med. 2000;343(23):1703–14.

35. Curnutte JT, Whitten DM, Babior BM. Defective superoxide production by granulocytes from patients with chronic granulomatous disease. N Engl J Med. 1974 Mar 14;290(11):593–7.

36. Cheung GYC, Bae JS, Otto M. Pathogenicity and virulence of *Staphylococcus aureus*. Virulence. 2021 Dec 31;12(1):547–69.

37. de Jong NWM, van Kessel KPM, van Strijp JAG. Immune Evasion by *Staphylococcus aureus*. Microbiol Spectr. 2019 Mar;7(2).

38. Thammavongsa V, Kim HK, Missiakas D, Schneewind O. Staphylococcal manipulation of host immune responses. Nat Rev Microbiol. 2015 Sep;13(9):529–43.

39. Tsai CM, Hajam IA, Caldera JR, Liu GY. Integrating complex host-pathogen immune environments into S. aureus vaccine studies. Cell Chem Biol. 2022 May 19;29(5):730–40.

40. Graves PR, Haystead TAJ. Molecular Biologist's Guide to Proteomics. Microbiol Mol Biol Rev. 2002 Mar;66(1):39–63.

41. Halder A, Verma A, Biswas D, Srivastava S. Recent advances in mass-spectrometry based proteomics software, tools and databases. Drug Discov Today Technol. 2021 Dec 1;39:69–79.

42. Rajczewski AT, Jagtap PD, Griffin TJ. An overview of technologies for MS-based proteomics-centric multi-omics. Expert Rev Proteomics. 2022 Mar;19(3):165–81.

43. Kollipara S, Agarwal N, Varshney B, Paliwal J. Technological Advancements in Mass Spectrometry and Its Impact on Proteomics. Anal Lett. 2011 May 1;44(8):1498–520.

44. Po A, Eyers CE. Top-Down Proteomics and the Challenges of True Proteoform Characterization. J Proteome Res. 2023 Dec 1;22(12):3663–75.

45. Melby JA, Roberts DS, Larson EJ, Brown KA, Bayne EF, Jin S, et al. Novel Strategies to Address the Challenges in Top-Down Proteomics. J Am Soc Mass Spectrom. 2021 Jun 2;32(6):1278–94.

46. Nakayasu ES, Gritsenko M, Piehowski PD, Gao Y, Orton DJ, Schepmoes AA, et al. Tutorial: best practices and considerations for mass-spectrometry-based protein biomarker discovery and validation. Nat Protoc. 2021 Aug;16(8):3737–60.

47. Masuda T, Mori A, Ito S, Ohtsuki S. Quantitative and targeted proteomics-based identification and validation of drug efficacy biomarkers. Drug Metab Pharmacokinet. 2021 Feb 1;36:100361.

48. Zoodsma M, de Nooijer AH, Grondman I, Gupta MK, Bonifacius A, Koeken VACM, et al. Targeted proteomics identifies circulating biomarkers associated with active COVID-19 and post-COVID-19. Front Immunol. 2022;13:1027122.

49. Zhang Y, Fonslow BR, Shan B, Baek MC, Yates JR. Protein Analysis by Shotgun/Bottom-up Proteomics. Chem Rev. 2013 Apr 10;113(4):2343–94.

50. Varnavides G, Madern M, Anrather D, Hartl N, Reiter W, Hartl M. In Search of a Universal Method: A Comparative Survey of Bottom-Up Proteomics Sample Preparation Methods. J Proteome Res. 2022 Oct 7;21(10):2397–411.

51. Winter D, Dehghani A, Steen H. Optimization of cell lysis and protein digestion protocols for protein analysis by LC-MS/MS. Methods Mol Biol Clifton NJ. 2015;1295:259–73.

52. Abhyankar W, de Koning LJ, Brul S, de Koster CG. Spore proteomics: the past, present and the future. FEMS Microbiol Lett. 2014 Sep 1;358(2):137–44.

53. Hayoun K, Gaillard JC, Pible O, Alpha-Bazin B, Armengaud J. High-throughput proteotyping of bacterial isolates by double barrel chromatography-tandem mass spectrometry based on microplate paramagnetic beads and phylopeptidomics. J Proteomics. 2020 Aug 30;226:103887.

54. Abele M, Doll E, Bayer FP, Meng C, Lomp N, Neuhaus K, et al. Unified Workflow for the Rapid and In-Depth Characterization of Bacterial Proteomes. Mol Cell Proteomics MCP. 2023 Aug;22(8):100612.

55. Falgenhauer E, von Schönberg S, Meng C, Mückl A, Vogele K, Emslander Q, et al. Evaluation of an E. coli Cell Extract Prepared by Lysozyme-Assisted Sonication via Gene Expression, Phage Assembly and Proteomics. Chembiochem Eur J Chem Biol. 2021 Sep 14;22(18):2805–13.

56. Michalik S, Depke M, Murr A, Gesell Salazar M, Kusebauch U, Sun Z, et al. A global *Staphylococcus aureus* proteome resource applied to the in vivo characterization of host-pathogen interactions. Sci Rep. 2017 Sep 8;7(1):9718.

57. Kassem S, van der Pan K, de Jager AL, Naber BAE, de Laat IF, Louis A, et al. Proteomics for Low Cell Numbers: How to Optimize the Sample Preparation Workflow for Mass Spectrometry Analysis. J Proteome Res. 2021 Sep 3;20(9):4217–30.

58. Nickerson JL, Sheridan LV, Doucette AA. Impact of Surfactants on Cumulative Trypsin Activity in Bottom-Up Proteome Analysis. J Proteome Res [Internet]. 2024 Jul 7 [cited 2024 Jul 10]; Available from: https://doi.org/10.1021/acs.jproteome.4c00162

59. Wiśniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. Nat Methods. 2009 May;6(5):359–62.

60. Sharma R, Dill BD, Chourey K, Shah M, VerBerkmoes NC, Hettich RL. Coupling a Detergent Lysis/Cleanup Methodology with Intact Protein Fractionation for Enhanced Proteome Characterization. J Proteome Res. 2012 Dec 7;11(12):6008–18.

61. Han X, Aslanian A, Yates JR. Mass spectrometry for proteomics. Curr Opin Chem Biol. 2008 Oct 1;12(5):483–90.

62. Shuken SR. An Introduction to Mass Spectrometry-Based Proteomics. J Proteome Res. 2023 Jul 7;22(7):2151–71.

63. Bateman NW, Goulding SP, Shulman NJ, Gadok AK, Szumlinski KK, MacCoss MJ, et al. Maximizing Peptide Identification Events in Proteomic Workflows Using Data-Dependent Acquisition (DDA) *. Mol Cell Proteomics. 2014 Jan 1;13(1):329–38.

64. Matafora V, Corno A, Ciliberto A, Bachi A. Missing Value Monitoring Enhances the Robustness in Proteomics Quantitation. J Proteome Res. 2017 Apr 7;16(4):1719–27.

65. Pino LK, Just SC, MacCoss MJ, Searle BC. Acquiring and Analyzing Data Independent Acquisition Proteomics Experiments without Spectrum Libraries. Mol Cell Proteomics. 2020 Jul 1;19(7):1088–103.

66. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics MCP. 2002 May;1(5):376–86.

67. Thompson A, Schäfer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, et al. Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. Anal Chem. 2003 Apr 1;75(8):1895–904.

68. He Y, Shishkova E, Peters-Clarke TM, Brademan DR, Westphall MS, Bergen D, et al. Evaluation of the Orbitrap Ascend Tribrid Mass Spectrometer for Shotgun Proteomics. Anal Chem. 2023 Jul 18;95(28):10655–63.

69. Hsu JL, Huang SY, Chow NH, Chen SH. Stable-Isotope Dimethyl Labeling for Quantitative Proteomics. Anal Chem. 2003 Dec 1;75(24):6843–52.

70. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, et al. Multiplexed Protein Quantitation in Saccharomyces cerevisiae Using Amine-reactive Isobaric Tagging Reagents *. Mol Cell Proteomics. 2004 Dec 1;3(12):1154–69.

71. Thielert M, Itang EC, Ammar C, Rosenberger FA, Bludau I, Schweizer L, et al. Robust dimethyl-based multiplex-DIA doubles single-cell proteome depth via a reference channel. Mol Syst Biol. 2023 Aug 21;19(9):e11503.

72. Tian X, de Vries MP, Permentier HP, Bischoff R. The Isotopic Ac-IP Tag Enables Multiplexed Proteome Quantification in Data-Independent Acquisition Mode. Anal Chem. 2021 Jun 15;93(23):8196–202.

73. Gil-Marqués ML, Pachón J, Smani Y. iTRAQ-Based Quantitative Proteomic Analysis of Acinetobacter baumannii under Hypoxia and Normoxia Reveals the Role of OmpW as a Virulence Factor. Microbiol Spectr. 2022 Mar 2;10(2):e02328-21.

74. Suzuki S, Subsomwong P, Narita K, Kawai N, Ishiai T, Teng W, et al. Differential proteomic analysis and pathogenic effects of outer membrane vesicles derived from Acinetobacter baumannii under normoxia and hypoxia. PLOS ONE. 2023 Mar 15;18(3):e0283109.

75. Großeholz R, Koh CC, Veith N, Fiedler T, Strauss M, Olivier B, et al. Integrating highly quantitative proteomics and genome-scale metabolic modeling to study pH adaptation in the human pathogen Enterococcus faecalis. Npj Syst Biol Appl. 2016 Sep 8;2(1):1–9.

76. Zhang DL, Bai YL, Bowman JP. Impact of Combined Acidic and Hyperosmotic Shock Conditions on the Proteome of *Listeria monocytogenes* ATCC 19115 in a Time-Course Study. J Food Qual. 2019;2019(1):3075028.

77. Al Dahouk S, Jubier-Maurin V, Neubauer H, Köhler S. Quantitative analysis of the Brucella suis proteome reveals metabolic adaptation to long-term nutrient starvation. BMC Microbiol. 2013 Sep 4;13(1):199.

78. Gallagher GE, Waldbauer JR. Proteome Expression and Survival Strategies of a Proteorhodopsin-Containing Vibrio Strain under Carbon and Nitrogen Limitation. mSystems. 2022 Apr 6;7(2):e01263-21.

79. Fu J, Qi L, Hu M, Liu Y, Yu K, Liu Q, et al. *Salmonella* proteomics under oxidative stress reveals coordinated regulation of antioxidant defense with iron metabolism and bacterial virulence. J Proteomics. 2017 Mar 22;157:52–8.

80. Sulaiman JE, Lam H. Application of proteomics in studying bacterial persistence. Expert Rev Proteomics. 2019 Mar 4;16(3):227–39.

81. Jasim R, Baker MA, Zhu Y, Han M, Schneider-Futschik EK, Hussein M, et al. A Comparative Study of Outer Membrane Proteome between Paired Colistin-Susceptible and Extremely Colistin-Resistant Klebsiella pneumoniae Strains. ACS Infect Dis. 2018 Dec 14;4(12):1692–704.

82. Li W, Zhang S, Wang X, Yu J, Li Z, Lin W, et al. Systematically integrated metabonomic-proteomic studies of *Escherichia coli* under ciprofloxacin stress. J Proteomics. 2018 May 15;179:61–70.

83. Rao AA, Patkari M, Reddy PJ, Srivastava R, Pendharkar N, Rapole S, et al. Proteomic analysis of *Streptomyces coelicolor* in response to Ciprofloxacin challenge. J Proteomics. 2014 Jan 31;97:222–34.

84. Jedrey H, Lilley KS, Welch M. Ciprofloxacin binding to GyrA causes global changes in the proteome of Pseudomonas aeruginosa. FEMS Microbiol Lett. 2018 Jul 1;365(13):fny134.

85. Planchon M, Léger T, Spalla O, Huber G, Ferrari R. Metabolomic and proteomic investigations of impacts of titanium dioxide nanoparticles on Escherichia coli. PLOS ONE. 2017 Jun 1;12(6):e0178437.

86. Zhang Y, Pan X, Liao S, Jiang C, Wang L, Tang Y, et al. Quantitative Proteomics Reveals the Mechanism of Silver Nanoparticles against Multidrug-Resistant Pseudomonas aeruginosa Biofilms. J Proteome Res. 2020 Aug 7;19(8):3109–22.

87. Yan X, He B, Liu L, Qu G, Shi J, Hu L, et al. Antibacterial mechanism of silver nanoparticles in Pseudomonas aeruginosa: proteomics approach†. Metallomics. 2018 Apr 1;10(4):557–64.

88. Ravikumar V, Mijakovic I, Pandit S. Antimicrobial Activity of Graphene Oxide Contributes to Alteration of Key Stress-Related and Membrane Bound Proteins. Int J Nanomedicine. 2022 Dec 28;17:6707–21.

89. Selkrig J, Li N, Hausmann A, Mangan MSJ, Zietek M, Mateus A, et al. Spatiotemporal proteomics uncovers cathepsin-dependent macrophage cell death during Salmonella infection. Nat Microbiol. 2020 Sep;5(9):1119–33.

90. Zhou C, Zou Y, Huang J, Zhao Z, Zhang Y, Wei Y, et al. TMT-Based Quantitative Proteomic Analysis of Intestinal Organoids Infected by *Listeria monocytogenes* Strains with Different Virulence. Int J Mol Sci. 2022 Jan;23(11):6231.

91. Noster J, Chao TC, Sander N, Schulte M, Reuter T, Hansmeier N, et al. Proteomics of intracellular Salmonella enterica reveals roles of Salmonella pathogenicity island 2 in metabolism and antioxidant defense. PLOS Pathog. 2019 Apr 22;15(4):e1007741.

92. Li Z, Liu Y, Fu J, Zhang B, Cheng S, Wu M, et al. Salmonella Proteomic Profiling during Infection Distinguishes the Intracellular Environment of Host Cells. mSystems. 2019 Apr 9;4(2):e00314-18.

93. Margalit A, Kavanagh K, Carolan JC. Characterization of the Proteomic Response of A549 Cells Following Sequential Exposure to Aspergillus fumigatus and Pseudomonas aeruginosa. J Proteome Res. 2020 Jan 3;19(1):279–91.

94. Menon D, Singh K, Pinto SM, Nandy A, Jaisinghani N, Kutum R, et al. Quantitative Lipid Droplet Proteomics Reveals Mycobacterium tuberculosis Induced Alterations in Macrophage Response to Infection. ACS Infect Dis. 2019 Apr 12;5(4):559–69.

95. Sorrentino JT, Golden GJ, Morris C, Painter CD, Nizet V, Campos AR, et al. Vascular Proteome Responses Precede Organ Dysfunction in a Murine Model of *Staphylococcus aureus* Bacteremia. mSystems. 2022 Aug;7(4):e00395-22.

96. Lapek JD, Mills RH, Wozniak JM, Campeau A, Fang RH, Wei X, et al. Defining Host Responses during Systemic Bacterial Infection through Construction of a Murine Organ Proteome Atlas. Cell Syst. 2018 May 23;6(5):579-592.e4.

97. Guiberson ER, Weiss A, Ryan DJ, Monteith AJ, Sharman K, Gutierrez DB, et al. Spatially Targeted Proteomics of the Host–Pathogen Interface during Staphylococcal Abscess Formation. ACS Infect Dis. 2021 Jan 8;7(1):101–13.

98. Willems P, Fels U, Staes A, Gevaert K, Van Damme P. Use of Hybrid Data-Dependent and -Independent Acquisition Spectral Libraries Empowers Dual-Proteome Profiling. J Proteome Res. 2021 Feb 5;20(2):1165–77.

99. Bonne Køhler J, Jers C, Senissar M, Shi L, Derouiche A, Mijakovic I. Importance of protein Ser/Thr/Tyr phosphorylation for bacterial pathogenesis. FEBS Lett. 2020 Aug;594(15):2339–69.

100. Prust N, van der Laarse S, van den Toorn HWP, van Sorge NM, Lemeer S. In-Depth Characterization of the *Staphylococcus aureus* Phosphoproteome Reveals New Targets of Stk1. Mol Cell Proteomics MCP. 2021;20:100034.

101. Karlsson R, Gonzales-Siles L, Gomila M, Busquets A, Salvà-Serra F, Jaén-Luchoro D, et al. Proteotyping bacteria: Characterization, differentiation and identification of pneumococcus and other species within the Mitis Group of the genus Streptococcus by tandem mass spectrometry proteomics. PLOS ONE. 2018 Dec 10;13(12):e0208804.

102. Singer SN, Ndumnego OC, Kim RS, Ndung'u T, Anastos K, French A, et al. Plasma host protein biomarkers correlating with increasing Mycobacterium tuberculosis infection activity prior to tuberculosis diagnosis in people living with HIV. EBioMedicine. 2022 Jan;75:103787.

103. Lanotte P, Perivier M, Haguenoer E, Mereghetti L, Burucoa C, Claverol S, et al. Proteomic Biomarkers Associated with Streptococcus agalactiae Invasive Genogroups. PLOS ONE. 2013 Jan 23;8(1):e54393.

104. Acar, Mustafa Burak et al. "A subtractive proteomics approach for the identification of immunodominant Acinetobacter baumannii vaccine candidate proteins." Frontiers in immunology vol. 13 1001633. 10 Nov. 2022

105. Walters MS, Mobley HL. Bacterial proteomics and identification of potential vaccine targets. Expert Rev Proteomics. 2010 Apr 1;7(2):181–4.

106. Acar MB, Ayaz-Güner Ş, Güner H, Dinç G, Ulu Kılıç A, Doğanay M, et al. A subtractive proteomics approach for the identification of immunodominant Acinetobacter baumannii vaccine candidate proteins. Front Immunol. 2022;13:1001633.

Chapter 2

Benchmarking proteomics acquisition and data analysis strategies

Ema Svetlicic, Carsten Jers, Ivan Mijakovic

Abstract:

Proteomics has advanced substantially, enabling detailed analysis of protein function and dynamics in complex biological samples. However, the pipelines are not standardized and variability across different laboratories was reported. This study benchmarks the performance of label-free quantification (LFQ) with data-dependent acquisition (DDA), tandem mass tag (TMT) labeling, and library free data-independent acquisition (DIA) approaches in proteomic analysis, evaluating their quantitative accuracy, reproducibility, and computational impact using software such as MaxQuant (MQ), Proteome Discoverer (PD), DIA-NN, and Spectronaut (SN). TMT labeling achieved the best quantitative performance and reproducibility but was the most time-consuming and expensive. DDA LFQ with PD provided similar accuracy and reproducibility, though with some missing values. Direct DIA using DIA-NN and Spectronaut identified the most proteins but had higher missing values and lower reproducibility, impacting quantification accuracy. The study also shows the importance of quality control in proteomics workflows to ensure accurate quantification and reproducibility, which can be effectively applied to biological settings.

2 Benchmarking proteomics acquisition and data analysis strategies for the purpose of host-pathogen interaction studies

2.1 Introduction

Proteomics has become an indispensable tool in biological and clinical research, offering insights into protein function, structure, and dynamics. The integration of mass spectrometry (MS) and bioinformatics has enabled the identification and quantification of thousands of proteins in complex biological samples. This achievement has been enabled by the constant development of new techniques and methodologies. Over the last decade, the field has seen an expansion in both academic and commercial methodologies of sample preparation (1–4), development of new instrumentation (5–8), and advances in computational tools driven by machine learning (9–11). Despite these advancements, proteomics workflows are not standardized, leading to variability in results across different laboratories and platforms (12,13). In sample preparation procedures, for example, choice of detergents, solubilization solutions, filtration and/or precipitation can lead to varying coverage of protein identification and biases related to molecular weight, protein pI, and hydrophobicity of the peptides (2,14,15). The choice of instrumentation and data acquisition should also be considered, because different MS instruments vary in their sensitivity, resolution and data acquisition capabilities, therefore affecting the depth of proteome identification and quantification (16,17). Furthermore, the stochastic nature of MS analysis means that results can differ even when the same sample is analyzed repeatedly on the same instrument (18,19). While there are several instruments widely used in MS based proteomics such as SCIEX TripleTOF 5600 (20) and TripleTOF 6600+ (21), Waters Synapt G2-Si TOF (22) and Bruker timsTOF (23), the most commonly used instruments in bottom-up proteomics are Thermo Fishers scientific Orbitrap instruments (24). Since its first commercial introduction in 2005 (25), the instruments have undergone numerous improvements and new models have been released. Most recently, the introduction of Orbitrap tribrid series consisting of three analyzers: quadrupole, linear ion trap and Orbitrap has greatly improved quantitative accuracy in tandem mass tags (TMT) based quantification (26). Moreover, integration of high-field asymmetric waveform ion mobility spectrometry (FAIMS) pro technology in combination with novel chromatographic solutions was shown to increase the detection sensitivity while shortening the LC gradients (27,28).

In addition to different instrumentation, data acquisition and subsequent bioinformatic analysis strategies affect the detection of peptides and the accuracy of quantification. For label free quantification (LFQ) in combination with data-dependent acquisition (DDA), commonly

used software packages are MaxQuant (MQ) (29) and Proteome Discoverer (PD, Thermo Fisher Scientific). While both softwares achieve comparable quantitative accuracy, MQ has been shown to exhibit more missing values and generally lower proteome coverage (30). Missing values, however, remain an issue associated with LFQ DDA analysis. Label-based approaches such as TMT circumvent the issues by simultaneous analysis of multiple samples in a single mass spectrometry run. The analysis is usually accompanied by offline fractionation, therefore increasing proteome coverage and reducing variability. However, TMT reagents are expensive, and the process requires a more complicated sample preparation step and more specialized instruments which increase the overall time and cost of the analysis (31,32). Data independent acquisition (DIA) approaches aim to increase reproducibility and proteome coverage because unlike DDA, which selects specific ions for fragmentation, DIA systematically fragments all ions within a given mass range. Moreover, due to its ability to fragment all ions in predefined mass windows, DIA increases throughput by allowing for shorter gradients and faster analysis (33). Traditionally, DDA analysis has been necessary to create spectral libraries for DIA analysis (34). This reliance on DDA carries over its inherent stochastic nature, which can introduce variability and limit reproducibility. Additionally, generating deep, project-specific libraries with DDA requires substantial amounts of sample material and considerable MS time, making it a resource-intensive process. In recent years, the development of spectral libraries through machine learning has become a major focus in the field. By creating spectral libraries *in silico*, researchers can generate spectral libraries from publicly available repositories of protein sequences, thereby eliminating the need for DDA analysis and its associated constraints (35–37). This type of DIA analysis is called library-free DIA or direct DIA. Several machine learning algorithms are available to generate spectral libraries from a protein sequence, however fully integrated library free search is for now only available in DIA-NN (38) and Spectronaut (SN, Biognosys) (39).

Since the library-free DIA potentially offers a high throughput, accurate and more cost-effective analysis, the aim of this study was to compare this newly developed method to the more traditional approaches LFQ DDA and TMT. Additionally, most commonly used software packages were compared to estimate the influence of computational analysis on the final result. Benchmarking experiments usually involve mixing proteins from different species in various ratios and performing relative quantifications on the proteome mixture. To make a comparison informative, it is necessary to perform global comparisons among software solutions by using global metrics, including the total number of proteins quantified, the number of missing values

produced, the coefficient of variation (CV) in technical replicates, and accuracy of reported quantification values.

2.2 Materials and Methods

2.2.1 Bacterial strains and cell culture

The strain used in this study was *S. aureus* USA300 FPR3757. Bacteria were grown overnight in Tryptic Soy Broth (TSB, Merck, Germany) at 37°C in 250 ml flasks with shaking at 200 rpm. The next day, bacterial culture was pelleted by centrifugation at 7000 g for 5min? and washed three times with ice cold PBS. The pellet was stored at -80 ºC until further used. Lung epithelial cell line A549 was cultivated in the growth medium DMEM/F-12 at 37 °C in saturated air humidity with 5% CO_2. The cell line was initially propagated in a T-75 flask until approximately 70% confluency and the medium was changed every two days. The cells were detached using a cell scraper, centrifuged and washed with ice cold PBS.

2.2.2 Sample preparation for proteome analysis

Pellets from 50 mL culture were resuspended in 100 µl of lysis buffer (6 M guanidinium hydrochloride, 10 mM Tris(2-carboxyethyl)phosphine hydrochloride, 40 mM 2-chloroacetamide, and 50 mM HEPES pH 8.5) vortexed and transferred into Bashing Bead Lysis Tube (Zymo Research, 0.1 & 0.5 mm). The tubes were subjected to bead beating using Precellys® 24 Touch homogenizer (Bertin Technologies) for 1 min at 4500 rpm. In total, 5 cycles of homogenization were performed with 2 minutes of cooling between cycles. The tubes were then vortexed for 30 s and subjected to sonication for 30 min (60 s “on”, 30 s “off”) using a Bioruptor (Diagenode, USA). The cell debris was removed by centrifugation at 10 000 g for 15 min. The protein amount was measured by Bicinchoninic acid (BCA) assay using BSA as a standard. Protein samples from *S. aureus* cells and A549 cells were mixed in following ratios: 1:1, 1:2, 1:5 and 1:10 which are termed groups A, B, C and D, respectively. The sample was diluted ten times with digestion buffer (50 mM HEPES pH 8.5) containing proteomics grade Trypsin/LysC (1:100, Sigma-Aldrich) and incubated overnight at 37 ºC. The next day, the samples were diluted 1:2 in 2% TFA. The peptides were desalted using SOLAµ cartridges as follows. The columns were primed with 100% methanol, equilibrated with 80% acetonitrile/0.1% TFA and washed two times with 3% acetonitrile/0.1% formic acid. The peptides were loaded and washed twice with 3% acetonitrile/0.1% formic acid. Finally, peptides were eluted with 60% acetonitrile/0.1% formic acid, dried and subsequently frozen at −20˚C until analysis by MS. For label free approaches (both DDA and DIA), peptides were

reconstituted in 0.1% formic acid and stored at -80 ºC until analysis. For label-based TMT approach, the manufacturer's instructions were followed. In brief, peptides were resuspended in 20 µl 50 mM HEPES by vortexing and sonicating followed by the addition of 20 µl of TMTpro 16plex Label Reagents diluted to 10 mM as described in (31). Samples were incubated for 1 hour at room temperature while shaking at 1200 rpm. The reaction was stopped by addition of 5 µl of 5 % hydroxylamine in 200 mM Triethylammonium bicarbonate buffer (TEAB) and incubation for 15 min at room temperature. The samples were then mixed in the same ratio and diluted with 2% TFA to bring acetonitrile concentration below 3%. The sample was then desalted using SOLAµ cartridges as described above. The multiplexed sample was fractioned using an offline Thermo Fisher Ultimate3000 liquid chromatography system at a flowrate of 5 µl/min over a 60 min gradient (from 5% to 35% acetonitrile), while collecting fractions every 2 min. The resulting 20 fractions were pooled into 10 final fractions (symmetrical pairing), dried in a vacuum centrifuge, and re-constituted in 1% TFA, 2% acetonitrile for individual MS analysis.

2.2.3 Label free quantification with data dependent acquisition

Peptides were reconstituted in 0.1% formic acid (FA) and loaded onto a 2 cm C18 trap column (Thermo Fisher 164946), connected in-line to a 15 cm C18 reverse-phase analytical column (Thermo EasySpray ES904) using 100% Buffer A (0.1% FA in water) at 750 bar, using the Thermo EasyLC 1200 HPLC system, and the column oven operating at 30 °C. Peptides were eluted over a 140 min gradient ranging from 6 to 60% of 80% ACN, 0.1% FA at 250 nl/min, and the Exploris 480 instrument (Thermo Fisher Scientific) with FAIMS Pro Interface (ThermoFisher Scientific) switched between compensation voltage of -50 V and -70 V with cycles of 38 and 28 scans. Full MS spectra were collected at a resolution of 60,000, with an AGC target of 300%, maximum injection time set to auto, and a scan range of 375–1500 m/z. MS1 precursors with an intensity of $>5\times10^3$ and charge state of 2-6 were selected for MS2 analysis. Dynamic exclusion was set to 60 s, the exclusion list was shared between CV values and Advanced Peak Determination was set to 'on'. Precursors selected for MS2 were isolated in the quadrupole with a 1.6 m/z window. Ions were collected with maximum injection time set to auto and normalized AGC target set to 75%. Fragmentation was performed with a HCD normalized collision energy of 28% and MS2 at a resolution of 15,000.

The raw files were analyzed using Proteome Discoverer 2.5 (PD, Thermo Fisher Scientific). The spectra were matched against the concatenated Uniprot database of *S. aureus strain* USA300 proteome (UP000001939) and Homo sapiens reviewed sequences (Accessed

December 2023). LFQ was enabled in the processing. Search was performed with trypsin/P digestion and allowed a maximum of two missed cleavages database. Dynamic modifications were set as Oxidation (M) and Acetyl on protein N-termini. Cysteine carbamidomethyl was set as a static modification. The results were filtered to a 1% false discovery rate (FDR). All other parameters were set to default.

The raw files were also processed in MaxQuant 2.4. Database search was performed using the Andromeda platform, which is built into MQ against a target-decoy human and *S. aureus* database. Briefly, trypsin was fixed as the protease with a maximum allowance of two missed cleavages. Variable modifications were set for methionine oxidation, N-terminal acetylation, and phosphorylation on serine, threonine, and tyrosine residues. Cysteine carbamidomethylation was set as a fixed modification. A mass tolerance was set to 4.5 ppm and 0.5 Daltons for MS and MS/MS scans, respectively. A false discovery rate of 1 % was applied at the peptide and protein level. Finally, the LFQ along with match between runs was set to infer quantitative information. All other parameters were set to default.

2.2.4 TMT based quantification

The Evosep One liquid chromatography system was used for LC separation. The standard 44 min method was used, where peptide elution is carried out with a 250 nl/min flow rate. A 15 cm × 75 µm ID column (PepSep) with 1.9 µm C18 beads (Dr. Maisch, Germany) and a 10 µm ID silica electrospray emitter (PepSep) was used. Mobile phases A and B were 0.1% formic acid in water and 0.1 formic acid % in Acetonitrile. Spectra were acquired with an Orbitrap Eclipse™ Tribrid™ Mass Spectrometer (ThermoFisher Scientific) running Tune 3.4, DD-SPS-MS3 mode, with FAIMS Pro interface (ThermoFisher Scientific) cycling between compensation voltage of -50 and -70 V, every 2 s. MS1 spectra were acquired at 120 000 resolution with a scan range from 375 to 1500 m/z, normalized AGC target of 100% and maximum injection time of 50ms. To filter MS1 precursors, monoisotopic peak determination was set to peptide, intensity threshold to 5x10e3 charge state to 2–7 and dynamic exclusion to 120 s with the single charge state option activated. Precursor ions were isolated in the quadrupole with a 0.7 m/z window, collected to a normalized AGC target of 300% or maximum injection time 35ms, and subsequently fragmented with 30 normalized HCD collision energy. Spectra were acquired in the ion trap in turbo mode. For the MS3 spectrum, the 10 most intense fragment ions were selected from the MS2 spectrum. The precursor ion for the MS3 scan was isolated from the MS1 scan with a 1.2 m/z window collected to a normalized AGC target of 300% or maximum injection time 86ms, and subsequently fragmented as defined in MS2

settings. MS3 spectra were acquired in Orbitrap with 50 000 resolution in a scan range of 100–500 m/z.

The raw files were analyzed using Proteome Discoverer 2.5. TMT SPS MS3 reporter ion quantitation was enabled in the processing and consensus steps, and spectra were matched against the concatenated database. Dynamic modifications were set as Oxidation (M), Deamidation (N, Q), and Acetyl on protein N-termini. Cysteine carbamidomethyl and TMTpro 16plex were set as a static modification. All results were filtered to a FDR of 1%.

2.2.5 Data independent acquisition

Peptides were reconstituted in 0.1% formic acid (FA) and loaded onto a 2 cm C18 trap column (Thermo Fisher 164946), connected in-line to a 15 cm C18 reverse-phase analytical column (Thermo EasySpray ES904) using 100% Buffer A (0.1% FA in water) at 750 bar, using the Thermo EasyLC 1200 HPLC system, and the column oven operating at 30 °C. Peptides were eluted over a 70 min gradient ranging from 6 to 60% of 80% ACN, 0.1% FA at 250 nl/min, The Exploris 480 instrument mass spectrometer was operated in positive mode with the FAIMSPro interface compensation voltage set to −45 V. MS1 scans were carried out at 120,000 resolution with an automatic gain control (AGC) of 300% and maximum injection time set to auto. For the DIA isolation window survey a scan range of 500–900 was used and 400–1000 was used for the rest of the experiments. Higher energy collisional dissociation (HCD) was used for precursor fragmentation with a normalized collision energy (NCE) of 33% and MS2 scan AGC target was set to 1000%.

DIA-NN and Spectronaut 16 and 17 versions were used to process raw data files. In DIA-NN was employed for DIA data analysis. Spectral library was generated from the FASTA file by enabling “FASTA digest for library free search”. The database was searched with maximal two missed cleavages. Peptide length was restricted from 7 to 30 peptides, and the precursor m/z range was set from 360 to 1300. N-terminal methionine excision was enabled. Cysteine carbamidomethylation was selected as a fixed modification, and methionine oxidation and N-terminal acetylation as variable modifications. The maximum number of variable modifications was set to five. All other parameters were default settings, including the 1% precursor FDR and enabled match between runs (MBR). The DIA-NN report “report_pg.txt” was used for all calculations.

In Spectronaut, direct DIA analysis was run on pipeline mode using modified BGS factory settings. Specifically, the imputation strategy was set to “None” and Quantity MS level

was changed to MS1. Trypsin and Lys-C were se selected as digestion enzymes and N-terminal protein acetylation and methionine oxidation were set as variable modifications. Carbamidomethylation of cysteines was set as fixed modification. All other parameters were set to default.

2.2.6 Data postprocessing and quantification

The protein table was further processed in R studio (V 2021.09.1) using packages diplyr, naniar, ggplot2 and corrplot. The processing included log2 transformation, removing missing values, Pearson correlation, Principal component analysis (PCA) and visualization. Differential expression analysis was performed using the limma package (V 3.48.3). A design matrix was constructed to model the experimental conditions. Specifically, contrasts were set for B-A, C-A and D-A. Linear models were fitted to the data using the lmFit() function from the limma package, incorporating the design matrix. The fitted linear model was used to compute the specified contrasts via the contrasts.fit() function. Empirical Bayes moderation of standard errors and log-fold changes was performed using the eBayes() function. A function "decideTests" was used to Identify which genes were significantly differentially expressed for each contrast from a fit object containing p-values and test statistics. The p values were adjusted with the Benjamini-Hochberg (BH) procedure. Proteins, which displayed a statistically significant abundance change (adjusted p-value ≤ 0.05) were termed differentially expressed proteins.

2.3 Results

2.3.1 Experimental overview

To assess the impact of data acquisition, software suite, peptide labeling on protein identification, reproducibility and quantitative accuracy, a benchmarking dataset was produced by mixing *S. aureus* and human proteins in known ratios. The proteins were mixed in ratios 1:1, 1:2, 1:5, 1:10, further referred to as samples A, B, C and D, respectively. The purpose of this assessment is to establish the most optimal method for identification of host-pathogen interactions using specifically *S. aureus* and A549 cell line. The tested methods included LFQ DDA and DIA method, which are relatively simply performed in terms of sample preparation. Library free DIA was chosen to evade the need for laborious generation of spectral libraries and to ensure the consistency of data analysis for potential future studies. In addition to testing various data acquisition methods, most commonly used softwares were tested. In case of DDA this were PD and MQ and for direct DIA, DIA-NN and Spectronaut were utilized. A TMT

approach was also assessed, which required labeling procedure of each sample, multiplexing into one sample and offline fractionation before MS analysis. In label free approaches, technical replicates were analyzed by injecting each sample three times. Moreover, the same samples were injected for LFQ DDA and DIA, to ensure that any variability would come from data acquisition and not variations in sample preparation. Since the TMT approach requires a different sample preparation, the peptide mixture was labeled in triplicates and subjected to offline fractionation. The output of the software analysis and protein quantification were protein tables which were processed by the same R script to include log2 transformation, missing, values removal and differential expression analysis. The schematic representation of the applied workflow is shown in Figure 2.1. The R script is available in Sciencedata.dk repository (link in Additional information).

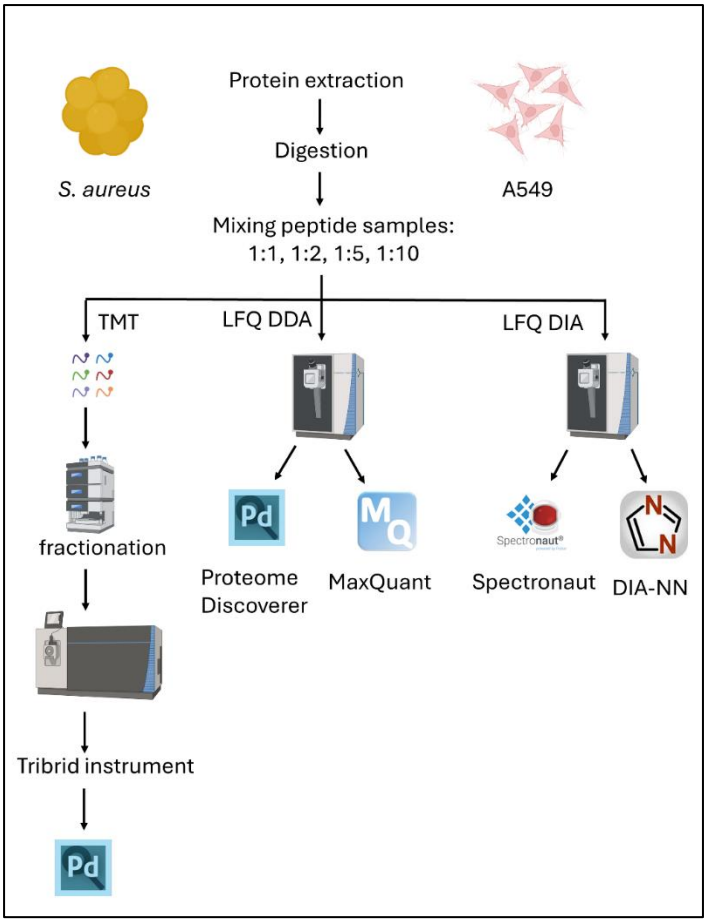


Figure 2.1: Workflow of the benchmarking experiment. Proteins were extracted from *S. aureus* and A549 cells, digested to peptides and mixed in defined ratios of 1:1, 1:2, 1:5 and 1:10, the mixed samples were named A, B, C and D, respectively. Each group was analyzed in triplicates. For LFQ DDA and DIA, Orbitrap Exploris 480 was utilized for analysis as described in materials and methods. Meanwhile TMT data were acquired on Orbitrap Eclipse Tribrid from pre-fractionated samples to perform quantification on MS3 level.

2.3.2 Protein identifications and variability

For each tested approach, the number of identified proteins for both *S. aureus* and human cells was assessed after a 1% FDR cutoff was applied. It can be seen on Figure 2.2 that DIA approaches yielded the highest number of protein identifications for all groups, followed by TMT. However, TMT based approach identified the same number of proteins across all groups, while in DIA data, the identification of human proteins was influenced by the proportion of proteins in the sample, therefore in sample A where bacterial and human proteins were mixed in 1:1 ratio, there are more than 4500 proteins identified, while there are

considerably less identified proteins than in Sample D (1:10) where around 3100 proteins were identified. This trend can also be seen in DDA data, where the difference in number of protein identifications between group A and D is up to 500 proteins. LFQ DDA approach had the fewest identified proteins, especially when in combination with MQ algorithm which identified on average 15% less bacterial and 10% less human proteins than TMT and DIA approaches, respectively, when compared to the entire proteome. The data completeness was assessed by plotting the missing values count (Figure 2.3A) which is important when comparing quantitative values between conditions. Interestingly, despite having lower number of identified proteins, the total amount of missing values was lower in LFQ DDA PD approach (5.8%) compared to DIA analysis (12.8%). The MQ algorithm led to the highest amount of missing values (18.5%). In case of the TMT approach, since peptides from all samples are coeluted, missing values are minimal or none as in this case. To enable comparison of protein intensities between groups, proteins with missing values were removed. Although imputation methods can replace missing values with estimated values derived from the intensity distribution, imputation was not performed in this study to avoid introducing an additional variable. Figure 2.3B depicts the number of proteins after missing value removal. By removing the proteins with missing values, protein counts of bacteria in all approaches did not decrease substantially. However, human protein count decreased, especially in DIA approaches where a high number of missing values were found in samples C and D where lower ratios of human proteins were analyzed. TMT based analysis, consistently identified a high number of human (>3500) and bacterial (>1700) proteins. Another important quality measure of MS analysis is reproducibility, measured by coefficient of variation (CV%) distribution which is calculated from the standard deviation of protein intensities divided by the mean. CV% distributions are depicted as violin plots in Figure 2.4, showing that the TMT approach achieved on average the lowest CV (15 %), compared to the highest CV when using SN (20%). All approaches showed a mean CV below 25%; however, certain proteins exhibited a CV of up to 200%, indicating that their quantification is highly variable across different measurements. To assess whether the CV% is influenced by protein abundances, scatter plots were generated for all groups (A, B, C, and D) and across all tested pipelines. The resulting plots, presented in Figure S2.1, demonstrate a trend where proteins with lower average intensities tend to exhibit higher CV%.

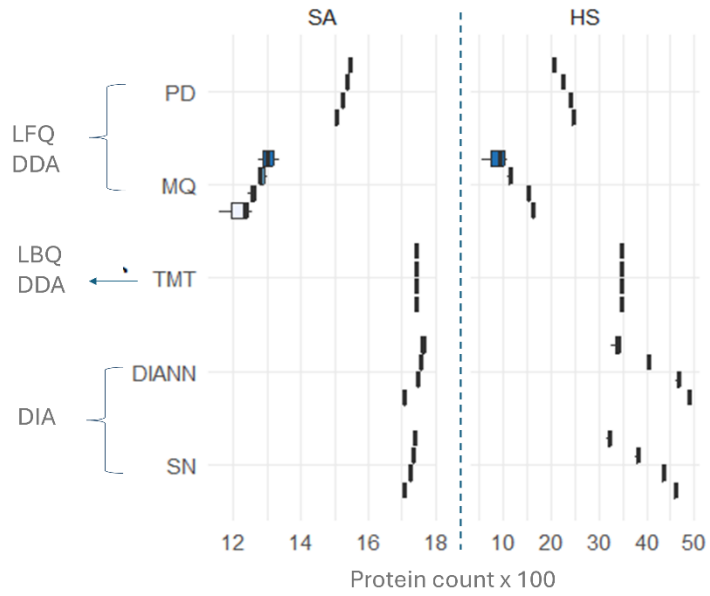


Figure 2.2: Protein counts distribution for each mix (group) and each analysis workflow separated by species: Each boxplot represents a group as follows (from top to bottom): D, C, B, and A . Protein counts are depicted as boxplots of replicates with a mean value, lower and upper quartile and minimum and maximum values. HS: *Homo sapiens*, SA: *Staphylococcus aureus* LFQ- label free quantification, LBQ- label based quantification, DDA- data dependent acquisition, DIA- data independent acquisition, PD- Proteome Discoverer, MQ-MaxQuant, TMT-tandem mass tags, DIANN- DIA-NN, SN-Spectronaut,

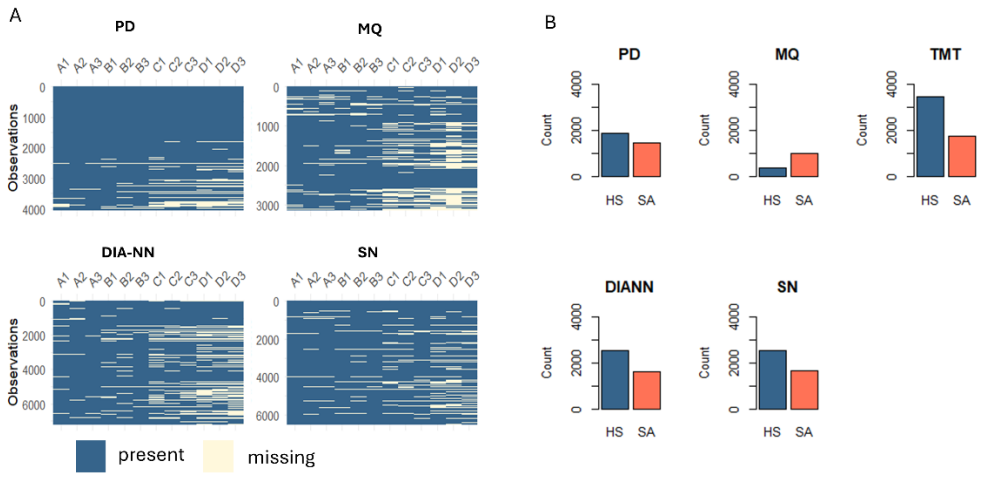


Figure 2.3: Impact of missing values on the datasets. A) Distribution of missing values across the tested approaches B) Protein counts in tested approaches after excluding proteins with missing values in any of the samples. HS: *Homo sapiens*, SA: *Staphylococcus aureus* LFQ- label free quantification, LBQ- label based quantification, DDA- data dependent acquisition, DIA- data independent acquisition, PD- Proteome Discoverer, MQ-MaxQuant, TMT-tandem mass tags, DIANN- DIA-NN, SN-Spectronaut,

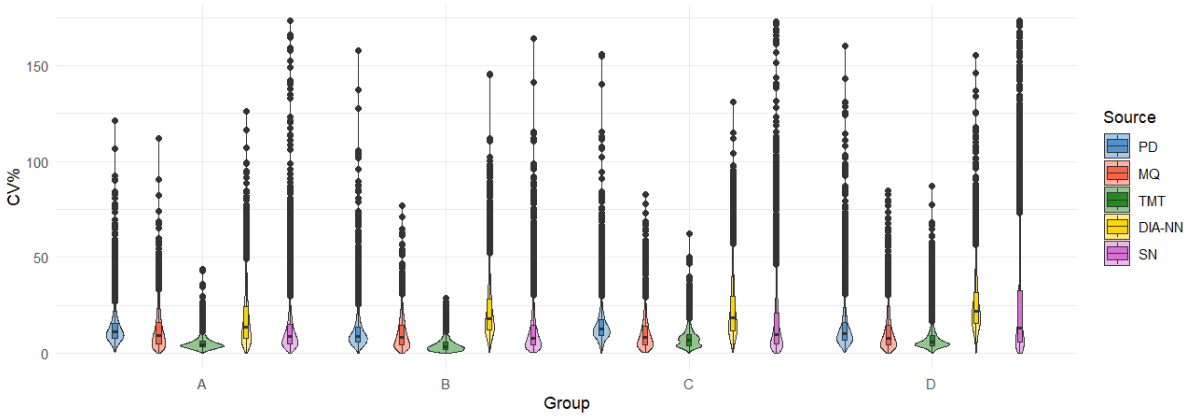


Figure 2.4: Violin plots depicting the distribution of the percent coefficient of variation (CV%). The central box shows the interquartile range (IQR), containing the middle 50% of the data, with the bottom and top edges marking the first (Q1) and third (Q3) quartiles. A line inside the box indicates the median (Q2). Whiskers extend from the box to the minimum and maximum values within 1.5 times the IQR from the quartiles. Data points outside this range are considered outliers and are plotted individually.

2.3.3 Quantitative accuracy

In quantitative proteomics, protein coverage and protein identifications are important, however for a successful experiment, a reliable and accurate quantification is crucial. The data was therefore log2 transformed, and differential expression analysis was performed to estimate the log2 ratio of groups B, C, and D relative to group A. The expected ratio for contrasts B-A, C-A and D-A were -2.45, -1.58, -0.59 for human and 0.85,0.69,0.42 for bacterial proteins, respectively. Statistically significant differential expression is measured by adjusted p-value for each protein and it would be expected that all proteins changed their expression, however this was not the observed outcome (Figure 2.5). The TMT-based approach shows the lowest amount of non-significantly regulated proteins with only a few hundred in B-A group where the expected ratio is the lowest. Both DIA approaches, DIA-NN and SN, exhibit a substantial number of proteins for which no statistically significant change in protein abundance was detected (adjuster p value >0.05). This lack of statistical significance coincides with the highest variability in these approaches, as measured by the CV%. The LFQ DDA strategy also shows numerous proteins with adjusted p value higher than 0.05, primarily in the B-A contrast and mostly for downregulated human proteins, while the upregulated proteins remain consistent with the number of identified bacterial proteins in the PD and MQ workflow.

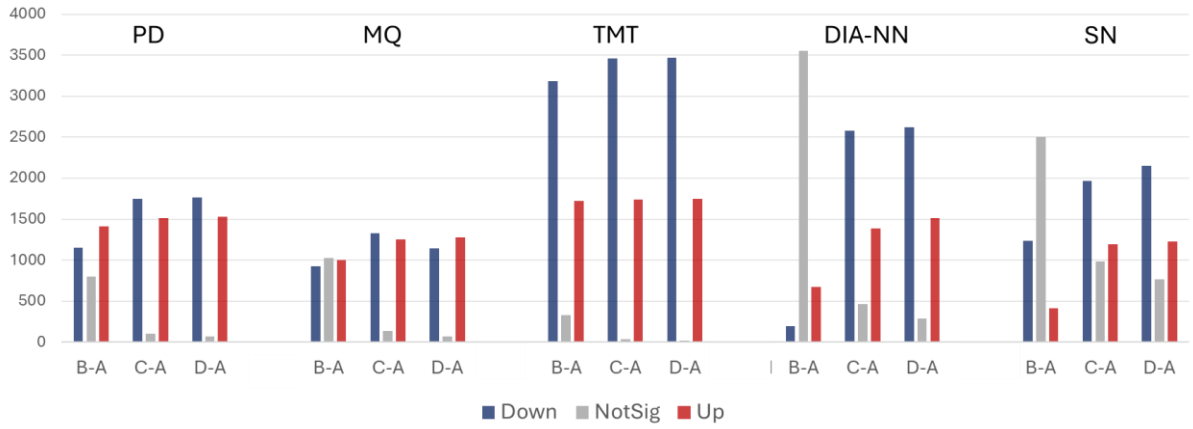


Figure 2.5: Results of limma differential expression analysis. The figure displays the number of differentially expressed proteins, determined from adjusted p-values and test statistics. To term protein up or downregulated, adjusted p values should be lower than 0.05. Down-downregulated NotSig- Non-significantly regulated Up- Upregulated. PD- Proteome Discoverer, MQ-MaxQuant, TMT-tandem mass tags, DIANN- DIA-NN, SN-Spectronaut

Next, calculated log2 fold changes for contrasts B-A, C-A and D-A were plotted (Figure 2.6) for each proteomics approach tested, resulting in boxplots depicting distribution of the values for all proteins. The plots also contain expected ratios highlighting the deviation of the measured data from the theoretical values. PD and MQ approaches show similar distribution of log2 ratios where the mean value in the log2 fold change matches the expected value for human proteins, while it is slightly overestimated for bacterial proteins. The expected ratio also matches the interquartile range in the boxplot of TMT, DIA-NN and SN approach. However, in the case of SN, there are several outliers exhibiting ratios which are 5-10 times higher than expected.

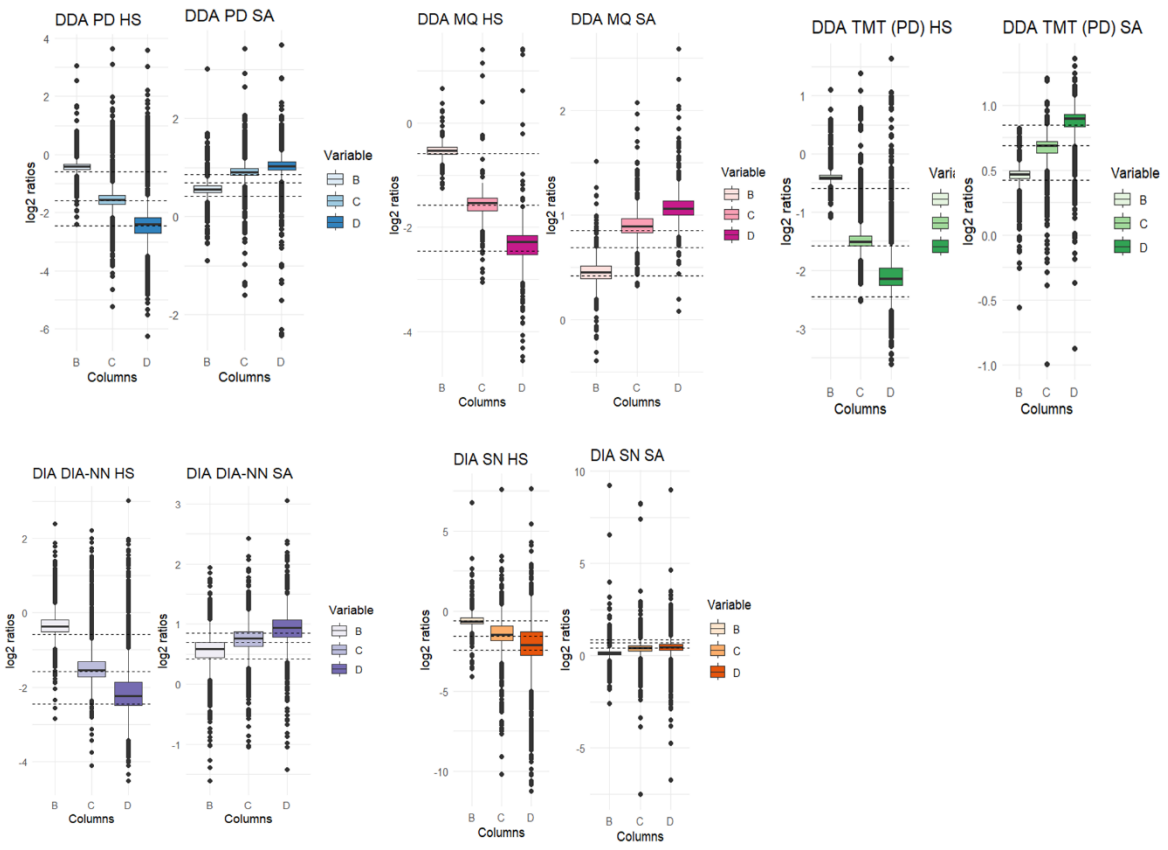


Figure 2.6: Log-transformed ratios (log2(X/A)) of proteins plotted for each MS workflow where X can be D, C or B. Data for human (HS) and *S. aureus* proteins (SA) are shown in separate plots. Dashed lines represent the expected log2(X/A) values for human and bacterial proteins. PD- Proteome Discoverer, MQ-MaxQuant, TMT-tandem mass tags, DIANN- DIA-NN, SN-Spectronaut

To have a more precise estimate of the log2 ratio accuracy, a mean relative error was calculated (MRE) and median relative error were calculated (MedRE) with following equations

$$\mathrm{MRE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\mathrm{Observed_i\text{-}Expected_i}}{\mathrm{Expected_i}}\right| \qquad [1]$$

$$\mathrm{MedRE} = \mathrm{median}\left(\left|\frac{\mathrm{Observed_i\text{-}Expected_i}}{\mathrm{Expected_i}}\right|\right) \qquad [2]$$

The measures provide an average percentage error, showing how much, on average, the observed values deviate from the expected values relative to the expected values. The calculated values for bacterial and human proteins for each approach are shown in Table 2.1. The TMT approach exhibited the lowest MRE and MedRE for both human and bacterial proteins, which was more pronounced for ratio of bacterial proteins than for human proteins where there was not substantial difference between approaches. The MRE and MedRE values also varied depending on the contrast where contrast C-A exhibited lowest errors. DIA-NN and SN approaches surprisingly exhibited comparable percentage errors, despite a high number of proteins with expression not significantly altered (adjusted p value > 0.05).

Table 2.1: Mean and median relative error calculated for log2 ratios of contrast B-A, C-A and D-A for all tested approached. Percentage errors were calculated from average deviation of measured ratio values from expected values. PD- Proteome Discoverer, MQ-MaxQuant, TMT-tandem mass tags, DIANN- DIA-NN, SN-Spectronaut

Approach	Mean relative error %			Median relative error%		
	B-A	C-A	D-A	B-A	C-A	D-A
Human proteins						
PD	-84.0227	-7.59171	284.8636	-82.7073	-1.95525	305.5288
MQ	-79.2293	-10.5596	252.2289	-78.9562	-7.53707	267.2787
TMT	-84.0111	-7.69132	248.0007	-83.5954	-4.19014	263.4118
DIANN	-86.5959	-8.29483	252.7615	-84.9962	-2.16133	279.6638
SN	-75.5652	-9.7502	254.5833	-73.3816	-6.29159	257.9917
Bacterial proteins						
PD	-33.7373	30.86637	143.1327	-34.492	31.25059	145.5706
MQ	-47.9407	27.86752	152.0992	-47.5011	28.10001	152.0595
TMT	-46.5151	-4.55663	105.0642	-44.9369	-0.10747	113.0365
DIANN	-33.6884	7.7526	117.4531	-31.4579	10.36462	124.0471
SN	-84.4386	-43.0404	1.35733	-83.3502	-37.6815	12.23034

2.4 Discussion and conclusions

In this study, a benchmark strategy was developed which provided an assessment of quantification accuracy and performance for bottom-up proteomics experiments using two species: *H. sapiens* (A549 cells) and *S. aureus*. Benchmarking proteomic experiments involves analyzing samples with predefined fold changes between groups by mixing them in different ratios and then comparing it to the experimental values (40). The comparison was performed on the protein level, while peptide and precursor comparison was omitted in this study, considering most researchers work with protein level data in bottom-up proteomics experiments and all software packages implement protein inference (41). In this experiment, several global metrics were compared, including protein identification, missing values, CV% and quantification accuracy among commonly used workflows available in our facilities. This study evaluated LFQ DDA workflow which utilizes two different software packages PD and MQ, a TMT workflow (DDA) as well as library free DIA workflow with DIA-NN and SN database search. While numerous published studies have benchmarked proteomics workflows

(33,40,42–44), the performance of these methods was compared with the available instrumentation and developed methods within our setting. Another aim was to evaluate the relatively new approach of library free DIA and to compare it to more traditional approaches like LFQ DDA and TMT based approach and to assess whether it can overcome some of their limitations and provide high-throughput, cost-effective analysis and more reproducible and accurate quantification. In this study, the qualitative measure of the number of identified proteins was in favor of DIA. In total, DIA identified the highest number of proteins both with DIA-NN and SN packages. However, there was variance in proteome coverage between different groups, so when removing all the missing values the number of proteins identified was similar to the TMT approach. Since the proteins with missing values were observed in groups C and D, with lower amounts of human proteins, it indicates that peptides with low intensities are still overlooked by DIA analysis. Nevertheless, the DIA approach identified substantially more proteins than the DDA approach in shorter LC run time, thus achieving higher throughput than both LFQ DDA and TMT. The advantages of direct DIA over DDA in terms of proteome coverage were previously reported (37,45,46). Interestingly, DIA approaches had a higher number of missing values than a PD workflow, but lower than a MQ workflow. In general, MQ workflow performed the poorest in terms of missing values which is consistent with previous studies where PD outperformed MQ in quantification yield, reproducibility and dynamic range, however not in the quantification accuracy which was slightly better in MQ (47). Quantitative accuracy in proteomics depends on several factors including technical variability such as inconsistencies in the instrument performance resulting in intergroup variability. Moreover, data processing algorithms can potentially distort protein quantification through inaccurate peak detection, integration or protein inference. Post-processing of the data such as normalization and imputation can also have a determinative impact on the quantification (48,49). While these methods are usually performed to account for different sample loading and missing values, in this study they were excluded from the analysis to avoid introducing variables that could affect different datasets differently, potentially skewing the quantification. In this study, on average, the highest quantitative accuracy was reported by TMT quantification where log2FC values were closest to the expected values both for human and bacterial proteins as calculated by MRE and MedRE values. Nevertheless, all tested approaches achieved similar quantitative accuracy, with variations observed depending on the specific contrast. For instance, the SN approach exhibited the lowest accuracy in bacterial quantification for contrast C-A but the highest accuracy of all approaches for contrast D-A. In a typical proteomics experiment, however log2FC is not

enough to estimate whether protein is differentially regulated, but rather a statistical analysis should take place. There are numerous statistical packages developed and applied to proteomics research (50). One of the most commonly used packages, limma performs well for proteomics experiments with small number of replicates, such as three replicates per condition (51). By performing the differential expression analysis and evaluating p-values it was observed that thousands of proteins in DIA analysis did not exhibit adjusted p-values lower than 0.5, thus terming the peptides not significantly changed. As a result, there are very few proteins with significantly altered expression which is not in accordance with the premixed sample ratios. DIA approaches had the highest intergroup variance, as measured by CV, which could explain the high p-values.

In conclusion, the best quantitative performance was achieved with TMT labeling as measured by differential expression analysis (log2 fold change and adjusted p values). Moreover, TMT approach performed best in number of missing values and reproducibility. However, it was also the most time-consuming and the most expensive method which involved a labeling protocol, offline fractionation and separate analysis of the produced fractions. On the other hand, DDA LFQ analysis in combination with PD software package produced comparable results in quantitative accuracy and reproducibility as TMT, while the missing values were found in the dataset, this method had the lowest amount of missing values among other workflows (MQ, DIA-NN and SN). In contrast, direct DIA analysis in combination with DIA-NN and SN showed highest number of protein identifications, but had a relatively high number of missing values and lower reproducibility. The latter, in particular, could have impacted the quantification accuracy, as there were numerous false negative proteins which were deemed not significant due to high p-values. At present, with the advantage of MS based proteomics and the development of new acquisition and computational methods, it is important to continuously optimize and assess the accuracy of these technologies across different laboratories. This will enable more precise and comprehensive protein quantification, enhance data reproducibility, and allow for more accurate identification of biologically significant changes.

2.5 References

1. Wiśniewski J. R. (2018). Filter-Aided Sample Preparation for Proteome Analysis. Methods in molecular biology (Clifton, N.J.), 1841, 3–10..

2. Varnavides G, Madern M, Anrather D, Hartl N, Reiter W, Hartl M. In Search of a Universal Method: A Comparative Survey of Bottom-Up Proteomics Sample Preparation Methods. J Proteome Res. 2022 Oct 7;21(10):2397–411.

3. Poulsen JW, Madsen CT, Young C, Poulsen FM, Nielsen ML. Using guanidine-hydrochloride for fast and efficient protein digestion and single-step affinity-purification mass spectrometry. J Proteome Res. 2013 Feb 1;12(2):1020–30.

4. Hughes CS, Sorensen PH, Morin GB. A Standardized and Reproducible Proteomics Protocol for Bottom-Up Quantitative Analysis of Protein Samples Using SP3 and Mass Spectrometry. Methods Mol Biol Clifton NJ. 2019;1959:65–87.

5. Stewart HI, Grinfeld D, Giannakopulos A, Petzoldt J, Shanley T, Garland M, et al. Parallelized Acquisition of Orbitrap and Astral Analyzers Enables High-Throughput Quantitative Analysis. Anal Chem. 2023 Oct 24;95(42):15656–64.

6. Senko MW, Remes PM, Canterbury JD, Mathur R, Song Q, Eliuk SM, et al. Novel Parallelized Quadrupole/Linear Ion Trap/Orbitrap Tribrid Mass Spectrometer Improving Proteome Coverage and Peptide Identification Rates. Anal Chem. 2013 Dec 17;85(24):11710–4.

7. He Y, Shishkova E, Peters-Clarke TM, Brademan DR, Westphall MS, Bergen D, et al. Evaluation of the Orbitrap Ascend Tribrid Mass Spectrometer for Shotgun Proteomics. Anal Chem. 2023 Jul 18;95(28):10655–63.

8. Meier F, Brunner AD, Koch S, Koch H, Lubeck M, Krause M, et al. Online Parallel Accumulation-Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer. Mol Cell Proteomics MCP. 2018 Dec;17(12):2534–45.

9. Yang KL, Yu F, Teo GC, Li K, Demichev V, Ralser M, et al. MSBooster: improving peptide identification rates using deep learning-based features. Nat Commun. 2023 Jul 27;14(1):4539.

10. Orsburn BC. Proteome Discoverer—A Community Enhanced Data Processing Suite for Protein Informatics. Proteomes. 2021 Mar 23;9(1):15.

11. Bouwmeester R, Gabriels R, Van Den Bossche T, Martens L, Degroeve S. The Age of Data-Driven Proteomics: How Machine Learning Enables Novel Workflows. PROTEOMICS. 2020;20(21–22):1900351.

12. Bell AW, Deutsch EW, Au CE, Kearney RE, Beavis R, Sechi S, et al. A HUPO test sample study reveals common problems in mass spectrometry–based proteomics. Nat Methods. 2009 Jun;6(6):423–30.

13. Tabb DL, Vega-Montoto L, Rudnick PA, Variyath AM, Ham AJL, Bunk DM, et al. Repeatability and Reproducibility in Proteomic Identifications by Liquid Chromatography−Tandem Mass Spectrometry. J Proteome Res. 2010 Feb 5;9(2):761–76.

14. Zheng W, Yang P, Sun C, Zhang Y. Comprehensive comparison of sample preparation workflows for proteomics. Mol Omics. 2022;18(6):555–67.

15. Klont F, Bras L, Wolters JC, Ongay S, Bischoff R, Halmos GB, et al. Assessment of Sample Preparation Bias in Mass Spectrometry-Based Proteomics. Anal Chem. 2018 Apr 17;90(8):5405–13.

16. Szabó D, Schlosser G, Vékey K, Drahos L, Révész Á. Collision energies on QTof and Orbitrap instruments: How to make proteomics measurements comparable? J Mass Spectrom. 2021;56(1):e4693.

17. Tian S, Zhan D, Yu Y, Wang Y, Liu M, Tan S, et al. Quartet protein reference materials and datasets for multi-platform assessment of label-free proteomics. Genome Biol. 2023 Sep 7;24(1):202.

18. Piehowski PD, Petyuk VA, Orton DJ, Xie F, Ramirez-Restrepo M, Engel A, et al. Sources of Technical Variability in Quantitative LC-MS Proteomics: Human Brain Tissue Sample Analysis. J Proteome Res. 2013 May 3;12(5):2128–37.

19. Glaab E, Schneider R. RepExplore: addressing technical replicate variance in proteomics and metabolomics data analysis. Bioinformatics. 2015 Jul 1;31(13):2235–7.

20. Andrews GL, Simons BL, Young JB, Hawkridge AM, Muddiman DC. Performance Characteristics of a New Hybrid Quadrupole Time-of-Flight Tandem Mass Spectrometer (TripleTOF 5600). Anal Chem. 2011 Jul 1;83(13):5442–6.

21. Wang Z, Mülleder M, Batruch I, Chelur A, Textoris-Taube K, Schwecke T, et al. High-throughput proteomics of nanogram-scale samples with Zeno SWATH MS. Kornmann B, editor. eLife. 2022 Nov 30;11:e83947.

22. Zhang Y, Cai Q, Luo Y, Zhang Y, Li H. Integrated top-down and bottom-up proteomics mass spectrometry for the characterization of endogenous ribosomal protein heterogeneity. J Pharm Anal. 2023 Jan 1;13(1):63–72.

23. Yu F, Haynes SE, Teo GC, Avtonomov DM, Polasky DA, Nesvizhskii AI. Fast Quantitative Analysis of timsTOF PASEF Data with MSFragger and IonQuant. Mol Cell Proteomics. 2020 Sep 1;19(9):1575–85.

24. Greguš M, Koller A, Ray S, Ivanov AR. Improved Data Acquisition Settings on Q Exactive HF-X and Fusion Lumos Tribrid Orbitrap-Based Mass Spectrometers for Proteomic Analysis of Limited Samples. J Proteome Res. 2024 Jun 7;23(6):2230–40.

25. Makarov A, Denisov E, Kholomeev A, Balschun W, Lange O, Strupat K, et al. Performance Evaluation of a Hybrid Linear Ion Trap/Orbitrap Mass Spectrometer. Anal Chem. 2006 Apr 1;78(7):2113–20.

26. Yu Q, Paulo JA, Naverrete-Perea J, McAlister GC, Canterbury JD, Bailey DJ, et al. Benchmarking the Orbitrap Tribrid Eclipse for Next Generation Multiplexed Proteomics. Anal Chem. 2020 May 5;92(9):6478–85.

27. Bekker-Jensen DB, Martínez-Val A, Steigerwald S, Rüther P, Fort KL, Arrey TN, et al. A Compact Quadrupole-Orbitrap Mass Spectrometer with FAIMS Interface Improves Proteome Coverage in Short LC Gradients *. Mol Cell Proteomics. 2020 Apr 1;19(4):716–29.

28. Stejskal K, Op de Beeck J, Dürnberger G, Jacobs P, Mechtler K. Ultrasensitive NanoLC-MS of Subnanogram Protein Samples Using Second Generation Micropillar Array LC Technology with Orbitrap Exploris 480 and FAIMS PRO. Anal Chem. 2021 Jun 29;93(25):8704–10.

29. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 2008 Dec;26(12):1367–72.

30. Zhao L, Cong X, Zhai L, Hu H, Xu JY, Zhao W, et al. Comparative evaluation of label-free quantification strategies. J Proteomics. 2020 Mar 20;215:103669.

31. Zecha J, Satpathy S, Kanashova T, Avanessian SC, Kane MH, Clauser KR, et al. TMT Labeling for the Masses: A Robust and Cost-efficient, In-solution Labeling Approach. Mol Cell Proteomics MCP. 2019 Jul;18(7):1468–78.

32. Halder A, Verma A, Biswas D, Srivastava S. Recent advances in mass-spectrometry based proteomics software, tools and databases. Drug Discov Today Technol. 2021 Dec 1;39:69–79.

33. Fröhlich K, Brombacher E, Fahrner M, Vogele D, Kook L, Pinter N, et al. Benchmarking of analysis strategies for data-independent acquisition proteomics using a large-scale dataset comprising inter-patient heterogeneity. Nat Commun. 2022 May 12;13(1):2622.

34. Searle BC, Pino LK, Egertson JD, Ting YS, Lawrence RT, MacLean BX, et al. Chromatogram libraries improve peptide detection and quantification by data independent acquisition mass spectrometry. Nat Commun. 2018 Dec 3;9(1):5128.

35. Mehta D, Scandola S, Uhrig RG. Library-free BoxCarDIA solves the missing value problem in label-free quantitative proteomics [Internet]. bioRxiv; 2021 [cited 2024 Jul 8]. p.2020.11.07.372276. Available from: https://www.biorxiv.org/content/10.1101/2020.11.07.372276v4

36. Sinitcyn P, Hamzeiy H, Salinas Soto F, Itzhak D, McCarthy F, Wichmann C, et al. MaxDIA enables library-based and library-free data-independent acquisition proteomics. Nat Biotechnol. 2021 Dec;39(12):1563–73.

37. Mehta D, Scandola S, Uhrig RG. Direct data-independent acquisition (direct DIA) enables substantially improved label-free quantitative proteomics in Arabidopsis [Internet]. bioRxiv; 2020 [cited 2024 Jul 8]. p. 2020.11.07.372276. Available from: https://www.biorxiv.org/content/10.1101/2020.11.07.372276v2

38. Demichev V, Messner CB, Vernardis SI, Lilley KS, Ralser M. DIA-NN: Neural networks and interference correction enable deep proteome coverage in high throughput. Nat Methods. 2020 Jan;17(1):41–4.

39. Baker CP, Bruderer R, Abbott J, Arthur JSC, Brenes AJ. Optimizing Spectronaut Search Parameters to Improve Data Quality with Minimal Proteome Coverage Reductions in DIA Analyses of Heterogeneous Samples. J Proteome Res. 2024 Jun 7;23(6):1926–36.

40. Ramus C, Hovasse A, Marcellin M, Hesse AM, Mouton-Barbosa E, Bouyssié D, et al. Benchmarking quantitative label-free LC–MS data processing workflows using a complex spiked proteomic standard dataset. J Proteomics. 2016 Jan 30;132:51–62.

41. Chen C, Hou J, Tanner JJ, Cheng J. Bioinformatics Methods for Mass Spectrometry-Based Proteomics Data Analysis. Int J Mol Sci. 2020 Jan;21(8):2873.

42. Dowell JA, Wright LJ, Armstrong EA, Denu JM. Benchmarking Quantitative Performance in Label-Free Proteomics. ACS Omega. 2021 Feb 2;6(4):2494–504.

43. Lou R, Cao Y, Li S, Lang X, Li Y, Zhang Y, et al. Benchmarking commonly used software suites and analysis workflows for DIA proteomics and phosphoproteomics. Nat Commun. 2023 Jan 6;14(1):94.

44. Staes A, Maia T, Dufour S, Bouwmeester R, Gabriels R, Martens L, et al. Benchmarking DIA data analysis workflows [Internet]. 2023 [cited 2024 Jul 6]. Available from: http://biorxiv.org/lookup/doi/10.1101/2023.06.02.543441

45. O'Connell JD, Paulo JA, O'Brien JJ, Gygi SP. Proteome-Wide Evaluation of Two Common Protein Quantification Methods. J Proteome Res. 2018 May 4;17(5):1934–42.

46. Muntel J, Kirkpatrick J, Bruderer R, Huang T, Vitek O, Ori A, et al. Comparison of Protein Quantification in a Complex Background by DIA and TMT Workflows with Fixed Instrument Time. J Proteome Res. 2019 Mar 1;18(3):1340–51.

47. Palomba A, Abbondio M, Fiorito G, Uzzau S, Pagnozzi D, Tanca A. Comparative Evaluation of MaxQuant and Proteome Discoverer MS1-Based Protein Quantification Tools. J Proteome Res. 2021 Jul 2;20(7):3497–507.

48. Harris L, Fondrie WE, Oh S, Noble WS. Evaluating Proteomics Imputation Methods with Improved Criteria. J Proteome Res. 2023 Nov 3;22(11):3427–38.

49. Välikangas T, Suomi T, Elo LL. A systematic evaluation of normalization methods in quantitative label-free proteomics. Brief Bioinform. 2018 Jan 1;19(1):1–11.

50. van Ooijen MP, Jong VL, Eijkemans MJC, Heck AJR, Andeweg AC, Binai NA, et al. Identification of differentially expressed peptides in high-throughput proteomics data. Brief Bioinform. 2018 Sep 28;19(5):971–81.

51. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015 Apr 20;43(7):e47.

2.6 Supporting information

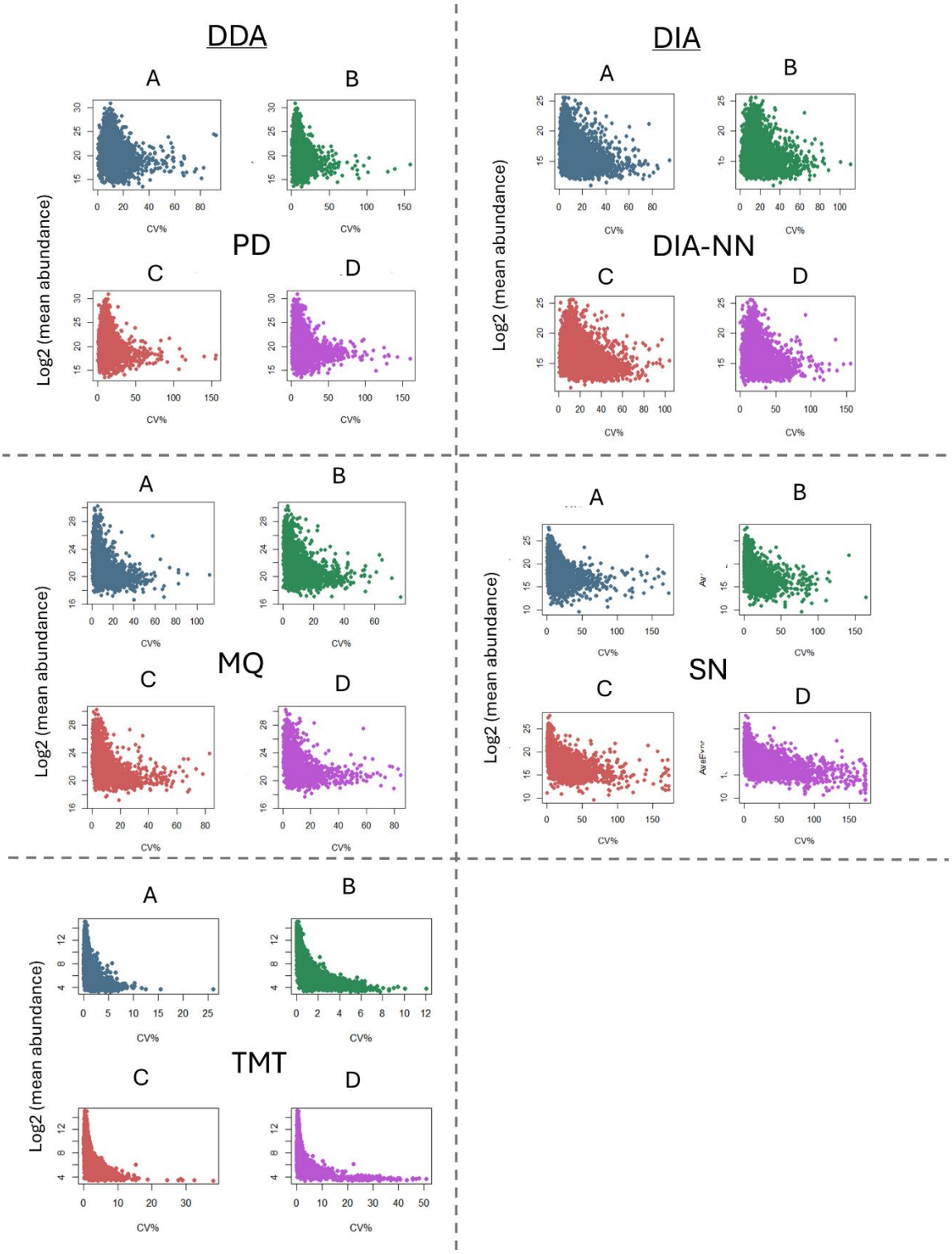


Figure S2.1: Scatter plot of log2 mean abundance versus coefficient of variation (CV%). The y-axis represents the log2-transformed mean abundance of proteins, while the x-axis represents the coefficient of variation (CV%). PD- Proteome Discoverer, MQ-MaxQuant, TMT-tandem mass tags, DIANN- DIA-NN, SN-Spectronaut

This Chapter is supported by the additional files which can be accessed through the following link:

https://sciencedata.dk/shared/f8f3bb3d8c4dc9516191d997aa72541f?download

Here is the brief description of the files:

File name	Description	Pipeline
DIA_DIANN_Postprocessing_script.R	R script	DIA DIA-NN
DIA_SN_Postprocessing_script.R	R script	DIA Spectronaut
LFQ_DDA_MQ_Postprocessing_script.R	R script	LFQ DDA MaxQuant
LFQ_DDA_PD_Postprocessing_script.R	R script	LFQ PD Proteome Discoverer
TMT_Postprocessing_script.R	R script	TMT
DIA_NN_report.pg_matrix.tsv	Protein table	DIA DIA-NN
proteinGroups_MQ.txt	Protein table	LFQ DDA MaxQuant
20230918_MK_Volta_MVL_FAIMS_2CV_15cm_140min_1756_DDA_500ng_protein.txt	Protein table	LFQ PD Proteome Discoverer
20230926_130044_1756_Ema_DIA_Spectronaut.txt	Protein table	DIA Spectronaut
20231005_EV_EVO_MVL_FAIMS_2CV_Endurance15_44min_1756_DDA_TMTpro_proteins.txt	Protein table	TMT
experimental_design_diann.txt	experimental design	DIA DIA-NN
experimental_design_mq.txt	experimental design	LFQ DDA MaxQuant
experimental_design_PD.txt	experimental design	LFQ PD Proteome Discoverer
experimental_design_sn.txt	experimental design	DIA Spectronaut
experimental_design_TMT.txt		TMT

Chapter 3

Proteome dynamics of *Staphylococcus aureus* during co-cultivation with human alveolar epithelial cells

Ema Svetlicic, Mohammed Ghalib Enayathullah, Carsten Jers, Ivan Mijakovic

Abstract:

Staphylococcus aureus is a versatile Gram-positive pathogen that is notably dangerous due to its ability to rapidly develop antibiotic resistance, evade host immune system and produce a variety of cell damaging toxins. OMICS technologies particularly proteomics, can offer new insights into the bacterium's adaptational network and response to environmental challenges. However, proteomic studies of bacteria in in infection conditions have been challenging due to highly abundant host proteins which can hinder detection of bacterial proteins. In this study, we employed a label-free quantitative proteomics approach to detect and quantify changes in the abundance of *S. aureus* proteins at different time points during infection of alveolar epithelial cells. By integrating specialized sample preparation, state of the art mass spectrometry and data analysis, we quantified more *S. aureus* proteins than in similar studies of *S. aureus*. This extensive data reveals how the bacterium modulates its proteome to adapt to the host environment, respond to host defenses and modulate the expression of known virulence factors

3 Proteome dynamics of *Staphylococcus aureus* during co-cultivation with human alveolar epithelial cells

3.1 Introduction

Staphylococcus aureus colonizes a vast number of tissues in the host, and while it can be a part of the normal microflora it also causes life threatening infections such as sepsis and pneumonia. The ability to infect and persist in the wide range of tissues and cells has been shaped by the complex mechanism of how the bacteria adapts to the host environment. To effectively develop therapies and other management strategies for infections caused by *S. aureus*, it is essential to gain an understanding of the complex pathophysiological interactions between the bacterium and its hosts (1). For the understanding of these interactions, infection models have been utilized, where the bacteria and its respective host can be put in a controlled environment and the progression of infection followed through various methods (2). *In vivo* animal models are able to mimic the physiological and immunological environment of a human organism, however bacterial pathogens, especially *S. aureus,* is highly adapted to the human host. This is an inherent limitation of the animal models, along with being complex, difficult to perform, expensive and with ethical considerations (3,4). Therefore, studies have utilized cell culture models to mimic the complex environment of an infection, such as epithelial, immune, endothelial, and osteocytes cell lines. They are cheap, easy to expand, and easy to handle, allowing convenient implementation and enabling the high-throughput screening of drugs and virulence factors (4–8). *S. aureus* has traditionally been regarded as an extracellular pathogen, because it was mostly observed extracellularly *in vivo* and due to its secretion of various extracellular toxins that are cytolytic to many host cells (9,10). However, it has been known for more than two decades that *S. aureus* is internalized and survives inside non-phagocytic cells such as alveolar epithelial cells (11). Pulmonary epithelium is the front line of lung and airway defense mechanisms against pathogens. It is therefore an important question how *S. aureus* is able to challenge those mechanisms and successfully colonize the lung tissue (12). Decades-long research on of *S. aureus* virulence factors such as adhesins, proteases, hemolysins and regulatory proteins has enabled the understanding of their role and mechanisms of action, however their implication and impact on pulmonary infections is not completely clear. (13). Adaptation of *S. aureus* to the host´s intracellular environment involves extensive modulation of metabolism, transcription and protein synthesis, which can be analyzed by OMICS approaches. However, using these approaches, especially proteomics, to study the pathogens response to the host has proven challenging, due to limited protein amounts and

detection interference from the more abundant host proteins. Therefore, presumably, the majority of the host-pathogen proteomics studies have been focused on the host response (14). Nevertheless, several studies have characterized the proteome response of *S. aureus* to infection-like conditions using *in vitro* or *in vivo* models (8,15–18). In one study, the temporal changes in metabolic proteins in both the host and the pathogen were studied using an infection model with human bronchial epithelial cells and internalized *S. aureus* (18). The study highlighted the metabolic adaptation of the host and bacteria to the infection conditions, especially in proteins involved in bacterial electron transport, pyruvate fermentation, acetate catabolism and amino acid metabolism. By studying specifically post internalization proteome dynamics, another study observed the activation of alternative sigma factor SigB upon internalization, subsequently showing the protein is required for intracellular growth (8). *In vivo* experiments are a major challenge as there is a low number of bacteria and thus a limited amount of proteins. However, in a study employing a murine model, the researchers identified more than 500 proteins, and observed that proteins required for oxidative stress, amino acid biosynthesis, fermentation and ribosomal proteins were found to be differently regulated, revealing a similar response to that observed in *in vitro* studies (17).

The aim of this study was to capture the early adaptation response of *S. aureus* and the proteome modulation upon attachment and internalization. We used a community-associated *S. aureus* USA300 strain to infect alveolar epithelial cells (A549) and the proteome response of *S. aureus* was analyzed over a time series capturing early response, attachment and internalization. The results highlight the known adaptation mechanism of *S. aureus* and reveal potential new proteins involved in virulence.

3.2 Materials and Methods

3.2.1 Bacterial strains and culture

The strain used in this study was *S. aureus* USA300 FPR3757 and the derivative of that strain expressing green fluorescent protein (GFP). Both strains exhibited similar growth. Bacteria were grown overnight in Tryptic Soy Broth (TSB, Merck, Germany) at 37°C with shaking at 200 rpm. The next day, bacterial culture was diluted approximately 150 times to OD_{600} of 0.02 in Dulbecco's Modified Eagle Medium/Nutrient Mixture F-12 (DMEM/F-12, Gibco™, USA) and grown at same conditions as described until the OD_{600} reached 0.5 which corresponds to mid-exponential phase. To determine the number of colony forming units

(CFU), serial dilutions of the bacterial suspensions were made in 96-well plates and 10 µl spots were plated on Tryptic soy agar (TSA) plates and incubated overnight at 37°C.

3.2.2 Eukaryotic Cell culture

Lung epithelial cell line A549 was cultivated in DMEM/F-12 at 37 °C in saturated air humidity with 5% CO_2. The cell line was initially propagated in a T-75 flask where the media was changed every two days until approximately a 70% confluency was obtained. The cells were then washed with PBS, detached using trypsin and were seeded in 6-well plates at $1x10^5$ cells/well and grown for 2 days to achieve full confluency. The cells were used between passages 84 and 86. Two wells containing the confluent monolayer were counted using Cell counter (Thermo Fischer).

3.2.3 Infection model setup

Prior to infection, the culture medium was removed from the A549 monolayer, and the cells were washed with warm PBS. The bacterial culture grown to OD_{600} of 0.5 was diluted 3 times in DMEM/F-12 media to achieve the CFU count of 1 $x10^8$ cells/ml. A volume of 1 mL of the diluted bacterial suspension was added to each well of a 6-well plate yielding a multiplicity of infection (MOI) of 100. The co-culture was incubated at the same conditions used for propagation of the A549 cell line for 30 min, 1h, 2 h and 4 h.

3.2.4 Enumeration of bacterial cells

At each time point, the number of colony forming units (CFU) of unattached bacteria, attached and internalized, and only internalized bacteria was estimated using drop technique. In brief, 10 µl of the bacterial suspension was added to the 96 well plate containing 90 µl of PBS and six serial dilutions (10-fold) were prepared. The 3rd, 4th, 5th and 6th serial dilutions were plated by depositing 10 µl of the dilution on an TSA agar plate and incubating it overnight at 37 °C. To estimate the number of unattached bacteria, the media of the co-culture after the infection was transferred to a clean tube and 10 µl was used for plating. Then the wells were washed with PBS two times to remove any loosely bound bacteria and 1 ml of 0.1% Triton-X was added to the well and incubated for 10 min at room temperature. A volume of 10 µl was again used for plating to estimate the number of internalized and attached bacteria. To only count the internalized bacteria, the medium was removed from the wells after infection, washed two times with 1 ml of PBS and 10 µl of lysostaphin (1 mg/ml) was added to each well and incubated at 37 °C for 30 min to kill any attached bacteria. After the incubation, the wells were washed again two times with PBS and 1 ml of 0.1% Triton-X was added and incubated for 10

min at room temperature. The suspension was plated to estimate the number of CFUs of only internalized bacteria.

3.2.5 Confocal microscopy

To image the infection model, the cell line was grown on coverslips inserted in 6-well plates. The infection setup was prepared exactly as described before, however the amount of bacteria was scaled to achieve the same MOI since the areas of the coverslip and the well differ. After infection, the coverslips were carefully transferred to a clean 6-well plate and incubated with 4% Paraformaldehyde (PFA) for 20 min to fix the cells. The coverslips were then washed two times with 400 µl of PBS. The cells were permeabilized with 200 µl of 0.1% Triton-X for 10 min at room temperature and washed twice with PBS. To each coverslip, a volume of 200 µl of 1 x Actin stain Alexa Fluor 568 Phalloidin plus (Invitrogen) solution and two drops of NucBlue solutions (Hoechst 33342, Invitrogen) was added and the coverslips were incubated in the dark for 30 min. Images were captured at green emission (emission maximum: 509 nm, excitation wavelength: 488 nm), blue emission (emission maximum: 460 nm, excitation wavelength: 360 nm), and red emission (emission maximum: 600 nm, excitation wavelength: 578 nm) using a Leica DM4000 microscope with 10X magnification.

3.2.6 Sample preparation for proteome analysis

The unbound bacteria from the infection culture were washed away with ice cold PBS two times. Then a volume of 500 µl of ice cold PBS was added to each well and the cell monolayer with internalized and attached bacteria was scraped using cell scrapers and transferred to a clean tube. The content of 3 wells was pooled to one tube. The tube was centrifuged at 5000 x g for 8 min at 4 ºC and supernatant was discarded. A volume of 100 µl of 0.1% Triton-X was added to the cell pellet and incubated for 5 min at room temperature to partially deplete human cells. A volume of 900 µl of cold PBS was added and the cells were washed two times with PBS and reconstituted in 100 µl of Tris-HCl buffer (pH 8.5) where 10 ul of lysostaphin (1mg/ml) was added and incubated for 10 min at 37 °C. The samples were immediately put on ice and to each sample, SDS was added to a final concentration of 2%. The tubes were then vortexed for 30 s and subjected to sonication for 30 min (60 s on, 30 s of) using a Bioruptor (Diagenode, USA) at high power. The cell debris was removed by centrifugation at 10 000 g for 15 min. The protein amount was measured by Bicinchoninic acid (BCA) assay. The protein sample (50 µg of total protein) was reduced by incubation with 25 mM tris(2-carboxyethyl)phosphine (TCEP) for 15 min at room temperature. The removal of SDS,

alkylation and digestion were performed using Filter-aided Sample Preparation (FASP) as described before (19). All centrifugation steps were conducted for 20-30 min at 10 000 g at room temperature. In brief, The Microcon-30kDa centrifugal filter unit (Merck, Germany) was washed with 200 µl of 8M urea and the sample was diluted 1:4 with 8M urea and added to the filter unit. The unit was centrifuged and washed trice with 200 µl of 8M urea. The flow through was discarded after each centrifugation step. Then, 200 µl of 50 mM triethylammonium bicarbonate (TEAB) was added to the filter and centrifuged. The alkylation buffer containing 40 mM chloroacetamide (CAA) was added and incubated for 20 min with shaking at 100 rpm. The filter was then centrifuged and washed once with 200 µl of TEAB. The protein digestion was performed by adding 100 µl of TEAB and 1 µl of trypsin/LysC (MS grade, Promega) to the filter unit. The filter tubes were incubated overnight at 37 ºC and the next day, peptides were collected by centrifugation. The filter was washed with 50 µl of TEAB and the filter was discarded. The digestion was quenched the following day by the addition of TFA to a final concentration of 1%. The peptides were desalted using SOLAµ cartridges as follows. The columns were primed with 100% methanol, equilibrated with 80% acetonitrile/0.1% TFA and washed two times with 3% acetonitrile/0.1% formic acid. The peptides were loaded and washed twice with 3% acetonitrile/0.1% formic acid. Finally, peptides were eluted with 60% acetonitrile/0.1% formic acid, dried and subsequently frozen at −20˚C until analysis by MS.

3.2.7 Mass spectrometry analysis

Peptides were reconstituted in 0.1% formic acid (FA) and loaded onto a 2 cm C18 trap column (Thermo Fisher 164946), connected in-line to a 15 cm C18 reverse-phase analytical column (Thermo EasySpray ES904) using 100% Buffer A (0.1% FA in water) at 750 bar, using the Thermo EasyLC 1200 HPLC system, and the column oven operating at 30 °C. Peptides were eluted over a 140 min gradient ranging from 6 to 60% of 80% acetonitrile, 0.1% FA at 250 nl/min, and the Exploris 480 instrument (Thermo Fisher Scientific) with high-field asymmetric-waveform ion-mobility spectrometry (FAIMS) Pro Interface (Thermo Fisher Scientific) switched between compensation voltage of -50 V and -70 V with cycles of 38 and 28 scans. Full MS spectra were collected at a resolution of 60,000, with an AGC target of 300%, maximum injection time set to auto, and a scan range of 375–1500 m/z. MS1 precursors with an intensity of $>5\times10^3$ and charge state of 2-6 were selected for MS2 analysis. Dynamic exclusion was set to 60 s, the exclusion list was shared between CV values and Advanced Peak Determination was set to ‘on’. Precursors selected for MS2 were isolated in the quadrupole with a 1.6 m/z window. Ions were collected with maximum injection time set to auto and

normalized AGC target set to 75%. Fragmentation was performed using higher-energy collisional dissociation (HCD) with a normalized collision energy of 28%, and MS2 spectra were acquired at a resolution of 15,000.

3.2.8 Proteome Data processing

The raw files were analyzed using Proteome Discoverer 2.5 (Thermo Fisher Scientific). The spectra were matched against the Uniprot database of *S. aureus* strain USA300 (UP000001939) proteome and Label-free quantitation (LFQ) was enabled in the processing. Search was performed with setting trypsin/P digestion and allowing a maximum of two missed cleavages. Dynamic modifications were set as Oxidation (M) and Acetyl on protein N-termini. Cysteine carbamidomethyl was set as a static modification. The results were filtered to a 1% false discovery rate (FDR). The abundance levels of all peptides and proteins were normalized using the total peptide amount normalization mode in Proteome Discoverer. The protein table was further processed in R studio (V 2021.09.1) using packages diplyr, naniar, ggplot2 and corrplot. The processing included log2 transformation, removing missing values, Pearson correlation, Principal component analysis (PCA) and visualization.

3.2.9 Differential expression analysis

Differential expression analysis was performed using the limma package (V 3.48.3) (20). A design matrix was constructed to model the experimental conditions. Specifically, contrasts were set for various time points relative to a control condition and between different time points: T30m-ctr, T1h-ctr, T2h-ctr, T4h-ctr, T1h-T30m, T2h-T30m, T4h-T30m, T2h-T1h, T4h-T1h and T4h-T2h. Linear models were fitted to the data using the lmFit() function from the limma package, incorporating the design matrix. The fitted linear model was used to compute the specified contrasts via the contrasts.fit() function. Empirical Bayes moderation of standard errors and log-fold changes was performed using the eBayes() function. A function "decideTests" was used to identify which genes were significantly differentially expressed for each contrast from a fit object containing p-values and test statistics. The p-values were adjusted with Benjamini-Hochberg (BH) procedure. Proteins, which displayed a statistically significant abundance change (adjusted p-value ≤ 0.05) and $|\log2FC| > 1$ were termed differentially expressed proteins.

3.2.10 Fuzzy C-Means Clustering

To identify the co-expression clusters over the course of the infection, differentially expressed proteins in at least one time point were clustered using the Mfuzz soft clustering

method (21). Standardization was carried out using the 'standardise' function. The parameter fuzzifier m, which prevents clustering of random data was determined using the 'mestimate' function. The number of clusters c was determined by manual inspection of the results of the mfuzz function and Dmin function and finally the parameters were set to m=2.14, c=4.

3.2.11 Functional Annotation Analysis

Gene Ontology (GO) enrichment analysis of the differentially expressed proteins was performed by the ShinyGO 0.80 online tool (22) (http://bioinformatics.sdstate.edu/go/) for each of the four clusters determined by fuzzy clustering. Biological process enrichment was determined using String Database *Staphylococcus aureus* subsp. aureus NCTC 8325. For this purpose, the USA300 strain proteins were mapped to NCTC 8325 proteins using standalone blastp+ (23). GO terms with a corrected FDR value of less than 0.05 were considered significantly enriched.

3.2.12 Validation using live cell imaging

S. aureus mutants were obtained from The Nebraska Transposon Mutant Library consisting of sequence-defined Tn insertions in the approximately 2,000 non-essential genes of S. aureus USA300 strain (24). The selected mutant strains along with the parent strain JE2 and the USA300 strain were grown in 200 µl of TSB medium in a 96 well plate overnight (37 ºC, 200 rpm). The next day the cultures were diluted to an OD_{600} of approximately 0.01. The OD_{600} measurements were conducted in a Synergy h1 plate reader (Biotek). The A549 cells were propagated as described before in this study and seeded in 96 well plates. The cells were grown for 3 days until full confluence (10^4 cells/well) and then they were infected with MOI 100 and incubated for 2 hours. The mock well contained only medium. After the incubation period, 40 µl of propidium iodide (PI, Invitrogen) was added to each well and incubated for 20 min. Whole wells were then imaged on ImageXpress pico automated cell imaging system (Molecular Devices) at 4x magnification using PI channels. Fluorescent cell counting analysis was employed using the CellReportXpress software to determine total number of death cells (PI positive). Before infection, the label free cell counting was applied to A549 cells to ensure that the cell numbers were consistent between wells.

3.3 Results and Discussion

3.3.1 *S. aureus* attaches to and replicates within A549 cells.

The A549 monolayer was "infected" with *S. aureus* at a MOI of 100 and the replication and survival of unattached, internalized and attached bacteria was estimated. The high MOI

was chosen to acquire sufficient bacterial biomass and protein for the proteomics experiment. The A549 alveolar epithelial cell line was chosen as an infection model in order to mimic human lung infection. The epithelial cell layer in the human lung forms an important primary barrier against infection (25). While the A549 cells do not fully replicate the complexity of the living organism, specifically immune system, *in vitro* models like A549 cells offer controlled experimental conditions and genetic uniformity, leading to consistent and reproducible results compared to the complex *in vivo* models (26,27). Another disadvantage of *in vivo* models for *S. aureus* infection lies in specificity of the bacterium towards its host. In other words, S. *aureus* employs virulence mechanisms which target some host species more efficiently or exclusively (34). One such example is the production of toxin LukAB, which is found to specifically target human cells and does not affect murine cells (35).

The infection model was monitored by light microscopy, and after 4 hours the cells were progressively dying which was observed as detachment of cells. Based on the results, sampling time points were set to 30 min, 1h, 2h and 4h. It was hypothesized that time point 30 min can capture early bacterial response to the host environment, including attachment to the host cells. The latter three time points (1h, 2h and 4h) can capture invasion, metabolic adaptation, expression of virulence factors and lysis of the host cells. A high number of unattached bacteria was observed at all time points, probably replicating in the media. The attached and internalized bacteria, however, were increasing over time, and reaching a maximum in the 4-hour sampling period (Figure 3.1). Additionally, the co-culture was monitored using confocal microscopy (Figure S3.1) where the attachment and internalization of the *S. aureus* was observed increasing over time. Moreover, it was observed that bacteria were not uniformly distributed in the human cells, but rather formed clusters.

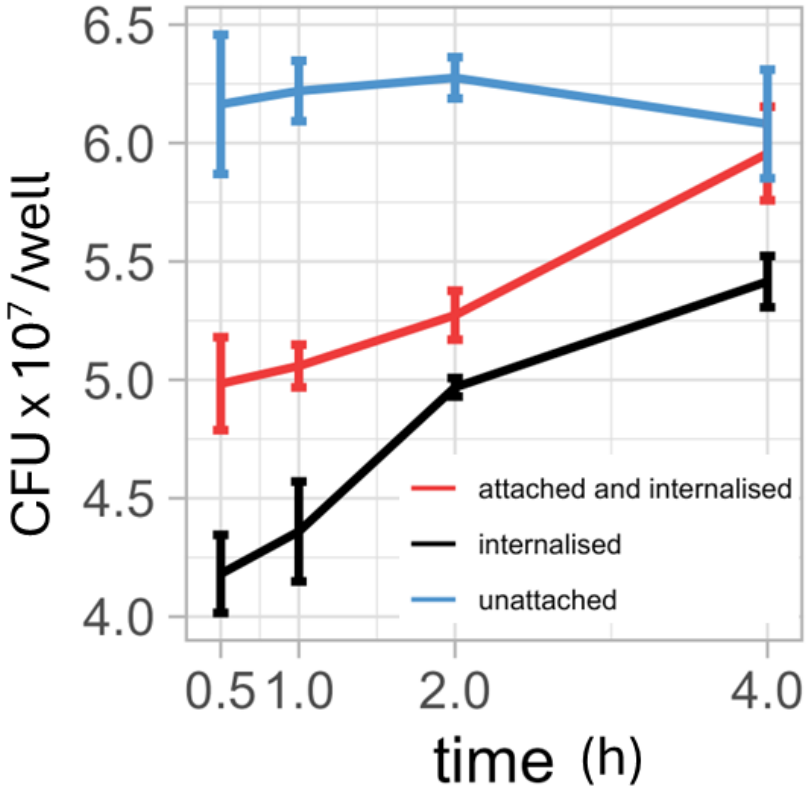


Figure 3.1: A549 cells (1×10^6) were infected by S. aureus cells (1×10^8) and incubated until sampling at selected time points 30 min (0.5 h), 1h, 2h and 4h. At each time point, unattached, attached and internalized bacteria were plated from the media in the plate well. The plating of the internalized bacteria was achieved by killing all extracellularly present bacteria using lysostaphin. The results show the colony forming units (CFU) from one well in a 6-well plate measured in triplicates.

3.3.2 Quantitative proteomics results show high correlation between replicates and distinct response in infection conditions

After incubation of *S. aureus* with A549 cells, samples taken at selected time points were prepared for proteomics analysis using a pipeline summarized in Figure 3.2A. Only bacteria which were attached and internalized in the host cells were sampled. This was achieved by removing the supernatant and washing the cells with ice cold PBS, to remove the unattached cells. To maximize the coverage of bacterial proteome, mammalian cells were partially depleted using low concentration of Triton-X which partially lysed human cells, but due to the more resistant cell wall did not substantially lyse bacteria (confirmed by plating). Protein extraction from Gram-positive bacteria such as *S. aureus* was also challenging, especially from

a low amount of biomass as in this experiment. To efficiently lyse bacteria and extract proteins, the enzyme lysostaphin was used along with SDS which resulted in sufficient protein amount (30 µg) after pooling material from three wells of a 6-well plate. The samples were analyzed using data dependent acquisition (DDA) in combination with label free quantification (LFQ). The proteome analysis resulted in identification and quantification of 1596 *S. aureus* proteins in all sample groups (Figure 3.2E), achieving a coverage of more than 60% of the theoretical proteome in the USA300 strain. To our knowledge, this study has achieved the highest proteome coverage for an *S. aureus* strain in host-pathogen research, quantifying substantially more proteins than in similar studies (21,23,24,36) with as little as 5% of missing values across all samples. After removing proteins with missing values, the coverage of the identified proteome was 61.3 %. Moreover, proteins were quantified with less than 25% coefficient of variation (CV) for biological replicates (Figure 3.2C). Pearson correlation analysis coefficients were above 0.9 for all conditions, and particularly high for biological replicates (>0.95), indicating a good reproducibility (Figure 3.2D). Principal component analysis (PCA) of the normalized protein expression data set showed that the three replicates of each experimental condition were well-grouped, moreover, the PCA plot showed different proteome response among conditions. The difference is especially noticeable in control (ctr) and infection time points (T30m, T1h, T2h, T4h). (Figure 3.2B). These results demonstrate a good performance of LFQ DDA analysis, even in complex samples such as host-pathogen systems. Besides improvements in chromatography and software tools, FAIMS technology, which enables the rapid and effective gas-phase separation of peptide ions before they enter the mass spectrometer, has previously been shown to improve protein coverage in DDA analysis (37,38).

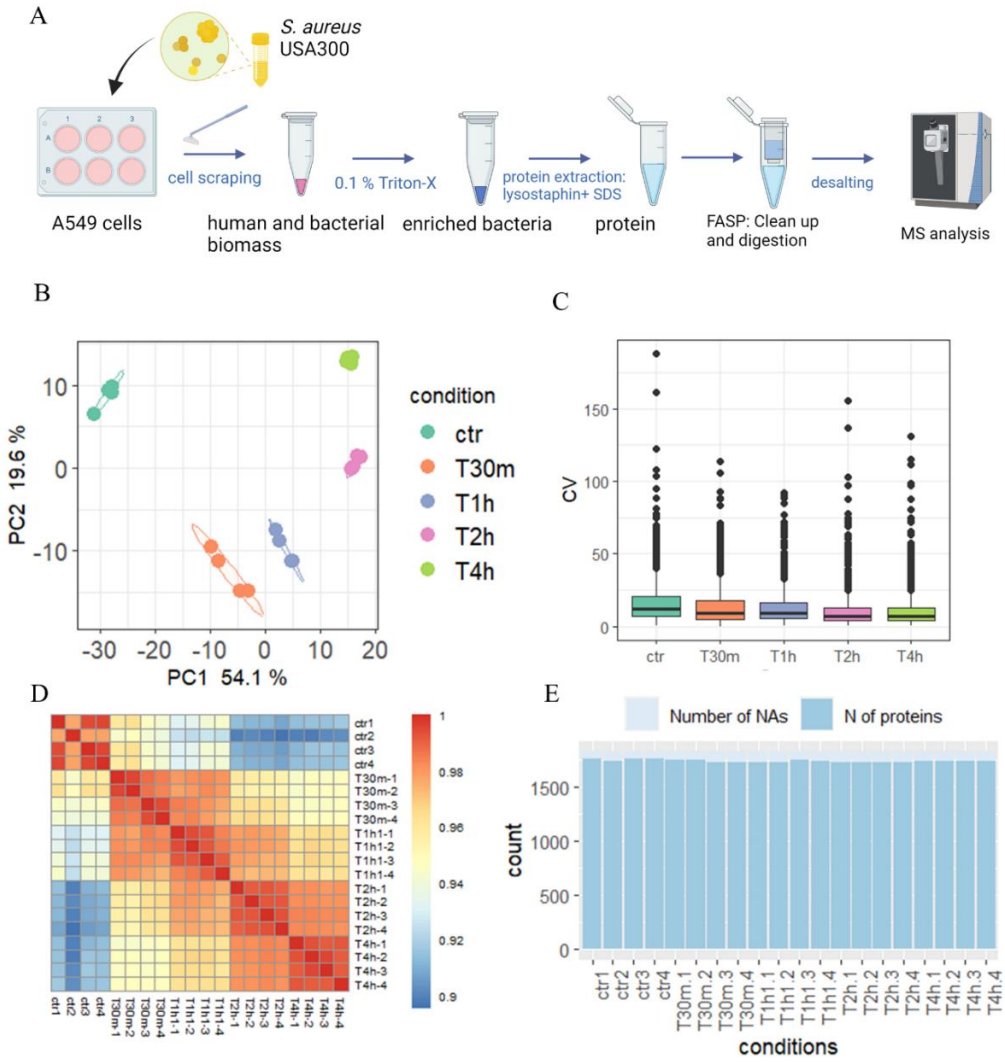


Figure 3.2: Global results of the proteomic analysis of *S. aureus* during co-cultivation with lung epithelial cell line A549. A) Summarized pipeline of the sample preparation for the proteomics analysis. B) Principal component analysis (PCA) on the log2 transformed protein intensities shows group separations along the first and second dimension (PC1 and PC2). C) Boxplot showing coefficient of variation (CV) calculated as the ratio of the standard deviation to the mean, expressed as a percentage, for the protein expression levels across different time points in the experiment. D) Pearson correlations between samples depicted as a heatmap. E) Number (N) of proteins quantified in each sample along with the count of missing values (denoted as NA).

3.3.3 Differential expression analysis reveals 513 proteins with significant abundance changes, including time-point specific and shared proteins

Differential expression analysis was performed on all possible contrasts (T30m-ctr, T1h-ctr, T2h-ctr, T4h-ctr, T1h-T30m, T2h-T30m, T4h-T30m, T2h-T1h, T4h-T1h and T4h-T2h). The Supplementary_dataset_3.2.xlsx contains the results of differential expression analysis for all 1596 quantified proteins along with annotation obtained from Uniprot and AureoWiki (28). To assign a protein differentially regulated, absolute log2 fold change (log2FC) had to be higher than 1 and adjusted p-value had to be lower than 0.05. Based on these criteria 513 proteins were differentially expressed (Table S3.1A). When considering only the adjusted p-value without the log2FC, approximately one-third of the proteome remained unchanged (Table S3.1B). However, further analysis focused on results with a two-fold change, as this is generally considered a meaningful threshold in biological systems (29). Volcano plots were used to visualize differentially expressed proteins between infection time points and control group (Figure 3.3A). The global protein expression in each infection time point was compared to the control group (bacteria grown in medium) and it was observed that the change in protein expression increased with latter infection time points. Nevertheless, even in the most affected time point T4h, only 20% of the proteome was modulated compared to control. This is in accordance with previous studies which observed, that *S. aureus* only changes a subset of its proteome to adapt to infection-like conditions (1,18).

Several proteins were highly upregulated with > 4-fold increases in their abundance in all time points: Aspartokinase, Cell division protein FtsL, Putative staphylocoagulase (SAUSA300_0773), Staphylococcal accessory regulator (SAUSA300_2247), Uncharacterized protein (SAUSA300_1566), Esterase-like protein (SAUSA300_2296) and Respiratory nitrate reductase (NarJ) (Supplementary_dataset_3.2.xlsx). Aspartokinase is involved in the biosynthesis of aspartate family amino acids (methionine, lysine, and threonine) and previously found to be overexpressed in an *S. aureus* CodY mutant, where CodY is recognized as a global regulator of virulence (30). FtsL is crucial for late phase of cell division and it is conserved in all bacteria (31). Putative staphylocoagulase is identified as von Willebrand factor-binding protein (Vwb) through homology with other *S. aureus* strains. The Vwb protein was shown to attach to endothelial cells and platelets by binding to the von Willebrand factor (vWF) present in the host's blood (32,33). Staphylococcal accessory regulator was identified as SarY protein, a member of Sar family of transcription factors related to virulence (34).

Proteins identified as differentially expressed were then compared across different conditions to determine the extent of overlap, i.e., those consistently differentially expressed at all infection time points compared to the control. As it can be seen on the Venn diagrams in Figure 3.3B, there are 72 and 21 proteins which are consistently upregulated and downregulated in all infection time points, respectively. GO analysis identified several enriched processes in the common upregulated proteins including oligopeptide transport, nitrate metabolic process and positive regulation of protein phosphorylation (Figure S3.2). Common downregulated proteins were enriched in ion transport, more specifically heme transport (Figure S3.2).

Additionally, several proteins were uniquely differentially expressed at each time point, meaning only differentially expressed at one infection time point (Table S3.2, Table S3.3). In the first time point T30m, where all bacteria were sampled, including unattached bacteria, there were 8 unique differentially expressed proteins, including a fibronectin-binding protein A (FnbA), which was downregulated in time points T2h and T4h. At the T1h time point, only three proteins were uniquely upregulated, including a pyruvate formate-lyase-activating enzyme (PflA), an iron-dependent protein that activates pyruvate formate-lyase (PFL), which in turn catalyzes the conversion of pyruvate to formate and acetyl-CoA under anaerobic conditions. The increased expression of PflA would indicate a metabolic shift from aerobic respiration to anaerobic fermentation pathways due to oxygen-limited environment caused by internalization in the host cells (35).

In T2h, the unique differentially expressed proteins are the ones known to cause lysis in host cells: Phenol-soluble modulin alpha 1 peptide (PsmA1), Phenol-soluble modulin alpha 4 peptide (PsmA4), and Delta-hemolysin. These proteins have increased expression in other time points compared to control, but their expression peaked at T2h, suggesting an increased host cell lysis activation in this time point. Other unique proteins in this time point include a Sensor protein kinase WalK, part of the essential TCS WalKR which regulates cell metabolism and triggers expression of several virulence factors (36), Single-stranded DNA-binding protein (Ssb) involved in DNA repair and several uncharacterized proteins. There were 43 proteins uniquely upregulated in T4h including numerous uncharacterized proteins. Notably, the putative pyridoxal phosphate-dependent acyltransferase (HisC) was highly upregulated at this time point, while it was downregulated compared to the control in all prior time points. There were 8 uniquely downregulated proteins (log2FC<-1) in T30m, out of which three belong to membrane proteins involved in transport (SdcS), phosphotransferase system

(SAUSA300_2576) and uncharacterized function (SAUSA300_1910). In the T1h time point, 6 proteins were uniquely differentially expressed including Sugar efflux transporter (SAUSA300_0659) , Heme sensor protein (HssS) and bifunctional autolysin (Atl). Among the 32 uniquely downregulated proteins in T2h, noteworthy examples include YoeB, a bacterial toxin associated with biofilm formation (37), and RsbV, a positive regulator of the σ^B factor (38). Additionally, proteins involved in DNA repair, such as Nth-a1 and RecD, were also downregulated. In T4h, Proteins involved in cell wall biosynthesis and peptidoglycan cleavage seemed to be specifically downregulated in this time point. These included Lipoteichoic acid synthase (LtaS, UDP-N-acetylmuramate--L-alanine ligase (MurC), Lysostaphin (LytM) and Probable transglycosylase (IsaA).

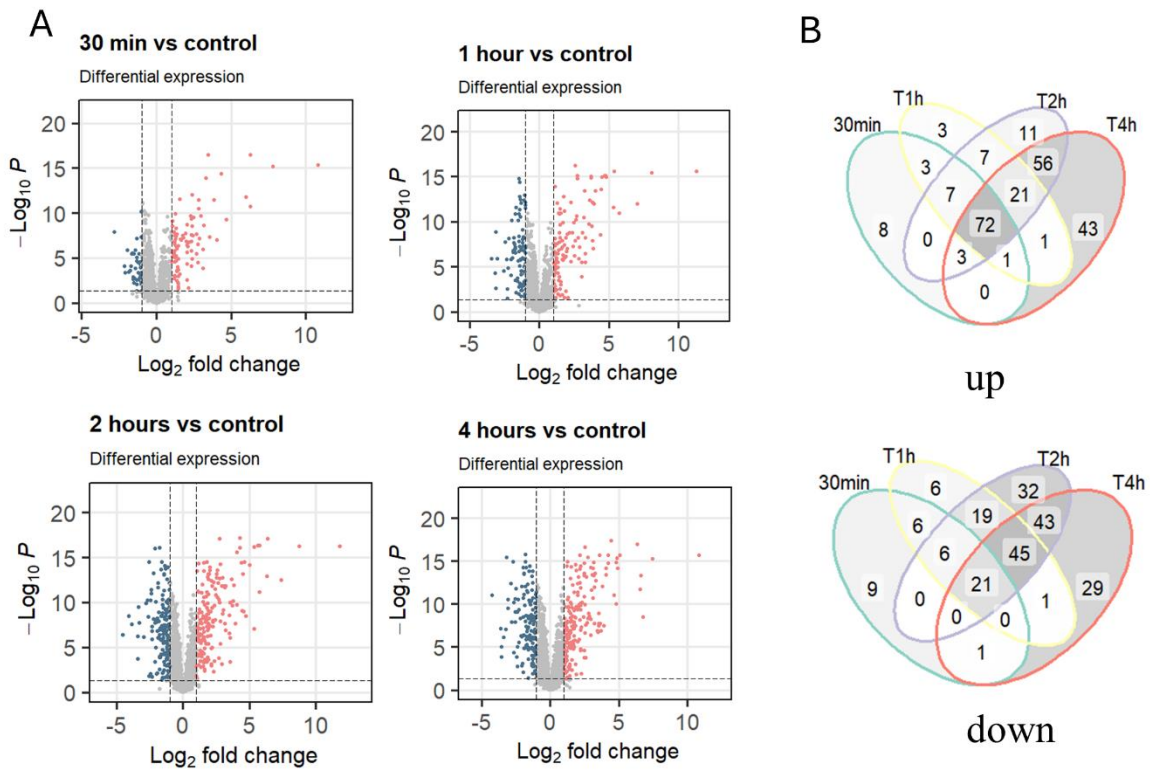


Figure 3.3: Differential expression analysis between infection time points and control sample. A) Volcano plots where red dots represent upregulated and blue dots represent downregulated proteins (|log2FC| >1, adj.p value < 0.05). Proteins statistically not significantly changed are

depicted in grey. B) Venn diagrams showing overlap of upregulated and downregulated proteins between infection time points and control group.

3.3.4 *Staphylococcus aureus* undergoes metabolic shift from aerobic respiration to fermentation and nitrate respiration during co-cultivation with A549 cells

The bacteria encounter oxygen limited conditions when internalized by the host cells which causes the bacteria to shift to anaerobic respiration (39). As mentioned above, the PflA protein which activates a PFL was upregulated at T1h time point, indicating a shift to anaerobic respiration. Accordingly, the expression pattern of enzymes involved in fermentation strongly supports that *S. aureus* undergoes mixed-acid and butanediol fermentation (Figure 3.4B) as seen in Figure 3.4A. Lactate dehydrogenase proteins Ldh1 and Ldh2 were increasing in abundance over time, peaking at T4h. Similarly, a PflB enzyme which catalyzes the non-oxidative decarboxylation of pyruvate to acetyl-CoA and formate was upregulated in all time points. The named reactions are a primary route for regenerating the NAD^+ used in glycolysis during *S. aureus* anaerobic growth (47). Upregulated proteins in infection also included BudA2 which converts pyruvate to acetolactate and ButA which catalyzes the conversion of acetoin to 2,3-butanediol and vice versa. 2,3-butanediol has been shown to contribute to bacterial growth at low pH (48). Interestingly, the enzyme acetolactate synthase (Als), which catalyzes the first step of the 2,3-butanediol fermentation pathway by converting pyruvate to acetolactate, was downregulated at all time points compared to the control. Moreover, a conversion of pyruvate to Acetyl-CoA under aerobic conditions is catalyzed by the PDH complex, which was downregulated in this study (Figure 3.4A), confirming its repression under anaerobic conditions (35). Abundances of enzymes involved in the conversion of acetyl-CoA to acetate, such as AckA (acetate kinase) and Pta (phosphotransacetylase) were not significantly different in infection time points compared to the control which coincides with growth on glucose where lactate is the preferred fermentation product, and only small amounts of acetate are produced (40). However, the proteins with highest expression increase in fermentation pathway is alcohol dehydrogenase AdhE which converts acetyl-CoA to ethanol via acetaldehyde. The AdhE is 64-fold higher in T4h than in the control group (log2FC>6). Previously, it was shown that excess acetyl-CoA is converted to ethanol by alcohol dehydrogenase. This process regenerates twice as much oxidizing power (NAD^+) compared to lactate production (41).

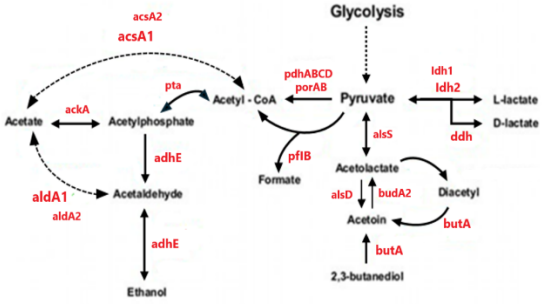
Furthermore, several transporter proteins were among the upregulated proteins (Figure 3.4C), including lactate permease (SAUSA300_2313, LctP2) and a formate/nitrite transporter

family protein (SAUSA300_2349). This suggests a mechanism for transporting metabolites out of the cell to prevent intracellular acidification and regenerate NAD+. In addition, NarT, a protein required for nitrate uptake was upregulated. Previously, it was shown that besides fermentation, *S. aureus* can grow anaerobically by nitrate respiration which allows *S. aureus* to generate ATP through oxidative phosphorylation, even in the absence of oxygen (51). This is supported in this study by upregulation of nitrate respiration genes NarH, NarJ and NirB together with upregulation of ATP synthase B subunit (Figure 3.4C). NarJ and NarH are part of the NarGHJ operon encoding respiratory nitrate reductase which catalyzes the reduction of nitrate (NO_3^-) to nitrite (NO_2^-). NirB is a nitrite reductase that catalyzes the reduction of nitrite to ammonia, using NADH or NADPH as electron donor (Figure 3.4D).

A

Gene Name	T30m	T1h	T2h	T4h
AdhE	-0.01*	4.70	6.39	6.38
Adh	0.68	2.75	4.28	4.42
PflB	0.46	2.59	2.77	2.54
Ldh1	0.36	1.54	1.92	2.07
ButA	-0.53	0.07*	1.44	1.91
Ldh2	-0.29	0.78	1.55	1.84
BudA2	0.66*	0.74	0.96	1.01
Ddh	-0.26	0.80	0.29	0.34
AtpH	-0.08*	-0.05*	-0.04*	-0.03*
Pta	-0.27	-0.12*	-0.10*	-0.03*
AckA	-0.25	-0.01*	-0.18	-0.04*
PdhA	-0.18	-0.28	-0.63	-0.73
PdhB	-0.42*	-0.44	-0.80	-0.75
AlsS	-0.68	-1.19	-1.15	-1.98

B



C

Gene Names	T30m	T1h	T2h	T4h
narT	2.03	3.79	2.82	2.96
SAUSA300_2313 (LctP1)	0.32	1.80	2.11	2.47
SAUSA300_2349	0.20	0.78	1.09	1.14
AtpB	0.53	0.66	0.66	0.51
LctP2	-0.05*	-0.07*	-0.31*	-0.02*
AtpD	-0.18	-0.19	-0.14	-0.22
AtpC	-0.33	-0.41	-0.52	-0.33

D

Gene Name	T30m	T1h	T2h	T4h
narJ	4.68	5.73	5.76	4.82
nirB	2.28	4.83	4.40	4.08
narH	-0.14	0.57	0.28	0.51

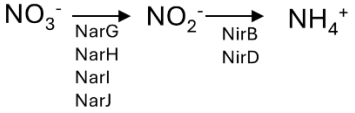


Figure 3.4: Differential expression of proteins involved in anaerobic metabolism in *S. aureus*. The values represent log2 fold changes between indicated time point and control group. The asterisk * indicates a non-significant change in expression (adjusted p value >0.05) A) Enzymes involved in mixed acid and butanediol fermentation B) The proposed reactions of *S. aureus* fermentation adapted from Chaffin et al. (42). C) Enzymes involved in transport of fermentation products E) Enzymes involved in nitrate respiration

3.3.5 Profiling of virulence factors and regulatory proteins implicated in virulence

S. aureus produces a diverse array of virulence factors whose expression increases following infection. The expression of these virulence factors is controlled by regulatory elements such as two-component systems (TCSs) and transcriptional regulatory systems in response to host environment encountered by the bacterium during infection (43). One of the most characterized TCS belongs to the agr quorum sensing system which plays an important role in the expression of various genes, particularly those involved in virulence, biofilm formation, and the adaptation to environmental changes (44). The signal sensed by the agr system is an autoinducing peptide (AIP). The agr locus consists of two adjacent transcripts, called RNAII and RNAIII, whose expressions are driven by the P2 and P3 promoters, respectively (Figure 3.5C). The RNAII transcript is an operon of four genes, agrBDCA, and RNAIII is a regulatory RNA that also contains the open reading frame encoding the delta-hemolysin gene (hld), which is upregulated in this study, peaking at T2h and decreasing in T4h (Figure 3.5A). The identified gene products of the agrBDCA operon in this study (AgrA and AgrC) were increasing in abundance with time, reaching their peak at T4h (Figure 3.5A). AgrC and AgrA form the sensor kinase and response regulator of the TCS, respectively. This TCS senses and transmits the quorum-sensing signal once the concentration of AIP reaches a threshold level (45).

AgrA directly regulates the expression of RNAII and RNAIII from the P2 and P3 promoters, respectively (Figure 3.5C). Moreover, it also regulates the expression of phenol-soluble modulins (PSMs) which were upregulated in this study (PSMa1 and PSMa4). Their abundance peaked at T2h and then decreased slightly in T4h, following a similar pattern as Hld protein. The Hld toxin and the PSMs belong to a class of membrane-damaging peptides, which are proposed to act as a surfactant to disrupt the cell membrane (46–48). Unlike most genes in the agr regulon that are controlled through RNAIII, the transcription of the PSM loci is directly

enhanced by the binding of the AgrA response regulator protein (49), therefore both Hld and PSMs are directly activated by AgrA which explains the similar expression profile.

Among the genes previously shown to be upregulated by RNAIII are a well-characterized α-toxin (Hla), serine proteases and lipase Geh (44,50). However, neither Hla nor serine proteases were identified in this dataset. The Hla pore forming cytotoxin can lyse multiple cell types, most notably erythrocytes, epithelial and endothelial cells (51,52) and its attenuation significantly reduced virulence in animal models (53). The lack of its identification in this study might be explained by the fact that only the cellular fraction was analyzed, and this is a known extracellular protein. Consequently, it would be relevant to also study the secreted proteins in future work to capture all exotoxins. A study by Surmann et al. applied global and targeted proteomics to identify *S. aureus* proteins secreted inside epithelial cells. However, a more sensitive targeted approach was necessary to identify Hla and still it was only identified in one of the tested time-points, indicating relatively low abundance (16).

On the other hand, proteins downregulated by RNAIII include Protein A (Spa), the immunodominant staphylococcal antigen A (IsaA) and Sodium/Proton Antiporter System Mnh (50). While MnhE, MnhD, MnhG and IsA were downregulated, Spa did not have a significantly changed expression between control and infection time points. The protein MnhA1 was upregulated in this study. This protein is suspected to contribute to pH and salt tolerance and deletion of the *mnhA1* gene led to a major loss of *S. aureus* virulence in mice, (54).

Agr activity is also regulated by many previously described regulatory proteins, several of which were identified in this study (Figure 3.5B). The expression levels of SarR, SarT, SrrAB, which are generally considered repressors of the agr system were upregulated in the infection time points(55,56). Other two identified negative regulators CodY (30) and ArlS (57) were downregulated in all time points. Positive regulators such as CcpA, SarA, and SarZ showed no significant differences in abundance compared to the control, while MgrA was upregulated at all time points, though the difference did not exceed two-fold.

Another well-characterized TCS in *S. aureus* is the SaeRS system which consists of the histidine kinase SaeS and the SaeR response regulator (58). SaeS is not significantly changed in infection time points compared to control and SaeR is only slightly upregulated (log2FC<1) (Figure 3.6A). Although not highly upregulated, this TCS regulates expression of several virulence factors including Panton-Valentine leucocidin (LukGH). Both LukG and LukH showed significant upregulation at all infection time points compared to the control, with their

expression levels increasing by more than 30-fold at T4h. Another upregulated protein which was previously shown to be regulated by SaeRS, is the previously mentioned Wvb where the abundance peaked at the first time point T30m, indicating its importance in bacterial adhesion. Other proteins shown to be important for adhesion such as cell wall-anchored proteins immunoglobulin-binding protein Sbi and Fibronectin binding protein B (FnbB) (59,60) were found to be upregulated in this study. FbnA was upregulated in the first time points and downregulated in the latter two time points. Moreover, a secreted chemotaxis inhibitory protein (Chp) which hinders the immune response by preventing neutrophils from migrating to infection sites (61) was highly upregulated in all infection points reaching its abundance maximum at T1h and T2h. Proteins downregulated or not statistically changed included extracellular fibrinogen-binding protein (Efb), and nuclease (Nuc2).

A

Protein	T30m	T1h	T2h	T4h
AgrA	0.01*	1.11	2.98	3.18
AgrC	-0.65	0.1*	1.26	1.77
MnhA1	0.88	1.32	1.84	1.16
Hld	0.28*	0.65	1.91	0.99
PsmA1	0.59	0.88	1.4	0.75
PsmA4	0.58	0.85	1.36	0.73
ScpA	0.53	0.38	0.09*	-0.02*
Spa	-0.25*	0.1*	0.2*	-0.07*
Rot	-0.29*	-0.16*	-0.63	-0.28*
Geh	-0.79	-1	-0.55	-0.3*
SpsB	-0.28	-0.42	-0.25	-0.3
MnhG1	-0.93	-1.02	-1.34	-0.99
MnhD1	-0.88	-1.04	-1.7	-1.06
IsaA	-0.82	-0.21*	-0.8	-1.49
MnhE1	-0.65	-1.05	-2.08	-1.63

B

Protein	T30m	T1h	T2h	T4h
SarR	0.4	0.34	0.87	1.81
SarT	2.06	1.88	2.22	1.67
SrrB	2.32	1.63	2.26	1.33
MgrA	0.7	0.32	0.9	0.78
RsbU	-0.25	0*	0.33	0.6
SrrA	-0.12*	0.25	0.34	0.52
ArlR	-0.13*	0.22*	0.19*	0.39
SarZ	0.05*	-0.16*	-0.1*	0.2*
SarA	0.11*	-0.13*	-0.26*	0.19*
CcpA	-0.11*	-0.07*	-0.1*	-0.08*
CodY	-0.35	-0.34	-0.35	-0.35
ArlS	-0.46	0.04*	-0.33*	-0.37*

C

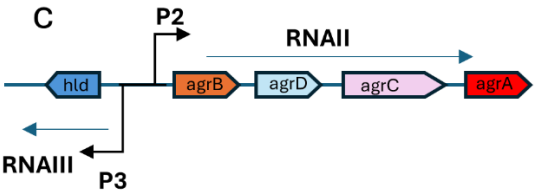


Figure 3.5: Time course of expression changes in virulence factors in *S. aureus* co-cultivated with A549 cells (infection condition) compared to control (medium only). Time course of expression changes for A) the Agr system and its direct targets. B) Regulatory proteins associated with regulation of the Agr system. Values in the boxes show log2 fold-change. The asterisk * indicates a non-significant change in expression (adjusted p-value >0.05) C) Schematic representation of agr operon

The SrrAB (staphylococcal respiratory response) TCS was initially identified as a global regulator of virulence factor production in *S. aureus* under low-oxygen conditions (62). In this study, SrrB protein was found to be upregulated in infection time points compared to the control, while SrrA was not substantially upregulated in infection time points (log2FC<0.6). However, the slight upregulation was statistically significant (adj.p.val < 0.05) at T1h, T2h, and T4h (Figure 3.6B). In contrast, several virulence factors known to be regulated by SrrAB were downregulated including alkyl hydroperoxide reductase (AhpC), and DNA protection during starvation protein (Dps) which are important for oxidative stress defense (63) and phosphatidylinositol-specific phospholipase (PLC) that was proposed to be involved in lysis of the host cell membrane (64).

A SaeRS

Protein	T30m	T1h	T2h	T4h
LukG	2.84	4.51	5.64	5.08
LukH	1.82	3.82	5.00	4.91
Vwb	6.31	5.38	5.77	4.85
Chp	2.67	4.42	4.58	3.54
Sbi	0.09*	1.11	2.33	2.63
FnbB	3.46	2.74	1.83	1.65
SaeR	0.15	0.33	0.40	0.70
Efb	-0.41*	-0.20*	0.45*	0.42*
SaeS	-0.04*	-0.09*	-0.26*	-0.06*
FnbA	1.18	0.59	-0.43	-0.32
Nuc2	0.48	0.38*	-0.07*	-0.41

B SrrAB

Protein	T30m	T1h	T2h	T4h
SrrB	2.32	1.63	2.26	1.33
SrrA	-0.12*	0.25	0.34	0.52
SatA	-0.52	-0.84	-0.62	-0.58
AhpC	-0.45	-0.76	-0.88	-0.89
Dps	-0.61	-0.81	-0.42	-0.01*
PLC	-0.47	-0.78	-0.81	-0.79

C ArlS

Protein	T30m	T1h	T2h	T4h
MgrA	0.7	0.32	0.9	0.78
ArlR	-0.13*	0.22*	0.19*	0.39
HtrA1	-0.51	-0.22*	-0.11*	-0.01*
CtpA	-0.49	-0.45	-0.49	-0.1*
ArlS	-0.46	0.04*	-0.33*	-0.37*
Ebh	-0.15*	-0.52*	-0.56*	-0.37*
HtrA2	-0.4	-0.3	-0.5	-0.42

D σ^B

Protein	T30m	T1h	T2h	T4h
RsbU	-0.25	0*	0.33	0.6
RpoD	0.2	-0.01*	-0.21	-0.31
RsbV	-0.22*	-0.69	-1.04	-0.65
RsbW	-0.44	-0.76	-0.99	-0.83
RpoF	-0.28	-0.56	-0.78	-0.97

Figure 3.6: Time course expression change of *S. aureus* TCSs (A-C) and the well-described transcriptional regulator σ^B (D) known to regulate expression of virulence factors. Values in the boxes show log2 fold-change. The asterisk * indicates a non-significant change in expression (adjusted p-value >0.05)

Another TCS described in *S. aureus* is ArlRS, identified as a repressor of autolysis in methicillin-sensitive *S. aureus* (MSSA) strains, but not in the Methicillin-Resistant *S. aureus*

(MRSA) (65). Moreover, ArlRS has been reported to positively regulate extracellular proteolytic activity. The majority of the identified proteins in this study known to be regulated by ArlRS were downregulated or not significantly different than in the control (Figure 3.6C). The only consistently upregulated protein was MgrA, a global regulator shown to modulate clumping and virulence by regulating surface exposed proteins. (66).

An important regulator of virulence is sigma factor B (σ^B) which was previously shown to react to stress conditions and regulate stress-response proteins as well as virulence factors (67). In *S. aureus*, σ^B encoded by the *sigB* gene (*rpoF*) gene is part of an operon which consists of genes *rsbU, rsbV,* and *rsbW* . The expression of all genes in the operon, besides *rsbU* was found to be downregulated in infection time points compared to control (Figure 3.6D)

3.3.6 Temporal proteomics profiling of *S. aureus* cultivated with A549 cells

A time series soft clustering analysis was performed using Mfuzz package and revealed the presence of four distinct clusters (Figure 3.7). Clusters represent a group of proteins that exhibit similar expression patterns over time, so they are potentially co-expressed. Each cluster presented enriched biological processes as identified by GO analysis. Cluster 1 contained 134 proteins which showed gradual increase in expression, peaking at T4h. Proteins in this cluster were enriched in branched-chain amino acid (BCAAs) biosynthetic process which included biosynthesis of valine, threonine and isoleucine. In *S. aureus*, the BCAA biosynthetic genes are encoded by the ilv-leu operon (68). The gene products of this operon, identified in cluster 1 of this study, include IlvA, IlvB, IlvC, IlvD, IlvE, LeuA, Asd and SAUSA300_1226. Moreover, cluster 1 is also enriched in aspartate family amino acid process, indicating an increase in synthesis of threonine and lysine, since the proteins identified in these cluster also include threonine synthase (ThrC), homoserine kinase (HrB), dihydrodipicolinate synthase (DapA) and dihydrodipicolinate reductase (DapB). The synthesis of BCAAs is regulated by *S. aureus* CodY that binds BCAAs which leads to repression of the majority of its targets including BCAA biosynthetic genes (69). However, the control of virulence exerted by CodY in *S. aureus* is indirect via agr quorum sensing system and SaeRS. Recently it was observed that CodY regulates expression of virulence genes through regulation of SaeRS which was correlated with the increase of branched chain fatty acids (BCFAs) derived from BCAAs. Similarly, lower levels of BCFAs were associated with decreased activity of SaeRS. Therefore, it was hypothesized that CodY might upregulate SaeRS activity via BCFA synthesis, with BCAA biosynthetic pathways likely involved (70)

Cluster 2 contained 118 proteins enriched in signal transduction by protein phosphorylation and nitrate metabolic processes. The expression profile of this cluster has an upregulation pattern, however unlike cluster 1, there is an immediate increase in protein expression in the first time point T30m, from which there is no substantial change until the last time point T4h, where there is a slight drop in the protein levels. Proteins enriched in nitrate assimilation included previously mentioned NarJ, NirB and NarT, while proteins enriched in signal transduction by phosphorylation expectedly included components of the TCSs: Sensor protein kinases WalK, SrrB, GraS and putative PEP-CTERM system histidine kinase (SAUSA300_1799). These TCSs have been implicated in antibiotic resistance (WalkKR, GraSR) and oxidative stress response (SrrAB) (71).

Clusters 3 and 4 are both characterized by downregulated proteins, where the expression decreased in the first three infection time points as compared to control. In Cluster 3, however, the protein level increased in the last time point T4h whereas for cluster 4 the protein level continued to decrease. A total of 71 proteins were associated with cluster 3, while cluster 4 contained 117 proteins. Biological processes enriched for these two clusters were connected to iron metabolism, more specifically iron-sulfur cluster assembly, siderophore metabolic process and iron ion transport. Iron acquisition is essential for *S. aureus* survival within the host because the host restricts iron availability as a defense mechanism against invading pathogens (72). Therefore *S. aureus* has many iron acquisition proteins which are usually upregulated in the infection-like conditions (73). However in a transcriptomics study of infected mouse livers it was shown that iron acquisition genes were downregulated even 6.5 hours after infection and then their expression was increased at 24 hours (74), indicating a role in later stages of infection.

In total 21 virulence proteins proteins were matched to clusters 1-4 as shown in Figure 3.8. The highest number of virulence factors was matched to cluster 2. These proteins are positively regulated by SaeRS (LukG, LukH, Vwb, Chp) and Agr (PsmA1, PsmA4, MnhA1). In addition, two regulatory proteins upregulating Agr were also matched to cluster 2 (SarT, SrrB). Cluster 2 also contained 20 uncharacterized proteins, with SAUSA300_1566 exhibiting a more than 64-fold abundance difference compared to control. The possible role of these proteins in infection is yet to be elucidated. Cluster 1 contained AgrA and AgrC proteins as well as regulatory protein SarA and SaeRS target Sbi. Only three virulence factors were matched to cluster 3 and they included FnbA, IsaA and SphX, which are shown to be regulated

by SaeRS, AgrAC and SrrrAB, respectively. Surface receptors MnhA, MnhF, and MnhG were placed in cluster 4 together with anti-anti-sigma factor RsbV.

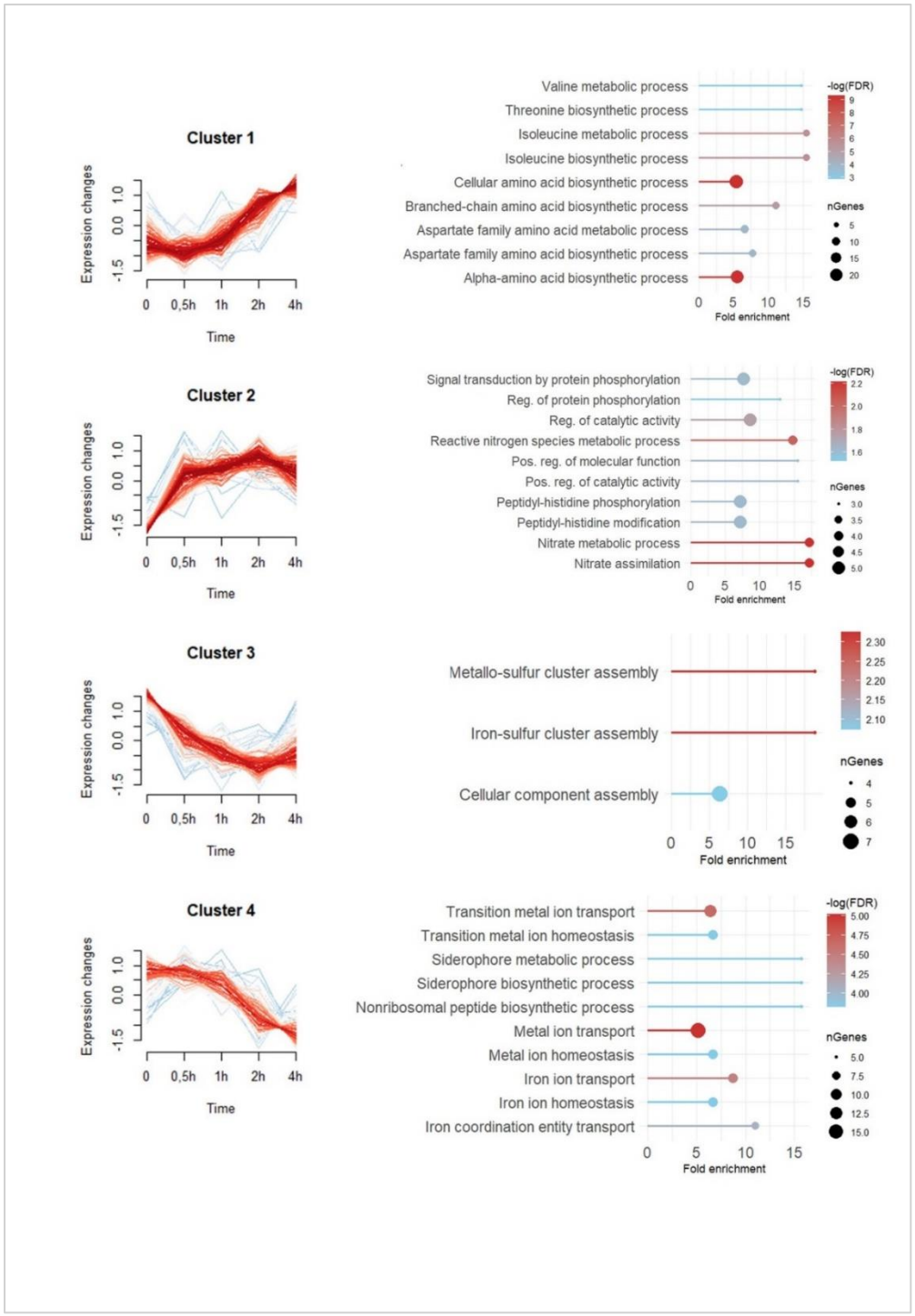


Figure 3.7: Soft clustering shows proteins with similar abundance patterns according to their temporal profiles. Enriched Gene ontology (GO) Biological process associated with each of the clusters are shown.

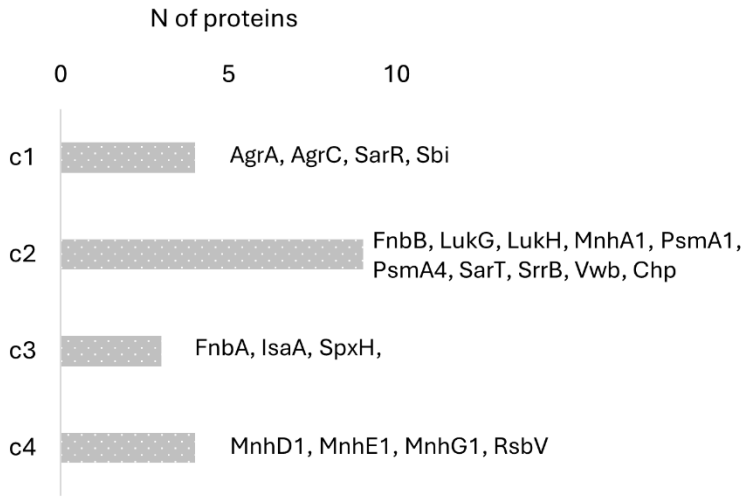


Figure 3.8: Virulence and regulatory proteins that belong to the four clusters identified by time series soft clustering analysis.

In addition to known virulence factors, each cluster contained numerous uncharacterized proteins and proteins whose role in infection is not yet elucidated. Moreover, while certain biological processes are enriched in the clusters, most proteins in these clusters do not belong to the identified enriched pathways. It is worth noting that many proteins in *S. aureus* still have unknown or poorly understood functions or poor annotation in the GO databases from which the pathway enrichments are obtained (75). To further explore what role the proteins in the clusters might have in infection, proteins with |log2FC|>3 in Cluster 1, 2 and 3 were chosen for an experimental screening of cytotoxicity against A4359 cells. Since there were only 6 proteins in cluster 4 with |log2FC|>3, proteins with |log2FC|>2.5 were also included. In total, 79 proteins were chosen for the screening experiment. The screening was performed by taking advantage of the Nebraska transposon mutant library which contains a collection of sequence-defined transposon (Tn) insertion mutants of *S. aureus* USA300. Out of 79 proteins chosen for screening, 64 were available in the mutant library. A list of the proteins with their respective identifiers and function is available in Table S3.4. To assess the growth of the mutants as compared to the parent strain (JE2), all the strains were grown in the medium used for mammalian cells growth and infection. The growth in microtiter plate was monitored in a plate reader over a period of 16 hours. Although there were some differences in the growth,

neither of the strains showed diminished growth as can be seen in Figure S3.3. Next, to test whether some of the strains would show altered cytotoxicity towards A549 cells, compared to the parent strain, a high-throughput assay was designed where A549 cell line was grown in 96-well plates and infected with a MOI 100 using bacterial cultures taken in the exponential phase and incubated for 2 hours. Following the incubation period, the human cells were washed and incubated with PI to assess the number of dead cells in the sample. The results of the high-throughput assay are shown in Figure 3.9. Although the method generally exhibited a high variation as can be seen from the large standard errors of triplicate measurements, the results revealed three proteins whose mutation significantly impacted the ability of *S. aureus* to kill host cells (p value < 0.05) as determined by two-sided t test. This included two proteins with increased virulence, namely nitrite reductase NirB and prephenate dehydrogenase TyrA, which were largely upregulated in all infection time points as compared to control (Figure 3.10). This result supports the indications that nitrite respiration plays a crucial role in the virulence of *S. aureus*. Moreover, TyrA has been shown to be directly repressed by CodY, indicating its role in virulence (30). High-affinity heme uptake system protein IsdE exhibited a lower killing capabilities. The level of this protein was downregulated in all infection time points as compared to control (Figure 3.10). Estimating cytotoxicity of bacteria towards host cells can be challenging. Most commonly used assays for cytotoxicity in general are MTT assay, Lactate dehydrogenase (LDH) and CellTiter-Glo which are based on reduction of MTT, measurement of LDH release and ATP levels, respectively. However, bacteria are also capable of these reactions, therefor when higher concentration of bacteria are present, it is difficult to estimate whether the results represent host cell death or bacterial death (76–78). Flow cytometry is another commonly used method to estimate cell death in combination with PI where host cells are separated from bacteria by size. However, constraints exist in regard to analyzing pathogens due to creation of aerosols in flow cytometry and specialized instrument configurations such special aerosol filter and enclosure in Class II biosafety cabinet (79) are not widely available. The method applied in this study using specialized imaging system enabled to differentiate between host cells and bacterial cells through nucleus size. However, this method also exhibited limitations in terms of variability, which may explain why some known mutations in toxins (e.g., LukGH) did not show a statistically significant effect on host killing. Nevertheless, the results pave the way for improving high-throughput assays, conducting individual testing of mutant strains, and thereby enhance the knowledge about the involvement of specific proteins in infection.

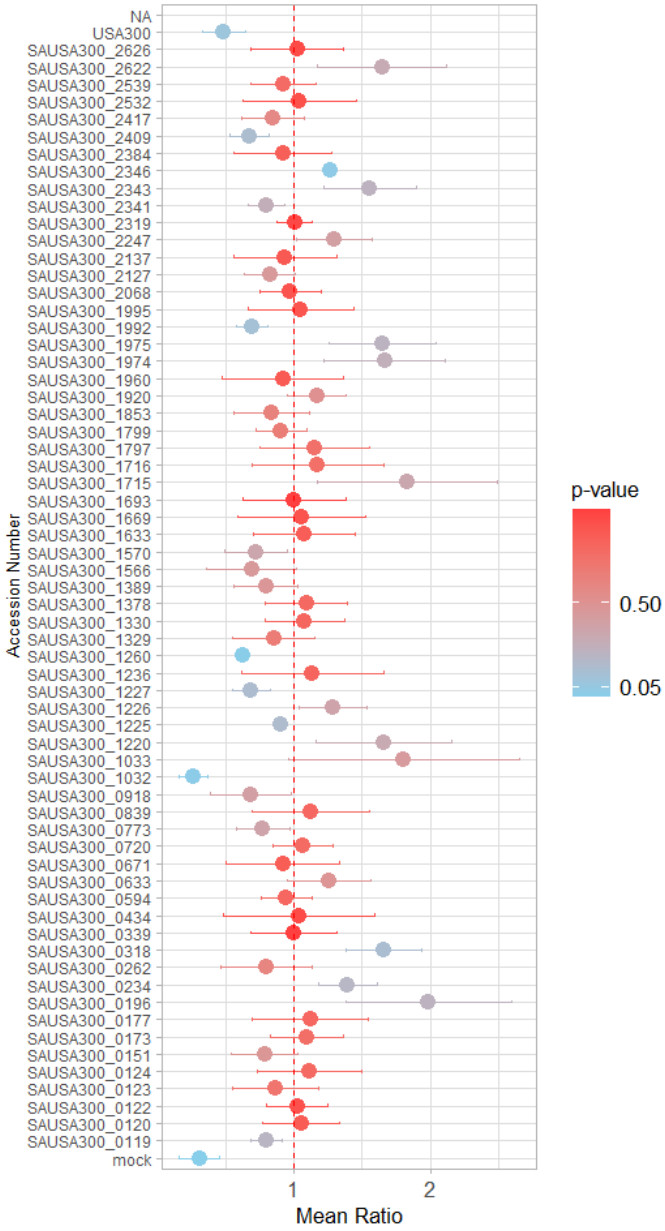


Figure 3.9: Results of the propidium iodide (PI) assay for various *Staphylococcus aureus* USA300 strains with Tn mutations from the Nebraska mutant library. The y-axis shows the names of the disrupted genes in each strain. The mean ratio on the x-axis represents the number of A549 cells stained with PI after infection by each mutant strain, divided by the number of A549 cells stained with PI after infection by the wild type (JE2).

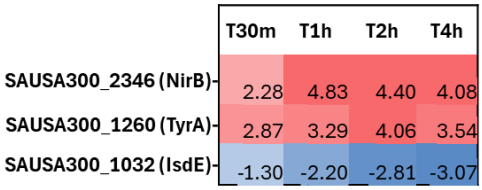


Figure 3.10: Proteome dynamics of proteins whose disruption caused significantly altered cytotoxicity towards A549 cells ($p<0.05$). The values represent log2 fold changes between the indicated infection time points and the control group.

3.4 Conclusions

In this study, a label-free quantitative proteomics approach was employed to detect and quantify changes in the abundance of *S. aureus* proteins at different time points during infection of alveolar epithelial cells. This approach identified the highest number of *S. aureus* proteins compared to previous similar studies, providing extensive data on how the bacterium modulates its proteome to adapt to the host environment and respond to host defenses. The findings highlight the pathogen's adaptation to the host's changing environment, including a switch to anaerobic respiration, upregulation of BCAAs implicated in virulence, and activation of several TCSs and virulence regulatory proteins. Additionally, we attempted to validate these findings using a high-throughput assay, which revealed several proteins that significantly alter cytotoxicity in host cells. Future studies should focus on further validating the roles of specific proteins in virulence and elucidating the functions of the many uncharacterized differentially expressed proteins. The comprehensive dataset generated in this study enables researchers to understand the expression patterns of proteins of interest and their interactions, paving the way for novel therapies and strategies to combat infections.

3.5 References

1. Howden BP, Giulieri SG, Wong Fok Lung T, Baines SL, Sharkey LK, Lee JYH, et al. *Staphylococcus aureus* host interactions and adaptation. Nat Rev Microbiol. 2023 Jun;21(6):380–95.

2. López-Jiménez AT, Mostowy S. Emerging technologies and infection models in cellular microbiology. Nat Commun. 2021 Nov 19;12(1):6764.

3. Aguilar C, Alves da Silva M, Saraiva M, Neyazi M, Olsson IAS, Bartfeld S. Organoids as host models for infection biology – a review of methods. Exp Mol Med. 2021 Oct;53(10):1471–82.

4. Reizner W, Hunter JG, O'Malley NT, Southgate RD, Schwarz EM, Kates SL. A systematic review of animal models for *Staphylococcus aureus* osteomyelitis. Eur Cell Mater. 2014 Mar 25;27:196–212.

5. Berends ETM, Zheng X, Zwack EE, Ménager MM, Cammer M, Shopsin B, et al. *Staphylococcus aureus* Impairs the Function of and Kills Human Dendritic Cells via the LukAB Toxin. mBio. 2019 Jan 2;10(1):10.1128/mbio.01918-18.

6. Grønnemose R, Garde C, Wassmann C, Klitgaard J, Nielsen R, Mandrup S, et al. Bacteria-host transcriptional response during endothelial invasion by *Staphylococcus aureus*. Sci Rep. 2021 Mar 16;11.

7. Zou T, Garifulin O, Berland R, Boyartchuk VL. *Listeria monocytogenes* infection induces prosurvival metabolic signaling in macrophages. Infect Immun. 2011 Apr;79(4):1526–35.

8. Pförtner H, Burian MS, Michalik S, Depke M, Hildebrandt P, Dhople VM, et al. Activation of the alternative sigma factor SigB of *Staphylococcus aureus* following internalization by epithelial cells - an in vivo proteomics perspective. Int J Med Microbiol IJMM. 2014 Mar;304(2):177–87.

9. Lowy FD. Is *Staphylococcus aureus* an intracellular pathogen? Trends Microbiol. 2000 Aug 1;8(8):341–3.

10. Tam K, Torres VJ. *Staphylococcus aureus* Secreted Toxins and Extracellular Enzymes. Microbiol Spectr. 2019 Mar;7(2).

11. Hébert A, Sayasith K, Sénéchal S, Dubreuil P, Lagacé J. Demonstration of intracellular *Staphylococcus aureus* in bovine mastitis alveolar cells and macrophages isolated from naturally infected cow milk. FEMS Microbiol Lett. 2000 Dec 1;193(1):57–62.

12. Eisele NA, Anderson DM. Host Defense and the Airway Epithelium: Frontline Responses That Protect against Bacterial Invasion and Pneumonia. J Pathog. 2011;2011:249802.

13. Pivard M, Moreau K, Vandenesch F. *Staphylococcus aureus* Arsenal To Conquer the Lower Respiratory Tract. mSphere. 2021 May 19;6(3):e00059-21.

14. Jean Beltran PM, Federspiel JD, Sheng X, Cristea IM. Proteomics and integrative omic approaches for understanding host-pathogen interactions and infectious diseases. Mol Syst Biol. 2017 Mar 27;13(3):922.

15. Surmann K, Simon M, Hildebrandt P, Pförtner H, Michalik S, Stentzel S, et al. A proteomic perspective of the interplay of *Staphylococcus aureus* and human alveolar epithelial cells during infection. J Proteomics. 2015 Oct 14;128:203–17.

16. Surmann K, Depke M, Dhople VM, Pané-Farré J, Hildebrandt P, Gumz J, et al. Analysis of *Staphylococcus aureus* proteins secreted inside infected human epithelial cells. Int J Med Microbiol. 2018 Aug 1;308(6):664–74.

17. Michalik S, Depke M, Murr A, Gesell Salazar M, Kusebauch U, Sun Z, et al. A global *Staphylococcus aureus* proteome resource applied to the in vivo characterization of host-pathogen interactions. Sci Rep. 2017 Sep 8;7(1):9718.

18. Palma Medina LM, Becker AK, Michalik S, Yedavally H, Raineri EJM, Hildebrandt P, et al. Metabolic Cross-talk Between Human Bronchial Epithelial Cells and Internalized *Staphylococcus aureus* as a Driver for Infection. Mol Cell Proteomics MCP. 2019 May;18(5):892–908.

19. Wiśniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. Nat Methods. 2009 May;6(5):359–62.

20. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015 Apr 20;43(7):e47.

21. Kumar L, E. Futschik M. Mfuzz: A software package for soft clustering of microarray data. Bioinformation. 2007 May 20;2(1):5–7.

22. Ge SX, Jung D, Yao R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinforma Oxf Engl. 2020 Apr 15;36(8):2628–9.

23. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009 Dec 15;10:421.

24. Fey PD, Endres JL, Yajjala VK, Widhelm TJ, Boissy RJ, Bose JL, et al. A genetic resource for rapid and comprehensive phenotype screening of nonessential *Staphylococcus aureus* genes. mBio. 2013 Feb 12;4(1):e00537-00512.

25. Leiva-Juárez MM, Kolls JK, Evans SE. Lung epithelial cells: therapeutically inducible effectors of antimicrobial defense. Mucosal Immunol. 2018 Jan;11(1):21–34.

26. Shi D, Mi G, Wang M, Webster TJ. In vitro and ex vivo systems at the forefront of infection modeling and drug discovery. Biomaterials. 2019 Apr;198:228–49.

27. Hirsch C, Schildknecht S. In Vitro Research Reproducibility: Keeping Up High Standards. Front Pharmacol. 2019;10:1484.

28. Fuchs S, Mehlan H, Bernhardt J, Hennig A, Michalik S, Surmann K, et al. AureoWiki‑The repository of the *Staphylococcus aureus* research and annotation community. Int J Med Microbiol IJMM. 2018 Aug;308(6):558–68.

29. Bai M, Deng J, Dai C, Pfeuffer J, Sachsenberg T, Perez-Riverol Y. LFQ-Based Peptide and Protein Intensity Differential Expression Analysis. J Proteome Res. 2023 Jun 2;22(6):2114–23.

30. Majerczyk CD, Dunman PM, Luong TT, Lee CY, Sadykov MR, Somerville GA, et al. Direct Targets of CodY in *Staphylococcus aureus*. J Bacteriol. 2010 Jun;192(11):2861–77.

31. Barbuti MD, Myrbråten IS, Morales Angeles D, Kjos M. The cell cycle of *Staphylococcus aureus*: An updated review. MicrobiologyOpen. 2023;12(1):e1338.

32. Alfeo MJ, Pagotto A, Barbieri G, Foster TJ, Vanhoorelbeke K, De Filippis V, et al. *Staphylococcus aureus* iron-regulated surface determinant B (IsdB) protein interacts with von Willebrand factor and promotes adherence to endothelial cells. Sci Rep. 2021 Nov 23;11(1):22799.

33. Claes J, Vanassche T, Peetermans M, Liesenborghs L, Vandenbriele C, Vanhoorelbeke K, et al. Adhesion of *Staphylococcus aureus* to the vessel wall under flow is mediated by von Willebrand factor–binding protein. Blood. 2014 Sep 4;124(10):1669–76.

34. Oriol C, Cengher L, Manna AC, Mauro T, Pinel-Marie ML, Felden B, et al. Expanding the *Staphylococcus aureus* SarA Regulon to Small RNAs. mSystems. 2021 Oct 12;6(5):10.1128/msystems.00713-21.

35. Fuchs S, Pané-Farré J, Kohler C, Hecker M, Engelmann S. Anaerobic Gene Expression in *Staphylococcus aureus*. J Bacteriol. 2007 Jun;189(11):4275–89.

36. Delauné A, Dubrac S, Blanchet C, Poupel O, Mäder U, Hiron A, et al. The WalKR System Controls Major Staphylococcal Virulence Genes and Is Involved in Triggering the Host Inflammatory Response. Infect Immun. 2012 Oct;80(10):3438–53.

37. Qi X, Brothers KM, Ma D, Mandell JB, Donegan NP, Cheung AL, et al. The *Staphylococcus aureus* toxin–antitoxin system YefM–YoeB is associated with antibiotic tolerance and extracellular dependent biofilm formation. J Bone Jt Infect. 2021 Jul 2;6(7):241–53.

38. Palma M, Cheung AL. sigma(B) activity in *Staphylococcus aureus* is controlled by RsbU and an additional factor(s) during bacterial growth. Infect Immun. 2001 Dec;69(12):7858–65.

39. Kuhn M, Goebel W. Internalization of *Listeria monocytogenes* by nonprofessional and professional phagocytes. Subcell Biochem. 2000;33:411–36.

40. Sun JL, Zhang SK, Chen JY, Han BZ. Metabolic profiling of *Staphylococcus aureus* cultivated under aerobic and anaerobic conditions with 1H NMR-based nontargeted analysis. Can J Microbiol. 2012 Jun;58(6):709–18.

41. Richardson AR, Libby SJ, Fang FC. A Nitric Oxide–Inducible Lactate Dehydrogenase Enables *Staphylococcus aureus* to Resist Innate Immunity. Science. 2008 Mar 21;319(5870):1672–6.

42. Chaffin DO, Taylor D, Skerrett SJ, Rubens CE. Changes in the *Staphylococcus aureus* Transcriptome during Early Adaptation to the Lung. PLOS ONE. 2012 Aug 2;7(8):e41329.

43. Bronner S, Monteil H, Prévost G. Regulation of virulence determinants in *Staphylococcus aureus*: complexity and applications. FEMS Microbiol Rev. 2004 May 1;28(2):183–200.

44. Jenul C, Horswill AR. Regulation of *Staphylococcus aureus* virulence. Microbiol Spectr. 2018 Feb;6(1):10.1128/microbiolspec.GPP3-0031–2018.

45. Haag AF, Bagnoli F. The Role of Two-Component Signal Transduction Systems in *Staphylococcus aureus* Virulence Regulation. Curr Top Microbiol Immunol. 2017;409:145–98.

46. Dinges MM, Orwin PM, Schlievert PM. Exotoxins of *Staphylococcus aureus*. Clin Microbiol Rev. 2000 Jan;13(1):16–34.

47. Omae Y, Sekimizu K, Kaito C. Inhibition of Colony-spreading Activity of *Staphylococcus aureus* by Secretion of δ-Hemolysin. J Biol Chem. 2012 May 1;287(19):15570–9.

48. Peschel A, Otto M. Phenol-soluble modulins and staphylococcal infection. Nat Rev Microbiol. 2013 Oct;11(10):667–73.

49. Queck SY, Jameson-Lee M, Villaruz AE, Bach THL, Khan BA, Sturdevant DE, et al. RNAIII-independent target gene control by the agr quorum-sensing system: insight into the evolution of virulence regulation in *Staphylococcus aureus*. Mol Cell. 2008 Oct 10;32(1):150–8.

50. Dunman PM, Murphy E, Haney S, Palacios D, Tucker-Kellogg G, Wu S, et al. Transcription profiling-based identification of *Staphylococcus aureus* genes regulated by the agr and/or sarA loci. J Bacteriol. 2001 Dec;183(24):7341–53.

51. Powers ME, Kim HK, Wang Y, Bubeck Wardenburg J. ADAM10 Mediates Vascular Injury Induced by *Staphylococcus aureus* α-Hemolysin. J Infect Dis. 2012 Aug 1;206(3):352–6.

52. Inoshima N, Wang Y, Wardenburg JB. Genetic Requirement for ADAM10 in Severe *Staphylococcus aureus* Skin Infection. J Invest Dermatol. 2012 May;132(5):1513–6.

53. Wardenburg JB, Patel RJ, Schneewind O. Surface Proteins and Exotoxins Are Required for the Pathogenesis of *Staphylococcus aureus* Pneumonia. Infect Immun. 2007 Feb;75(2):1040–4.

54. Vaish M, Price-Whelan A, Reyes-Robles T, Liu J, Jereen A, Christie S, et al. Roles of *Staphylococcus aureus* Mnh1 and Mnh2 Antiporters in Salt Tolerance, Alkali Tolerance, and Pathogenesis. J Bacteriol. 2018 Feb 7;200(5):e00611-17.

55. Reyes D, Andrey DO, Monod A, Kelley WL, Zhang G, Cheung AL. Coordinated regulation by AgrA, SarA, and SarR to control agr expression in *Staphylococcus aureus*. J Bacteriol. 2011 Nov;193(21):6020–31.

56. Yarwood JM, McCormick JK, Schlievert PM. Identification of a novel two-component regulatory system that acts in global regulation of virulence factors of *Staphylococcus aureus*. J Bacteriol. 2001 Feb;183(4):1113–23.

57. Dunman PM, Murphy E, Haney S, Palacios D, Tucker-Kellogg G, Wu S, et al. Transcription Profiling-Based Identification of*Staphylococcus aureus* Genes Regulated by the agrand/or sarA Loci. J Bacteriol. 2001 Dec 15;183(24):7341–53.

58. Liu Q, Yeo WS, Bae T. The SaeRS Two-Component System of *Staphylococcus aureus*. Genes. 2016 Oct 3;7(10):81.

59. Smith EJ, Corrigan RM, van der Sluis T, Gründling A, Speziale P, Geoghegan JA, et al. The immune evasion protein Sbi of *Staphylococcus aureus* occurs both extracellularly and anchored to the cell envelope by binding lipoteichoic acid. Mol Microbiol. 2012 Feb;83(4):789–804.

60. Josse J, Laurent F, Diot A. Staphylococcal Adhesion and Host Cell Invasion: Fibronectin-Binding and Other Mechanisms. Front Microbiol. 2017 Dec 5;8:2433.

61. de Haas CJC, Veldkamp KE, Peschel A, Weerkamp F, Van Wamel WJB, Heezius ECJM, et al. Chemotaxis Inhibitory Protein of *Staphylococcus aureus*, a Bacterial Antiinflammatory Agent. J Exp Med. 2004 Mar 1;199(5):687–95.

62. Tiwari N, López-Redondo M, Miguel-Romero L, Kulhankova K, Cahill MP, Tran PM, et al. The SrrAB two-component system regulates *Staphylococcus aureus* pathogenicity through redox sensitive cysteines. Proc Natl Acad Sci U S A. 2020 May 19;117(20):10989–99.

63. Morikawa K, Ohniwa RL, Kim J, Maruyama A, Ohta T, Takeyasu K. Bacterial nucleoid dynamics: oxidative stress response in *Staphylococcus aureus*. Genes Cells. 2006;11(4):409–23.

64. Nakamura Y, Kanemaru K, Shoji M, Totoki K, Nakamura K, Nakaminami H, et al. Phosphatidylinositol-specific phospholipase C enhances epidermal penetration by *Staphylococcus aureus*. Sci Rep. 2020 Oct 20;10(1):17845.

65. Memmi G, Nair DR, Cheung A. Role of ArlRS in Autolysis in Methicillin-Sensitive and Methicillin-Resistant *Staphylococcus aureus* Strains. J Bacteriol. 2012 Jan 24;194(4):759–67.

66. Crosby HA, Schlievert PM, Merriman JA, King JM, Salgado-Pabón W, Horswill AR. The *Staphylococcus aureus* Global Regulator MgrA Modulates Clumping and Virulence by Controlling Surface Protein Expression. PLoS Pathog. 2016 May;12(5):e1005604.

67. Kullik I, Giachino P, Fuchs T. Deletion of the Alternative Sigma Factor ςB in *Staphylococcus aureus* Reveals Its Function as a Global Regulator of Virulence Genes. J Bacteriol. 1998 Sep;180(18):4814–20.

68. Lei T, Yang J, Ji Y. Determination of essentiality and regulatory function of staphylococcal YeaZ in branched-chain amino acid biosynthesis. Virulence. 2014 Dec 17;6(1):75–84.

69. Kaiser JC, King AN, Grigg JC, Sheldon JR, Edgell DR, Murphy MEP, et al. Repression of branched-chain amino acid synthesis in *Staphylococcus aureus* is mediated by isoleucine via CodY, and by a leucine-rich attenuator peptide. PLoS Genet. 2018 Jan 22;14(1):e1007159.

70. Pendleton A, Yeo WS, Alqahtani S, DiMaggio DA, Stone CJ, Li Z, et al. Regulation of the Sae Two-Component System by Branched-Chain Fatty Acids in *Staphylococcus aureus*. mBio. 2022 Sep 22;13(5):e01472-22.

71. Qiu Y, Xu D, Xia X, Zhang K, Aadil RM, Batool Z, et al. Five major two components systems of *Staphylococcus aureus* for adaptation in diverse hostile environment. Microb Pathog. 2021 Oct 1;159:105119.

72. Andrews SC, Robinson AK, Rodríguez-Quiñones F. Bacterial iron homeostasis. FEMS Microbiol Rev. 2003 Jun;27(2–3):215–37.

73. Friedman DB, Stauff DL, Pishchany G, Whitwell CW, Torres VJ, Skaar EP. *Staphylococcus aureus* Redirects Central Metabolism to Increase Iron Availability. PLOS Pathog. 2006 Aug 25;2(8):e87.

74. Hamamoto H, Panthee S, Paudel A, Ohgi S, Suzuki Y, Makimura K, et al. Transcriptome change of *Staphylococcus aureus* in infected mouse liver. Commun Biol. 2022 Jul 20;5(1):1–10.

75. Hu S, Luo Y, Zhang Z, Xiong H, Yan W, Jiang M, et al. Protein function annotation based on heterogeneous biological networks. BMC Bioinformatics. 2022 Nov 18;23(1):493.

76. Xu T, Marr E, Lam H, Ripp S, Sayler G, Close D. Real-time toxicity and metabolic activity tracking of human cells exposed to Escherichia coli O157:H7 in a mixed consortia. Ecotoxicology. 2015 Dec 1;24(10):2133–40.

77. Van den Bossche S, Vandeplassche E, Ostyn L, Coenye T, Crabbé A. Bacterial Interference With Lactate Dehydrogenase Assay Leads to an Underestimation of Cytotoxicity. Front Cell Infect Microbiol. 2020 Sep 15;10:494.

78. Benov L. Improved Formazan Dissolution for Bacterial MTT Assay. Microbiol Spectr. 2021 Dec 22;9(3):e01637-21.

79. Reifel KM, Swan BK, Jellison ER, Ambrozak D, Baijer J, Nguyen R, et al. Procedures for Flow Cytometry-Based Sorting of Unfixed Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Infected Cells and Other Infectious Agents. Cytometry. 2020 Jul;97(7):674–80.

80. Krogh AKH, Haaber J, Bochsen L, Ingmer H, Kristensen AT. Aggregating resistant *Staphylococcus aureus* induces hypocoagulability, hyperfibrinolysis, phagocytosis, and neutrophil, monocyte, and lymphocyte binding in canine whole blood. Vet Clin Pathol. 2018 Dec;47(4):560-574.

3.6 Supporting information

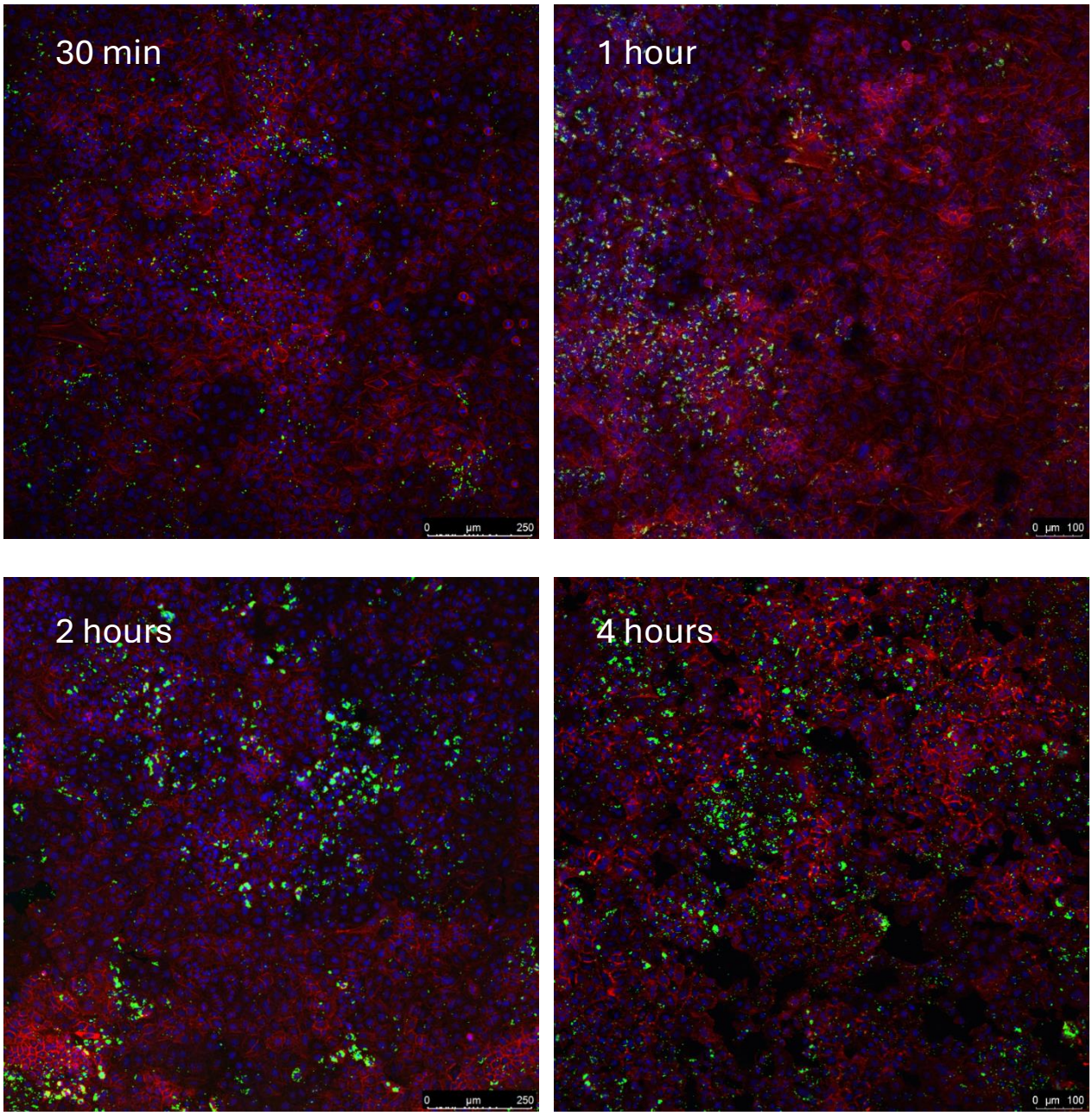


Figure S3.1: Confocal microscopy images depicting GFP expressing *Staphycloccus aureus USA*300 (green) (80) co-cultivated with A549 cell line in which nucleus was stained blue and actin was stained red.

Table S3.1: Results of differential expression analysis for all contrasts. A) The results show the number of proteins termed upregulated (log2FC>1), downregulated (log2FC<1) and non-significantly changed proteins (adjusted p-value >0.05). B) Number of proteins termed upregulated and downregulated based only on adjusted p-value (<0.05) while log2FC is disregarded.

A	T30m-ctr	T1h-ctr	T2h-ctr	T4h-ctr	T1h-T30m	T2h-T30m	T4h-T30m	T2h-T1h	T4h-T1h	T4h-T2h
Down	43	104	166	140	11	61	90	27	51	22
NotSig	1459	1377	1253	1259	1556	1435	1362	1515	1434	1545
Up	94	115	177	197	29	100	144	54	111	29
B	**T30m-ctr**	**T1h-ctr**	**T2h-ctr**	**T4h-ctr**	**T1h-T30m**	**T2h-T30m**	**T4h-T30m**	**T2h-T1h**	**T4h-T1h**	**T4h-T2h**
Down	547	579	701	597	237	500	474	308	336	138
NotSig	759	695	506	563	1179	720	675	1024	875	1207
Up	290	322	389	436	180	376	447	264	385	251

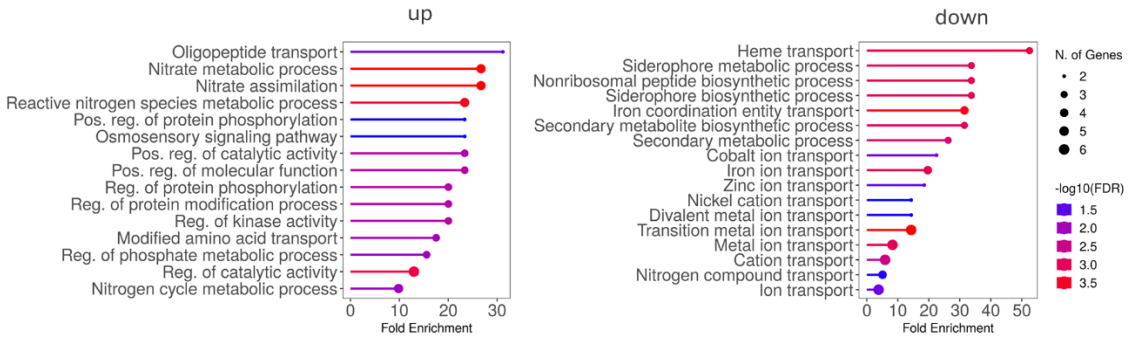


Figure S3.2: Gene ontology (GO) enrichment of biological processes of proteins upregulated and downregulated in all infection time points (T30, T1h, T2h and T4h) as compared to the control (adjusted p-value <0.05, |log2FC|>1). ShinyGO 0.80 was used to infer the enriched GO terms

Table S3.2 Uniquely upregulated proteins for time points 30 min, 1h, 2h and 4 h (T30m, T1h, T2h and T4h). The table shows Uniprot accession number, log2FC and respective contrasts, protein description and gene name. If the protein had a log2FC higher than 1 in other time points it means that in that particular time point, the change was not statistically significant (adj.p val>0.05).

TP	Accession	T30m.ctr	T1h.ctr	T2h.ctr	T4h.ctr	Description	Gene name
30	Q2FE03	1.18	0.59	-0.43	-0.32	Fibronectin-binding protein A	*fnbA*
30	Q2FJC0	1.05	0.57	0.50	0.39	Pyridoxal 5'-phosphate synthase subunit PdxT	*pdxT*
30	A0A0H2XFY2	1.45	0.65	0.44	0.44	ABC transporter, ATP-binding protein	SAUSA300_0271
30	Q2FJQ9	1.49	0.19	0.64	0.06	Probable acetyl-CoA acyltransferase	SAUSA300_0355
30	Q2FJ96	1.29	0.95	0.37	0.05	Putative ribosomal protein L7Ae-like	SAUSA300_0529
30	A0A0H2XIG6	1.36	0.34	-0.31	-0.15	Oxidoreductase, short-chain dehydrogenase/reductase family	SAUSA300_1451
30	A0A0H2XH33	1.15	0.71	0.81	0.50	Uncharacterized protein	SAUSA300_2529
30	Q2FHJ8	1.14	0.84	0.74	0.32	tRNA (guanine-N(1)-)-methyltransferase	trmD
1	Q2FJB6	0.55	1.08	0.23	-0.04	Protein-arginine kinase	*mcsB*
1	Q2FK43	0.08	2.25	0.73	-0.24	Pyruvate formate-lyase-activating enzyme	*pflA*
1	A0A0H2XHY3	0.65	1.15	0.39	0.31	Glycerate kinase	SAUSA300_2377
2	A0A0H2XF86	0.45	0.91	1.19	1.00	Molybdopterin molybdenumtransferase	*moeA*
2	A0A0H2XFS9	0.76	0.99	1.32	0.77	Sensor protein kinase WalK	*phoR*

2	P0C7Y0	0.59	0.88	1.40	0.75	Phenol-soluble modulin alpha 1 peptide	*psmA1*
2	P0C817	0.58	0.85	1.36	0.73	Phenol-soluble modulin alpha 4 peptide	*psmA4*
2	Q2FJJ6	0.56	0.85	1.04	0.91	Uncharacterized lipoprotein SAUSA300_0419	SAUSA300_0419
2	A0A0H2XE39	0.65	0.57	1.06	0.87	Uncharacterized protein	SAUSA300_0592
2	A0A0H2XF63	0.57	0.72	1.30	0.72	Uncharacterized protein	SAUSA300_1906
2	A0A0H2XH16	0.28	0.65	1.91	0.99	Delta-hemolysin	SAUSA300_1988
2	A0A0H2XFJ4	0.11	0.99	1.28	0.53	Uncharacterized protein	SAUSA300_2004
2	A0A0H2XF98	0.59	0.66	1.37	0.77	Precorrin-2 dehydrogenase	SAUSA300_2553
2	A0A0H2XGI6	0.84	0.74	1.31	0.12	Single-stranded DNA-binding protein	ssb
4	A0A0H2XFQ6	0.66	0.74	0.96	1.01	Alpha-acetolactate decarboxylase	*budA*
4	A0A0H2XGD4	0.14	0.58	0.89	1.04	Aminoacyltransferase FemA	*fmhA*
4	A0A0H2XFK8	0.45	0.99	0.89	1.03	Coproporphyrinogen III oxidase	*hemG*
4	A0A0H2XGU8	-2.82	-3.14	-1.35	1.66	Putative pyridoxal phosphate-dependent acyltransferase	*hisC*
4	Q2FK46	0.09	0.45	0.76	1.09	Sensor protein kinase HptS	*hptS*
4	A0A0H2XFY1	0.32	0.40	0.62	1.14	Peptide methionine sulfoxide reductase regulator MsrR	*msrR*
4	Q2FFJ3	0.26	0.59	0.81	1.30	Sodium/proline symporter	*putP*
4	Q2FEJ8	0.40	0.34	0.87	1.81	HTH-type transcriptional regulator SarR	*sarR*
4	A0A0H2XI02	0.25	0.93	0.97	1.26	5'-nucleotidase family protein	SAUSA300_0025
4	A0A0H2XI33	0.19	0.12	0.71	1.06	5' nucleotidase family protein	SAUSA300_0147

4	A0A0H2XFU2	-0.17	-0.10	0.78	1.33	Uncharacterized protein	SAUSA300_0198
4	A0A0H2XF57	-0.34	-0.12	0.16	1.18	Uncharacterized protein	SAUSA300_0261
4	A0A0H2XEY5	0.05	-0.05	0.70	1.35	Uncharacterized protein	SAUSA300_0384
4	A0A0H2XH20	-0.05	0.26	0.74	1.31	Uncharacterized protein	SAUSA300_0460
4	A0A0H2XFV1	-0.56	0.05	0.97	1.75	HAD-superfamily hydrolase, subfamily IA, variant 1	SAUSA300_0540
4	A0A0H2XEG8	0.33	0.39	0.92	1.38	Acetyltransferase, GNAT family	SAUSA300_0785
4	Q2FI62	-0.16	0.30	0.82	1.03	Putative phosphoesterase SAUSA300_0916	SAUSA300_0916
4	A0A0H2XFA8	-0.31	-0.30	0.67	1.49	Aminotransferase	SAUSA300_0952
4	A0A0H2XGR8	0.29	0.19	0.84	1.26	GCN5-related N-acetyltransferase	SAUSA300_0956
4	Q2FI23	0.02	0.05	0.79	1.39	Uncharacterized protein SAUSA300_0957	SAUSA300_0957
4	Q2FHW8	0.47	0.16	0.17	1.67	UPF0358 protein SAUSA300_1012	SAUSA300_1012
4	A0A0H2XJ11	0.25	0.65	0.56	1.04	Uncharacterized protein	SAUSA300_1057
4	A0A0H2XHB3	0.06	0.21	0.69	1.21	Purine nucleoside phosphorylase	SAUSA300_1081
4	A0A0H2XIM6	0.07	0.40	0.87	1.04	Uncharacterized protein	SAUSA300_1560
4	Q2FGA9	0.04	0.22	0.59	1.12	UPF0297 protein SAUSA300_1574	SAUSA300_1574
4	A0A0H2XIF3	0.11	-0.11	0.72	1.45	ThiF domain-containing protein	SAUSA300_1585
4	A0A0H2XHN7	-0.16	-0.29	0.15	1.21	CBS domain protein	SAUSA300_1651
4	A0A0H2XI85	-0.20	0.00	0.86	1.29	Chorismate mutase/phospho-2-dehydro-3-deoxyheptonate aldolase	SAUSA300_1683
4	A0A0H2XGB4	-1.20	-0.68	-0.06	1.12	Uncharacterized protein	SAUSA300_1905

4	A0A0H2XHE0	-0.72	-0.57	0.26	1.04	Uncharacterized protein	SAUSA300_2093
4	A0A0H2XEW9	-0.03	0.00	-0.38	1.17	Cation efflux family protein	SAUSA300_2099
4	A0A0H2XGP0	0.03	0.14	0.98	1.45	Uncharacterized protein	SAUSA300_2209
4	A0A0H2XI51	-0.32	-0.07	0.80	1.19	Lactose phosphotransferase system repressor	SAUSA300_2261
4	Q2FEC8	-0.76	-0.53	0.72	1.49	Uncharacterized lipoprotein SAUSA300_2315	SAUSA300_2315
4	A0A0H2XGW2	0.58	0.61	0.97	1.00	Transcription regulatory protein	SAUSA300_2326
4	A0A0H2XDX7	-0.11	0.80	0.97	1.69	Putative membrane protein	SAUSA300_2385
4	A0A0H2XJ90	-0.42	-0.11	0.56	1.11	D-lactate dehydrogenase	SAUSA300_2496
4	A0A0H2XHD4	-0.47	-0.32	0.19	1.48	Amino acid permease family protein	SAUSA300_2538
4	A0A0H2XIM9	0.35	0.75	0.90	1.05	Sensor histidine kinase	SAUSA300_2558
4	A0A0H2XI01	0.13	0.34	0.55	1.08	N-acetylmuramoyl-L-alanine amidase domain protein	SAUSA300_2579
4	A0A0H2XHN5	-0.41	-0.53	-0.42	1.22	Uncharacterized protein	SAUSA300_2615
4	A0A0H2XFK9	0.29	0.50	0.98	2.55	Putative cobalt ABC transporter, ATP-binding protein	SAUSA300_2617
4	Q2FHA3	-0.11	-0.15	0.68	1.07	Homoserine kinase	thrB

Table S3.3: Uniquely downregulated proteins for time points 30 min, 1h, 2h and 4 h (T30m, T1h, T2h and T4h). The table shows Uniprot accession number, log2FC and respective contrasts, protein description and gene name. If the protein had a log2FC lower than -1 in other time points it means that in that particular time point, the change was not statistically significant (adj.p val>0.05).

TP	Accession	T30m.ctr	T1h.ctr	T2h.ctr	T4h.ctr	Description	Description
30	A0A0H2XGB4	-1.20	-0.68	-0.06	1.12	Uncharacterized protein	SAUSA300_1905
30	A0A0H2XFJ2	-1.33	-0.54	0.49	0.90	Sodium-dependent dicarboxylate transporter SdcS	SAUSA300_0676
30	A0A0H2XGV3	-1.62	-0.80	0.00	0.68	Acetyltransferase, GNAT family	SAUSA300_2316
30	A0A0H2XII5	-1.07	-0.04	0.78	0.35	Mannitol-specific phosphotransferase enzyme IIA component	SAUSA300_2576
30	A0A0H2XDY8	-1.09	-0.70	-0.33	0.10	Prephenate dehydratase	pheA
30	A0A0H2XHG8	-1.03	-0.76	-0.86	-0.08	Putative membrane protein	SAUSA300_1910
30	A0A0H2XG26	-1.05	-0.93	-0.62	-0.34	Inosine-uridine preferring nucleoside hydrolase	SAUSA300_0237
30	A0A0H2XH86	-1.11	-0.66	-0.95	-0.43	6-phospho-beta-glucosidase	bglA
30	A0A0H2XGQ0	-1.03	-0.75	-0.47	-0.92	Uncharacterized protein	SAUSA300_1186
1	A0A0H2XFQ5	-0.76	-1.01	-0.97	-0.91	Uncharacterized protein	SAUSA300_2138
1	A0A0H2XH23	-0.63	-1.01	-0.92	0.51	Putative lipoprotein	SAUSA300_2355
1	A0A0H2XH13	-0.33	-1.47	-0.33	-0.47	Sugar efflux transporter	SAUSA300_0659
1	Q2FED4	-0.87	-1.17	-0.89	-0.72	Heme sensor protein HssS	hssS
1	Q2FDQ4	-0.82	-1.02	-0.97	-0.81	Fructose-bisphosphate aldolase class 1	fda
1	A0A0H2XIJ3	-0.85	-1.37	-0.96	-0.84	Bifunctional autolysin	atl
2	A0A0H2XJ67	-0.95	-0.78	-1.11	0.19	Putative mRNA interferase YoeB	SAUSA300_2352
2	A0A0H2XH11	-1.29	-0.29	-2.33	0.12	Glycine betaine aldehyde dehydrogenase	betB
2	A0A0H2XI18	-0.49	-0.82	-1.13	-0.09	Uncharacterized protein	SAUSA300_2236
2	Q2FEM4	-0.51	-0.16	-1.33	-0.33	GTP 3',8-cyclase	moaA

2	A0A0H2XEK8	-1.33	-1.11	-1.54	-0.34	Membrane-embedded CAAX protease MroQ	MroQ
2	A0A0H2XI82	-0.60	-0.70	-1.17	-0.36	o-succinylbenzoate synthase	menC
2	Q2FIS8	-0.82	-0.88	-1.23	-0.50	7-cyano-7-deazaguanine synthase	queC
2	Q2FFZ0	-0.57	-0.36	-1.17	-0.52	tRNA (guanine-N(7)-)-methyltransferase	trmB
2	A0A0H2XI21	-0.99	-0.79	-1.09	-0.56	Uncharacterized protein	SAUSA300_0942
2	A0A0H2XJ96	-0.39	-0.68	-1.06	-0.59	NADH-dependent flavin oxidoreductase, Oye family	SAUSA300_0322
2	A0A0H2XDT8	-0.03	0.12	-1.07	-0.60	Release factor glutamine methyltransferase	prmC
2	A0A0H2XJ47	-0.09	-0.82	-1.18	-0.61	Veg protein	SAUSA300_0471
2	A0A0H2XG98	-0.22	-0.69	-1.04	-0.65	Anti-sigma factor antagonist	rsbV
2	A0A0H2XF32	-0.42	-0.36	-1.02	-0.67	Endonuclease III	nth
2	A0A0H2XGK3	-0.64	-0.89	-1.57	-0.69	ATP-dependent RecD-like DNA helicase	recD2
2	A0A0H2XF42	-0.98	-0.93	-1.02	-0.73	Cytochrome D ubiquinol oxidase, subunit I	SAUSA300_0986
2	A0A0H2XDP9	-0.19	-0.61	-1.14	-0.79	ABC transporter, ATP-binding protein	SAUSA300_2357
2	A0A0H2XHV8	-0.10	-0.11	-1.14	-0.80	Putative aluminium resistance protein	SAUSA300_1199
2	A0A0H2XGQ7	-0.15	-0.56	-1.30	-0.80	Uncharacterized protein	SAUSA300_0830
2	Q2FEK0	-0.55	-0.58	-1.49	-0.80	Urease accessory protein UreG	ureG
2	Q2FDU3	-0.44	-0.55	-1.06	-0.81	4,4'-diaponeurosporene oxygenase	crtP
2	A0A0H2XE11	-0.50	-0.44	-1.08	-0.82	Methyltransferase	SAUSA300_0464
2	A0A0H2XHT5	-0.82	-0.90	-1.18	-0.83	Uncharacterized protein	SAUSA300_1084

2	A0A0H2XK42	-0.20	-0.75	-1.02	-0.89	Amino acid ABC transporter, amino acid-binding protein	SAUSA300_2359
2	A0A0H2XJ20	-0.44	-0.87	-1.50	-0.90	Uncharacterized protein	SAUSA300_2592
2	A0A0H2XFM8	-0.63	-0.95	-1.28	-0.91	Acetyltransferase, GNAT family	SAUSA300_1312
2	A0A0H2XFG6	-0.56	-0.93	-1.01	-0.94	Uncharacterized protein	SAUSA300_1698
2	A0A0H2XGU3	-0.37	-0.66	-1.03	-0.97	Uncharacterized protein	SAUSA300_1859
2	Q2FIR0	-0.71	-0.43	-1.86	-0.97	Protein NrdI	nrdI
2	A0A0H2XGB6	-0.32	-0.75	-1.18	-0.99	ABC transporter, permease protein	SAUSA300_2358
2	A0A0H2XIW2	-0.32	-0.48	-1.14	-0.99	Membrane protein, TerC family	SAUSA300_0922
2	A0A0H2XEB7	-0.55	-0.79	-1.11	-0.99	Integral membrane protein	SAUSA300_0729
4	A0A0H2XEB5	-0.10	-0.37	-0.57	-1.02	RNA methyltransferase, TrmA family	rumA
4	A0A0H2XH84	-0.66	-0.63	-0.91	-1.02	Putative XpaC protein	SAUSA300_1298
4	Q2FHD9	-0.27	-0.63	-0.91	-1.02	Glycerol kinase	glpK
4	A0A0H2XHJ8	-0.22	-0.40	-0.54	-1.05	Peptidase, M16 family	SAUSA300_1172
4	A0A0H2XG46	-0.47	-0.79	-0.95	-1.06	Putative oxidoreductase	SAUSA300_0329
4	A0A0H2XH18	-0.52	-0.29	-0.88	-1.07	Lysostaphin	SAUSA300_2162
4	A0A0H2XIU9	-0.07	0.06	-0.63	-1.08	Uncharacterized protein	SAUSA300_0700
4	Q2FIF6	-0.15	-0.53	-0.92	-1.09	UPF0051 protein SAUSA300_0822	SAUSA300_0822
4	Q2FI79	-0.08	-0.26	-0.58	-1.13	Adapter protein MecA	mecA
4	A0A0H2XHZ1	-0.53	-0.85	-0.77	-1.13	Putative lysophospholipase	SAUSA300_0070
4	Q2FIE9	0.07	0.06	-0.89	-1.13	Lipoyl synthase	lipA
4	A0A0H2XGP9	0.22	-0.10	-0.66	-1.15	Uncharacterized protein	SAUSA300_0931
4	A0A0H2XFH1	-0.04	-0.28	-0.86	-1.17	Ribosomal silencing factor RsfS	rsfS

4	Q2FFZ8	0.08	-0.07	-0.95	-1.17	UDP-N-acetylmuramate--L-alanine ligase	murC
4	A0A0H2XE16	0.34	-0.09	-0.14	-1.24	5-formyltetrahydrofolate cyclo-ligase	SAUSA300_1510
4	A0A0H2XK28	-0.85	-0.72	-0.59	-1.25	PTS system mannitol-specific EIICB component	mtlF
4	A0A0H2XHA6	-0.13	-0.55	-0.28	-1.25	Uncharacterized protein	SAUSA300_2522
4	A0A0H2XFG7	-0.10	-0.65	-0.69	-1.27	Glycerol uptake facilitator	glpF
4	A0A0H2XG96	0.93	1.34	1.08	-1.27	Uncharacterized protein	SAUSA300_0086
4	A0A0H2XHW4	-0.14	-0.18	-0.82	-1.28	Uncharacterized protein	SAUSA300_1069
4	A0A0H2XJM3	0.15	0.12	-0.84	-1.30	Uncharacterized protein	SAUSA300_1888
4	A0A0H2XF46	-0.45	-0.20	-0.67	-1.36	Uncharacterized protein	SAUSA300_0858
4	Q2FIS2	-0.19	-0.72	-0.54	-1.40	Lipoteichoic acid synthase	ltaS
4	Q2FDT8	-0.82	-0.20	-0.80	-1.48	Probable transglycosylase IsaA	isaA
4	A0A0H2XH27	-0.34	-0.04	-0.62	-1.77	Uncharacterized protein	SAUSA300_1180
4	A0A0H2XDY7	-0.38	-0.46	-0.97	-1.86	Uncharacterized protein	SAUSA300_1019
4	A0A0H2XEB3	-0.78	-0.88	-1.06	-2.06	Uncharacterized protein	SAUSA300_0439
4	A0A0H2XI49	-0.13	-0.47	-0.85	-2.08	Sucrose operon repressor	scrR
4	A0A0H2XEI1	-0.09	-0.31	-0.49	-2.41	Uncharacterized protein	SAUSA300_1236

Table S3.4: Mutant strain chosen for the validation of the proteomics dataset. The mutants were obtained from Nebraska transposon (Tn) mutant library. The table indicates the gene name which is disrupted by Tn, respective proteins and biological process in which the protein is annotated

Cluster Numer	Gene name	Nebraska Accession	Protein names	Gene Ontology (biological process)
C1	SAUSA300_1227	NR-47429	Aspartokinase (EC 2.7.2.4)	lysine biosynthetic process via diaminopimelate [GO:0009089]; phosphorylation [GO:0016310]; threonine biosynthetic process [GO:0009088]
C1	SAUSA300_1715	NR-48198	Carboxyltransferase domain-containing protein	
C1	SAUSA300_2539	NR-46548	Staphylococcal accessory regulator	
C1	SAUSA300_1226	NR-47092	Aldehyde-alcohol dehydrogenase	alcohol metabolic process [GO:0006066]; carbon utilization [GO:0015976]
C1	SAUSA300_1992	NR-48074	Uncharacterized leukocidin-like protein 1	cytolysis in another organism [GO:0051715]
C1	SAUSA300_2127	NR-48158	Uncharacterized leukocidin-like protein 2	cytolysis in another organism [GO:0051715]
C1	SAUSA300_1669	NR-48194	Putative staphylocoagulase	
C1	SAUSA300_1693	NR-47294	Respiratory nitrate reductase, delta subunit	chaperone-mediated protein complex assembly [GO:0051131]; nitrate assimilation [GO:0042128]
C1	SAUSA300_0594	NR-47924	Amino acid permease	
C1	SAUSA300_0151	NR-46657	Aminotransferase	
C2	SAUSA300_0318	NR-47580	Alcohol dehydrogenase (ADH) (EC 1.1.1.1)	
C2	SAUSA300_2384	NR-46909	L-threonine dehydratase catabolic TdcB (EC 4.3.1.19) (Threonine deaminase)	L-threonine catabolic process to propionate [GO:0070689]
C2	SAUSA300_0671	NR-46907	Nitrite reductase [NAD(P)H], large subunit (EC 1.7.1.4)	nitrate assimilation [GO:0042128]
C2	SAUSA300_0262	NR-47620	nitric oxide dioxygenase (EC 1.14.12.17)	
C2	SAUSA300_2068	NR-46904	Aminotransferase, class V (EC 1.12.-.-)	
C2	SAUSA300_1799	NR-47596	Prephenate dehydrogenase (PDH) (EC 1.3.1.12)	tyrosine biosynthetic process [GO:0006571]
C2	SAUSA300_0196	NR-47210	Chemotaxis inhibitory protein (CHIPS)	

C2	SAUSA300_1260	NR-47816	UPF0340 protein SAUSA300_2068	
C2	SAUSA300_0234	NR-48286	Riboflavin biosynthesis protein RibD	riboflavin biosynthetic process [GO:0009231]
C2	SAUSA300_2247	NR-46753	histidine kinase (EC 2.7.13.3)	
C2	SAUSA300_0177	NR-47290	Homoserine dehydrogenase (EC 1.1.1.3)	isoleucine biosynthetic process [GO:0009097]; methionine biosynthetic process [GO:0009086]; threonine biosynthetic process [GO:0009088]
C2	SAUSA300_2341	NR-48054	nitrate reductase (quinone) (EC 1.7.5.1)	nitrate metabolic process [GO:0042126]
C2	SAUSA300_1220	NR-46805	Putative transporter	p-aminobenzoyl-glutamate transmembrane transport [GO:1902604]
C2	SAUSA300_1225	NR-47931	Accessory gene regulator protein A	
C2	SAUSA300_1920	NR-48258	Metallo-beta-lactamase domain-containing protein	
C2	SAUSA300_2417	NR-46973	Threonine synthase (EC 4.2.3.1)	threonine biosynthetic process [GO:0009088]
C2	SAUSA300_1378	NR-47827	ABC transporter, ATP-binding protein, MsbA family	
C2	SAUSA300_1566	NR-48401	Type I restriction enzyme endonuclease subunit (R protein) (EC 3.1.21.3)	DNA restriction-modification system [GO:0009307]
C2	SAUSA300_1389	NR-46992	Multidrug resistance efflux pump SepA	
C2	SAUSA300_0773	NR-47724	Ribokinase (RK) (EC 2.7.1.15)	D-ribose catabolic process [GO:0019303]
C2	SAUSA300_1974	NR-47928	DNA-binding response regulator, LuxR family	phosphorelay signal transduction system [GO:0000160]; regulation of DNA-templated transcription [GO:0006355]
C2	SAUSA300_2346	NR-47400	Putative Na+/H+ antiporter	
C2	SAUSA300_2343	NR-46869	Lipoprotein	
C2	SAUSA300_1329	NR-46754	Putative N-acetylmannosamine-6-phosphate 2-epimerase (EC 5.1.3.9)	carbohydrate metabolic process [GO:0005975]; N-acetylmannosamine metabolic process [GO:0006051]; N-acetylneuraminate catabolic process [GO:0019262]
C2	SAUSA300_1975	NR-47843	Acyl-CoA dehydrogenase	

C2	SAUSA300_1330	NR-47058	PhiSLT ORF636-like protein	
C3	SAUSA300_0839	NR-48337	Oligopeptide ABC transporter, permease protein	nickel cation transport [GO:0015675]; transmembrane transport [GO:0055085]
C3	SAUSA300_1236	NR-46861	Siderophore biosynthesis protein, IucC family	siderophore biosynthetic process [GO:0019290]
C3	SAUSA300_0633	NR-46949	Putative phage-related DNA recombination protein	DNA metabolic process [GO:0006259]
C3	SAUSA300_1570	NR-47603	DUF4242 domain-containing protein	
C3	SAUSA300_0720	NR-46812	Luciferase-like domain-containing protein	
C3	SAUSA300_1853	NR-47943	Sucrose operon repressor	regulation of DNA-templated transcription [GO:0006355]
C3	SAUSA300_1633	NR-47885	Peptidase, U32 family (EC 3.4.-.-)	organic substance metabolic process [GO:0071704]
C3	SAUSA300_1995	NR-47285	Transcriptional regulator	
C3	SAUSA300_1797	NR-47987	Bacillithiol transferase BstA	
C3	SAUSA300_2622	NR-47537	Glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.-)	glucose metabolic process [GO:0006006]
C3	SAUSA300_2532	NR-47200	Iron-regulated surface determinant protein F	
C3	SAUSA300_0918	NR-48205	CAP domain-containing protein	
C4	SAUSA300_2626	NR-46791	Ornithine cyclodeaminase (EC 4.3.1.12)	siderophore biosynthetic process [GO:0019290]
C4	SAUSA300_0120	NR-47574	Putative iron compound ABC transporter, ATP-binding protein	
C4	SAUSA300_0434	NR-46603	Siderophore biosynthesis protein, IucA/IucC family	siderophore biosynthetic process [GO:0019290]
C4	SAUSA300_0119	NR-48168	Ferredoxin--NADP reductase (FNR) (Fd-NADP(+) reductase) (EC 1.18.1.2)	
C4	SAUSA300_0123	NR-46757	NIF system FeS cluster assembly NifU C-terminal domain-containing protein	iron-sulfur cluster assembly [GO:0016226]

C4	SAUSA300_0339	NR-47306	HPCH/HPAI aldolase family protein	
C4	SAUSA300_1960	NR-48372	Cystathionine gamma-synthase (EC 4.4.1.8)	transsulfuration [GO:0019346]
C4	SAUSA300_0124	NR-47465	Ferrichrome transport ATP-binding protein fhuA	
C4	SAUSA300_0122	NR-46728	Iron-regulated surface determinant protein E	heme transport [GO:0015886]
C4	SAUSA300_1716	NR-47126	Aspartate 1-decarboxylase (EC 4.1.1.11)	alanine biosynthetic process [GO:0006523]; pantothenate biosynthetic process [GO:0015940]
C4	SAUSA300_2137	NR-46922	Uncharacterized protein	
C4	SAUSA300_2409	NR-47427	Alanine racemase N-terminal domain-containing protein	
C4	SAUSA300_0173	NR-47554	FAD-dependent urate hydroxylase	
C4	SAUSA300_2319	NR-47706	Processive diacylglycerol beta-glucosyltransferase (EC 2.4.1.315)	enterobacterial common antigen biosynthetic process [GO:0009246]; glycolipid biosynthetic process [GO:0009247]; lipoteichoic acid biosynthetic process [GO:0070395]
C4	SAUSA300_1033	NR-47715	Siderophore biosynthesis protein, IucC family	siderophore biosynthetic process [GO:0019290]
C4	SAUSA300_1032	NR-48350	tRNA uridine(34) hydroxylase (EC 1.14.-.-) (tRNA hydroxylation protein O)	tRNA modification [GO:0006400]

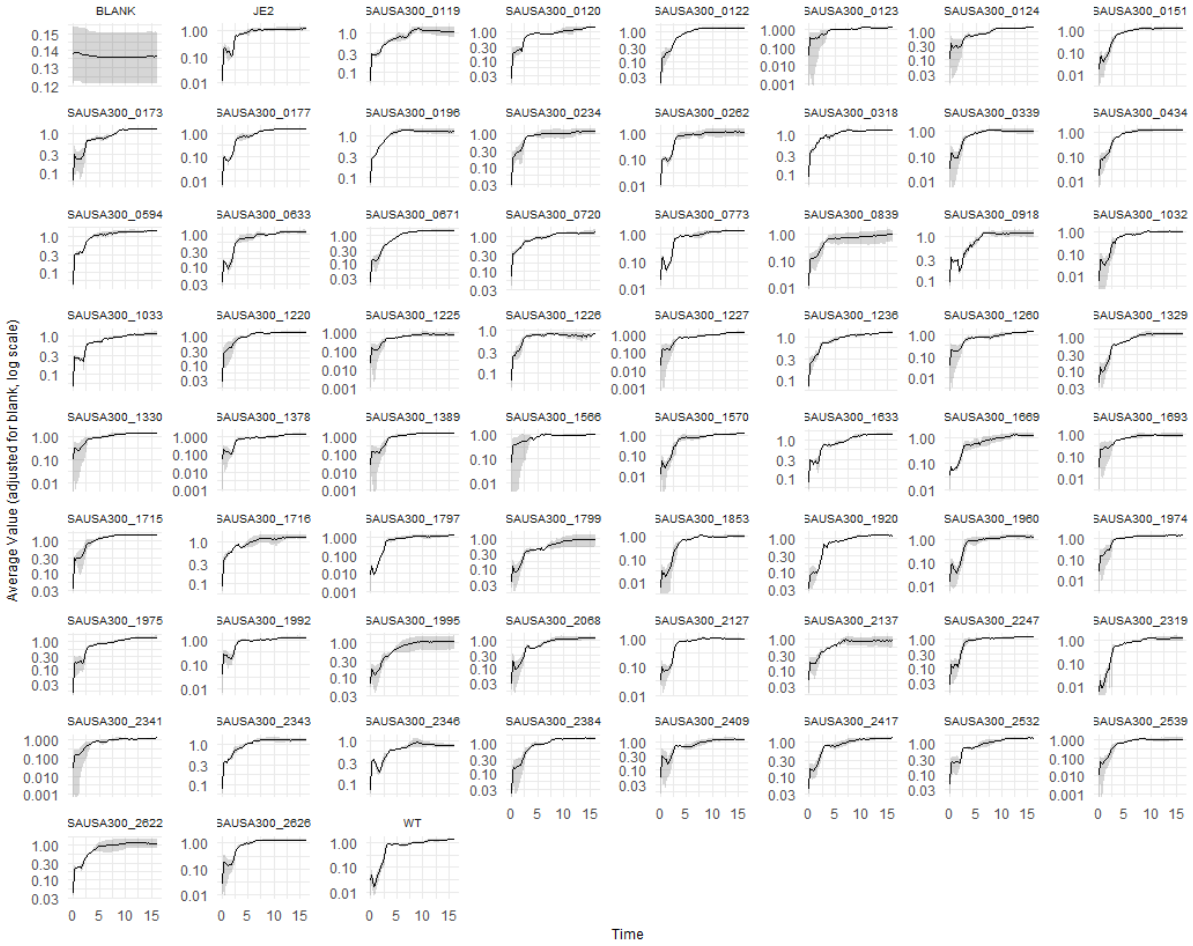


Figure S3.3: Growth curves of the selected mutants grown in DMEM/F12 media in a 96 well plate and monitored in a plate reader instrument.

This chapter is supported by additional datasets which can be obtained through link:

https://sciencedata.dk/shared/a86a5cd107a014f365dee6664b4f3009?download

Supplementary_Dataset_3.1.xlsx: Protein table with abundance values for all identified proteins. Result of the Proteome Discoverer search

Supplementary_Dataset_3.2.xlsx: Results of differential expression analysis for all quantified proteins including annotation obtained from Uniprot and AureoWiki.

Supplementary_Dataset_3.3.xlsx: Results of differential expression analysis of T1-T4 versus control

Supplementary_Dataset_3.4.xlsx: Results of the mfuzz soft clustering, the complete list of proteins matched to clusters 1-4

Infection_4_time_points.Rmd: R script utilized for the analysis of the dataset in this chapter/

Chapter 4

Mass spectrometry-based analysis of surface proteins in *Staphylococcus aureus* clinical strains: Identification of promising peptide targets for diagnostics

Ema Svetlicic, Leonarda Acha Alarcon, Roger Karlsson, Carsten Jers, Ivan Mijakovic

Abstract:

Surface proteins of Gram-positive bacteria are critical for adherence to host tissues, evasion of the immune system, and interaction with the environment. They can be utilized as biomarkers in diagnostics, vaccine development, and as therapeutic targets due to their accessibility and role in pathogenicity. If utilized as diagnostic targets, surface biomarkers should be highly conserved across different strains of the pathogen, unique to the species to avoid cross-reactivity, abundantly expressed on the bacterial surface, and accessible to antibodies or detection reagents. Mass spectrometry-based proteomics methods advanced the studies of surface proteins, often in combination with selective enrichment strategies such as tryptic "shaving. In this study, 11 clinical strains of *Staphylococcus aureus* underwent tryptic shaving to identify common surface proteins. Further bioinformatics analysis confirmed that these proteins are encoded in the core genome of numerous *S. aureus* strains, contained species-specific peptides, and verified structural accessibility. The analysis identified 26 peptides in 15 proteins as potential clinical application targets.

4 Mass spectrometry-based analysis of surface proteins in *Staphylococcus aureus* clinical strains: Identification of promising peptide targets for diagnostics

4.1 Introduction

Staphylococcus aureus is a Gram-positive bacterium which can cause invasive, potentially life-threatening, infections such as sepsis, endocarditis and pneumonia. The treatment of invasive infections has been reported as extremely challenging due to the pathogen´s evasion of the host immune system in combination with growing antimicrobial resistance (1). The mechanism involves a wide range of protein virulence factors, many of which are either secreted into the surrounding environment or attached to the cell wall in which case they are referred to as surface proteins (2). In general, the surfaces of Gram-positive bacteria display a diverse set of biomolecules responsible for interaction with the surroundings, metabolite and nutrient transport, attachment to surfaces or other cells. While these surface biomolecules are versatile and include a thick peptidoglycan layer, capsular polysaccharides and phospholipids, surface proteins or more specifically cell wall proteins have been particularly important in potential clinical applications due to their availability, structural and functional diversity as well as antigenicity properties (3,4). Therefore, these proteins can be utilized as targets for detection of pathogens from clinical and food samples (5–8) and as targets for vaccine development (3,9,10). The known surface proteins of Gram-positive bacteria are usually characterized by signal sequences such as LPXTG and transmembrane domains consisting of hydrophobic stretches. Identification of surface-associated proteins and their exposed epitopes represents a major challenge due to their low abundance and hydrophobicity (11). The mass spectrometry based proteomics techniques, greatly improved their identification through the application of specific sample preparation protocols such as biotinylation and tryptic “shaving” (12). The biotinylation approach, which uses nonpermeable biotin molecule to tag surface proteins of intact cells enables identification of surface proteins after the cells are lysed and proteins tagged with biotin are retained on the streptavidin column. The method has been applied in Gram-negative pathogen, *Helicobacter pylori* for identification of vaccine candidates (13) or to elucidate proteins that mediate bacterium-host interactions (14). Recently it was also used for the elucidation of *Vibrio cholerae* surface proteins expressed in infection *in vivo* (15). Furthermore, in Gram-positive bacteria the biotinylation approach was used for several applications such as identification of vaccine candidates in *Enterococcus faecium* (16), for study of lipoprotein processing in *Streptococcus pneumoniae* (17) and investigation of differential surface protein expression of *S. aureus* under iron limiting conditions (18). However, it was observed in the latter study that biotin exhibits limited affinity for surface

proteins attached via a sortase, resulting in decreased detection sensitivity for these relevant protein antigens. Tryptic shaving or commonly referred to as surface shaving is a process in which live bacteria are incubated with the protease trypsin which results in digestion of ideally only surface exposed proteins and release of their peptides (19). The tryptic peptides are subsequently separated from the cells by centrifugation or filtration and analyzed by mass spectrometry. The method is relatively easy to perform and can also provide information of protein topology since only exposed residues can be accessed by trypsin. However, the drawback of the method is concerns about cell lysis during sample preparation that leads to release of the cytoplasmic content resulting in proteins falsely being identified as extracellular (19). More notably, surface shaving was more successful in Gram-positive bacteria due to thicker peptidoglycan layer and higher resistance to lysis (13). To avoid the loss of membrane integrity, several strategies have been developed to achieve a controlled digestion using either immobilized trypsin, where the cell suspension is passed through a column containing trypsin (20) or by cell immobilization in which case the trypsin solution flows through a channel with cells attached (21,22). In addition to mass spectrometry-based identification of surface proteins, various prediction algorithms have been developed for the recognition of the sequence patterns in the proteome which help to determine the subcellular locations of bacterial proteins. To name a few, pSort (23) and CELLO (24) utilize sequence information and training datasets to categorize localization of proteins, TMHMM (25) detects transmembrane domains in the sequence, SignalP detects signal sequences which indicate a secretion (26), while tools such as SecretomeP (27) predict non-classical protein secretion. Extensive explanation and overview of subcellular localization tools can be found elsewhere (28). Several studies, however, reported differences between predicted and observed localization, most notably in the proteins predicted as cytoplasmic, which were observed extracellularly possibly due to non-classical excretion, translocation mechanisms and moonlighting properties (29). Therefore, the prediction algorithms should be used interchangeably with experimental methods and proteins predicted as cytoplasmic should not be automatically disregarded as potential diagnostic or vaccine targets. For a good biomarker it is important that the target proteins are conserved within the species and not homologous to host proteins or other microorganism proteins contained in the sample. Rather than using alignment methods, which can be computationally consuming, new approaches based on short substrings of DNA or peptides sequences called k-mers were developed in the last decade. By analyzing the frequency and distribution of k-mers across genomes or metagenomic datasets, researchers were able to develop tools such as Kraken (30) and CLARK (31) which are now widely used for taxonomic classification of

sequenced reads. A recently published paper by paper by Mouratidis et al. .identified the shortest set of peptide kmers that are found in only one species and are absent in every other assembled reference proteomes in RefSeq, termed quasi-prime peptides (32). The study identified quasiprimes for several pathogens, including *S. aureus* and highlighted the potential utilization as biomarkers. The peptides were made available through a new database KmerDB (33) based on the research by Mouratidis et al. Peptides represent shorter, defined sequences compared to full proteins, which can contain multiple epitopes. Raising antibodies against peptides allows for more precise targeting of specific linear epitopes within the protein sequence, minimizing the risk of cross-reactivity with other regions of the protein (34). Through the utilization of tryptic shaving, pangenomics, KmerDB, along with structural availability, several peptide candidates as potential targets for diagnostics and were identified. The schematic pipeline of this study is depicted in Figure 4.1.

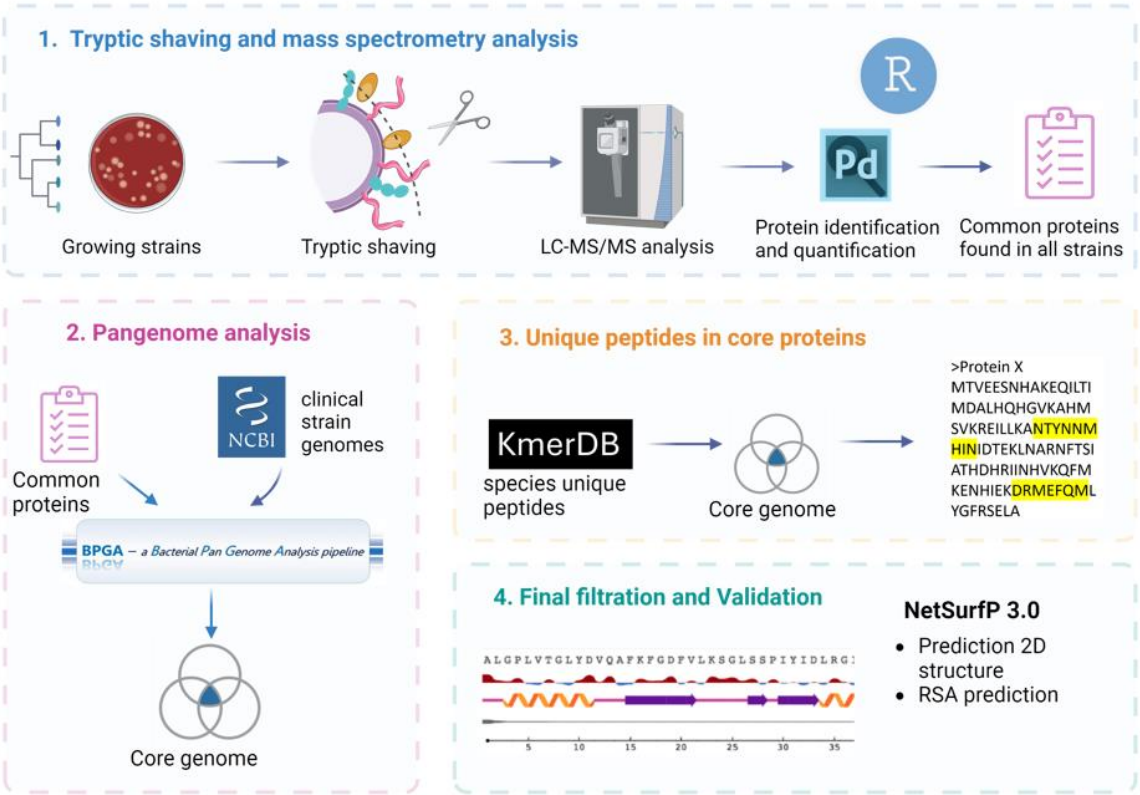


Figure 4.1: Biomarker discovery pipeline consisting of four main phases. The first phase is generation of surface exposed peptides using tryptic shaving of several evolutionary distinct bacterial strains. The peptides are analyzed by LC-MS/MS and identified by proteomic specialized software. Phase 1 concludes with filtering the strains to obtain only proteins common to all analyzed strains. Phase 2 is further filtering the list of common proteins to

include only proteins found in the majority of clinical strains. This is achieved by aligning sequences of common proteins to the core genome of NCBI derived strains. Phase 3 includes matching of the unique peptides of the bacterial species (quasiprime peptides) to the sequences of the core proteins identifying proteins which contain them. Phase 4 consists of determining the 2D structure of target proteins and filtering sequences of quasiprime peptides which are determined to be exposed on the surface.

4.2 Materials and methods

4.2.1 Bacterial strains and cultivation

Bacterial strains were obtained from the Culture Collection University of Gothenburg (CCUG) (www.ccug.se): *S. aureus* strains CCUG 10778, CCUG 15915, CCUG 1799, CCUG 1800T, CCUG 1801, CCUG 2353, CCUG 2354, CCUG 259252, CCUG 41586, CCUG 41879 and CCUG 54581, were cultivated on Blood Agar medium made of Columbia Agar Base plus 5% horse blood at 37˚C, with 5% CO_2, overnight (18 hours). The biomass was collected and suspended in phosphate-buffered saline (PBS). Bacterial cell suspension optical densities (OD) were measured at a wavelength of 600 nm and adjusted in 1 ml PBS to an OD_{600} of 1.00 (10^9 cfu/ml). The biomass was washed with PBS three times and resuspended in 150 µl of PBS.

4.2.2 Surface shaving of the selected strains

Cell suspension in a volume of 150 µl was slowly injected into the LPI™ HexaLane Flow Cell (Nanoxis Consulting AB, Sweden) using a pipette to fill the FlowCell channel. The Flow Cell was incubated for 30 min at room temperature to allow for the immobilization of cells to the channel surface. The FlowCell channels were washed with 200 µl of PBS to remove unbound bacteria. Enzymatic digestion of the proteins was performed by injecting trypsin (20 µg/ml in 20 mM PBS, pH 8.0) into the FlowCell channels and incubating for 20 min at room temperature. The generated peptides were eluted by injecting 100 µl PBS buffer (pH 8.0) into the channels. The eluted peptides were collected at the outlet ports and peptide solutions were incubated at room temperature overnight at 37 ºC with the addition of 1 µg of trypsin. The peptides were desalted using SOLAµ cartridges as following. The columns were primed with 100% methanol, equilibrated with 80% acetonitrile/0.1% TFA and washed two times with 3% acetonitrile/0.1% formic acid/. The peptides were loaded and washed two times with 3% acetonitrile/0.1% formic acid. Finally, peptides were eluted with 60% acetonitrile/0.1% formic acid-. dried and subsequently frozen at −20˚C until analysis by MS.

4.2.3 LC-MS/MS analysis

For each sample, peptides were loaded onto a 2 cm C18 trap column (ThermoFisher 164946), connected in-line to a 15 cm C18 reverse-phase analytical column (Thermo EasySpray ES904) using 100% Buffer A (0.1% FA in water) at 750 bar, using the Thermo EasyLC 1200 HPLC system, and the column oven operating at 45 °C. Peptides were eluted over a 60 min gradient ranging from 10 to 60% of 80% ACN, 0.1% FA at 250 nl/min, and the Exploris 480 instrument (Thermo Fisher Scientific) was run in a DDA-MS2 top10 method. Full MS spectra were collected at a resolution of 70,000, with an AGC target of 3×10^6 or maximum injection time of 20 ms and a scan range of 300–1750 m/z. The MS2 spectra were obtained at a resolution of 17,500, with an AGC target value of 1×10^6 or maximum injection time of 60 ms, a normalized collision energy of 25 and an intensity threshold of 1.7e4. Dynamic exclusion was set to 60 s, and ions with a charge state < 2 or unknown were excluded.

4.2.4 Proteomics downstream analysis

The raw files were analyzed using Proteome Discoverer 2.5. The spectra were matched against the Uniprot database of *S. aureus* strain USA300 (UP000001939). The strain is a community associated methicillin-resistant *S. aureus* frequently isolated from patients, thus it was chosen as a reference proteome. Dynamic modifications were set as Oxidation (M) and Acetyl on protein N-termini. All results were filtered to a 1% FDR. The peptide table was further processed in R Studio using a custom-made script employing the packages tidyr, dplyr, ggplot2, corrplot, reshape2. The processing included calculating iBAQ values for each protein using the formula: (Sum of peptide intensity Intensity)/(Theoretical number of peptides). The resulting protein table was then filtered to exclude proteins not quantified in any of the samples. Moreover, if a protein appeared in only one out of three replicates, it was filtered out. Finally, proteins that were detected in all 11 strains were kept as well as the proteins represented with more than 2 peptide to spectrum matches (PSMs).

4.2.5 Cellular localization prediction

The FASTA sequence of the proteome of *S. aureus* USA 300 (UniProt ID UP000001939) was searched with Secretome P1.0, pSort v3.0.2 and SignalP 4.0 web tools with settings for Gram-positive bacteria.

4.2.6 Pangenome analysis

Genome sequences of *S. aureus* strains were downloaded from National Center for Biotechnology Information (NCBI) database. The isolates were filtered using Pathogen

Detection browser (https://www.ncbi.nlm.nih.gov/pathogens/isolates) in order to include complete genomes with contig which originate from clinical strains isolated from *human source* a in the last ten years. The filtration resulted in 490 genomes which were re-annotated using Prokka v1.14.5 (35). Pangenome analysis was conducted using Bacterial Pangenome Analysis Pipeline (BGPA) (36) . Genes were considered core genes if they were detected in all strains with 70% homology. Core genome FASTA was used as database for matching protein sequences from tryptic shaving. The matching was performed using standalone blast+ (37) and proteins found in tryptic shaving were considered part of the core proteome if they exhibited >95% sequence similarity.

4.2.7 Unique sequences in the core genome

The list of quasiprime peptides was obtained from a database kmerDB (https://pavlopoulos-lab.org/kMerDB/) which contains species-specific heptamer peptides termed quasiprime peptides. Heptamer strings were matched to the common core protein FASTA file obtained in pangenome analysis and surface shaving experiments.

4.2.8 Surface accessibility of quasi-prime peptides

To assess whether the unique peptides are accessible to the potential antibodies, a relative solvent accessibility (RSA) and secondary structure of the target proteins were predicted using NetSurfP - 3.0 (38). The aim was to identify the quasi-prime peptides that have seven continuous surface exposed amino acid residues (RSA >20%).

4.3 Results

4.3.1 Tryptic shaving experiments identify 388 proteins common proteins in 11 strains of *S. aureus*

A detailed description of the 11 strains is present in Table S4.1. The strains were subjected to surface shaving with immobilized cells using the Lipid-based Protein Immobilization (LPI) technology as described before (21). In addition, this study employed semi-quantitative analysis using the iBAQ (Intensity-Based Absolute Quantification) method. The named quantification approach enables the estimation of protein abundance based on the sum of the intensity of peptides identified in mass spectrometry divided by the number of theoretical peptides of the protein. In total, 873 proteins were identified in at least one of the samples, but the number of identified proteins differed for the various strains (Figure 4.2A). Generally, the biological replicates had a similar number of identified proteins (Figure 4.2A). Moreover, a binary heatmap visualizing present/absent proteins showed clear clustering of biological replicates and variation in identified proteins among different strains, which

clustered into three distinct clades (Figure 4.2B). To determine which proteins are common to all strains, proteins detected in less than two biological replicates were filtered out and finally only proteins quantified in all 11 strains were kept. Furthermore, proteins with only one peptide to spectrum match (PSMs) were removed, resulting in 388 common proteins. Prediction algorithm pSort was employed for prediction of protein localization for proteins in *S. aureus* USA300 proteome, in protein sequences of all proteins identified (N=873) as well as proteins identified in all strains (N= 388) (Figure 4.2C). The highest number of proteins in all categories belonged to cytoplasmic proteins. Even among proteins identified in the surface shaving experiments, the algorithm predicted more than 70% cytoplasmic proteins. Along with pSort, the tools signalP and SecretomeP were employed to identify signal peptides and non-classically secreted peptides, respectively (Figure 4.2D). The analysis showed overlap between proteins predicted as secreted, however there was a substantial disagreement between the tools as 61 proteins termed cytoplasmic by pSort were classified as non-classically secreted proteins by SecretomeP. There were less discrepancies between pSort and SignalP, where signal sequences were predicted in only 5 proteins predicted to be cytoplasmic by pSort. Cell lysis cannot be excluded and probably happened to some extent due to tryptic digestion of membrane proteins. The proteins labeled as cytoplasmic were not excluded from further analysis as they were identified in all strains and could potentially be expressed on the surface. In addition to qualitative assessment of surface proteins in the analysed strains, iBAQ was employed to estimate a semi-absolute amount of protein within the sample. The distribution of protein amounts (iBAQ values) across all samples is shown in boxplot in Figure 4.3A. Next, an assessment of the most abundant proteins was made by calculating the mean of the abundances through all groups. The ten most abundant proteins based on total iBAQ mean across all strains were: Immunoglobulin G-binding protein A (Spa), Acyl carrier protein (AcP), 30S ribosomal proteins (RpsC, RpsL RpsE), UPF0337 domain containing protein (SAUSA300_0816), Secretory antigen (SsaA), Phage protein (SAUSA300_1904), UPF0478 protein (SAUSA300_1685) and Elongation factor Tu (Tuf) (Figure 4.3B).

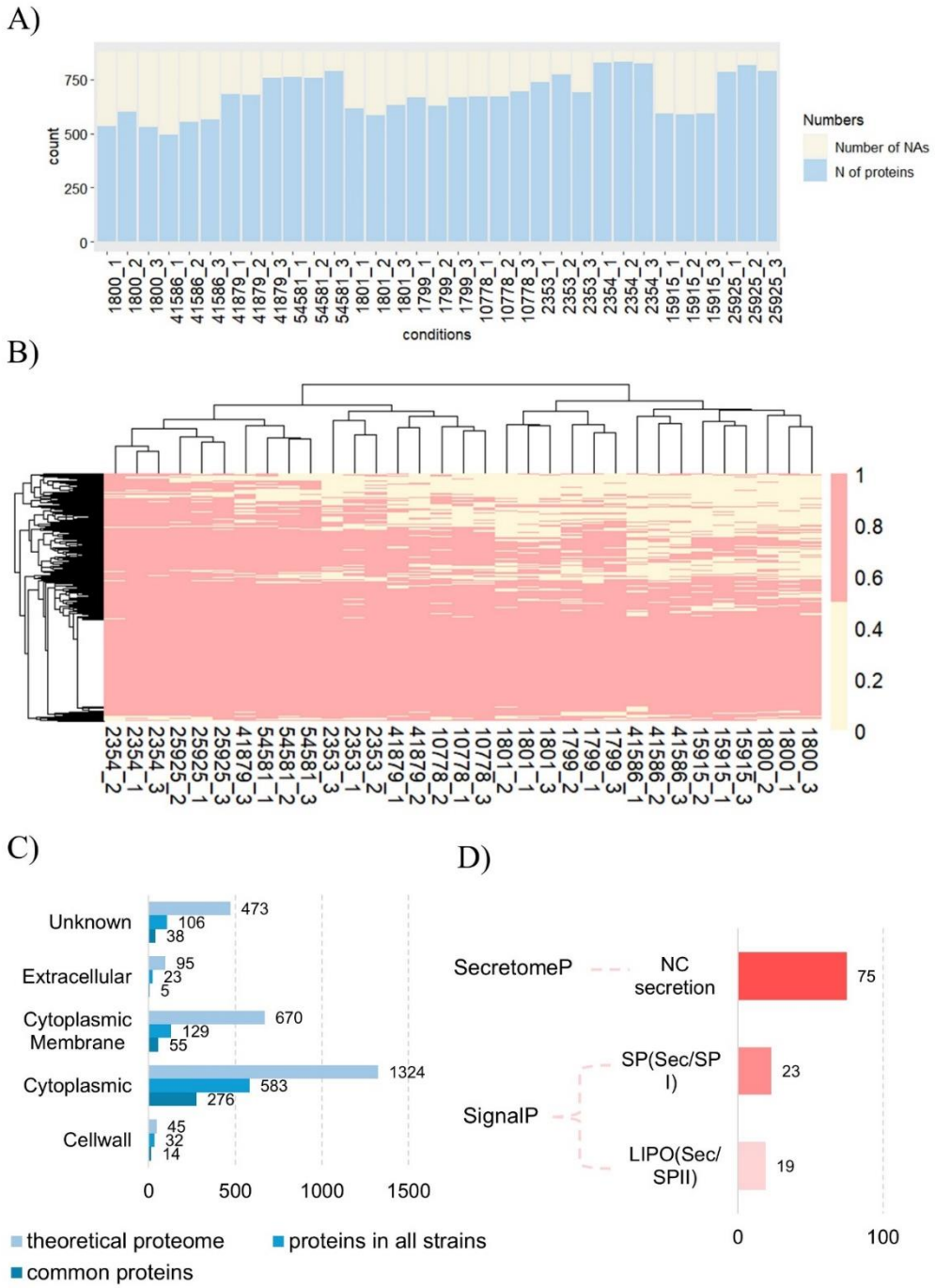


Figure 4.2: Results of the tryptic shaving experiment of 11 *Staphylococcus aureus* strains analyzed in biological triplicates. A) Identified proteins across all samples. B) Two-way hierarchical clustering heatmap visualizing present/absent proteins across samples. C) Predicted localization of proteins in theoretical *S. aureus* USA300 proteome, of proteins identified in all strains in the tryptic shaving experiment and the subgroup of these proteins found in all strains. D) Predicted localization of common proteins by SignalP and SecretomeP.

NC- Non-classical, Sec/SPI: "standard" secretory signal peptides transported by the Sec translocon and cleaved by Signal Peptidase I, LIPO (Sec/SPII): Lipoprotein signal peptide

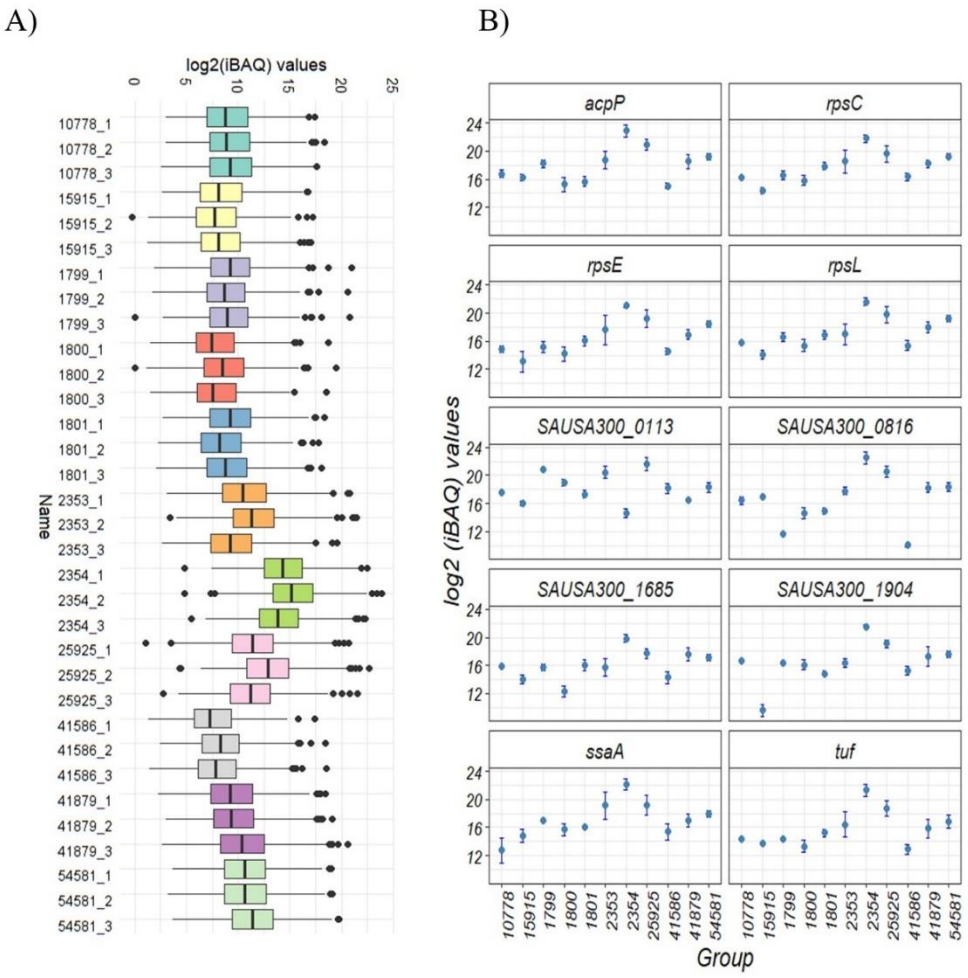


Figure 4.3*:* Quantitative overview of proteins identified by tryptic shaving across samples A) Boxplot depicting log2 (iBAQ) values across all samples B) log2 (iBAQ) intensities of the 10 most abundant proteins (corresponding gene names are shown) depicted as mean values with standard deviation in the various strains.

4.3.2 Identification of surface proteins conserved in *S. aureus* full genome sequences

To verify conservation in a greater number of *S. aureus* strains, proteins identified by tryptic shaving in all strains were then matched against a generated core genome of several hundred *S. aureus* clinical strains whose genomes were obtained from (NCBI). The strains

included in the analysis yielded a pangenome size of 5363 proteins. The BPGA pipeline used in the analysis classifies proteins into core proteins which are present in all strains, unique proteins which are present in only one strain and the remaining proteins are categorized as accessory proteins. The rarefaction curve of core proteins shows that the core proteome, after an initial sharp drop, plateaued and exhibited a minimal decrease upon addition of more genomes (Figure 4.4A). The core proteome consisted of 1806 proteins, the accessory proteome had 2862 proteins and there were 695 unique proteins (Figure 4.4B). The common proteins from surface shaving experiment were matched against the core proteome to find the overlap between the experimental data and a broader core proteome dataset. Of the experimentally identified 388 common proteins, there were 346 proteins with above 95% similarity to proteins in the core genome. These 346 proteins were termed common core proteins.

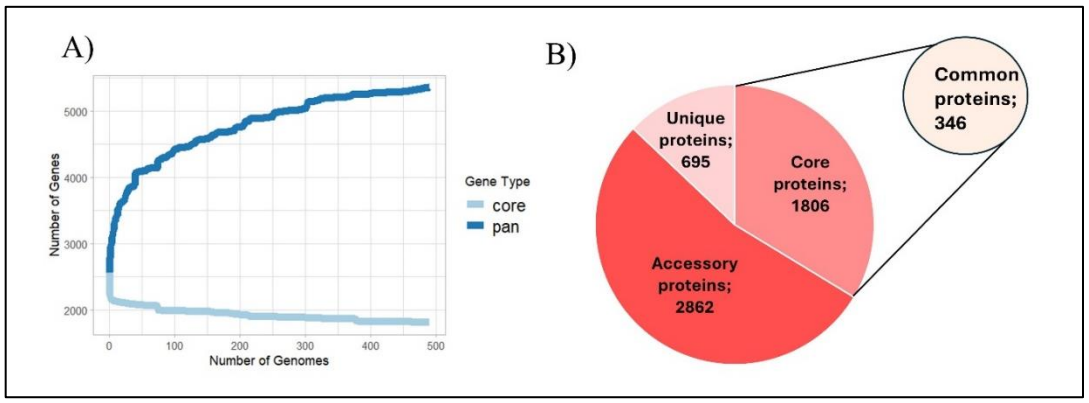


Figure 4.4: Pangenome analysis of clinical strains of *S. aureus* whose proteome sequence was obtained from NCBI Pathogen Detection resources. A) Rarefaction curve showing plateauing of the core genome. B) Results of the pangenome analysis in terms of number of core, accessory and unique genes, where core genes represent the genes found in all tested strains.

4.3.3 Quasi-prime peptide mapping identifies 26 peptide targets in 15 proteins

Potential diagnostic biomarkers should be unique to the targeted species and not present in other species. Previously mentioned, KmerDB which consists of more than 5000 quasi-prime heptamer peptide sequences unique to *S. aureus* (quasi-prime peptides). The sequences were downloaded, and string-matched to the fasta sequences of the 346 common core proteins. In total, 285 heptamers were matched to 132 proteins in the common core proteins. (Supplementary_dataset_4.4.xlsx). The protein surface exposure measured by RSA is a critical feature of biomarkers, because the receptor or antibody should be able to access the residues. The RSA was calculated for all 132 proteins using a sequence-based tool NetSurfP 3 and the

RSA values of all residues in the quasiprime peptides were plotted to observe the distribution of exposed residues (Figure 4.5A). Figure 4.5B shows the distribution of RSA values in quasiprime peptides found in the core proteins. Around 40% of residues had RSA <25% and were termed buried, however most residues pass the threshold and are predicted to be exposed (60%). The residues having RSA<0.25 were filtered out from the dataset. The data was further filtered to include only peptides with seven consecutive surface exposed residues. This criterion greatly narrowed down the possible biomarker candidates, as only 26 surface exposed peptides in 15 proteins remained which are shown in Table 4.1. To determine the abundance of the 15 target proteins within the tryptic shaving samples, the average log2(iBAQ) values were plotted for each strain (Figure 4.5C). On average, the most abundant protein biomarker candidates within the samples seemed to be Phage protein, Alkaline shock protein 23, Immunoglobulin-binding protein Sbi, UPF0478 protein and Aerobic glycerol-3-phosphate dehydrogenase.

Table 4.1: Quasi-prime peptides, comprised of seven unique amino acids (K-mers) specific to *S. aureus*, identified within the core proteome established in this study. The proteins are represented by their common name, gene locus and Uniprot Accession number.

Accession	Gene name/ locus	Protein name	K-mers	Start	End
A0A0H2XEA7	spsB	Signal peptidase I (EC 3.4.21.89)	KHNFNPE	145	151
A0A0H2XEU5	SAUSA300_0508	Excinuclease ABC subunit B	CAEGHHP	6	12
			HHPWNQA	10	16
			MVCQTCA	1	7
A0A0H2XG38	SAUSA300_1904	Phage protein	CVFKFVF	7	13
			FIKCVFK	4	10
			MWNFIKC	1	7
			WNFIKCV	2	8
A0A0H2XH40	SAUSA300_1304	Uncharacterized protein	ETEMTMP	204	210
A0A0H2XHD3	SAUSA300_1729	NERD domain-containing protein	QEDYNHM	73	79
A0A0H2XHM5	ftsK	DNA translocase FtsK	NIVNHHQ	241	247
A0A0H2XHQ1	ftsA	Cell division protein FtsA	HEHVQDK	446	452
			HQEHKQN	439	445
A0A0H2XJD9	SAUSA300_1003	Uncharacterized protein	MVAPQYY	155	161
A0A0H2XJZ5	pbpA	Penicillin-binding protein 1	LRISYIM	31	37
Q2FDP9	betA	Oxygen-dependent choline dehydrogenase (CDH)	YIDYYKH	546	552
			YKHGVHD	550	556
Q2FE79	sbi	Immunoglobulin-binding protein Sbi	DNKAPHD	166	172
			QDNKAPH	165	171
Q2FEV0	asp23	Alkaline shock protein 23	NQEPQFK	33	39
			VMTQKEW	144	150
Q2FFZ9	SAUSA300_1685	UPF0478 protein SAUSA300_1685	DKWQNRH	130	136
			KWQNRHY	131	137
Q2FH99	katA	Catalase (EC 1.11.1.6)	YNQRQDD	417	423
Q2FHD8	glpD	Aerobic glycerol-3-phosphate dehydrogenase (EC 1.1.5.3)	AQHGNNQ	546	552
			QTSQYHD	463	469

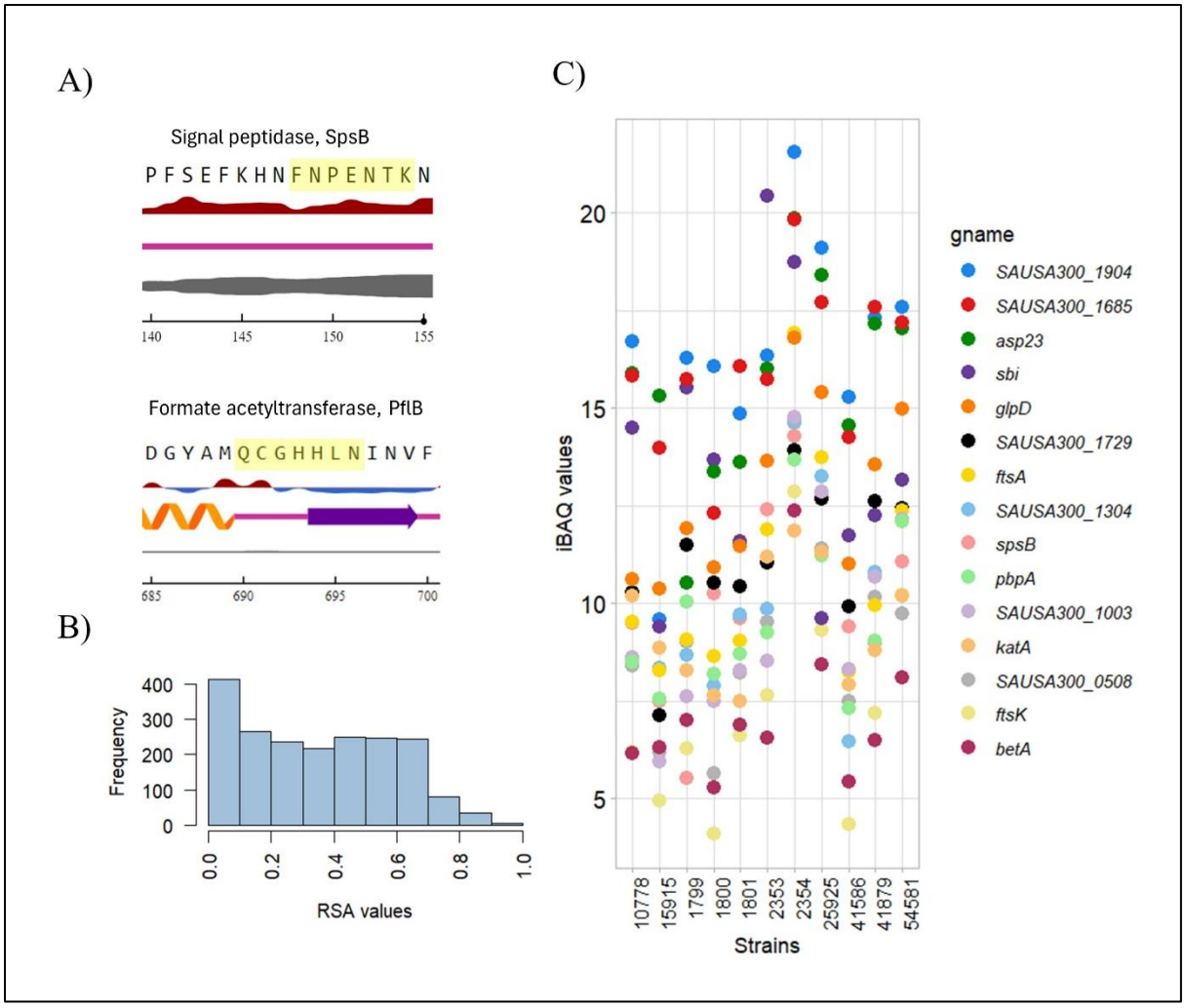


Figure 4.5: Relative surface accessibility (RSA) of quasiprime peptides. A) Results of the NetSurfP 3 for the representative two proteins. The first line shows a protein sequence with the quasiprime peptide sequence highlighted in yellow. The second line represents the RSA values, which can be blue or red, where red indicates an exposed residue and blue buried. The third line shows the predicted secondary structure of the protein region, either a coil in pink, a helix in orange or beta-sheet in blue. Finally the fourth line shows probability of the region being disordered, where the thicker line indicates higher probability. B) Histogram of RSA distributions across quasiprime peptides found in the core proteins C) Protein abundance measured as log2 (iBAQ) values of the proteins that contain quasiprime peptides with 7 consecutive exposed residues.

4.4 Discussion and conclusions

The present study sets out to identify surface proteins which are conserved in and unique to *S. aureus* as potential biomarkers for diagnostics. The number of identified proteins is quite higher than in the previous research of *S. aureus* surfaceome (20,39), which might be attributed

to differences in growth conditions, cell lysis, mass spectrometry analysis and proteomics software used. Nevertheless, the results are comparable as the study also shows heterogeneous protein expression between different *S. aureus* strains and more importantly an overlap between identified proteins. The proteins overlapping between the two studies among others include probable malate:quinone oxidoreductase, 30S and 50S ribosomal proteins, fructose-bisphosphate aldolase FbA, pyruvate dehydrogenase complexes E1 (PhdAB), elongation factor Tu (EF-Tu) and G (FusA), enolase (Eno), bifunctional purine biosynthesis protein PurH, pyruvate dehydrogenase PdhAB and glyceraldehyde 3-phosphate dehydrogenase GAPDH. These proteins are historically considered intracellular, however for several of the named proteins moonlighting properties were observed in *S. aureus* as well as other bacteria. A glycolytic enzyme GAPDH was one of the first moonlighting proteins discovered which role was idnetifed to be as scavenging factor for iron when latter is limited in infection, therefor play an important part in pathogenesis (40). Moreover, it was found that EF-Tu moonlights on the surface and binds to host proteins (41). Eno and FbA were also revealed to be exported to the cell surface under various growth conditions in different species (42,43). In fact, excretion of cytoplasmic proteins is a common physiological feature in *S. aureus* as well as other bacteria, nevertheless prediction tools such as pSORT used in this study, localize these proteins only in the cytoplasm. The reason for the discrepancy is as previously noted poor understanding of these mechanisms and lack of any apparent recognition patterns (44). Therefore, the experimental analysis approach could shed light on the surface associated proteins. However, the major limitation of this study is that some degree of cell lysis during sample preparation cannot be excluded, and the portion of proteins identified could be the result of trypsin penetrating to cytoplasmic membrane and causing the lysis. The protein abundance was measured with a semi absolute approach using iBAQ. The most abundant protein was SpA, a well characterized known surface anchored protein which binds to Immunoglobulin G, thus hindering the immune response of the host (45). While it was detected in all samples subjected to tryptic shaving, the protein was not part of the core genome generated *in silico* by pangenome analysis. The SpA was previously found in approximately 90% of isolates (46) and in our study was part of the accessory genome, thus it was excluded from the further analysis. Another highly abundant protein was the acyl carrier protein ACP, a protein involved in fatty acid and polyketide synthesis, not previously discussed to be surface anchored. Part of the most abundant proteins found in this study were ribosomal proteins, more specifically 30S ribosomal proteins. Although, known as intracellular proteins having a crucial role in protein synthesis, they have been identified in the secretome, associated with extracellular vesicles and identified

in surfacome profiling experiment (47–49). While it is known that many ribosomal proteins are involved in other functions, the moonlighting properties have not been observed on the surface of bacteria and their exact translocation (if any) is not known. (49). The discussed proteins were all part of the 388 common proteins identified in identified in all 11 strains in at least two biological replicates. The common proteins were matched against the core genome generated by the pangenome analysis of several hundred clinical strains of *S. aureus*. This approach enabled the expansion of the study to include the strains from various studies while still including high quality genome assemblies. The core proteome consisted of 1806 proteins corresponding to about 65% of an average *S. aureus* genome, which is a high portion when comparing to other organisms (46). Moreover, the majority of the core genome includes housekeeping proteins which are responsible for transcription, translation, RNA processing and modification [40]. Most of the common proteins identified by tryptic shaving were part of the core genome (89%), however the remaining proteins identified in the surface shaving were filtered out of the further processing as the study focuses on finding new diagnostic markers and it is favored that they are present in all of the strains. Next, the proteins were checked for uniqueness. Instead of using alignment-based algorithms such as BLAST, which can be computationally intensive, this study utilized an already established set of unique peptides (kmers) present in only *S. aureus*. The usage of kmers for taxonomic classification is widely used in annotation of sequence reads in metagenomic research due to their speed and comparable sensitivity (50,51). The unique peptides were detected in the previous study by analyzing proteomes from 21 875 species and identifying the unique kmers of 7 amino acids (32). The distribution of these heptamers was examined across all proteomes and the species-specific kmers called quasi-prime peptides were identified and made available in kmerDB. By matching those quasi-prime peptides to the core proteins identified by tryptic shaving in this study, we identified 132 proteins that contained a total of 285 unique peptides. For the application of these proteins in healthcare such as diagnostics, it is critical that the peptides are exposed on the surface of the protein, in order to be targeted by a receptor or a drug. A commonly used measure of surface exposure is RSA which was calculated for all the 132 proteins. While the majority of residues had RSA values above the threshold (25%) indicating surface exposure, only 15 proteins contained quasi-prime peptides with all seven residues predicted as exposed. Some of these proteins, such as PbpA and Sbi have already been targeted in the vaccine research (52,53). In addition, Sbi was one of the most abundant proteins in all samples confirming its potential in clinical applications. Another highly abundant protein was Asp23, a stress response factor previously identified as a T cell antigen (54) and the

moonlighting protein GlpD. The proteins Phage protein (SAUSA300_1904) and UPF0478 protein (SAUSA300_1685) have not been experimentally characterized so far and not much is known about their properties or function. However,, its uniqueness and surface-exposed localization might make it an interesting target. These biomarkers, being surface-exposed and unique to the pathogen, can be utilized in the creation of rapid assays that can identify the presence of the bacteria in clinical samples, improving early diagnosis and treatment. Characteristics of the ideal diagnostic test, especially in the developing world, have been defined almost two decades ago. The tests should be affordable, sensitive, specific, rapid, equipment-free, and accessible to those who need them (55). Incorporation of biomarkers into rapid diagnostic devices using biorecognition elements such as antibodies, bacteriophages, aptamers, and antimicrobial peptides are continuously developed (56–60), with some already available on the market (61). For the biomarkers to be utilized in diagnostics, an extensive experimental screening of the identified biomarkers should take place to evaluate the biorecognition capabilities of receptors, as well as their sensitivity and specificity to *S. aureus*. Some limitations of this tudy should also be mentioned. While quasi-prime peptides offer short, unique sequences favorable for recognition, sequence variants and sequencing errors can result in false identifications of the unique k-mers (32,32). Therefore, it is recommended to rely on several biomarkers which is shown to improve diagnostic accuracy (62). Moreover, future work could also evaluate species specific peptides in excreted *S. aureus* proteins, especially numerous toxins which this bacterium produces.

In conclusion, this study presents a list of proteins from diverse *S. aureus* clinical strains identified by tryptic shaving and thus presumed to be surface exposed. In addition to identification, the study employed semi-absolute quantification using iBAQ values proven to be accurate in quantitative estimation (63). The data gathered from these analyses was utilized for the selection of potential diagnostic targets by several bioinformatic approaches 1) pangenome analysis to ensure the high coverage of the target proteins among *S. aureus* strains 2) specificity to the species by matching species unique peptides (quasi-prime peptides) 3) Prediction of RSA values for the quasi-prime peptides and selection of fully exposed peptides. The bioinformatic analysis uncovered 26 peptides in 15 proteins which could serve as targets for clinical applications. The results presented in this study can greatly reduce the workload and cost of such screening. Moreover, the pipeline in the analysis can be applied to other pathogen species and thus aid in future strategies to identify diagnostics targets.

4.5 References:

1. Howden BP, Giulieri SG, Wong Fok Lung T, Baines SL, Sharkey LK, Lee JYH, et al. *Staphylococcus aureus* host interactions and adaptation. Nat Rev Microbiol. 2023 Jan 27;1–16.

2. Foster TJ, Geoghegan JA, Ganesh VK, Höök M. Adhesion, invasion and evasion: the many functions of the surface proteins of *Staphylococcus aureus*. Nat Rev Microbiol. 2014 Jan;12(1):49–62.

3. Diaz-Dinamarca DA, Manzo RA, Soto DA, Avendaño-Valenzuela MJ, Bastias DN, Soto PI, et al. Surface Immunogenic Protein of Streptococcus Group B is an Agonist of Toll-Like Receptors 2 and 4 and a Potential Immune Adjuvant. Vaccines. 2020 Jan 16;8(1):29.

4. Dwivedi P, Alam SI, Tomar RS. Secretome, surfome and immunome: emerging approaches for the discovery of new vaccine candidates against bacterial infections. World J Microbiol Biotechnol. 2016 Jul 27;32(9):155.

5. Marcos-Fernández R, Ruiz L, Blanco-Míguez A, Margolles A, Sánchez B. Cell wall hydrolase as a surface-associated protein target for the specific detection of *Lactobacillus rhamnosus* using flow cytometry. Innov Food Sci Emerg Technol. 2020 Jan 1;59:102240.

6. Pyclik M, Górska S, Brzozowska E, Dobrut A, Ciekot J, Gamian A, et al. Epitope Mapping of *Streptococcus agalactiae* Elongation Factor Tu Protein Recognized by Human Sera. Front Microbiol. 2018 Feb 6;9:125.

7. Dobrut A, Wójcik-Grzybek D, Młodzińska A, Pietras-Ożga D, Michalak K, Tabacki A, et al. Detection of immunoreactive proteins of *Escherichia coli, Streptococcus uberis,* and *Streptococcus agalactiae* isolated from cows with diagnosed mastitis. Front Cell Infect Microbiol. 2023;13:987842.

8. Squeglia F, Marasco D, Ruggiero A, Testa G, Esposito L, Berisio R. Structural and Binding Properties of the Active Cell Wall Hydrolase RipA from M. tuberculosis, a Promising Biosensing Molecule for Early Warning Bacterial Detection. Curr Med Chem. 2022;29(24):4282–92.

9. Klimka A, Mertins S, Nicolai AK, Rummler LM, Higgins PG, Günther SD, et al. Epitope-specific immunity against *Staphylococcus aureus* coproporphyrinogen III oxidase. Npj Vaccines. 2021 Jan 18;6(1):1–12.

10. Romero-Saavedra F, Laverde D, Wobser D, Michaux C, Budin-Verneuil A, Bernay B, et al. Identification of Peptidoglycan-Associated Proteins as Vaccine Candidates for Enterococcal Infections. PLOS ONE. 2014 Nov 4;9(11):e111880.

11. Solis N, Cordwell SJ. Current methodologies for proteomics of bacterial surface-exposed and cell envelope proteins. PROTEOMICS. 2011;11(15):3169–89.

12. Pauwels J, Fijałkowska D, Eyckerman S, Gevaert K. Mass spectrometry and the cellular surfaceome. Mass Spectrom Rev. 2022 Sep;41(5):804–41.

13. Hornburg D, Kruse T, Anderl F, Daschkin C, Semper RP, Klar K, et al. A mass spectrometry guided approach for the identification of novel vaccine candidates in gram-negative pathogens. Sci Rep. 2019 Nov 22;9(1):17401.

14. Voss BJ, Gaddy JA, McDonald WH, Cover TL. Analysis of Surface-Exposed Outer Membrane Proteins in Helicobacter pylori. J Bacteriol. 2014 Jul;196(13):2455–71.

15. Zoued A, Zhang H, Zhang T, Giorgio RT, Kuehl CJ, Fakoya B, et al. Proteomic analysis of the host–pathogen interface in experimental cholera. Nat Chem Biol. 2021 Nov;17(11):1199–208.

16. Romero-Saavedra F, Laverde D, Wobser D, Michaux C, Budin-Verneuil A, Bernay B, et al. Identification of Peptidoglycan-Associated Proteins as Vaccine Candidates for Enterococcal Infections. PLoS ONE. 2014 Nov 4;9(11):e111880.

17. Pribyl T, Moche M, Dreisbach A, Bijlsma JJE, Saleh M, Abdullah MR, et al. Influence of Impaired Lipoprotein Biogenesis on Surface and Exoproteome of *Streptococcus pneumoniae*. J Proteome Res. 2014 Feb 7;13(2):650–67.

18. Hempel K, Herbst FA, Moche M, Hecker M, Becher D. Quantitative Proteomic View on Secreted, Cell Surface-Associated, and Cytoplasmic Proteins of the Methicillin-Resistant Human Pathogen *Staphylococcus aureus* under Iron-Limited Conditions. J Proteome Res. 2011 Apr 1;10(4):1657–66.

19. Olaya-Abril A, Jiménez-Munguía I, Gómez-Gascón L, Rodríguez-Ortega MJ. Surfomics: shaving live organisms for a fast proteomic identification of surface proteins. J Proteomics. 2014 Jan 31;97:164–76.

20. Dreisbach A, Hempel K, Buist G, Hecker M, Becher D, van Dijl JM. Profiling the surfacome of *Staphylococcus aureus*. PROTEOMICS. 2010;10(17):3082–96.

21. Wolden R, Pain M, Karlsson R, Karlsson A, Aarag Fredheim EG, Cavanagh JP. Identification of surface proteins in a clinical Staphylococcus haemolyticus isolate by bacterial surface shaving. BMC Microbiol. 2020 Apr 7;20(1):80.

22. Sui P, Miliotis T, Davidson M, Karlsson R, Karlsson A. Membrane protein digestion - comparison of LPI HexaLane with traditional techniques. Methods Mol Biol Clifton NJ. 2011;753:129–42.

23. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinforma Oxf Engl. 2010 Jul 1;26(13):1608–15.

24. Yu CS, Chen YC, Lu CH, Hwang JK. Prediction of protein subcellular localization. Proteins. 2006 Aug 15;64(3):643–51.

25. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001 Jan 19;305(3):567–80.

26. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol. 2019 Apr;37(4):420–3.

27. Bendtsen JD, Kiemer L, Fausbøll A, Brunak S. Non-classical protein secretion in bacteria. BMC Microbiol. 2005 Oct 7;5:58.

28. Jiang Y, Wang D, Wang W, Xu D. Computational methods for protein localization prediction. Comput Struct Biotechnol J. 2021 Oct 19;19:5834–44.

29. Jeffery C. Intracellular proteins moonlighting as bacterial adhesion factors. AIMS Microbiol. 2018 May 31;4(2):362–76.

30. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014 Mar 3;15(3):R46.

31. Ounit R, Wanamaker S, Close TJ, Lonardi S. CLARK: fast and accurate classification of metagenomic and genomic sequences using discriminative k-mers. BMC Genomics. 2015 Mar 25;16(1):236.

32. Mouratidis I, Chan CSY, Chantzi N, Tsiatsianis GC, Hemberg M, Ahituv N, et al. Quasi-prime peptides: identification of the shortest peptide sequences unique to a species. NAR Genomics Bioinforma. 2023 Jun;5(2):lqad039.

33. Mouratidis I, Baltoumas FA, Chantzi N, Patsakis M, Chan CSY, Montgomery A, et al. kmerDB: A database encompassing the set of genomic and proteomic sequence information for each species. Comput Struct Biotechnol J. 2024 Apr 21;23:1919–28.

34. Lee BS, Huang JS, Jayathilaka LP, Lee J, Gupta S. Antibody Production with Synthetic Peptides. Methods Mol Biol Clifton NJ. 2016;1474:25–47.

35. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014 Jul 15;30(14):2068–9.

36. Chaudhari NM, Gupta VK, Dutta C. BPGA- an ultra-fast pan-genome analysis pipeline. Sci Rep. 2016 Apr 13;6:24373.

37. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009 Dec 15;10:421.

38. Høie MH, Kiehl EN, Petersen B, Nielsen M, Winther O, Nielsen H, et al. NetSurfP-3.0: accurate and fast prediction of protein structural features by protein language models and deep learning. Nucleic Acids Res. 2022 Jul 5;50(W1):W510–5.

39. Dreisbach A, Wang M, van der Kooi-Pol MM, Reilman E, Koedijk DGAM, Mars RAT, et al. Tryptic Shaving of *Staphylococcus aureus* Unveils Immunodominant Epitopes on the Bacterial Cell Surface. J Proteome Res. 2020 Aug 7;19(8):2997–3010.

40. Hemmadi V, Biswas M. An overview of moonlighting proteins in *Staphylococcus aureus* infection. Arch Microbiol. 2021 Mar 1;203(2):481–98.

41. Widjaja M, Harvey KL, Hagemann L, Berry IJ, Jarocki VM, Raymond BBA, et al. Elongation factor Tu is a multifunctional and processed moonlighting protein. Sci Rep. 2017 Sep 11;7(1):11227.

42. Hussain M, Kohler C, Becker K. Enolase of Staphylococcus lugdunensis Is a Surface-Exposed Moonlighting Protein That Binds to Extracellular Matrix and the Plasminogen/Plasmin System. Front Microbiol. 2022;13:837297.

43. Mendonça M, Moreira GMSG, Conceição FR, Hust M, Mendonça KS, Moreira ÂN, et al. Fructose 1,6-Bisphosphate Aldolase, a Novel Immunogenic Surface Protein on Listeria Species. PLOS ONE. 2016 Aug 4;11(8):e0160544.

44. Ebner P, Prax M, Nega M, Koch I, Dube L, Yu W, et al. Excretion of cytoplasmic proteins (ECP) in *Staphylococcus aureus*. Mol Microbiol. 2015;97(4):775–89.

45. Bear A, Locke T, Rowland-Jones S, Pecetta S, Bagnoli F, Darton TC. The immune evasion roles of *Staphylococcus aureus* protein A and impact on vaccine development. Front Cell Infect Microbiol. 2023 Sep 27;13:1242702.

46. Bosi E, Monk JM, Aziz RK, Fondi M, Nizet V, Palsson BØ. Comparative genome-scale modelling of *Staphylococcus aureus* strains identifies strain-specific metabolic capabilities linked to pathogenicity. Proc Natl Acad Sci U S A. 2016 Jun 28;113(26):E3801–9.

47. Voss BJ, Gaddy JA, McDonald WH, Cover TL. Analysis of Surface-Exposed Outer Membrane Proteins in Helicobacter pylori. J Bacteriol. 2014 Jun 3;196(13):2455–71.

48. Wang J, Wang J, Wang Y, Sun P, Zou X, Ren L, et al. iTRAQ®-based quantitative proteomics reveals the proteomic profiling of methicillin-resistant *Staphylococcus aureus*-derived extracellular vesicles after exposure to imipenem. Folia Microbiol (Praha). 2021 Apr 1;66(2):221–30.

49. Wang W, Jeffery CJ. An analysis of surface proteomics results reveals novel candidates for intracellular/surface moonlighting proteins in bacteria. Mol Biosyst. 2016 Apr 26;12(5):1420–31.

50. Van Etten J, Stephens TG, Bhattacharya D. A k-mer-Based Approach for Phylogenetic Classification of Taxa in Environmental Genomic Data. Syst Biol. 2023 Sep 1;72(5):1101–18.

51. Panyukov VV, Kiselev SS, Ozoline ON. Unique k-mers as Strain-Specific Barcodes for Phylogenetic Analysis and Natural Microbiome Profiling. Int J Mol Sci. 2020 Jan 31;21(3):944.

52. Yang Y, Back CR, Gräwert MA, Wahid AA, Denton H, Kildani R, et al. Utilization of Staphylococcal Immune Evasion Protein Sbi as a Novel Vaccine Adjuvant. Front Immunol. 2018;9:3139.

53. Olatunde SK, Oladipo EK, Owolabi JB. Designing a novel *in-silico* multi-epitope vaccine against penicillin-binding protein 2A in *Staphylococcus aureus*. Inform Med Unlocked. 2022 Jan 1;33:101080.

54. Lawrence PK, Rokbi B, Arnaud-Barbe N, Sutten EL, Norimine J, Lahmers KK, et al. CD4 T Cell Antigens from *Staphylococcus aureus* Newman Strain Identified following Immunization with Heat-Killed Bacteria. Clin Vaccine Immunol. 2012 Apr;19(4):477–89.

55. Urdea M, Penny LA, Olmsted SS, Giovanni MY, Kaspar P, Shepherd A, et al. Requirements for high impact diagnostics in the developing world. Nature. 2006 Nov;444(1):73–9.

56. Wang CH, Wu JJ, Lee GB. Screening of highly-specific aptamers and their applications in paper-based microfluidic chips for rapid diagnosis of multiple bacteria. Sens Actuators B Chem. 2019 Apr 1;284:395–402.

57. Santos SB, Cunha AP, Macedo M, Nogueira CL, Brandão A, Costa SP, et al. Bacteriophage-receptor binding proteins for multiplex detection of Staphylococcus and Enterococcus in blood. Biotechnol Bioeng. 2020;117(11):3286–98.

58. Ilhan H, Guven B, Dogan U, Torul H, Evran S, Çetin D, et al. The coupling of immunomagnetic enrichment of bacteria with paper-based platform. Talanta. 2019 Aug 15;201:245–52.

59. Zhang X, Ren C, Hu F, Gao Y, Wang Z, Li H, et al. Detection of Bacterial Alkaline Phosphatase Activity by Enzymatic In Situ Self-Assembly of the AIEgen-Peptide Conjugate. Anal Chem. 2020 Apr 7;92(7):5185–90.

60. Park C, Lee J, Kim Y, Kim J, Lee J, Park S. 3D-printed microfluidic magnetic preconcentrator for the detection of bacterial pathogen using an ATP luminometer and antibody-conjugated magnetic nanoparticles. J Microbiol Methods. 2017 Jan 1;132:128–33.

61. Trotter AJ, Aydin A, Strinden MJ, O'Grady J. Recent and emerging technologies for the rapid diagnosis of infection and antimicrobial resistance. Curr Opin Microbiol. 2019 Oct;51:39–45.

62. Ma H, Yang J, Xu S, Liu C, Zhang Q. Combination of multiple functional markers to improve diagnostic accuracy. J Appl Stat. 2022;49(1):44–63.

63. Krey JF, Wilmarth PA, Shin JB, Klimek J, Sherman NE, Jeffery ED, et al. Accurate Label-Free Protein Quantitation with High- and Low-Resolution Mass Spectrometers. J Proteome Res. 2014 Feb 7;13(2):1034–44.

4.6 Supporting information:

Table S4.1: Strains chosen for the tryptic shaving experiment

N°	Specie	Accession N° Genbank	CCUG code	Assembly status	Other culture collections	Place of origen	Sample type	Deposit date
1	S. aureus	GCA_002025145.1	CCUG 10778	Complete	ATCC 6538	London UK	Human lesion	29599
2	S. aureus	GCA_001879295.1	CCUG 15915	Contig	ATCC 29213	Rockville USA	Wound	30901
3	S. aureus	GCA_900457405.1	CCUG 1799	Contig	NCTC 8531	London UK	Human blood, Osteomy elitis	26525
4	S. aureus	GCA_006094915.1	CCUG 1800T	Complete	ATCC 12600	London UK	Pleural fluid	26525
5	S. aureus	GCA_900636395.1	CCUG 1801	Complete	NCTC 6131	London UK	clinical	26525
6	S. aureus	GCA_900474695.1	CCUG 2353	Complete	NCTC 5663	London UK	Unknown	26772
7	S. aureus	GCA_900457855.1	CCUG 2354	Contig	NCTC 7428	London UK	Unknown	26772
8	S. aureus	GCA_900458145.1	CCUG 25925	Contig	NCTC 10655	London UK	Human leg abscess	32871
9	S. aureus	GCA_003052445.1	CCUG 41586	Contig	ATCC 43300	Kansas USA	Human clinical isolate	36159
10	S. aureus	GCA_013427085.1	CCUG 41879	Contig	ATCC 33591	New York USA	Human blood	36224
11	S. aureus	GCA_900457575.1	CCUG 54581	Contig	NCTC 10833	Czekia	clinical	39167

This manuscript is supported by additional datasets which can be obtained through link:

https://sciencedata.dk/shared/11489cee00d2c52f965566da7c6d0409?download

Supplementary_Dataset_4.1.xlsx: Protein table with abundance values for all identified proteins. Result of the Proteome Discoverer search

Supplementary_Dataset_4.2.xlsx: Uniprot ids and annotations of 346 common core proteins identified by tryptic shaving and pangenome analysis

Supplementary_Dataset_4.3.xlsx: Relative solvent accessibility (RSA) values for all common core protein amino acid residues

Supplementary_Dataset_4.4.xlsx: Quasiprime peptides which are contained in the common core proteins

Chapter 5

Conclusions and outlook

5 Conclusions and outlook

The objective of this thesis was to employ mass spectrometry (MS) based bottom-up proteomics to investigate the protein expression profiles of *Staphylococcus aureus*. The named bacterium is one of the leading causes of diverse infections, ranging from minor skin infections to severe conditions such as pneumonia and sepsis. The emergence of methicillin-resistant *S. aureus* (MRSA) strains, which exhibit resistance to many standard antibiotics, underscores the public health threat posed by this pathogen (1). Therefore, advancing our understanding of *S. aureus* is crucial for developing novel treatments, enhancing infection control measures, and mitigating the impact of antibiotic-resistant infections.

By analyzing the proteomic data, this study aimed to identify proteins involved in virulence, adaptation to the host environment, and biomarkers that could aid in the rapid diagnostics of the bacterium. MS based proteomics has dramatically evolved in the last two decades, enabling the identification of more proteins in shorter gradient time, identification of a greater number of proteins from limited sample amounts, and more accurate quantification These advances were driven by the development and improvement of mass spectrometry (MS) and chromatography instrumentation, which have enabled enhanced sensitivity and resolution, improving the depth of proteomic analysis (2,3). In bottom-up proteomics, the aim is to compare relative differences in protein expression. This relative quantification is essential for understanding biological processes and therefore, quantitative accuracy is critical (4).

In Chapter 2 of this thesis, commonly used proteomics analysis pipelines were used in a benchmarking study to establish which pipeline performs the best in terms of identified proteome depth, reproducibility and quantitative accuracy. Benchmarked pipelines included classical label-free quantification (LFQ) data-dependent acquisition (DDA) approach, in combination with software tools such as Proteome Discoverer (PD) and MaxQuant (MQ). Furthermore, a label-based quantification using tandem mass tags (TMT) employing multiplexing, offline fractionation and an MS3 quantification on a tribrid MS instrument was tested. Lastly, a library free data independent acquisition (DIA) was evaluated, in conjunction with software packages DIA-NN and Spectronaut. The best approach was TMT in terms of all parameters tested (quantitative performance, proteome depth, missing values, reproducibility), followed by LFQ DDA PD approach. While Data-Independent Acquisition (DIA) identified a higher number of proteins, the results of the differential expression analysis of the samples did not match the expected theoretical values, particularly in terms of statistical significance (p value >0.05). The results of the benchmarking study were used as a guide for the next study in

this thesis which investigated proteome dynamics of *S. aureus* in infection like conditions. While it was established TMT performed the best, complexity of the sample preparation, cost and time were taken in consideration. Moreover, the number of samples in the study (N=20) meant the sample size exceeds the number of labeling tags in the TMTpro 18-plex set. This can be mitigated by using multiple TMT sets and bridging them with a reference sample, however this was shown to introduce batch effects and missing values (5), a phenomenon observed in our lab also with certain TMT multibatch datasets. All these factors influenced our decision to use label-free quantification (LFQ) with data-dependent acquisition (DDA) and Proteome Discoverer (PD) for the other experiments in this thesis. The benchmarking study provided insights into the overall quality and added validity to the subsequent research, where biological changes were examined.

In Chapter 3 of this thesis, bottom-up proteomics was applied to study temporal proteome response of *S. aureus* during co-cultivation with human alveolar epithelial cell line (A549) which aimed to mimic infection like conditions. With the applied LFQ DDA proteomics approach in combination with PD, more than 60% of the proteome was quantified, achieving the highest proteome depth among similar studies of *S. aureus* under infection-like conditions. Differential expression analysis between infection time points and control revealed a subset of *S. aureus* proteins with significantly altered abundance. Proteins that are differentially expressed include those involved in adapting to the host environment such as proteins involved in anaerobic metabolism of fermentation and nitrate respiration which were upregulated in infection conditions compared to control. Moreover, a significant change in expression of several virulence factors connected to the two component systems (TCSs) and their respective regulatory proteins was observed and described. Changes in the abundance of virulence-related proteins were highlighted in the time-series soft clustering analysis which identified four clusters of protein expression profiles. The expression of several virulence factors was associated with a cluster enriched in the synthesis pathway of branched-chain amino acids (BCAAs), which has recently been linked to the major virulence regulator CodY (6). Additional experimental validation identified proteins whose gene disruption significantly altered cytotoxicity toward host cells. However, further and improved methods of validation are required to elucidate the role of these proteins in infection. It is worth mentioning that numerous uncharacterized proteins showed significantly altered abundances at infection time points and had similar expression profiles to virulence factors. In the future, these proteins could be further

investigated to uncover their roles in pathogenicity and their potential as targets for therapeutic intervention.

The main aim of Chapter 4 was to identify surface proteins that are conserved and unique to *S. aureus* as potential biomarkers for diagnostics. This was achieved by selecting a diverse range of evolutionary distinct *S. aureus* clinical strains and conducting a proteomic analysis using tryptic surface shaving, a mass spectrometry DDA approach and employing various bioinformatics tools. The study identified several hundred proteins, some of which are traditionally considered intracellular but also appear on the surface of bacteria. The approach then focused on those consistently present across different strains of and unique to *S. aureus*. The core proteins of S *aureus* were identified through pangenome analysis and proteins were termed unique by matching species-specific peptides (quasi-prime peptides (7)) against core proteins, ensuring they were present only in *S. aureus.* By leveraging a combination of pangenome analysis, species-specific peptide identification, and surface exposure prediction, the research identified 26 peptides in 15 proteins as promising diagnostic targets. Biomarkers can aid in diagnosing *S. aureus* by enabling the development of rapid diagnostic kits based on specific recognition of the biomarkers. Antibodies, bacteriophages, aptamers, and antimicrobial peptides all exhibit biorecognition capabilities that can be harnessed for this purpose (8). Future research should concentrate on determining whether any of the recognition elements can accurately identify the biomarkers predicted in this study, as well as assessing their specificity and sensitivity. Lastly, it is worth mentioning that the pipelines developed in this thesis can be applied to other pathogenic bacteria, particularly Gram-positive bacteria, thus aiding in the development of strategies to combat several bacterial pathogens.

5.1 References:

1. Tong SYC, Davis JS, Eichenberger E, Holland TL, Fowler VG. *Staphylococcus aureus* Infections: Epidemiology, Pathophysiology, Clinical Manifestations, and Management. Clin Microbiol Rev. 2015 Jul;28(3):603–61.

2. Halder A, Verma A, Biswas D, Srivastava S. Recent advances in mass-spectrometry based proteomics software, tools and databases. Drug Discov Today Technol. 2021 Dec 1;39:69–79.

3. Sivanich MK, Gu TJ, Tabang DN, Li L. Recent advances in isobaric labeling and applications in quantitative proteomics. Proteomics. 2022 Oct;22(19–20):e2100256.

4. Dupree EJ, Jayathirtha M, Yorkey H, Mihasan M, Petre BA, Darie CC. A Critical Review of Bottom-Up Proteomics: The Good, the Bad, and the Future of This Field. Proteomes. 2020 Sep;8(3):14.

5. Brenes A, Hukelmann J, Bensaddek D, Lamond AI. Multibatch TMT Reveals False Positives, Batch Effects and Missing Values. Mol Cell Proteomics MCP. 2019 Oct;18(10):1967–80.

6. Kaiser JC, King AN, Grigg JC, Sheldon JR, Edgell DR, Murphy MEP, et al. Repression of branched-chain amino acid synthesis in *Staphylococcus aureus* is mediated by isoleucine via CodY, and by a leucine-rich attenuator peptide. PLOS Genet. 2018 Jan 22;14(1):e1007159.

7. Mouratidis I, Chan CSY, Chantzi N, Tsiatsianis GC, Hemberg M, Ahituv N, et al. Quasi-prime peptides: identification of the shortest peptide sequences unique to a species. NAR Genomics Bioinforma. 2023 Jun;5(2):lqad039.

8. Sande MG, Rodrigues JL, Ferreira D, Silva CJ, Rodrigues LR. Novel Biorecognition Elements against Pathogens in the Design of State-of-the-Art Diagnostics. Biosensors. 2021 Nov;11(11):418.

Data availability

The mass spectrometry proteomics data for Chapter 2,3 and 4 is accessible through a following link:

https://dtudk-my.sharepoint.com/:f:/r/personal/emasve_dtu_dk/Documents/raw_data?csf=1&web=1&e=uEpKwB