PROTEOMICS OF MICROBIAL HUMAN PATHOGENS

EDITED BY : Nelson C. Soares, Jonathan M. Blackburn and German Bou
PUBLISHED IN : Frontiers in Microbiology

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: **researchtopics@frontiersin.org**

PROTEOMICS OF MICROBIAL HUMAN PATHOGENS

Topic Editors:
Nelson C. Soares, University of Cape Town, South Africa
Jonathan M. Blackburn, University of Cape Town, South Africa
German Bou, Complejo Hospitalario Universitario A Coruña, Spain
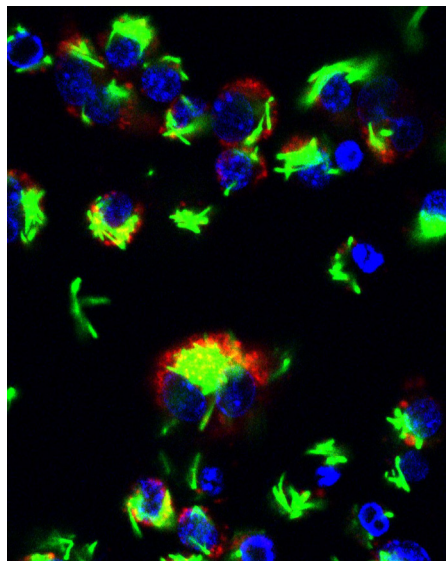
The murine macrophage cell line RAW 264.7 was infected with GFP-labelled Mycobacterium smegmatis at a multiplicity of infection (MOI) of 10:1. The nuclei of RAW cells were visualised with Hoechst and lysosomal compartments with LysoTracker® Red DND-99. Infection of marcophages was followed using a Zeiss Axiovert 200M LSM 510 Meta Confocal Microscope
Credit: Dr. Clemens Hermann and Ms Nazla Hassen (UCT-iBMS)

According to the World Health Organization (WHO), in 2012 infectious diseases and related conditions account for more than 70% of premature deaths across 22 African countries and estimated 450 000 people worldwide developed multi-drug resistant tuberculosis. This alarming situation, of great public health concern, calls for the urgent development of novel and efficient responding strategies. The employment of important research platforms, such as genomics and proteomics, has contributed significant insight into the mechanisms underlying microbial infection and microbe-host interaction. In this Frontiers Research Topic, we aim to produce a timely and pertinent discussion regarding the current status of "Proteomics of microbial Human pathogens" and the role of proteomics in combating the challenges posed by microbial infection and indeed acquired anti-microbial resistance.

As the field of proteomics progressed from 2-DE gel based approaches to modern LC-MS/MS based workflows, remarkable advances have been reported in terms of data quantity and quality. Given the immediate and enormous advantages that high resolution and accurate mass spectrometers have brought to the field, proteomics has now evolved into a robust platform capable of generating large amounts of comprehensive data comparable to that reported previously in genomics studies. For example, detection of the complete yeast proteome has been reported and other small proteomes, such as those of bacteria, are within reach. Mass spectrometry-based proteomics has become an essential tool for biologists and biochemists, and is now considered by many as an essential component of modern structural biology.

Additionally, the introduction of high-resolution mass spectrometers has driven the development of various different strategies aimed at accurate quantification of absolute and relative amount of protein(s) of interest. Emerging targeted mass spectrometry methodologies such as; Selected Reaction Monitoring (SRM), Parallel Reaction Monitoring (PRM) and SWATH, are perhaps the latest breakthrough within the proteomics community. Indeed, through a label free approach, targeted mass spectrometry offers an unequalled capability to characterize and quantify a specific set of proteins reproducibility, in any biological sample. Usefully, Aebersold and colleagues have recently generated and validated a number of assays to quantify 97% of the 4,012 annotated Mycobacterium tuberculosis (Mtb) proteins by SRM. As such, the Mtb Proteome library represents a valuable experimental resource that will undoubtedly bring new insight to the complex life cycle of Mtb. Finally, as reviewed recently in Frontiers Research Topic, mass spectrometry-based proteomics has had a tremendous impact on our current understanding of post translational modification (PTM) in bacteria including the key role of PTMs during interaction of pathogenic bacteria and host interactions.

We believe that our understanding of microbial Human pathogens has benefited enormously from both 2-DE gel and modern LC-MS/MS based proteomics. It is our wish to produce an integrated discussion surrounding this topic to highlight the existing synergy between these research fields. We envisage this Research Topic as a window to expert opinions and perspectives on the realistic practicalities of proteomics as an important tool to address healthcare problems caused by microbial pathogens.

Table of Contents

frontiers
in Microbiology

Check for
updates

Editorial: Proteomics of Microbial Human Pathogens

Nelson C. Soares[1]*, German Bou[2] and Jonathan M. Blackburn[1]

[1] Division of Chemical and Systems Biology, Department of Integrative Biomedical Sciences, Faculty of Health Sciences, Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa, [2] Servicio de Microbiologia-Instituto de Investigación Biomédica, Complejo Hospitalario Universitario A Coruña, A Coruña, Spain

The Editorial on the Research Topic

Proteomics of Microbial Human Pathogens

Despite remarkable advances in treatment and prevention, infectious diseases remain amongst the leading causes of death worldwide, particularly in the developing world, and drug resistant pathogens are ominously on the rise. By way of one specific example, the global incidence of human tuberculosis (TB) disease—caused by infection with the pathogen *Mycobacterium tuberculosis*—is estimated to be ∼9 million new cases (∼0.1% of the global population) per annum, causing 1.5–2 million deaths per annum and with an estimated 450,000 people developing multi-drug resistant tuberculosis each year); moreover, the WHO estimates that ∼1/3rd of the World's population carries a latent *M. tuberculosis* infection, thus representing a huge reservoir of potential future TB cases. In the context of this Research Topic, it is also relevant that the burden of TB disease is very uneven globally, with certain countries in Africa having a much higher incidence than observed in the developed world—for example, in South Africa the national incidence of TB disease is around 1% of the population and a single city, Cape Town, reports more cases of TB annually than the whole of North America and Europe combined. Furthermore, roughly 21% of all deaths in South Africa are associated with TB disease today, despite proactive efforts at TB control since the beginning of the twentieth century and despite the TB control program in Cape Town achieving 75% case finding and 85% completion rates for smear positive disease. Similarly depressing statistics abound for numerous other infectious diseases in the developing world, with disease burden being driven by both microbial adaption (to drug resistance and to differing ecologies), as well as by the emergence of new, often zoonotic pathogens (e.g., ebola and zika viruses). According to the World Health Organization (WHO), infectious diseases today account for more than 70% of premature deaths across 22 African countries (Who, 2014), with co-infections, for example with HIV or helminths, being rife. It is therefore critical that we develop new molecular knowledge now that will inform new strategies in the global fight against human microbial pathogens—including those that are not prevalent in the developed world—with the use of state-of-the art "omics" technologies set to take center-stage through providing a detailed understanding of the mechanistic basis of pathogenicity and by creating a comprehensive description of the molecular interplays that mediate host-pathogen interactions.

During the last two decades, powerful high throughput research platforms—namely genomics, transcriptomics, and proteomics—have contributed significantly to our understanding of the life styles of pathogenic microbes, including important aspects of host interactions and infection (Raskin et al., 2006; Merhej et al., 2013; Yang et al., 2015). Of all the omics platforms, proteomics is perhaps the research platform that has undergone the greatest transformation, including a structural transformation, as it evolved from initial two-dimensional gel electrophoresis (2-DE gel) based workflows to sophisticated shotgun proteomics analysis, capable

of generating comprehensive data with a coverage of individual proteomes that begins to approach that of corresponding genomics/transcriptomics datasets now. Modern mass spectrometry (MS)-based approaches today enable the detection and relative/absolute quantification of several thousand proteins in a single run, providing the opportunity to make a major impact on our general understanding of microbial pathogens. The goal of this Research Topic is thus to provide a survey of recent mass spectrometry-based research on the proteomics of human pathogens, which term we use broadly to refer to microorganisms that cause disease in humans, including *inter alia* virus, bacteria, fungi, and protozoa.

In this Research topic, we have particularly sought to highlight the powerful synergies that can be found between the fields of proteomics and microbiology, as exemplified by the review of Baarda and Sikora, which outlines the treasure hunt for counter-measures against old disease (Baarda and Sikora). Their review starts by acknowledging early contributions of 2-DE gel approaches, and how this approach implicated novel proteins—such as peroxidoxin, outer membrane protein Rmp, and the 50S ribosomal proteins L7/L12—in the *Neisseria gonorrhoeae* acquired resistance to spectinomycin. They then highlight how the employment of recent quantitative proteomics (SILAC, iTRAQ, and iCAT) has illuminated the pathways utilized by *N. gonorrhoeae* to adapt to different lifestyles, including during host interaction itself. A similar trend, suggesting that progression in the field of proteomics strongly augments classical microbiological research on microbial pathogens, is also evident in several others reviews (Ravikumar et al.; Pérez -Llarena and Bou; Soufi and Soufi), as well as opinion articles, in this Research Topic (Soares and Blackburn).

Elsewhere in this Research Topic, Perez-Llarena and Bou describe proteomics as a tool to study bacterial virulence (Pérez-Llarena and Bou) and, following this lead, this Research Topic carries several original reports that employ proteomics based approaches to investigate different aspects of bacterial/fungi virulence, including biofilm formation (Arnal et al.), motility (Merino et al.), virulence factors (Diehl et al.), altered virulence between strains (Peters et al.), and determinants that mediate host interaction in *Candida albicans* (Marin et al.). For example, Peters et al. used a combination of discovery and targeted liquid chromatography-tandem mass spectrometry (LC-MS/MS)-based proteomics to compare the proteomes of six clinically relevant mycobacterial strains within the *M. tuberculosis* complex (MTBC) (Peters et al.); MTBC members show a high degree of genetic conservation (~99.9%), yet clinically they exhibit differing pathogenicity and virulence. Peters et al. identified an average of 3290 protein groups for each MTBC organism, corresponding to >80% coverage of the theoretical proteomes; thereafter, the authors identified quantitative differences between strains for specific proteins that could be linked to enhanced bacterial fitness in the more virulent W. Beijing lineage of *M. tuberculosis* (Peters et al.).

This Research Topic also includes a number of manuscripts that explore novel applications of mass spectrometry within microbial research (Hartmann et al.); (Potgieter et al.); (Soares and Blackburn) (Soufi and Soufi), providing evidence that

proteomics has now evolved into a robust platform capable of generating comprehensive datasets comparable in size and utility to those described in genomic studies and enabling the application of mass spectrometry-based proteomics to genome reannotation and refinement (Krug et al., 2011). For example, in this topic, Potgieter et al. report the results of a proteogenomic analysis of *M. smegmatis*, integrating mass spectrometry-based proteomics with genomic six frame translation and *ab initio* gene prediction databases to identify a total of 2887 open reading frames (ORFs), including 2810 ORFs previously annotated to a reference protein and 63 ORFs not previously annotated to any reference protein (Potgieter et al.).

Finally, several papers in this Research Topic have described the enormous contribution of proteomics to the knowledge-based regarding bacterial protein post translation modifications (PTMs) and their intimate association with virulence (Ravikumar et al.); (Calder et al.); (Pérez -Llarena and Bou); (Soares and Blackburn); (Soufi and Soufi) and it will be interesting to see whether future investigation of bacterial PTMs provide a means to discover targets for novel therapies (Soufi and Soufi). In an Opinion article, we challenge researchers in the field to join efforts and to take advantage of recent advances in mass spectrometry instrumentation and in the development of targeted quantitative workflows that enable detection and accurate quantification of bacterial PTMs (Soares and Blackburn) since we believe that this represents a golden opportunity to access the dynamics of bacterial PTMs at sites of infection and that, through this, researchers will gain meaningful insights into the functional role of such PTMs during host-pathogen interactions.

CONCLUSION

Overall, this Research Topic demonstrates the huge potential for modern mass spectrometry-based (phospho)proteomics to yield major breakthroughs in our understanding of host-pathogen interactions. Therefore, it seems reasonable to suppose that as proteomics-based experimentation is married ever more closely with biologically relevant models of human microbial disease, the analytical power of quantitative mass spectrometry will yield testable hypotheses about key molecular host-pathogen interactions and might identify new candidate drug or vaccine targets. However, in order to reach this Holy Grail, the burgeoning microbial proteomics field will need to not only infer but then experimentally validate true biological significance from amongst the vast datasets generated. Tighter integration and iteration between computational modeling of proteomic data and model-driven further experimentation, including use of cell biology, imaging and CRISPR technologies therefore seems likely to represent the future.

AUTHOR CONTRIBUTIONS

NS and JB wrote the original draft of the paper, read and edited the final draft. GB critically discussed the content of the manuscript.

ACKNOWLEDGMENTS

REFERENCES

Krug, K., Nahnsen, S., and Macek, B. (2011). Mass spectrometry at the interface of proteomics and genomics. *Mol. Biosyst.* 7, 284–291. doi: 10.1039/C0MB00168F

Merhej, V., Georgiades, K., and Raoult, D. (2013). Postgenomic analysis of bacterial pathogens repertoire reveals genome reduction rather than virulence factors. *Brief. Funct. Genomics* 12, 291–304. doi: 10.1093/bfgp/elt015

Raskin, D. M., Seshadri, R., Pukatzki, S. U., and Mekalanos, J. J. (2006). Bacterial genomics and pathogen evolution. *Cell* 124, 703–714. doi: 10.1016/j.cell.2006.02.002

Who (2014). *World Health Statistics 2014*. Geneva. WHO.

Yang, Y., Hu, M., Yu, K., Zeng, X., and Liu, X. (2015). Mass spectrometry-based proteomic approaches to study pathogenic bacteria-host interactions. *Protein Cell* 6, 265–274. doi: 10.1007/s13238-015-0136-6

Identification of Quantitative Proteomic Differences between *Mycobacterium tuberculosis* Lineages with Altered Virulence

Julian S. Peters[1][†], Bridget Calder[2][†], Giulia Gonnelli[3], Sven Degroeve[3], Elinambinina Rajaonarifara[2], Nicola Mulder[2], Nelson C. Soares[2], Lennart Martens[3] and Jonathan M. Blackburn[2]*

[1] Centre of Excellence for Biomedical TB Research, Witwatersrand University, Johannesburg, South Africa, [2] Department of Integrative Biomedical Sciences, Institute of Infectious Diseases and Molecular Medicine, University of Cape Town, Cape Town, South Africa, [3] VIB, Ghent University, Ghent, Belgium

Evidence currently suggests that as a species *Mycobacterium tuberculosis* exhibits very little genomic sequence diversity. Despite limited genetic variability, members of the *M. tuberculosis* complex (MTBC) have been shown to exhibit vast discrepancies in phenotypic presentation in terms of virulence, elicited immune response and transmissibility. Here, we used qualitative and quantitative mass spectrometry tools to investigate the proteomes of seven clinically-relevant mycobacterial strains—four *M. tuberculosis* strains, *M. bovis*, *M. bovis* BCG, and *M. avium*—that show varying degrees of pathogenicity and virulence, in an effort to rationalize the observed phenotypic differences. Following protein preparation, liquid chromatography mass spectrometry (LC MS/MS) and data capture were carried out using an LTQ Orbitrap Velos. Data analysis was carried out using a novel bioinformatics strategy, which yielded high protein coverage and was based on high confidence peptides. Through this approach, we directly identified a total of 3788 unique *M. tuberculosis* proteins out of a theoretical proteome of 4023 proteins and identified an average of 3290 unique proteins for each of the MTBC organisms (representing 82% of the theoretical proteomes), as well as 4250 unique *M. avium* proteins (80% of the theoretical proteome). Data analysis showed that all major classes of proteins are represented in every strain, but that there are significant quantitative differences between strains. Targeted selected reaction monitoring (SRM) assays were used to quantify the observed differential expression of a subset of 23 proteins identified by comparison to gene expression data as being of particular relevance to virulence. This analysis revealed differences in relative protein abundance between strains for proteins which may promote bacterial fitness in the more virulent W. Beijing strain. These differences may contribute to this strain's capacity for surviving within the host and resisting treatment, which has contributed to its rapid spread. Through this approach, we have begun to describe the proteomic portrait of a successful mycobacterial pathogen. Data are available via ProteomeXchange with identifier PXD004165.

Keywords: *Mycobacterium tuberculosis*, virulence, proteomics, SRM, fitness, stress response

INTRODUCTION

Tuberculosis disease is caused by the bacterium *Mycobacterium tuberculosis*, and remains one of the leading causes of death by a single pathogen worldwide. Despite the presence of a vaccine and a number of antibiotics for the disease, it continues to cause about 2 million deaths and 8 million new cases worldwide per year. The emergence of multiple and extremely drug resistant strains, together with HIV co-infection, are fuelling the pandemic-especially in developing countries. Furthermore, latent and subclinical tuberculosis infection compounds tuberculosis control strategies by creating an unseen pathogenic reservoir. Even though various strains of *M. tuberculosis* have whole genome sequences available, the bacterium still closely guards the secrets of its success as a human pathogen.

According to whole-genome analysis, members of the *M. tuberculosis* complex (MTBC) exhibit the greatest degree of genetic conservation above all other pathogenic bacteria (99.9%). This strict level of observed genetic homogeneity initially led to the assumption that genetic variety amongst different strains would not be of any clinical significance (Homolka et al., 2008). However, subsequent research has led to the understanding that traits manifested by members of the MTBC are influenced by the genetic and evolutionary background of the strains (Gagneux and Small, 2007). Although, thousands of strains have been identified, only a few seem to drive widespread disease outbreaks and multi-drug resistance (Bifani et al., 2002). *M. tuberculosis* isolates have been observed to exhibit vast discrepancies in phenotypic presentation especially with regard to clinical outcome and epidemiological behavior (Shimono et al., 2003; Baker et al., 2004; Gagneux and Small, 2007; Nicol and Wilkinson, 2008). The East Asian/Beijing *M. tuberculosis* lineage is particularly of interest due to its increasing prevalence in the global TB community, implying an apparent selective advantage compared to existing strains (Parwati et al., 2010). It is therefore cause for concern that modern Beijing lineages appear to be accumulating mutations which enhance pathogenicity, apparently under positive selection pressure (Merker et al., 2015). The exact mode by which increased pathogenicity is conferred in this lineage remains undetermined and is likely to be a combination of factors (Ribeiro et al., 2014), however some proposed mechanisms include enhanced stress response, drug resistance and altered host-pathogen interactions, as has been reviewed previously (Hanekom et al., 2011; Warner et al., 2015). On the other hand, some closely related strains in the *M. tuberculosis* complex (MTBC) have attenuated virulence in humans (such as the vaccine strain BCG), or are not typically human pathogens and will only opportunistically infect immunocompromised humans (Desforges and Horsburgh, 1991; Wang and Behr, 2014; Halstrom et al., 2015).

Whilst genetic variation across multiple strains has been studied in depth, the clinical and epidemiological consequences of genetic differences between mycobacterial strains remains poorly understood (Malik and Godfrey-Faussett, 2005). As a consequence, it is not known whether the proteome is comparatively static between different strains of *M. tuberculosis* or whether quantitative differences in the expressed proteomes could contribute in some way to differential virulence. Here, we used liquid chromatography mass spectrometry (LC-MS)-based proteomics to define and compare the proteomic complement of 6 clinically relevant mycobacterial strains within the MTBC as well as a strain of *Mycobacterium avium* as an outlier. While these strains are all pathogenic in principal, the extent to which they cause disease in humans varies greatly. We therefore aimed to identify protein expression profiles that might correlate with altered virulence amongst these strains by comparing more pathogenic strains in the MTBC to less pathogenic *Mycobacterium bovis*, BCG and *M. avium* strains.

MATERIALS AND METHODS

M. tuberculosis isolates H37Rv, W-Beijing, CAS and LAM3 were obtained from the Medical Microbiology Division of the University of Cape Town. The clinical strains representing lineages 2 (Beijing), 3 (CAS), and 4 (LAM3/F11) were isolated from pediatric patients from Red Cross war memorial hospital, Cape Town. *M. tuberculosis* H37Rv was used in all assays as a reference strain. Phylogeny of the isolates was determined using spoligotyping and MIRU-VNTR described in Sarkar et al. (2012) and is shown in Supplementary Table 1. The Danish strain of *M. bovis* BCG was used in this study. The *M. avium* strain was obtained from the National Health Laboratory Services (NHLS) laboratory and was verified using line probe assays. *M. bovis* was obtained from Stellenbosch University Health Sciences Department in Tygerberg Hospital, Cape Town.

Cell Culture

Cells were maintained in wholly synthetic Sautons media (2% glycerol, 0.4% L-asparagine, 0.2% glucose, 0.2% citric acid, 0.05% mono-potassium phosphate, 0.05% magnesium sulfate, 0.015% Tween 80, 0.005% ferric citrate, 0.00001% zinc sulfate at pH 7.4). Briefly, 190 ml of Sautons medium was inoculated with a 10 ml starter culture (approximately 10^8 bacteria/ml). The flasks were sealed and incubated at 37°C and 5% CO_2 with gentle agitation until OD_{600} reached 0.9 (approximately 6 weeks).

Protein Extraction

Proteins were extracted in a Biosafety level 3 facility in line with health and safety guidelines. Briefly, the cell pellet was separated from the culture filtrate by centrifugation at 4000 × g for 15 min in a bench-top centrifuge. Cell lysis was carried out by boiling the cell pellet in 1% SDS buffer (1% SDS, 100 mM Tris-HCl pH 7.6, 0.1 mM dithiothreitol (DTT), 1 mM PMSF) for 30 min. Cell debris was separated from the protein containing supernatant by centrifugation at 10,000 × g for 15 min in a bench top centrifuge and the supernatant containing the protein was transferred into a clean tube. Protein extracts from cell lysates were concentrated and buffer exchanged to 2 M urea buffer using 3 kDa MWCO filters (Millipore). Culture filtrate proteins were concentrated and buffer exchanged into 2 M urea using 15 ml 10 kDa MWCO filters. Protein concentration was determined using the BCA assay kit (Thermo Scientific). A 10 kDa MWCO filter was used for the culture filtrate instead of a 3 kDa because this is the lowest filter size available for large volumes, however, according

to manufacturer's product specifications (Millipore), proteins as low as 3 kDa are still retained on 10 kDa MWCO filters.

Protein Separation (1D SDS PAGE)

Proteins were separated according to molecular weight using an SDS PAGE gel system. The separating gels were made from 10% acrylamide: bis-acrylamide, 0.375 M Tris-HCl (pH 8.8), 7.5% SDS, 0.5% ammonium persulphate and 0.1% TEMED. The stacking gels consisted of 4% acrylamide: bis-acrylamide, 0.125 M Tris-HCl (pH 6.8), 0.1% SDS, 0.5% ammonium persulphate and 0.1% TEMED. 40 μg of each sample (culture filtrate and intracellular protein) was mixed with an equal volume of 2x sample buffer and heated at 65°C for 5 min. Electrophoresis was performed from anode to cathode at 100 V using a BioRad mini-Protean II gel system until the bromophenol blue dye reached the bottom of the gel.

Protein Visualization

Visualization of the proteins on the gel was performed using Coomasie brilliant blue R250 for 1 h (50% methanol, 10% acetic acid and 0.1% Coomasie brilliant blue R250). Destaining of the gels was carried out by incubating on a shaker overnight at room temperature in destaining solution (10% methanol, 10% acetic acid).

In Gel Trypsin Digestion

Each gel lane for each strain sample was divided into 5 pieces (i.e., 5 culture filtrate fractions and 5 intracellular protein fractions, hence a total of 10 fractions per strain). Each gel piece was cut into smaller cubes and washed twice with water followed by 50% (v/v) acetonitrile for 10 min. The acetonitrile was replaced with 50 mM ammonium bicarbonate and incubated for 10 min. Washes with 50 mM ammonium bicarbonate were repeated twice to remove acetonitrile. All the gel pieces were then incubated in 100% acetonitrile until they turned white, after which the gel pieces were dried *in vacuo*. Proteins were reduced with 10 mM DTT for 1 h at 57°C. This was followed by brief washing steps of ammonium bicarbonate followed by 50% acetonitrile before proteins were alkylated with 55 mM iodoacetamide for 1 h in the dark. Following alkylation, the gel slices were washed with ammonium bicarbonate for 10 min followed by 50% acetonitrile for 20 min, before being dried *in vacuo*. The proteins in the gel cubes were digested with trypsin (Promega) at 37°C overnight in a 1:50 trypsin: protein ratio. The resulting peptides were extracted twice with 70% acetonitrile in 0.1% formic acid for 30 min and then dried and stored at −20°C. Dried peptides were dissolved in 5% acetonitrile in 0.1% formic acid and 10 μl injections were made for nano-LC chromatography.

Mass Spectrometry

All experiments were performed on a Thermo Scientific EASY-nLC II coupled to an LTQ Orbitrap Velos mass spectrometer (Thermo Scientific, Bremen, Germany) equipped with a nano-electrospray source. For liquid chromatography, separation was performed on an EASY-Column (2 cm, ID 100 μm, 5 μm, C18) pre-column followed by a, EASY-column (10 cm, ID 75 μm, 3 μm, C18) column with a flow rate of 300 nl/min. The gradient

used was from 5–15% B in 5 min, 15–35% B in 90 min, 35–60% B in 10 min, 60–80% B in 5 min, and kept at 80% B for 10 min. Solvent A was 100% water in 0.1% formic acid; solvent B was 100% acetonitrile in 0.1% formic acid. MS/MS data was acquired from the Orbitrap Velos in Top 20 CID mode.

Post MS Data Analysis

Raw data was captured from the mass spectrometer and converted to MS2 files using MakeMS2 software (Thermo Scientific). The data was then analyzed using Crux (McIlwain et al., 2014) and Mascot (Cottrell and London, 1999), and the output of MS2PIP (Degroeve and Martens, 2013) was used additional features for the Percolator algorithm. Spectra were obtained from each fraction of the gel (a total of 5 fractions per strain) and were viewed using Peaks v5.3. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Vizcaíno et al., 2016) partner repository with the dataset identifier PXD004165.

Protein Preparation for SRM-MS

Proteins were extracted in the BSL3 facility in line with health and safety guidelines as described in Section Protein extraction. Protein concentration was determined using a BCA assay according to manufacturer's protocol (#23227, Thermo Fisher Scientific). Protein preparation was performed using a filter aided sample preparation (FASP) method. Briefly, 200 μg of each protein sample was placed into a 10 KDa molecular weight cut off filter (MWCO) (Millipore). Protein cysteine residues were alkylated in the dark for 30 min in 10 mM iodoacetamide. Iodoacteamide was then removed by centrifugation at 14,000 × g for 15 min. Buffer exchange was performed twice with 8 M urea in 0.1 mM Tris-HCl pH 8.5 by centrifugation at 14,000 ×g for 15 min in a refrigerated bench-top centrifuge at 18°C. The urea buffer was then exchanged for 0.05 M ammonium bicarbonate by centrifugation. Sequencing-grade modified trypsin (#608-274-4330, Promega) was added at a ratio of 1:100 enzyme: substrate and incubated overnight at 37°C in a wet chamber. The peptides were finally collected through the filter into a clean collection tube by centrifugation at 14 000 × g for 10 min.

To stop the tryptic digest the pH was lowered to 2 using 50% trifluoro acetic acid (TFA) followed by an incubation for 15 min at 37°C with shaking at 500 rpm. The peptide solution was desalted with C18 reversed-phase columns (Pierce #89870-25) according to the manufacturer's instructions. Briefly, the C18 columns were activated with 50% methanol, followed by equilibration with 5% ACN: 0.5% TFA. After loading the sample, the columns were washed 3 times with 5% ACN: 0.5% TFA. Finally, peptides were eluted with 70% ACN, dried under vacuum, and re-solubilized in 0.1% FA to a final concentration of 8 μg/μl.

SRM Mass Spectrometry

All SRM experiments were performed on a TSQ Vantage triple quadrupole mass spectrometer (Thermo Fisher Scientific) equipped with a heated electrospray II ion source. For liquid chromatography and separation of peptides, a Synergi 4 μ Hydro

RP 150 × 4.60 mm 80 Å pore size C18 column (serial # 630710-14) was used with a column flow rate of 300 μl/min. The gradient used was from 5–15% B in 5 min, 15–35% B in 90 min, 35–60% B in 10 min, 60–80% B in 5 min and kept at 80% B for 10 min. Solvent A was 100% water in 0.1% formic acid, and solvent B was 100% acetonitrile in 0.1% formic acid.

The mass spectrometer was operated in positive mode using electrospray ionization with a voltage of 3500 V. The capillary temperature was set to 350°C and the collision gas pressure to 1.2 mTorr. Up to 336 transitions per run were acquired with a cycle time of 3 s and a dwell time of at least 20 ms. Collision energies were calculated per individual peptide transition ion using Skyline software and further optimized by a series of energy ramping experimental steps (10 steps of 5 V) to obtain the optimum energy of each transition. MS/MS data was acquired from the CID mode. Raw data was captured from the mass spectrometer and analyzed using Skyline software.

The proteins chosen for SRM analysis were based on relevance in pathogenicity and/or virulence as stated in literature as shown in Supplementary Table 2. To design SRM assays, peptides were chosen for analysis based on their prior observation in our discovery experiments. Although, each protein was typically observed by two or more peptides in the discovery experiment, the two best performing peptides were chosen to confirm each protein in the SRM assay, with the exception of 2 proteins (Rv1818c and Rv0833) (Supplementary Table 3). Due to the small size and non-redundant nature of the *M. tuberculosis* proteome, 3 ion transitions were set as the minimum number required to identify a peptide in an SRM assay.

SRM Data Analysis

Data analysis was carried out using Skyline software (MacCoss Lab software). Intra-assay ambiguity (CV) for each peptide was based on the calculated average protein concentration for a set of technical duplicate injections of each sample. Inter-assay CV was calculated for each peptide from across three biological replicates of the 7 strains. Quantitation of each peptide was carried out using the area under the curve for the peptide transitions assayed in Skyline. Retention times for the peptide standards were obtained by pre-assessment on the MS. For peptides without standards, the retention times were obtained from predictions made by Skyline software and gated at 5 s from the predicted retention time.

Protein Inference

The protein databases used to generate theoretical spectra were strain specific individual non-redundant fasta files obtained from Ensembl (www.ensembl.org), with the exception of LAM for which there is no Ensembl annotation, and W-Beijing whose annotated file is not sufficient for downstream cross strain comparison. For LAM, the UniProt fasta file was used (www.uniprot.org) and for W-Beijing, the H37Rv Ensembl fasta file was used. The parameters were standard across both searches and included carbamidomethylation of cysteine residues as a fixed modification and oxidation of methionine residues as a variable modification. Two missed cleavages were allowed and peptide mass tolerance was set at 10 ppm whilst fragment mass

tolerance was set to 0.5 Da. Decoy databases were used for FDR analysis and a cut-off was set at 5% for protein identifications.

The PSMs obtained from the search were used to predict the expressed protein repertoire of each sample. MS/MS spectra from the 10 fractions per strain were searched with each individual search engine. Combining results from multiple search engines yields higher protein identifications (Shteynberg et al., 2013), and therefore all the proteins identified from each individual search engine were combined and redundancy was removed to give one complete non-redundant dataset per study organism.

RESULTS

Total protein extracts from the discovery MS approach were quantified using BCA assay with concentrations ranging between 2.5–15 μg/μl for total cellular proteins and 10–40 μg/μl for culture filtrate proteins. Each tryptically digested sample was analyzed on the Orbitrap Velos to produce an LC-MS and MS/MS dataset. To assess the efficiency of the tryptic digest a descriptive analysis software package in Protein Pilot was used and the results are summarized in Supplementary Table 4.

As shown in Supplementary Figure 1A, the MS1 scan of H37Rv confirms a successful tryptic digest with the total ion chromatogram in showing a steady elution of peptides across the LC gradient. The 2D MS chromatogram (Supplementary Figure 1B) demonstrates the complexity of the sample, showing a significant number of discrete tryptic peptides eluting at the marked time point indicated on the 1D chromatogram (∼2600 s).

For the analysis of the MS2 spectra a software pipeline was implemented that combines the results of different peptide identification strategies. At the core of our pipeline is the semi-supervised learning algorithm implemented in Percolator (Brosch et al., 2009) that has been shown to obtain high identification sensitivity. The first tool in the pipeline is Crux (McIlwain et al., 2014) which is a reimplementation of the popular tool Sequest with added post-processing by the Percolator algorithm. The second tool is Mascot (Cottrell and London, 1999) for which again Percolator was used to post-process the MOWSE identification scores[1].

The Percolator tool allows for adding new features that can be exploited by the semi-supervised learning algorithm to further increase peptide identification sensitivity. It has been shown that adding features obtained from MS2 peak intensity predictions can significantly increase sensitivity (Sun et al., 2007). Therefore, we employed the MS2PIP (Degroeve and Martens, 2013) tool to predict the b- and y-ion peak intensities for all peptides suggested by Mascot (top ranked peptide for each MS2 spectrum). We then computed several features from the difference between the predicted and the observed MS2 peak intensities, such as the Pearson correlation. We observed that adding these features to the Percolator algorithm for Mascot did indeed increase identification sensitivity significantly.

---

[1]http://www.sanger.ac.uk/science/tools/mascotpercolator

Comparison of protein numbers obtained at 1% and 5% FDR showed that the use of 5% FDR allows a substantial increase in the absolute number of true positives with an insignificant increase in the absolute number of false positives, hence providing an apparently favorable trade-off in true positives over false positives (**Figure 1**). The final proteome obtained from each strain using all algorithms represented a high proportion of the theoretical protein fasta files for each strain, as shown in **Table 1**.
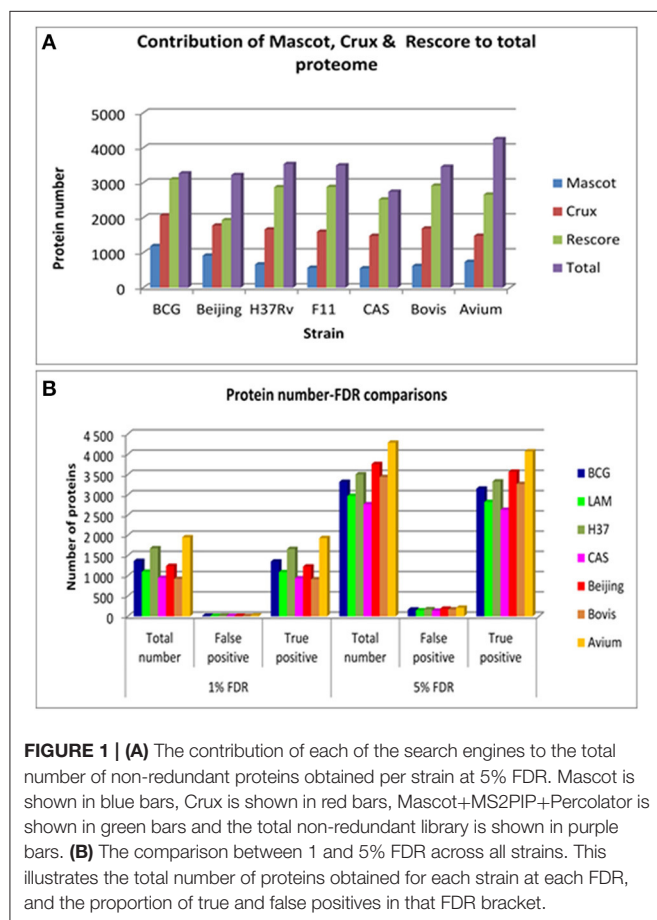
FIGURE 1 | (A) The contribution of each of the search engines to the total number of non-redundant proteins obtained per strain at 5% FDR. Mascot is shown in blue bars, Crux is shown in red bars, Mascot+MS2PIP+Percolator is shown in green bars and the total non-redundant library is shown in purple bars. **(B)** The comparison between 1 and 5% FDR across all strains. This illustrates the total number of proteins obtained for each strain at each FDR, and the proportion of true and false positives in that FDR bracket.

TABLE 1 | Total non-redundant number of proteins obtained in the experiment compared to the total number of proteins in the theoretical fasta file.

Strain	Total # proteins in fasta file	Total # detected experimentally (5% FDR)	% of total theoretical proteome found
M. avium	5314	4250	80
M. tuberculosis H37Rv	4049	3539	87
M. tuberculosis Beijing	4049	3224	80
M. tuberculosis CAS	4049	2746	68
M. tuberculosis LAM	3904	3500	90
M. bovis BCG	4041	3272	81
M. bovis	4001	3461	87

Data Alignment for Downstream Comparison

To carry out an effective cross strain comparison, it was crucial to ensure that as much of the theoretical proteome as possible was observed by discovery MS. After obtaining a non-redundant dataset for each strain using strain specific databases, it became necessary to convert all the protein IDs into a standard protein ID by orthology mapping to allow effective cross strain comparison. To achieve this, the total non-redundant IDs from each strain obtained by searching against its individual Ensembl fasta file were then mapped back to the Ensembl H37Rv protein IDs. These were all in turn mapped to UniProt accession numbers and Tuberculist "Rv" loci numbers to facilitate downstream analysis with tools such as GO analysis and pathway mapping. It was observed that there is a slight deficiency in ortholog mapping data between databases (Ensembl and UniProt), as well as shortcomings in ortholog mapping between strains. These discrepancies lead to a minor loss of information, as represented in **Figure 2**. The discrepancy in ortholog mapping was much more pronounced when mapping protein IDs to H37Rv from the more distantly related *M. avium,* which lies outside the MTBC group; this resulted in loss of approximately 50% of the biological information in downstream comparisons to *M. avium*. Other strains showed relatively small losses in the number of experimentally observed proteins mapped to H37Rv orthologs, for instance the LAM strain had approximately 600 proteins with no orthologs in H37Rv which were therefore not included in the cross-species comparison.

Qualitative Cross Species Comparison

With congruent IDs, strains were cross-compared to obtain a comprehensive qualitative comparison as summarized in **Figure 3** using the Venn diagram tool Venny [267]. Protein IDs from the 4 *Mycobacterium tuberculosis* strains were compared (**Figure 3A**) and, as expected, the majority of observed proteins were found to be shared amongst all *M. tuberculosis* strains, with less than 5% being strain specific. A total of 1938 proteins comprise the shared proteins among the 4 *M. tuberculosis* strains, perhaps representing a *M. tuberculosis* core proteome. A second diagram was generated comparing the collective, non-redundant proteomes of the four *M. tuberculosis* strains to those of *M. bovis*, BCG, *M. avium* (**Figure 3B**). Surprisingly, *M. avium* had many proteins in common with the MTBC strains, with only 12 unique proteins apparently unique to *M. avium*; this may simply reflect though the deficiencies in ortholog mapping between more distantly related organisms. The four *M. tuberculosis* strains also share 989 common proteins with *M. bovis* and BCG which they do not share with *M. avium*. A group of 168 was observed uniquely in the *M. tuberculosis* strains and not found in *M. bovis*, BCG or *M. avium*; these proteins were therefore earmarked as candidate virulence factors to be further explored.

Protein Expression Profiling

We subsequently sought to quantitatively assess a subset of the candidate virulence factors identified by discovery MS and cross-strain comparison using selected reaction monitoring (SRM)—a sensitive, reproducible and quantitative MS technique. Since

FIGURE 2 | The losses in number of identified proteins obtained after experimental ortholog mapping to *M. tuberculosis* H37Rv for each strain. Avium—*Mycobacterium avium*, H37Rv—*Mycobacterium tuberculosis* H37Rv, Beijing—*Mycobacterium tuberculosis* W. Beijing, CAS—*Mycobacterium tuberculosis* CAS, LAM—*Mycobacterium tuberculosis* LAM, BCG—*Mycobacterium bovis* BCG, Bovis—*Mycobacterium bovis*.

FIGURE 3 | Qualitative cross strain/species data analysis. (A) Venn diagram showing the overlap in the numbers of protein identifications between the 4 strictly *Mycobacterium tuberculosis* strains CAS, LAM, W-Beijing, and H37Rv. **(B)** Venn diagram representing the comparison between the collective non-redundant proteins obtained in the 4 strictly *M. tuberculosis* strains (CAS, LAM, W-Beijing, and H37Rv) denoted "Human clinical strains" and those of *M. bovis*, BCG, and *M. avium*.

the design of SRM robust assays can be a lengthy process and the capacity to highly multiplex hundreds of SRM assays remains challenging, we devised a strategy to create a short list of candidate proteins with possible relevance in differential clinical phenotypes observed between the *M. tuberculosis* isolates for subsequent quantitative proteomic analysis. To do this, we compared our proteomic data on each of the 168 proteins (Supplementary Table 5) that were observed only in the *M. tuberculosis* strains with 771 gene expression data sets contained in the TBDB, representing varying *in vitro* models of TB disease. We focussed attention in particular on 7 categories of experiment from the TBDB that aimed to more closely reflect *in vivo* conditions (e.g., starvation models, macrophage infection models, hypoxic models, etc.), the logic being that consistent over-expression of a protein in one of those categories might plausibly confer a selective advantage to the bacterium *in vivo*; the categories chosen are listed in **Table 2**.

For each of the 168 proteins, we assessed whether they were significantly over- or under-expressed in each of the 7 chosen categories of gene expression models and carried out statistical analysis on the gene expression values obtained from the TBDB experiments using packages in R (strategy depicted in **Figure 4**). Proteins from the 168 protein set whose gene expression showed an average fold change of ≥2 SD from the mean across all datasets in an individual condition were taken as significantly differentially expressed in that condition. Proteins that had significant fold change for fewer than 4 out of the 7 categories were removed from this list, resulting in a final shortlist of 23 proteins, summarized in **Table 3**. In order to create SRM assays for the shortlisted proteins, the *M. tuberculosis* proteome library (Schubert et al., 2013) was consulted and validated SRM assays for the proteins of interest were extracted.

A minimum of 2 peptides per protein and 3 transitions per peptide were assessed by SRM for the shortlisted proteins in all seven strains. Intra-assay and inter-assay coefficients of variation were determined for each individual peptide. Intra
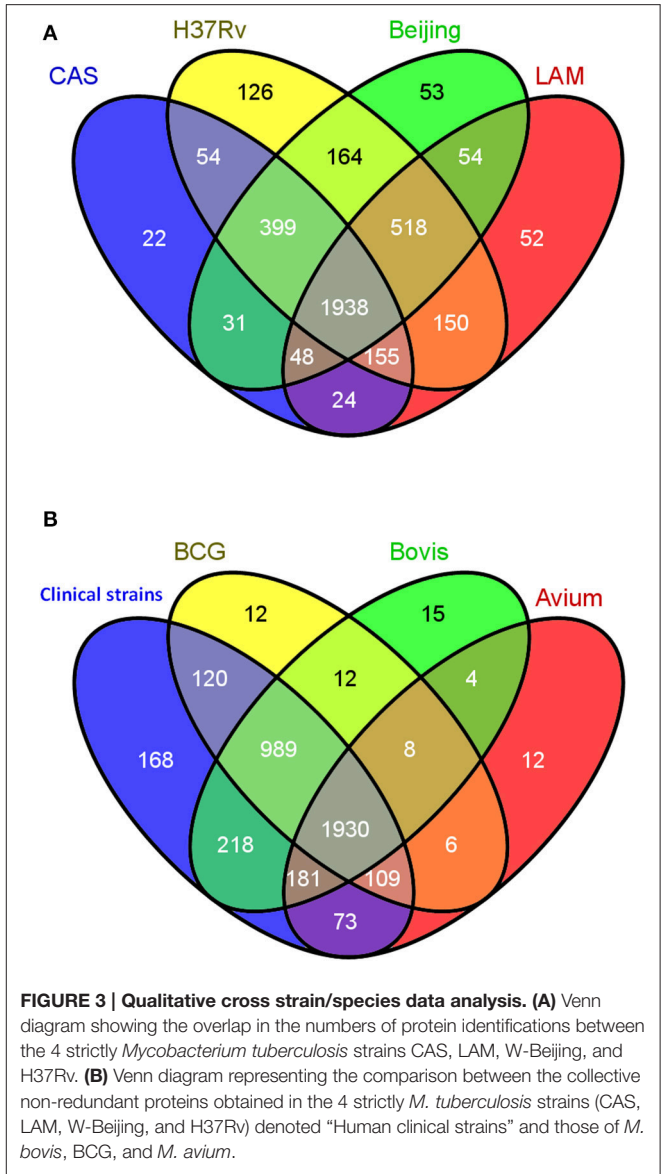
TABLE 2 | The number of significantly differentially expressed proteins from the 168 proteins unique to human clinical strains segregated into over- or under-expressed per category assessed (www.tbdb.org).

TBDB Experiment	Over-expressed	Under-expressed
Acid media	58	38
Macrophage	18	53
Hypoxia	59	49
NO treatment	78	32
Starvation	51	26
Persistence	23	43
Antibiotics	40	54

assay variability was based on 3 technical replicates per strain whilst inter assay variability was assessed based on 2 biological replicates. The signal of each peptide observed was obtained by

FIGURE 4 | Gene expression analysis approach of the short list of 168 proteins expressed in human clinical strains. Expression ratio data was obtained from the chosen TBDB experiments and statistical packages in R were used to convert the ratios into fold changes. Analysis of the fold changes was done in R to indicate proteins that have significant fold changes (≥2 from the mean) which are shown in the scatterplot above and below the shaded area.

summing the peak areas of each measured transition for that peptide and then normalizing by the total number of cells per strain at the point of protein extraction. An ANOVA test was used to determine if there was a significant difference in the expression of each of the peptides representing each protein. The single factor ANOVA results (**Table 3**) shows that of the 23 peptides assessed, 18 had a significant difference in expression between the 7 strains. The proteins were then broadly classified into 4 groups denoting some aspect of the organisms' success in the host (Supplementary Figure 2) in order to aid further interpretation of the data.

The four proteins assayed with roles in drug response are PyrB, PyrC, CarA, and CarB, and all form part of the pyrimidine biosynthetic operon in *M. tuberculosis*. All with the exception of carB (Rv1384) are more abundant in *M. avium* compared to the other strains, but within the MTBC the Beijing strain has the highest expression of PyrB, PyrC, and CarA.

Amongst proteins which are known to modulate the host immune response, the uncharacterized hypothetical protein Rv0966 is highly expressed in the LAM strain compared to the other strains. In this functional category, Rv2136c, Rv1002, Rv2703, and Rv2108 are more abundant in the Beijing strain. Rv1818c appears to be more abundant in the CAS strain while in the other strains it is present at comparatively low amounts.

Amongst proteins responsible for the growth of *M. tuberculosis* in the host, a possible toxin with unknown function, VapC2 (Rv0301), has the highest relative expression in the Beijing strain. PE_PGRS13 (Rv0833), MetC (Rv3340), and conserved hypothetical protein Rv3412 are all more

abundant in the LAM strain than in BCG or *M. avium*, although similarly highly abundant in H37Rv and *M. bovis*. PPE65 (Rv3621c) appears to be upregulated in both LAM and *M. avium* strains.

In terms of adaptation to stress, the Acyl-CoA dehydrogenase MbtN (Rv1346) is particularly abundant in the Beijing strain. The protein Rv0901 is apparently less abundant in *M. bovis* and BCG strains whereas it is relatively abundant in the other strains, particularly so in H37Rv and *M. avium*. Aspartate kinase (Rv3709c) is more abundant in H37Rv and LAM strains comparatively, and almost entirely absent in *M. avium* and *M. bovis*.

The contribution to total signal detected for each protein as measured by SRM in each of the 7 strains is represented in **Figure 5**. This analysis demonstrates that proteins which were detected in a particular strain in the discovery experiment, e.g., Rv0301 and Rv1002c which were observed only in the Beijing strain, tend to contribute the highest signal when measured by SRM. This appears to be clearly the case for 14 of the 23 proteins assessed (Rv0301, Rv1381, Rv1383, Rv1002c, Rv3709c, Rv1346, Rv3412, Rv2108, Rv3621, Rv0833, Rv0966c, Rv1818c, Rv2136c, and Rv3340).

DISCUSSION

Although, genetically similar, different strains of *M. tuberculosis* present very different clinical phenotypes in terms of virulence. We therefore postulated that there may be a proteomic mechanism underpinning the differences in pathogenicity

TABLE 3 | The 23 mycobacterial proteins identified in discovery proteomic experiments which were subsequently selected for quantitative targeted proteomic analysis.

Rv locus	Function and prediction from gene expression data	p-value
Rv0301	VapC2; toxin; possible mRNAse; impaired growth when expressed | macrophage infection models	6,75E-07
Rv0899	ArfA; outer membrane porin A; tolerance to acidic conditions, impaired growth at pH 5.5	5,55E-08
Rv0901	ArfC; unknown function; tolerance to acidic conditions, impaired growth at pH 5.5	6,42E-06
Rv0966	Unknown function; highly activated in the early stages of tuberculosis blood brain barrier invasion (CNS TB)	1,96E-05
Rv1002c	Pmt; probable mannosyltransferase; conserved membrane protein growth and survival in host	9,35E-07
Rv1346	MbtN; mycobactin biosynthesis; adaptation to intracellular environment | stress response models	0,020716
Rv1380	PyrB; pyrimidine biosynthesis; essential for growth; high confidence drug target | drug response models	0,039441
Rv1381	PyrC; pyrimidine biosynthesis; growth and survival in host | drug response models	4,59E-09
Rv1383	CarA; pyrimidine biosynthesis; growth and survival in host | drug response models	1,08E-06
Rv1384	CarB; pyrimidine biosynthesis; growth and survival in host	4,47E-09
Rv1980c	Mpt64; Unknown function; tolerance to starvation, highly immunogenic; vaccine and drug target potential	0,422015
Rv1997	CtpF; Metal cation-transporting ATPase; implicated in dormancy/persistance, response to hypoxia, NO	6,10E-06
Rv2108	PPE36; unknown function; immuno-active membrane component; diagnostic and vaccine target	0,876293
Rv2126c	PE_PGRS37; unknown function; possible virulence/adaptation	0,988059
Rv2136c	UppP; undecaprenyl pyrophosphatase; high confidence drug target | host immune evasion and virulence models	5,32E-12
Rv2156	MraY; peptidoglycan biosynthesis; growth and survival in host, high confidence drug target	0,97868
Rv2703	SigA; primary sigma factor in *M. tuberculosis*; host immune response modulator, virulence, growth in host, high confidence drug target | host immune evasion and virulence models	0,026023
Rv3340	MetC; methionine biosynthesis; growth and survival in host	0,649041
Rv3412	Unknown function; hypothetical protein; essential for cholesterol metabolism, essential during infection	3,27E-07
Rv3621c	PPE65; unknown function; possible virulence/adaptation	0,035777
Rv3709c	Ask; asparate kinase; survival in host	1,23E-05
Rv1818	PE-PGRS33; unknown function; modulation of host immune response, response to oxygen and starvation	1,61E-06
Rv0833	PE_PGRS13; unknown function; possible virulence/adaptation	9,79E-08

FIGURE 5 | Comparison of relative protein quantity per strain for the 23 mycobacterial proteins as determined by SRM. Each color in the stacked bar represents the percentage contribution of that strain to the total detected amount of each protein.
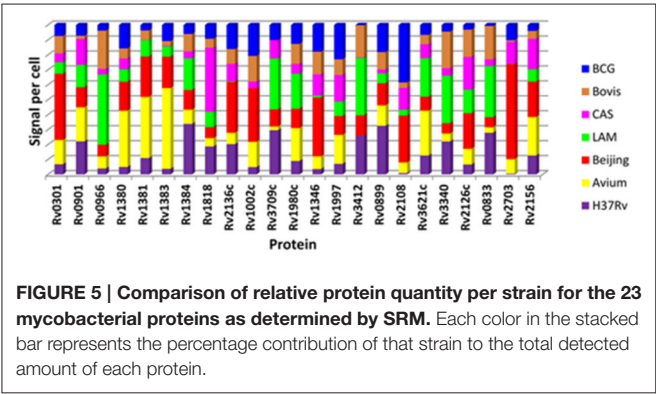
observed between strains of *M. tuberculosis* and we explored this possibility using discovery and targeted mass spectrometry techniques. In order to reveal candidate proteins that might be involved in differential pathogenicity, we aimed to compare proteomes between individual pathogenic and non-pathogenic mycobacteria, noting that both BCG and *M. avium* can cause TB-like disease in immune compromised individuals, suggesting that their pathogenicity is attenuated, not lost entirely. To underpin our intended cross-strain and cross-species proteomic comparisons, we first carried out an exhaustive mass spectrometry-based discovery proteomics analysis of

the 7 mycobacterial strains. Through use of a sophisticated bioinformatic strategy, combining data from multiple search engines, we obtained >80% coverage at the protein level for 6 out of the 7 theoretical proteomes. Combining data across the 4 *M. tuberculosis* strains, we identified a total of 3788 *M. tuberculosis* proteins with high confidence, meaning that we failed to observe only 235 out of the predicted 4023 proteins in the *M. tuberculosis* proteome. To our knowledge, this level of discovery proteomic coverage across multiple *M. tuberculosis* strains is unprecedented, although we note that recently reported SWATH-based analyses on *M. tuberculosis* H37Rv have come close to this figure (Schubert et al., 2015). Surprisingly, although the 23 proteins subsequently quantified by SRM were initially observed only in the *M. tuberculosis* strains by discovery MS analysis, they were in fact all identified in all 7 strains by SRM analysis, albeit with relative quantifications that correlated to a large degree with our discovery data. Furthermore, SRM analysis of 45 predicted *M. tuberculosis* proteins that had not been observed in our discovery MS analysis of any of the 7 mycobacterial strains revealed that half were in fact expressed in each of the *M. tuberculosis* strains (data not shown), presumably reflecting low absolute expression levels for those proteins, below the detection limit for discovery MS. Taken together, our data suggests that the total expressed complement of proteins is remarkably similar in the different clinical strains of *M. tuberculosis* and moreover that virtually the entire *M. tuberculosis* proteome is expressed in all strains,

at least under optimal *in vitro* conditions. However, our data also clearly demonstrates that significant, quantitative differences in expression levels exist between strains which may directly influence the phenotype of these strains. While it is possible that protein quantity does not track with the enzyme activity of that protein in a cell, due to allosteric effects or post-translational modifications, enzyme activity was not measured in our study.

Cross-strain and cross-species comparisons of our semi-quantitative discovery proteomics data were limited only by the relatively poor ortholog mapping found between *M. avium* and *M. tuberculosis* H37Rv and enabled identification of 168 proteins that were originally observed only in the *M. tuberculosis* strains and were therefore considered candidates that might contribute to the differential virulence of these mycobacterial strains. However, we were conscious that our proteomic data had been generated under one set of culture conditions that were likely far removed from the true host environment in TB disease. We therefore cross-correlated our proteomic data with 771 gene expression models deposited in the TBDB, covering a wide range of different *in vitro* culture conditions and exogenous stresses on *M. tuberculosis* that can be thought of as mimicking various aspects of the host environment (e.g., macrophage infection; hypoxia; starvation; etc.). The genes for 23 of the 168 proteins were found to be significantly up- or down-regulated in macrophage-based and related *in vitro* gene expression models of TB disease and we therefore carried out quantitative analysis of their protein expression *in vitro* across the 7 strains, focussing particularly on differential expression in the Beijing and LAM lineages that are known to have particularly virulent clinical phenotypes (Pillay and Sturm, 2007; Cowley et al., 2008).

One of *M. tuberculosis*'s many features as a pathogen is its ability to evade the host innate and acquired immune responses such that it is capable of attaining latency and can potentially remain relatively quiescent in alveolar macrophages for decades (Flynn and Chan, 2003). Here, we identified four proteins which modulate the host immune response which are significantly more abundant in the Beijing strain compared to other strains and may therefore have functional significance in conferring virulence in *M. tuberculosis*—Rv2136c, Rv1002c, Rv2703, and Rv2108. Mutant *M. tuberculosis* with insertionally inactivated Rv2136c, a known virulence factor of the MTBC (Forrellad et al., 2013), has severe hypersensitivity to acid and a number of other stresses (Vandal et al., 2009). Rv2136c (uppP) is an undecaprenyl pyrophosphate phosphatase which recycles undecaprenyl pyrophosphate back to undecaprenyl phosphate so that it can act again as a receptor for the UDP-MurNAc-pentapeptide to make C55-PP-MurNAc-pentapeptide (lipid 1). The antibiotic Bacitracin inhibits peptidoglycan synthesis by sequestering undecaprenyl diphosphate, thereby reducing the pool of lipid carrier available, whilst increased expression of uppP provides resistance; by extension, Rv2703c might therefore be involved in virulence by speeding up the recycling of key lipid intermediates and hence cell wall biosynthesis, thus conferring a selectable advantage on the virulent Beijing strains. Similarly, sigma factor A (sigA), Rv2703, is the primary sigma factor in this bacterium and is essential for growth. Increased initiation of transcription,

and thus RNA processing capacity, may therefore be another mechanism by which this strain has achieved hypervirulence, perhaps coupled to the observed increased expression of several proteins involved in pyrimidine biosynthesis (pyrB, pyr C, CarA). Although the function of Rv1002c is unknown, it is essential for growth in H37Rv (Sassetti et al., 2003), whereas the PPE family protein PPE36 (Rv2108) has no known function and is non-essential for growth in H37Rv. Both of these proteins therefore represent attractive targets for further investigation.

Once inside the host macrophage, *M. tuberculosis* becomes dependent on the intracellular environment for sources of carbon (McKinney et al., 2000; Eisenreich et al., 2010) and iron. The acyl-CoA dehydrogenase, MbtN (Rv1346), is involved in the production of mycobactins which are thought to be vital for the acquisition of iron within the macrophage and are therefore considered to be virulence factors (De Voss et al., 2000). It is notable therefore that MbtN protein was significantly more abundant in the Beijing strain, suggesting that this strain's capacity to acquire iron intracellularly may be superior. The capacity to produce essential amino acids in the host may also provide a selective advantage in terms of virulence, as demonstrated by the increased relative abundance of the O-acetylhomoserine sulfhydrylase MetC (Rv3340) and aspartate kinase Ask (Rv3709c) in the LAM strain.

Another important feature of intracellular *M. tuberculosis* in the persistent phase is the toxin/antitoxin system, of which *M. tuberculosis* has a remarkable 79 encoded loci (Sala et al., 2014). Here we found that the possible toxin VapC2 (Rv0301) was significantly more abundant in the Beijing strain than other clinical strains. Vap C has been reported to suppress translation by hydrolysis of mRNA, so its increased expression in the Beijing strain may represent an evolutionary advantage by providing an efficient means to erase previous transcriptional profiles, thus allowing *M. tuberculosis* to rapidly reprogram the proteome and hence change the metabolic state of the cell in response to rapidly changing external stresses during the bacterium's host-based lifecycle.

Finally, proteins with unknown function are clear targets for further investigation—for example the PE_PGRS and PPE family proteins which are known virulence factors yet currently have no functional categorization. Both PE_PGRS13 (Rv0833) and PPE65 (Rv3621c) as well as conserved hypothetical proteins Rv0966 and Rv3412 are abundant in the LAM strain. Interestingly, Rv0901—a possible exported or membrane protein with unknown function—is abundant in all measured strains except for *M. bovis* and BCG, both of which have attenuated pathogenicity. The loss of this protein could therefore alter the pathogenicity of the bacterium, indicating that this protein is a potential therapeutic target.

An important caveat for any quantitative *in vitro* study on *M. tuberculosis* that aims to correlate gene or protein expression levels with *in vivo* clinical phenotypes is that significant differences in expression observed under specific *in vitro* conditions may not accurately reflect the situation at the site of disease due to altered environmental influences on expression. This is further compounded by the fact that it is not currently technically possible to isolate sufficient *M.*

tuberculosis bacilli from the site of disease in a human lung for a discovery proteomics experiment and by the fact that the clinical definitions of "virulence" and "pathogenicity" are themselves largely qualitative. In order to mitigate this caveat, we therefore correlated our quantiative proteomic data with over 700 gene expression models of TB disease—themselves acquired under a number of different *in vitro* conditions that each mimic in some way aspects of the stress likely to be experienced by *M. tuberculosis* at the site of disease—in order to provide a logical means to infer biological significance from *in vitro* data with greater confidence. A testable prediction from the *in vitro* quantitative proteomic data presented here is thus that the observed differential expression of specific mycobacterial proteins across these 7 strains when cultured under a common set of environmental conditions will affect clinical phenotype *in vivo*. However, the true role of these proteins in virulence will need to be validated in due course by targeted analysis of limiting numbers of *M. tuberculosis* bacilli isolated from the site of disease.

CONCLUSION

Through our combined discovery- and quantitative proteomic analysis of differential protein expression in 7 mycobacterial strains of varying pathogenicity and virulence, we have uncovered previously unknown, statistically significant quantitative differences in the expression of numerous proteins which begin to shed new light on differential virulence in *M. tuberculosis* strains. In particular, our data suggests strain specific bacterial fitness in the W-Beijing lineage, including: the ability to rapidly remodel the *M. tuberculosis* proteome in response to altered environments; up-regulation of key sigma factors to support rapid transcriptional responses; up-regulation of enzymes involved in pyrimidine biosynthesis and cell wall biosynthesis to promote rapid growth; enhanced mycobactin biosynthesis to promote iron scavenging in the host. These individually selectable traits may then conceivably work together to provide the W-Beijing lineage with an enhanced ability to establish primary infection and active TB disease in a new host. These are testable hypotheses and further research is underway on this now.

AUTHOR CONTRIBUTIONS

JP carried out research, wrote initial draft. BC analyzed data, refined and rewrote initial draft to produce final draft. GG, SD, ER, NM, NS, LM assisted with data analysis and JB is principal investigator.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2016.00813

REFERENCES

Baker, L., Brown, T., Maiden, M. C., and Drobniewski, F. (2004). Silent nucleotide polymorphisms and a phylogeny for *Mycobacterium tuberculosis*. *Emerg. Infect. Dis.* 10, 1568–1577. doi: 10.3201/eid1009.040046

Bifani, P. J., Mathema, B., Kurepina, N. E., and Kreiswirth, B. N. (2002). Global dissemination of the *Mycobacterium tuberculosis* W-Beijing family strains. *Trends Microbiol.* 10, 45–52. doi: 10.1016/S0966-842X(01)02277-6

Brosch, M., Yu, L., Hubbard, T., and Choudhary, J. (2009). Accurate and sensitive peptide identification with Mascot Percolator. *J. Proteome Res.* 8, 3176–3181. doi:10.1021/pr800982s

Cottrell, J. S., and London, U. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20, 3551–3567.

Cowley, D., Govender, D., February, B., Wolfe, M., Steyn, L., Evans, J., et al. (2008). Recent and rapid emergence of W-Beijing strains of *Mycobacterium tuberculosis* in Cape Town, South Africa. *Clin. Infect. Dis.* 47, 1252–1259. doi: 10.1086/592575

Degroeve, S., and Martens, L. (2013). MS2PIP: a tool for MS/MS peak intensity prediction. *Bioinformatics* 29, 3199–3203. doi: 10.1093/bioinformatics/btt544

Desforges, J. F., and Horsburgh, C. R. Jr. (1991). *Mycobacterium avium* complex infection in the acquired immunodeficiency syndrome. *N. Engl. J. Med.* 324, 1332–1338.

De Voss, J. J., Rutter, K., Schroeder, B. G., Su, H., Zhu, Y., and Barry, C. E. (2000). The salicylate-derived mycobactin siderophores of *Mycobacterium tuberculosis* are essential for growth in macrophages. *Proc. Natl. Acad. Sci. U.S.A.* 97, 1252–1257. doi: 10.1073/pnas.97.3.1252

Eisenreich, W., Dandekar, T., Heesemann, J., and Goebel, W. (2010). Carbon metabolism of intracellular bacterial pathogens and possible links to virulence. *Nat. Rev. Microbiol.* 8, 401–412. doi: 10.1038/nrmicro2351

Flynn, J. L., and Chan, J. (2003). Immune evasion by *Mycobacterium tuberculosis*: living with the enemy. *Curr. Opin. Immunol.* 15, 450–455. doi: 10.1016/S0952-7915(03)00075-X

Forrellad, M. A., Klepp, L. I., Gioffré, A., Sabio y García, J., Morbidoni, H. R., de la Paz Santangelo, M., et al. (2013). Virulence factors of the *Mycobacterium tuberculosis* complex. *Virulence* 4, 3–66. doi: 10.4161/viru.22329

Gagneux, S., and Small, P. M. (2007). Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. *Lancet Infect. Dis.* 7, 328–337. doi: 10.1016/S1473-3099(07)70108-1

Halstrom, S., Price, P., and Thomson, R. (2015). Review: environmental mycobacteria as a cause of human infection. *Int. J. Mycobacteriol.* 4, 81–91. doi: 10.1016/j.ijmyco.2015.03.002

Hanekom, M., Gey van Pittius, N. C., McEvoy, C., Victor, T. C., Van Helden, P. D., and Warren, R. M. (2011). *Mycobacterium tuberculosis* Beijing genotype: a template for success. *Tuberculosis (Edinb).* 91, 510–523. doi: 10.1016/j.tube.2011.07.005

Homolka, S., Post, E., Oberhauser, B., George, A. G., Westman, L., Dafae, F., et al. (2008). High genetic diversity among *Mycobacterium tuberculosis* complex strains from Sierra Leone. *BMC Microbiol.* 8:103. doi: 10.1186/1471-2180-8-103

Malik, A. N., and Godfrey-Faussett, P. (2005). Effects of genetic variability of *Mycobacterium tuberculosis* strains on the presentation of disease. *Lancet Infect. Dis.* 5, 174–183. doi: 10.1016/s1473-3099(05)01310-1

McIlwain, S., Tamura, K., Kertesz-Farkas, A., Grant, C. E., Diament, B., Frewen, B., et al. (2014). Crux: rapid open source protein tandem mass spectrometry analysis. *J. Proteome Res.* 13, 4488–4491. doi: 10.1021/pr500741y

McKinney, J. D., zu Bentrup, K. H., Muñoz-Elias, E. J., Miczak, A., Chen, B., Chan, W.-T., et al. (2000). Persistence of *Mycobacterium tuberculosis* in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. *Nature* 406, 735–738. doi: 10.1038/35021074

Merker, M., Blin, C., Mona, S., Duforet-Frebourg, N., Lecher, S., Willery, E., et al. (2015). Evolutionary history and global spread of the *Mycobacterium tuberculosis* Beijing lineage. *Nat. Genet.* 47, 242–249. doi: 10.1038/ng.3195

Nicol, M. P., and Wilkinson, R. J. (2008). The clinical consequences of strain diversity in *Mycobacterium tuberculosis*. *Trans. R. Soc. Tropical Med. Hygiene* 102, 955–965. doi: 10.1016/j.trstmh.2008.03.025

Parwati, I., van Crevel, R., and van Soolingen, D. (2010). Possible underlying mechanisms for successful emergence of the *Mycobacterium tuberculosis* Beijing genotype strains. *Lancet Infect. Dis.* 10, 103–111. doi: 10.1016/S1473-3099(09)70330-5

Pillay, M., and Sturm, A. W. (2007). Evolution of the extensively drug-resistant F15/LAM4/KZN strain of *Mycobacterium tuberculosis* in KwaZulu-Natal, South Africa. *Clin. Infect. Dis.* 45, 1409–1414. doi: 10.1086/522987

Ribeiro, S. C., Gomes, L. L., Amaral, E. P., Andrade, M. R., Almeida, F. M., Rezende, A. L., et al. (2014). *Mycobacterium tuberculosis* strains of the modern sublineage of the Beijing family are more likely to display increased virulence than strains of the ancient sublineage. *J. Clin. Microbiol.* 52, 2615–2624. doi: 10.1128/JCM.00498-14

Sala, A., Bordes, P., and Genevaux, P. (2014). Multiple toxin-antitoxin systems in *Mycobacterium tuberculosis*. *Toxins* 6, 1002–1020. doi: 10.3390/toxins6031002

Sarkar, R., Lenders, L., Wilkinson, K. A., Wilkinson, R. J., and Nicol, M. P. (2012). Modern lineages of *Mycobacterium tuberculosis* exhibit lineage-specific patterns of growth and cytokine induction in human monocyte-derived macrophages. *PLoS ONE* 7:e43170. doi: 10.1371/journal.pone.0043170

Sassetti, C. M., Boyd, D. H., and Rubin, E. J. (2003). Genes required for mycobacterial growth defined by high density mutagenesis. *Mol. Microbiol.* 48, 77–84. doi: 10.1046/j.1365-2958.2003.03425.x

Schubert, O. T., Ludwig, C., Kogadeeva, M., Zimmermann, M., Rosenberger, G., Gengenbacher, M., et al. (2015). Absolute proteome composition and dynamics during dormancy and resuscitation of *Mycobacterium tuberculosis*. *Cell Host Microbe* 18, 96–108. doi: 10.1016/j.chom.2015.06.001

Schubert, O. T., Mouritsen, J., Ludwig, C., Röst, H. L., Rosenberger, G., Arthur, P. K., et al. (2013). The Mtb proteome library: a resource of assays to quantify the complete proteome of *Mycobacterium tuberculosis*. *Cell Host Microbe* 13, 602–612. doi: 10.1016/j.chom.2013.04.008

Shimono, N., Morici, L., Casali, N., Cantrell, S., Sidders, B., Ehrt, S., et al. (2003). Hypervirulent mutant of *Mycobacterium tuberculosis* resulting from disruption of the mce1 operon. *Proc. Natl. Acad. Sci. U.S.A.* 100, 15918–15923. doi: 10.1073/pnas.2433882100

Shteynberg, D., Nesvizhskii, A. I., Moritz, R. L., and Deutsch, E. W. (2013). Combining results of multiple search engines in proteomics. *Mol. Cell. Proteomics* 12, 2383–2393. doi: 10.1074/mcp.R113.027797

Sun, S., Meyer-Arendt, K., Eichelberger, B., Brown, R., Yen, C.-Y., Old, W. M., et al. (2007). Improved validation of peptide MS/MS assignments using spectral intensity prediction. *Mol. Cell Proteomics* 6, 1–17. doi: 10.1074/mcp.M600320-MCP200

Vandal, O. H., Roberts, J. A., Odaira, T., Schnappinger, D., Nathan, C. F., and Ehrt, S. (2009). Acid-susceptible mutants of *Mycobacterium tuberculosis* share hypersusceptibility to cell wall and oxidative stress and to the host environment. *J. Bacteriol.* 191, 625–631. doi: 10.1128/JB.00932-08

Vizcaíno, J. A., Csordas, A., del-Toro, N., Dianes, J. A., Griss, J., Lavidas, I., et al. (2016). 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* 44, D447–D456. doi: 10.1093/nar/gkv1145

Wang, J., and Behr, M. A. (2014). Building a better bacillus: the emergence of *Mycobacterium tuberculosis*. *Front. Microbiol.* 5:139. doi: 10.3389/fmicb.2014.00139

Warner, D. F., Koch, A., and Mizrahi, V. (2015). Diversity and disease pathogenesis in *Mycobacterium tuberculosis*. *Trends Microbiol.* 23, 14–21. doi: 10.1016/j.tim.2014.10.005

The polar and lateral flagella from *Plesiomonas shigelloides* are glycosylated with legionaminic acid

Susana Merino[1], Eleonora Aquilini[1], Kelly M. Fulton[2], Susan M. Twine[2] and Juan M. Tomás[1]*

[1] *Departamento de Microbiología, Facultad de Biología, Universidad de Barcelona, Barcelona, Spain,* [2] *National Research Council, Ottawa, ON, Canada*

Plesiomonas shigelloides is the unique member of the *Enterobacteriaceae* family able to produce polar flagella when grow in liquid medium and lateral flagella when grown in solid or semisolid media. In this study on *P. shigelloides* 302-73 strain, we found two different gene clusters, one exclusively for the lateral flagella biosynthesis and the other one containing the biosynthetic polar flagella genes with additional putative glycosylation genes. *P. shigelloides* is the first *Enterobacteriaceae* were a complete lateral flagella cluster leading to a lateral flagella production is described. We also show that both flagella in *P. shigelloides* 302-73 strain are glycosylated by a derivative of legionaminic acid (Leg), which explains the presence of Leg pathway genes between the two polar flagella regions in their biosynthetic gene cluster. It is the first bacterium reported with *O*-glycosylated Leg in both polar and lateral flagella. The flagella *O*-glycosylation is essential for bacterial flagella formation, either polar or lateral, because gene mutants on the biosynthesis of Leg are non-flagellated. Furthermore, the presence of the lateral flagella cluster and Leg *O*-flagella glycosylation genes are widely spread features among the *P. shigelloides* strains tested.

Keywords: *Plesiomonas shigelloides*, polar flagella, lateral flagella, *O*-glycosylation, legionaminic acid

Introduction

Plesiomonas shigelloides is a Gram-negative bacilli flagellated bacterium. This facultative anaerobic bacterium is ubiquitous, has been isolated from different water sources (freshwater or surface water), and animals (wild and domestic; Farmer et al., 1992). In humans, *P. shigelloides* is associated with diarrheal disease in humans (Brenden et al., 1988). Sometimes could also be the cause of gastroenteritis, including acute secretory gastroenteritis (Mandal et al., 1982), an invasive shigellosis-like disease (McNeeley et al., 1984), and a cholera-like illness (Tsukamoto et al., 1978). Extra intestinal infections, such as meningitis, bacteremia (Billiet et al., 1989), and pseudoappendicitis (Fischer et al., 1988), are also associated with *P. shigelloides* infection. Of particular concern are the severe cases of meningitis and bacteremia (Fujita et al., 1994) caused by *P. shigelloides*.

Plesiomonas shigelloides was initially classified in the *Vibrionaceae* family; however, molecular studies by Martinez-Murcia et al. (1992) indicated that is related to the enterobacterial genus *Proteus* phylogenetically. Huys and Sings (1999) during studies of *Aeromonas* spp. genotyping using by the amplified fragment length polymorphism found that *P. shigelloides* clearly falls out of the major *Aeromonas* cluster. According to these features the genus *Plesiomonas* was reclassified to

the family *Enterobacteriaceae,* being the only oxidase-positive member of this family (Garrity et al., 2001). In order to distinguishing different strains of *P. shigelloides*, two major serotyping schemes, one based on *O*-antigen lipopolysaccharide (O) and the other one on flagellar (H) antigens. With a total of 102 somatic antigens and 51 flagellar antigens recognized (Aldova and Shimada, 2000).

The flagella biosynthesis, in terms of resources and energy, is a costly commitment for the bacterium (Macnab, 1996). The flagella number is variable, and the distribution most frequently found on pathogenic bacteria are monotrichous (single flagellum) or pertitrichous (multiple flagella around the cell; Macnab, 1996). The flagella expression is dependable of the growth conditions. When grown in plates, several bacterial species produced more flagella than when they grow in liquid medium. Some species, like *Proteus mirabilis*, have been observed to show an increase in the numbers of flagella. *Vibrio parahaemolyticus*, have a single polar flagellum in liquid medium, instead when grown on solid medium, produces the polar flagellum (Fla) and peritrichous (or lateral) flagella (Laf; Allison and Hughes, 1991; Allison et al., 1992; Merino et al., 2014). Lateral flagella, were shown in about seven other *Vibrio* species (some of which evokes a disease spectrum similar to *V. parahaemolyticus*; Shinoda et al., 1992), while only a reduced number of bacterial species, including *Rhodospirillum centenum* (a purple photosynthetic bacterium; McClain et al., 2002), *Azospirillum* spp. (nitrogen-fixing rhizobacteria that colonize plants; Moens et al., 1996), *Helicobacter mustelae* (the causative agent of chronic gastritis and ulcer disease in ferrets; O'Rourke et al., 1992), *P. shigelloides* (Inoue et al., 1991), and *Aeromonas* spp. (opportunistic and gastroenteric pathogens of man; Gavín et al., 2002). Other species that show lateral flagella include *Bradyrhizobium japonicum* (Kanbe et al., 2007), *Photobacterium profundum* (Eloe et al., 2008), and *Rhodobacter sphaeroides* (Poggio et al., 2007). Furthermore, *Selenomonas ruminantium* subsp. *lactilytica* is a solely laterally flagellate bacterium (Haya et al., 2011).

Protein glycosylation is one of the most common protein post-translational modifications and consists in the covalent attachment of carbohydrates to amino acids. This mechanism was thought to occur exclusively in eukaryotes. However, protein glycosylation systems have been identified in all forms of life including prokaryotes. *N*-glycosylation is the covalent linkage to asparagine residues of carbohydrates, while *O*-glycosylation to serine or threonine residues. *O*-glycosylation in bacteria has been largely reviewed recently (Iwashkiw et al., 2013). As more bacterial genomes are now available together with bioinformatic analysis coupled with functional analysis, the elucidation of glycosylation pathways achieved increasing, including the identification of many genes that participate in flagellin glycosylation (Iwashkiw et al., 2013). The number of *O*-glycosylation genes involved is diverse in each bacterial species (Goon et al., 2003; Schirm et al., 2003; Faridmoayer et al., 2007; Iwashkiw et al., 2012). In spite of these advances, the knowledge of glycans structure and composition of which modify from Gram-negative bacteria flagellins is restricted to certain species and has been observed to be strain-dependent [as reviewed by Merino and Tomás (2014)].

In this work we study the genetics of *P. shigelloides* flagella (polar and lateral), and their flagella post-translational modifications, the first report of flagellar glycosylation in enteric bacteria.

Materials and Methods

Bacterial Strains, their Growth Conditions, and Plasmids Used

The bacterial strains, as well as the plasmids used, are listed on **Table 1**. Bacteria were grown in TSB broth and TSA medium supplemented if necessary with kanamycin (25 μg/ml), tetracycline (20 μg/ml), and rifampicin (100 μg/ml) when needed.

MiniTn5Km-1 Mutagenesis

Conjugal transfer of miniTn5Km-1 transposition element from *Escherichia coli* S17-1λ*pir*Km-1 to *P. shigelloides* 302-73R (wild type strain rifampicin-resistant) was carried out in a conjugal drop as previously described (Aquilini et al., 2013).

Construction of a *P. shigelloides* Genomic Library

Plesiomonas shigelloides strain 302-73 (serotype O1) genomic DNA was isolated and partially digested with *Sau3A* as described by Sambrook et al. (1989). The *P. shigelloides* strain 302-73 genomic library, using cosmid pLA2917 (Allen and Hanson, 1985), was performed as described (Guasch et al., 1996).

General DNA Methods

General DNA manipulations were done essentially as previously described described (Sambrook et al., 1989; Aquilini et al., 2014).

Southern Blot Hybridizations

Southern blotting was performed by capillary transfer (Sambrook et al., 1989) from the gel to a nylon membrane (Hybond N1, Amersham). Probe labeling, hybridization, and detection were carried out as previously described (Aquilini et al., 2014) using the enhanced chemiluminescence labeling and detection system (Amersham) according to the manufacturer's instructions.

DNA Sequencing and *In Silico* Analysis of Sequence Data

These studies were previously described (Wilhelms et al., 2013). The dideoxy-chain termination method (Sanger et al., 1977), BLAST (Altschul et al., 1997; Bateman et al., 2002), and Clustal W were used.

Complementation Studies

Complementation of the different mutants carrying the miniTn5 was done as previously described (Aquilini et al., 2013) by conjugal transfer of positive recombinant clones from the genomic library.

Antisera

Anti-*P. shigelloides* polar flagellum and lateral flagella serum were independently obtained using purified polar flagellum or lateral

TABLE 1 | Bacterial strains and plasmids used.

Strain or plasmid	Relevant characteristics[a/]	Source or reference
Escherichia coli		
DH5α	F- *endA hsdR17* (rk⁻ mk⁺) *supE44 thi-1 recA1 gyr-A96 80lacZ*	Hanahan (1983)
S17-1λpirKm1	*thi thr1 leu tonA lacY supE recA::RP4-2 (Tc::Mu)Km*ʳ λ*pir* with miniTn5Km1	De Lorenzo et al. (1990)
MC1061λpir	*thi thr1 leu6 proA2 his4 argE2 lacY1 galK2 ara14 xyl5 supE44* λ *pir*	Rubirés et al. (1997)
Plesiomonas Shigelloides		
302-73	Wild type, serotype O12:K80	Pieretti et al. (2010)
302-73R	302-73, spontaneous Rifʳ	Aquilini et al. (2013)
A	302-73*flgE*:mini-Tn5Km1 Rifʳ Kmʳ	This study
B	302-73*flhA*:mini-Tn5Km1 Rifʳ Kmʳ	This study
C	302-73*fliI*:mini-Tn5Km1 Rifʳ Kmʳ	This study
D	302-73*flgK*:mini-Tn5Km1 Rifʳ Kmʳ	This study
E	302-73*lafA*:mini-Tn5Km1 Rifʳ Kmʳ	This study
F	302-73*flhA*$_L$:mini-Tn5Km1 Rifʳ Kmʳ	This study
G	302-73*flgE*$_L$:mini-Tn5Km1 Rifʳ Kmʳ	This study
H	302-73*ptmA*:mini-Tn5Km1 Rifʳ Kmʳ	This study
I	302-73*legH*:mini-Tn5Km1 Rifʳ Kmʳ	This study
Plasmids		
pLA2917	Tcʳ, Kmʳ	Allen and Hanson (1985)
COS-FLAregI-1	pLA2917 with 20-kb chromosomal 302-73 *Sau*3A insert carrying part of the polar flagella biosynthesis region I,Tcʳ	This study
COS-LAFI	pLA2917 with 20-kb chromosomal 302-73 *Sau*3A insert carrying part of the lateral flagella biosynthesis region, Tcʳ	This study
COS-LEG	pLA2917 with 20-kb chromosomal 302-73 *Sau*3A insert carrying complete Leg biosynthesis region, Tcʳ	This study
pRK2073	Helper plasmid, Spʳ	Canals et al. (2006)
pGEM-T	PCR cloning vector, Ampʳ	Promega
pDM4	*pir* dependent with *sacAB* genes, oriR6K, CmᴿR	Milton et al. (1996)
pDM4Δ*pgmL*	pDM4 with truncated in frame *pgmL*	This study
pDM4Δ*legF*	pDM4 with truncated in frame *legF*	This study
pBAD33	Arabinose inducible expression vector, CmᴿR	ATCC
pBAD33-*pgmL*	pBAD33 with complete *pgmL*	This study
pBAD33-*legF*	pBAD33 with complete *legF*	This study

[a/] = resistant.

flagella obtained after cesium chloride, and assayed as previously described for other surface molecules (Tomás et al., 1991; Merino et al., 1992).

Motility Assays (Swarming and Swimming)

The studies were performed as previously described (Wilhelms et al., 2012). Bacterial colonies were picked with a sterile toothpick and deposited into the center of swarm agar or swim agar plate. The plates were incubated up for 16–24 h at 25°C and motility was examined by the migration of bacteria through the agar from the center toward the plate periphery. Swimming motility in liquid medium was observed by phase-contrast microscopy at a magnification of x 400 as previously (Wilhelms et al., 2012).

Transmission Electron Microscopy (TEM)

Transmission electron microscopy (TEM) studies were performed as previously described (Wilhelms et al., 2012).

Flagella Purification

Plesiomonas shigelloides strain 302-73 was grown in TSB for the polar flagellum purification. For the isolation of lateral flagella the strains were grown on TSA and recovered with 100 mM Tris (pH = 7.8). Purified flagella were isolated as previously described (Merino et al., 2014).

Cytoplasmic Fraction

Plesiomonas shigelloides cytoplasmic fraction from strain 302-73 cells grown in TSB at 37°C was obtained as previously described (Wilhelms et al., 2012).

Immunological Methods

Western blot of cytoplasmic fractions or purified flagella was performed as previously described (Wilhelms et al., 2012). Immunoblotting was carried out as described (Towbin and Gordon, 1984) using specific anti-polar or lateral flagellins polyclonal serum (Canals et al., 2006; 1:2000).

Electrospray Liquid Chromatography Mass Spectrometry

Mass spectrometry studies of intact flagellin proteins were carried out using 1 µg or less of protein, as described in our previous work (Wilhelms et al., 2012). Briefly, purified flagellin samples were injected onto a protein microtrap (Michrom Bioresources Inc., Auburn, CA, USA) connected to a gradient HPLC pump

(Agilent 1100 HPLC). To resolve the proteins, a gradient of 5–60% solvent B (1 mL/min) over 60 min was used, where Solvent A was 0.1% formic acid in HPLC grade water and solvent B was 0.1% formic acid in acetonitrile. A precolumn splitter was used to direct ~35 µl/min of the HPLC mobile phase through the trap or column and into the electrospray interface of the QTOF2 (Waters, Milford, MA, USA) or Orbitrap XL Mass Spectrometer (Thermal, CA, USA) to allow real-time monitoring of ion elution profiles. Intact masses of proteins were calculated using MaxEnt (Waters, Beverly, MA, USA) software by spectral deconvolution.

To identify potential glycopeptides, flagellin (50–200 µg) was digested and analyzed as previously described (Wilhelms et al., 2012). Unmodified peptides were identified using MASCOT (Matrix Science, London, UK) as described (Wilhelms et al., 2012). Glycopeptide MS/MS spectra were *de novo* sequenced as previously described (Wilhelms et al., 2012).

Construction of Defined in Frame Legionaminic Acid Mutants and their Complementation

The chromosomal in-frame *pgmL* and *legF* deletion mutants, 302Δ*pgmL* and 302Δ*legF*, respectively, were constructed by allelic exchange as described (Milton et al., 1996), and used

by us (Merino et al., 2014). The primers used to obtain the mutants are listed in **Table 2**. Two DNA fragments (A–B and C–D) were obtained after asymmetric polymerase chain reactions (PCRs), then were annealed at their overlapping region, and a single DNA fragment obtained after PCR using primers A and D. pDM4Δ*pgmL* and pDM4Δ*legF* plasmids were obtained as previously described (Merino et al., 2014).These plasmids were transferred by triparental matings using the *E. coli* MC1061 (λ*pir*), the mobilizing strain *E. coli* HB101/pRK2073 and *P. shigelloides* mutant 302-73R as recipient strain. Colonies grown on plates with chloramphenicol and rifampicin, were confirmed for genome integration of vector by PCR analysis. Colonies grown rifampicin resistant (Rif[R]) and chloramphenicol sensitive (Cm[S]) after sucrose treatment, PCR confirmed for mutation were chosen.

Plasmids pBAD33-*pgmL* and pBAD33-*legF* were constructed carrying the wild type genes *pgmL* and *legF* by PCR amplification of genomic DNA by using specific primer pairs and ligated to the plasmid pBAD33 from ATCC (American Type Culture Collection; see the list of primers in **Table 2**). Plasmids pBAD33-*pgmL* and pBAD33-*legF* were introduced in *E. coli* DH5α by electroporation, and then by triparental matings were introduced in the corresponding mutants. Induction or repression of genes in pBAD33 was achieved as described in ATCC.

TABLE 2 | (A) Primers used in the construction of chromosomal in-frame deletion mutants. **(B)** Primers used for mutant complementation using vector pBAD33.

A	
Primers[a,b]	**Amplified fragment**
pgmL	
A: 5′-CGC<u>GGATCC</u>GAACGCTTGAGTCGTGAGT-3′	AB (687 bp)
B: 5′-*TGTTTAAGTTTAGTGGATGGG*ACCCAGCTTCAACACAAAG-3′	
C: 5′-*CCCATCCACTAAACTTAAACA*GAAGGCGAAGATCTGGAG-3′	CD (695 bp)
D: 5′-CGC<u>GGATCC</u>TACCAATTCCACCACCAC-3′	
	AD (1403 bp)
legF	
A: 5′-GA<u>AGATCT</u>TGCCGTTGGCTACTGTC-3′	AB (684 bp)
B: 5′-*TGTTTAAGTTTAGTGGATGGG*ACCCCGAGCAAATATAAACG-3′	
C: 5′-*CCCATCCACTAAACTTAAACA*AGTCCCAAAGTCACGTTCTG-3′	CD (685 bp)
D: 5′-GA<u>AGATCT</u>ATATGCCACCAGGGCTAAC-3′	
	AD (1390 bp)

[a]*Italic letters show overlapping regions.*
[b]*Underlined letters show BamHI or BglII restriction site.*

B	
Plasmid	**Primer**
pBAD33-*pgmL*[a]	PgmL-FOR: 5′-TCCCCCGGGTACACGATGTGCAAG-3′
	PgmL-REV: 5′-GC<u>TCTAGA</u>CCACAACCTGCTGTGAC-3′
pBAD33-*legF*[b]	LegF-FOR: 5′-TCCCCCGGGCCTGAGTGGGACAAAAAT-3′
	LegF-REV: 5′-GC<u>TCTAGA</u>TCAATGTCAGCAGCAACG-3′

[a]*Primers contain SmaI(bold) and XbaI(underlined), the PCR amplified product (1496 bp) was ligated to SmaI- XbaI digested pBAD33.*
[b]*Primers contain SmaI(bold) and XbaI(underlined), the PCR amplified product (982 bp) was ligated to SmaI- XbaI digested pBAD33.*

Results

Plesiomonas shigelloides 302-73 [serogroup O1 (Pieretti et al., 2010)] grown in liquid medium or semisolid medium (swimming agar plates) showed the typical three-four flagella located in single point of one cell pole (lophotricus; **Figure 1**). However, when the agar concentration was increased, the flagellar distribution shifted from single pole to more disperse. The agar concentration seems to be involved in this change in flagella distribution. When the bacteria were grown in solid or semisolid media (swarming agar plates), a complete different flagella distribution was observed. As can be seen in **Figure 1** the flagella showed a typical peritrichous distribution over the entire cell surface.

A similar pattern of flagellar distribution with changes in growth medium was observed with 12 *P. shigelloides* strains. Among these strains eight represented five different serotypes (O1, O2, O3, O17, and O54) while four were non-serotyped strains. The source of the strains was from clinical stools (7) and fish (5), from Japan four of them, four from Spain, three from Brazil, and one from Poland.

MiniTn5Km-1 Mutagenesis

A spontaneous rifampicin-resistant *P. shigelloides* mutant (named 302-73R) derived from the wild type strain 302-73 was isolated by our group. *P. shigelloides* 302-73R showed identical pattern of flagella production as described previously for wild type strain. We selected insertional mutants, as described in Materials and Methods, and grouped by their inability to swim, to swarm, or both negative characteristics.

Among an initial screening of 2500 colonies four mutants were selected (initially named A, B, C, and D), based upon

FIGURE 1 | *Plesiomonas shigelloides* 302-73 wild type strain serotype O1. TEM from cells grown in liquid medium **(A)** and swarming agar plates **(B)**. Motility in swimming **(C)** and swarming **(D)** agar plates.

inability to swim but retaining the ability to swarm. A further, three mutants (initially named E, F, and G) were selected based upon inability to swarm but retaining ability to swim. Lastly, two mutants (initially named H and I) were selected that were unable to swim or swarm. Mutants A, B, C, and D, when observed by EM in appropriate conditions showed lateral flagella but not polar (**Figure 2**), while mutants E, F, and G (**Figure 3**), showed polar but not lateral flagella by EM when grown in appropriate conditions. Mutants H and I were unable to produce polar or lateral flagella observed by EM in any growth conditions (**Figure 4**). The presence of a single copy of the minitransposon in their genome was determined by Southern blot analysis. We were unable to clone the minitransposon-containing DNA fragment from the mutants using methodologies that were successful in other bacteria (Aquilini et al., 2013).

Complementation of the mutants, using a cosmid based genomic library of *P. shigelloides* 302-73 (see Materials and Methods) reversed the phenotype observed, either to swim or swarm in motility plates.

Polar Flagella Mutants

We found several recombinant positive clones able to complement A, B, C, and D mutants. The complementation was studied by the recovery of swimming behavior under appropriate conditions. All complemented mutants were able to produce polar flagella when observed by EM growing in liquid conditions (**Figure 2**). Sequencing the recombinant positive clones complete inserts revealed the complete region to correspond to PLESHI_03205 to PLESHI_03505 in the complete *P. shigelloides* 302-73 genome (Piqué et al., 2013).

Polar flagella gene cluster, as shown in **Figure 5A**, are based in two gene regions (I and II) adjacent to a group of putative biosynthetic Leg genes. In region I there are several genes encoding chemotaxis proteins, including the σ^{28} factor *fliA,* cluster from *flhB* to *G, fliK* to *R, fliE* to *J, flrA* and *C,* and *flaC* to *J* (transcribed in the same direction). This region I, similar to *V. parahaemolyticus* region two by gene distribution and transcription sense, also lacks the motor genes (McCarter, 2001). Region II, downstream of the putative biosynthetic Leg genes group, contains cluster *flgP,O,T,* or *flgA,M,N* with the typical transcription sense in the different Gram-negative bacteria described, two genes encoding chemotaxis proteins, and cluster *flgB* to *L.* By gene distribution and transcription sense this region II is similar to region 1 of *V. parahaemolyticus* and *Aeromonas hydrophila* (McCarter, 2001; Canals et al., 2006).

Table 3 shows the ORFs with their predicted function based on their homology to proteins of known function. Proteins of unknown function were not included. The last gene in this region encoded an ORF (named Gt), which showed homology to domains of a glycosyltransferase. This was provisionally assigned to the polar flagella cluster and not to the putative biosynthetic Leg genes. Once the DNA fragment was completely sequenced, several primers were used to derive the DNA sequence to locate the miniTn5 [A = *flgE,* B = *flhA,* C = *fliI,* and D = *flgK* (**Figure 5A**)].
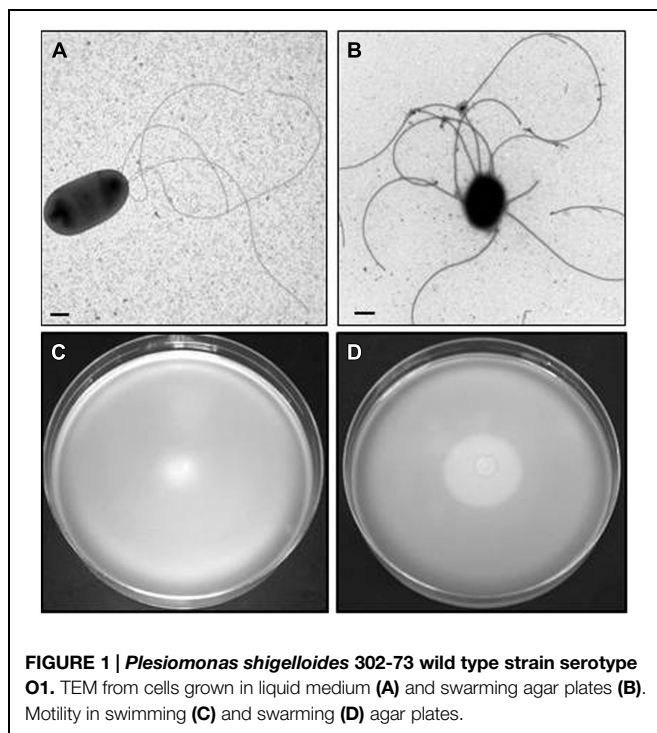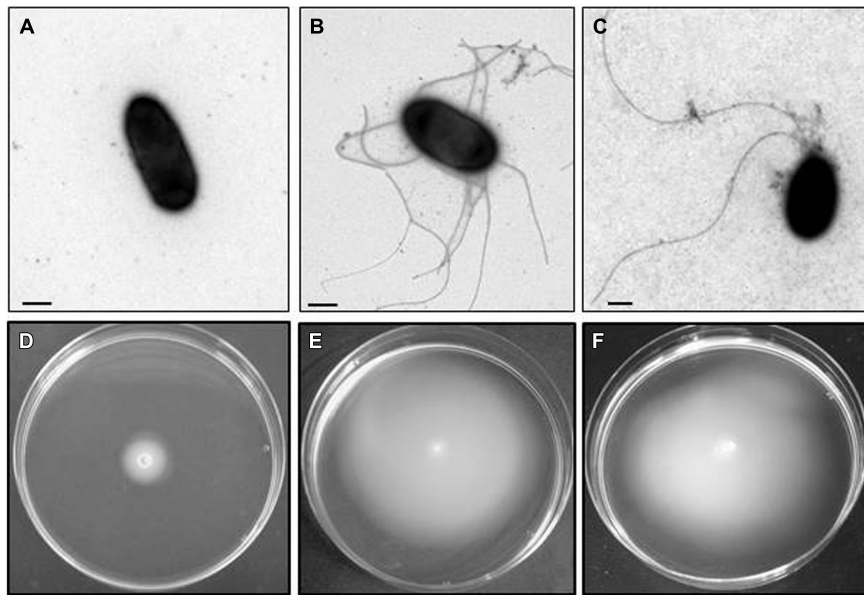
FIGURE 2 | *Plesiomonas shigelloides* A mutant (as an example for the insertional polar flagella mutants). TEM of the A mutant grown in liquid medium **(A)** and swarming agar plates **(B)** and complemented mutant with COS-FLAregI-1harboring the corresponding wild type gene grown in liquid medium **(C)**. Bar, correspond to 0.5 μm. Motility of the A mutant in swimming **(D)** and swarming **(E)** agar plates. The complemented mutant with COS-FLAregI-1harbouring the corresponding wild type gene in swimming agar plate **(F)**.
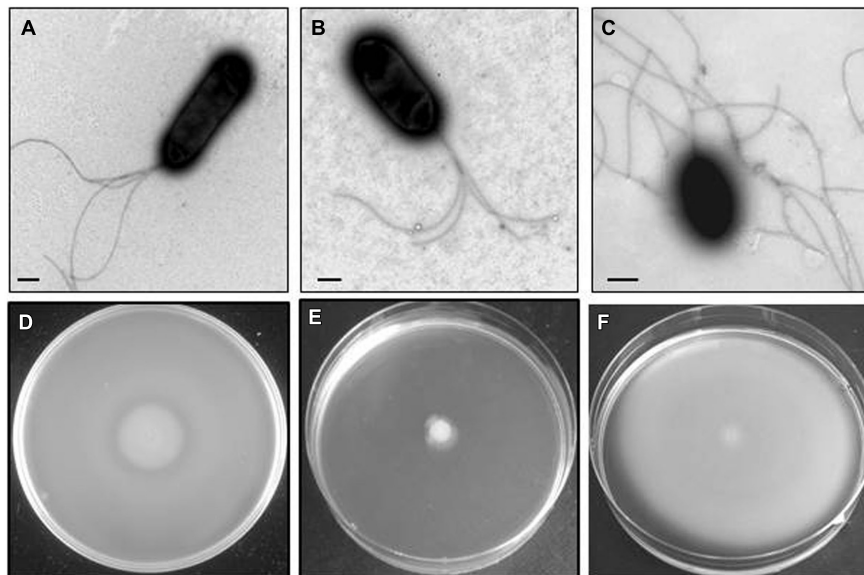
FIGURE 3 | *Plesiomonas shigelloides* E mutant (as an example for the insertional lateral flagella mutants). TEM of the E mutant grown in liquid medium **(A)** and swarming agar plates **(B)** and complemented mutant with COS-LAFI harboring the corresponding wild type gene grown in semisolid medium **(C)**. As could be observed in B the polar flagella are constitutively expressed in semisolid medium. Bar, correspond to 0.5 μm. Motility of the E mutant in swimming **(D)** and swarming **(E)** agar plates. The complemented mutant with COS-FLAregI-1harbouring the corresponding wild type gene in swarming agar plate **(F)**.

Lateral Flagella Mutants

Several recombinant positive clones complemented E, F, and G mutants separately. Some clones were observed to complement two mutants. The complementation was studied on the basis of recovery of swarming behavior on appropriate growth plates. All complemented mutants were able to produce lateral flagella when observed by EM growing in semisolid conditions (**Figure 3**). We used the same strategy previously indicated to sequence the entire

FIGURE 4 | *Plesiomonas shigelloides* **H mutant (as an example for the insertional Leg biosynthetic cluster mutants).** TEM of the H mutant grown in liquid medium **(A)** and swarming agar plates **(B)** and complemented mutant with COS-LAFI harboring the corresponding wild type gene grown in liquid **(E)** and semisolid medium **(F)**. Bar, correspond to 0.5 μm. Motility of the H mutant in swimming **(C)** and swarming **(D)** agar plates. The complemented mutant with COS-LEG harboring the corresponding wild type gene in swimming **(G)** and swarming **(H)** agar plates.

DNA region contained in the recombinant positive clones. This complete region correspond to PLESHI_07125 to PLESHI_07305 in the complete *P. shigelloides* 302-73 genome (Piqué et al., 2013).

Lateral flagella gene cluster shows 37 genes grouped in a single region (**Figure 5B**). Five typical group of genes (*lafA* to *U*; *flgB_L* to *L_L*; *flgA_L*,*M_L*,*N_L*; *fliE_L* to *J_L*; and *fliM_L* to *R_L* plus *flhB-A_L*) when compared to the most similar *A. hydrophila* AH-3 lateral flagella

region were found. All the genes were found in a unique region similar to *A. hydrophila* or enteric bacteria. In contrast, in the equivalent region in *V. parahaemolyticus* is found in two separate regions (Canals et al., 2006; Merino et al., 2006). The group of genes *fliE_L* to *J_L* and *fliM_L* to *R_L* plus *flhB-A_L* are adjacent in all the lateral flagella clusters described. The groups of genes have been shown to be transcribed in the same direction in *A. hydrophila* and divergently in *Vibrio*, enteric bacteria and *P. shigelloides* (Merino and Tomás, 2009). **Table 4** shows the ORFs with their predicted function based on their homology to proteins of known function. All the protein analogies that were from unknown or not well-established homology were discarded. Between the group of genes *flgB-L_L* and *lafA-U*, there is a gene encoding for a hypothetical protein without the classical motility accessory factors domains found in *A. hydrophyla* Maf-5. However, this encoded protein showed a minimal similarity with this Maf-5, and the gene was denoted *maf-5* (Parker et al., 2014). Once the DNA fragment was completely sequenced, we used several primers derived from the DNA sequence to locate the miniTn5 in *lafA* (E), *flhA_L* (F), and *flgE_L* (G; **Figure 5B**).

Mutants Unable to Produce Flagella
A single recombinant positive clone was observed to complement both mutants H and I as they recover swimming and swarming in plates. The complemented mutants were able to produce polar and lateral flagella when observed by EM growing in appropriate conditions (**Figure 4**). Sequencing the entire DNA region in the recombinant positive clone showed this region to contain the group of putative biosynthetic Leg genes (**Figure 5A**) between region I and II codifying for the polar flagella. This complete region corresponds to PLESHI_03365 to PLESHI_03405 in the complete *P. shigelloides* 302-73 genome (Piqué et al., 2013).**Table 5** shows the ORFs with their predicted function based on their homology to proteins of known function.

The *Campylobacter jejuni* CMP-Leg biosynthetic pathway described involves two segments: synthesis of a GDP-sugar building block and synthesis of the final CMP-nonulosonate which are linked by the *N*-acetyl transferase GlmU (Schoenhofen et al., 2009). We found all the genes encoding for the necessary two segments of the CMP-Leg biosynthetic pathway in this region besides the one encoding phosphoglucosamine mutase (PgmL) included in the first segment of the biosynthesis. Once the DNA fragment was completely sequenced, we used several primers derived from the DNA sequence to establish that the miniTn5 was located in *ptmA* (H) and *legH* (I; **Figure 5A**).

Flagella Purification
Polar flagellins were purified from the wild type strain after grown in liquid medium and a mixture of polar and lateral flagellins after grown in swarm agar plates (**Figure 6A**). Lateral flagellin was also isolated from insertion mutant A (unable to produce constitutive polar flagella with unaltered lateral flagella).

Intact Mass Analysis of Purified Flagellins
Purified polar flagellin preparations showed a well-resolved ion envelop of multiple charged protein ions, which deconvoluted into three distinct masses at 40201, 40652, and 40931 Da. The

FIGURE 5 | *Plesiomonas shigelloides* **302-73 wild type strain serotype O1 polar (A) and lateral (B) flagella clusters.** The polar flagella cluster shows two different regions adjacent to Leg biosynthetic genes. The inverted triangles (▼) indicate the position of the different insertional mutants obtained by miniTn5 mutagenesis.
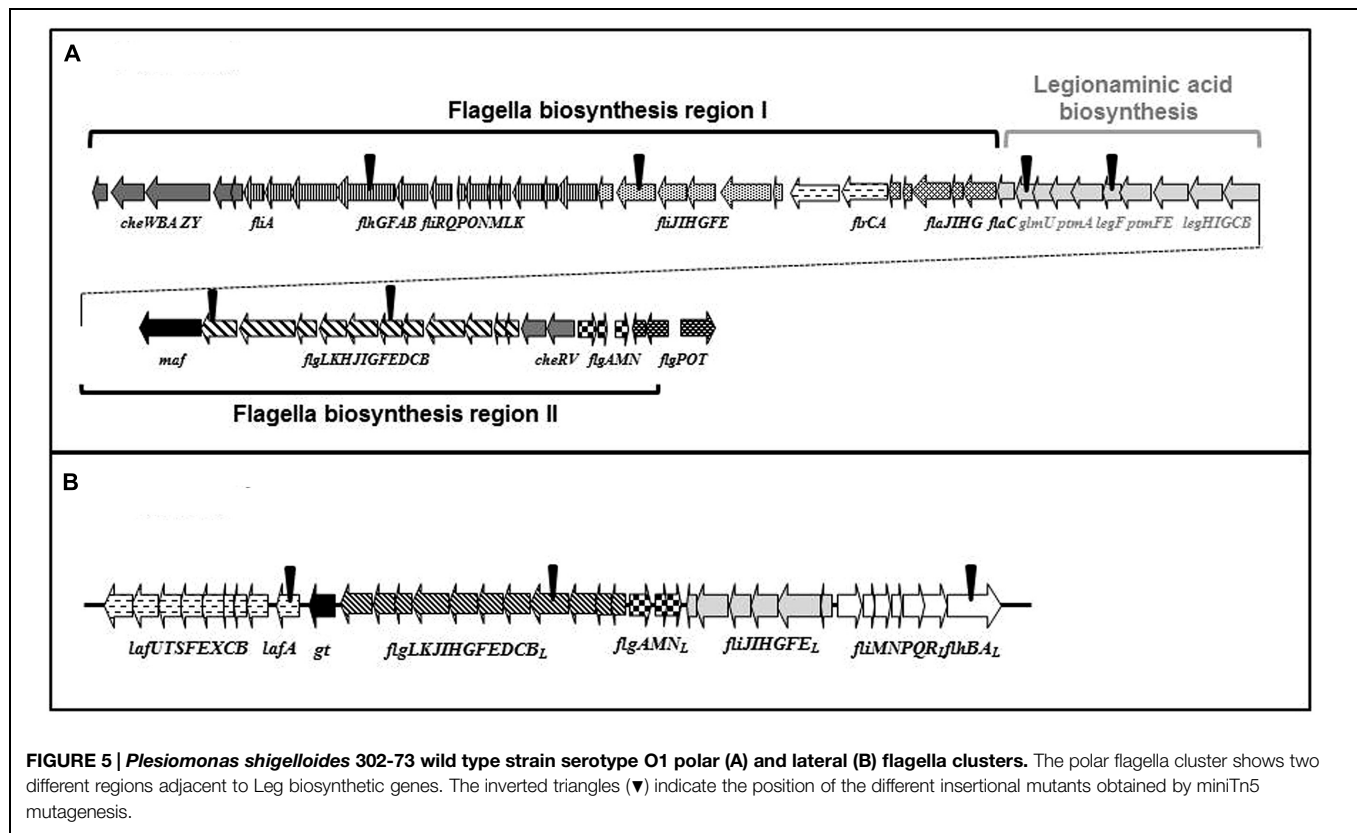
mass of the translated gene sequence for polar flagellin was 38710 Da, giving mass excesses of 1491, 1942, and 2221 Da, respectively (data not shown). During front end CID experiments of the purified polar flagellin preparation, labile glycan related ions were observed at *m/z* 359 and 317. Using increasing cone voltages, fragmentation of this ion at *m/z* 359 was observed, as shown in **Figure 7**. The fragment ions observed at *m/z* 317, 299, 281, 222, and 181 were characteristic fragment ions of nonulosonic acids, such as pseudaminic or legionaminic acid.

From the observed mass of 316.124, the top ranked plausible elemental formula was $C_{13}H_{21}N_2O_6$, suggestive that this moiety is a carbohydrate. The additional glycan ion observed at *m/z* 359, gave a top ranked plausible elemental formula $C_{15}H_{23}N_2O_8$, suggesting this species to be a nonulosonic acid with an additional of an acetyl group. An intense fragment ion was observed at *m/z* 341, most likely a loss of water from the glycan ion observed at *m/z* 359.

The preparation containing purified polar and lateral flagellins showed a more complex elution profile when HPLC separated, with two sequentially eluting protein peaks. The area under each peak was combined separately and each showed a complex ion envelope. The ion envelope of the first eluted protein deconvoluted into two distinct masses at 39325, 40678 Da. The second eluting protein ion envelope deconvoluted to give a single protein mass at 30940. It is possible that the larger MW proteins correspond to the polar flagellin and the 30 kDa protein the lateral flagellin. The A mutant that is unable to produce polar flagella

showed only this second eluting peak when grown in swarming conditions (**Figure 6A**). In each case, the measured molecular mass is greater than that of the translated gene sequence for each protein. This suggests that both polar and lateral flagellins are post-translational modified. Front end CID experiments showed almost identical profiles when compared with the polar flagellin preparation, with intense ions observed at *m/z* 359, 317. These data suggest that both polar and lateral flagellins are modified with the same nonulosonic acid sugar, with or without acetylation.

Bottom Up Mass Spectrometry Studies of Flagellins

Tandem mass spectrometry studies of tryptic digests of purified polar flagellins identified a number of unmodified peptides. *De novo* sequencing of the MS/MS data showed a number of spectra that were identified as flagellin peptides and harboring mass excess of 316 Da. Also observed was an intense ion at *m/z* 317, suggestive of a glycan oxonium ion. **Figure 8A** shows the MS/MS spectrum of the polar flagellin glycopeptide AIASLSTATINK, modified with a putative 316 Da glycan. Peptide type y and b fragment ions are annotated and confirm the peptide sequence. In addition, low *m/z* fragment ions that did not correspond to peptide type y or b ions were also observed at *m/z* 317, 299, 281, 240, 221, 196, and 181. Combined with the mass excess, glycan oxonium ion and putative glycan fragment ions, the data suggest the flagellin peptides to be modified with a legionaminic acid like glycan.

TABLE 3 | Characteristics of the _P. shigelloides_ 302-73 strain polar flagella gene regions I and II.

ORF	Protein name	Protein Size	Predicted function	Homologous protein with known function	Percentage identity/similarity
Polar flagella region I					
1	CheW	162	Chemotaxis protein	CheW (VP2225) of _Vibrio parahaemolyticus_	84/88
2	CheB	377	Chemotaxis protein	CheB-2 (AHA_1386) of _Aeromonas hydrophila_ ATCC7966	72/76
3	CheA	728	Chemotaxis protein	CheA (VP2229) of _Vibrio parahaemolyticus_	66/69
4	CheZ	241	Chemotaxis protein	CheZ (ASA_1356) of _Aeromonas salmonicida_ A449	51/59
				CheZ (VP2230) of _Vibrio parahaemolyticus_	48/54
5	CheY	127	Chemotaxis protein	CheY (AHA_1383) of _Aeromonas hydrophila_ ATCC7966	91/93
6	FliA	240	σ^{28}	FliA of _Aeromonas hydrophila_ AH-3	67/71
7	FlhG	296	Flagella number regulator	FlhG of _Vibrio alginolyticus_	72/78
8	FlhF	527	Polar flagella site determinant	FlhF (VP2234) of _Vibrio parahaemolyticus_	66/73
9	FlhA	698	Export/assembly	FlhA (VP2235) of _Vibrio parahaemolyticus_	78/81
10	FlhB	377	Export/assembly	FlhB (VP2236) of _Vibrio parahaemolyticus_	60/66
11	FliR	264	Export/assembly	FliR (VP2237) of _Vibrio parahaemolyticus_	51/61
12	FliQ	89	Export/assembly	FliQ (VP2238) of _Vibrio parahaemolyticus_	67/78
13	FliP	261	Export/assembly	FliP (VP2239) of _Vibrio parahaemolyticus_	76/82
14	FliO	139	Export/assembly	FliO of _Vibrio cholerae_	39/41
15	FliN	128	Motor switch	FliN (AHA_1373) of _Aeromonas hydrophila_ ATCC7966	72/78
16	FliM	347	Motor switch	FliM of _Aeromonas hydrophila_ AH-3	80/87
17	FliL	164	Flagella protein	FliL (AHA_1371) of _Aeromonas hydrophila_ ATCC7966	42/46
18	FliK	487	Hook length	FliK of _Vibrio cholerae_	57/68
19	FliJ	146	Export/assembly	FliJ (VP2245) of _Vibrio parahaemolyticus_	43/51
20	FliI	439	Export ATP synthase	FliI (VP2246) of _Vibrio parahaemolyticus_	78/82
21	FliH	322	Export/assembly	FliH (VP2247) of _Vibrio parahaemolyticus_	37/46
22	FliG	342	Motor switch	FliG (AHA_1366) of _Aeromonas hydrophila_ ATCC7966	75/82
23	FliF	569	M-ring	FliF (VP2249) of _Vibrio parahaemolyticus_	53/61
24	FliE	107	MS ring/rod adapter	FliE (VP2250) of _Vibrio parahaemolyticus_	58/62
25	FlrC	558	σ^{54}-dependent two-components response regulator	FlaM (VP2251) of _Vibrio parahaemolyticus_	61/65
26	FlrA	509	σ^{54}-dependent flagella regulator	FlrA of _Vibrio cholerae_	58/65
27	FlaJ	134	Chaperone	FlaJ (VP2254) of _Vibrio parahaemolyticus_	63/69
28	FlaI	94	Flagella rod protein	FlaI (VP2255) of _Vibrio parahaemolyticus_	38/50
29	FlaH	446	Hook-associated protein-2	FlaH (VP2256) of _Vibrio parahaemolyticus_	36/45
30	FlaG	132	Filament length control	FlaG of _Vibrio alginolyticus_	36/41
31	FlaC	377	Flagellin	FlaC (VP0788) of _Vibrio parahaemolyticus_	51/55
Polar flagella region II					
1	Gt	691	Glicosyltransthase	BRAO375_790043 of _Bradyrhizobium_ sp.	36/42
				AZOBR_p1140113 of _Azospirillum brasilense_	32/40
2	FlgL	417	Hook-associated protein 3	FlgL of _Vibrio cholerae_	34/42
3	FlgK	639	Hook-associated protein 1	FlgK (VP0785) of _Vibrio parahaemolyticus_	35/41
4	FlgH	230	L-ring	FlgH of _Vibrio cholerae_	56/63
5	FlgJ	322	Peptidoglycan hydrolase	FlgJ (VP0784) of _Vibrio parahaemolyticus_	58/66
6	FlgI	355	P-ring	FlgI (VP0783) of _Vibrio parahaemolyticus_	73/81
7	FlgG	262	Rod	FlgG of _Vibrio cholerae_	69/76
8	FlgF	248	Rod	FlgF (AHA_2838) of _Aeromonas hydrophila_ ATCC7966	56/74
9	FlgE	431	Hook	FlgE (VP0778) of _Vibrio parahaemolyticus_	52/58
10	FlgD	305	Rod	FlgD of _Vibrio cholerae_	49/56
11	FlgC	137	Rod	FlgC (VP0776) of _Vibrio parahaemolyticus_	74/80
12	FlgB	136	Rod	FlgB (VP0775) of _Vibrio parahaemolyticus_	59/62
13	CheR	278	Chemotaxis	CheR (VP0774) of _Vibrio parahaemolyticus_	70/76
14	CheV	313	Chemotaxis	CheV (AHA_2844) of _Aeromonas hydrophila_ ATCC7966	73/81
15	FlgA	214	P-ring assembly	FlgA of _Vibrio cholerae_	39/48
16	FlgM	105	Anti-σ^{28}	FlgM of _Vibrio cholerae_	44/51
17	FlgN	139	Chaperone	FlgN of _Vibrio alginolyticus_	40/45
18	FlgP	151	Flagella lipoprotein	FlgP of _Vibrio cholerae_	51/55
19	FlgO	267	Flagella lipoprotein	FlgO (VP0768) of _Vibrio parahaemolyticus_	48/55
20	FlgT	391	Flagella protein	FlgT (VP0767) of _Vibrio parahaemolyticus_	39/45

TABLE 4 | Characteristics of the *P. shigelloides* 302-73 strain lateral flagella cluster.

ORF	Protein name	Protein size	Predicted function	Homologous protein with known function	Percentage identity/similarity
1	LafU	455	Proton motor	LafU of *Aeromonas hydrophila* AH-3	41/62
2	LafT	284	Proton motor	LafT (VPA1556) of *Vibrio parahaemolyticus*	49/65
3	LafS	249	σ^{28}	LafS (VPA1555) of *Vibrio parahaemolyticus*	45/64
4	LafF	158	Unknown	LafF (VPA1554) of *Vibrio parahaemolyticus*	30/52
5	LafE	404	Hook length control	LafE (VPA1553) of *Vibrio parahaemolyticus*	42/67
6	LafX	96	Chaperone	LafD (VPA1552) of *Vibrio parahaemolyticus*	21/37
7	LafC	131	Chaperone	LafC of *Aeromonas hydrophila* AH-3	52/68
8	LafB	438	Hook-associated protein 2	LafB of *Aeromonas hydrophila* AH-3	31/49
9	LafA	275	Lateral Flagellin	LafA (VPA1548) of *Vibrio parahaemolyticus*	49/65
10	Maf-5	349	Motility accessory factor	Maf-5 of *Campylobacter jejuni* subsp. *jejuni* 00-2415	25/43
				Maf-5 of *Aeromonas hydrophila* AH-3	15/30
11	FlgL$_L$	300	Hook-associated protein 3	FglL$_L$ of *Aeromonas hydrophila* AH-3	38/43
12	FlgK$_L$	467	Hook-associated protein 1	LfgK (VPA0273) of *Vibrio parahaemolyticus*	34/40
13	FlgJ$_L$	328	Peptidoglycan hydrolase	LfgJ (VPA0272) of *Vibrio parahaemolyticus*	45/51
14	FlgI$_L$	364	P-ring	FglI$_L$ of *Aeromonas hydrophila* AH-3	66/72
15	FlgH$_L$	219	L-ring	FglH$_L$ of *Aeromonas hydrophila* AH-3	58/63
16	FlgG$_L$	261	Rod	LfgG (VPA0269) of *Vibrio parahaemolyticus*	68/74
17	FlgF$_L$	241	Rod	FglF$_L$ of *Aeromonas hydrophila* AH-3	54/61
18	FlgE$_L$	391	Hook	FglE$_L$ of *Aeromonas hydrophila* AH-3	44/50
19	FlgD$_L$	243	Rod	LfgD (VPA0266) of *Vibrio parahaemolyticus*	39/49
20	FlgC$_L$	140	Rod	FglC$_L$ of *Aeromonas hydrophila* AH-3	58/64
21	FlgB$_L$	125	Rod	LfgB (VPA0264) of *Vibrio parahaemolyticus*	55/59
22	FlgA$_L$	231	P-ring assembly	LfgA (VPA0263) of *Vibrio parahaemolyticus*	43/52
23	FlgM$_L$	91	Anti-σ^{28}	LfgM (VPA0262) of *Vibrio parahaemolyticus*	32/35
24	FlgN$_L$	142	Chaperone	FlgN$_L$ of *Aeromonas hydrophila* AH-3	46/53
25	FliJ$_L$	147	Export/assembly	FliJ (VPA1532) of *Vibrio parahaemolyticus*	26/57
26	FliI$_L$	443	Export ATP synthase	FliI (VPA1533) of *Vibrio parahaemolyticus*	59/66
27	FliH$_L$	253	Export/assembly	FliH (VPA1534) of *Vibrio parahaemolyticus*	39/47
28	FliG$_L$	337	Motor switch	FliG (VPA1535) of *Vibrio parahaemolyticus*	44/54
29	FliF$_L$	569	M-ring	FliF (VPA1536) of *Vibrio parahaemolyticus*	42/49
30	FliE$_L$	115	Basal body component	FliE (VPA1537) of *Vibrio parahaemolyticus*	49/56
31	FliM$_L$	300	Motor switch	FliM (VPA1540) of *Vibrio parahaemolyticus*	41/52
32	FliN$_L$	121	Motor switch	FliN (VPA1541) of *Vibrio parahaemolyticus*	57/64
33	FliP$_L$	245	Export/assembly	FliP (VPA1542) of *Vibrio parahaemolyticus*	73/79
34	FliQ$_L$	89	Export/assembly	FliQ (VPA1543) of *Vibrio parahaemolyticus*	60/69
35	FliR$_L$	263	Export/assembly	FliR (VPA1544) of *Vibrio parahaemolyticus*	58/65
36	FlhB$_L$	371	Export/assembly	FlhB (VPA1545) of *Vibrio parahaemolyticus*	43/49
37	FlhA$_L$	701	Export/assembly	FlhA (VPA1546) of *Vibrio parahaemolyticus*	59/66

The purified polar and lateral flagellins were also digested with trypsin and analyzed by tandem mass spectrometry, identifying a number of unmodified flagellin peptides. Once again, *de novo* sequencing showed several flagellin peptides from both polar and lateral flagellins to be modified with putative glycan moieties. The lateral flagellin (LafA) harbored peptides modified with glycans of 316 and 358 Da (**Figure 8B**). In some cases peptides were showed to harbor both glycans. It was not clear from the data whether two monosaccharides were modifying two separate amino acids, or whether a single disaccharide was modifying at one site.

The polar flagellin was also observed to be modified with 316 and 358 Da glycan moieties. In some cases, glycan chains comprised of multiple 358 Da glycans were observed; in other cases a single modification of 316 or 358 Da was noted. Very low levels of peptides harboring distinct glycan masses were observed, such as the peptide AIASLSTATINK, was observed to be modified with either 316 Da glycan, or a 523 or 481 Da glycan. Glycan related ions were observed in each case, with intense ions observed at *m/z* 524 and 184 or *m/z* 424 and 184. The ion at *m/z* 184 was also observed in front end CID experiments with the intact polar and lateral flagellin preparations, and gave a top ranked plausible elemental formula of $C_9H_{12}O_4$, suggesting that it is a related nonulosonic acid type sugar. The low abundance of these glycopeptides made any further analyses challenging.

TABLE 5 | Characteristics of the _P. shigelloides_ 302-73 strain gene region for legionaminic acid biosynthesis between polar flagella regions I and II.

ORF	Protein name	Protein size	Predicted function	Homologous protein with known function	Percentage identity/similarity
1	GlmU	189	Acetyltransferase	WeiJ of _Escherichia coli_	59/65
2	PtmA	254	Flagella modification protein	PtmA of _Vibrio fischeri_	71/79
				PtmA of _Campylobacter coli_	36/41
3	LegF	229	CMP-NeuAc synthase	NeuA of _Vibrio fischeri_	74/80
				Elg7 of _Escherichia coli_	70/73
				LegF of _Campylobacter coli_	31/37
4	PtmF	326	Oxidoreductase	(VF_0146) of _Vibrio fischeri_	55/62
				WeiH of _Escherichia coli_	52/58
				PtmF of _Campylobacter coli_	46/49
5	PtmE	352	Nucleotydil transferase	(VF_0145) of _Vibrio fischeri_	63/71
				Elg6 of _Escherichia coli_	61/70
6	LegH	217	_O_-acetyltransferase	NeuD (VF_0144) of _Vibrio fischeri_	49/55
				Elg5 of _Escherichia coli_	45/52
7	LegI	359	_N_-acetylneuraminate synthase	NeuB (VF_0143) of _Vibrio fischeri_	74/79
				Elg4 of _Escherichia coli_	70/74
8	LegG	382	UDP-_N_-acetylglucosamine 2-epimerase	NeuC of _Vibrio parahaemolyticus_	68/72
				Elg3 of _Escherichia coli_	65/78
				LegG of _Campylobacter jejuni_	42/47
9	LegC	382	Aminotransferase	WvaN of _Vibrio parahaemolyticus_	69/75
				PglE of _Vibrio parahaemolyticus_	69/75
10	LegB	395	Dehydratase	WvaM of _Vibrio parahaemolyticus_	75/80
				Elg1 of _Escherichia coli_	75/78

Legionaminic Acid Biosynthetic Mutants

The insertional mutants in _ptmA_ (H) and _legH_ (I) were unable to produce polar or lateral flagella under induced conditions, as shown by TEM or by immunodetection (**Figure 6B**) or lateral flagellins (**Figure 6C**) in purified flagella. The introduction of the _P. shigelloides_ wild type genes was observed to recover the production of polar and lateral flagella in the mutants. This was demonstrated using immunodections, as shown in **Figures 6B,C**. These data prompted us to examine the production of the polar flagellin in the mutants by immunodetection. Western blot analysis shows presence of polar flagellin the cytoplasmic subcellular fraction. Interestingly, only a single protein band was observed, with a lower than expected molecular weight (**Figure 6D**). Wild type flagellin typically migrates as two distainct bands, both detectable by Western blot. We speculate that the single, lower molecular weight species is a non-glycosylated form of flagellin. The complemented mutants showed the same cytoplasmic polar flagellin molecular weight bands as observed with wild type strain. Similarly, where lateral flagellin was detected in the cytoplasmic fraction, it was observed at a lower molecular weight, likely the non-modified form of the protein. Then, the lack of polar and lateral flagella formation observed in the mutants is not by the lack of flagellin protein or the master regulator transcription.

In order to prove at the genomic level that mutations in the CMP-Leg biosynthetic pathway were responsible for the phenotypic traits shown by insertional mutants H and I, two in-frame _pgmL_ and _legF_ deletion mutants were generated, 302Δ_pgmL_ and 302Δ_legF_, respectively. Our genomic studies indicates that all the genes of the Leg pathway are included in the cluster between polar region I and II, with the exception of the PgmL ortholog which is found in another region of the chromosome [703.5 peg 1785 (Piqué et al., 2013)]. PgmL or GlmM, phosphoglucosamine mutase, is involved in the first step to produce GDP-GlcNAc. LegF, CMP-legionaminic acid synthase is the final enzyme of the second step to produce CMP-Leg. Using TEM, neither mutant was observed to produce polar or lateral flagella under induced conditions. Both show the same phenotypic traits as insertional mutants H and I. When mutants 302Δ_pgmL_ and 302Δ_legF_ were complemented with their single corresponding wild type gene (pBAD33-_pgmL_ and pBAD33-_legF_, respectively) under inducing conditions (plus arabinose) all the wild type phenotypic traits (production of polar and lateral flagella or swimming and swarming motilities) were fully recovered. Control plasmid pBAD33 alone under inducing conditions (plus arabinose) was unable to do it.

Lateral Flagella and Leg _O_-Flagella Glycosylation Gene Distribution on _P. shigelloides_

In order to test if the presence of lateral flagella and Leg _O_-flagella glycosylation genes is a specific feature for the strain studied, the 12 previously mentioned _P. shigelloides_ strains used for PCR studies were eight strains representing five different serotypes (O1, O2, O3, O17, and O54) plus 4 non-serotyped strains described in Material and Methods. Initially, genomic DNA from 302-73 strain was used as template for PCR amplification with two sets of oligonucleotides: 5′-ATCGCGTCTGAAAGGCTAC-3′ and 5′-CTGCGCCATAGAACTACCC-3′ which amplified a 2160 bp DNA fragment from lateral flagella cluster (partial _lafA_ and complete _maf-5_); and another oligonucleotide set (5′-CGGGTTAAAGCTATCCCATC-3′ and 5′-CCAATGACAGC

FIGURE 6 | (A) *Plesiomonas shigelloides* 302-73 wild type strain serotype O1 purified flagella according to Section "Materials and Methods" when grown in liquid medium (1) and swarming agar plates (2). As could be observed in 2, and previously indicated in **Figure 3**, polar flagella are constitutively expressed in semisolid medium. Purified flagella from *P. shigelloides* insertional polar A mutant grown in swarming agar plates (3). **(B)** Western blot with specific polar flagella antiserum of purified flagella from wild type (1), *P. shigelloides* insertional polar H mutant (2), and complemented mutant with COS-LEG harboring the corresponding wild type gene (3) obtained in liquid medium growth. **(C)** Western blot with specific lateral flagella antiserum of purified flagella from wild type (1), *P. shigelloides* insertional polar H mutant (2), and complemented mutant with COS-LEG harboring the corresponding wild type gene (3) obtained in swarming agar plates. **(D)** Western blot with specific polar flagella antiserum of cytoplasmic fractions obtained as described in Section "Materials and Methods" of wild type (1), *P. shigelloides* insertional polar H mutant (2), and complemented mutant with COS-LEG harboring the corresponding wild type gene (3) obtained in liquid medium growth. The low molecular weight band could correspond to the non-glycosylated form, and the upper band (not present in the mutant) to the glycosylated form.

TGAATCTCC-3′) amplified a 1985 bp DNA fragment from Leg biosynthesis genes (partial *legH* and complete *legI*). DNA fragments of the same size (2160 and 1985 bp, respectively) were PCR amplified for all the genomic DNAs from the strains studied, as shown by the results shown in **Figure 9**. DNA sequence of the amplified fragments confirmed the presence of the lateral and Leg biosynthetic genes. In addition, in all the amplified *maf-5* and *legI* fragments the presence of a sequence coding for the N-terminal amino acid residues of *lafA* and *legH* genes, respectively, were found adjacent to *maf-5* or *legI*, suggesting that in the analyzed strains the genomic location is the same as that found in *P. shigelloides* wild type strain 302-73 (**Figure 5**).

Discussion

Motility is an essential mechanism in adaptation to different environments for free living bacteria. Bacteria showed three

flagella types classified according to their location on a cell: peritrichous, polar, and lateral. It has been reported dual flagella systems in some polar flagellated bacteria when grow in viscous environments or surfaces. This fact allows bacteria to swarm on solid or semisolid media by a mixed flagellation (polar and lateral flagella). *P. shigelloides* has been observed to express mixed flagellation (Inoue et al., 1991).

Two *P. shigelloides* 302-73 different gene clusters were described, one exclusively involved in lateral flagella biosynthesis, and a second containing the polar flagella genes distributed in two regions spaced by putative glycosylation genes. It is characteristic of the bacteria with dual flagella systems to separate both in different gene clusters (McCarter, 2001; Canals et al., 2006; Merino et al., 2006; Merino and Tomás, 2009). Of note, *P. shigelloides* is the first *Enterobacteriaceae* with lateral flagella production as shown herein.

Plesiomonas shigelloides lateral gene cluster is nearly identical to the lateral gene cluster of *A. hydrophila* according to the gene grouping and transcription direction, with the exception of the group of genes $fliM_L$ to R_L plus $flhB-A_L$ which are transcribed in opposite direction (Canals et al., 2006). However, no *lafK* ortholog could be detected in *P. shigelloides* lateral gene cluster. This gene has been reported in all the lateral gene clusters, including the non-functional in the *Enterobacteriaceae* (Canals et al., 2006; Merino and Tomás, 2009). A non-functional Flag-2 flagella cluster with large similarity to *V. parahaemolyticus* lateral flagella system, was found in different *E. coli* enteroaggregative or *Yersinia pestis* or *pseudotuberculosis* strains (Ren et al., 2004). However, as we proved, *P. shigelloides* lateral gene cluster is fully functional.

The transcriptional hierarchy of *V. parahaemolyticus* lateral flagella is one of the *Gammaproteobacteria* model. LafK (σ^{54}-associated transcriptional activator) is the master regulon in this model, controlling Class II lateral flagella genes transcription. Class II genes contains the σ^{28} factor ($fliA_L$) which is involved in transcription of Class III lateral flagella genes (Stewart and McCarter, 2003). In *V. parahaemolyticus* the absence of polar flagellum induces the expression of lateral flagella in liquid medium, and LafK is able to compensate the lack of FlaK (σ^{54}-associated polar transcriptional activator) and activate polar flagellum class promoters. *A. hydrophila* lateral flagella transcriptional hierarchy represents the second *Gammaproteobacteria* model. Class I gene transcription in *A. hydrophila* lateral flagella is σ^{70}-dependent as LafK in contrast to describe in *V. parahaemolyticus* (Stewart and McCarter, 2003). It is important to point out that *A. hydrophila* lateral flagella genes are transcribed in liquid and solid or semisolid media, and unlike *V. parahaemolitycus* the genes are not induced by mutation of polar flagellum genes. The transcription hierarchy of *A. hydrophila* lateral flagella is complex because LafK is not strictly their master lateral flagella regulator, and many clusters of genes are LafK independently transcribed (Wilhelms et al., 2013). *A. hydrophila* LafK protein is unable to not compensate the lack of FlrA, which is the polar-flagellum regulator (σ^{54}-associated transcriptional activator for polar flagellum), a situation that happens in *V. parahaemolyticus* (Wilhelms et al., 2013). This point is in agreement with *A. hydrophila* FlrA mutation not affect
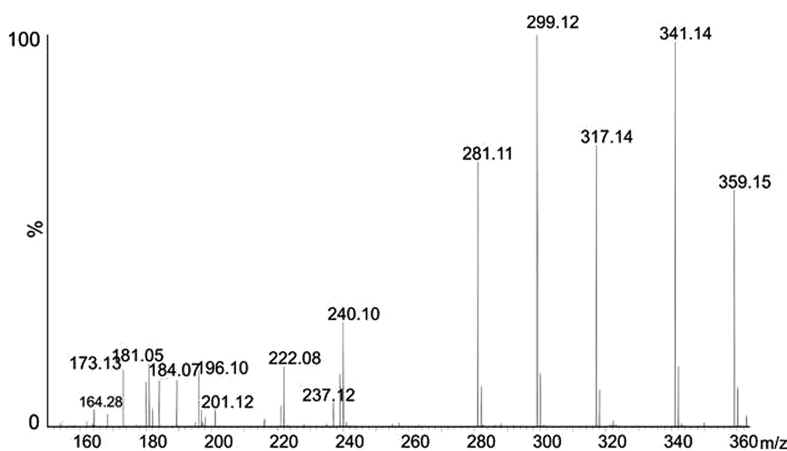
FIGURE 7 | Glycan fragmentation pattern. Front end collision induced dissociation of polar flagellin protein, showing low *m/z* region. Fragment ions characteristic of an acetylated nonulosonic acid sugar are present.
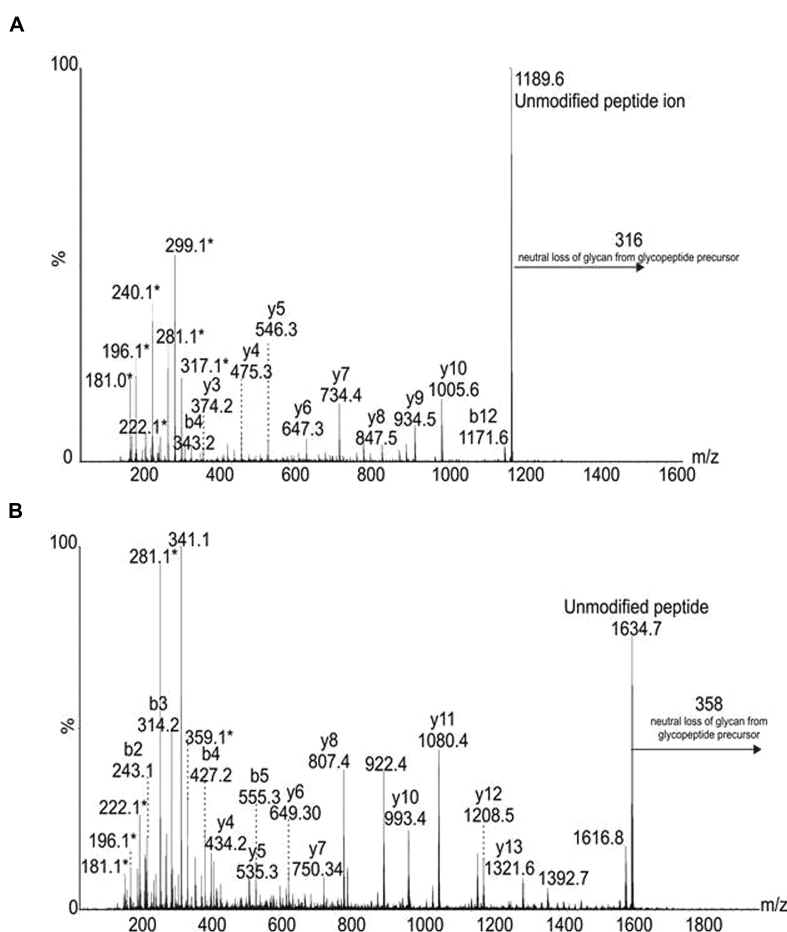
FIGURE 8 | Tandem mass spectra of flagellin glycopeptides (A) from polar flagellin, peptide AIASLSTAINK, modified with a 316 Da glycan. Peptide type y and b ions are indicated, confirming the peptide sequence. In addition, glycan related fragment ions are indicated with an asterisk (*), and include a glycan oxonium ion at *m/z* 317, and related ions at *m/z* 299, 281, 240, 221, and 181. **(B)** From lateral flagellin, the peptide type y and b fragment ions indicate a sequence of ELALQSANGTNTTADK. In this case the peptide is modified with a 358 Da glycan, as observed with a neutral loss of 358 from the glycopeptide precursor and the glycan oxonium ion at *m/z* 359. A loss of water is observed from this oxonium ion, in addition glycan related fragment ions are indicated with an asterisk (*).

FIGURE 9 | Polymerase chain reaction (PCR) amplified bands from genomic *P. shigelloides* DNAs of strains: 302-73, serotype O1 (1); 306-73, serotype O2 (2); 307-73, serotype O3 (3); 343-73, serotype O17 (4); PCM2647, serotype O54 (5); C1, non-serotyped from Spain (6); P12, non-serotyped from Spain (7); and It645, non-serotyped from Brazil (8); using the set of oligonucleotides for lateral flagella cluster (A, 2160 bp DNA fragment) and Leg biosynthesis genes (B, 1985 bp DNA fragment).

lateral flagella besides that abolishes polar flagellum formation in liquid and on solid surfaces (Wilhelms et al., 2013).

The *P. shigelloides* polar flagella gene regions show greater similarity to those reported in *Vibrio* or *Aeromonas* than the regions in *Enterobacteriaceae* [e.g., *E. coli* or *S. typhimurium* (Chilcott and Hughes, 2000)]. Bacteria with peritrichous flagella, such as *E. coli* and *Salmonella*, showed three hierarchy levels. The σ^{70} is required for transcription of class I and II genes, and class I promoter responds to different regulatory factors and transcribes the FlhDC master activator, which allowed the class II σ^{70}-dependent promoter expression. At the top of the *Vibrio* sp. or *A. hydrophyla* polar flagella hierarchy is σ^{54}-associated transcriptional activator (FlrA, named FleQ in *Pseudomonas aeruginosa*) which activates class II genes σ^{54}-dependent promoters. Class II promoters encode a two component signal-transducing system (*Vibrio* sp. or *A. hydrophyla* FlrBC and FleSR in *P. aeruginosa*) whose regulator (FlrC/FleR) activates class III genes σ^{54}-dependent promoters.

In the *P. shigelloides* polar flagella region I only *flrA* and *C* orthologs were observed. *P. shigelloides* FlrA shows the characteristic three domains (FleO, σ^{54}-interaction domain and family regulatory protein Fis) like in *Vibrio* sp. or *A. hydrophila* (Kim and McCarter, 2004; Wilhelms et al., 2011). Class II promoters encode a two component signal-transducing system (FlrBC of *Vibrio* sp. or *A. hydrophila* and FleSR in *P. aeruginosa*) whose regulator (FlrC/FleR) activates class III σ^{54}-dependent promoters. However, when analysis of *P. shigelloides* FlrC encoded protein, revealed the corresponding domains for FlrB and C. Thus, *P. shigelloides* FlrC contains two domains of *Vibrio* sp. or *A. hydrophila* FlrB (PAS domain and His Kinase A) as well as two domains of *Vibrio* sp. or *A. hydrophila* FrlC (σ^{54}-interaction domain and family regulatory protein Fis).

We suggest that *P. shigelloides* FlrC could be able to activate class III genes σ^{54}-dependent promoters as observed in *Vibrio* sp. or *A. hydrophila*. No FlrB ortholog was observed in the *P. shigelloides* 302-73 genome (Piqué et al., 2013). It could be suggested that in *P. shigelloides*, FlrB and C functions are developed by a single bifunctional protein encoded by the single *flrC* as it happens for some LPS-core biosynthetic genes (Jiménez et al., 2009). Taken together, the data presented hererin, no *lafK* or separate *flrB* in *P. shigelloides*, indicate that their lateral and polar flagella transcriptional hierarchy represents a different *Gammaproteobacteria* model that requires further study.

Among this large *P. shigelloides* polar flagella gene cluster, genes were identified between the two polar flagella regions, the presence of genes putatively linked to glycosylation. These genes were not found in other *Enterobacteriaceae* studied. *O*-glycosylation could be performed by a mechanism dependent or not of an oligosaccharyltransferase (OTase; Kim and McCarter, 2004; Iwashkiw et al., 2013). The *O*-glycosylation frequently affects protein stability, flagella filament assembly, bacterial adhesion, biofilm formation, and virulence in general as has been described in several bacteria (Lindenthal and Elsinghorst, 1999; Logan, 2006; Faridmoayer et al., 2008; Egge-Jacobsen et al., 2011; Iwashkiw et al., 2013; Lithgow et al., 2014). The predominant *O*-glycans linked to flagellins are mainly derivatives of pseudaminic acid (PseAc, where Ac represents an acetamido group) and in a minor extent an acetamidino form of legionaminic acid (LegAm, where Am represents acetamidino; Merino et al., 2014). Both are nine-carbon sugars related to sialic acid. The flagellin glycosylation pathways in both cases have been elucidated, including the Pse pathway of *Helicobacter pylori* and *C. jejuni* (Fox, 2002), the Leg pathway of *C. jejuni* (Schoenhofen et al., 2009). Until today the Leg flagella glycosylation has been restricted to *C. jejuni* or *coli*. The CMP-legionaminic acid biosynthetic pathway in *C. jejuni* involves two steps: synthesis of a GDP-GlcNAc and synthesis of the final CMP-Leg (Schoenhofen et al., 2009). The insertional mutants obtained *ptmA* (H) and *legH* (I), represent key eznymes in the first and second steps of the CMP-Leg biosynthesis, confirming the observation data that both mutants are unable to produce polar or lateral flagella. Furthermore, the in frame mutants obtained in *pgmL* and *legF*, one enzyme of the first step and the last enzyme of the second step of the CMP-Leg biosynthesis, respectively, clearly confirmed the legionaminic acid polar and lateral glycosylation as both mutants are unable to produce polar or lateral flagella as it happens with the insertional mutants.

Mass spectrometry studies show that both flagella in *P. shigelloides* strain 302-73 are glycosylated by a derivative of Leg, and is also indicated by the presence of Leg biosynthetic pathway genes nearby the polar flagella gene regions. It is the first *Enterobacteriaceae* reported to harbor *O*-glycosylation modification on both polar and lateral flagella. Moreover, it is also the first bacteria reported to express a lateral flagella glysosylated by Leg. We also demonstrated that flagella *O*-glycosylation is essential for bacterial flagella formation, either polar or lateral. However, the flagella *O*-glycosylation is not determinant for cytoplasmic flagellin production as can be observed by immunodetection studies.

The *P. shigelloides* homologous recombination rates are extremely high (Salerno et al., 2007), like naturally transformable species as *Streptococcus pneumoniae*. In the rest of *Enterobacteriaceae* the recombination rate is much lower. The high recombination observed in this bacterium could offer a reason for *P. shigelloides* variety of LPS-core structures (Salerno et al., 2007). The PCR experiments using several *P. shigelloides* strains and lateral flagella or Leg pathway genes, with the motility and EM studies, demonstrated that presence of lateral flagella and Leg *O*-flagella glycosylation is a widely spread feature, not a strain specific observation. Furthermore, the maintenance of these genes among the different strains besides the recombination rate

observed for *P. shigelloides*, indicates the importance of glycosylated polar and lateral flagella production for this bacterium.

Acknowledgments

References

Aldova, E., and Shimada, T. (2000). New O and H antigens of the international antigenic scheme for *Plesiomonas shigelloides*. *Folia Microbiol.* 45, 301–304. doi: 10.1007/BF02817550

Allen, L. N., and Hanson, R. S. (1985). Construction of broad host-range cosmid cloning vector: identification of genes necessary for growth of *Methylobacterium organophilum* on methanol. *J. Bacteriol.* 161, 955–962.

Allison, C., Coleman, N., Jones, P. L., and Hughes, C. (1992). The ability of *Proteus mirabilis* to invade human urothelial cells is coupled to motility and swarming differentiation. *Infect. Immun.* 60, 4740–4746.

Allison, C., and Hughes, C. (1991). Bacterial swarming: an example of prokaryotic differentiation and multicellular behaviour. *Sci. Prog.* 75, 403–422.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389

Aquilini, E., Merino, S., Regué, M., and Tomás, J. M. (2014). Genomic and proteomic studies on *Plesiomonas shigelloides* lipopolysaccharide core biosynthesis. *J. Bacteriol.* 196, 556–567. doi: 10.1128/JB.01100-13

Aquilini, E., Merino, S., and Tomás, J. M. (2013). The *Plesiomonas shigelloides* wbO1 gene cluster and the role of O1-antigen LPS in pathogenicity. *Microb. Pathog.* 63, 1–7. doi: 10.1016/j.micpath.2013.05.010

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R., et al. (2002). The Pfam protein families database. *Nucleic Acids Res.* 30, 276–280. doi: 10.1093/nar/30.1.276

Billiet, J., Kuypers, S., Van Lierde, S., and Verhaegen, J. (1989). *Plesiomonas shigelloides* meningitis and septicaemia in a neonate: report of a case and review of the literature. *J. Infect.* 19, 267–271. doi: 10.1016/S0163-4453(89)90809-8

Brenden, R. A., Miller, M. A., and Janda, J. M. (1988). Clinical disease spectrum and pathogenic factors associated with *Plesiomonas shigelloides* infections in humans. *Rev. Infect. Dis.* 10, 303–316. doi: 10.1093/clinids/10.2.303

Canals, R., Altarriba, M., Vilches, S., Horsburgh, G., Shaw, J. G., Tomás, J. M., et al. (2006). Analysis of the lateral flagellar gene system of *Aeromonas hydrophila* AH-3. *J. Bacteriol.* 188, 852–862. doi: 10.1128/JB.188.3.852-862.2006

Chilcott, G. S., and Hughes, K. T. (2000). Coupling of flagellar gene expression to flagellar assembly in *Salmonella enterica* serovar Typhimurium and *Escherichia coli*. *Microbiol. Mol. Biol. Rev.* 64, 694–708. doi: 10.1128/MMBR.64.4.694-708.2000

De Lorenzo, V., Herrero, M., Jakubzik, U., and Timmis, K. N. (1990). Mini-Tn5 transposon derivatives for insertion mutagenesis, promoter probing, and chromosomal insertion of cloned DNA in gram-negative eubacteria. *J. Bacteriol.* 172, 6568–6572.

Egge-Jacobsen, W., Salomonsson, E. N., Aas, F. E., Forslund, A. L., Winther-Larsen, H., Maier, J., et al. (2011). O-linked glycosylation of the PilA pilin protein of *Francisella tularensis*: identification of the endogenous protein-targeting oligosaccharyltransferase and characterization of the native oligosaccharide. *J. Bacteriol.* 193, 5487–5497. doi: 10.1128/JB.00383-11

Eloe, E. A., Lauro, F. M., Vogel, R. F., and Bartlett, D. H. (2008). The deep-sea bacterium *Photobacterium profundum* SS9 utilizes separate flagellar systems

for swimming and swarming under high-pressure conditions. *Appl. Environ. Microbiol.* 74, 6298–6305. doi: 10.1128/AEM.01316-08

Faridmoayer, A., Fentabil, M. A., Haurat, M. F., Yi, W., Woodward, R., Wang, P. G., et al. (2008). Extreme substrate promiscuity of the *Neisseria* oligosaccharyl transferase involved in protein O-glycosylation. *J. Biol. Chem.* 283, 34596–34604. doi: 10.1074/jbc.M807113200

Faridmoayer, A., Fentabil, M. A., Mills, D. C., Klassen, J. S., and Feldman, M. F. (2007). Functional characterization of bacterial oligosaccharyltransferases involved in O-linked protein glycosylation. *J. Bacteriol.* 189, 8088–8098. doi: 10.1128/JB.01318-07

Farmer, J. J. III, Arduino, M. J., and Hickman-Brenner, F. W. (1992). "The genera *Aeromonas* and *Plesiomonas*," in *The Prokaryotes*, Vol. 3, eds A. Balows, H. G. Trupper, M. Dworkin, H. Wim, and K. H. Schleifer (New York, NY: Springer-Verlag), 3012–3043.

Fischer, K., Chakraborty, T., Hof, H., Kirchner, T., and Wamsler, O. (1988). Pseudoappendicitis caused by *Plesiomonas shigelloides*. *J. Clin. Microbiol.* 26, 2675–2677.

Fox, J. G. (2002). The non-H pylori helicobacters: their expanding role in gastrointestinal and systemic diseases. *Gut* 50, 273–283. doi: 10.1136/gut.50.2.273

Fujita, K., Shirai, M., Ishioka, T., and Kakuya, F. (1994). Neonatal *Plesiomonas shigelloides* septicemia and meningitis: a case and review. *Acta Paediatr. Jpn.* 36, 450–452. doi: 10.1111/j.1442-200X.1994.tb03222.x

Garrity, G. M., Winters, M., and Searles, D. B. (2001). "Taxonomic outline of the procaryotic genera," in *Bergey's Manual of Systematic Bacteriology*, 2nd Edn, eds G. M. Garrity, D. R. Boone, and R. W. Castenholz (New York, NY: Springer-Verlag), 13.

Gavín, R., Rabaan, A. A., Merino, S., Tomás, J. M., Gryllos, I., and Shaw, J. G. (2002). Lateral flagella of *Aeromonas* species are essential for epithelial cell adherence and biofilm formation. *Mol. Microbiol.* 43, 383–397. doi: 10.1046/j.1365-2958.2002.02750.x

Goon, S., Kelly, J. F., Logan, S. M., Ewing, C. P., and Guerry, P. (2003). Pseudaminic acid, the major modification on *Campylobacter flagellin*, is synthesized via the Cj1293 gene. *Mol. Microbiol.* 50, 659–671. doi: 10.1046/j.1365-2958.2003.03725.x

Guasch, J. F., Piqué, N., Climent, N., Ferrer, S., Merino, S., Rubires, X., et al. (1996). Cloning and characterization of two *Serratia marcescens* genes involved in core lipopolysaccharide biosynthesis. *J. Bacteriol.* 178, 5741–5747.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* 166, 557–580. doi: 10.1016/S0022-2836(83)80284-8

Haya, S., Tokumaru, Y., Abe, N., Kaneko, J., and Aizawa, S. (2011). Characterization of lateral flagella of *Selenomonas ruminantium*. *Appl. Environ. Microbiol.* 77, 2799–2882. doi: 10.1128/AEM.00286-11

Huys, G., and Sings, J. (1999). Evaluation of a fluorescent amplified fragment length polymorphism (FAFLP) methodology for the genotypic discrimination of *Aeromonas* taxa. *FEMS Microbiol. Lett.* 177, 83–92. doi: 10.1111/j.1574-6968.1999.tb13717.x

Inoue, K., Kosako, Y., Suzuki, K., and Shimada, T. (1991). Peritrichous flagellation in *Plesiomonas shigelloides* strains. *Jpn. J. Med. Sci. Biol.* 44, 141–146. doi: 10.7883/yoken1952.44.141

Iwashkiw, J. A., Seper, A., Weber, B. S., Scott, N. E., Vinogradov, E., Stratilo, C. H., et al. (2012). Identification of a general O-linked protein glycosylation system in *Acinetobacter baumannii* and its role in virulence and biofilm formation. *PLoS Pathog.* 8:e1002758. doi: 10.1371/journal.ppat.1002758

Iwashkiw, J. A., Vozza, N. F., Kinsella, R. L., and Feldman, M. F. (2013). Pour some sugar on it: the expanding world of bacterial protein O-linked glycosylation. *Mol. Microbiol.* 89, 14–28. doi: 10.1111/mmi.12265

Jiménez, N., Vilches, S., Lacasta, A., Regué, M., and Merino, S. (2009). A bifunctional enzyme in a single gene catalyzes the incorporation of GlcN into the *Aeromonas* core LPS. *J. Biol. Chem.* 284, 32995–33005. doi: 10.1074/jbc.M109.038828

Kanbe, M., Yakasagi, J., Zehner, S., Göttfert, M., and Aizawa, S. I. (2007). Characterization of two sets of sub-polar flagella in *Bradyrhizobium japonicum*. *J. Bacteriol.* 189, 1083–1089. doi: 10.1128/JB.01405-06

Kim, Y. K., and McCarter, L. L. (2004). Cross-regulation in *Vibrio parahaemolyticus*: compensatory activation of polar flagellar genes by the lateral flagellar regulator LafK. *J. Bacteriol.* 186, 4014–4018. doi: 10.1128/JB.186.12.4014-4018.2004

Lindenthal, C., and Elsinghorst, E. A. (1999). Identification of a glycoprotein produced by enterotoxigenic *Escherichia coli*. *Infect. Immun.* 67, 4084–4091.

Lithgow, K. V., Scott, N. E., Iwashkiw, J. A., Thomson, E. L. S., Foster, L. J., Feldman, M. F., et al. (2014). A general protein O-glycosylation system within the *Burkholderia cepacia* complex is involved in motility and virulence. *Mol. Microbiol.* 92, 116–137. doi: 10.1111/mmi.12540

Logan, S. M. (2006). Flagellar glycosylation—a new component of the motility repertoire? *Microbiology* 152, 1249–1262. doi: 10.1099/mic.0.28735-0

Macnab, R. M. (1996). "Flagella and motility," in *Escherichia coli and Salmonella typhimurium: Celular and Molecular Biology*, eds F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik et al. (Washington, DC: ASM Press), 123–145.

Mandal, B. K., Whale, K., and Morrison, B. C. (1982). Acute colitis due to *Plesiomonas shigelloides*. *Br. Med. J.* 285, 1539–1540. doi: 10.1136/bmj.285.6354.1539-a

Martinez-Murcia, A. J., Benlloch, S., and Collins, M. D. (1992). Phylogenetic interrelationships of members of the genera *Aeromonas* and *Plesiomonas* as determined by 16S ribosomal DNA sequencing: lack of congruence with results of DNA-DNA hybridizations. *Int. J. Syst. Bacteriol.* 42, 412–421. doi: 10.1099/00207713-42-3-412

McCarter, L. L. (2001). Polar flagellar motility of the Vibrionaceae. *Microbiol. Mol. Biol. Rev.* 65, 445–462. doi: 10.1128/MMBR.65.3.445-462.2001

McClain, J., Rollo, D. R., Rushing, B. G., and Bauer, C. E. (2002). *Rhodospirillum centenum* utilizes separate motor switch components to control lateral and polar flagellum rotation. *J. Bacteriol.* 184, 2429–2438. doi: 10.1128/JB.184.9.2429-2438.2002

McNeeley, D., Ivy, P., Craft, J. C., and Cohen, I. (1984). *Plesiomonas*: biology of the organism and diseases in children. *Pediatr. Infect. Dis. J.* 3, 176–181. doi: 10.1097/00006454-198403000-00023

Merino, S., Camprubí, S., Albertí, S., Benedí, V. J., and Tomás, J. M. (1992). *Klebsiella pneumoniae* resistance mechanisms to complement mediated killing. *Infect. Immun.* 60, 2529–2535.

Merino, S., Fulton, K. M., Twine, S. M., Wilhelms, M., Molero, R., and Tomás, J. M. (2014). *Aeromonas hydrophila* flagella glycosylation: involvement of a lipid carrier. *PLoS ONE* 9:e89630. doi: 10.1371/journal.pone.0089630

Merino, S., Shaw, J. G., and Tomás, J. M. (2006). Bacterial lateral flagella: an inducible flagella system. *FEMS Microbiol. Lett.* 263, 127–135. doi: 10.1111/j.1574-6968.2006.00403.x

Merino, S., and Tomás, J. M. (2009). "Lateral flagella systems," in *Pili and Flagella: Current Research and Future Trends*, ed. K. Jarrell (Norfolk, VA: Caister Academic Press), 171–188.

Merino, S., and Tomás, J. M. (2014). Gram-negative flagella glycosylation. *Int. J. Mol. Sci.* 15, 2840–2857. doi: 10.3390/ijms15022840

Milton, D. L., O'Toole, R., Horstedt, P., and Wolf-Watz, H. (1996). Flagellin a is essential for the virulence of *Vibrio anguillarum*. *J. Bacteriol.* 178, 1310–1319.

Moens, S., Schloter, M., and Vanderleyden, J. (1996). Expression of the structural gene, laf1, encoding the flagellin of the lateral flagella in *Azospirillum brasilense* Sp7. *J. Bacteriol.* 178, 5017–5019.

O'Rourke, J., Lee, A., and Fox, J. G. (1992). An ultrastructural study of *Helicobacter mustelae* and evidence of specific association with gastric mucosa. *J. Med. Microbiol.* 36, 420–427. doi: 10.1099/00222615-36-6-420

Parker, J. L., Lowry, R. C., Couto, N. A. S., Wright, P. C., Stafford, G. P., and Shaw, J. G. (2014). Maf dependent bacterial flagellin glycosylation occurs before chaperone binding and flagelar T3SS export. *Mol. Microbiol.* 92, 258–272. doi: 10.1111/mmi.12549

Pieretti, G., Carillo, S., Lindner, B., Lanzetta, R., Parrilli, M., Jimenez, N., et al. (2010). The complete structure of the core of the LPS from *Plesiomonas shigelloides* 302-73 and the identification of its O-antigen biological repeating unit. *Carbohydr. Res.* 345, 2523–2528. doi: 10.1016/j.carres.2010.09.007

Piqué, N., Aquilini, E., Alioto, T., Miñana-Galbis, D., and Tomás, J. M. (2013). Genome sequence of *Plesiomonas shigelloides* strain 302-73 (serotype O1). *Genome Announc.* 1, e00404–e00413. doi: 10.1128/genomeA.00404-13

Poggio, S., Abreu-Goodger, C., Fabela, S., Osorio, A., Dreyfus, G., Vinuesa, P., et al. (2007). A complete set of flagellar genes acquired by horizontal transfer coexists with the endogenous flagellar system in *Rhodobacter sphaeroides*. *J. Bacteriol.* 189, 3208–3216. doi: 10.1128/JB.01681-06

Ren, C. P., Beatson, S., Parkhill, J., and Pallen, M. J. (2004). The Flag-2 locus, an ancestral gene cluster, is potentially associated with a novel flagellar system from *Escherichia coli*. *J. Bacteriol.* 187, 1207–1209.

Rubirés, X., Saigí, F., Piqué, N., Climent, N., Merino, S., Rubires, X., et al. (1997). A gene (wbbL) from *Serratia marcescens* N28b (O4) complements the rfb-50 mutation of *Escherichia coli* K-12 derivatives. *J. Bacteriol.* 179, 7581–7586.

Salerno, A., Delétoile, M., Lefevre, I., Ciznar, I., Krovacek, K., Grimont, P., et al. (2007). Recombining population structure of *Plesiomonas shigelloides* (Enterobacteriaceae) revealed by multilocus sequence typing. *J. Bacteriol.* 189, 7808–7818. doi: 10.1128/JB.00796-07

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, 2nd Edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* 74, 5463–5467. doi: 10.1073/pnas.74.12.5463

Schirm, M., Soo, E. C., Aubry, A. J., Austin, J., Thibault, P., and Logan, S. M. (2003). Structural, genetic and functional characterization of the flagellin glycosylation process in *Helicobacter pylori*. *Mol. Microbiol.* 48, 1579–1592. doi: 10.1046/j.1365-2958.2003.03527.x

Schoenhofen, I. C., Vinogradov, E., Whitfield, D. M., Brisson, J. R., and Logan, S. M. (2009). The CMP-legionaminic acid pathway in *Campylobacter*: biosynthesis involving novel GDP-linked precursors. *Glycobiology* 19, 715–725. doi: 10.1093/glycob/cwp039

Shinoda, S., Yakiyama, I., Yasui, S., Kim, Y., Ono, B., and Nakagami, S. (1992). Lateral flagella of vibrios: serological classification and genetical similarity. *Microbiol. Immunol.* 36, 303–309. doi: 10.1111/j.1348-0421.1992.tb01667.x

Stewart, B. J., and McCarter, L. L. (2003). Lateral flagellar gene system of *Vibrio parahaemolyticus*. *J. Bacteriol.* 185, 4508–4518. doi: 10.1128/JB.185.15.4508-4518.2003

Tomás, J. M., Camprubí, S., Merino, S., Davey, M. R., and Williams, P. (1991). Surface exposure of O1 serotype lipopolysaccharide in *Klebsiella pneumoniae* strains expressing different K antigens. *Infect. Immun.* 59, 2006–2011.

Towbin, H., and Gordon, J. (1984). Immunoblotting and dot immunobinding-current status and outlook. *J. Immunol. Methods* 72, 313–340. doi: 10.1016/0022-1759(84)90001-2

Tsukamoto, T., Kinoshita, Y., Shimada, T., and Sakazaki, R. (1978). Two epidemics of diarrhoeal disease possibly caused by *Plesiomonas shigelloides*. *J. Hyg.* 80, 275–280. doi: 10.1017/S0022172400053638

Wilhelms, M., Fulton, K. M., Twine, S. M., Tomás, J. M., and Merino, S. (2012). Differential glycosylation of polar and lateral flagellins in *Aeromonas hydrophila* AH-3. *J. Biol. Chem.* 287, 27851–27862. doi: 10.1074/jbc.M112.376525

Wilhelms, M., Gonzalez, V., Tomás, J. M., and Merino, S. (2013). *Aeromonas hydrophila* lateral flagella gene transcriptional hierarchy. *J. Bacteriol.* 195, 1436–1445. doi: 10.1128/JB.01994-12

Wilhelms, M., Molero, R., Shaw, J. G., Tomás, J. M., and Merino, S. (2011). Transcriptional hierarchy of *Aeromonas hydrophila* polar-

flagellum genes. *J. Bacteriol.* 193, 5179–5190. doi: 10.1128/JB.053 55-11

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

CrossMark

Bordetella pertussis Isolates from Argentinean Whooping Cough Patients Display Enhanced Biofilm Formation Capacity Compared to Tohama I Reference Strain

Laura Arnal[1†], Tom Grunert[2†], Natalia Cattelan[1], Daan de Gouw[3,4], María I. Villalba[1], Diego O. Serra[1,5], Frits R. Mooi[6], Monika Ehling-Schulz[2]* and Osvaldo M. Yantorno[1]*

[1] CINDEFI–Centro Científico Tecnológico CONICET La Plata, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, Buenos Aires, Argentina, [2] Functional Microbiology, Institute of Microbiology, Department of Pathobiology, University of Veterinary Medicine Vienna, Vienna, Austria, [3] Laboratory of Pediatric Infectious Diseases, Department of Pediatrics, Radboud University Medical Centre, Nijmegen, Netherlands, [4] Laboratory of Medical Immunology, Department of Laboratory Medicine, Radboud University Medical Centre, Nijmegen, Netherlands, [5] Mikrobiologie, Institut for Biologie, Humboldt-Universitat zu Berlin, Berlin, Germany, [6] Netherlands Centre for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven, Netherlands

Pertussis is a highly contagious disease mainly caused by *Bordetella pertussis*. Despite the massive use of vaccines, since the 1950s the disease has become re-emergent in 2000 with a shift in incidence from infants to adolescents and adults. Clearly, the efficacy of current cellular or acellular vaccines, formulated from bacteria grown in stirred bioreactors is limited, presenting a challenge for future vaccine development. For gaining insights into the role of *B. pertussis* biofilm development for host colonization and persistence within the host, we examined the biofilm forming capacity of eight argentinean clinical isolates recovered from 2001 to 2007. All clinical isolates showed an enhanced potential for biofilm formation compared to the reference strain Tohama I. We further selected the clinical isolate *B. pertussis* 2723, exhibiting the highest biofilm biomass production, for quantitative proteomic profiling by means of two-dimensional fluorescence difference gel electrophoresis (2D-DIGE) coupled with mass spectrometry, which was accompanied by targeted transcriptional analysis. Results revealed an elevated expression of several virulence factors, including adhesins involved in biofilm development. In addition, we observed a higher expression of energy metabolism enzymes in the clinical isolate compared to the Tohama I strain. Furthermore, all clinical isolates carried a polymorphism in the *bvgS* gene. This mutation was associated to an increased sensitivity to modulation and a faster rate of adhesion to abiotic surfaces. Thus, the phenotypic biofilm characteristics shown by the clinical isolates might represent an important, hitherto underestimated, adaptive strategy for host colonization and long time persistence within the host.

Keywords: whooping cough, *Bordetella pertussis,* clinical isolates, biofilm, proteomic, real time PCR

INTRODUCTION

Bordetella pertussis is a human-restricted pathogen specifically adapted to infect the respiratory tract producing whooping cough or pertussis. Despite the success of mass immunization in reducing the incidence of the disease in the 1950s, after six decades of sustained high vaccination coverage, whooping cough remains endemic with epidemic cycles every 2–5 years (Mooi et al., 1998, 2001, 2009, 2014; Heikkinen et al., 2007; King et al., 2008). Although the disease has been associated to an acute infection, mainly affecting unvaccinated infants aged <6 months, in the last two decades, a shift in the incidence toward vaccinated children, adolescents, and adults has become increasingly evident (Cherry, 1999; Hellenbrand et al., 2009; de Greeff et al., 2010). This new scenario represents a significant health concern, since these individuals could provide reservoirs for *B. pertussis* transmission. Several reasons for the resurgence and persistence of pertussis in the population have been discussed, including: waning immunity over time, variation between circulating isolates and vaccine strains as a result of constant pathogen adaptation, and reduced efficiency of vaccine formulations (He et al., 2003; Hewlett and Edwards, 2005; Brinig et al., 2006; Bottero et al., 2007; Berbers et al., 2009; Elomaa et al., 2009; Lavine et al., 2011). For the commercial production of both, cellular and acellular vaccines, *B. pertussis* cells are grown in stirred bioreactor operated in batch culture. This planktonic (free-floating) mode of growth does not reflect the lifestyle of the pathogen in its host, where bacteria must primarily adhere to ciliated respiratory epithelial cells; in this hostile environment, bacteria must resist mucociliary clearance and avoid the immune system's mechanisms, adjusting their growth state and virulence accordingly to survive.

Numerous reports provide evidence that the ability of pathogens to adhere and grow attached to tissues' surfaces in microbial communities, known as biofilm, is crucial for the development and progression of human infections (Costerton et al., 1999; Hall-Stoodley et al., 2004). Generally, biofilm development, which is often associated to an enhanced resistance to antimicrobial agents and host defenses, is considered as an important survival strategy for bacteria (Ito et al., 2009; Hoiby et al., 2010; Gurung et al., 2013). In addition, the switch from planktonic to biofilm lifestyle is accompanied by significant changes in bacterial metabolism and phenotypic features, which represent a unique challenge for the development of novel prophylactic therapeutics. We as well as others have shown the capacity of *Bordetella* spp. to grow adhered to abiotic and biotic surfaces as biofilms (Irie et al., 2004; Mishra et al., 2005; Serra et al., 2008, 2011). The two-component sensory transduction system BvgAS, which controls the expression of nearly all known virulence factors, was reported to play an important role in the regulation of biofilm formation for these bacteria (Irie et al., 2004; Mishra et al., 2005). However, the role of biofilm in *B. pertussis* pathogenesis is not yet fully understood and, with a few exceptions (de Gouw et al., 2014), this mode of growth is still largely ignored when new antigens are selected for the formulation of novel pertussis vaccines. Thus, the aim of this study was to compare the biofilm formation by a well-characterized reference strain and eight *B. pertussis* clinical isolates, retrieved from children with pertussis during a 7-years period in Argentina. A comparative analysis, employing proteomics, targeted transcriptomics, and complementary genetic studies including the reference strain Tohama I (which has been sub-cultured *in vitro* since the 1950s), and the clinical isolate Bp2723 (selected by its high capacity of biofilm growth) were carried out to gain insight into the mechanisms responsible for the different behavior of sessile cells exposed to similar external conditions. Our results support the hypothesis that the phenotypic heterogeneity between the reference strain and clinical isolates reflects a specific adaptation of clinical *B. pertussis* to its host. Interestingly, a single nucleotide polymorphism in the *bvgS* gene in all clinical isolates was found, which might implicate an intrinsic feature of the circulating cells. This mutation allowed a fast adaptive response of modulated cells (vir-), incubated under non-modulating conditions, to adhere on abiotic surfaces. Our results foster the hypothesis that these bacteria have developed a repertoire of mechanisms that enable adaptive response to growth adhered to surfaces, allowing these cells to persist in unfriendly environments.

MATERIALS AND METHODS

Bacterial Strains and Culture Conditions

Bordetella pertussis Tohama I strain -isolated in Japan in the 1950s- was obtained from the Collection of Institute Pasteur, Paris, France (CIP 8132); BPSM, a streptomycin resistant (Smr) strain derivative from Tohama I; Bp$_{K705E}$, a mutant derivative of BPSM in which the amino acid lysine (K) at position 705 of the BvgS has been replaced by glutamic acid (E) (Herrou et al., 2009); and eight *B. pertussis* clinical isolates collected at La Plata Children's Hospital (Hospital Interzonal de Agudos Especializado en Pediatría Sor Maria Ludovica, La Plata, Argentina) from 2001 to 2007 (**Table 1**) were used throughout this study. Stock cultures were grown on Bordet–Gengou agar (Difco Laboratories, Detroit, MI, USA) plates supplemented

TABLE 1 | *Bordetella pertussis* reference strain and clinical isolates used in this study.

Strain	Year of isolation	Patient age (weeks)	Source	PtxA	Fim	Prn
Bp Tohama I	1954	–	Japan	2	2	1
Bp2723	2001	8	Argentina	1	3	1
Bp1918	2003	12	Argentina	1	3	2
Bp2930	2004	17	Argentina	1	3	2
Bp3495	2004	4	Argentina	1	3	2
Bp7470	2005	8	Argentina	1	3	2
Bp162	2006	8	Argentina	1	3	2
Bp492	2006	6	Argentina	1	3	2
Bp892	2007	12	Argentina	1	3	2
Bp$_{PK705E}$	2009		Japan	2	2	1
BPSM	1994		Tohama derivative	2	2	1

with 1% w/v Bactopectone (Difco) and 15% v/v defibrinated sheep blood (Instituto Biológico, Ministerio de Salud, La Plata, Argentina; BGA) for 72 h at 37°C. Colonies were cultured for others 48 h and then inoculated into 250-mL Erlenmeyer flasks containing 50-mL of Stainer–Scholte (SS) broth and incubated for 24 h at 37°C on a rotatory shaker (160 rpm). The suspension was used as inoculum for 1-L Erlenmeyer flasks containing 200-mL of SS broth. The initial optical density at 650 nm (OD_{650}) was adjusted to 0.2 and the flaks were incubated for 24 h with agitation. Bacteria were harvested (15°C, 8000 × g, 20 min) at exponential phase, frozen using liquid Nitrogen for 30 s and stored for 48 h at −80°C before being freeze-dried. To study the growth kinetic in liquid medium, samples were collected every 2 h and the OD_{650} was measured. Three independent experiments were performed for each strain, averages and standard deviations of the experimental data obtained are indicated in the figures. Specific growth rates (μ) were obtained from curves ln OD_{650} vs. time.

Biofilm Cultures

The biofilm growth was performed as indicated previously (de Gouw et al., 2014). Briefly, for each *B. pertussis* strain, a bacterial suspension of planktonic cells (24 h of growth), adjusted to an $OD_{650} = 1.0$ (1.0×10^9 CFUs/mL) was incubated with 20 g of polypropylene beads (4.2 mm diameter and 2 mm high, average density: 0.901 g/L, Petroken SA, Argentina) contained in glass tubes for 4 h at 37°C under static conditions. The cell suspension was drained and 20-mL of fresh medium were added to each reactor (glass tubes). The tubes were incubated for 72 h on roller drums under 60 rpm agitation. The culture medium was replaced every 24 h by fresh broth. Before harvest, the beads were washed three times with phosphate buffer saline (PBS) and then used for crystal violet staining. Growth kinetic for sessile cells of *B. pertussis* 2723 strain was evaluated analyzing colony forming units (CFU) every 24 h until 120 h of development. These experiments were performed by triplicate. In a similar way, a semi-continuous biofilm culture was performed to obtain samples able to be analyzed by confocal laser scanning microscopy (CLSM). Duplicates of Bp Tohama I and Bp2723 biofilms were grown attached to glass cover slips. In a first step, bacteria coming from a 24 h planktonic culture were incubated during 4 h with the cover slips inside a Petri dish. Then, broth was changed for fresh medium and incubated under agitation (60 rpm) for 72 h. Every 24 h the medium was changed for fresh broth. After 72 h the cover slips were carefully washed with PBS and stained for CLSM analysis.

For proteomic studies sessile cells were cultured on polypropylene beads contained in column bioreactors as was previously described (Serra et al., 2008) with minor modifications. Briefly, *B. pertussis* Tohama I strain or the clinical isolate *B. pertussis* 2723 were grown in 200-mL SS broth for 24 h and then used to inoculate column reactors. After 4 h of static incubation to allow cell attachment, the suspension was drained to remove non-adhered cells and 200-mL of fresh SS broth were added to each column. Bioreactors were incubated with a constant air supply (0.1 vvm) at 37°C for 72 h (mature biofilm stage). Every 24 h the broth was replaced by fresh medium.

Afterward, polypropylene beads were washed three times with PBS prior to harvest the cells. The biofilm was detached from the surface by lightly agitation on PBS, subsequently, cells were harvested (15°C, 8000 × g, 20 min) and immediately frozen in liquid Nitrogen and stored at −80°C before being freeze-dried.

Fourier Transform Infrared Spectroscopy (FT-IR)

For infrared analysis of biofilms each strain was grown in 6-well plates. After 72 h incubation, the biofilms were washed three times with distilled water and the biomass attached to the wells were suspended in bi-distilled water, adjusting the OD_{650} to 10. Samples were prepared from three independent assays by triplicate in each case. One hundred microliters of each bacterial suspension were transferred to an optical cell of zinc selenide (ZnSe) and vacuum dried (3.6 kPa) to obtain transparent films on the cell. FT-IR absorption spectra from 4,000 to 600 cm^{-1} were acquired with a FT-IR spectrometer (Spectrum One, Perkin Elmer, USA) as was reported (Serra et al., 2007). Infrared analysis was carried out as by means of OPUS 7.0 software (Bruker Optics, USA).

Quantification of Biofilm Biomass

Biofilm biomass was quantified using the crystal violet assay described by O'Toole and Kolter (1998) adapted to the system. The absorbance of the solubilized dye was measured at 590 nm (OD_{590}). Triplicates were made for each strain and t-Student's test was used to compare absorbance against *B. pertussis* Tohama I's biofilm. Samples were considered significantly different when $p \leq 0.05$. CLSM was used to study the architecture and quantitative information of 72 h biofilms. An inverted confocal microscope Carl Zeiss LSM510-Axiovert 100M (Germany) was used as previously reported (Serra et al., 2007). Briefly, biofilms coming from semi-continuous culture, adhered to glass cover slips were first washed very carefully in PBS, and fixed with 4% paraformaldheyde. Then, adhered cells were rinsed in PBS, stained for 20 min with Acridine Orange and washed three times. In order to obtain quantitative information of mature biofilm structure, images were analyzed by COMSTAT software (Heydorn et al., 2000).

Preparation of Soluble Protein Fraction

Cytosolic proteins were obtained following the protocol described by Ehling-Schulz et al. (2002) with minor modifications. Planktonic and sessile freeze-dried bacteria were suspended in detergent buffer containing 7 M urea, 2 M thiourea, 4% CHAPS and 30 mM TRIS (Sigma, St. Louis, MO, USA), cooled and then passed through a precooled French pressure cell (SLM AMINCO) working at 140 mPa two times. Cellular debris was harvested by centrifugation (4°C, 10000 × g, 15 min). The supernatant was transferred to ultracentrifuge tubes (Beckman, USA) and centrifuged at 30000 × g for 40 min at 15°C. The supernatant containing cytosolic proteins was stored in aliquots at −80°C. Protein concentration was estimated using the 2-D Quant kit following the manufacture's protocol (GE Healthcare, Amersham Biosciences).

2D-DIGE

For two-dimensional difference gel electrophoresis (2D-DIGE) protein samples were minimally labeled as previously described (Radwan et al., 2008) with minor modifications. CyDye DIGE™ fluorescent dyes (GE Healthcare Life Science, Munich, Germany) were used to label 33 μg of proteins per sample using 8 nmol dye/mg proteins. For each mode of growth three biological replicates were used. Biofilm and planktonic samples from each strain were labeled with Cye3 and Cye5. The internal standard comprising a pool of equal amounts from all samples was labeled with Cye2. Isoelectric focusing was carried out on an IPGphor III (GE Healthcare, Amersham Biosciences) system using 18 cm IPG Dry strips with linear pH gradients of 4–7 and 6–9 (all GE Healthcare, Amersham Biosciences). The IPG strips were rehydrated over night with rehydration buffer [7 M urea, 2 M thiourea, 4% (w/v) CHAPS, 0.4% (w/v) DTT, 0.5% carrier ampholytes] at room temperature. DTT 0.4% (w/v) and carrier ampholytes 0.5% (v/v) were added to the mixed proteins samples in detergent buffer and the final volume was adjusted to 50 μL with rehydration buffer. Protein samples were then loaded onto the strips via loading cups. pH 4–7 strips were focused for a total of 36 KVh and pH 6–9 strips were focused for a total of 27 KVh. The IPG strips were reduced with 1% w/v DTT for 15 min and alkylated using 4% (w/v) iodoacetamide for 15 min in equilibration buffer (6 M urea, 30% (v/v) glycerin, 2% (w/v) SDS) and SDS-PAGE (12.5% T) was subsequently performed overnight (13 mA per gel) using an Ettan Dalt Six Electrophoresis Chamber (GE Healthcare, Amersham, Biosciences).

Imaging and Data Processing

Fluorescence images of the gels were acquired using a Typhoon 9400 scanner (GE Healthcare). Data analysis was performed with the DeCyder software version 7.0 (GE Healthcare). Spot detection, matching, and normalization were performed using a standard algorithm of the software. Spot matching as well as spot quality of proteins of interest were manually checked eliminating false positives. To assess the reproducibility of gel replicates, principal component analysis (PCA) was carried out employing the DeCyder analysis module. For PCA all spots within the ANOVA 95th confidence interval were included. Spots showing more than threefold changes ($p \leq 0.05$) in abundance between the strains growing as biofilm or in planktonic mode were considered as significant differences, manually excised from silvers stained gels (Blum et al., 1989; Miller and Gemeiner, 1992) and subjected to mass-spectrometry for protein identification.

In-gel Trypsin Digestion and MS-Based Protein Identification

Protein identification was carried out using a matrix-assisted laser desorption/ionization time-of-flight (MALDI–TOF) mass spectrometer (Bruker Daltonics, Ultraflex I) in MS and MS/MS modes. Spot distaining, in gel digestion and sample purification using Zip-Tipμ-C18 (Millipore) pipette tips were performed as reported previously (Blum et al., 1989). Samples were applied on a disposable target plate (Bruker Daltonics, PAC target) pre-spotted with α-cyano-4-hydroxycinnamic acid as matrix. Spectral pre-processing and peak annotation were carried out using FlexAnalysis 3.0 and Biotools 3.2 (Bruker Daltonics). Processed MS and MS/MS spectra were submitted to MASCOT server (Matrix Science) searching the database NCBInr restricting to *B. pertussis* Tohama I strain. Peptide mass fingerprint (PMF) search parameters were set for mass accuracy: <150 ppm, fixed modification: carbamidomethylation, variable modifications: methionine oxidation and acetylation at the protein N-terminal end, and missed cleavages: one. Based on the measured PMF at least one peptide was selected for MS/MS experiments. Search parameters were identical to PMF experiments, except for product ion tolerance (±1.0 Da). A protein was considered as identified, if the scores of database searches clearly exceeded the algorithm's significance threshold ($p < 0.05$) for PMF data and for sequence tag ion analyses of at least one peptide.

RNA Isolation, cDNA Synthesis, and Quantitative Real-time PCR (qRT-PCR)

Total RNA was isolated from planktonic and biofilm bacteria using Trizol reagent (Life Technologies, Invitrogen) following manufacturer's instructions. The RNA was treated with DNase I (Promega, Madison, WI, USA) to remove contaminating DNA and cDNA synthesis was performed using hexamers primers (Promega, Madison, WI, USA) and M-MLV retrotranscriptase enzyme (Invitrogen, Carlsbad, CA, USA) following supplier's protocol. Specific primers were used to determine transcript levels of the selected genes (**Table 2**). SYBR premix (Thermo Scientific) was used for qPCR assays following manufacturer's instructions. Reactions were carried out on triplicate samples, including technical duplicates. Relative mRNA expression ratios of selected genes were normalized to the expression of 16S rRNA. Calculations for comparison between samples were performed using the $\Delta\Delta CT$ (where CT is threshold cycle) method as described by Conover et al. (2012). In the case of differences in primers efficiency a modification was done following the method described by Pfaffl (2001).

DNA Sequencing and Data Analysis

The *bvgS* gene was sequenced for all clinical isolates used in this study. In addition, the following genes and their promoter regions were sequenced for clinical isolate Bp2723: *fhaB*, *ptxS1*, *fim3*, *prn*, *bsp22*, *bcrH2*, *vag8*, *brkA*, and *bvgA* using the primers listed in **Table 2**. For chromosomal PCR amplification the procedure described by Van Loo and Mooi (2002) was employed. Briefly, 1 μL of DNA was added to 19 μL of buffer comprising 50% Hotstar Taq Master mix kit (Qiagen), 1 μM concentration of each primer, and 5–10% dimethyl sulfoxide. Amplification of genes was performed in a Hybaid Omnigene incubator. The PCR fragments were purified with QIAquick PCR purification kit (Qiagen) and sequenced with the primers used for amplification and internal primers (not shown). Sequence reactions were performed with an ABI Prism Big Dye terminator reaction Kit, and the reactions were analyzed using a 377 or 3700 ABI DNA Sequencer (Perkin–Elmer, Applied Biosystems). The resulting sequences were searched against NCBI nucleotide or non-redundant protein database by using BLAST tool.

TABLE 2 | List of primers used for amplification of virulence genes, *bvgA* and *bvgS* genes of *B. pertussis* reference strain and the clinical isolate Bp 2723 through quantitative RT-PCR.

Gene	Primer sequence	Gene	Primer sequence
bvgA	F: 5′AGACCGTCAGCACC TACA	*prn*	F: 5′TGTTCCGCATGAA TGTCTTC
	R: 5′GAGGTCTATCAGTTC CACCA		R: 5′TGTTGGCAAGG GTAAAGGTC
bvgS	F: 5′ATTACGTCAACCG CTACTTC	*fim2*	F: 5′GCCGCAGTTCCGGA TAAA
	R: 5′GTTCAGGATGGACAT CAGTT		R: 5′CGTTTGGGTCGACT CGTTG
vag8	F: 5′GGTTCACTGGTAG AGAGCAC	*fim3*	F: 5′GGTGCGGGAAGC TGTAGTTC
	R: 5′GTTGAGCAGGGACAC ATTAC		R: 5′CGTAGTGGTGGTTGA TGCTGT
bsp22	F: 5′GAACTCGAAAGTGCC TACAC	*ompQ*	F: 5′CAACCAGCCTTTA TGCCTATG
	R: 5′ATGTCCATCTGTTG CGTATT		R: 5′GTCATTCCCACGC CAAAC
brkA	F: 5′GACGCAGGAGTTC AAAAG	*bipA*	F: 5′GACAGCGGTTTC TACCTGGA
	R: 5′TACGAAGCATAGAGG TTGTG		R: 5′CGCCACCTTGAA GTCATTCT
bcrH2	F:5′CTATGCCTGCAGAA GACC	*ptxS*	F: 5′TGTTCCGCATG AATGTCTTC
	R: 5′GAATCTGGATAGAGC GTGAG		R: 5′GACAGCGGTTTC TACCTGGA
fhaB	F: 5′GCCACGATTTCACG GTGCA	*16S*	F: 5′TCAGCATGTCGC GGTGAAT
	R: 5′CAGCGTCGCGTC ATGCT		R: 5′TGTGACGGGC GGTGTGTA

FIGURE 1 | Growth kinetics of *Bordetella pertussis* Tohama I strain and four clinical isolates in Stainer–Scholte (SS) liquid medium. *B. pertussis* clinical isolates and *B. pertussis* Tohama I strain were grown planktonically for 48 h in SS broth. The biomass of clinical isolates after 30 h of culture was almost 70% higher than produced by the reference strain. Results are represented as the mean of values obtained from three independent experiments. The error bars indicate the standard deviations.

RESULTS

Planktonic Growth and Biofilm Formation Capacity on Abiotic Surfaces by *B. pertussis* Tohama I Strain and Clinical Isolates

Using a collection of eight clinical isolates recovered in Argentina from whooping cough patients and the reference strain *B. pertussis* Tohama I, a comparative growth analysis under planktonic conditions in SS broth was performed. **Figure 1** shows the growth kinetics for Tohama I strain and the clinical isolates *B. pertussis* 2723 (Bp 2723), *B. pertussis* 892, *B. pertussis* 1918, and *B. pertussis* 492. The four isolates are depicted as representative for the clinical isolates, which showed similar growth behavior (data not shown). At stationary phase the planktonic biomass of the isolates -measured by optical density- was approximately 70% higher than the biomass reached by *B. pertussis* Tohama I strain. From batch cultures, specific growth rate (μ) for each strain was calculated. *B. pertussis* 2723, as well as the other clinical isolates, exhibited similar specific growth rates of $0.091 \pm 0.003 \, \text{h}^{-1}$ while the μ of Tohama I strain under the same experimental conditions was significantly lower ($0.052 \pm 0.002 \, \text{h}^{-1}$). Next, we compared the adhesion and the mature biofilm biomass of the clinical isolates on abiotic surface. All clinical isolates showed higher adhesion to polypropylene beads after 4 h of static incubation (data not shown) and higher biofilm biomass production after 72 h of culture (mature biofilm) compared to the reference strain (**Figure 2**). However, the final sessile biomass was different for each isolate. The *B. pertussis* 2723 strain was selected for further analysis since it exhibited fivefold more biofilm biomass as well as 70% more biomass under planktonic conditions compared to the reference strain Tohama I. The growth kinetics of this clinical isolate and *B. pertussis* Tohama I strain growing as biofilm were also studied. The clinical isolate Bp2723 showed a final biomass of 2.6×10^{10} CFU/cm^2 while the biomass for reference strain was 2.5×10^{9} CFU/cm^2. In addition, the specific growth rate for Bp2723 was $0.033 \pm 0.002 \, \text{h}^{-1}$ and for Tohama I strain $0.028 \pm 0.001 \, \text{h}^{-1}$. Next, and to differentiate the structure of mature biofilms produced by both Bp2723 clinical isolate and Tohama I strain, micrograph images of biofilms were obtained from CLSM stacks (Supplementary Figure S1). The images were then analyzed by using COMSTAT software. The images obtained showed characteristic biofilm architecture with channels for both, Bp Tohama I and Bp2723. A more profound analysis revealed that the clinical strain produced a bigger biofilm, characterized by a maximum thickness of around the double of that achieved by the reference strain. Five architectural parameters calculated for the two biofilms are provided in **Table 3**. These parameters show not only significant differences in the thickness of the biofilms but also in the covered surface and the roughness coefficient that were higher for the clinical strain biofilm. This analysis revealed an apparent higher complexity of the biofilm produced by Bp2723 compared to the biofilm produced by the reference strain and confirm its enhanced biomass production under this culture condition. To gain insight into the nature of the biofilm developed by the clinical isolate we employed FT-IR spectroscopy for the comparison of the chemical composition of the biofilms developed by the reference strain and the clinical isolate growing in similar environmental

FIGURE 2 | Biofilm formation by *B. pertussis* Tohama I strain and clinical isolates on an abiotic surface. Adhesion and biofilm formation capacity of *B. pertussis* Tohama I reference strain and eight clinical isolates after 72 h of growth (mature biofilm). Biofilm biomass was quantified by staining with crystal violet and absorbance was measurement at OD 590 nm. Each value represents the mean from three independent experiments and the bars indicate standard deviation. Statistically significant differences in Student's test between each clinical isolate and Bp Tohama I biomass absorbance are indicated by symbols when present (*$p < 0.05$).

FIGURE 3 | Principal component analysis (PCA) scores scatter plots of biofilms FT-IR spectra from *B. pertussis* Tohama I strain (green) and *B. pertussis* 2723 isolate (red) in the 4000–650 cm^{-1} region.

conditions. FT-IR spectroscopy, one of the most frequently used spectroscopic techniques to compare biochemical composition among biological samples, in association with PCA showed clear chemical differences between the FT-IR spectra obtained from both biofilms communities (**Figure 3**). Differences in biomass composition were detected between spectral areas assigned to protein and carbohydrates. The protein:carbohydrate ratio was 1.684 for Bp2723 and 1.142 for *B. pertussis* Tohama I strain, respectively. Based on this information we decided to explore the molecular basis for these significant phenotypic differences between the reference strain and the clinical isolate selected.

2D-DIGE Analysis and Protein Identification

Bordetella pertussis Tohama I strain and the clinical isolate *B. pertussis* 2723 were grown in parallel as planktonic cells for 24 h in SS medium (exponential phase) and as biofilm on polypropylene beads for 72 h (mature stage). To investigate the differentially expressed proteins in the reference strain and the clinical isolate under both culture conditions, a comparative

proteomic analysis was performed. The soluble cellular protein fraction was isolated from three replicates per strain and growth condition, and subjected to differential 2D DIGE analysis in two pH ranges (4–7 and 6–9) and a PCA was carried out. As shown in Supplementary Figure S2, the statistical analysis of each biological replicate, clearly indicates a distinct clustering of the four groups demonstrating a high reproducibility between the replicate samples. In addition, the analysis demonstrates that the highest variation was found to be strain-dependent (PC1), whereas PC2 discriminates the different growth conditions. Representative 2D electrophoresis patterns of bacterial proteins based on the internal standard sample for both pH ranges are depicted in **Figure 4**. The global proteome analysis showed that out of a total of 1275 spots analyzed, 65 proteins (5.1%) were differentially expressed in *B. pertussis* 2723 compared to the reference strain growing attached to surface and under planktonic culture conditions. Forty eight differentially expressed protein spots were selected for protein identification based on a combination of selection criteria as published elsewhere (Radwan et al., 2008). MALDI–ToF–MS–MS analysis resulted in the identification of 35 different proteins and/or protein species (Supplementary Table S1). The clinical isolate showed, in comparison to Tohama I strain 10 up-regulated proteins ($p < 0.05$) and five down-regulated proteins ($p < 0.05$) growing under biofilm conditions, and 27 proteins up-regulated and eight proteins down-regulated growing under planktonic conditions. These differentially expressed proteins can be assigned to five functional categories, namely metabolism-energy production, amino acid and protein synthesis, transport, virulence, and cellular process (Supplementary Table S1). More specifically, within the metabolic group, four proteins (aconitate hydratase –*acnB*-, dihydrolipoamide succinyl transferase component of 2-oxoglutarate dehydrogenase complex –*odhB*-, citrate synthase –*gltA*- and enoyl-CoA

TABLE 3 | Biometric parameters obtained from 72 h biofilms formed by *Bordetella pertussis* wild type (Bp Tohama I) and the clinical isolate *B. pertussis* 2723.

Quantitative parameters	Bp Tohama I	Bp2723
Thickness (μm)	10.33 (0.10)	14.55 (0.38)
Maximum thickness (μm)	17.78 (0.005)	36.56 (0.57)
Roughness coefficient	0.18 (0.005)	0.36 (0.07)
Covered surface	0.78 (0.03)	0.94 (0.04)
Surface to biovolume ratio (μm^2/μm^3)	1.72 (0.008)	1.93 (0.42)

The features of the biofilms were quantified using the program COMSTAT. Parameters correspond to mean values. Standard errors are indicated.
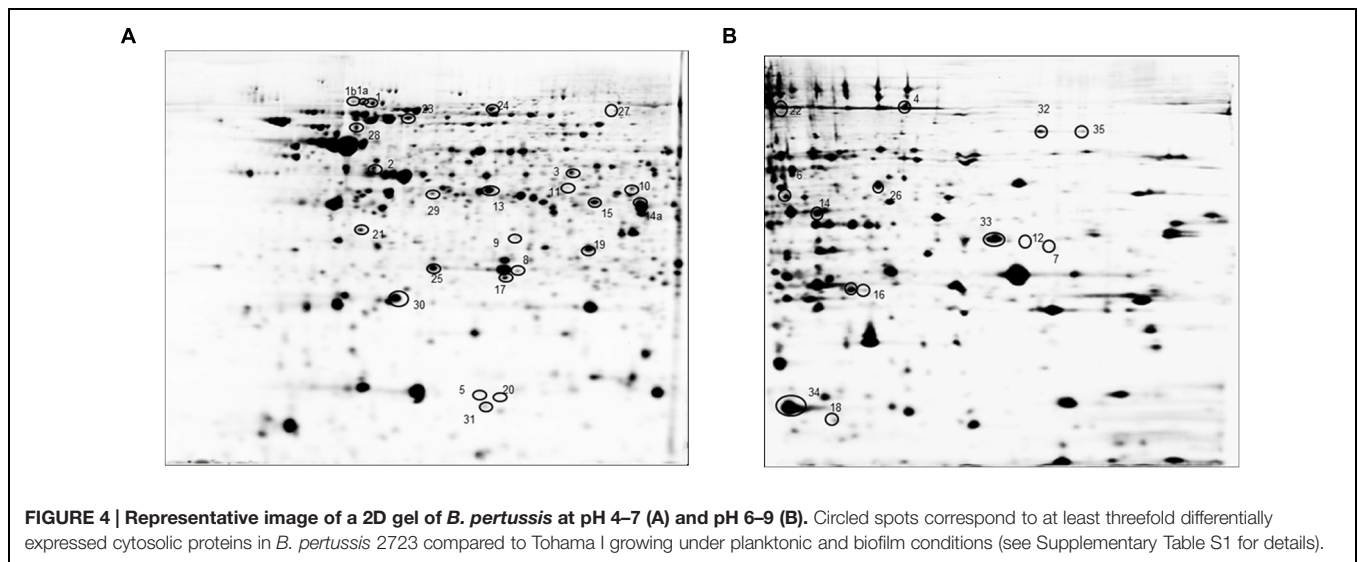
FIGURE 4 | Representative image of a 2D gel of *B. pertussis* at pH 4–7 (A) and pH 6–9 (B). Circled spots correspond to at least threefold differentially expressed cytosolic proteins in *B. pertussis* 2723 compared to Tohama I growing under planktonic and biofilm conditions (see Supplementary Table S1 for details).

hydratase/isomerase –*acnA*-), related to the energy production were found in higher abundance in the clinical isolate under both culture conditions. Three out of four proteins belonging to "amino acids and protein biosynthesis" pathways were found down regulated in the clinical isolate 2723. These proteins correspond to enzymes involved in phenylalanine, tyrosine, tryptophan, and lysine biosynthesis. The fourth protein, cystathionine beta-lyase -*metC*-, which is related to methionine synthesis and sulfur metabolism, was up-regulated in *B. pertussis* 2723. In addition, three proteins involved in amino acids transport [leu/ile/val (branched chain amino acid-) –*Q7VYN1*- binding protein –*livJ*-, amino acid- binding periplasmic protein –*Q7VS83*-, ABC transporter ATP binding protein –*Q7VTG4*-] and four proteins related to stress response and adaptation (putative Zinc protease –*Q7VVY4*-, antioxidant protein –*Q7VZE7*-, chaperone protein DnaJ –*Q7VVY3*-, and protein tex –*Q45388*-) were found to be up-regulated in the clinical isolate under both culture conditions (Supplementary Table S1). Noteworthy, four BvgAS-activated virulence factors, BcrH2, OmpQ, Vag8, and BrkA, were found under both growth conditions in higher abundance in the clinical isolate 2723 than in the reference strain. One of these proteins, namely BcrH2, is a chaperone protein member of the Type III secretion system (T3SS). T3SS correspond to an injection system that delivers virulence factors into the host cells changing its physiological functions (Mattoo et al., 2004; Medhekar et al., 2009). It is known to protrude from bacterial outer membrane of many Gram-negative bacteria. Another BvgAS regulated protein found to be differentially expressed in the clinical isolate is OmpQ, an outer membrane porin protein that so far has no clear assigned function (Finn et al., 1995). In addition, Vag8 and BrkA were also found in higher abundance in the clinical isolate. Vag8 is an autotransporter recently described to participate in serum resistance and described to bind C1; and BrkA, another serum resistance protein of *B. pertussis*, was reported as necessary for efficient colonization of mice (Finn and Amsbaugh, 1998; Barnes and Weiss, 2001; Marr et al., 2011). These latter observations were

particularly interesting since we detected four BvgAS-activated proteins highly expressed in the clinical isolate compared to the Tohama I strain in soluble fraction. As BvgAS-activated genes play an essential role in biofilm development, we next studied the expression of BvgAS regulated adhesins in both strains under biofilm and planktonic culture conditions.

qRT-PCR for *bvgAS* and its Positively Regulated Genes

qRT-PCR was performed to measure mRNA expression levels of BvgAS-regulated genes: *fhaB*, *fim* (*fim2* for Tohama I and *fim3* for Bp2723), *prn*, and *bipA* in cells grown under biofilm and planktonic conditions. In addition *vag8*, *brkA*, *bcrH2*, and *ompQ* genes, encoding the four proteins which showed increased levels of expression by *B. pertussis* 2723 compared to Tohama I strain in the cytosolic proteome (Supplementary Table S1), as well as *bvgS, bvgA,* and *ptx* genes were included in the analysis. This approach revealed a higher mRNA expression levels of *vag8*, *bcrH2*, *prn,* and *brkA* in planktonic cells of the clinical isolate compared to Tohama I strain (**Figure 5A**). Interestingly, significantly higher mRNA expression levels of *fhaB*, *fim3*, *ptxS1*, *bipA*, *vag8*, *prn*, *bvgA*, *bsp22*, and *brkA*, genes were found in sessile cells of the clinical isolate compared to the reference strain ($p < 0.05$; **Figure 5B**). The higher expression of *fhaB*, *prn*, *fim3*, and *bipA* genes, might probably be associated with increased biofilm formation capacity depicted by the clinical isolate, since these genes encode adhesins reported to participate in the development of *B. pertussis* mature biofilm (Serra et al., 2011; Sugisaki et al., 2013; de Gouw et al., 2014).

Sequence Analysis of *bvgA*, *bvgS*, Virulence Genes, and Promoters

To determine whether specific nucleotide changes within virulence genes, or their promoter regions had occurred in Bp2723, the respective DNA sequences and their flanking regions were analyzed. No changes were detected when compared with
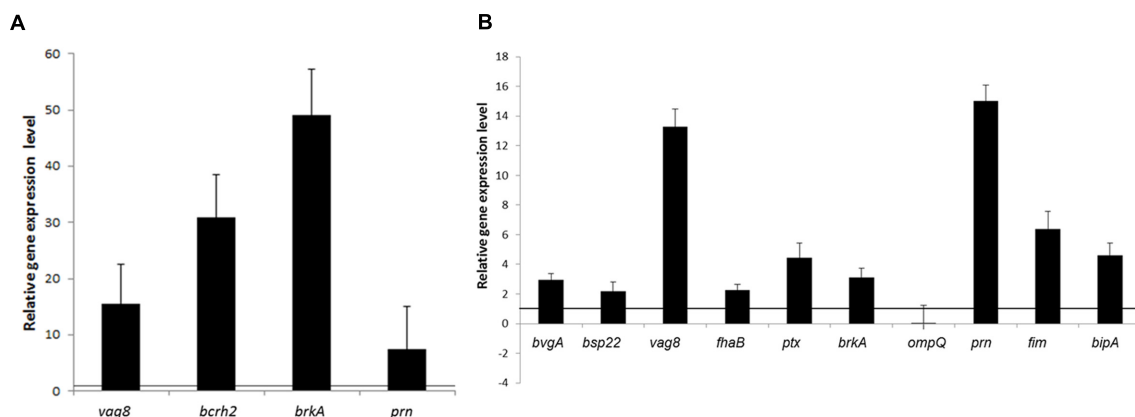
FIGURE 5 | Relative transcription levels of virulence genes. Transcription levels of selected genes positively regulated by the BvgAS system were quantified by qRT-PCR. Transcription levels from the clinical isolate *B. pertussis* 2723 are shown relative to those from the Tohama I strain growing in planktonic **(A)** and biofilm culture **(B)**. The black line indicates equal relative gene expression level, lower values indicate low expression in the clinical strain Bp2723 and higher value indicates up regulation of the genes. In the case of *fim* genes, expression of *fim2* gene in Tohama I was compared to the corresponding *fim 3* of the clinical isolate.

the published sequences of Tohama I strain (GenBank accession number BX470248.1). However, one nucleotide polymorphism, which corresponds to a single exchange of an A for a G at position 2113 of the *bvgS* gene (**Figure 6**), resulting in the replacement of K by E in the amino acid sequence of BvgS protein was observed for Bp 2723. Based on this result, we sequenced the *bvgS* gene of all clinical isolates used in this study and surprisingly found the same mutation in each clinical isolate (**Figure 6**). Interestingly, the same nucleotide exchange has been previously described by Herrou et al. (2009) for strains circulating in the Netherlands.

Effect of E at 705 Position of BvgS Sensor on the Biofilm Formation Ability of *B. pertussis*

To evaluate whether the mutation detected in the *bvgS* gene could affect biofilm formation ability, *B. pertussis* 2723 and *B. pertussis* Tohama I strain were tested for their capacity for biofilm development. Growth was evaluated in tube-bioreactors containing polypropylene beads. The mutant strain Bp_{K705E} with an E at position 705 of BvgS and the wild type strain (BPSM) with a K at position 705 were also included in this analysis. Our results did not reveal any significant differences in the mature biofilm biomass of Bp_{K705E} strain, the wild type

BPSM and Tohama I strain after 72 h of culture (data not shown), suggesting that this point mutation in the *bvgS* gene is not associated to the increased biofilm biomass shown by the clinical isolate. Since this mutation has been previously described to confer a sensitive response to modulatory agents, such as $MgSO_4$ and nicotinic acid, we investigated whether this mutation could trigger a faster attachment of bacteria when they are transferred from a modulatory environment to a non-modulatory one. The four strains described above were therefore modulated using SS culture medium with the addition of 40 mM $MgSO_4$ and then incubated in a non-modulating SS medium under static conditions with polypropylene beads for 4 h. Under this condition, the surface adhesion of each strain was monitored every 30 min. Noteworthy, the clinical isolate Bp2723 and the mutant Bp_{K705E} strain adhered to polypropylene beads significantly faster than Tohama I and BPSM wild type strains (**Figure 7**). The kinetic adhesion of the Tohama I and BPSM wild type strains showed a lag period of 120 min while an increase of biomass attached was already observed for the clinical isolate and Bp_{K705E} strain after 30 min. These findings suggest that the *bvgS* allele coding for the 705 E protein variant might be associated with an accelerated expression of adhesins involved in the initial adhesion steps.
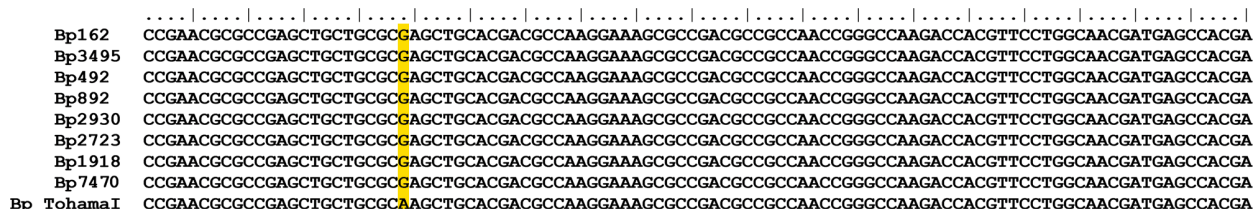

FIGURE 6 | Comparative sequence analysis of *bvgS* gene. The *bvgS*-PAS domain sequence for *B. pertussis* Tohama I strain and the eight *B. pertussis* clinical isolates analyzed are shown. The single nucleotide mutation at position 2113 is highlighted in yellow.
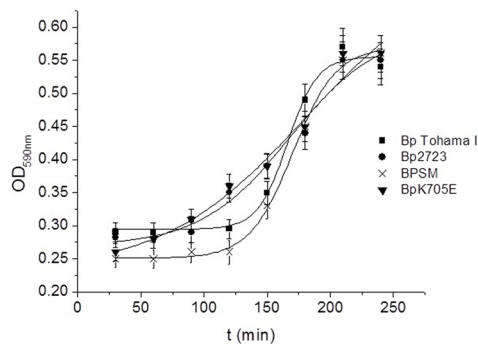
FIGURE 7 | Adhesion kinetic of modulated *B. pertussis* to polypropylene beads. *B. pertussis* Tohama I strain (square spots), Bp 2723 (circle spots), the mutant strain for the BvgS, BpK705E (horizontal triangles) and its wild type strain, BPSM (vertical triangles) were modulated in agar plates and planktonic cultures under 40 mM MgSO$_4$. Then the cells were harvested and suspended in non-modulating media to perform the adhesion assay to polypropylene beads in static conditions. Every 30 min. a sample was taken to measure the adhered biomass to beads using the violet crystal staining.

DISCUSSION

In the current work, we investigated the biofilm formation capacity of eight argentinean *B. pertussis* clinical isolates recovered over a seven years time period in a local children hospital compared to a laboratory adapted strain grown under biofilm conditions. Clinical strains showed an increased ability to grow attached to polypropylene surfaces compared to the laboratory strain. The architecture of biofilm developed by the reference strain was compared with the one developed by the isolate Bp2723 growing in similar culture conditions by using CLSM. We found a marked variability in three dimensional biofilm structures between the two strains studied. The measurement of parameters extracted from confocal stack images analyzed by COMSTAT software such as thickness, roughness coefficient, surface to biovolume ratio, revealed that the biofilm developed by clinical isolate is significantly thicker than the formed by the reference strain. In addition to the differences between the architectures of both biofilms a FT-IR spectroscopic analysis also showed clearly phenotypic variations between these two biofilms. Thus, to better understand the different growth performances between clinical isolates and Tohama I strain we carried out a proteome investigation. The major functional groups of differentially expressed proteins in the clinical isolate included energy metabolism, transport, stress and regulation, and virulence factors. Seven proteins involved in metabolism and energy production were found up-regulated in sessile cells of the clinical isolate *B. pertussis* 2723. Among them the citrate synthase that catalyzes the first reaction in the TCA cycle represents an important control point for determining the metabolic rate of the cell (Park et al., 1994). The higher expression of proteins associated to energy production under respiratory conditions was attributed primarily to the cell biosynthetic needs to produce higher biomass quantities.

Our results indicate that the clinical isolate 2723 could have different metabolic and energetic requirements than the reference strain, which is supported by the higher final biomass reached for this isolate both in liquid medium and biofilm growth conditions. However, this higher capacity of biofilm formation is not associated with a higher growth rate. The TCA cycle is not only a central point in the metabolism of living organisms but also important for the survival of infectious biofilms. Therefore, its inhibition could be a promising strategy for the control of biofilms (Yahya et al., 2014). Similar results were previously reported in comparative proteomic studies of two *Burkholderia cenocepacia* isolates retrieved from a chronically infected cystic fibrosis patient. *B. cenocepacia* isolate obtained after 3 years of persistent infection and antibiotic therapy, showed an up-regulation of citrate synthase (Madeira et al., 2011), which was reported to be important for biofilm formation and virulence (Subramoni et al., 2011). The results from the *B. cenocepacia* study as well as the results from our current study on *B. pertussis*, point toward a tight link between primary metabolism and biofilm formation capability.

Our proteome analysis revealed an increased expression of the Bvg-activated factors BcrH2, OmpQ, BrkA, and Vag8 in the clinical isolate *B. pertussis* 2723. These proteins are positively regulated by the *B. pertussis* BvgAS two-component signal transduction system. This system is known for its key role in the regulation of *Bordetella* virulence gene expression including adhesins and toxins, and it has also been shown to be determinant for the ability of *Bordetella* species to produce biofilms (Irie et al., 2004; Mishra et al., 2005; Serra et al., 2007). If BvgAS is not active *B. pertussis* is unable to adhere to respiratory tract and colonize the host (Bassinet et al., 2000; Scheller et al., 2015). Using quantitative real-time PCR assays we analyzed the relative expression level of adhesin genes, known to be positively regulated by the BvgAS system. After 72 h of biofilm growth these genes were transcribed at higher levels in the clinical isolated compared to Tohama I strain. In addition the *bvgA* regulatory gene showed three times higher transcription in *B. pertussis* 2723 sessile cells compared to reference strain, although no significant increase was found on the protein level. The latter results underpin the importance and value of combinatory analysis on transcriptional and translation levels. Although beyond the scope of our current work, these studies should be expanded in the future to post-translation and metabolic level to gain a holistic picture of pathogen host adaptation mechanisms.

The adhesins tested, namely FHA, Fim3, Prn, and BipA, showed higher transcriptional levels by qRT-PCR in the clinical isolate grown as biofilm compared to the reference strain Tohama I. FHA, one of the main adhesins described for *B. pertussis,* is involved in different steps of biofilm formation *in vitro* and *in vivo* in mouse nasopharynx, contributing not only to the first adhesion to the surface but also enhancing cell–cell interactions (Serra et al., 2011). Recently, we also reported that BipA is a common signature of *B. pertussis* biofilms (de Gouw et al., 2014). Interestingly, although Prn negative strains are now increasingly being isolated from patients with whooping cough (Barkoff et al., 2012; Pawloski et al., 2014), our results showed that this protein is up-regulated in biofilm culture for *B. pertussis* 2723. Proteomics

and targeted transcriptomics approaches provide a picture of the changes between reference strain and a clinical *B. pertussis* isolate growing in similar culture conditions. These results are in agreement with the ones from FT-IR analysis, which show a chemical abundance of proteins in the mature structure of the biofilm produced by the clinical isolate. Therefore, it is tempting to speculate that: (i) the high expression of adhesins, mediating a faster and enhanced attachment, as well as (ii) the higher expression of enzymes involved in energy metabolism, leading to the augmented biomass, are responsible for the robust biofilm structure of the clinical isolate. When bacteria are under stress conditions, they often get together to form biofilms, which suggests that this bacterial lifestyle increases the fitness of the cells in harsh environments. Differential gene expression patterns between Tohama I strain and clinical isolates planktonic cells were previously attributed to either sequence divergence in *cis*-regulatory regions or variation in the levels, activity, or encoding of transcriptional regulatory proteins (Cummings et al., 2006). However, in our current study, the higher transcription of adhesin genes could not be assigned to specific polymorphisms in the sequences of structural genes or promoters, suggesting that trans-acting factors could be involved. A single nucleotide mutation was found in *bvgS* gene of all clinical isolates tested, resulting in an exchange of lysine by glutamic acid at position 705 in the linker domain of the sensor protein. Interestingly, Herrou et al. (2009) reported the same mutation previously. Using an experimental infection model, they demonstrated that this mutation in the BvgS sensor of *B. pertussis* BPSM strain does not lead to a better pulmonary survival of the pathogen, though a faster response to modulatory agents like $MgSO_4$ and nicotinic acid was observed. The latter results are in agreement with our current findings. After being modulated by $MgSO_4$, the clinical isolate Bp2723 were transferred to SS medium and incubated under non-modulating conditions with polypropylene beads. Under the latter culture condition cells showed faster adherence to abiotic surfaces than the reference strain. In our experimental conditions, both, the clinical isolate Bp2723 and the Bp_{K075E} strain showed an accelerated adhesion kinetic to polypropylene beads compared to Tohama I and the wild type BPSM strain. This high phase variation capability might represent an important adaptive advantage during pathogen colonization of its host. Fast adhesion suggests that the phases are tuned to different environmental niches favoring spatially defined regulation. This timing for attachment could promote persistence by protecting bacteria from the clearance occasioned by hydrodynamic forces in upper respiratory tract and the killing activity of host defense mechanisms. It is important to note that the increase of adhered cells (Bp2723 and Bp_{K075E} strain) detected after 30 min of incubation under our experimental conditions is not due to growth on the surface since these sessile bacteria have a reduced specific growth rate ($\mu = 0.033$ h^{-1}, and duplication time of 21.0 h). *B. pertussis* transmission from human to human can probably occur by aerosolized respiratory droplets containing modulated bacteria. Thus, the results shown in our work could be important for the understanding of the rapid adaptation of clinical isolates to new environment. Our combinatory approach of proteomics, targeted transcriptional and genetic analysis revealed that multiple realms of regulation are governing the adaptation of *B. pertussis* to biofilm lifestyle.

CONCLUSION

The divergent biofilm responses between *B. pertussis* clinical isolates and the laboratory adapted reference strain suggest that clinical isolates probably evolved in order to increase their potential and capacity to form biofilm and eventually to adapt rapidly to the fluctuations that they encounter at the site of infection. To date the emergence of pertussis remains a critical issue that should gear researchers to develop novel control measures by considering particularly the biofilm as a *B. pertussis* way of life.

AUTHOR CONTRIBUTIONS

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2015.01352

REFERENCES

Barkoff, A. M., Mertsola, J., Guillot, S., Guiso, N., Berbers, G., and He, Q. (2012). Appearance of *Bordetella pertussis* strains not expressing the vaccine antigen pertactin in Finland. *Clin. Vaccine Immunol.* 19, 1703–1704. doi: 10.1128/CVI.00367-12

Barnes, M. G., and Weiss, A. A. (2001). BrkA protein of *Bordetella pertussis* inhibits the classical pathway of complement after C1 deposition. *Infect. Immun.* 69, 3067–3072. doi: 10.1128/IAI.69.5.3067-3072.2001

Bassinet, L., Gueirard, P., Maitre, B., Housset, B., Gounon, P., and Guiso, N. (2000). Role of adhesins and toxins in invasion of human tracheal epithelial cells by *Bordetella pertussis*. *Infect. Immun.* 68, 1934–1941. doi: 10.1128/IAI.68.4.1934-1941.2000

Berbers, G. A., de Greeff, S. C., and Mooi, F. R. (2009). Improving pertussis vaccination. *Hum. Vaccin.* 5, 497–503. doi: 10.4161/hv.8112

Blum, H., Gross, H. J., and Beier, H. (1989). The expression of the TMV-specific 30-kDa protein in tobacco protoplasts is strongly and selectively enhanced by actinomycin. *Virology* 169, 51–61. doi: 10.1016/0042-6822(89)90040-8

Bottero, D., Gaillard, M. E., Fingermann, M., Weltman, G., Fernandez, J., Sisti, F., et al. (2007). Pulsed-field gel electrophoresis, pertactin, pertussis toxin S1 subunit polymorphisms, and surfaceome analysis of vaccine and clinical *Bordetella pertussis* strains. *Clin. Vaccine Immunol.* 14, 1490–1498. doi: 10.1128/CVI.00177-07

Brinig, M. M., Cummings, C. A., Sanden, G. N., Stefanelli, P., Lawrence, A., and Relman, D. A. (2006). Significant gene order and expression differences in *Bordetella pertussis* despite limited gene content variation. *J. Bacteriol.* 188, 2375–2382. doi: 10.1128/JB.188.7.2375-2382.2006

Cherry, J. D. (1999). Epidemiological, clinical, and laboratory aspects of pertussis in adults. *Clin. Infect. Dis.* 28, 112–117. doi: 10.1086/515058

Conover, M. S., Redfern, C. J., Ganguly, T., Sukumar, N., Sloan, G., Mishra, M., et al. (2012). BpsR modulates *Bordetella biofilm* formation by negatively regulating the expression of the Bps polysaccharide. *J. Bacteriol.* 194, 233–242. doi: 10.1128/JB.06020-11

Costerton, J. W., Stewart, P. S., and Greenberg, E. P. (1999). Bacterial biofilms: a common cause of persistent infections. *Science* 284, 1318–1322. doi: 10.1126/science.284.5418.1318

Cummings, C. A., Bootsma, H. J., Relman, D. A., and Miller, J. F. (2006). Species- and strain-specific control of a complex, flexible regulon by *Bordetella* BvgAS. *J. Bacteriol.* 188, 1775–1785. doi: 10.1128/JB.188.5.1775-1785.2006

de Gouw, D., Serra, D. O., de Jonge, M. I., Hermans, P. W. M., Wessels, H. J. C. T., Zomer, A., et al. (2014). The vaccine potential of *Bordetella pertussis* biofilm derived membrane proteins. *Emerg. Microbes Infect.* 3:e58. doi: 10.1038/emi.2014.58

de Greeff, S. C., Mooi, F. R., Westerhof, A., Verbakel, J. M., Peeters, M. F., Heuvelman, C. J., et al. (2010). Pertussis disease burden in the household: how to protect young infants. *Clin. Infect. Dis.* 50, 1339–1345. doi: 10.1086/652281

Ehling-Schulz, M., Schulz, S., Wait, R., Gorg, A., and Scherer, S. (2002). The UV-B stimulon of the terrestrial cyanobacterium Nostoc commune comprises early shock proteins and late acclimation proteins. *Mol. Microbiol.* 46, 827–843. doi: 10.1046/j.1365-2958.2002.03209.x

Elomaa, A., He, Q., Minh, N. N., and Mertsola, J. (2009). Pertussis before and after the introduction of acellular pertussis vaccines in Finland. *Vaccine* 27, 5443–5449. doi: 10.1016/j.vaccine.2009.07.010

Finn, T. M., and Amsbaugh, D. F. (1998). Vag8, a *Bordetella pertussis* bvg-regulated protein. *Infect. Immun.* 66, 3985–3989.

Finn, T. M., Li, Z., and Kocsis, E. (1995). Identification of a *Bordetella pertussis* bvg-regulated porin-like protein. *J. Bacteriol.* 177, 805–809.

Gurung, J., Khyriem, A. B., Banik, A., Lyngdoh, W. V., Choudhury, B., and Bhattacharyya, P. (2013). Association of biofilm production with multidrug resistance among clinical isolates of *Acinetobacter baumannii* and *Pseudomonas aeruginosa* from intensive care unit. *Indian J. Crit. Care Med.* 17, 214–218. doi: 10.4103/0972-5229.118416

Hall-Stoodley, L., Costerton, J. W., and Stoodley, P. (2004). Bacterial biofilms: from the natural environment to infectious diseases. *Nat. Rev. Microbiol.* 2, 95–108. doi: 10.1038/nrmicro821

He, Q., Makinen, J., Berbers, G., Mooi, F. R., Viljanen, M. K., Arvilommi, H., et al. (2003). *Bordetella pertussis* protein pertactin induces type-specific antibodies:

one possible explanation for the emergence of antigenic variants? *J. Infect. Dis.* 187, 1200–1205. doi: 10.1086/368412

Heikkinen, E., Kallonen, T., Saarinen, L., Sara, R., King, A. J., Mooi, F. R., et al. (2007). Comparative genomics of *Bordetella pertussis* reveals progressive gene loss in Finnish strains. *PLoS ONE* 2:e904. doi: 10.1371/journal.pone.0000904

Hellenbrand, W., Beier, D., Jensen, E., Littmann, M., Meyer, C., Oppermann, H., et al. (2009). The epidemiology of pertussis in Germany: past and present. *BMC Infect. Dis.* 9:22. doi: 10.1186/1471-2334-9-22

Herrou, J., Debrie, A. S., Willery, E., Renauld-Mongenie, G., Locht, C., Mooi, F., et al. (2009). Molecular evolution of the two-component system BvgAS involved in virulence regulation in Bordetella. *PLoS ONE* 4:e6996. doi: 10.1371/journal.pone.0006996

Hewlett, E. L., and Edwards, K. M. (2005). Clinical practice. Pertussis–not just for kids. *N. Engl. J. Med.* 352, 1215–1222. doi: 10.1056/NEJMcp041025

Heydorn, A., Nielsen, A. T., Hentzer, M., Sternberg, C., Givskov, M., Ersbøll, B. K., et al. (2000). Quantification of biofilm structures by the novel computer program COMSTAT. *Microbiology* 146, 2395–2407. doi: 10.1099/00221287-146-10-2395

Hoiby, N., Bjarnsholt, T., Givskov, M., Molin, S., and Ciofu, O. (2010). Antibiotic resistance of bacterial biofilms. *Int. J. Antimicrob. Agents* 35, 322–332. doi: 10.1016/j.ijantimicag.2009.12.011

Irie, Y., Mattoo, S., and Yuk, M. H. (2004). The Bvg virulence control system regulates biofilm formation in *Bordetella bronchiseptica*. *J. Bacteriol.* 186, 5692–5698. doi: 10.1128/JB.186.17.5692-5698.2004

Ito, A., Taniuchi, A., May, T., Kawata, K., and Okabe, S. (2009). Increased antibiotic resistance of *Escherichia coli* in mature biofilms. *Appl. Environ. Microbiol.* 75, 4093–4100. doi: 10.1128/AEM.02949-08

King, A. J., van Gorkom, T., Pennings, J. L., van der Heide, H. G., He, Q., Diavatopoulos, D., et al. (2008). Comparative genomic profiling of Dutch clinical *Bordetella pertussis* isolates using DNA microarrays: identification of genes absent from epidemic strains. *BMC Genomics* 9:311. doi: 10.1186/1471-2164-9-311

Lavine, J. S., King, A. A., and Bjornstad, O. N. (2011). Natural immune boosting in pertussis dynamics and the potential for long-term vaccine failure. *Proc. Natl. Acad. Sci. U.S.A.* 108, 7259–7264. doi: 10.1073/pnas.1014394108

Madeira, A., Santos, P. M., Coutinho, C. P., Pinto-de-Oliveira, A., and Sa-Correia, I. (2011). Quantitative proteomics (2-D DIGE) reveals molecular strategies employed by Burkholderia cenocepacia to adapt to the airways of cystic fibrosis patients under antimicrobial therapy. *Proteomics* 11, 1313–1328. doi: 10.1002/pmic.201000457

Marr, N., Shah, N. R., Lee, R., Kim, E. J., and Fernandez, R. C. (2011). *Bordetella pertussis* autotransporter Vag8 binds human C1 esterase inhibitor and confers serum resistance. *PLoS ONE* 6:e20585. doi: 10.1371/journal.pone.0020585

Mattoo, S., Yuk, M. H., Huang, L. L., and Miller, J. F. (2004). Regulation of type III secretion in Bordetella. *Mol. Microbiol.* 52, 1201–1214. doi: 10.1111/j.1365-2958.2004.04053.x

Medhekar, B., Shrivastava, R., Mattoo, S., Gingery, M., and Miller, J. F. (2009). Bordetella Bsp22 forms a filamentous type III secretion system tip complex and is immunoprotective in vitro and in vivo. *Mol. Microbiol.* 71, 492–504. doi: 10.1111/j.1365-2958.2008.06543.x

Miller, I., and Gemeiner, M. (1992). Two-dimensional electrophoresis of cat sera: protein identification by crossreacting antibodies against human serum proteins. *Electrophoresis* 13, 450–453. doi: 10.1002/elps.1150130193

Mishra, M., Parise, G., Jackson, K. D., Wozniak, D. J., and Deora, R. (2005). The BvgAS signal transduction system regulates biofilm development in Bordetella. *J. Bacteriol.* 187, 1474–1484. doi: 10.1128/JB.187.4.1474-1484.2005

Mooi, F. R., Van Der Maas, N. A. T., and De Melker, H. E. (2014). Pertussis resurgence: waning immunity and Pathogen Adaptation-Two Sides of the same coin. *Epidemiol. Infect.* 142, 685–694. doi: 10.1017/S0950268813000071

Mooi, F. R., van Loo, I. H., and King, A. J. (2001). Adaptation of *Bordetella pertussis* to vaccination: a cause for its reemergence? *Emerg. Infect. Disl.* 7, 526–528. doi: 10.1017/S0950268813000071

Mooi, F. R., van Loo, I. H., van Gent, M., He, Q., Bart, M. J., Heuvelman, K. J., et al. (2009). *Bordetella pertussis* strains with increased toxin production associated with pertussis resurgence. *Emerg. Infect. Dis.* 15, 1206–1213. doi: 10.3201/eid1508.081511

Mooi, F. R., van Oirschot, H., Heuvelman, K., van der Heide, H. G., Gaastra, W., and Willems, R. J. (1998). Polymorphism in the *Bordetella pertussis* virulence

factors P.69/pertactin and pertussis toxin in The Netherlands: temporal trends and evidence for vaccine-driven evolution. *Infect. Immun.* 66, 670–675.

O'Toole, G. A., and Kolter, R. (1998). Initiation of biofilm formation in *Pseudomonas* fluorescens WCS365 proceeds via multiple, convergent signalling pathways: a genetic analysis. *Mol. Microbiol.* 28, 449–461. doi: 10.1046/j.1365-2958.1998.00797.x

Park, S. J., McCabe, J., Turna, J., and Gunsalus, R. P. (1994). Regulation of the citrate synthase (gltA) gene of *Escherichia coli* in response to anaerobiosis and carbon supply: role of the arcA gene product. *J. Bacteriol.* 176, 5086–5092.

Pawloski, L. C., Queenan, A. M., Cassiday, P. K., Lynch, A. S., Harrison, M. M., and Shang, W. (2014). Prevalence and molecular characterization of pertactin-deficient *Bordetella pertussis* in the US. *Clin. Vaccine Immunol.* 21, 119–125. doi: 10.1128/CVI.00717-13

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 29:e45. doi: 10.1093/nar/29.9.e45

Radwan, M., Miller, I., Grunert, T., Marchetti-Deschmann, M., Vogl, C., O'Donoghue, N., et al. (2008). The impact of tyrosine kinase 2 (Tyk2) on the proteome of murine macrophages and their response to lipopolysaccharide (LPS). *Proteomics* 8, 3469–3485. doi: 10.1002/pmic.200800260

Scheller, E. V., Melvin, J. A., Sheets, A. J., and Cotter, P. A. (2015). Cooperative roles for fimbria and filamentous hemagglutinin in bordetella adherence and immune modulation. *mBio* 6:e500–e515. doi: 10.1128/mBio.00500-15

Serra, D., Bosch, A., Russo, D., Rodríguez, M. E., Zorreguieta, A., Schmitt, J., et al. (2007). Continuous nondestructive monitoring of *Bordetella pertussis* biofilms by Fourier transform infrared spectroscopy and other corroborative techniques. *Anal. Bioanal. Chem.* 387, 1759–1767. doi: 10.1007/s00216-006-1079-9

Serra, D. O., Conover, M. S., Arnal, L., Sloan, G. P., Rodriguez, M. E., Yantorno, O. M., et al. (2011). FHA-Mediated cell-substrate and cell-cell adhesions are critical for *Bordetella pertussis* biofilm formation on abiotic surfaces and in the mouse nose and the trachea. *PLoS ONE* 6:e28811. doi: 10.1371/journal.pone.0028811

Serra, D. O., Lucking, G., Weiland, F., Schulz, S., Gorg, A., Yantorno, O. M., et al. (2008). Proteome approaches combined with Fourier transform infrared spectroscopy revealed a distinctive biofilm physiology in *Bordetella pertussis*. *Proteomics* 8, 4995–5010. doi: 10.1002/pmic.200800218

Subramoni, S., Nguyen, D. T., and Sokol, P. A. (2011). Burkholderia cenocepacia ShvR-regulated genes that influence colony morphology, biofilm formation, and virulence. *Infect. Immun.* 79, 2984–2997. doi: 10.1128/IAI.00170-11

Sugisaki, K., Hanawa, T., Yonezawa, H., Osaki, T., Fukutomi, T., Kawakami, H., et al. (2013). Role of (p)ppGpp in biofilm formation and expression of filamentous structures in *Bordetella pertussis*. *Microbiology* 159, 1379–1389. doi: 10.1099/mic.0.066597-0

Van Loo, I. H., and Mooi, F. R. (2002). Changes in the Dutch *Bordetella pertussis* population in the first 20 years after the introduction of whole-cell vaccines. *Microbiology* 148, 2011–2018. doi: 10.1099/00221287-148-7-2011

Yahya, M. F. Z. R., Hamid, U. M. A., Norfatimah, M. Y., and Kambol, R. (2014). In silico analysis of essential tricarboxylic acid cycle enzymes from biofilm-forming bacteria. *Trends Bioinformatics* 7:26. doi: 10.3923/tb.2014.19.26

Candida albicans Shaving to Profile Human Serum Proteins on Hyphal Surface

Elvira Marín[1], Claudia M. Parra-Giraldo[1][†], Carolina Hernández-Haro[1], María L. Hernáez[2], César Nombela[1,3][†], Lucía Monteoliva[1,3]* and Concha Gil[1,2,3]

[1] *Departamento de Microbiología II, Facultad de Farmacia, Universidad Complutense de Madrid, Madrid, Spain,* [2] *Unidad de Proteómica, Facultad de Farmacia, Universidad Complutense de Madrid, Madrid, Spain,* [3] *Instituto Ramón y Cajal de Investigación Sanitaria, Madrid, Spain*

Candida albicans is a human opportunistic fungus and it is responsible for a wide variety of infections, either superficial or systemic. *C. albicans* is a polymorphic fungus and its ability to switch between yeast and hyphae is essential for its virulence. Once *C. albicans* obtains access to the human body, the host serum constitutes a complex environment of interaction with *C. albicans* cell surface in bloodstream. To draw a comprehensive picture of this relevant step in host-pathogen interaction during invasive candidiasis, we have optimized a gel-free shaving proteomic strategy to identify both, human serum proteins coating *C. albicans* cells and fungi surface proteins simultaneously. This approach was carried out with normal serum (NS) and heat inactivated serum (HIS). We identified 214 human and 372 *C. albicans* unique proteins. Proteins identified in *C. albicans* included 147 which were described as located at the cell surface and 52 that were described as immunogenic. Interestingly, among these *C. albicans* proteins, we identified 23 GPI-anchored proteins, Gpd2 and Pra1, which are involved in complement system evasion and 7 other proteins that are able to attach plasminogen to *C. albicans* surface (Adh1, Eno1, Fba1, Pgk1, Tdh3, Tef1, and Tsa1). Furthermore, 12 proteins identified at the *C. albicans* hyphae surface induced with 10% human serum were not detected in other hypha-induced conditions. The most abundant human proteins identified are involved in complement and coagulation pathways. Remarkably, with this strategy, all main proteins belonging to complement cascades were identified on the *C. albicans* surface. Moreover, we identified immunoglobulins, cytoskeletal proteins, metabolic proteins such as apolipoproteins and others. Additionally, we identified more inhibitors of complement and coagulation pathways, some of them serpin proteins (serine protease inhibitors), in HIS vs. NS. On the other hand, we detected a higher amount of C3 at the *C. albicans* surface in NS than in HIS, as validated by immunofluorescence.

Keywords: *Candida albicans*, shaving, surface proteins, human serum, complement pathways, coagulation pathways, GPI-anchored proteins, host-pathogen interaction

INTRODUCTION

The yeast *Candida albicans* is the most important opportunistic fungi, causing a wide variety of infections ranging from superficial to invasive candidiasis, and it is often found in the mucosal microbiota. *C. albicans* infections produce high morbidity and mortality in intensive care, post-surgery and cancer patients or other types of immunocompromised patients (Leleu et al., 2002; Almirante et al., 2005; Hube, 2006; Brown et al., 2012). The high mortality of these infections is determined by the diagnostic limitations, the scarcity of antifungal agents and the emergence of resistance to them (Viudes et al., 2002; Wilson et al., 2002). For these reasons, the incidence of invasive candidiasis is still high.

The cell wall of *C. albicans* is a dynamic and complex multi-layered structure located external to the plasma membrane. It is responsible for maintenance of the shape that characterizes each growth form (mainly yeast and hyphae) of the fungus (Klis et al., 2009). The cell wall mediates the first contact with the environment and integrates multiple cues into complex signaling networks that are coordinated by various transcription factors. Consequently, *C. albicans* differentially expresses cell surface proteins and virulence factors. The dimorphic transition that is the ability of *C. albicans* to reversibly switch from yeast to hyphal growth is essential for virulence; strains that are locked in either form are avirulent (Yan et al., 2013; Lu et al., 2014). The hyphal form allows the pathogen to penetrate into tissues to acquire nutrients or escape from the host defense, and yeast cells disseminate across the human body. There are many signals capable of inducing the dimorphic phenotypic switch: N-acetyl-D-glucosamine, physiological temperature and pH (37°C and neutral), high CO_2 concentration, hypoxia and nutrient starvation; nonetheless, growing in human serum is the most physiological condition to study this process (Kumamoto and Vinces, 2005; Karkowska-Kuleta et al., 2009; Sudbery, 2011; Mayer et al., 2013; Ene et al., 2014).

C. albicans is able to adhere to host cells and tissues; for this reason, it exposes surface proteins such as adhesins and many other pathogenic factors (Chaffin, 2008). Adhesion to host tissue is a prerequisite for tissue invasion and infection. Among adhesins, there are the Als family (agglutinin-like sequence), the Hwp family and the Iff/Hyr family (de Groot et al., 2013). Pathogenic factors include tissue-digesting hydrolytic enzymes such as the Sap family (Secreted Aspartyl Proteinases) (Kretschmar et al., 2002; Buu and Chen, 2014), lipases and phosholipases. The Sap family is composed of 10 proteins with different expression patterns; they act optimally at acidic pH. Their main role is related to digesting molecules for nutrient acquisition and for digesting or distorting host cell membranes to avoid or resist antimicrobial attack by the host immune system (Naglik et al., 2003). Some adhesins are attached to the wall through a C-terminus glycosylphosphatidylinositol (GPI) anchorage sequence (De Groot et al., 2003; Pardo et al., 2004; Klis et al., 2009).

Human serum is a very complex body fluid; its qualitative and quantitative composition was a hot topic of study during years (Anderson and Anderson, 2002; Mitchell, 2010). In the Plasma Proteome Database, proteomic data of 10,546 proteins detected in human serum and plasma are collected (Nanjappa et al., 2014). The dynamic range of abundance of these proteins has more than 10 orders of magnitude of variation. The complement system is an efficient component of the innate immune response and establishes a link with the adaptive system. The three conventional pathways by which complement is activated (classical, alternative and lectin) are composed of around 30 proteins (Engleberg et al., 2012). It mediates several physiological processes such as tissue regeneration and clearance of immune complexes. C2 and factor B (FB) are the main heat-sensitive complement components (Soltis et al., 1979). C2 is the central component in the activation of classical and lectin pathways and FB is for alternative pathway; both are necessary for the formation of C3- and C5-convertases in the respective complement cascades aforementioned.

C. albicans develops several strategies to evade the complement system, delay the detection by the host immune systems and even capable of manipulate the host defenses for its own purpose (Collette and Lorenz, 2011). In this way, *C. albicans* acquires host-derived inhibitory proteins such as factor H (FH), factor H-like protein 1 (FHL-1) and C4b-binding protein (C4BP), to prevent activation of complement pathways on its surface. These three protein regulators are closely related molecules composed of variable numbers of short consensus repeat (SCR) domains. There are four *C. albicans* proteins described with the ability to fix FH on their surface: glycerol 3-phosphate-dehydrogenase 2 (Gpd2), high-affinity gluthathione transporter 1 (Hgt1), phosphoglycerate mutase 1 or complement regulator-acquiring surface protein 1 (Gpm1/CRASP1) and pH-regulated antigen 1 (Pra1/CRASP2) (Poltermann et al., 2007; Luo et al., 2009, 2013). With the exception of Hgt1, the other three are able also to bind plasminogen. Additionally, Pra1 and Hgt1 bind C4BP (Lesiak-Markowicz et al., 2011; Luo et al., 2011). Factor H mediates binding to polyanions (sialic acids and human heparin) present on *C. albicans* cell wall (Soares et al., 2000; Rodríguez de Córdoba et al., 2004; Green et al., 2013). Pra1 also binds fibrinogen and C3 directly (Zipfel et al., 2011).

The aim of this study is to know the human serum protein coat on *C. albicans* surface as well as which *C. albicans* proteins could be candidates for diagnostic markers, potential vaccines or therapeutic targets, as they are detected when human serum is present. The interaction between human serum proteins and other microorganism surface was tested by different proteomic approaches in *Staphylococcus aureus* and *Streptococcus pyogenes* (Dreisbach et al., 2011; Sjöholm et al., 2014). In this work, we carried out a gel-free proteomic approach by shaving live *C. albicans* cells with trypsin after serum interaction to obtain exposed peptides of *C. albicans* and human proteins. With this shaving strategy that was applied in our laboratory to *C. albicans* and *S. cerevisiae* cultured in synthetic media (Hernáez et al., 2010; Insenser et al., 2010; Gil-Bona et al., 2015), it is not necessary to perform sub-cellular fractionation. In the present work, we have optimized this protocol for the detection of *C. albicans* and human proteins together in conditions that mimic *in vivo* interaction during systemic infections; overall, 372 *C. albicans* and 214 human proteins were identified. Among

C. albicans proteins, 147 are described as located at the cell surface. Interestingly Pra1 and Gpd2 are able to attach inhibitors of complement cascades and plasminogen; while a further 7 proteins are also capable of attaching plasminogen. Regarding human proteins, all main proteins of complement pathways were identified, and many proteins involved in coagulation and metabolism, as well as immunoglobulins or cytoskeletal proteins. Furthermore, in this work proteins of complement and coagulation pathways were observed surrounded the surface of *C. albicans* hyphal cells.

MATERIALS AND METHODS

Candida albicans Strain and Growth Conditions

The *C. albicans* strain used in this work was SC5314, from a clinical isolate (Gillum et al., 1984). This strain was grown on YPD plates (2% D-glucose, 2% peptone, 1% yeast extract, and 2% agar) and incubated at 30°C. A colony was picked and grown in YPD medium at 30°C and 200 rpm during 6–8 h. The $OD_{600\,nm}$ was measured and diluted on minimum medium (MM) at 0.00003 $OD_{600\,nm}$ (2% D-glucose, 0.17% yeast nitrogen base, 0.5% $(NH_4)_2SO_4$, 0.192% synthetic complete mixture (Kaiser) drop-out -ura and 0.01% uracil). Culture was grown at 30°C and 200 rpm overnight to obtain and collected at a concentration ~10^6 cells/ml. 1.5×10^7 cells were resuspended in 9 ml of Lee medium at pH 6.7 (Lee et al., 1975). One milliliter of human serum was added and the culture was grown during 5 h at 37°C and 200 rpm. Human sera were from healthy donor volunteers who did not have any clinical or microbiological evidence of infection. They were used in the experiments as normal serum (NS) or heat inactivated serum (HIS). Serum inactivation was done by heating at 56°C during 30 min in a water bath.

Surface Shaving

Surface shaving procedure was adapted from previous experiment of our group (Hernáez et al., 2010; Gil-Bona et al., 2015). Briefly, after incubation of *C. albicans* with 10% human serum (normal serum-NS or heat inactivated serum-HIS), the cultures were centrifuged at 3500 rpm 5 min. Cell pellet was resuspended in 1 ml of phosphate buffer saline (PBS) with 0.1% Tween 20, centrifuged again and washed 6 times more with PBS. The last pellet was resuspended in 400 µl of ammonium bicarbonate (AMBI buffer; NH_4HCO_3) 25 mM pH 8.0 and 7.5 µl dithiothreitol (DTT) 1 M and 9 µg of recombinant sequencing grade trypsin (Roche) were added. DTT was added during the digestion as reducing agent to consent deep digestion of non-covalently or disulphide bridges associated protein. The samples were incubated at 37°C, 30 min and 600 rpm. After incubation, samples were centrifuged at 5000 rpm 5 min and the supernatant was filtered with a filter unit of 0.22 µm. The cell pellet was resuspended in 400 µl of fresh AMBI buffer and 100 µl of trifluoroacetic acid (TFA) 0.1% (v/v) were added to stop the proteolytic reaction. It was centrifuged and the supernatant was filtered again. Both peptide-supernatants were put together and processed for further proteomic analysis. Subsequently,

originated peptides were cleaned up with a Poros R2 resin (ABSciex, Framingham, MA). Peptides were eluted with 80% acetonitrile in 0.1% TFA, dried in a Speed-Vac and resuspended in 0.1% formic acid. The samples were stored at −20°C before nano-LC-MS/MS analysis.

LTQ-Orbitrap Velos Analysis, Protein Identification, and Bioinformatics Analysis

Peptides were analyzed by nano-LC-MS/MS analysis. Peptide samples were concentrated and desalted using C18-A1 ASY-column 2 cm pre-column (Thermo Scientific) and then eluted onto a Biosphere C18 column (Nano-Separations). Peptides were separated with a 140 min gradient (110 min from 0 to 40% Buffer B; Buffer A: 0.1% formic acid/2% acetonitrile; Buffer B: 0.1% formic acid in acetonitrile) at a flow-rate of 250 nl/min on a nano-Easy HPLC (Proxeon) coupled to a nano-electrospray ion source (Proxeon). Mass spectrometry experiments were performed using a LTQ-Orbitrap Velos (Thermo Scientific) in the positive ion mode. Full-scan MS spectra (m/z 400/1400) were acquired in the Orbitrap apparatus with a target value of 1,000,000 at a resolution of 60,000 at m/z 400 and the 15 most intense ions were selected for collision induced dissociation (CID) fragmentation in the LTQ with a target value of 10,000 and normalized collision energy of 38%. Precursor ion charge state screening and monoisotopic precursor selection were enabled. Singly charged ions and unassigned charge states were rejected. Dynamic exclusion was enabled with a repeat count of 1 and exclusion duration of 30 s.

Protein identification was carried out with mass spectra raw files using a licensed version of search engine MASCOT 2.3.0 with Proteome Discoverer software version 1.4.1.14 (Thermo Scientific). Searchers were made against *C. albicans* SC5314 (6221 sequences) and human database (20233 sequences) available on *Candida* Genome Database (Assembly 21 of CGD, http://www.candidagenome.org/) and Uniprot-SwissProt (http://www.uniprot.org) respectively, to identify peptides and proteins. Search parameters were oxidized methionine as variable modification, peptide mass tolerance 10 ppm, 2 missed trypsin cleavage sites and MS/MS fragment mass tolerance of 0.8 Da. In all protein identification, the FDR was <1%, using a Mascot Percolator, with a q-value of 0.01. Gene Ontology (GO) terms at CGD (http://www.candidagenome.org/cgi-bin/ GO/goTermFinder) were used to classify *C. albicans* proteins.

As an estimation of the relative protein abundances the normalized spectral abundance factor (NSAF) was used (Zybailov et al., 2007), and the average of the normalized values was calculated. Mass spectrometry proteomics data have been deposited in *C. albicans* PeptideAtlas (Vialas et al., 2014) with the data set identifier PASS00446.

Cell Permeability

Before and after trypsin treatment, *C. albicans* cell wall permeability was evaluated by staining with propidium iodide (PI). 10^6 cells/ml were incubated with 10 µl of PI 5 mM and the positive fluorescent cells (red staining) were checked, at least 200 cells of each sample were counted in a fluorescence microscope.

Cells treated with 70% ethanol-PBS (v/v) were used as positive control.

Epifluorescence and Confocal Fluorescence Microscopy

For detecting complement proteins on *C. albicans* surface after interaction with human serum, a solution of 10^6 cells/ml on Lee medium at pH 6.7 with 10% of human serum (NS or HIS) was prepared. 10^5 cells of this suspension were laid during 30 min, 1.5 or 5 h at $37°C$ on glass coverslips pre-coated with poly-L-lysine (1 mg/ml). Cells were fixed with formaldehyde 4% (w/v) (in PBS) for 20 min at room temperature (RT). Glass slides were washed twice with PBS. To stain cell membrane, PKH26 dye was added following technical instructions (Sigma Aldrich). Then, slides were blocked during 45 min at RT with buffer B (PBS plus bovine serum albumin (BSA) at 1 mg/ml). The slides were washed twice again with PBS and then incubated for 1.5 h at RT in the same buffer with an anti-C3, anti-factor B polyclonal serum (dilution 1:75 and 1:50, respectively) or only buffer. The slides were washed three times with PBS and further incubation for 1 h with an anti-rabbit IgG conjugated with Alexa-488 diluted at 1:500 in buffer B was done. Nuclei were stained with DAPI dye (5 µg/ml; 5 min at RT). Mounting medium Fluoromount-G (SouthernBiotech) was added to the preparations. Cells were then examined with epifluorescence and confocal microscopy images were collected using an Olympus FV1200 microscope.

Flow Cytometry Analysis

C. albicans cells were grown as previously described on Immunofluorescence assay section. In this case the incubation with human serum (NS or HIS) was done during 30 min to avoid larger hyphae formation. After that, cells were washed with PBS and fixed with formaldehyde 4% in PBS during 20 min at RT. Then, three washes with PBS were performed and samples were blocked with 0.1% BSA in PBS at $4°C$ overnight with gentle shaking. Samples were incubated with primary antibody or with buffer alone as negative control, for 2 h at RT. The primary antibodies were prepared in PBS with 0.1% BSA, anti-C3 and anti-FB at 1:75 and 1:50 dilutions, respectively. Four PBS washes were performed and incubated with secondary antibody below, an anti-rabbit conjugated with Alexa-488 diluted at 1:500, 2 h at RT with gently shaking. Then, samples were washed with PBS 4 times more. These cells were analyzed with a Guava easyCyte cytometer of Millipore.

RESULTS

Gel-free Proteomic Approach to Study Protein Interactions between *C. albicans* and Human Serum

In order to break through the broad spectrum of human serum proteins attached to *C. albicans* cells and the fungus surface proteins, we analyzed *C. albicans* cell surface after 5 h of incubation with 10% human serum by a gel-free proteomic strategy. Normal serum (NS) was used to mimic physiological conditions of *C. albicans* during invasive candidiasis and heat

inactivated serum (HIS) to determine human proteins deposited on the *C. albicans* surface without conventional complement cascade activation. As human serum is the most important inductor of *C. albicans* dimorphic transition, after 5 h of incubation all *C. albicans* cells were in the hyphal form. To obtain the protein peptides, a shaving methodology was used, which consists of direct cell digestion with trypsin and MS/MS identification. Before digestion, cells were washed with PBS plus Tween-20 and several additional washes with PBS were done to remove human serum proteins weakly attached to *C. albicans* cells. We optimized parameters such as time and trypsin concentration during shaving of *C. albicans* cells, in order to control cell integrity and identify the highest number of proteins (from both, *C. albicans* and human). With short times of trypsin treatment, between 5 and 10 min, only human proteins were identified (data not shown). Otherwise, after 30 min, human and *C. albicans* proteins were detected; thus, this duration of trypsin incubation was selected. Before and after trypsin treatment, cells were treated with propidium iodide (PI) and the cell permeability state was analyzed under the microscope. Less than 1% positive staining was obtained in all conditions. Cell integrity was assessed with the purpose of avoiding cytoplasmic protein contamination after trypsin treatment.

Peptides obtained after *C. albicans* cells interaction with 10% human serum (NS or HIS) by shaving were recovered and identified by LC-MS/MS on a LTQ Orbitrap. Four biological replicates were made using NS and three using HIS. Mass spectra raw files were searched against *C. albicans* SC5314 and human databases (CGD and Uniprot-SwissProt, respectively) and proteins identified in at least two biological replicates with two or more peptides in one of them were selected. *C. albicans*-identified proteins are listed in **Table 1** and Table S1, and human proteins in **Table 2** and Table S2.

This analysis rendered the identification of 372 unique *C. albicans* proteins, 371 in NS and 134 in HIS (Table S1). There were 133 proteins identified in both conditions, accounting for 36% of the total *C. albicans*-identified proteins. The Orf19.8 (an ortholog of *C. dubliniensis* CD36*)* was the only protein identified with HIS and not with NS. Noteworthy, around ∼33% of the *C. albicans* proteins were identified in at least six of the seven analyzed samples. Furthermore, of the 372 *C. albicans*-identified proteins, 60 proteins presented a signal peptide (SP) at CGD. On the other hand, 214 human proteins were identified; of these, 128 were common to both conditions (60%), 43 were present only after incubation with NS and the other 43 only with HIS (**Table 2** and Table S2).

Analysis of *C. albicans* Cell Surface Protein Pattern after Interaction with 10% human Serum

Only those *C. albicans* identified proteins with a Gene Ontology (GO) annotation at CGD pointing either to surface localization, either to surface exposure or with the ability to induce human serum antibodies, were included in **Table 1**. We focused our study on these 155 *C. albicans* proteins classified into 5 categories: glycosylphosphatidylinositol (GPI)-anchored proteins; proteins

TABLE 1 | Selected *C. albicans* proteins identified after incubation with 10% NS or HIS related with cell surface and immunogenicity.

Categories	Common in NS and HIS	Only identified in NS	Only in HIS	Total
GPI-anchored proteins[a]	11	12	–	23
	Als3, Cht2, Ecm33, **Hyr1**, Ihd1, Pga45, Phr1, Pir1, Rbt1, **Rbt5**, Sod5	**Als1**, Crh11, Pga10, **Pga4**, Phr2, Plb3, Plb4.5, Rhd3, Sap9, Ssr1, **Utr2**, Ywp1	–	
Proteins involved in cell wall organization or biogenesis[b](GO: 71554)	8	18	–	26
	Act1, Bmh1, Gda1, Msb2, **Sod1,** Srb1, **Tsa1, Yps7**	Agm1, **Bgl2**, Eng1, Frs1, Gfa1, Gsc1, Kex2, Mnn26, Mnt1, Pmi1, Pmt2, Ras1, Rho1, Rvs161, Sim1, **Slk19,** Sur7, Ypt31	–	
Cell surface proteins[b](GO: 9986)	40	21	1	62
	Adh1, Cdc19, Cdc48, **Cef3,** Ddr48, **Ece1,** Efb1, **Eft2,** Egd2, **Eno1, Fba1, Glx3, Hsp70, Hsp90, Ino1, Ipp1,** Kar2, **Met6,** Mp65, Orf19.2478.1, **Pdc11, Pdi1, Pgk1,** Pma1, Pra1, Rpl10, Rpl17B, Rpl3, Rpl4B, **Rps7A,** Rps8A, **Sam2, Ssa2, Ssb1, Tal1, Tdh3,** Tef1, **Tkl1, Tos1, Tpi1**	Ape2, **Atp1, Atp2,** Coi1, Csp37, Gpd2, Gph1, **Hem13,** Hsp104, Hsp21, Rpl13, Rpl14, Rpl19A, Rpl20B, Rpl6, Rps10, Rps6A, **Ssc1,** Ssz1, Ugp1, Xyl2	Cht3	
Plasma membrane proteins[b] (GO: 5886)	8	28	–	36
	Ahp1, Orf19.6553, Cof1, Met15, Pfy1, **Pgi1,** Pil1, Tif	**Ade17,** Adh2, Bcy1, Bzz1, Cdc42, Chc1, Faa4, **Gca1,** Hgt6, Ist2, Kin2, Lsp1, Mdg1, Orf19.1564, Orf19.3003, Orf19.4216, Orf19.6160, Pet9, Pga63, Pst3, Rdi1, Sec26, Sec4, **Shm2,** Sso2, Tom70, Ypt1, Ykt6	–	
Others immunogenic proteins	6	2	–	8
	Asc1, Gnd1, Grp2, Hxk2, Msi3, Sah1	**Acs2, Mdh1**	–	
				155

The classification is hierarchical and exclusive. Proteins were included if they were detected in at least two replicates of one condition with at least 2 peptides in one replicate. There are four replicates of Normal Serum (NS) and three replicates of Heat Inactivated Serum (HIS). Protein names from Candida Genome Database (CGD) (Inglis et al., 2012).
[a]GPI-cell wall proteins reviewed in De Groot et al. (2003), Pardo et al. (2004), Chaffin (2008), Klis et al. (2009).
[b]Gene Ontology (GO) terms (in parentheses the GO identifier) at CGD.
Proteins able to induce human serum antibodies in patients with invasive candidiasis are indicated in bold (Pitarch et al., 2004, 2006, 2008, 2011; Mochon et al., 2010).

involved in cell wall organization or biogenesis; cell surface proteins; plasma membrane proteins and other immunogenic proteins not present in the previous group. The groups are hierarchical and exclusive.

The first category includes 23 proteins attached to the plasma membrane or cell wall by a GPI-anchor; most are involved in biogenesis and maintenance of the *C. albicans* cell wall. All GPI-anchored proteins identified have a detected SP for entry into the secretory pathway at CGD. In the second category, there are 26 other proteins involved in cell wall organization or biogenesis. Among the 62 proteins included in the cell surface category, there are proteins located at the cell surface, although most of them do not have SP and could not reach this cell localization via the secretory pathway. Several of them have been described as "moonlighting" or multifunctional proteins. It is interesting that antibodies against a total of 52 *C. albicans*-identified proteins have been detected in human blood or serum from patients with invasive candidiasis (these immunogenic proteins are highlighted in **Table 1**). Many of these are abundant at the cell surface (as they were detected with more than 10 unique peptides) and were not secreted by the classical secretory pathway (not have SP), such as Eno1, Fba1, Hsp70, and 17 additional proteins.

It is also remarkable that two out of the four proteins described in *C. albicans* with the ability to fix the complement inhibitor factor H (FH) were identified. Pra1 was identified in both samples (NS and HIS), but Gpd2 was only in NS samples. Furthermore, we detected many *C. albicans* proteins described with the ability to bind plasminogen: Adh1, Eno1, Fba1, Gpd2, Pgk1, Pra1, Tdh3, Tef1, and Tsa1.

Analysis of Human Serum Proteins that Interact with *C. albicans* Cell Surface Relevant to *C. albicans*-host Interaction

The 214 identified human proteins were classified into the following classes according to their function: complement pathway, coagulation pathway, metabolism, immunoglobulins, cytoskeleton, and others. Proteins belonging to complement and coagulation cascades are listed in **Table 2** and the rest of the proteins are in Table S2. The percentages of proteins identified in each category in the two conditions tested are shown in **Figure 1**. The complement and coagulation pathways represent higher percentages of the total identified proteins when HIS was used and, on the other hand, the immunoglobulin

TABLE 2 | Human proteins of complement and coagulation pathways identified on the surface of *C. albicans* after incubation with human serum.

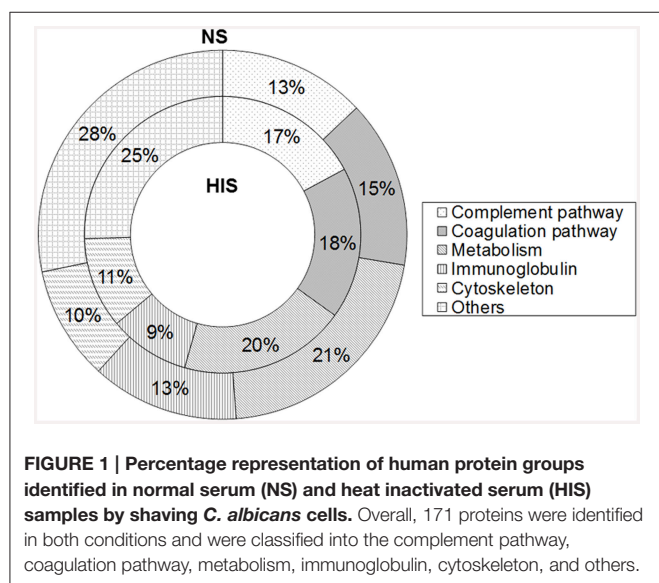| Accession number[a]	Gene symbol[a]	Description/Function[a]	No. replicates (peptides on each replicate)[b]	
			NS	HIS
COMPLEMENT PATHWAY				
P01023	A2M	Alpha-2-macroglobulin	2 (15,11)	3 (10,9,70)
P04003	C4BPA	C4b-binding protein alpha chain	4 (5,3,4,9)	3 (15,8,4)
P20851	C4BPB	C4b-binding protein beta chain	1 (2)	2 (2,3)
P10909	Clusterin	Extracellular chaperone or ApoJ	4 (11,8,13,19)	3 (11,13,18)
Q9BWP8	COLEC11	Collectin kidney protein 1 or collectin-11	4 (5,5,2,4)	0
P02745	C1qA	Complement C1q subcomponent subunit A	0	2 (1,4)
P02746	C1qB	Complement C1q subcomponent subunit B	3 (1,1,1)	3 (5,5,4)
P02747	C1qC	Complement C1q subcomponent subunit C	3 (1,1,2)	3 (7,5,6)
P00736	C1r	Complement C1r subcomponent	2 (2,7)	3 (21,15,14)
Q9NZP8	C1r-like	Complement C1r-like protein	0	3 (6,5,5)
P09871	C1s	Complement C1s subcomponent	2 (1,10)	3 (23,14,11)
P06681	C2	Complement component C2	1 (2)	3 (19,15,10)
P01024	C3	Complement component C3	4 (103,106,141,147)	3 (141,105,138)
P0C0L4	C4-A	Complement component C4-A	4 (32,32,94,98)	3 (90,61,76)
P0C0L5	C4-B	Complement component C4-B	4 (32,33,95,98)	2 (90,62)
P01031	C5	Complement component C5	3 (34,23,21)	3 (68,46,70)
P13671	C6	Complement component C6	3 (20,18,4)	3 (21,14,16)
P10643	C7	Complement component C7	4 (13,9,8,6)	3 (6,11,16)
P07357	C8-A	Complement component C8 alpha chain	4 (13,11,6,7)	3 (11,8,8)
P07358	C8-B	Complement component C8 beta chain	4 (5,6,16,27)	3 (22,16,13)
P07360	C8-G	Complement component C8 gamma chain	4 (6,6,4,1)	3 (6,4,9)
P02748	C9	Complement component C9	4 (17,15,17,22)	3 (24,17,14)
P00751	FB	Complement factor B	4 (8,3,8,10)	3 (47,31,28)
P00746	FD	Complement factor D	0	3 (2,2,4)
P08603	FH	Complement factor H	2 (5,3)	3 (18,17,13)
P05156	FI	Complement factor I	1 (1)	3 (3,3,6)
P27918	FP	Complement factor P or Properdin	4 (5,5,2,2)	1 (3)
Q15485	Ficolin-2	Collagen/fibrinogen domain containing protein 2	0	2 (1,3)
O75636	Ficolin-3	Collagen/fibrinogen domain containing protein 3	1 (1)	3 (5,2,7)
O00187	MASP2	Mannan-binding lectin serine protease 2	3 (2,2,5)	1 (3)
P80108	GPLD1	Phosphatidylinositol-glycan-specific phospholipase D	0	2 (4,15)
P05155	C1INH	Plasma protease C1 inhibitor (Serpin G1)	0	3 (20,14,24)
P04004	VTN	Vitronectin or protein S	4 (10,12,10,11)	3 (14,11,13)
COAGULATION PATHWAY				
P01009	SERPINA1	Alpha-1-antitrypsin (AAT)	4 (4,3,6,12)	3 (37,35,47)
P08697	SERPINF2	Alpha-2-antiplasmin	4 (2,2,9,13)	3 (17,16,18)
P01008	SERPINC1	Antithrombin-III (AT-III)	4 (2,2,5,10)	3 (13,16,25)
P02749	ApoH	Beta-2-glycoprotein 1; APC inhibitor; Apolipoprotein H	2 (1,1)	3 (8,3,4)
Q96IY4	CPB2	Carboxypeptidase B2; Thrombin-activable fibrinolysis inhibitor	4 (2,1,12,10)	3 (5,3,3)
P00740	F9	Coagulation factor IX	4 (3,2,3,5)	3 (5,4,4)
P12259	F5	Coagulation factor V	4 (8,2,11,22)	3 (11,4,8)
P08709	F7	Coagulation factor VII	2 (2,3)	2 (3,2)
P00742	F10	Coagulation factor X	2 (2,2)	3 (6,6,6)
P00748	F12	Coagulation factor XII	1 (1)	2 (4,8)
P00488	F13A	Coagulation factor XIII A chain	3 (1,2,1)	2 (4,7)
P05160	F13B	Coagulation factor XIII B chain	0	3 (2,1,2)
P02671	FGA	Fibrinogen alpha chain	4 (13,12,3,7)	3 (6,6,10)
P02675	FGB	Fibrinogen beta chain	4 (13,10,1,7)	3 (6,11,14)

(Continued)

TABLE 2 | Continued

Accession number[a]	Gene symbol[a]	Description/Function[a]	No. replicates (peptides on each replicate)[b]	
			NS	HIS
P02679	FGG	Fibrinogen gamma chain	3 (6,6,4)	3 (2,5,10)
P02751	FN1	Fibronectin; Binds cell surfaces and various compounds (collagen, fibrin, heparin…)	4 (60,62,15,16)	3 (29,38,24)
P68871	HBB	Hemoglobin subunit beta	4 (7,9,3,10)	3 (9,7,11)
Q9Y251	HPA	Heparanase	2 (1,3)	0
P05546	HCF2	Heparin cofactor 2 (Serpin D1)	4 (8,11,34,28)	3 (29,23,27)
P04196	HRG	Histidine-rich glycoprotein	4 (22,23,26,23)	3 (21,24,17)
Q14520	HABP2	Hyaluronan-binding protein 2	0	3 (4,3,5)
P01042	KNG1	Kininogen-1	4 (6,2,11,15)	3 (11,14,11)
P03952	KLKB1	Plasma kallikrein	4 (7,8,2,1)	3 (8,9,10)
P05154	SERPINA5	Plasma serine protease inhibitor	0	2 (3,6)
P00747	PLG	Plasminogen	4 (5,6,21,34)	3 (41,41,15)
P40197	GP5	Platelet glycoprotein V	2 (2,2)	3 (5,2,3)
Q9UK55	SERPINA10	Protein Z-dependent protease inhibitor	4 (1,1,10,9)	3 (6,5,7)
P00734	F2	Prothrombin	4 (16,18,17,25)	3 (21,17,10)
P04070	PROC	Vitamin K-dependent protein C	1 (3)	2 (3,4)
P07225	PROS1	Vitamin K-dependent protein S	3 (1,1,11)	3 (6,3,14)
P04275	VWF	von Willebrand factor	4 (18,16,19,72)	3 (52,36,9)

[a] *Accession number, gene symbol and function from UniProtKB/Swiss-Prot database (Uniprot, 2014).*
[b] *Proteins were included if they were identified in at least two replicates of one condition with at least 2 peptides in one replicate.*

FIGURE 1 | Percentage representation of human protein groups identified in normal serum (NS) and heat inactivated serum (HIS) samples by shaving *C. albicans* cells. Overall, 171 proteins were identified in both conditions and were classified into the complement pathway, coagulation pathway, metabolism, immunoglobulin, cytoskeleton, and others.

group represents a higher percentage when NS was used.

Close to all components of the three complement pathways (33 proteins) were identified, most of them in the two conditions tested. However, there are some complement proteins that have been detected only in the NS sample (Collectin-11) or in the HIS sample [complement C1q subcomponent subunit A (C1qA), complement C1r-like protein, complement factor D (FD), ficolin-2, phosphatidylinositol-glycan-specific phospholipase D (GPLD1) and plasma protease C1 inhibitor (SERPING1 or C1INH)] (**Table 3**). The identification of many complement inhibitors such as alpha-2-macroglobulin (A2M), C1INH, FH, and factor I (FI) is also interesting. In relation to the coagulation pathway, 31 proteins were identified, with most of them also being detected in the two conditions. Although heparanase was only identified in NS sample and coagulation factor XIII B chain, hyaluronan-binding protein 2 and plasma serine protease inhibitor (SERPINA5) were only identified in the HIS one. Interestingly, antithrombin-III (SERPINC1) and plasminogen (PLG), two coagulation proteins that are relevant complement inhibitors, were identified. Focusing on human immunoglobulins IgG and IgM, they were identified in NS and HIS samples and with a similar number of peptides in both of them (**Table S2**). The category of metabolic proteins includes many apolipoproteins detected on the surface of *C. albicans* (**Table S2**).

Interestingly, 13 proteins belonging to the serpin family were identified. Serpins are a relevant group of proteins with similar structures able to inhibit proteases. Some of them had only been observed in the HIS sample, such as SERPINA5, C1INH and cortisol-binding globulin (SERPINA6) (**Table 4**).

To estimate the relative abundance of human proteins identified on *C. albicans* surface, the normalized spectral abundance factor (NSAF) (Zybailov et al., 2007) of each protein was calculated and the average NSAF values among samples were assessed. We compared these NSAF values with the ranking of the original protein abundance in human plasma using data of the dynamic range of most abundant proteins adapted from Mitchell (2010). As shown in Figure S1, the relative abundance

TABLE 3 | Comparison of complement proteins identified on the cell surface of *C. albicans* by shaving after incubation with NS vs. HIS.

Complement proteins	Normal serum (NS)	Heat inactivated serum (HIS)
Lectin pathway	7 proteins **Collectin-11**, MASP2, Ficolin-3*, C2*, C3, C4-A, C4-B	7 proteins MASP2*, **Ficolin-2**, Ficolin-3, C2*, C3, C4-A, C4-B
Classical pathway	8 proteins C1qB, C1qC, C1r, C1s, C2*, C3, C4-A, C4-B	10 proteins **C1qA**, C1qB, C1qC, C1r, **C1r-like**, C1s, C2, C3, C4-A, C4-B
Alternative pathway	3 proteins C3, FB, FP	4 proteins C3, FB, **FD**, FP*
Membrane attack complex (MAC)	7 proteins C5, C6, C7, C8-A, C8-B, C8-G, C9	7 proteins C5, C6, C7, C8-A, C8-B, C8-G, C9
Complement inhibitors	7 proteins A2M, C4BPA, C4BPB*, clusterin, FH, FI*, vitronectin	9 proteins A2M, **C1INH**, C4BPA, C4BPB, clusterin, FH, FI, **GPLD1**, vitronectin

Proteins indicated in bold are only identified in one condition.
Proteins with *are identified only in one replicate of the indicated condition.

TABLE 4 | Identified proteins belonging to the SERPIN family.

Accesion number	Gene name	Protein	Substrate	No. replicates (peptides on each replicate)	
				NS	HIS
P01009	SERPINA1	Alpha 1-antitrypsin	Human neutrophil elastase	4 (4,3,6,12)	3 (37,35,47)
P01011	SERPINA3	Alpha 1-antichymotrypsin	Cathepsin G	2 (1,2)	3 (27,19,30)
P29622	SERPINA4	Kallistatin	Kallikrein	2 (5,10)	3 (13,19,16)
P05154	SERPINA5	Plasma serine protease inhibitor	Protein C	0	2 (3,6)
P08185	SERPINA6	Cortisol-binding globulin	Non-inhibitory; cortisol binding	0	3 (11,8,12)
P05543	SERPINA7	Thyroxine-binding globulin	Non-inhibitory; thyroxine binding	2 (2,1)	3 (6,4,14)
P01019	SERPINA8	Angiotensionogen	Non-inhibitory; Precursor of angiotensin I	4 (6,5,5,5)	3 (11,13,19)
Q9UK55	SERPINA10	Protein Z-related protease inhibitor	Factor Z and XI	4 (1,1,10,9)	3 (6,5,7)
P01008	SERPINC1	Antithrombin-III	Thrombin and factors IXa, Xa, and XIa	4 (2,2,5,10)	3 (13,16,25)
P05546	SERPIND1	Heparin cofactor II	Thrombin	4 (8,11,34,28)	3 (29,23,27)
P36955	SERPINF1	Pigment epithelium derived factor	Non-inhibitory; Potent anti-angiogenic molecule	1 (1)	3 (11,10,19)
P08697	SERPINF2	Alpha 2-antiplasmin	Plasmin and trypsin	4 (2,2,9,13)	3 (17,16,18)
P05155	SERPING1	Plasma protease C1 inhibitor	C1r, C1s, F12a, kallikrein, F11a, plasmin, MASP1, MASP2	0	3 (20,14,24)

of proteins in the plasma is different to the relative abundance at the hypha surface after interaction with human serum. For example, C3 protein of complement pathways was identified as one of the most abundant proteins on the surface of *C. albicans* in both conditions and it was not among the most abundant proteins in plasma. Also, in the apolipoprotein family, the relative abundance of some proteins is different in the plasma vs. the fungal surface. ApoB100 and ApoE are less abundant in human plasma than ApoCIII and both proteins showed higher NSAF than ApoCIII.

The abundance ranking of the proteins involved in the complement and coagulation pathways identified on *C. albicans* surface is shown in **Figure 2**. C3, C4-B, and C4-A had the highest NSAF results among complement proteins in both serum conditions. Complement components of Membrane Attack

Complex (MAC: C6, C7, C8-A, C8-B, C8-G, and C9) had higher NSAF in NS than in HIS samples and only C5 had higher NSAF in HIS than in NS samples. In the coagulation pathway, it is notable that SERPINA1 (alpha-1-antitrypsin) is much more abundant in the HIS sample than other proteins such as HRG (histidine-rich glycoprotein), HCF2 (SERPIND1) or F2 (prothrombin), which are more abundant in the NS sample than SERPINA1.

During complement activation, C3, C4, and C5 are proteolyzed by cleavage at a specific site and small fragments are produced; these fragments are called anaphylotoxins or complement peptides, and are indicated with the letter "a" (Walport, 2001a,b). The analysis of the identified peptides belonging to: (i) anaphylotoxin regions (C3a, C4a, and C5a), (ii) released domains during proteolysis activation of complement

Complement pathway	NS	HIS	Coagulation pathway	NS	HIS
C3	0.055	0.032	HRG	0.086	0.016
C4-B	0.023	0.013	HCF2	0.029	0.012
C4-A	0.023	0.011	F2	0.012	0.005
Clusterin	0.010	0.005	HBB	0.009	0.009
C9	0.009	0.006	FN1	0.005	0.002
VTN	0.007	0.005	SERPINF2	0.004	0.008
C8-B	0.005	0.004	CPB2	0.004	0.001
C8-G	0.004	0.003	PLG	0.003	0.005
C5	0.004	0.005	SERPINA1	0.003	0.120
Collectin-11	0.004	-	SERPINA10	0.003	0.001
C8-A	0.003	0.002	FGB	0.003	0.002
C6	0.003	0.002	KNG1	0.003	0.003
C7	0.002	0.001	VWF	0.002	0.002
FP	0.002	0.001	SERPINC1	0.002	0.006
C4BPA	0.002	0.002	FGA	0.002	0.001
C4BPB	0.002	0.001	FGG	0.002	0.002
FB	0.002	0.010	KLKB1	0.001	0.001
C1s	0.002	0.003	F12	0.001	0.001
A2M	0.002	0.004	F10	0.001	0.001
C1r	0.001	0.004	F9	0.001	0.001
C1qC	0.001	0.003	PROS1	0.001	0.001
MAPS2	0.001	0.0004	F5	0.001	0.0003
FH	0.001	0.001	PROC	0.001	0.001
C1qB	0.001	0.002	F7	0.001	0.001
C2	0.0003	0.003	APOH	0.001	0.002
FI	0.0002	0.001	GP5	0.001	0.0005
Ficolin-3	0.0001	0.002	HPA	0.001	-
C1INH	-	0.008	F13A	0.0002	0.001
C1qA	-	0.002	HABP2	-	0.001
C1r-like	-	0.001	SERPINA5	-	0.001
GPLD1	-	0.001	F13B	-	0.0002
FD	-	0.001			
Ficolin-2	-	0.0004			

FIGURE 2 | Protein ranking of complement and coagulation proteins identified on normal serum (NS) and heat inactivated serum (HIS) samples according to the average of NSAF. *C. albicans* was incubated with 10% human serum at 37°C (normal serum-NS or heat inactivated serum-HIS) and NSAF was calculated for proteins belonging to complement and coagulation proteins. The proteins were ordered according to their NSAF obtained in NS samples, from more to less abundant. Gradient color ranges from red (more abundant) to green (less abundant).

pathways (C2b and FBa), and (iii) of the MAC-interaction regions (MACPF) is shown in **Table 5**. The tendency was the identification of more peptides belonging to these regions in HIS samples than in NS samples. It is remarkable that peptides from C2b and C5a were not identified in NS samples.

Validation of Proteomic Results

To validate the proteomic results, firstly the deposition level of C3 and FB on the *C. albicans* surface was analyzed by immunofluorescence microscopy and flow cytometry assays. C3 is a common component of the three complement cascades and together with A2M, are the two highest abundant complement proteins in human serum. Factor B (FB) is an abundant complement protein specific of the alternative pathway. *C. albicans* cells were incubated with NS and HIS and treated with anti-C3 and anti-FB antibodies (**Figure 3**). We observed the deposition of C3 and FB along the surface of all *C. albicans* cells when incubated with human serum (NS or HIS) and some areas with higher intensity in the case of C3 after 5 h of incubation with NS (**Figure 3A**, left panels). These differences

TABLE 5 | Analysis of peptides corresponding to cleavage fragments produced during complement activation and to contact regions in MAC formation.

Region	No. peptides			
	Normal Serum-NS		Heat Inactivated Serum-HIS	
	Total	Only in NS	Total	Only in HIS
C2b	0	0	4	4
C3a	8	2	10	3
C4a	5	1	4	1
C5a	0	0	5	5
Ba	1	0	5	4
C6-MACPF	13	0	19	6
C7-MACPF	4	0	15	12
C8-A-MACPF	4	0	8	4
C8-B-MACPF	19	1	19	0

Identified peptides belong to the indicated released region or remain part of the full-length protein.
Gray background highlights identification in HIS condition.

between the fluorescence intensity of FB and C3 can be correlated with the number of peptides of these proteins in the samples at 5 h. When cells were incubated with HIS, the intensity of FB was slightly higher than in the case of NS incubation (it is well observed in short incubation times) (**Figure 3A**, right panels). To achieve flow cytometry assays, the incubation time of *C. albicans* cells with human serum was decreased to avoid the formation of large filaments. Within 30 min of incubation, the deposition of C3 was higher among the population in NS samples than in HIS (Figure S3). The intensity of FB was lower than that of C3, as expected, and less homogenous in both samples (NS and HIS) (upper panels). Mean fluorescence intensity (MFI) of two independent experiments were collected and showed that, in all cases, the MFI for C3 and FB was higher when cells were incubated with NS than with HIS at short times, although the differences are not significant in the case of FB (bottom panel).

All of these data together showed that C3 and FB deposition was clearly observed at short incubation times, and C3 enhanced its accumulation on the surface with longer incubation time and more dramatically when the incubation was with NS. However, FB accumulation on *C. albicans* surface did not appear to increase with time, maybe because it can be masked by the rest of the complement proteins (**Figure 3A**).

Furthermore, to address whether C3 and FB are close to the cytoplasmic membrane, fluorescence co-localization was evaluated with confocal microscopy images by the analysis of transversal sections across cells (**Figure 3B**). As observed, both C3 and FB proteins are localized around the plasmatic membrane, staining when *C. albicans* cells were incubated with NS and also with HIS.

Afterwards, in order to evaluate the deposition of human IgGs on *C. albicans* surface after 5 h of incubation with 10% human serum, IgGs were detected in NS and HIS samples, with higher intensity on the yeast surface than on the hyphal surface in both samples (Figure S2).
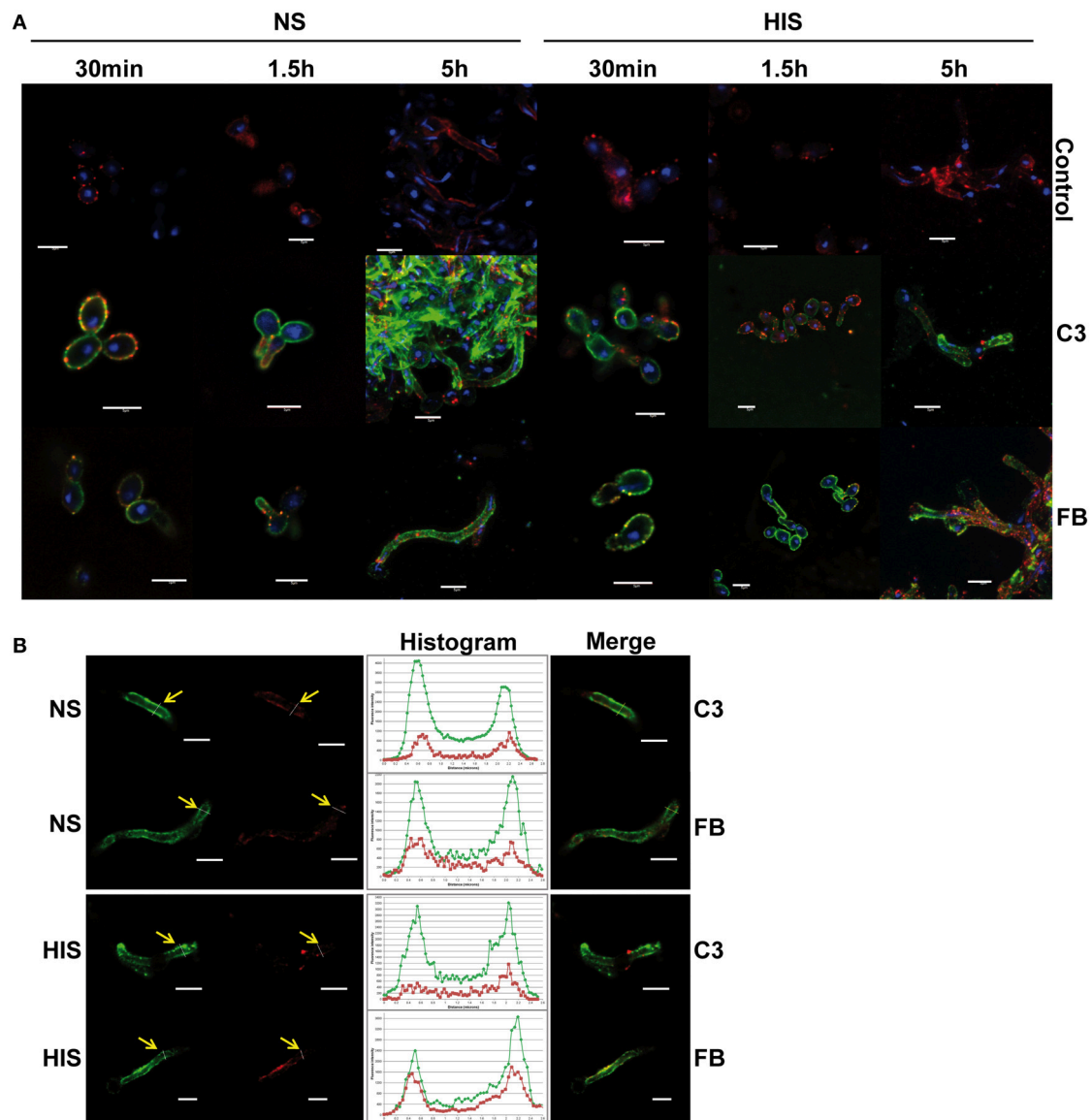
FIGURE 3 | Detection of C3 and factor B on *C. albicans* surface after incubation with human serum. (A) Immunofluorescence assay to detect C3 and factor B (FB) on *C. albicans* surface incubated 30 min, 1.5 and 5 h with normal serum (NS; left panels) or heat inactivated serum (HIS; right panels). Control images showed the background of secondary antibody (anti-rabbit-A488). **(B)** Co-localization of C3 and FB with *C. albicans* plasmatic membrane (staining with PKH26) was examined by the analysis of confocal microscopy images, after 5 h of incubation with NS and HIS. Representative images are shown. Yellow arrows indicate position of transversal sections of illustrative histograms. Co-localization analysis was performed using FiJi software. Red staining of plasmatic membrane was done with PKH26, blue staining of DNA with DAPI and green staining with anti-rabbit-A488. Line in bottom right corner indicated in all figures corresponds to a 5 μm.

DISCUSSION

Our work shows, for the first time to our knowledge, the identification by a shaving approach of 147 *C. albicans* surface proteins and 214 human serum proteins attached to the surface of the fungus in the same report. The more relevant *C. albicans* proteins identified are GPI-anchored proteins and proteins involved in cell wall organization or biogenesis. The most relevant human proteins identified belong to complement and coagulation pathways. Indeed, all main human complement

proteins were identified on the surface of *C. albicans*. The identified proteins were reproducible among the experiments and would represent the physiological mode of action of the immune system against *C. albicans* or components of the immune system used for *C. albicans* to evade, disturb or use on its own purpose the host defenses.

Analysis of *C. albicans* Surface Proteins

With the objective of unraveling the protein surface pattern of *C. albicans* hyphal cells, we analyzed our data in comparison to

the data obtained by our group using the same methodology of cell shaving over yeast and hyphal cells (growth also in Lee's medium, but without human serum) (Gil-Bona et al., 2015). This analysis rendered 304 proteins identified in all *C. albicans* assays (yeasts and hyphae with and without human serum) that we consider the *C. albicans* common core of cell surface proteins and not morphotype specific (Table S3). Also, a hyphal protein pattern of 55 proteins identified on *C. albicans* hypha forms, but not identified in *C. albicans* yeasts, was obtained. Among them, Als3, Hyr1, and Sod5 were associated with hyphae in different works (Heilmann et al., 2011; Sosinska et al., 2011; Sudbery, 2011). They are proteins that are expected to be detected, indicating that the other 52 proteins are at least more abundant or accessible in hypha than in yeast cells. Another example is Ece1, which is a hyphal-specific protein with unknown function (Fan et al., 2013) (Table S3).

Other proteins were associated with hyphae in some works and described as morphotype-independent in others, for example, Phr1 and Rbt1 (Heilmann et al., 2011; Ragni et al., 2011; Monniot et al., 2013). We identified both proteins in all shaving *C. albicans* experiments. Furthermore, Phr2 (a homolog of Phr1) was also identified in all conditions. The differences observed at the level of detection of Phr1 and Phr2 among the samples were possibly related to the optimum pH for expression of each one (Sosinska et al., 2011; Dühring et al., 2015) (Table S1). On the other hand, Rhd3 and Ywp1 were reported as yeast-specific and were found to be part of the common core pattern in our study (Heilmann et al., 2011). A relevant protein found in *C. albicans* common core pattern is Msb2; this stabilizes the fungal cell wall and inactivates human antimicrobial peptides (Dühring et al., 2015).

It is interesting to know that GPI-anchored proteins identified in this work are part of the common core of *C. albicans* surface proteins or of the hyphal-form specific ones. There are 17 proteins belonging to the common core in our analysis, and also Ece1 and Sod5, which are part of the hyphal-specific group, whose genes were up-regulated under blood growth (Fradin et al., 2003).

Remarkably, 12 proteins were only identified on the *C. albicans* hyphal surface when it was grown with 10% human serum and we highlighted Pra1, Sap5, Tef1, and 5 more Orfs with unknown function (Table S3). In this work, two members of the Sap family, Sap9 and Sap5, were identified. Sap9 is part of *C. albicans* common core while Sap5 shows up in hyphae induced with the human serum group. Both genes were up-regulated in *C. albicans* biofilm associated with bloodstream infections (Joo et al., 2013). Ten out of these twelve proteins are identified for the first time on the surface of *C. albicans*; Pra1 and Tef1 have been previously identified in the cell wall of *C. albicans*.

Many proteins described as "moonlighting" or multifunctional proteins in *C. albicans* were identified in this work (Nombela et al., 2006; Chaffin, 2008) (indicated with * in Table S1). Among them, Gpd2 has been described as a virulence factor on microorganism surface because it mediates complement evasion; meanwhile, there is a glycerol 3-phosphate-dehydrogenase in the cytoplasm that is involved in the intracellular accumulation of glycerol to control osmotic pressure (Luo et al., 2013). Tsa1 is involved in oxidative stress response in the cytoplasm and in hyphal cell wall biogenesis (Urban et al., 2005). Tdh3 is a glycolytic enzyme (glyceraldehyde-3-phosphate dehydrogenase) in the cytoplasm and able to attach fibronectin and laminin at the cell surface (Gozalbo et al., 1998).

Regarding *C. albicans* proteins that interact with complement inhibitor proteins (FH and PLG), the expression of Gpd2 and Pra1 is more prominent at the hyphal tip (Zipfel et al., 2011; Luo et al., 2013). Pra1 is also able to bind C4BP and it is also known that the hyphal tip is a prominent binding site for C4BP (Meri et al., 2004; Luo et al., 2011). For these reasons and because it facilitates host tissue penetration, the hyphal tip is considered an important factor of pathogenesis. We identified these two proteins on the *C. albicans* surface grown with 10% human serum that activate hyphal growth. Focusing on the mechanism to evade complement activation mediated by plasminogen coated surfaces, as plasminogen is an important complement cascade inhibitor in the fluid phase and on the human surface (Barthel et al., 2012), the identification of nine out of the eleven *C. albicans* proteins with the ability to bind PLG (Adh1, Eno1, Fba1, Gpd2, Pgk1, Pra1, Tdh3, Tef1, and Tsa1) is another highlight of this study (Crowe et al., 2003; Jong et al., 2003; Barthel et al., 2012).

Human Serum Proteins with the Ability to Interact with *C. albicans* Surface

We detected human plasma proteins from different abundance strata on *C. albicans* surface; most of them are from the most abundant proteins or from intermediate stratum (Anderson and Anderson, 2002; Mitchell, 2010). Due to the heat treatment of HIS serum, FB and C2 become non-functional; in this situation, other human serum proteins could interact further with the surface of *C. albicans*. Remarkably, *C. albicans* surface could be less accessible to trypsin digestion when it was grown with 10% HIS as the number of coated of human proteins is greater. This could explain the observation of higher sum of peptide-spectrum match (PSMs) of human serum proteins in the HIS condition in comparison with the NS condition. In any case, it is not possible to discard that some proteins differentially observed in HIS are a result of changes in their binding properties due to protein denaturation.

Albumin is the most abundant protein in human plasma with close to two orders of magnitude more than the rest of proteins (Mitchell, 2010), but it was not detected as the highest abundant protein on *C. albicans* surface incubated with human serum (either NS or HIS) (Figure S1). Also, there are some proteins only detected on *C. albicans* surface in HIS samples that have intermediate abundance in human plasma such as C1INH, C1qA, and RBP4. Serotransferrin and fibrinogen (FGG, FGA and FGB) were more abundant in NS condition and in the other side apolipoprotein A-I and transthyretin in HIS condition. All of these data indicate that the interaction between human proteins and *C. albicans* surface did not respond solely to their protein abundance in human serum. The study of C3 and FB deposition on *C. albicans* surface after incubation with 10% human serum was carried out by immunofluorescence with specific antibodies. The clearly deposition of C3 and FB was observed at short times of interaction in **Figure 3A**.

A graphical summary of the identified proteins at *C. albicans* surface from the different complement pathways in NS and HIS conditions is presented in **Figure 4**. The enrichment on the surface of *C. albicans* in proteins belonging to the lectin, classical and alternative complement pathways is visibly observed in normal conditions (NS); almost all of the main proteins of complement pathways were identified (**Figures 4A,B**). Peptides

belonging to anaphylotoxin region are less frequently identified in NS than in HIS condition; this could be related to that C2b, C3a, FBa, and C5a fragments were released to the medium under this condition (**Table 5**). Furthermore, two peptides (CCEDGMR and CCEDGMRENPMR), which are part of the LRK26 peptide derived from C3a and with antimicrobial effects on *C. albicans* surface (Sonesson et al., 2007), were only identified in NS.
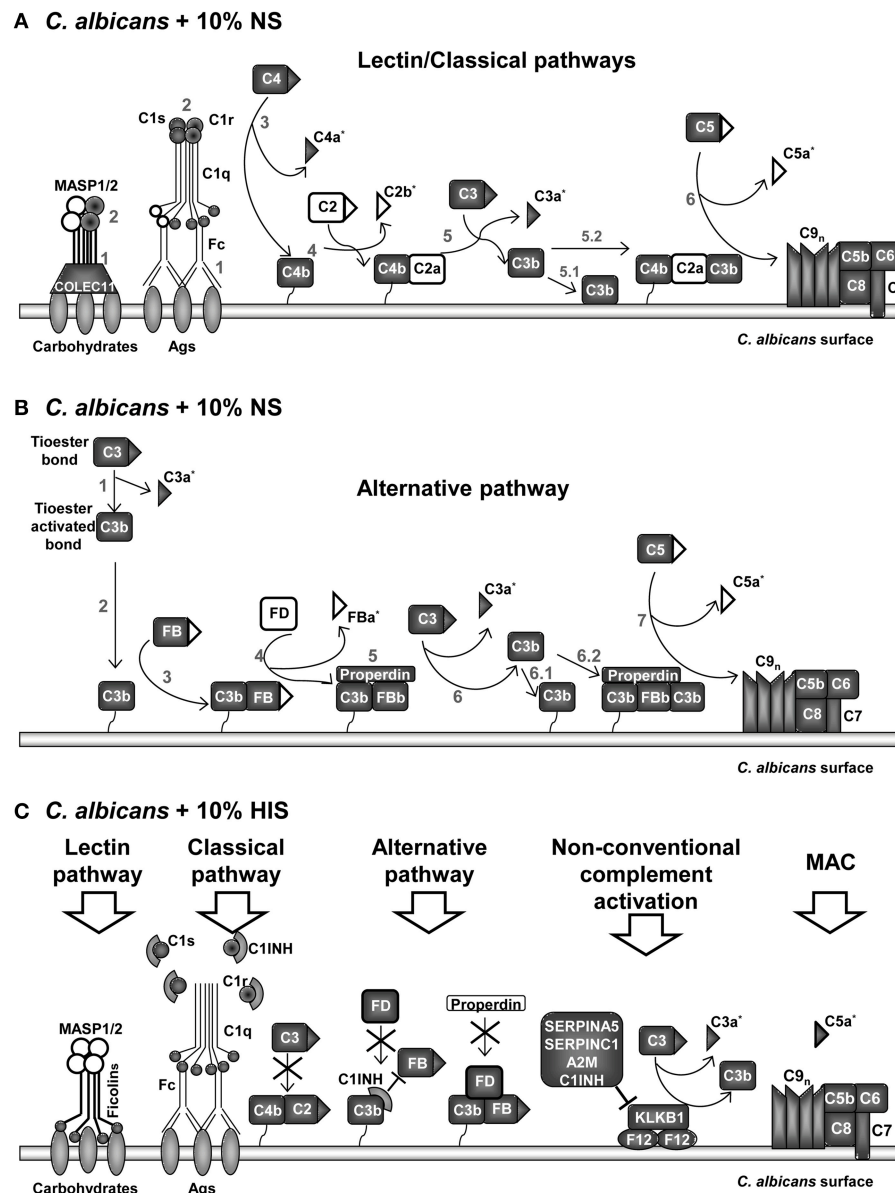
FIGURE 4 | Model of interactions among proteins of the complement pathways identified on *C. albicans* surface after incubation with human serum. Proteins identified in normal serum (NS) samples in **(A,B)**. Proteins belonging to lectin and classical pathways are represented in **(A)** and to alternative pathway in **(B)**. Proteins mainly identified on heat inactivated serum (HIS) in **(C)**. Identified proteins on samples are indicated with gray background and not-identified in white. Complement fragments released during complement activation are indicated with triangles. Peptides belonging to complement fragments that can be released or be part of a full-length protein are indicated with *. Numbers in **(A,B)** correspond to sequential steps during complement pathways activation. **(A)** Collectin-11 was only identified in NS samples. MASP2 was identified in NS and MASP1 was not identified in any sample. Fc (constant region of Ig) of IgGs or IgMs start classical complement pathway. **(B)** Properdin was identified in NS and Factor D (FD) was not. **(C)** Ficolin-2, Ficolin-3, C1qA, C1INH, and FD were identified only in HIS samples. C1INH dissociates the C1qrs complex, blocks alternative C3 convertase and inhibits C3 activation by kallikrein. C2 was identified in HIS, including 4 peptides of C2b. Five peptides belonging to FBa were identified in HIS condition. Signally, peptides belonging to C5a fragment were only identified in HIS samples.

Peptides belonging to the C5a region were only identified in HIS samples, just as more peptides belonging to the MACPF region of C6, C7, and C8-A. Furthermore, lower NSAF for all components of MAC, except C5, were obtained with HIS (**Figure 2**), which could be correlated with the enhanced identification of C5a peptides in this condition. In contrast, no identification of peptides belonging of C5a region and higher NSAF for the rest of components of MAC in NS samples may indicate MAC formation. MACPF-domains would not be accessible to trypsin digestion because they are points of interaction for MAC formation and would be protected in this condition. Furthermore, we identified clusterin and vitronectin (VTN) with higher NSAF in NS condition and both are able to interact with MAC in plasma, and thus prevent the cytolysis of host cells (Chauhan and Moore, 2006). Further experiments are required to determine the formation and the functionality of MAC on *C. albicans* surface.

In the same line of thought, in NS samples, properdin (FP) and not factor D (FD) were identified, while the opposite happened with HIS samples (**Figure 4B** vs. **Figure 4C**). The results are in agreement with a functional alternative complement pathway in NS condition on the surface of *C. albicans*. On the other hand, in HIS samples, FD could still be attached to the C3bB complex and we suppose that C3 convertase is not formed. In this situation, FP could not stabilize the complex and would be not identified on HIS samples. C1INH is able to bind to immobilized C3b and block progression of the complement cascade (Jiang et al., 2001). Serpin role of C1INH produces dissociation of C1qrs complex and the identification of these proteins in the HIS condition was enhanced, possibly for this reason. C3 and C4 (C4-A and C4-B) were identified with less NSAF and C2 and FB with higher NSAF in HIS samples (**Figure 4C**). In a current study about *Streptococcus pyogenes* interaction with human plasma, C2, C1qA, and FD were not detected on the surface of the microorganism (Sjöholm et al., 2014). In our experiments, when *C. albicans* cells were incubated with HIS, we detected not only these proteins, but also C1INH, and the enrichment of FB, C1qB, and C1qC. This is an interesting observation that may be due to the absence of functional convertases; FB and C2 might be more accessible to trypsin digestion. We assume that the unsuitable formation of complement convertases, or at least in a less extended manner, is responsible for the enhanced peptide identification of C2, FB, FD, C5, and the C1qrs complex in HIS samples.

Coagulation and complement pathways are highly connected, with many cross-talks (Kazatchkine and Jouvin, 1984; Amara et al., 2008, 2010; Peerschke et al., 2008; Diamond, 2013; Verschoor and Langer, 2013). Many complement inhibitors were identified on the surface of *C. albicans* and in general more abundant in HIS samples. The innate immune system strives to respond and activates complement cascades by non-conventional mechanisms (thrombin (F2), PLG, Factor XII (F12), kallikrein (KLKB1), ficolin-2 and -3) and bypass locking points in absence of functional FB and C2 (Nielsen et al., 1996; Rittirsch et al., 2008) (**Figures 2**, **4C**). In HIS samples, KLKB1 inhibitors were enriched, such as SERPINA5, SERPINC1, A2M, and C1INH (Gadjeva et al., 1998; Paréj et al., 2013).

A relevant group of proteins is serpin family (serine protease inhibitors), which are broadly distributed among eukaryote organisms. Serpins are able to inhibit proteases by a specific mechanism where covalent complexes are formed with target proteases (Law et al., 2006). Since serpin has to be cleaved to inhibit the target substrates, inhibition also consumes the serpin; serpins are therefore irreversible enzyme inhibitors. For these reasons, the identification of many serpins on *C. albicans* surface and more interestingly in HIS samples is an interesting observation because they may be inactivating their substrates attached to the surface of *C. albicans* or themselves coating *C. albicans* surface in the sense to avoid interactions between their substrates and *C. albicans* (**Table 4**).

Interestingly, the apolipoproteins classified in the metabolism group are involved in lipid transport and also play a relevant role in host defense as part of the innate immune system (Grunfeld and Feingold, 2008). Furthermore, *C. albicans* grows better in lipoprotein-depleted plasma due to two factors: greater availability of lipids and reduced neutralizing candidacidal factor (Vonk et al., 2004). ApoE has a relevant role in systemic candidiasis infection and showed higher NSAF in NS samples (Figure S1). ApoE-/- mice showed an increased mortality in comparison with wild type (85 vs. 52%) (Vonk et al., 2004), and also an impaired immune response in *Klebsiella pneumonia* infection (de Bont et al., 2000).

In summary, the mechanisms developed by *C. albicans* to evade, interfere or use for its benefit host defenses are very wide and part of its pathogenic evolution. The analysis of *C. albicans* surface after incubation with human serum to obtain a partial breakdown of the immunological crosstalk between *C. albicans* and human serum proteins is a relevant target of study. *C. albicans* mimic human immune evasion model using many surface proteins to fix plasminogen and complement inhibitors (FH and C4BP) to protect against complement activation. Furthermore, the identification of Pra1, Sap5, Tef1, and 9 more proteins only identified in *C. albicans* hyphal surface induced with 10% human serum and not in other published conditions with an equivalent approach is a relevant highlight of this analysis.

AUTHOR CONTRIBUTIONS

EM designed and performed the experiments, analysis of results, and writing of the manuscript. CP designed and performed the experiments. CH performed the experiments. MH performed the experiments and analysis of results. CN research support. LM designed the experiments, analysis of results and critically revised the manuscript. CG designed the experiments, analysis of results, and critically revised the manuscript.

ACKNOWLEDGMENTS

by the ISCIII and FEDER. CP was the recipient of a fellowship from the program Continuous training for teachers of the Pontificia Javeriana University (Colombia) in conjunction with COLCIENCIAS-LASPAU University of Harvard. The proteomic analyses were carried out in the Proteomic Units of Complutense University of Madrid and of Centro de Investigaciones Biológicas (CSIC), members of ProteoRed network. The Immunofluorescence Microscopy images were carried out in Centro Nacional de Microscopía Electrónica (ICTS)-UCM. The authors would like to thank Dr. Fernando Vivanco [Fundación Jimenez Diaz, Universidad Complutense,

Madrid (Spain)] and Dra. Pilar Sánchez-Corral Gómez [Unidad de Investigación de Inmunopatología, Hospital Universitario La Paz, Madrid (Spain)] for the gift of C3 and FB antibodies, respectively. These results are lined up with the Spanish Initiative on the Infectious Diseases-Human Proteome Project (B/D-HPP).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2015.01343

REFERENCES

Almirante, B., Rodríguez, D., Park, B. J., Cuenca-Estrella, M., Planes, A. M., Almela, M., et al. (2005). Epidemiology and predictors of mortality in cases of *Candida* bloodstream infection: results from population-based surveillance, Barcelona, Spain, from 2002 to 2003. *J. Clin. Microbiol.* 43, 1829–1835. doi: 10.1128/JCM.43.4.1829-1835.2005

Amara, U., Flierl, M. A., Rittirsch, D., Klos, A., Chen, H., Acker, B., et al. (2010). Molecular intercommunication between the complement and coagulation systems. *J. Immunol.* 185, 5628–5636. doi: 10.4049/jimmunol.0903678

Amara, U., Rittirsch, D., Flierl, M., Bruckner, U., Klos, A., Gebhard, F., et al. (2008). Interaction between the coagulation and complement system. *Adv. Exp. Med. Biol.* 632, 71–79. doi: 10.1007/978-0-387-78952-1_6

Anderson, N. L., and Anderson, N. G. (2002). The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell. Proteomics* 1, 845–867. doi: 10.1074/mcp.R200007-MCP200

Barthel, D., Schindler, S., and Zipfel, P. F. (2012). Plasminogen is a complement inhibitor. *J. Biol. Chem.* 287, 18831–18842. doi: 10.1074/jbc.M111.323287

Brown, G. D., Denning, D. W., Gow, N. A., Levitz, S. M., Netea, M. G., and White, T. C. (2012). Hidden killers: human fungal infections. *Sci. Transl. Med.* 4, 165rv113. doi: 10.1126/scitranslmed.3004404

Buu, L. M., and Chen, Y. C. (2014). Impact of glucose levels on expression of hypha-associated secreted aspartyl proteinases in *Candida albicans*. *J. Biomed. Sci.* 21:22. doi: 10.1186/1423-0127-21-22

Collette, J. R., and Lorenz, M. C. (2011). Mechanisms of immune evasion in fungal pathogens. *Curr. Opin. Microbiol.* 14, 668–675. doi: 10.1016/j.mib.2011.09.007

Crowe, J. D., Sievwright, I. K., Auld, G. C., Moore, N. R., Gow, N. A., and Booth, N. A. (2003). *Candida albicans* binds human plasminogen: identification of eight plasminogen-binding proteins. *Mol. Microbiol.* 47, 1637–1651. doi: 10.1046/j.1365-2958.2003.03390.x

Chaffin, W. L. (2008). *Candida albicans* cell wall proteins. *Microbiol. Mol. Biol. Rev.* 72, 495–544. doi: 10.1128/MMBR.00032-07

Chauhan, A. K., and Moore, T. L. (2006). Presence of plasma complement regulatory proteins clusterin (Apo J) and vitronectin (S40) on circulating immune complexes (CIC). *Clin. Exp. Immunol.* 145, 398–406. doi: 10.1111/j.1365-2249.2006.03135.x

de Bont, N., Netea, M. G., Demacker, P. N., Kullberg, B. J., van der Meer, J. W., and Stalenhoef, A. F. (2000). Apolipoprotein E-deficient mice have an impaired immune response to *Klebsiella pneumoniae*. *Eur. J. Clin. Invest.* 30, 818–822. doi: 10.1046/j.1365-2362.2000.00715.x

de Groot, P. W., Bader, O., de Boer, A. D., Weig, M., and Chauhan, N. (2013). Adhesins in human fungal pathogens: glue with plenty of stick. *Eukaryotic Cell* 12, 470–481. doi: 10.1128/EC.00364-12

De Groot, P. W., Hellingwerf, K. J., and Klis, F. M. (2003). Genome-wide identification of fungal GPI proteins. *Yeast* 20, 781–796. doi: 10.1002/yea.1007

Diamond, S. L. (2013). Systems biology of coagulation. *J. Thromb. Haemost* 11 (Suppl. 1), 224–232. doi: 10.1111/jth.12220

Dreisbach, A., van der Kooi-Pol, M. M., Otto, A., Gronau, K., Bonarius, H. P., Westra, H., et al. (2011). Surface shaving as a versatile tool to profile global interactions between human serum proteins and the *Staphylococcus aureus* cell surface. *Proteomics* 11, 2921–2930. doi: 10.1002/pmic.201100134

Dühring, S., Germerodt, S., Skerka, C., Zipfel, P. F., Dandekar, T., and Schuster, S. (2015). Host-pathogen interactions between the human innate immune system and *Candida albicans*—Understanding and modeling defense and evasion strategies. *Front. Microbiol.* 6:625. doi: 10.3389/fmicb.2015.00625

Ene, I. V., Brunke, S., Brown, A. J., and Hube, B. (2014). Metabolism in Fungal Pathogenesis. *Cold Spring Harb. Perspect. Med.* 4. doi: 10.1101/cshperspect.a019695

Engleberg, N. C., Dermody, T. S., and DiRita, V. (eds.). (2012). "Innate immunity" in *Schaechter's Mechanisms of Microbial Disease* (Philadelphia, PA: Lippincott Williams & Wilkins), 66–90.

Fan, Y., He, H., Dong, Y., and Pan, H. (2013). Hyphae-specific genes *HGC1*, *ALS3*, *HWP1*, and *ECE1* and relevant signaling pathways in *Candida albicans*. *Mycopathologia* 176, 329–335. doi: 10.1007/s11046-013-9684-6

Fradin, C., Kretschmar, M., Nichterlein, T., Gaillardin, C., d'enfert, C., and Hube, B. (2003). Stage-specific gene expression of *Candida albicans* in human blood. *Mol. Microbiol.* 47, 1523–1543. doi: 10.1046/j.1365-2958.2003.03396.x

Gadjeva, M., Dodds, A. W., Taniguchi-Sidle, A., Willis, A. C., Isenman, D. E., and Law, S. K. (1998). The covalent binding reaction of complement component C3. *J. Immunol.* 161, 985–990.

Gil-Bona, A., Parra-Giraldo, C. M., Hernáez, M. L., Reales-Calderon, J. A., Solis, N. V., Filler, S. G., et al. (2015). *Candida albicans* cell shaving uncovers new proteins involved in cell wall integrity, yeast to hypha transition, stress response and host-pathogen interaction. *J. Proteomics* 127(Pt B), 340–351. doi: 10.1016/j.jprot.2015.06.006

Gillum, A. M., Tsay, E. Y., and Kirsch, D. R. (1984). Isolation of the *Candida albicans* gene for orotidine-5'-phosphate decarboxylase by complementation of *S. cerevisiae ura3* and *E. coli pyrF* mutations. *Mol. Gen. Genet.* 198, 179–182. doi: 10.1007/BF00328721

Gozalbo, D., Gil-Navarro, I., Azorín, I., Renau-Piqueras, J., Martínez, J. P., and Gil, M. L. (1998). The cell wall-associated glyceraldehyde-3-phosphate dehydrogenase of *Candida albicans* is also a fibronectin and laminin binding protein. *Infect. Immun.* 66, 2052–2059.

Green, J. V., Orsborn, K. I., Zhang, M., Tan, Q. K., Greis, K. D., Porollo, A., et al. (2013). Heparin-binding motifs and biofilm formation by *Candida albicans*. *J. Infect. Dis.* 208, 1695–1704. doi: 10.1093/infdis/jit391

Grunfeld, C., and Feingold, K. R. (2008). HDL and innate immunity: a tale of two apolipoproteins. *J. Lipid Res.* 49, 1605–1606. doi: 10.1194/jlr.E800011-JLR200

Heilmann, C. J., Sorgo, A. G., Siliakus, A. R., Dekker, H. L., Brul, S., de Koster, C. G., et al. (2011). Hyphal induction in the human fungal pathogen *Candida albicans* reveals a characteristic wall protein profile. *Microbiology* 157, 2297–2307. doi: 10.1099/mic.0.049395-0

Hernáez, M. L., Ximénez-Embún, P., Martínez-Gomariz, M., Gutiérrez-Blázquez, M. D., Nombela, C., and Gil, C. (2010). Identification of *Candida albicans* exposed surface proteins *in vivo* by a rapid proteomic approach. *J. Proteomics* 73, 1404–1409. doi: 10.1016/j.jprot.2010.02.008

Hube, B. (2006). Infection-associated genes of *Candida albicans*. *Future Microbiol.* 1, 209–218. doi: 10.2217/17460913.1.2.209

Inglis, D. O., Arnaud, M. B., Binkley, J., Shah, P., Skrzypek, M. S., Wymore, F., et al. (2012). The *Candida* genome database incorporates multiple *Candida* species: multispecies search and analysis tools with curated gene and protein

information for *Candida albicans* and *Candida glabrata*. *Nucleic Acids Res.* 40, D667–D674. doi: 10.1093/nar/gkr945

Insenser, M. R., Hernáez, M. L., Nombela, C., Molina, M., Molero, G., and Gil, C. (2010). Gel and gel-free proteomics to identify *Saccharomyces cerevisiae* cell surface proteins. *J. Proteomics* 73, 1183–1195. doi: 10.1016/j.jprot.2010.02.005

Jiang, H., Wagner, E., Zhang, H., and Frank, M. M. (2001). Complement 1 inhibitor is a regulator of the alternative complement pathway. *J. Exp. Med.* 194, 1609–1616. doi: 10.1084/jem.194.11.1609

Jong, A. Y., Chen, S. H., Stins, M. F., Kim, K. S., Tuan, T. L., and Huang, S. H. (2003). Binding of *Candida albicans* enolase to plasmin(ogen) results in enhanced invasion of human brain microvascular endothelial cells. *J. Med. Microbiol.* 52, 615–622. doi: 10.1099/jmm.0.05060-0

Joo, M. Y., Shin, J. H., Jang, H. C., Song, E. S., Kee, S. J., Shin, M. G., et al. (2013). Expression of *SAP5* and *SAP9* in *Candida albicans* biofilms: comparison of bloodstream isolates with isolates from other sources. *Med. Mycol.* 51, 892–896. doi: 10.3109/13693786.2013.824623

Karkowska-Kuleta, J., Rapala-Kozik, M., and Kozik, A. (2009). Fungi pathogenic to humans: molecular bases of virulence of *Candida albicans, Cryptococcus neoformans* and *Aspergillus fumigatus*. *Acta Biochim. Pol.* 56, 211–224.

Kazatchkine, M. D., and Jouvin, M. H. (1984). Interactions between the alternative complement pathway and proteases of the coagulation system. *Adv. Exp. Med. Biol.* 167, 235–239. doi: 10.1007/978-1-4615-9355-3_19

Klis, F. M., Sosinska, G. J., de Groot, P. W., and Brul, S. (2009). Covalently linked cell wall proteins of *Candida albicans* and their role in fitness and virulence. *FEMS Yeast Res.* 9, 1013–1028. doi: 10.1111/j.1567-1364.2009.00541.x

Kretschmar, M., Felk, A., Staib, P., Schaller, M., Hess, D., Callapina, M., et al. (2002). Individual acid aspartic proteinases (Saps) 1-6 of *Candida albicans* are not essential for invasion and colonization of the gastrointestinal tract in mice. *Microb. Pathog.* 32, 61–70. doi: 10.1006/mpat.2001.0478

Kumamoto, C. A., and Vinces, M. D. (2005). Contributions of hyphae and hypha-co-regulated genes to *Candida albicans* virulence. *Cell. Microbiol.* 7, 1546–1554. doi: 10.1111/j.1462-5822.2005.00616.x

Law, R. H., Zhang, Q., McGowan, S., Buckle, A. M., Silverman, G. A., Wong, W., et al. (2006). An overview of the serpin superfamily. *Genome Biol.* 7, 216. doi: 10.1186/gb-2006-7-5-216

Lee, K. L., Buckley, H. R., and Campbell, C. C. (1975). An amino acid liquid synthetic medium for the development of mycelial and yeast forms of *Candida albicans*. *Sabouraudia* 13, 148–153. doi: 10.1080/00362177585190271

Leleu, G., Aegerter, P., and Guidet, B. (2002). Systemic candidiasis in intensive care units: a multicenter, matched-cohort study. *J. Crit. Care* 17, 168–175. doi: 10.1053/jcrc.2002.35815

Lesiak-Markowicz, I., Vogl, G., Schwarzmuller, T., Speth, C., Lass-Flörl, C., Dierich, M. P., et al. (2011). *Candida albicans* Hgt1p, a multifunctional evasion molecule: complement inhibitor, CR3 analogue, and human immunodeficiency virus-binding molecule. *J. Infect. Dis.* 204, 802–809. doi: 10.1093/infdis/jir455

Lu, Y., Su, C., and Liu, H. (2014). *Candida albicans* hyphal initiation and elongation. *Trends Microbiol.* 22, 707–714. doi: 10.1016/j.tim.2014.09.001

Luo, S., Blom, A. M., Rupp, S., Hipler, U. C., Hube, B., Skerka, C., et al. (2011). The pH-regulated antigen 1 of *Candida albicans* binds the human complement inhibitor C4b-binding protein and mediates fungal complement evasion. *J. Biol. Chem.* 286, 8021–8029. doi: 10.1074/jbc.M110.130138

Luo, S., Hoffmann, R., Skerka, C., and Zipfel, P. F. (2013). Glycerol-3-phosphate dehydrogenase 2 is a novel factor H-, factor H-like protein 1-, and plasminogen-binding surface protein of *Candida albicans*. *J. Infect. Dis.* 207, 594–603. doi: 10.1093/infdis/jis718

Luo, S., Poltermann, S., Kunert, A., Rupp, S., and Zipfel, P. F. (2009). Immune evasion of the human pathogenic yeast *Candida albicans*: Pra1 is a Factor H, FHL-1 and plasminogen binding surface protein. *Mol. Immunol.* 47, 541–550. doi: 10.1016/j.molimm.2009.07.017

Mayer, F. L., Wilson, D., and Hube, B. (2013). *Candida albicans* pathogenicity mechanisms. *Virulence* 4, 119–128. doi: 10.4161/viru.22913

Meri, T., Blom, A. M., Hartmann, A., Lenk, D., Meri, S., and Zipfel, P. F. (2004). The hyphal and yeast forms of *Candida albicans* bind the complement regulator C4b-binding protein. *Infect. Immun.* 72, 6633–6641. doi: 10.1128/IAI.72.11.6633-6641.2004

Mitchell, P. (2010). Proteomics retrenches. *Nat. Biotechnol.* 28, 665–670. doi: 10.1038/nbt0710-665

Mochon, A. B., Jin, Y., Kayala, M. A., Wingard, J. R., Clancy, C. J., Nguyen, M. H., et al. (2010). Serological profiling of a *Candida albicans* protein microarray reveals permanent host-pathogen interplay and stage-specific responses during candidemia. *PLoS Pathog.* 6:e1000827. doi: 10.1371/annotation/eff399e1-51e0-43b1-bdf4-1a11e9ada9bd

Monniot, C., Boisramé, A., Da Costa, G., Chauvel, M., Sautour, M., Bougnoux, M. E., et al. (2013). Rbt1 protein domains analysis in *Candida albicans* brings insights into hyphal surface modifications and Rbt1 potential role during adhesion and biofilm formation. *PLoS ONE* 8:e82395. doi: 10.1371/journal.pone.0082395

Naglik, J. R., Challacombe, S. J., and Hube, B. (2003). *Candida albicans* secreted aspartyl proteinases in virulence and pathogenesis. *Microbiol. Mol. Biol. Rev.* 67, 400–428. doi: 10.1128/MMBR.67.3.400-428.2003

Nanjappa, V., Thomas, J. K., Marimuthu, A., Muthusamy, B., Radhakrishnan, A., Sharma, R., et al. (2014). Plasma Proteome Database as a resource for proteomics research: 2014 update. *Nucleic Acids Res.* 42, D959–D965. doi: 10.1093/nar/gkt1251

Nielsen, E. W., Johansen, H. T., Høgåsen, K., Wuillemin, W., Hack, C. E., and Mollnes, T. E. (1996). Activation of the complement, coagulation, fibrinolytic and kallikrein-kinin systems during attacks of hereditary angioedema. *Scand. J. Immunol.* 44, 185–192. doi: 10.1046/j.1365-3083.1996.d01-298.x

Nombela, C., Gil, C., and Chaffin, W. L. (2006). Non-conventional protein secretion in yeast. *Trends Microbiol.* 14, 15–21. doi: 10.1016/j.tim.2005.11.009

Pardo, M., Monteoliva, L., Vázquez, P., Martínez, R., Molero, G., Nombela, C., et al. (2004). *PST1* and *ECM33* encode two yeast cell surface GPI proteins important for cell wall integrity. *Microbiology* 150, 4157–4170. doi: 10.1099/mic.0.26924-0

Paréj, K., Dobó, J., Závodszky, P., and Gál, P. (2013). The control of the complement lectin pathway activation revisited: both C1-inhibitor and antithrombin are likely physiological inhibitors, while alpha2-macroglobulin is not. *Mol. Immunol.* 54, 415–422. doi: 10.1016/j.molimm.2013.01.009

Peerschke, E. I., Yin, W., and Ghebrehiwet, B. (2008). Platelet mediated complement activation. *Adv. Exp. Med. Biol.* 632, 81–91. doi: 10.1007/978-0-387-78952-1_7

Pitarch, A., Abian, J., Carrascal, M., Sánchez, M., Nombela, C., and Gil, C. (2004). Proteomics-based identification of novel *Candida albicans* antigens for diagnosis of systemic candidiasis in patients with underlying hematological malignancies. *Proteomics* 4, 3084–3106. doi: 10.1002/pmic.200400903

Pitarch, A., Jiménez, A., Nombela, C., and Gil, C. (2006). Decoding serological response to *Candida* cell wall immunome into novel diagnostic, prognostic, and therapeutic candidates for systemic candidiasis by proteomic and bioinformatic analyses. *Mol. Cell. Proteomics* 5, 79–96. doi: 10.1074/mcp.M500243-MCP200

Pitarch, A., Jiménez, A., Nombela, C., and Gil, C. (2008). Serological proteome analysis to identify systemic candidiasis patients in the intensive care unit: analytical, diagnostic and prognostic validation of anti-*Candida* enolase antibodies on quantitative clinical platforms. *Proteomics Clin. Appl.* 2, 596–618. doi: 10.1002/prca.200780039

Pitarch, A., Nombela, C., and Gil, C. (2011). Prediction of the clinical outcome in invasive candidiasis patients based on molecular fingerprints of five anti-*Candida* antibodies in serum. *Mol. Cell. Proteomics* 10:M110.004010. doi: 10.1074/mcp.M110.004010

Poltermann, S., Kunert, A., von der Heide, M., Eck, R., Hartmann, A., and Zipfel, P. F. (2007). Gpm1p is a factor H-, FHL-1-, and plasminogen-binding surface protein of *Candida albicans*. *J. Biol. Chem.* 282, 37537–37544. doi: 10.1074/jbc.M707280200

Ragni, E., Calderon, J., Fascio, U., Sipiczki, M., Fonzi, W. A., and Popolo, L. (2011). Phr1p, a glycosylphosphatidylinsitol-anchored beta(1,3)-glucanosyltransferase critical for hyphal wall formation, localizes to the apical growth sites and septa in *Candida albicans*. *Fungal Genet. Biol.* 48, 793–805. doi: 10.1016/j.fgb.2011.05.003

Rittirsch, D., Flierl, M. A., and Ward, P. A. (2008). Harmful molecular mechanisms in sepsis. *Nat. Rev. Immunol.* 8, 776–787. doi: 10.1038/nri2402

Rodríguez de Córdoba, S., Esparza-Gordillo, J., Goicoechea de Jorge, E., Lopez-Trascasa, M., and Sánchez-Corral, P. (2004). The human complement factor H: functional roles, genetic variations and disease associations. *Mol. Immunol.* 41, 355–367. doi: 10.1016/j.molimm.2004.02.005

Sjöholm, K., Karlsson, C., Linder, A., and Malmström, J. (2014). A comprehensive analysis of the *Streptococcus pyogenes* and human plasma protein interaction network. *Mol. Biosyst.* 10, 1698–1708. doi: 10.1039/C3MB 70555B

Soares, R. M., de A Soares, R. M., Alviano, D. S., Angluster, J., Alviano, C. S., and Travassos, L. R. (2000). Identification of sialic acids on the cell surface of *Candida albicans*. *Biochim. Biophys. Acta* 1474, 262–268. doi: 10.1016/S0304-4165(00)00003-9

Soltis, R. D., Hasz, D., Morris, M. J., and Wilson, I. D. (1979). Studies on the nature of heat-labile anti-complementary activity in normal human serum. *Clin. Exp. Immunol.* 37, 310–322.

Sonesson, A., Ringstad, L., Nordahl, E. A., Malmsten, M., Mörgelin, M., and Schmidtchen, A. (2007). Antifungal activity of C3a and C3a-derived peptides against *Candida*. *Biochim. Biophys. Acta* 1768, 346–353. doi: 10.1016/j.bbamem.2006.10.017

Sosinska, G. J., de Koning, L. J., de Groot, P. W., Manders, E. M., Dekker, H. L., Hellingwerf, K. J., et al. (2011). Mass spectrometric quantification of the adaptations in the wall proteome of *Candida albicans* in response to ambient pH. *Microbiology* 157, 136–146. doi: 10.1099/mic.0.044206-0

Sudbery, P. E. (2011). Growth of *Candida albicans* hyphae. *Nat. Rev. Microbiol.* 9, 737–748. doi: 10.1038/nrmicro2636

Uniprot, C. (2014). Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* 42, D191–D198. doi: 10.1093/nar/gkt1140

Urban, C., Xiong, X., Sohn, K., Schröppel, K., Brunner, H., and Rupp, S. (2005). The moonlighting protein Tsa1p is implicated in oxidative stress response and in cell wall biogenesis in *Candida albicans*. *Mol. Microbiol.* 57, 1318–1341. doi: 10.1111/j.1365-2958.2005.04771.x

Verschoor, A., and Langer, H. F. (2013). Crosstalk between platelets and the complement system in immune protection and disease. *Thromb. Haemost.* 110, 910–919. doi: 10.1160/TH13-02-0102

Vialas, V., Sun, Z., Loureiro Y Penha, C. V., Carrascal, M., Abián, J., Monteoliva, L., et al. (2014). A *Candida albicans* PeptideAtlas. *J. Proteomics* 97, 62–68. doi: 10.1016/j.jprot.2013.06.020

Viudes, A., Pemán, J., Cantón, E., Ubeda, P., López-Ribot, J. L., and Gobernado, M. (2002). Candidemia at a tertiary-care hospital: epidemiology, treatment, clinical outcome and risk factors for death. *Eur. J. Clin. Microbiol. Infect. Dis.* 21, 767–774. doi: 10.1007/s10096-002-0822-1

Vonk, A. G., De Bont, N., Netea, M. G., Demacker, P. N., van der Meer, J. W., Stalenhoef, A. F., et al. (2004). Apolipoprotein-E-deficient mice exhibit an increased susceptibility to disseminated candidiasis. *Med. Mycol.* 42, 341–348. doi: 10.1080/13693780410001657135

Walport, M. J. (2001a). Complement. First of two parts. *N. Engl. J. Med.* 344, 1058–1066. doi: 10.1056/NEJM200104053441406

Walport, M. J. (2001b). Complement. Second of two parts. *N. Engl. J. Med.* 344, 1140–1144. doi: 10.1056/NEJM200104123441506

Wilson, L. S., Reyes, C. M., Stolpman, M., Speckman, J., Allen, K., and Beney, J. (2002). The direct cost and incidence of systemic fungal infections. *Value Health* 5, 26–34. doi: 10.1046/j.1524-4733.2002.51108.x

Yan, L., Yang, C., and Tang, J. (2013). Disruption of the intestinal mucosal barrier in *Candida albicans* infections. *Microbiol. Res.* 168, 389–395. doi: 10.1016/j.micres.2013.02.008

Zipfel, P. F., Skerka, C., Kupka, D., and Luo, S. (2011). Immune escape of the human facultative pathogenic yeast *Candida albicans*: the many faces of the *Candida* Pra1 protein. *Int. J. Med. Microbiol.* 301, 423–430. doi: 10.1016/j.ijmm.2011.04.010

Zybailov, B. L., Florens, L., and Washburn, M. P. (2007). Quantitative shotgun proteomics using a protease with broad specificity and normalized spectral abundance factors. *Mol. Biosyst.* 3, 354–360. doi: 10.1039/b701483j

Structure and Interactions of a Dimeric Variant of sHIP, a Novel Virulence Determinant of *Streptococcus pyogenes*

Carl Diehl [1,2], Magdalena Wisniewska [1,3], Inga-Maria Frick [4], Werner Streicher [1,5], Lars Björck [4], Johan Malmström [4] and Mats Wikström [1*†]

[1] Protein Function and Interactions Group, Faculty of Health and Medical Sciences, The Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Copenhagen, Denmark, [2] SARomics Biostructures, Lund, Sweden, [3] Malopolska Centre of Biotechnology, Krakow, Poland, [4] Division of Infection Medicine, Department of Clinical Sciences, Lund University, Lund, Sweden, [5] Novozymes A/S, Bagsvaerd, Denmark

Streptococcus pyogenes is one of the most significant bacterial pathogens in the human population mostly causing superficial and uncomplicated infections (pharyngitis and impetigo) but also invasive and life-threatening disease. We have previously identified a virulence determinant, protein sHIP, which is secreted at higher levels by an invasive compared to a non-invasive strain of *S. pyogenes.* The present work presents a further characterization of the structural and functional properties of this bacterial protein. Biophysical and structural studies have shown that protein sHIP forms stable tetramers both in the crystal and in solution. The tetramers are composed of four helix-loop-helix motifs with the loop regions connecting the helices displaying a high degree of flexibility. Owing to interactions at the tetramer interface, the observed tetramer can be described as a dimer of dimers. We identified three residues at the tetramer interface (Leu84, Leu88, Tyr95), which due to largely non-polar side-chains, could be important determinants for protein oligomerization. Based on these observations, we produced a sHIP variant in which these residues were mutated to alanines. Biophysical experiments clearly indicated that the sHIP mutant appear only as dimers in solution confirming the importance of the interfacial residues for protein oligomerisation. Furthermore, we could show that the sHIP mutant interacts with intact histidine-rich glycoprotein (HRG) and the histidine-rich repeats in HRG, and inhibits their antibacterial activity to the same or even higher extent as compared to the wild type protein sHIP. We determined the crystal structure of the sHIP mutant, which, as a result of the high quality of the data, allowed us to improve the existing structural model of the protein. Finally, by employing NMR spectroscopy in solution, we generated a model for the complex between the sHIP mutant and an HRG-derived heparin-binding peptide, providing further molecular details into the interactions involving protein sHIP.

Keywords: proteomics, structural biology, virulence factors, protein-protein interactions, host-pathogen relationship

INTRODUCTION

Streptococcus pyogenes, also known as group A streptococci (GAS) is a significant human pathogen that infects and colonizes the skin and the upper respiratory tract where it causes relatively mild clinical conditions such as impetigo and pharyngitis. Some infections caused by invasive strains, such as the AP1 strain of the M1 serotype, can lead to severe and potentially life-threatening diseases such as necrotizing fasciitis and streptococcal toxic shock syndrome (STSS), whereas acute rheumatic fever and glomerulonephritis are sequelae to acute *S. pyogenes* infections. *S. pyogenes* causes an estimated 700 million cases of mild and non-invasive infections each year, of which ∼650,000 progress to severe invasive infections with a mortality of at least 25% (Carapetis et al., 2005; Ralph and Carapetis, 2013). *S. pyogenes* produces a number of proteins that enable the bacterium to attach to host tissues, evade the immune response, and spread by penetrating host tissue layers. These virulence factors are predominantly secreted or surface associated proteins, and they include the family of M proteins (Lancefield, 1962; Swanson et al., 1969; Phillips et al., 1981), fibronectin-binding proteins (Talay et al., 1994; Kreikemeyer et al., 1995; Jaffe et al., 1996; Courtney et al., 1999; Rocha and Fischetti, 1999; Terao et al., 2001), super-antigenic exotoxins (Stevens et al., 1989; Abe et al., 1991; Tomai et al., 1992; Mollick et al., 1993; Norrby-Teglund et al., 1994), and the secreted streptococcal inhibitor of complement referred to as protein SIC (Åkesson et al., 1996; Fernie-King et al., 2002; Frick et al., 2003).

In a recent study, we quantitatively analyzed and compared *S. pyogenes* proteins in the growth medium of a strain that is virulent to mice (AP1), with a non-virulent strain (SF370). We found that one protein in particular was present at significantly higher levels in the stationary growth medium from the virulent strain. The amount of sHIP in the medium fraction is similar to the secreted mitogenic exotoxin SmeZ (Kamezawa et al., 1997), and shows an overall abundance profile resembling the profile observed for the surface associated proteins H and M1 (Lancefield, 1962; Åkesson et al., 1990; Wisniewska et al., 2014). Through the use of affinity pull-down mass spectrometry analysis of human plasma, we could demonstrate that the new bacterial protein interacts with the antimicrobial human protein histidine-rich glycoprotein (HRG), and the name sHIP (streptococcal Histidine-rich glycoprotein Interacting Protein) was therefore introduced (Wisniewska et al., 2014). HRG is an abundant plasma glycoprotein of approximately 60 kDa that interacts with several other protein ligands such as tropomyosin, heparin, plasminogen, plasmin, fibrinogen, and IgG (Jones et al., 2005). HRG has been shown to exhibit broad antimicrobial potency (Rydengård et al., 2007), including activity against *S. pyogenes* (Shannon et al., 2010). It was shown that sHIP binds both intact HRG and HRG-derived peptides (peptides containing consensus heparin-binding sequences) with high affinity. Moreover, the antibacterial activity of HRG is blocked by protein sHIP, which represents a new mechanism that can contribute to the virulence of AP1 bacteria. In addition, we could show that patients with severe *S. pyogenes* infection, in contrast to patients with superficial and uncomplicated infections, are more prone to develop antibodies against sHIP, which suggests that sHIP represents a novel virulence determinant (Wisniewska et al., 2014). Furthermore, the determination of the three-dimensional structure of sHIP, showed that it has a tetrameric organization composed of four helix-loop-helix motifs. A similar structural unit can be found in the adhesion factor FadA from *Fusobacterium nucleatum*. However, the two proteins differ significantly in their respective oligomeric organization. In FadA, the helix-loop-helix motifs form elongated fibers whereas the sHIP monomers are organized into a compact tetrameric structure. In order to understand the molecular prerequisites for the observed interaction between sHIP and HRG, we have performed extensive crystallization trials but not been able to obtain any diffracting crystals of complexes between sHIP and HRG or any HRG derived peptides. Due to the size of the tetrameric sHIP, NMR experiments in solution have proven to be very challenging and would benefit from access to a sHIP variant with lower molecular weight and retained activity. We have previously made the observation that the tetrameric structure of sHIP can also be described as a dimer of dimers, and postulated that three residues in the dimer-dimer interface could represent important determinants for the stabilization of the tetrameric form of sHIP. In this study, we have mutated these three interfacial residues and thereby been able to produce a stable dimeric variant of sHIP that was shown to be active both in biophysical binding experiments and an antimicrobial assay. The generation of stable dimers enabled further characterization of the interaction between sHIP and a peptide from HRG through NMR experiments, providing the first molecular details of an interaction involving this novel virulence determinant.

MATERIALS AND METHODS

Mutagenesis

The DNA sequence corresponding to residues Lys3-Met98 in the *S. pyogenes* protein sHIP (UniprotID: Q99XU0) was expressed as described previously and denoted sHIPwt (Wisniewska et al., 2014). A protein sHIP variant, a quadruple mutant named sHIPqp, in which Leu84, Leu88, Tyr95 were mutated to alanine, and Cys65 was mutated to serine, was constructed using the Quikchange XL site-directed mutagenesis kit (Agilent) following the manufacturer's instructions. We have previously observed that the wildtype protein can form larger aggregates through the exposed Cys65 residue in the absence of a reducing agent. The mutation Cys65Ser was introduced in order to avoid the need to use a reducing agent in subsequent experiments. As previously noted, this mutation has no effect on the structure and function of sHIP (Wisniewska et al., 2014). We are therefore using the Cys65Ser mutant in all studies involving both the wildtype sHIP and in the sHIP variant. All mutations were verified by DNA sequence analysis. The resulting expression construct contained a His-tag and a TEV protease cleavage site preceding the protein sequence of interest.

Protein Expression and Purification

The protein sHIP variants sHIPwt and sHIPqp were expressed in *E. coli* strain Rosetta BL21 (DE3) and purified in two steps

using standard immobilized metal ion affinity chromatography (IMAC) followed by proteolytic removal of the His-tag, and reverse phase chromatography (RPC), as described previously (Wisniewska et al., 2014). For the NMR studies, protein was produced using minimal medium for cell growth in order to introduce the stable isotopes ^{15}N and ^{13}C enabling triple-resonance NMR experiments (Neidhardt et al., 1974). The His-tag was removed by treatment with tobacco etch virus (TEV) protease giving the following amino acid sequences for the wildtype protein sHIP, sHIPwt:

SMKQDQLIVEKMEQTYEAFSPKLANLIEALDAFKEHYEE
YATLRNFYSSDEWFRLANQPWDDIPCGVLSEDLLFDMIG
DHNQLLADILDLAPIMYKHM,

and the quadruple mutant protein sHIP, sHIPqp:

SMKQDQLIVEKMEQTYEAFSPKLANLIEALDAFKEHYEE
YATLRNFYSSDEWFRLANQPWDDIPSGVLSEDLLFDMIG
DHNQLAADIADLAPIMAKHM, respectively.

The purity and mono-dispersity of the recombinant proteins were verified by SDS-PAGE electrophoresis and mass spectrometry.

Crystallization, X-Ray Data Collection, and Structure Determination

Crystallization of the protein sHIP variant was carried out using the sitting drop vapor diffusion method at 18°C. The best crystals were obtained from 20% PEG 6000, 0.2 M calcium chloride and 0.1 M Tris, pH 8.0. Prior to plunge freezing, the crystals were soaked for approximately 30 s in a drop of a reservoir solution containing 20% v/v ethylene glycol as cryo-protectant. The crystals belonged to the space group P22121 and contained one monomer per asymmetric unit. The highest resolution data set of 180 frames (1° oscillation range) was collected from an orthorhombic crystal at Max II lab beamline I911-2, Lund, Sweden, using a MAR CCD detector. Data were indexed and integrated using iMOSFLM (Battye et al., 2011) and scaled using SCALA (Evans, 2006) from the CCP4 suite of programs. The structure was solved by molecular replacement using wild type sHIP as the search model (PDB ID: 4MER), and refined using REFMAC (Murshudov et al., 1997). Refinement rounds were complemented with manual rebuilding using COOT (Emsley et al., 2010). Water molecules were automatically inserted using Arp/wArp and visually inspected. For further details on data processing and refinement statistics, see **Table 1**. Geometry of the model was analyzed with Molprobity (Chen et al., 2010). Molecular graphics images were generated using the program Pymol. The atomic coordinates and structure factors for the sHIP mutant (sHIPqp) crystal structure have been deposited with the protein data bank (PDB ID: 4PZ1).

Analytical Ultracentrifugation

Sedimentation velocity experiments were performed at 20°C and 50000 rpm using a Beckman XL-I instrument. Samples containing either sHIPwt or sHIPqp were in 20 mM MES buffer

TABLE 1 | Crystallographic data collection and refinement statistics.

Protein	sHIPqp mutant
DATA COLLECTION	
Beamline	I911-2
Wavelength (Å)	1.038
Space group	P22121
Cell dimensions	
a, b, c (Å)	29.37 33.28 89.58
α, β, γ (°)	90 90 90
Resolution (Å)	19.76–1.73 (1.82–1.73)
R_{meas}(%)	6.4 (14.6)
I/sigI	19.9 (10.3)
Completeness (%)	98.8 (97.8)
Multiplicity	6.4 (6.2)
REFINEMENT	
Resolution (Å)[a]	19.76–1.73 (1.77–1.73)
No. reflections	9077 (657)
R_{work}/R_{free}	18.9/21.1 (20.3/38)
NO. ATOMS	
Protein	792
Chloride	2
Calcium	1
Water	78
B-FACTORS (Å^2)	
Protein	14.8
Water	20.56
RMSD STEREOCHEMISTRY[b]	
Bond lengths (Å)	0.011
Bond angles (°)	1.295
RAMACHANDRAN PLOT (%)	
Most favored regions	100

[a]Values for the highest resolution shell values are indicated in parentheses.
[b]RMSD, root mean square deviation.
alpha, beta, gamma refers to the unit cell axes dimensions, while alpha, beta, and gamma in symbol refers to the inclination angles of the axes in the unit cell. The space group number P22121 refers to the description of the symmetry of the crystal.

(2-(N-morpholino) ethanesulfonic acid), pH 5.5. The data were analyzed with SEDFIT using a continuous c (S) distribution (Schuck, 2000). HYDROPRO (Ortega et al., 2011) was used to calculate the theoretical sedimentation coefficient using the PDB IDs 4MER (sHIPwt) and 4PZ1 (sHIPqp), respectively.

Synthetic Peptide

The sequence for the HRG peptide used in the study referred to as HRGsingle, corresponding to a single heparin-binding motif, have the amino acid sequence GHHPHG. The peptide was synthesized by Biosyntan GmbH (Berlin, Germany) and purity and molecular mass were confirmed by MALDI-TOF MS.

Antimicrobial Assay

The *S. pyogenes* strain AP1 (40/58), was from the World Health Organization Collaborating Centre for Reference and Research on Streptococci, Prague, Czech Republic. The bacteria were cultivated in THY (Todd-Hewitt broth (Difco) supplemented with 0.2% yeast extract (Oxoid) at 37°C and 5% CO_2 until
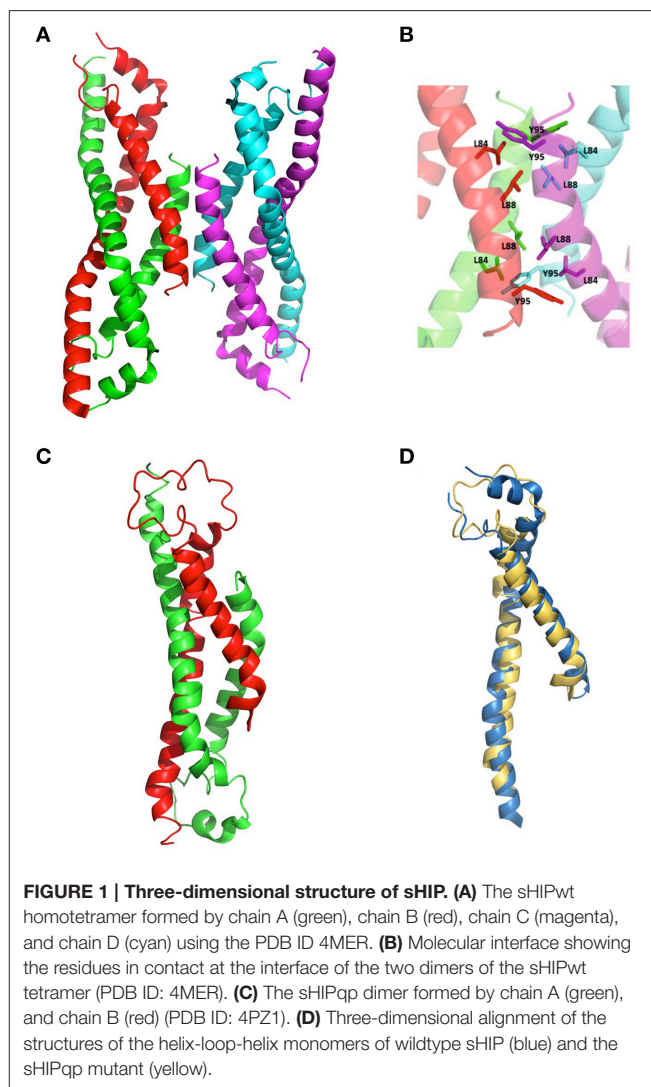
reaching mid-log phase (OD 620 nm approximately 0.4). The bacterial cells were washed and resuspended in 10 mM Tris-HCl, pH 7.5, containing 5 mM glucose, to a concentration of 2×10^9 cfu (colony forming unit)/ml. Subsequently, the bacteria were diluted to a concentration of 2×10^6 cfu/ml in 10 mM MES buffer, pH 5.5, containing 5 mM glucose. Fifty microliters of the bacterial solution was incubated with recombinant His-tagged HRG (Creative Biomart) at a concentration of 0.45 μM together with various concentrations of protein sHIPwt or protein sHIPqp for 40 min at 37°C. Serial dilutions of the incubation mixtures were plated on TH agar, plates were incubated over night at 37°C and the number of cfu's were determined.

Isothermal Titration Calorimetry (ITC)

ITC experiments were performed using a VP-ITC200 instrument (GE Healthcare). The samples were extensively dialyzed against 20 mM MES buffer, pH 5.5. 0.5 mM HRGsingle peptide was titrated into 0.05 mM of sHIPwt, or sHIPqp respectively. All experiments were performed at 25°C and run until saturation was achieved. The data were fitted using a model describing one independent binding site using the software provided by the manufacturer (Wiseman et al., 1989).

NMR Experiments

Protein samples were prepared in 20 mM MES pH 5.5 with 7% (v/v) D2O for the spectrometer lock and 0.02% (w/v) NaN$_3$ to prevent bacterial growth in the samples. All NMR experiments were performed at 25°C. For the studies of the sHIPqp/HRGsingle peptide complex, the sample contained 1.0 mM of double-labeled (^{15}N, ^{13}C) protein sHIPqp and 1.1 mM of the HRGsingle peptide. The backbone and side-chain resonance assignment experiments were performed at 800 MHz proton resonance frequency, whereas the experiments in order to acquire Nuclear Overhauser Enhancement (NOE) distance restraints were acquired at 900 MHz proton resonance frequency. The assignment of the signals from sHIPqp and the HRGsingle peptide, and collection of NOE restraints were performed by using a combination of standard 2D and 3D triple-resonance experiments including HNCA (Clore et al., 1990), HNCOCA (Bax and Ikura, 1991), HNCACB (Wittekind and Mueller, 1993), CBCACONH (Grzesiek and Bax, 1992), HCCONH (Montelione et al., 1992), CCONH (Montelione et al., 1992), ^{15}N-NOESY-HSQC (Marion et al., 1989), ^{13}C-NOESY-HSQC (separate for aliphatic and aromatic signals; Marion et al., 1989), 2D double-filtered NOESY (Ikura et al., 1992), and 2D single-filtered NOESY (Ikura et al., 1992). In addition, 2D NOESY and TOCSY experiments were acquired for the HRGsingle peptide. Spectra were processed with NMRpipe (Delaglio et al., 1995), applying zero-filling and linear prediction in the indirect dimensions and solvent filter and polynomial baseline correction in the direct dimension. CCPNMR (Vranken et al., 2005) was used for visualization of spectra, resonance connectivity analysis, and distance restraints (NOE) assignments. Structure calculations were performed using an iterative procedure within the ARIA/CNS suite of programs (Brunger et al., 1998). The assessment of the quality of the NMR generated structural models, were performed using the Protein Structure Validation

FIGURE 1 | Three-dimensional structure of sHIP. (A) The sHIPwt homotetramer formed by chain A (green), chain B (red), chain C (magenta), and chain D (cyan) using the PDB ID 4MER. **(B)** Molecular interface showing the residues in contact at the interface of the two dimers of the sHIPwt tetramer (PDB ID: 4MER). **(C)** The sHIPqp dimer formed by chain A (green), and chain B (red) (PDB ID: 4PZ1). **(D)** Three-dimensional alignment of the structures of the helix-loop-helix monomers of wildtype sHIP (blue) and the sHIPqp mutant (yellow).

Suite (PSVS, **Table 3**; Bhattacharya et al., 2007). Molecular graphics images were generated using the program Pymol (**Figure 5**).

RESULTS

The sHIPqp Mutant Forms Stable Dimers of Two Helix-Loop-Helix Motifs

The wildtype sHIP is composed of tetramers that can be described as the dimer of two dimers each harboring two helix-loop-helix motifs (**Figure 1A**). The hypothesis was that the three non-polar residues at the tetramer interface (Leu84, Leu88, Tyr95) represent important determinants for the protein oligomerization (**Figure 1B**), and these residues were therefore mutated to alanines in the sHIPqp mutant protein. The orthorhombic crystals of the sHIPqp mutant contain one monomer in the asymmetric unit, with a solvent content of 36%. The crystal packing shows the formation of a dimer with a total buried surface area of 6070 Å^2 as calculated by
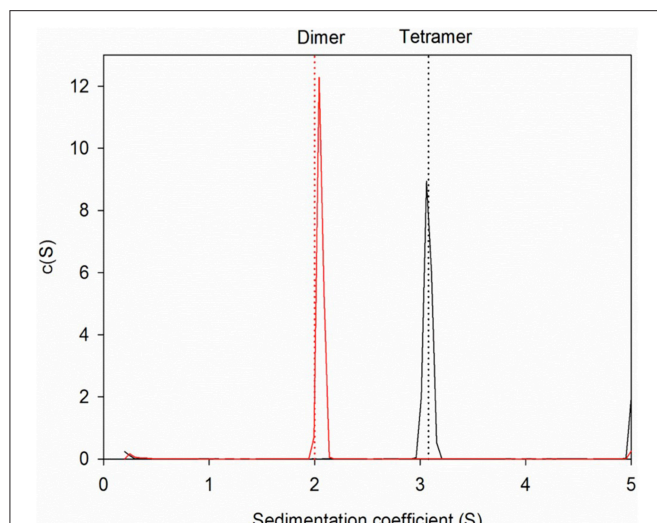
FIGURE 2 | Sedimentation velocity experiments for protein sHIP. The dotted lines indicate the theoretical sedimentation coefficient for a tetramer and a dimer respectively calculated using HYDROPRO and the structural coordinates for protein sHIPwt (PDB ID: 4MER) and protein sHIPqp (PDB ID: 4PZ1). The wild-type sHIPwt is shown in black, and the sHIPqp mutant in red respectively.
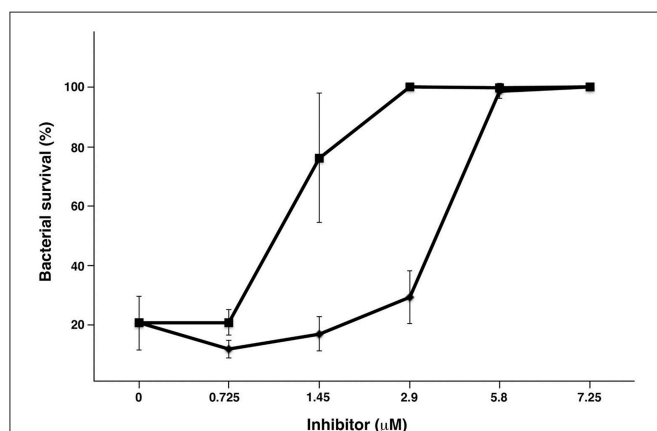

FIGURE 3 | Protein sHIPqp mutant inhibits the bactericidal activity of HRG more efficiently than protein sHIP wildtype. The bactericidal effect of HRG at 0.45 μM against *S. pyogenes* AP1 bacteria (2 × 10^6 cfu/ml) was inhibited with sHIP wildtype (◆), and sHIPqp mutant (■) at indicated concentrations. Experiments were repeated at least three times and mean values ± SD are shown.

PISA server (Krissinel and Henrick, 2007). The final model comprises residues 3-98, corresponding to the entire sHIPqp molecule (**Figure 1C**). The structure is very well defined, with an average B factor of 12.2 Å^2 (10.8 Å^2 and 13.6 Å^2 for main chain and side chains, respectively). Final electron density maps are of very high quality including the loop region, which was partially not visible in sHIP wild type (**Figure 1D**). All non-glycine residues exhibit main-chain angles in the favored regions of the Ramachandran plot as defined by the use of the program Molprobity (Chen et al., 2010; **Table 1**). To provide

a quantitative analysis of the oligomeric states of the sHIP variants, analytical ultracentrifugation was employed on sHIPwt and the sHIPqp mutant. These experiments clearly show that the wildtype sHIP (sHIPwt) appear as tetramers in solution whereas the sHIP mutant (sHIPqp) exist as dimers (**Figure 2**). The largest differences in the monomeric helix-loop-helix motif between the two sHIP variants are observed in the loop region connecting the two helices. In the sHIPqp mutant, all residues, including side-chains, are very well defined, while in sHIPwt, some residues could not be defined due to the absence of interpretable electron density maps in this region (**Figure 1D**).

The sHIPqp Mutant Inhibits the Antibacterial Activity of HRG

It is known that HRG has antibacterial activity including activity against *S. pyogenes* (Shannon et al., 2010), and when challenged by HRG, sHIP has been shown to rescue *S. pyogenes* bacteria (Wisniewska et al., 2014). To examine whether a sHIP mutant has the same ability, we performed an antimicrobial assay comparing sHIPwt and sHIPqp. As shown in **Figure 3**, sHIPqp inhibits the bactericidal activity of HRG to the same or even higher extent that wildtype sHIP indicating that the dimeric mutant sHIPqp retains its inhibitory capacity.

sHIP Binds a Single Heparin-Binding Motif in HRG

In our previous studies we have shown that protein sHIP binds to both HRG and HRG-derived peptides with nanomolar to micromolar affinity, and that the histidine-rich region is important for the interaction (Wisniewska et al., 2014). To further dissect the molecular details in the interaction, we investigated whether the sHIP variants are also able to bind a single heparin-binding motif (GHHPHG) in HRG (denoted as HRGsingle). Isothermal titration calorimetry (ITC) experiments showed that both sHIPwt and sHIPqp binds HRGsingle with similar micro-molar affinities, clearly indicating that the heparin-binding motif represents an important recognition motif in HRG for the interaction with protein sHIP (**Figure 4**, **Table 2**).

The HRG-Binding Site in sHIP Involves Both Helix-Loop-Helix Motifs of the Dimer

The wildtype protein sHIP is a tetrameric protein with a molecular weight of 46 kDa complicating NMR experiments in solution due to broad and overlapping resonances. The creation of the mutant dimeric form of sHIP, sHIPqp, and through the verification of its functional properties, we now have access to a smaller, yet biologically relevant system with a molecular weight of 23kDa, which enabled detailed NMR studies in solution.

A molecular model was created with a combination of the structure of sHIPqp (PDB ID: 4PZ1), and distance restraints (NOEs) extracted from the NMR experiments of the complex between sHIPqp and the HRGsingle peptide. A total of 1714 NOEs were used as distance restraints in the structure calculations of which 321 are long-range restraints and 56 represent NOEs detected between sHIPqp and the HRGsingle peptide (**Table 3**). It should be noted that this model represent
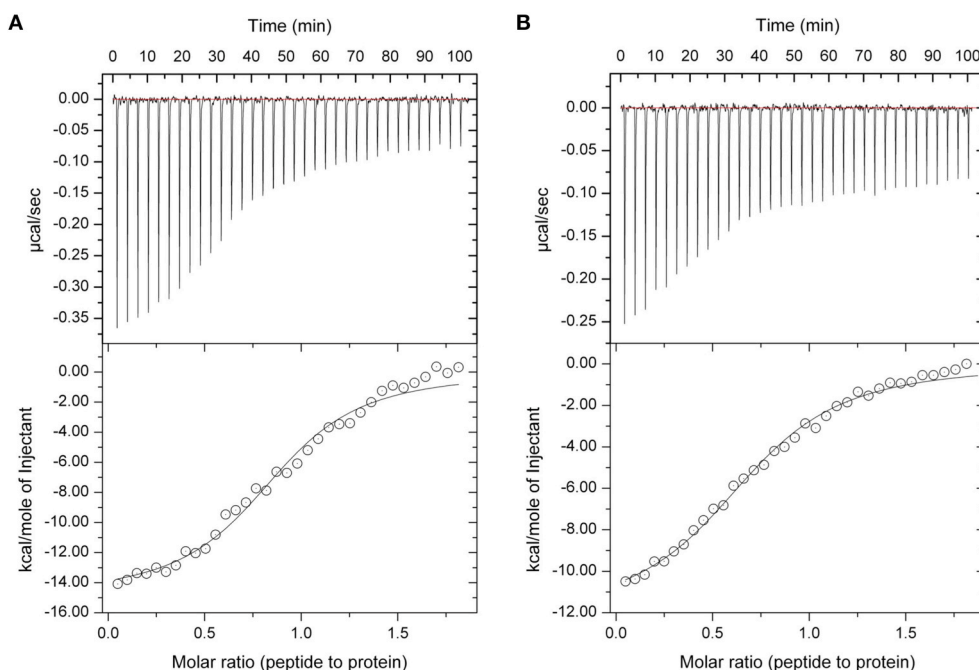
FIGURE 4 | Isothermal titration Calorimetry (ITC) experiments showing the binding between sHIP and the HRGsingle peptide. (A) sHIPwt and HRGsingle. **(B)** sHIPqp and HRGsingle. The raw data of the experiments are presented on the top panel. The area underneath each injection peak is equal to the total heat released for that injection. When this integrated heat is plotted against the molar ratio of titrant added to the peptide solution in the cell, a complete binding isotherm for the interaction is obtained (bottom panel). A model for either one binding site was used to fit the data. The solid line is the calculated curve using the best-fit parameters.

TABLE 2 | Summary of thermodynamic parameters from ITC experiments of the binding of the two sHIP variants sHIPwt and sHIPqp to the HRGsingle peptide respectively.

Complex	N (sites)	K_d (µM)	ΔH (kcal/mol)
sHIPwt/HRGsingle	0.90 ± 0.02	1.9 ± 0.2	−15.1 ± 0.4
sHIPqp/HRGsingle	0.79 ± 0.03	2.8 ± 0.3	−12.2 ± 0.3

a preliminary description of the sHIPqp-HRGsingle complex that will need further refinement before considered a finalized complex structure. However, the assessment of the generated structure shows that it is of sufficient quality to represent a most relevant descriptive model for the complex (**Table 3**). The model shows that the HRGsingle peptide binds in a pocket created by residues from both monomeric helix-loop-helix motifs, suggesting that the dimeric form represents the smallest sHIP variant attainable with retained HRG-binding capacity. Each sHIP dimer binds two HRGsingle peptides in two identical sites (**Figure 5A**). The binding site includes the N-terminal part of helix 1 of one monomer (residues Asp5-Met12), and the end region of helix 1 of the second monomer (residues Ala41-Ala56). For the first monomer three residues are involved in NOE contacts with the HRGsingle peptide with primary contribution from Ile8. For the second monomer, thirteen residues give rise to detectable NOEs dominated by contacts involving Phe46, Glu51, and Arg54. Moreover, the binding site does not involve

any residues in the proximity of the dimer-dimer interface in agreement with the observation that the mutant sHIPqp maintain its full inhibitory capacity of HRG in antimicrobial assays (**Figure 3**). The two external Glycine residues of the HRGsingle peptide, Gly-1 and Gly-6, are not involved in any interactions with sHIP. However, the core part of the peptide, (HHPH), forms multiple contacts with a binding pocket created by residues from both chains of the structural framework created by the sHIP dimer. In particular, residues His-2 and His-5 are positioned deeply in the binding pocket, a confirmation of the peptide backbone that is made possible due to the unique backbone properties of the connecting Pro-4 residue (**Figure 5B**).

DISCUSSION

Bacterial infections represent a major challenge to human health worldwide, a challenge increasing at an elevated pace due to the alarming upsurge of antibiotic resistance. In order to treat bacterial infections in the future, new therapeutic strategies need to be developed that will require a better understanding of the complex molecular interplay between bacterial pathogens and their human host. *S. pyogenes* represents one of the most significant bacterial pathogens in the human population. Using mass spectrometry based proteomics, we previously identified a novel protein and virulence factor from this bacterium (Wisniewska et al., 2014). In this study we have extended the structural and functional characterization of this protein to

TABLE 3 | NMR structural restraints and structure statistics of the sHIPqp/HRGsingle model.

Summary of experimental restraints[a]	
NOE-based distance constraints:	
Total	1714
intra-residue [i = j]	403
sequential [| i - j | = 1]	467
medium range [1 < | i - j | < 5]	523
long range [| i - j | > 5]	321
sHIPqp/HRGsingle intermolecular NOEs	56
NOE constraints per restrained residue	17.0
Dihedral-angle constraints:	697
Total number of restricting constraints	2411
Total number of restricting constraints per restrained residue[b]	23.9
Restricting long-range constraints per restrained residue[b]	3.2
Structure Quality Factors–overall statistics	
[b]Procheck G-factor (phi/psi only)	−0.88
[b]Procheck G-factor (all dihedral angles)	−1.04
MolProbity clashscore	73.12
[c]Ramachandran plot (%)	
Most favored regions	70.1
Additionally allowed regions	28.3
Generously allowed regions	0.5
Disallowed regions	1.1

[a]Analyzed for residues 2 to 98 (sHIPqp chain A), 102-198 (sHIPqp chain B), 301-306 (HRGsingle peptide 1), and 401-406 (HRGsingle peptide 2).

[b]PROCHECK (Laskowski et al., 1993, 1996).

[c]Ramachandran statistics calculated by PROCHECK (Laskowski et al., 1993, 1996).

FIGURE 5 | Molecular model of the complex between sHIPqp and the HRGsingle peptide. (A) sHIPqp is shown in a ribbon representation with chain A in green and chain B in red. The HRGsingle peptide is shown in a stick representation. **(B)** sHIPqp is shown in a space filling representation with chain A in green and chain B in red. The HRGsingle peptide is shown in a stick representation.

further our understanding of the interactions between sHIP and the human host.

The structure determination of the sHIPqp mutant showed that the fold of the sHIP mutant and the sHIP wild type monomers is essentially identical, displaying an elongated helix-loop-helix motif (**Figure 1**). However, the most significant difference is related to the oligomeric states of the two protein variants. For wildtype sHIP, four monomers assemble into a compact homotetramer, which can be described as a dimer of dimers due to the interactions at the dimer-dimer interface. Our previous studies showed that this interface consists of amino acids with largely non-polar side-chains, such as Leu84, Leu88, Tyr95, and Met98, which can be responsible for protein oligomerization. The mutation of these residues to alanine's resulted in a shift in quaternary structure from tetramers to dimers further evidenced by sedimentation velocity experiments that clearly show that the sHIPqp mutant exists as dimers in solution (**Figure 2**).

To gain further understanding of the interaction between sHIP and HRG, a molecular model was created for the complex between sHIPqp and a single heparin-binding motif of HRG (HRGsingle), a peptide shown to be active in binding to both sHIPwt and sHIPqp (**Figure 4**). The model shows that the HRGsingle peptide interacts with both monomers of the helix-loop-helix dimers, and that a proline situated in the middle of the peptide sequence allows two histidines, His-2 and His-5, to enter deeply into the binding pocket created by the
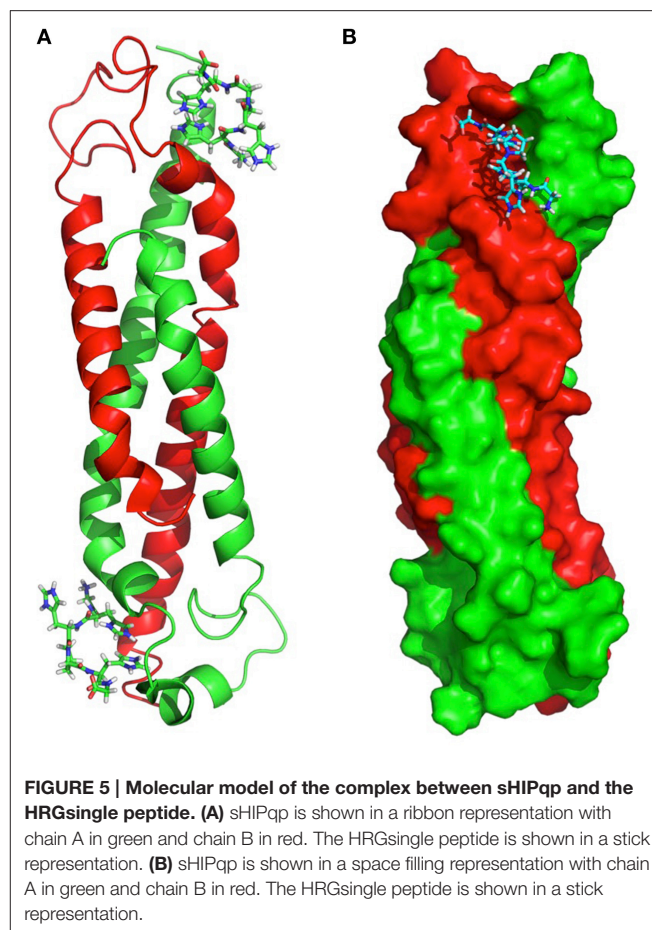
two sHIP monomers (**Figures 5A,B**). However, even if both monomers contribute to contacts with the HRGsingle peptide, one of the monomers clearly dominates the overall contribution to the binding epitope for the sHIP dimer. Future studies should involve mutagenesis experiments of both sHIP and HRG peptides to further define the critical determinants for the interaction.

CONCLUDING REMARKS

Through structure-based protein engineering of a selected set of residues in the dimer-dimer interface of sHIP, we were able to generate a dimeric form of the protein with retained functional properties. The access to dimeric sHIP provides the opportunity to utilize NMR techniques in order to further elucidate interactions involving this protein. In conclusion, we have through a combination of protein engineering, biophysics, and structural biology methods, provided the first molecular details describing the interaction between the antimicrobial protein HRG and the novel virulence factor sHIP.

AUTHOR CONTRIBUTIONS

CD and MagW performed the NMR and X-ray experiments, respectively. IF performed the antimicrobial assay, WS the

REFERENCES

Abe, J., Forrester, J., Nakahara, T., Lafferty, J. A., Kotzin, B. L., and Leung, D. Y. (1991). Selective stimulation of human T cells with streptococcal erythrogenic toxins A and B. *J. Immunol.* 146, 3747–3750.

Åkesson, P., Cooney, J., Kishimoto, F., and Björck, L. (1990). Protein H–a novel IgG binding bacterial protein. *Mol. Immunol.* 27, 523–531. doi: 10.1016/0161-5890(90)90071-7

Åkesson, P., Sjöholm, A. G., and Björck, L. (1996). Protein SIC, a novel extracellular protein of *Streptococcus pyogenes* interfering with complement function. *J. Biol. Chem.* 271, 1081–1088. doi: 10.1074/jbc.271.2.1081

Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R., and Leslie, A. G. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D Biol. Crystallogr.* 67, 271–281. doi: 10.1107/S0907444910048675

Bax, A., and Ikura, M. (1991). An efficient 3D NMR technique for correlating the proton and 15N backbone amide resonances with the alpha-carbon of the preceding residue in uniformly 15N/13C enriched proteins. *J. Biomol. NMR* 1, 99–104. doi: 10.1007/BF01874573

Bhattacharya, A., Tejero, R., and Montelione, G. T. (2007). Evaluating protein structures determined by structural genomics consortia. *Proteins* 66, 778–795. doi: 10.1002/prot.21165

Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., et al. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* 54, 905–921. doi: 10.1107/S0907444998003254

Carapetis, J. R., Steer, A. C., Mulholland, E. K., and Weber, M. (2005). The global burden of group A streptococcal diseases. *Lancet Infect. Dis.* 5, 685–694. doi: 10.1016/S1473-3099(05)70267-X

Chen, V. B., Arendall, W. B. III, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., et al. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 66, 12–21. doi: 10.1107/S0907444909042073

Clore, G. M., Bax, A., Driscoll, P. C., Wingfield, P. T., and Gronenborn, A. M. (1990). Assignment of the side-chain proton and carbon-13 resonances of interleukin-1. beta. using double-and triple-resonance heteronuclear three-dimensional NMR spectroscopy. *Biochemistry* 29, 8172–8184. doi: 10.1021/bi00487a027

Courtney, H. S., Hasty, D. L., Li, Y., Chiang, H. C., Thacker, J. L., and Dale, J. B. (1999). Serum opacity factor is a major fibronectin-binding protein and a virulence determinant of M type 2 *Streptococcus pyogenes. Mol. Microbiol.* 32, 89–98. doi: 10.1046/j.1365-2958.1999.01328.x

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol.* 6, 277–293. doi: 10.1007/BF00197809

Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* 66, 486–501. doi: 10.1107/S0907444910007493

Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* 62, 72–82. doi: 10.1107/S0907444905036693

Fernie-King, B. A., Seilly, D. J., Davies, A., and Lachmann, P. J. (2002). Streptococcal inhibitor of complement inhibits two additional components of the mucosal innate immune system: secretory leukocyte proteinase inhibitor and lysozyme. *Infect. Immun.* 70, 4908–4916. doi: 10.1128/IAI.70.9.4908-4916.2002

Frick, I. M., Åkesson, P., Rasmussen, M., Schmidtchen, A., and Björck, L. (2003). SIC, a secreted protein of *Streptococcus pyogenes* that inactivates antibacterial peptides. *J. Biol. Chem.* 278, 16561–16566. doi: 10.1074/jbc.M301995200

Grzesiek, S., and Bax, A. (1992). Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.* 114, 6291–6293. doi: 10.1021/ja00042a003

Ikura, M., Clore, G. M., Gronenborn, A. M., Zhu, G., Klee, C. B., and Bax, A. (1992). Solution structure of a calmodulin-target peptide complex by multidimensional NMR. *Science* 256, 632–638. doi: 10.1126/science.1585175

Jaffe, J., Natanson-Yaron, S., Caparon, M. G., and Hanski, E. (1996). Protein F2, a novel fibronectin-binding protein from *Streptococcus pyogenes*, possesses two binding domains. *Mol. Microbiol.* 21, 373–384. doi: 10.1046/j.1365-2958.1996.6331356.x

Jones, A. L., Hulett, M. D., and Parish, C. R. (2005). Histidine-rich glycoprotein: a novel adaptor protein in plasma that modulates the immune, vascular and coagulation systems. *Immunol. Cell Biol.* 83, 106–118. doi: 10.1111/j.1440-1711.2005.01320.x

Kamezawa, Y., Nakahara, T., Nakano, S., Abe, Y., Nozaki-Renard, J., and Isono, T. (1997). Streptococcal mitogenic exotoxin Z, a novel acidic superantigenic toxin produced by a T1 strain of *Streptococcus pyogenes. Infect. Immun.* 65, 3828–3833.

Kreikemeyer, B., Talay, S. R., and Chhatwal, G. S. (1995). Characterization of a novel fibronectin-binding surface protein in group A streptococci. *Mol. Microbiol.* 17, 137–145. doi: 10.1111/j.1365-2958.1995.mmi_17010137.x

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 372, 774–797. doi: 10.1016/j.jmb.2007.05.022

Lancefield, R. C. (1962). Current knowledge of type-specific M antigens of group A streptococci. *J. Immunol.* 89, 307–313.

Laskowski, R. A., Macarthur, M. W., Moss, D. S., and Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 26, 283–291. doi: 10.1107/S0021889892009944

Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* 8, 477–486. doi: 10.1007/BF00228148

Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., et al. (1989). Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear 1H-15N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta. *Biochemistry* 28, 6150–6156. doi: 10.1021/bi00441a004

Mollick, J. A., Miller, G. G., Musser, J. M., Cook, R. G., Grossman, D., and Rich, R. R. (1993). A novel superantigen isolated from pathogenic strains of *Streptococcus pyogenes* with aminoterminal homology to staphylococcal enterotoxins B and C. *J. Clin. Invest.* 92, 710–719. doi: 10.1172/JCI116641

Montelione, G. T., Lyons, B. A., Emerson, S. D., and Tashiro, M. (1992). An efficient triple resonance experiment using carbon-13 isotropic mixing for determining sequence-specific resonance assignments of isotopically-enriched proteins. *J. Am. Chem. Soc.* 114, 10974–10975. doi: 10.1021/ja00053a051

Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* 53, 240–255. doi: 10.1107/S0907444996012255

Neidhardt, F. C., Bloch, P. L., and Smith, D. F. (1974). Culture medium for enterobacteria. *J. Bacteriol.* 119, 736–747.

Norrby-Teglund, A., Newton, D., Kotb, M., Holm, S. E., and Norgren, M. (1994). Superantigenic properties of the group A streptococcal exotoxin SpeF (MF). *Infect. Immun.* 62, 5227–5233.

Ortega, A., Amorós, D., and García de la Torre, J. (2011). Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys. J.* 101, 892–898. doi: 10.1016/j.bpj.2011.06.046

Phillips, G. N. Jr., Flicker, P. F., Cohen, C., Manjula, B. N., and Fischetti, V. A. (1981). Streptococcal M protein: alpha-helical coiled-coil structure and arrangement on the cell surface. *Proc. Natl. Acad. Sci. U.S.A.* 78, 4689–4693. doi: 10.1073/pnas.78.8.4689

Ralph, A. P., and Carapetis, J. R. (2013). Group a streptococcal diseases and their global burden. *Curr. Top. Microbiol. Immunol.* 368, 1–27. doi: 10.1007/82_2012_280

Rocha, C. L., and Fischetti, V. A. (1999). Identification and characterization of a novel fibronectin-binding protein on the surface of group A streptococci. *Infect. Immun.* 67, 2720–2728.

Rydengård, V., Olsson, A. K., Mörgelin, M., and Schmidtchen, A. (2007). Histidine-rich glycoprotein exerts antibacterial activity. *FEBS J.* 274, 377–389. doi: 10.1111/j.1742-4658.2006.05586.x

Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* 78, 1606–1619. doi: 10.1016/S0006-3495(00)76713-0

Shannon, O., Rydengård, V., Schmidtchen, A., Mörgelin, M., Alm, P., Sørensen, O. E., et al. (2010). Histidine-rich glycoprotein promotes bacterial entrapment in clots and decreases mortality in a mouse model of sepsis. *Blood* 116, 2365–2372. doi: 10.1182/blood-2010-02-271858

Stevens, D. L., Tanner, M. H., Winship, J., Swarts, R., Ries, K. M., Schlievert, P. M., et al. (1989). Severe group A streptococcal infections associated with a toxic shock-like syndrome and scarlet fever toxin A. *N. Engl. J. Med.* 321, 1–7. doi: 10.1056/NEJM198907063210101

Swanson, J., Hsu, K. C., and Gotschlich, E. C. (1969). Electron microscopic studies on streptococci. I. M antigen. *J. Exp. Med.* 130, 1063–1091. doi: 10.1084/jem.130.5.1063

Talay, S. R., Valentin-Weigand, P., Timmis, K. N., and Chhatwal, G. S. (1994). Domain structure and conserved epitopes of Sfb protein, the fibronectin-binding adhesin of *Streptococcus pyogenes*. *Mol. Microbiol.* 13, 531–539. doi: 10.1111/j.1365-2958.1994.tb00448.x

Terao, Y., Kawabata, S., Kunitomo, E., Murakami, J., Nakagawa, I., and Hamada, S. (2001). Fba, a novel fibronectin-binding protein from *Streptococcus pyogenes*, promotes bacterial entry into epithelial cells, and the fba gene is positively transcribed under the Mga regulator. *Mol. Microbiol.* 42, 75–86. doi: 10.1046/j.1365-2958.2001.02579.x

Tomai, M. A., Schlievert, P. M., and Kotb, M. (1992). Distinct T-cell receptor V beta gene usage by human T lymphocytes stimulated with the streptococcal pyrogenic exotoxins and pep M5 protein. *Infect. Immun.* 60, 701–705.

Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., et al. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* 59, 687–696. doi: 10.1002/prot.20449

Wiseman, T., Williston, S., Brandts, J., and Lin, L. (1989). Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.* 179, 131–137. doi: 10.1016/0003-2697(89)90213-3

Wisniewska, M., Happonen, L., Kahn, F., Varjosalo, M., Malmström, L., Rosenberger, G., et al. (2014). Functional and structural properties of a novel protein and virulence factor (Protein sHIP) in *Streptococcus pyogenes*. *J. Biol. Chem.* 289, 18175–18188. doi: 10.1074/jbc.M114.565978

Wittekind, M., and Mueller, L. (1993). HNCACB, a High-Sensitivity 3D NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha- and Beta-Carbon Resonances in Proteins. *J. Magn. Reson. Ser. B* 101, 201–205. doi: 10.1006/jmrb.1993.1033

Proteogenomic Analysis of *Mycobacterium smegmatis* Using High Resolution Mass Spectrometry

Matthys G. Potgieter[1], Kehilwe C. Nakedi[2], Jon M. Ambler[1], Andrew J. M. Nel[2], Shaun Garnett[2], Nelson C. Soares[2], Nicola Mulder[1]* and Jonathan M. Blackburn[2]*

[1] Computational Biology Division, Department of Integrative Biomedical Sciences, IDM, University of Cape Town, Cape Town, South Africa, [2] Division of Chemical and Systems Biology, Department of Integrative Biomedical Sciences, IDM, University of Cape Town, Cape Town, South Africa

Biochemical evidence is vital for accurate genome annotation. The integration of experimental data collected at the proteome level using high resolution mass spectrometry allows for improvements in genome annotation by providing evidence for novel gene models, while validating or modifying others. Here, we report the results of a proteogenomic analysis of a reference strain of *Mycobacterium smegmatis* (mc^{2}155), a fast growing model organism for the pathogenic *Mycobacterium tuberculosis*—the causative agent for Tuberculosis. By integrating high throuput LC/MS/MS proteomic data with genomic six frame translation and *ab initio* gene prediction databases, a total of 2887 ORFs were identified, including 2810 ORFs annotated to a Reference protein, and 63 ORFs not previously annotated to a Reference protein. Further, the translational start site (TSS) was validated for 558 Reference proteome gene models, while upstream translational evidence was identified for 81. In addition, N-terminus derived peptide identifications allowed for downstream TSS modification of a further 24 gene models. We validated the existence of six previously described interrupted coding sequences at the peptide level, and provide evidence for four novel frameshift positions. Analysis of peptide posterior error probability (PEP) scores indicates high-confidence novel peptide identifications and shows that the genome of *M. smegmatis* mc^{2}155 is not yet fully annotated. Data are available via ProteomeXchange with identifier PXD003500.

Keywords: *Mycobacterium smegmatis* mc^{2}155, mass spectrometry, proteogenomics, genome annotation, proteomics

INTRODUCTION

Evidence for the existence of protein coding genes include *ab initio* gene predictions, transcriptomic analysis, and comparative genomics information (Krug et al., 2011). Although, gene annotation of model organisms often relies on transcript sequencing, it has become apparent that evidence of transcription may not equal evidence of translation (Castellana and Bafna, 2010). Proteomics data, on the other hand, gives direct evidence of which genes are translated. Using proteomics data, true genes can be separated from pseudogenes, and the translational frame can be determined. The location of TSSs (translational start sites) can be identified, while signal cleavage and other post-translational modifications (PTMs) can be identified (Kucharova and Wiker, 2014). By mapping to uncharacterized genomic coordinates, novel genes can be identified (Borchert et al., 2010).

Genomic six frame translation allows for the creation of a database containing, in the ideal case, all possible putative proteins, but at the cost of including many spurious entries. This leads to decreased sensitivity of identifications at the same FDR as standard proteomic databases with a higher proportion of non-spurious entries (Castellana and Bafna, 2010). Due to the large database sizes involved in proteogenomics, and large numbers of spectra that need to be assigned, many false positive identifications are obtained even at a low error rate. To limit the number of spurious entries, thus increasing the sensitivity of identifications, proteogenomic databases can be compacted, by excluding entries below a minimum length cutoff or only focusing on genomic regions identified by *ab initio* gene prediction tools (Castellana and Bafna, 2010).

Automated gene prediction at the early stage of genome annotation is prone to errors, with rates of incorrect TSS prediction of up to 44% reported (Gallien et al., 2009), while short protein-coding genes are difficult to predict (Renuse et al., 2011). Accurately identifying the TSS of a gene is complicated by the existence of different possible start codons, with many non-standard start codons identified in prokaryotes (Castellana and Bafna, 2010). Mycobacteria are known to use GTG and TTG as initiator *Met* (*fMet*) start codons, in addition to translation of these codons with *Val* and *Leu,* respectively (Kelkar et al., 2011).

Wang et al. (2005) used databases of gene predictions to identify mass spectra, identifying 901 proteins in *Mycobacterium smegmatis* using the partially sequenced genome available at the time—validating many predicted genes. Gallien et al. (2009) identified 946 proteins in *M. smegmatis*, characterizing 443 N-terminal peptides, and revealed an error rate of 19% in predicted TSSs. Kelkar et al. (2011) reported 41 novel protein-coding genes in *Mycobacterium tuberculosis* strain H37Rv. By identifying N-terminal peptides, the authors were able to correct the TSS of 33 proteins, and validate the TSS of 727 annotated proteins. Strikingly, the authors identified eight proteins with evidence for translation initiation at two different sites.

The occurrence of false positive peptide identifications from tandem mass spectrometry data implies the need for statistical concepts such as the false discovery rate (FDR) and posterior error probability (PEP) score to gauge the reliability of identifications. The FDR is controlled by scoring and ranking peptide spectral matches (PSMs), and identifying the score above which a maximum allowable proportion of PSMs from a decoy database are identified—indicating the expected maximum proportion of incorrect PSMs—while posterior error probability (PEP) indicates the probability of error of a single PSM (Käll et al., 2008). The *MaxQuant* algorithms calculate PEP scores based on search engine PSM scores and peptide length, and determine a cutoff PEP score at a specific FDR (Cox and Mann, 2008).

Krug et al. (2013) compared the number and median PEP scores of novel peptide identifications with reverse sequence hit peptides in a proteogenomic analysis of *Escherichia coli*, and determined that the absolute number as well as the median PEP-values for both groups were very similar. The authors concluded that as a genome approaches complete annotation, the likelihood of any novel peptide identification being a false positive identification increases (Krug et al., 2013).

Annotation of coding regions for organisms is a continuously evolving process. The genome of *M. tuberculosis* was sequenced in 1998 (Cole et al., 1998), yet a few years later, the same authors re-annotated the genome and identified 71 more ORFs than before (Camus et al., 2002). To date, the *M. tuberculosis* genome is annotated with 4018 ORFs[1]. This highlights the importance of continuous review and re-annotation of genomes as technologies improve. The *M. smegmatis* mc[2]155 genome may not yet be fully annotated, and the genome sequence has been shown to contain multiple errors (Deshayes et al., 2007), one of the reasons being the high GC content and genome annotation shortfalls, such as short protein validation and incorrect TSS assignment of genes. That poses a problem by limiting our understanding of the many cellular processes coded by these genes. It is imperative that the genome of *M. smegmatis* mc[2]155 be fully annotated since this bacterium, due to its non-pathogenic and fast growing status, is used frequently as a model organism to study the biology of *M. tuberculosis*, the causative agent of Tuberculosis. Tuberculosis continues to be a burden on the health system, with an estimated 9.6 million cases of infection and 1.5 million deaths in 2014, despite a globally decreasing incidence of ∼1.5% every year since 2000 (WHO, 2015). Limited understanding of the biology of *M. tuberculosis* is an obstacle to improving current treatment and eradication of the disease. Thus, it is hoped that refinements in proteogenomics pipelines and the study of model organisms of *M. tuberculosis* such as *M. smegmatis* mc[2]155, may further our understanding of this pathogenic organism.

In this study, we developed and applied a compacted six frame genomic database, to map high resolution and high accuracy tandem mass spectrometry coupled to liquid chromatography data to re-evaluate the genome annotation of *M. smegmatis* mc[2]155. We also used a database of *ab initio* gene predictions using *GeneMarkS*[2] in the proteogenomics pipeline to identify novel open reading frames, gene model validations, and gene model modifications. By analyzing the PEP score distribution of novel, annotated, and reverse sequence hit peptides, we explored the current annotation status of *M. smegmatis* mc[2]155.

MATERIALS AND METHODS

Bacterial Cultures
Wild type strain of *M. smegmatis* mc[2]155 were grown in 7H9 Middlebrook (BD, Maryland, USA) broth supplemented with 0.05% Tween 80, OADC (Becton Dickinson), and 0.2% glycerol (v/v). Cells were grown at 37°C with continuous agitation (120 rpm).

Protein Extraction
Cells were harvested from three biological replicates each during the exponential and early stationary phase (OD$_{600}$ ∼ 1.2 and 1.8, respectively) by centrifugation at 4000 g for 15 min at 4°C. Cell pellets were washed with PBS (10 mM phosphate buffer, 2.7 mM potassium chloride, and 137 mM sodium chloride, pH 7.4) and flash frozen in liquid nitrogen and stored at −80°C. Pellets were

---

[1]http://tuberculist.epfl.ch/

[2]GeneMark™—*Free Gene Prediction Software*." http://exon.gatech.edu/GeneMark/(Accessed September 21, 2015).

suspended in lysis buffer [500 mM Tris-HCl, 0.1% (w/v) SDS, 0.15% sodium deoxycholate], 1× protease inhibitor cocktail, 1× phosphatase inhibitor cocktail (Roche, Mannheim Germany), and 50 μg/ml lysozyme (Repaske, 1956) and disrupted by sonication at maximum power for six cycles of 30 s, with 1 min cooling on ice between cycles (Rezwan et al., 2007). Lysates were further clarified using centrifugation at 4000 g for 5 min and filtering through a 20 μm pore size low-protein binding filter (Merck, NJ, USA). Proteins were precipitated using the chloroform–methanol precipitation method as previously described (Wessel and Flügge, 1984). Protein precipitate was suspended in denaturing buffer (10 mM Tris-HCl, 6 M urea, 2 M thiourea, pH 8). Protein concentration was determined using the modified Bradford assay as described by Ramagli (1999).

In-Solution Trypsin Digestion

Fifty micrograms of precipitated protein was reduced with 1 mM DTT for 1 h followed by another hour of incubation in the presence of 5.5 mM IAA. Alkylated protein samples were pre-digested for 3 h at room temperature with lysyl endopeptidase LysC (Waco, Neuss, Germany). Pre-digested samples were diluted four-fold with 20 mM ammonium bicarbonate pH 8 prior to trypsination. Sequencing grade modified trypsin (Promega, Madison, USA) was used at a protease:protein ratio of 1:50 (w/w) for 14 h at room temperature with gentle agitation. Trypsination was terminated with 0.1% formic acid final concentration (Sigma Aldrich, St Louis, USA). Ten micrograms of peptides were desalted using a homemade stage tip containing Empore Octadecyl C18 solid-phase extraction disk (Supelco; Rappsilber et al., 2003). Activation, equilibration, and peptide wash and elution were all carried out using centrifugation at 5000 g for 5 min. Activation and equilibration of the C18 disk was carried out using three rinses with 80% acetonitrile (ACN), followed by three rinses with 2% ACN, respectively. Peptide rich solution was loaded onto the disk and centrifuged. Desalting was carried out using three washes of 2% ACN, followed by three washes of 2% ACN containing 0.1% formic acid (Sigma). Elution of desalted peptides into glass capillary tubes was carried out using three rounds of 100 μL of 60% ACN, 0.1% formic acid. Peptides were dried in a vacuum and resuspended in 2% ACN, 0.1% formic acid at 50 ng/μL.

LC/MS/MS Analysis

Data acquisition was performed on the Orbitrap Q-Exactive mass spectrometer (Thermo Scientific) in a data-dependent manner, coupled to the Dionex Ultimate 3000 UHPLC (Thermo Scientific). One microgram of peptides were loaded on to an inhouse packed pre-column (100 μm ID × 20 mm) connected to an in-house packed analytical column (75 μm × 400 mm) both packed with C18 Luna 5 μm 100 Å beads (04A-5452) for liquid chromatography separation. The flow rate was set to 300 nl/min with the gradient of 2% to 25% ACN for 125 min, then up to 35% in 5 min. To wash the column ACN was increased to 80% for 20 min followed by a column equilibration at 2% ACN for 10 min. A top 10 method with 30 s dynamic exclusion was used to acquire mass spectra with automatic switching between

MS and MS/MS scans. The LC/MS/MS methods used have been described previously (Nakedi et al., 2015).

Proteogenomic Databases

The genome of *M. smegmatis* mc^{2}155 was accessed from the *European Nucleotide Archive* under the accession number CP000480, Genome Assembly GCA_000015005.1 (Fleischmann et al., 2006). To facilitate genomic database compaction each open reading frame in the six frames was translated *in silico* starting at the most upstream start codon—using possible start codons ATG, GTG, and TTG—and sequences below a minimum translated length of 20 amino acids were excluded from the database. For sequences at the end of each of the genomic frames that did not end in a stop codon, the sequence up till the last in-frame codon was translated and included in the database. The genomic coordinates of the six frame translated sequences were included in the *FASTA* headers of the database. Translated sequences that occurred more than once in the database were combined into a single entry with multiple genomic coordinates. The Reference proteome for *M. smegmatis* mc^{2}155 was obtained from UniProt[3] (Proteome ID UP000000757), and any translated six frame sequence overlapping with or identical to a Reference protein sequence was mapped to that protein in the *FASTA* header. Where a Reference sequence was located downstream of a translated six frame TSS, the number of amino acids difference in the N-termini of the two sequences was recorded, and six frame sequences identical to a protein in the Reference proteome were labeled. Overlapping open reading frames were included in the database, leading to a final database size of 79,481 entries.

A second genomic database for targeted identification of TSSs was generated using *ab initio* prediction of protein-coding genes with the *GeneMarkS* software package (version 4.28)—at the same time providing supporting evidence for novel protein identifications and gene model modifications. This software uses a Hidden Markov Model algorithm, combined with models of protein coding and non-coding regions and gene regulation sites, to predict the occurrence of genes in a DNA sequence (Besemer et al., 2001). The predicted genes were translated and the final database contained 6655 sequences.

Database Search Using Maxquant

The *MaxQuant* software package (version 1.5.0.30) was used to search the raw MS spectra using the Andromeda search engine separately against the six frame database, UniProt Reference proteome and *GeneMarkS* database—using reverse decoy databases and a selection of known contaminants provided by *MaxQuant*. Trypsin and LysC were selected as enzymes, and a maximum of three missed cleavages were allowed. Specific enzyme mode was selected, which allowed for the detection of non-tryptic N-terminal peptides of database entries, peptides with N-Met cleavage, and fully tryptic peptides. Carbamidomethyl was set as a fixed modification and Acetyl (Protein N-term) and Oxidation (M) were set as variable modifications. A minimum peptide length of seven amino acids, and a minimum of one unique peptide identification per protein group, was required. The default *MaxQuant* false discovery

---

[3] *UniProt.* http://www.uniprot.org/ (Accessed September 21, 2015).

rate (FDR) cutoff of 0.01 (1%) was used at the PSM, peptide and protein group levels. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (Vizcaíno et al., 2014) via the *PRIDE* partner repository (Vizcaíno et al., 2013) with the dataset identifier PXD003500 and the data is freely available[4].

Proteogenomic Analysis

We relied heavily on previously published proteogenomic protocols, in particular the methodology described by Kelkar et al. (2011)—differing in that we used a separate search database (*GeneMarkS*) for targeted TSS identification, and did not exclude all novel peptides mapping to multiple genomic locations, allowing for the identification of paralogous translated ORF sequences.

All *MaxQuant* results for the different databases were combined using an in-house python script. Use was made of the *Biopython* software package[5] (Cock et al., 2009). Pettersen et al. (2015) only considered proteins identified in at least two replicates in their proteogenomic analysis of enterotoxigenic *Escherichia coli*—similarly, we only considered peptides identified in at least two replicates for further analysis. All peptides were searched against the genome dynamically translated in the six reading frames, and peptides unique to a single position in the genome were identified. Mycobacteria are known to have duplications in protein-coding genes due to the effect of transposable elements (Dale, 1995). To allow for the detection of paralogous sequences, peptides specific to a single repeating translated six frame ORF sequence in the genome were identified as paralogous sequence peptides, and also included in the analysis, but identified as belonging to paralogous translated ORF sequences.

Peptides not found in the Reference proteome were identified as genome search specific peptides (GSSPs). GSSPs are peptides mapping to genomic regions not considered to be coding regions, or not included in overlapping or adjacent gene models (Kelkar et al., 2011). We used the UniProt Reference proteome as a benchmark for annotation status to discriminate annotated peptides from GSSPs. Only GSSPs were used for upstream gene model modifications and novel gene model identifications, while N-terminal peptides mapping downstream in the gene model of Reference proteins were used for downstream TSS assignment. N-terminal peptides mapping to the start of Reference protein gene models were used for TSS validation. All protein groups with no peptides unique to the protein sequence after performing the genome search, were excluded from further analysis. All results in the different databases mapping to the same ORF were identified for comparison, and the unique identified peptide set across all three databases for each ORF was obtained. All database entries were mapped to their respective ORFs in the genome using an in-house python script (see **Figure 1**).

TSS peptides were identified by the non-tryptic nature of their N-terminals, *N-Met* cleavage, or N-terminal acetylation. Reference protein start sites were validated by identifying TSS

peptides mapping to the genomic coordinates of the start of a Reference protein sequence. All open reading frames mapping to Reference proteins, with peptides identified upstream in the ORF from the annotated TSS of the Reference protein, were identified for upstream gene model modification, while N-terminal peptides mapping downstream of a Reference protein start were used for downstream gene model modification. The coordinates of novel ORFs were compared to the coordinates of known cDNA features (sequenced transcripts obtained from Ensembl[6] using a sense strand genome coordinate search (in-house python script), and ORFs overlapping with cDNA evidence in the genome were identified. A nucleotide BLAST of novel ORFs was performed against the transcript sequences, requiring same strand alignment with an *E*-value cutoff of 0.0001. A protein BLAST of novel translated ORFs, as well as ORFs for which gene model modification were identified, was performed against the NCBI non-redundant (nr) BLAST database, and the highest scoring alignment by *E*-value was obtained for each sequence, using an *E*-value cutoff of 0.0001. The leading gi IDs of protein BLAST results were converted to UniProtKB IDs using the UniProt Retrieve/ID Mapping tool. Novel ORFs were ranked by number of identified GSSPs, protein BLAST evidence, cDNA nucleotide BLAST evidence, and cDNA coordinate overlap.

Interrupted CoDing Sequences (ICDSs) are erroneously shortened gene model predictions either due to the presence of unrecognized true genomic events (such as programmed frameshifts or in-frame stop codons), or artificially due to sequencing errors (Perrodou et al., 2006). We followed the methodology described by Perrodou et al. (2006) and Deshayes et al. (2007) to identify ICDSs based on shared homologous sequence evidence (protein BLAST) of adjacent or overlapping non-paralogous ORFs, but differ in that we only focused on ORFs identified at the peptide level.

All peptides identified in the three databases in at least two replicates, as well as the *GeneMarkS FASTA* database, were processed into general feature format (GFF) files for visualization using the Ensembl genome browser for this strain[7]—allowing for manual examination of novel annotations, gene model modifications, and ICDSs (see **Supplementary Data Sheets 1–4**).

The PEP score distributions of novel, annotated and reverse peptide identifications from each database were analyzed with in-house python and R scripts. Use was made of the python module *matplotlib* (Hunter, 2007) to produce boxplots of the PEP score distributions, while the R *ggplot2* package (Wickham, 2009), as well as the *density* and *qqnorm* functions from the R *stats* package (R Core Team, 2015) were used to plot the PEP score distributions and investigate for normality. As non-normal PEP score distributions were found, the non-parametric Kruskal–Wallis analysis of variance test was chosen to investigate for differences between the groups—using the *kruskal.test* function from the R *stats* package, followed by *post-hoc* analysis with a two-sided Dunn test (Dunn, 1964) with Bonferroni correction using the *dunnTest* function from the R *FSA* package (Ogle, 2016).

---

[4]http://www.ebi.ac.uk/pride/archive/

[5]*Biopython-Biopython*. Available online at: http://biopython.org/wiki/Main_Page (Accessed September 21, 2015).

[6]ftp://ftp.ensemblgenomes.org/pub/bacteria/current/fasta/bacteria_7_collection/mycobacterium_smegmatis_str_mc2_155/cdna/

[7]http://bacteria.ensembl.org/Mycobacterium_smegmatis_str_mc2_155/Info/Index
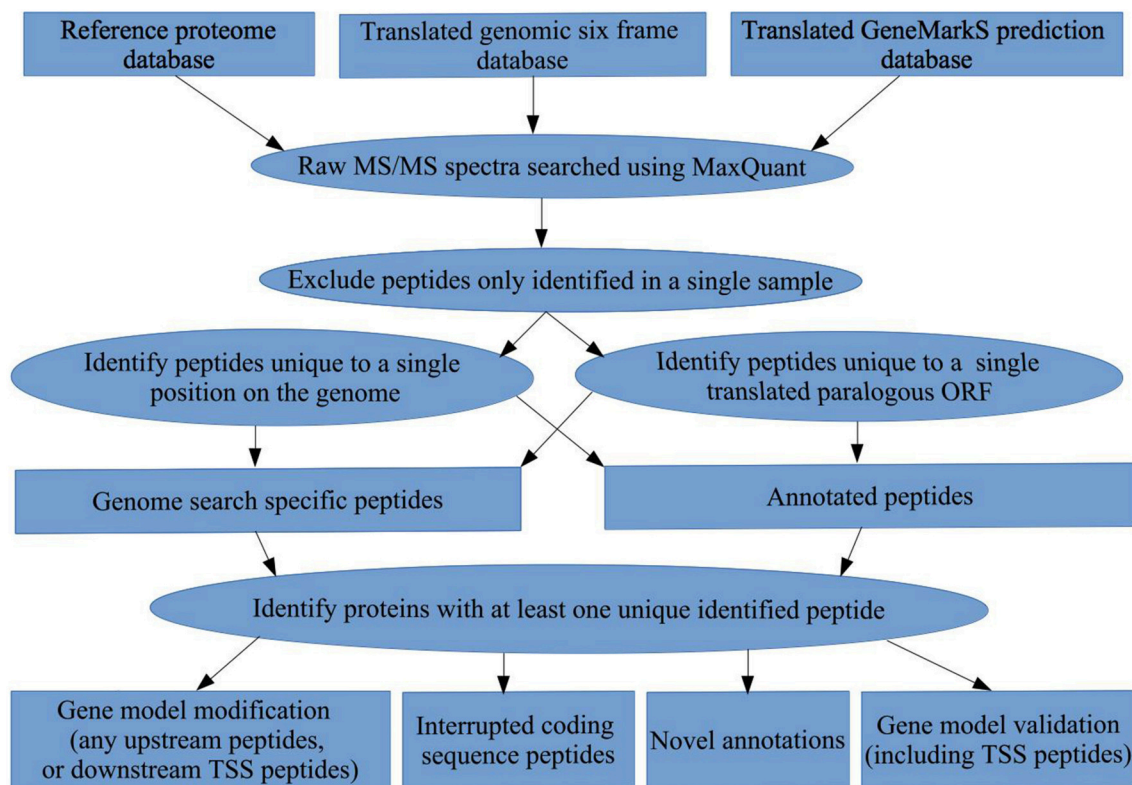
FIGURE 1 | Bioinformatics workflow. A summary of the proteogenomic pipeline applied to obtain gene model validations, gene model modifications, novel ORF annotations, and interrupted coding sequence identifications in *M. smegmatis* mc^{2}155, using a translated genomic six frame database, a translated *GeneMarkS* gene prediction database, and the Reference proteome.

Venn diagrams were produced using the online venn diagram plotting tool Venny 2.1.0 (Oliveros, 2016).

RESULTS

From the 276,472 MS/MS spectra submitted to *MaxQuant* at 1% FDR, 26,125 peptide sequences were identified from 172,570 spectra using the translated six frame database, 27,895 peptides from 176,518 spectra using the Reference proteome database, and 27,735 peptides from 176,301 spectra using the *GeneMarkS* database. Only protein groups were considered where the leading protein had at least one unique peptide identification. By mapping identified proteins from the different databases to their corresponding ORFs in the genome, 2887 ORFs were identified at the peptide level (identical translated ORF sequences with multiple occurrences in the genome having been combined into a single entry with multiple genomic coordinates at the database generation phase; see **Supplementary Table 1** and **Figure 2**).

TSS Peptide Identifications

TSS peptide identifications can be divided into peptides with and without *N-Met* cleavage. We identified 137 TSS peptides with non-tryptic *Met* N-termini at a genomic start codon position (of which three were acetylated at the N-terminal), and 549 *N-Met* cleaved peptides (of which 127 were acetylated at the N-terminal).

The distribution of penultimate amino acids of identified *N-Met* cleaved peptides—*Thr* (188), *Ser* (134), *Ala* (128), *Pro* (51), *Gly* (18), *Val* (15), *Asn* (13), *Leu* (1), and *Arg* (1)—corresponds to the non-random nature of *Met*-AP cleavage (Link et al., 1997; Frottin et al., 2006). Further, the identified start codon distribution of all non-paralogous ORFs with an identified N-terminal—ATG 63.96%, TTG 1.14%, and GTG 34.90%—corresponds to the high percentage of GTG start codons reported previously in *M. tuberculosis* (Cole et al., 1998) and *M. smegmatis* (Gallien et al., 2009), but with a lower proportion of identified TTG start codons. We identified three ORFs with evidence for multiple initiation, corresponding to the observations of Kelkar et al. (2011; see **Supplementary Table 2**).

Two ORFs were detected where the most upstream evidence was an *N-Met* cleaved peptide and the penultimate position corresponded to another possible start codon. For one of these proteins, the second amino acid in the peptide was also a *Met* located at an ATG codon, thus not allowing for discrimination between N-Met cleavage and downstream initiation. For the second protein, the N-terminal was identified by an *N-Met* cleaved peptide with V as the penultimate amino acid, mapping to a GTG codon at the annotated start site of a Reference proteome sequence. Due to the known initiation of translation with *fMet*, this was concluded to be an instance of *N-Met* cleavage, leading to modification of the gene model to include the adjacent upstream start codon.

FIGURE 2 | Peptide and ORF identification venn diagrams. Panel **(A)** compares peptide identifications from three different database searches. Panel **(B)** shows the genomic open reading frames (ORFs) with at least one unique peptide identified, from the different databases searches. Only peptides identified in more than one replicate were considered. Two identified Reference proteins spanning multiple ORFs (A4ZHT6 and A4ZHR8), not mapping on a one to one basis to a genomic ORF, are included in the diagram as separate ORFs.

Gene Model Validations

We identified 2810 genomic ORFs annotated to Reference proteins with at least one unique peptide—with a median of six unique peptides identified per ORF. Using N-terminal peptide evidence, we validated the TSS for 558 Reference protein gene models. The start codon distribution of validated Reference protein TSSs also reflected the higher proportion of GTG start codons in Mycobacteria—with 65.05% ATG, 0.72% TTG, and 34.23% GTG. Prominent Reference proteins identified include A0R1H7_MYCS2—Fatty acid synthase—with 168 unique peptides and 71.32% sequence coverage, RPOC_MYCS2—DNA-directed RNA polymerase subunit

beta—with 96 unique peptides and 79.5% sequence coverage, Q3L891_MYCS2—Linear gramicidin synthetase subunit D, predicted protein—with 89 unique peptides and 54.98% sequence coverage, Q3L885_MYCS2—Type I modular polyketide synthase, predicted protein—with 88 unique peptides and 38.44% sequence coverage, and A0R617_MYCS2—Polyketide synthase, predicted protein—with 82 unique peptides and 62.14% sequence coverage. The predicted and unreviewed Reference proteome entry A0R0A1_MYCS2 (Glyoxalase/bleomycin resistance protein/dioxygenase) was identified with 99.28% sequence coverage from 13 identified peptides and 152 MS/MS spectra (see **Figure 3** and **Supplementary Table 3**).

Gene model validation was also obtained for the Reference protein MSHD_MYCS2 (Mycothiol acetyltransferase, inferred from homology). This is an interrupted coding sequence (ICDS) identified by Deshayes et al. (2007)—which they confirmed by resequencing to span a sequencing error in the *M. smegmatis* mc^{2}155 genome (*GenBank* accession DQ866865). We present first-time peptide evidence for this protein with eight peptides identified from 50 MS/MS scans using the Reference proteome database, including a peptide spanning the frameshift position (245) in the corrected sequence (see **Figure 4** and **Supplementary Table 4**).

Upstream Gene Model Modifications

Upstream peptide evidence was identified for 81 Reference proteome gene models. Upstream peptide identifications can be grouped into N-terminal (TSS) or fully tryptic upstream peptides—for 39 of the upstream gene model modifications the TSS was detected exactly, while for the 42 sequences with only non-tryptic upstream peptides, the next upstream start-codon in the sequence was located. The putative new gene models thus obtained were searched against the NCBI nr database using protein BLAST, to identify orthologous gene models, or alternative gene models in the same strain. For 58 of the modified gene models, nr BLAST alignment yielded a sequence of the same length, supporting the gene model modification, while 14 nr BLAST alignments in the group without exact TSS identification indicated a TSS further upstream. Only one nr BLAST result in the group with an identified TSS indicated a further upstream TSS than the one identified. The start codon GTG was overrepresented in upstream TSS identifications, making up 38.46%—supporting the observation that ATG may be over predicted as translational start codon (Gallien et al., 2009). The median length of upstream N-terminal extension was five amino acids in the group with upstream TSS identifications. The Reference protein A0R4J1_MYCS2 (Phosphoribosylamine–glycine ligase, inferred from homology) was identified with the N-terminal of three identified peptides extending upstream of the predicted TSS—extending the N-terminal of the protein by 29 amino acids. The modified sequence is identical to the predicted protein I7GFT2_MYCS2 of the same strain (not included in the Reference proteome; see **Figure 5** and **Supplementary Table 5**).

Downstream Gene Model Modifications

Downstream TSS evidence was identified for 24 Reference proteome gene models—with all cases corresponding to

FIGURE 3 | Reference protein gene model validation —A0R0A1. A validation of the gene model of the Reference protein A0R0A1 (Glyoxalase/bleomycin resistance protein/dioxygenase). This is an unreviewed predicted entry associated with dioxygenase activity (GO molecular function), of length of 138. Panel **(A)** illustrates validation of the gene model at the peptide level visualized on the Ensembl genome browser, with a sequence coverage of 99.28% from 13 peptides and a total of 152 MS/MS spectra identified in the Reference proteome database search. Panel **(B)** shows the identified translation start site (TSS) peptide, illustrating the cleavage of initiator methionine (N-Met) by methionine aminopeptidase (MAP) with Val as the penultimate amino acid. Panel **(C)** shows a representative MS/MS spectrum for the identified TSS peptide.

FIGURE 4 | Reference protein frameshift validation —MSHD. Shows the identification of A4ZHT6 (Mycothiol acetyltransferase), a Reference protein inferred from homology. **(A)** This sequence corresponds to an interrupted coding sequence (ICDS) identified by Deshayes et al. (2007), which they confirmed by resequencing to span a sequencing error in the *M. smegmatis* mc²155 genome (GenBank accession DQ866865). We present peptide evidence for this protein with eight peptides identified from 50 MS/MS scans after searching the Reference proteome database, as well as identifying a peptide spanning the frameshift position (245) in the corrected sequence—highlighted in red. **(B)** A representative spectrum of the highlighted peptide, which was identified with four MS/MS scans, supporting the sequence correction at the peptide level.

FIGURE 5 | Upstream gene model modification —A0R4J1. Shows evidence for gene model modification of the Reference protein A0R4J1 (Phosphoribosylamine–glycine ligase) by N-terminal extension, with the N-terminal of three identified peptides extending upstream of the predicted TSS, extending the N-terminal of the protein by 29 amino acids. The modified sequence is identical to the predicted protein I7GFT2 for the same strain (not included in the Reference proteome). Panel **(A)** shows the peptide RDPEVEGLAVAPGNAGTSSIADQYDVDVTSGEAVVK—highlighted in red—extending upstream of the annotated TSS along with two other peptides. Panel **(B)** shows the sequence of the modified gene model of A0R4J1, with the annotated TSS position indicated by the blue arrow. Panel **(C)** shows a representative MS/MS spectrum of RDPEVEGLAVAPGNAGTSSIADQYDVDVTSGEAVVK, identified from eight MS/MS spectra in the translated six frame database search.

an alternative TSS prediction in the *GeneMarkS* database. The predicted gene model A0QWY3_MYCS2 (Quinone oxidoreductase) was shortened by 16 amino acids, with the identification of a TSS peptide MHAIEVAETGGPEVLNYIER PEPSPGPGEVLIK with a non-tryptic N-terminal downstream of the annotated TSS. The downstream TSS peptide corresponded to the N-terminal of the *GeneMarkS* predicted sequence for this ORF, thus allowing this semi-tryptic TSS peptide to be included in the *GeneMarkS* database search space. The modified sequence is identical to I7G8G0_MYCS2, predicted for the same strain but not included in the Reference proteome. The downstream TSS peptide was identified from 34 MS/MS scans (see **Figure 6** and **Supplementary Table 6**).

Novel ORF Identifications

Due to the high prevalence of gene prediction errors, inclusion in the UniProt Reference proteome was used as a benchmark

for annotation status in this study. Thus, all identified ORFs not annotated to a Reference proteome entry were considered novel identifications. Peptide evidence mapping to ORFs not annotated to a Reference proteome entry was identified for 72 ORFs, of which 44 were identified with two or more peptides. ORFs with one or more identified peptides that occurred adjacent to another novel ORF identification, were examined as possible evidence for Interrupted CoDing Sequences (ICDSs). Nineteen novel ORFs that were identified with a single peptide, and were supported by either protein BLAST alignment or a previously identified transcript overlapping on the genome on the same strand, are presented as lower ranking evidence for genome annotation. Nine ORFs with only one identified peptide and no supporting evidence, were excluded from the further analysis due to the high likelihood of these identifications being erroneous—leading to a total of 63 novel ORF identifications (see **Supplementary Table 7**).

FIGURE 6 | Downstream gene model modification —A0QWY3. Illustrates the downstream modification of an annotated predicted gene model A0QWY3 by 16 amino acids, with the identification of a TSS peptide MHAIEVAETGGPEVLNYIERPEPSPGPGEVLIK with a non-tryptic N-terminal downstream of the annotated TSS. In this case, the downstream TSS peptide corresponded to the N-terminal of the GeneMarkS predicted sequence for this ORF, thus allowing this semi-tryptic TSS peptide to be included in the GeneMarkS database search space. The modified sequence is identical to I7G8G0, predicted for the same strain but not included in the Reference proteome. Panel **(A)** visualizes the identified peptides using the Ensembl genome browser, with the downstream TSS peptide highlighted in red. The downstream gene model modification of A0QWY3 corresponds with the simultaneous upstream gene model modification of A0QWY4 on the opposite strand, supporting both reannotations. Panel **(B)** illustrates the modified gene model, with the TSS peptide highlighted in red. Panel **(C)** illustrates a representative MS/MS spectrum for the downstream TSS peptide, which was identified from 34 MS/MS scans using the *GeneMarkS* database search.

Validation of Previously Identification Interrupted Coding Sequences (ICDSs)

Twelve non-Reference proteome ORFs identified at the peptide level corresponded to six interrupted coding sequences previously reported by Deshayes et al. (2007), with GenBank accessions DQ866867, DQ866856, DQ866859, DQ866863, DQ866858, and DQ866873—see **Supplementary Figures 1A–E,G,** respectively, for Ensembl genome browser visualizations, and **Supplementary Table 8** rows 2–11 and 14–15 for detailed information. The above authors had shown by resequencing that these frameshifts corresponded to genome sequencing errors, and they also reported peptide-level evidence for two of these sequences using nano-LC/MS/MS analysis (DQ866873 and DQ866856). Thus, we were able to identify four of these ICDSs with first-time peptide evidence, and validate two ICDSs previously identified at the peptide level.

Novel ICDSs

Four likely novel ICDS sequences were identified with peptide evidence spanning either side of a possible genomic frameshift region from eight non-Reference proteome ORFs. In an interesting case, three novel peptides were identified from an ORF with the closest nr protein BLAST alignment to a predicted protein in *Mycobacterium goodii*—A0A0K0X632_9MYCO (Peptidase M75). A non-Reference proteome predicted ORF I7FS93_MYCS2 was identified partially overlapping and upstream of this ORF with peptide evidence. This upstream

ORF also aligned with high confidence to the same sequence in *M. goodii*. To our knowledge an ICDS has not previously been identified at this position, but peptide evidence from two overlapping reading frames and alignment to an orthologous sequence of a related species, supports the identification of a novel ICDS at this site—although the existence of separate adjacent protein coding genes cannot entirely be excluded (see **Supplementary Figures 1F** and **Supplementary Table 8** rows 12–13).

Another novel ORF with only one identified peptide— TAILDAAAQLIAER—was identified upstream and partially overlapping the predicted ORF I7G4B8_MYCS2 that was identified with three peptides (see **Supplementary Figure 1J** and **Supplementary Table 8** rows 20–21). Both ORFs aligned with high confidence to the predicted protein L8F9F0_MYCSM of *M. smegmatis* MKD8—the genome sequence of which has recently been announced (Gray et al., 2013). Thus, orthologous sequence evidence combined with evidence at the peptide level strongly supports the existence of an ICDS in this position, and further investigation is needed to ascertain whether this is an occurrence of authentic mutation or sequencing error, or in fact two separate protein-coding genes. The identification of two more ICDS sequences with peptide evidence on either side of a possible frameshift position was also facilitated by alignment to orthologous sequences submitted by Gray et al. (2013) as a result of their genome sequencing efforts of *M. smegmatis* MKD8— emphasizing the iterative nature of genome annotation as new data becomes available (see **Supplementary Figures 1H,I**, and **Supplementary Table 8** rows 16–19).

Novel ORFs Identified with Two or More Peptides

We identified 44 non-Reference proteome ORFs with two or more peptides (with a median of five identified unique peptides per ORF in this group). Protein BLAST alignments were obtained for 43 of these, and the gi accession numbers thus obtained were mapped to their corresponding entries in UniProt using the UniProt "Retrieve/ID mapping" tool[8]. Of the 39 sequences that were mapped with this tool, 31 were predicted and eight were inferred from homology. Three of these sequences were annotated to *M. smegmatis* MKD8, one to *M. thermoresistibile* strain ATCC 19527, and four to *M. smegmatis* non-specifically. One identified ORF alignment—discussed above as part of an identified ICDS—aligned to a sequence annotated to *M. goodii*. The remaining 30 identified ORFs aligned to Non-Reference proteome sequences from *M. smegmatis* mc^{2}155—with 26 predicted and four inferred from homology. An interesting novel ORF was identified with two peptides from eight MS/MS spectra mapping to an intergenic region with a genomic position from 6,434,313 to 6,434,801 on the reverse strand. The sequence yielded a protein nr BLAST alignment to G7CE94_MYCTH (Lipoprotein LppV), a 182 amino acids long predicted protein annotated to *M. thermoresistibile*, with an *E*-value of 2.13747e-44 (see **Figure 7** and **Supplementary Table 7**).

---

[8]*Retrieve/ID Mapping (UniProt)*. http://www.uniprot.org/uploadlists/

Database Comparison

The PEP scores of the identified peptides were obtained from the *MaxQuant* peptides.txt file—which contains the identified peptides with the PEP score calculated using the peptide length and the *Andromeda* score for the best associated MS/MS spectrum. Density plots of the distribution of PEP scores for each database revealed non-gaussian PEP score distributions. Analysis of variance of novel, annotated and reverse peptide PEP scores for each database was performed using the Kruskal–Wallis non-parametric analysis of variance test followed by *post-hoc* analysis with a two-sided Dunn's test and Bonferroni correction (see **Supplementary Data Sheet 5**). A statistically significant difference between group means (annotated, novel and reverse peptides) for each of the three databases was found (*p*-values < 2.2e-16). *Post-hoc* analysis indicated a significant difference between reverse and annotated peptide group PEP scores for both the *GeneMarkS*, six frame database (adjusted *p*-value 6.68e-21), six frame database (adjusted *p*-value 3.67e-19), and Reference proteome (adjusted *p*-value 1.05e-31) peptide results, while the assignment of statistically significant differences between annotated and novel peptide group PEP scores varied between the six frame database (adjusted *p*-value 1.89e-06), and *GeneMarkS* database (adjusted *p*-value 8.75e-01), although the values were much higher than the annotated-reverse comparisons. Further, the comparison between novel and reverse peptide group PEP scores indicated a significant difference for both the six frame database (adjusted *p*-value 9.84e-13) and *GeneMarkS* database (adjusted *p*-value 5.50e-17) comparisons—with much lower adjusted *p*-values than those of the annotated-novel group comparisons of both databases. PEP score analysis was also performed on the combined PSMs for each database obtained from the msms.txt file, supporting the below analysis (see **Supplementary Table 9, Supplementary Figure 2**, and **Supplementary Data Sheet 6**.
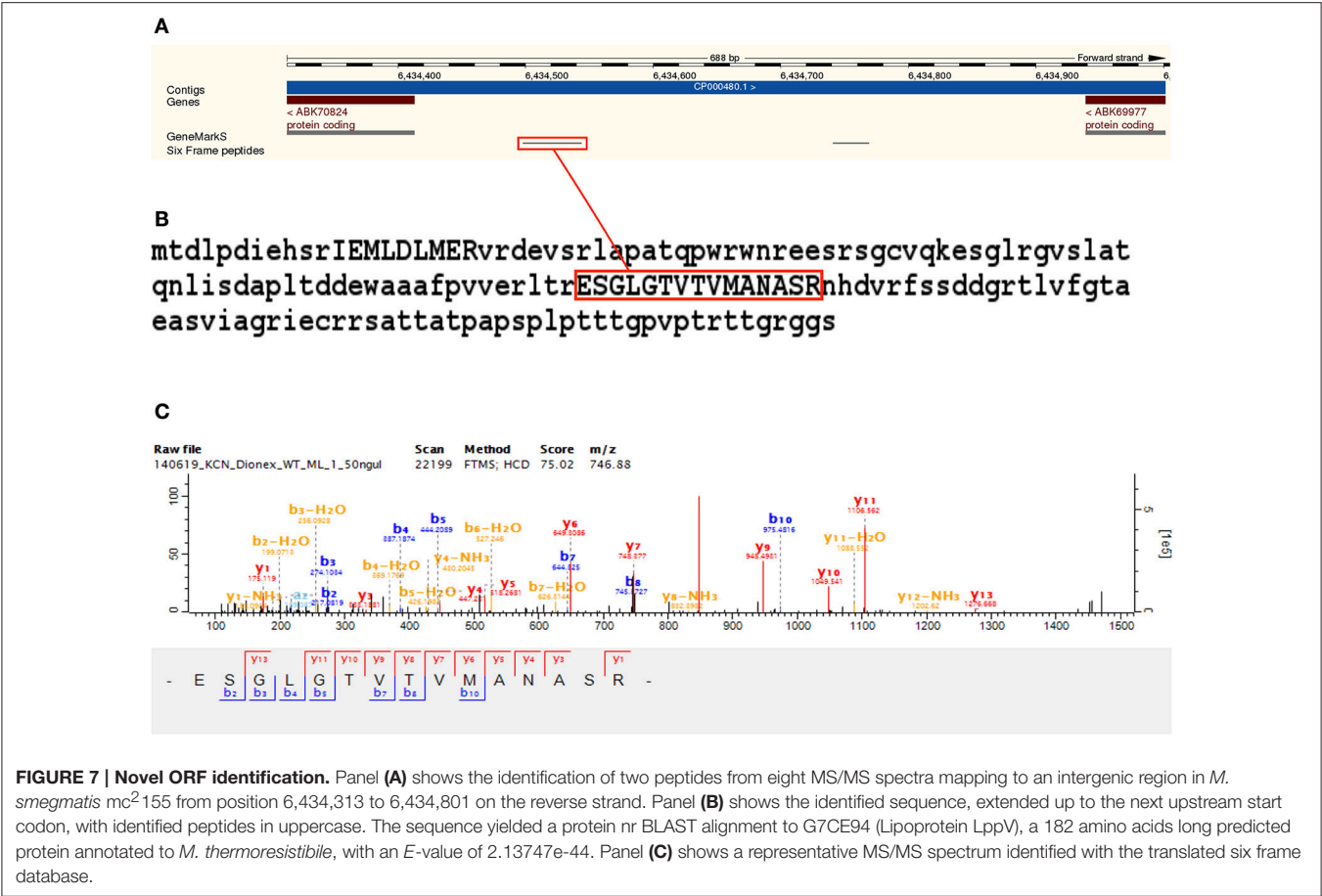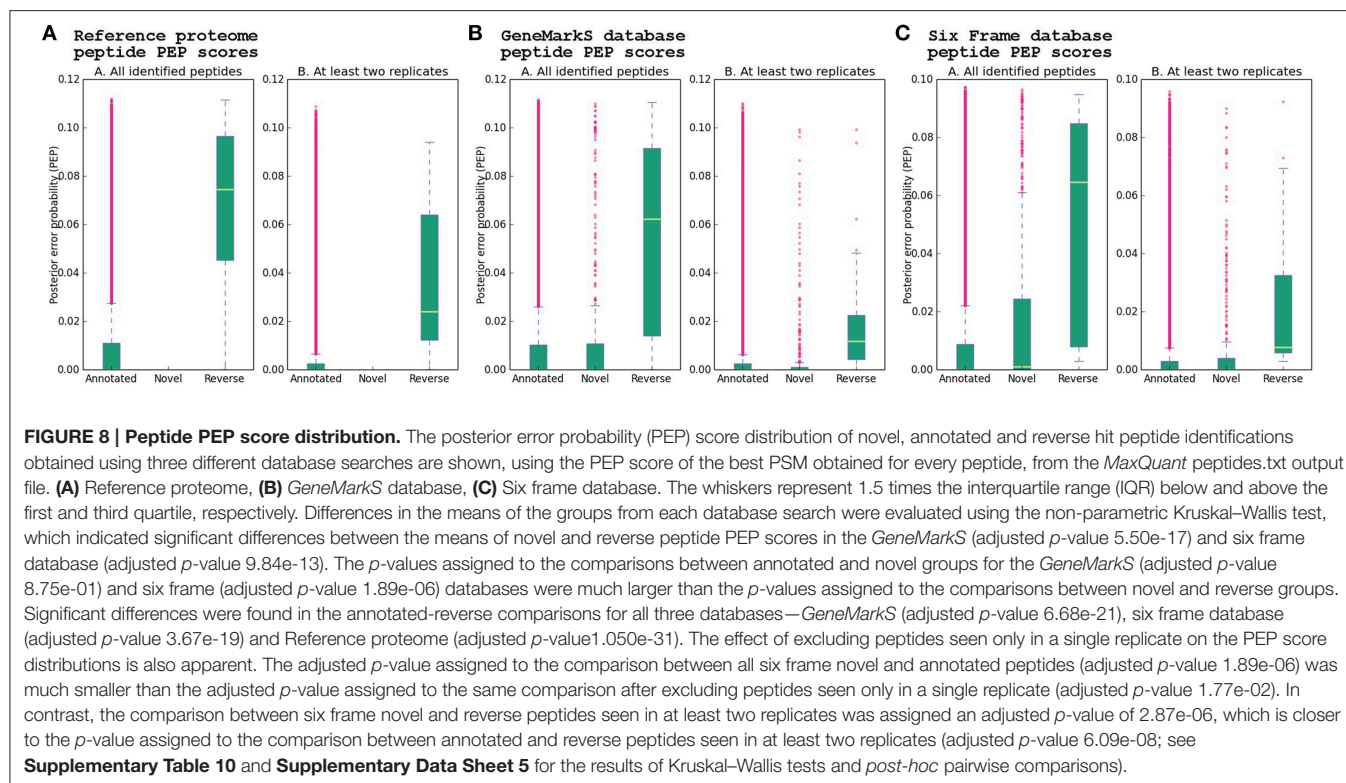
Reference Proteome Database

Using the Reference proteome database, 27,720 annotated peptides were identified, with a median PEP score of 1.32E-04. Further, 72 reverse hit peptides were reported by *MaxQuant*, with a median PEP score of 7.46E-02. Kruskal–Wallis and Dunn's test *post-hoc* analysis showed a statistically significant difference between the two groups (adjusted *p*-value 1.05e-31 before excluding peptides seen in only one replicate). After selecting peptides identified in at least two replicates, 2788 database sequences were identified with at least one unique peptide. For the group of Reference proteins with TSS validation and single genomic coordinates, a start site distribution of ATG 353 (64.53%), TTG 4 (0.73%), and GTG 190 (34.73%) was determined (see **Figures 2, 8A, Table 1**, and **Supplementary Table 10**).

GeneMarkS Database

The *GeneMarkS* database allowed for the identification of 27,168 annotated and 407 novel peptides, with a median peptide PEP score of 1.16E-04 and 2.22E-04 respectively, and 59 reverse hit peptides with a median peptide PEP score of 6.24E-02. The PEP score distribution of annotated and novel peptides were not significantly different using Kruskal–Wallis and *post-hoc* tests (adjusted *p*-value 8.75e-01), and from the boxplot

FIGURE 7 | Novel ORF identification. Panel **(A)** shows the identification of two peptides from eight MS/MS spectra mapping to an intergenic region in *M. smegmatis* mc²155 from position 6,434,313 to 6,434,801 on the reverse strand. Panel **(B)** shows the identified sequence, extended up to the next upstream start codon, with identified peptides in uppercase. The sequence yielded a protein nr BLAST alignment to G7CE94 (Lipoprotein LppV), a 182 amino acids long predicted protein annotated to *M. thermoresistibile*, with an *E*-value of 2.13747e-44. Panel **(C)** shows a representative MS/MS spectrum identified with the translated six frame database.

TABLE 1 | Database identifications.

	Reference proteome	Six frame database	*GeneMarkS* database	Combined
Reference protein ORF identifications	2788	2753	2773	2810
Reference TSS validations	554	282	439	558
All upstream gene model modifications	0	70	42	81
Upstream TSS identifications	0	29	28	39
Downstream TSS identifications	0	0	24	24
Novel ORF identifications	0	63	42	63

A table comparing Reference protein ORF identifications, TSS validations, upstream gene model modifications, upstream TSS identifications, downstream TSS identifications and novel ORF identifications using three different databases.

visualizations appear similar to the PEP distribution of annotated peptides obtained from the Reference proteome database. A significant difference was found between annotated and reverse peptide PEP score distributions (adjusted *p*-value 6.68e-21). A comparison between novel and reverse peptide PEP score distributions revealed a significant difference (adjusted *p*-value 5.50e-17), and the number of reverse hit peptides identified was markedly lower than the number of novel peptides identified. Thus, it is very likely that most novel identifications are true positive identifications (see **Figure 8B**). After excluding peptides only identified in a single sample, 2815 *GeneMarkS* sequences

were identified with at least one unique peptide. Of these, 42 were not annotated to a Reference protein. Further, in the group of 558 Reference protein TSS validations, 442 sequences had a corresponding entry in the *GeneMarkS* database with the correct TSS prediction (of which 439 were identified)—while only one did not have a corresponding entry, and 115 had a corresponding entry with an alternative TSS assignment. Thus, in the group of TSS validations with a corresponding *GeneMarkS* entry, *GeneMarkS* had correctly predicted the TSS in 79.35% cases. Further in the set of all identified genomic ORFs annotated to a Reference protein (2810), GeneMarkS had

FIGURE 8 | Peptide PEP score distribution. The posterior error probability (PEP) score distribution of novel, annotated and reverse hit peptide identifications obtained using three different database searches is shown, using the PEP score of the best PSM obtained for every peptide, from the *MaxQuant* peptides.txt output file. **(A)** Reference proteome, **(B)** *GeneMarkS* database, **(C)** Six frame database. The whiskers represent 1.5 times the interquartile range (IQR) below and above the first and third quartile, respectively. Differences in the means of the groups from each database search were evaluated using the non-parametric Kruskal–Wallis test, which indicated significant differences between the means of novel and reverse peptide PEP scores in the *GeneMarkS* (adjusted *p*-value 5.50e-17) and six frame database (adjusted *p*-value 9.84e-13). The *p*-values assigned to the comparisons between annotated and novel groups for the *GeneMarkS* (adjusted *p*-value 8.75e-01) and six frame (adjusted *p*-value 1.89e-06) databases were much larger than the *p*-values assigned to the comparisons between novel and reverse groups. Significant differences were found in the annotated-reverse comparisons for all three databases—*GeneMarkS* (adjusted *p*-value 6.68e-21), six frame database (adjusted *p*-value 3.67e-19) and Reference proteome (adjusted *p*-value1.050e-31). The effect of excluding peptides seen only in a single replicate on the PEP score distributions is also apparent. The adjusted *p*-value assigned to the comparison between all six frame novel and annotated peptides (adjusted *p*-value 1.89e-06) was much smaller than the adjusted *p*-value assigned to the same comparison after excluding peptides seen only in a single replicate (adjusted *p*-value 1.77e-02). In contrast, the comparison between six frame novel and reverse peptides seen in at least two replicates was assigned an adjusted *p*-value of 2.87e-06, which is closer to the *p*-value assigned to the comparison between annotated and reverse peptides seen in at least two replicates (adjusted *p*-value 6.09e-08; see **Supplementary Table 10** and **Supplementary Data Sheet 5** for the results of Kruskal–Wallis tests and *post-hoc* pairwise comparisons).

predicted the ORF as protein-coding in 99.68% of cases. Of the 81 upstream gene model modifications, 80 had a corresponding *GeneMarkS* sequence identified, of which 42 of the modifications were supported by upstream peptide identifications from the *GeneMarkS* database (with 28 TSS identifications). All of the 24 downstream gene model modifications were identified with TSS evidence from the *GeneMarkS* database, with a mean N-terminal shortening of ∼10 amino acids in this group (see **Figure 2**, **Table 1**, and **Supplementary Table 10**).

Six Frame Database

The translated six frame database allowed for the identification of 25,427 annotated and 553 novel peptides, with a median peptide PEP score of 1.07E-04 and 1.04E-03 respectively, and 48 reverse hit peptides with a median PEP score of 6.47E-02. Using Kruskal–Wallis and Dunn's test *post-hoc* analysis, a significant difference between the group PEP scores of annotated and reverse (adjusted *p*-value 3.67e-19), annotated and novel (adjusted *p*-value 1.89e-06), and novel and reverse (adjusted *p*-value 9.84e-13) peptide identifications was found. After selecting peptides seen in at least two replicates, a much larger although still statistically significant adjusted *p*-value of 1.77e-02 for the comparison between novel and annotated groups was obtained—possibly indicating a relatively higher proportion of false positive identifications in the set of novel peptide identifications than in the annotated set. *Post-hoc* comparison between novel and reverse peptide PEP score distributions revealed adjusted *p*-values of 9.84e-13 and 2.87e-06 before and after excluding peptides seen in only one replicate respectively—much lower than the *p*-values obtained from the

comparison between novel and annotated peptide groups—indicating a much greater difference between the distribution of novel and reverse peptide PEP scores than between novel and annotated peptides, see **Figure 8C**. After selecting peptides seen in at least two replicates, 2825 ORFs were identified using the six frame database. Seventy-two ORFs identified using the six frame database mapped to non-Reference proteome ORFs. Of the 558 Reference protein sequences with TSS validations, a corresponding six frame sequence was identified in 544 cases. In this group, 299—or 54.96%—of the six frame sequences (translated from the most upstream start codon in the ORF) correctly assigned the TSS. On average, the six frame database overestimated the correct TSS by ∼32 amino acids. Further, in the group of 81 upstream gene model modifications, 80 cases had a corresponding six frame sequence identified. In this group, 87.5% of the upstream gene model modifications were supported by the identification of upstream peptide evidence using the six frame database. Of the 38 sequences with identified TSS sites in this group, 29 were correctly assigned in the six frame database (see **Figure 2**, **Table 1**, and **SupplementaryTable 10**).

DISCUSSION

Validation of genomic information by proteomic data has gained momentum over the past few years. The improvement in genome annotation can give rise to a better understanding of the biology of pathogenic organisms and ultimately, new strategies for disease prevention and treatment. Mass-spectrometry data

has proven invaluable in terms of providing evidence of the translation of genes expressed under differing conditions.

By mapping identified proteins from the different databases to their corresponding ORFs in the genome, 2887 ORFs were identified at the peptide level. We showed that the identified start codon distribution of all identified ORFs with TSS identifications corresponded to previously reported findings in Mycobacteria, with a relatively higher percentage of GTG, although a relatively lower proportion of TTG start codons was identified. The identified penultimate amino acids of N-Met cleaved peptides corresponded to previously published findings, although the preponderance of *Thr* was notable.

We identified 2810 Reference proteins at the peptide level, and using N-terminal peptide evidence we validated the TSS for 558 of these. Further, 81 gene model modifications were indicated by the identification of peptides mapping upstream, and 24 by N-terminal peptides mapping downstream, of the annotated TSS of a Reference protein.

We provide experimental evidence for 63 novel ORFs not previously verified at the protein level. A total of 44 ORFs were identified with two or more peptides—of which 30 aligned to predicted or inferred non-Reference proteome sequences of *M. smegmatis* mc^{2}155. A single ORF was identified with two peptides but did not yield any protein BLAST alignments. Further, 19 novel ORFs were identified with a single peptide but were supported by either protein BLAST alignment or a previously identified transcript overlapping the ORF on the genome on the same strand, and are presented as lower-ranking evidence for annotation. We also identified six previously reported interrupted coding sequences caused by sequencing errors, with peptides identified on either side of a frameshift position—four of which do not appear to have been previously identified at the peptide level. Further, we identified evidence for four novel ICDSs, three of which were supported by alignment to orthologous sequence evidence from a newly sequenced strain of *M. smegmatis,* emphasizing the importance of continuous review of genome annotation as new information becomes available.

In this study, we analyzed LC/MS/MS data with three different sequence databases to improve the genome annotation of *M. smegmatis* mc^{2}155; two genomic—a six frame translation and a *GeneMarkS* gene prediction database—and the UniProt Reference proteome[9]. In both genomic databases, the PEP score distribution of novel peptides was much closer to that of annotated peptides than reverse sequence hits, and almost identical in the *GeneMarkS* database—indicating that most novel peptide identifications are likely to be true positive identifications. A higher number of total peptide identifications was attained using *GeneMarkS* than the six frame database—closely approaching the number attained using the Reference proteome. We report a high percentage of accurate TSS predictions using *GeneMarkS*—79.35% in the group of Reference protein TSS validations where a corresponding *GeneMarkS* entry was predicted. This indicates that proteogenomic database

generation with targeted ORF and TSS predictions using *de novo* gene prediction tools such as *GeneMarkS* can fruitfully decrease the database search space, increasing the sensitivity of peptide identifications. However, database compaction is a compromise between reducing spurious possibilities, and minimizing the exclusion of non-spurious entries from the search space. ORF and TSS assignments will depend on the particular gene prediction tool used, and will not be detected if incorrectly assigned at the database generation phase. By including a six frame translation database in the pipeline, ORFs missed by the gene prediction algorithm, as well as upstream peptide evidence and TSS site identifications not included in the gene prediction database, may be identified.

Considering the PEP score distributions of novel and annotated peptides, and the much smaller number of reverse sequence hits than novel peptide identifications, it is evident that the proteome of *M. smegmatis* mc^{2}155 is not yet fully characterized and more work needs to be done to identify novel protein coding genes before a complete genome annotation status is attained. We hope that the evidence we present here will lead to the addition of new sequences to the Reference proteome of this strain, supporting the downstream functional characterization of these proteins, thus leading to a better understanding of the biology of an important model organism of the infectious *M. tuberculosis*. The gene model modifications we present for many Reference proteome sequences—many of which have not yet been identified at the protein level—facilitates improved functional and structural characterization of these proteins, allowing for accurate conclusions to be drawn by downstream comparative proteomics analyses.

AUTHOR CONTRIBUTIONS

MP, MSc student, writing, Bioinformatics component; KN, writing, Proteomics component; JA, technical support on Bioinformatics component; AN, technical support on Proteomics component; SG, technical support on Proteomics component; NS, writing, technical support on Proteomics component; JB, corresponding author, technical supervision, and assistance, Proteomics component; NM, corresponding author, technical supervision and assistance, proof reading, Bioinformatics component.

ACKNOWLEDGMENTS

---

[9]*Mycobacterium smegmatis Reference Proteome. UniProt Proteomes.* http://www.uniprot.org/proteomes/UP000000757 (Accessed August 9, 2015).

number 86934, including an MSc bursary to MP. Computations were performed using facilities provided by the University of Cape Town's ICTS High Performance Computing team: http://hpc.uct.ac.za.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2016.00427

Supplementary Figure 1 | Novel ORF ICDSs. Shows peptide evidence for six previously identified ICDSs—of which only two have previously been reported at the protein level—and four novel ICDS sequences. **(A)** DQ866867: A predicted protein spanning a sequencing error reported by Deshayes et al. (2007), showing peptides on either side of a frame shift position. **(B)** DQ866856: A protein spanning a sequencing error reported by Deshayes et al. (2007), which they also identified at the peptide level, showing peptides on either side of the frameshift position. **(C)** DQ866859: Peptide evidence for a predicted protein spanning a sequencing error reported by Deshayes et al. (2007). **(D)** DQ866863: Peptide evidence for a predicted protein spanning a sequencing error reported by Deshayes et al. (2007). **(E)** DQ866858: Peptide evidence for a predicted protein spanning a sequencing error reported by Deshayes et al. (2007), showing an upstream peptide identified in an overlapping ORF (highlighted in red). **(F)** Novel ICDS: An upstream ORF first predicted by Deshayes et al. in 2007 (I7FS93_MYCS2), with downstream peptide evidence in an overlapping novel ORF (three peptides). Both ORFs align to A0A0K0X632_9MYCO (peptidase M75 of *M. goodii*), with very low *E*-values (0.0 and 5.03e-09, respectively). The downstream peptides are highlighted in red. **(G)** DQ866873: an ICDS spanning a sequencing error, and also detected at the protein level, reported by Deshayes et al. (2007). Five upstream peptides in an overlapping novel ORF are highlighted in red. **(H)** Novel ICDS: Two adjacent ORFs in alternate reading frames, both aligned to the protein L8F5X0_MYCSM—predicted by Gray et al. (2013) from the genome of *M. smegmatis* MKD8. **(I)** Novel ICDS: Two adjacent ORFs in alternate reading frames, both aligned to the protein L8FGK0_MYCSM—predicted by Gray et al. (2013) from the genome of *M. smegmatis* MKD8. **(J)** Novel ICDS: Two adjacent ORFs in alternate reading frames, both aligned to the protein L8F9F0_MYCSM—predicted by Gray et al. (2013) from the genome of *M. smegmatis* MKD8. A single upstream peptide in an alternate frame is highlighted in red.

Supplementary Figure 2 | PSM PEP score distribution. Shows the PEP score distributions of all annotated, novel and reverse sequence PSMs from the different databases. The whiskers represent 1.5 times the interquartile range (IQR) below and above the first and third quartile, respectively. In the *GeneMarkS* and six frame

database groups the PEP score distribution of novel PSMs is noticeably closer to the group of annotated PSMs than reverse hits, indicating the high confidence of novel PSM identifications. PEP score distributions are further improved by excluding the PSMs of peptides that were only identified in a single replicate. Using the Kruskal–Wallis test followed by *post-hoc* analysis, significant differences were found for the comparisons between all novel and reverse PSM PEP scores for both the *GeneMarkS* (adjusted *p*-value 2.20e-42) and six frame databases (adjusted *p*-value 8.01e-36), respectively. Similarly, significant differences were found between all annotated and all reverse group PSM PEP scores for the Reference proteome (adjusted *p*-value 1.71e-70), *GeneMarkS* (adjusted *p*-value 2.00e-48) and six frame database (adjusted *p*-value 1.83e-44). The *p*-values assigned to the comparisons between all annotated and all novel PSM PEP scores were much higher for both the *GeneMarkS* (adjusted *p*-value 4.26e-02) and six frame database (adjusted *p*-value 6.83e-09; see **Supplementary Data Sheet 6**) for the results of Kruskal–Wallis tests and *post-hoc* pairwise comparisons.

Supplementary Table 1 | Combined ORF identifications.

Supplementary Table 2 | Multiple initiation.

Supplementary Table 3 | Validated TSSs.

Supplementary Table 4 | Frameshift validation.

Supplementary Table 5 | Upstream start annotations.

Supplementary Table 6 | Downstream start annotations.

Supplementary Table 7 | Novel ORF annotations.

Supplementary Table 8 | Novel ORF ICDSs.

Supplementary Table 9 | PSM PEP scores.

Supplementary Table 10 | Peptide PEP scores.

Data Sheet 1 | GeneMarkS sequence database (GFF format).

Data Sheet 2 | Six frame database peptide identifications (GFF format).

Data Sheet 3 | GeneMarkS database peptide identifications (GFF format).

Data Sheet 4 | Reference proteome peptide identifications (GFF format).

Data Sheet 5 | Analysis of best peptide PSM PEP scores.

Data Sheet 6 | Analysis of all peptide PSM PEP scores.

REFERENCES

Besemer, J., Lomsadze, A., and Borodovsky, M. (2001). GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* 29, 2607–2618. doi: 10.1093/nar/29.12.2607

Borchert, N., Dieterich, C., Krug, K., Schütz, W., Jung, S., Nordheim, A., et al. (2010). Proteogenomics of *Pristionchus pacificus* reveals distinct proteome structure of nematode models. *Genome Res.* 20, 837–846. doi: 10.1101/gr.103119.109

Camus, J.-C., Pryor, M. J., Médigue, C., and Cole, S. T. (2002). Re-annotation of the genome sequence of *Mycobacterium Tuberculosis* H37Rv. *Microbiology* 148, 2967–2973. doi: 10.1099/00221287-148-10-2967

Castellana, N., and Bafna, V. (2010). Proteogenomics to discover the full coding content of genomes: a computational perspective. *J. Proteomics* 73, 2124–2135. doi: 10.1016/j.jprot.2010.06.007

Cock, P. J., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., et al. (2009). Biopython: freely available python tools for computational molecular biology and bioinformatics. *Bioinformatics (Oxford, England)* 25, 1422–1423. doi: 10.1093/bioinformatics/btp163

Cole, S. T. R., Brosch, J., Parkhill, T., Garnier, C., Churcher, D., Harris, S. V., et al. (1998). Deciphering the biology of *Mycobacterium Tuberculosis* from the complete genome sequence. *Nature* 393, 537–544. doi: 10.1038/31159

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367–1372.

Dale, J. W. (1995). Mobile genetic elements in mycobacteria. *Eur. Resp. J. Suppl.* 20, 633s–648s.

Deshayes, C., Perrodou, E., Gallien, S., Euphrasie, D., Schaeffer, C., Van Dorsselaer, A., et al. (2007). Interrupted coding sequences in *Mycobacterium smegmatis*: authentic mutations or sequencing errors? *Genome Biol.* 8:R20. doi: 10.1186/gb-2007-8-2-r20

Dunn, O. J. (1964). Multiple comparisons using rank sums. *Technometrics* 6, 241–252. doi: 10.1080/00401706.1964.10490181

Fleischmann, R. D., Dodson, R. J., Haft, D. H., Merkel, J. S., Nelson, W. C., and Fraser, C. M. (2006). "CP000480; SV 1; Circular; Genomic DNA; STD; PRO; 6988209 BP.," *ENA*. Available online at: http://www.ebi.ac.uk/ena/data/view/CP000480&display=text

Frottin, F., Martinez, A., Peynot, P., Mitra, S., Holz, R. C., Giglione, C., et al. (2006). The proteomics of N-terminal methionine cleavage. *Mol. Cell. Proteomics* 5, 2336–2349. doi: 10.1074/mcp.M600225-MCP200

Gallien, S., Perrodou, E., Carapito, C., Deshayes, C., Reyrat, J. M., Van Dorsselaer, A., et al. (2009). Ortho-proteogenomics: multiple proteomes investigation through orthology and a New MS-based protocol. *Genome Res.* 19, 128–135. doi: 10.1101/gr.081901.108

Gray, T. A., Palumbo, M. J., and Derbyshire, K. M. (2013). Draft genome sequence of MKD8, a conjugal recipient *Mycobacterium smegmatis* strain. *Genome Announc.* 1, e0014813. doi: 10.1128/genomeA.001 48-13

Hunter, J. D. (2007). Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* 9, 90–95. doi: 10.1109/MCSE.2007.55

Käll, L., Storey, J. D., MacCoss, M. J., and Noble, W. S. (2008). Posterior error probabilities and false discovery rates: two sides of the same coin. *J. Proteome Res.* 7, 40–44. doi: 10.1021/pr700739d

Kelkar, D. S., Kumar, D., Kumar, P., Balakrishnan, L., Muthusamy, B., Yadav, A. K., et al. (2011). Proteogenomic analysis of *Mycobacterium Tuberculosis* by high resolution mass spectrometry. *Mol. Cell. Proteomics* 10, 3–4, 7. doi: 10.1074/mcp.M111.011627

Krug, K., Carpy, A., Behrends, G., Matic, K., Soares, N.C., and Macek, B. (2013). Deep coverage of the *Escherichia coli* proteome enables the assessment of false discovery rates in simple proteogenomic experiments. *Mol. Cell. Proteomics* 12, 3420–3430. doi: 10.1074/mcp.M113.029165

Krug, K., Nahnsen, S., and Macek, B. (2011). Mass spectrometry at the interface of proteomics and genomics. *Mol. Biosyst.* 7, 284–291. doi: 10.1039/C0M B00168F

Kucharova, V., and Wiker, H. G. (2014). Proteogenomics in microbiology: taking the right turn at the junction of genomics and proteomics. *Proteomics* 14, 2360–2675. doi: 10.1002/pmic.201400168

Link, A. J., Robison, K., and Church, G. M. (1997). Comparing the predicted and observed properties of proteins encoded in the genome of *Escherichia coli* K-12. *Electrophoresis* 18, 1259–1313. doi: 10.1002/elps.1150180807

Nakedi, K. C., Nel, A. J., Garnett, S., Blackburn, J. M., and Soares, N. C. (2015). Comparative Ser/Thr/Tyr phosphoproteomics between two mycobacterial species: the fast growing *Mycobacterium smegmatis* and the slow growing *Mycobacterium bovis* BCG. *Front. Microbiol.* 6:237. doi: 10.3389/fmicb.2015.00237

Ogle, D. H. (2016). *FSA: Fisheries Stock Analysis. R package version 0.8.5.*

Oliveros, J. C. (2016). "Venny 2.1.0.," *Venny. An Interactive Tool for Comparing Lists with Venn's Diagrams. (2007-2015).* Available online at: http://bioinfogp.cnb.csic.es/tools/venny/ (Accessed February 15, 2016).

Perrodou, E., Deshayes, C., Muller, J., Schaeffer, C., Van Dorsselaer, A., Ripp, R., et al. (2006). ICDS database: interrupted CoDing sequences in prokaryotic genomes. *Nucleic Acids Research* 34(suppl 1), D338–D343. doi:10.1093/nar/gkj060

Pettersen, V. K., Steinsland, H., and Wiker, H. G. (2015). Improving genome annotation of enterotoxigenic *Escherichia coli* TW10598 by a label-free quantitative MS/MS approach. *Proteomics* 15, 3826–3834. doi: 10.1002/pmic.201500278

R Core Team (2015). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, (Vienna, Austria) Available online at: https://www.R-project.org/

Ramagli, L. S. (1999). Quantifying protein in 2-D PAGE solubilization buffers. *Methods Mol. Biol.* 112, 99–103.

Rappsilber, J., Ishihama, Y., and Mann, M. (2003). Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* 75, 663–670. doi: 10.1021/ac026117i

Renuse, S., Chaerkady, R., and Pandey, A. (2011). Proteogenomics. *Proteomics* 11, 620–630. doi: 10.1002/pmic.201000615

Repaske, R. (1956). Lysis of gram-negative bacteria by lysozyme. *Biochim. Biophys. Acta* 22, 189–191. doi: 10.1016/0006-3002(56)90240-2

Rezwan, M., Lanéelle, M. A., Sander, P., and Daffé, M. (2007). Breaking down the wall: fractionation of mycobacteria. *J. Microbiol. Methods* 68, 32–39. doi: 10.1016/j.mimet.2006.05.016

Vizcaíno, J. A., Côté, R. G., Csordas, A., Dianes, J. A., Fabregat, A., Foster, J. M., et al. (2013). The Proteomics Identifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* 41, D1063–D1069. doi: 10.1093/nar/gks1262

Vizcaíno, J. A., Deutsch, E. W., Wang, R., Csordas, A., Reisinger, F., Ríos, D., et al. (2014). ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* 32, 223–226. doi: 10.1038/nbt.2839

Wang, R., Prince, J. T., and Marcotte, E. M. (2005). Mass spectrometry of the *M. Smegmatis* proteome: protein expression levels correlate with function, operons, and codon bias. *Genome Res.* 15, 1118–1126. doi: 10.1101/gr.3994105

Wessel, D., and Flügge, U. I. (1984). A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* 138, 141–143. doi: 10.1016/0003-2697(84)90782-6

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis.* New York, NY: Springer-Verlag.

WHO (2015). *Global Tuberculosis Report 2015.* WHO. Available online at: http://www.who.int/tb/publications/global_report/en/ (Accessed January 19, 2016).

CrossMark

Proteomics of *Neisseria gonorrhoeae*: the treasure hunt for countermeasures against an old disease

Benjamin I. Baarda and Aleksandra E. Sikora*

Department of Pharmaceutical Sciences, College of Pharmacy, Oregon State University, Corvallis, OR, USA

Neisseria gonorrhoeae is an exquisitely adapted, strictly human pathogen and the causative agent of the sexually transmitted infection gonorrhea. This ancient human disease remains a serious problem, occurring at high incidence globally and having a major impact on reproductive and neonatal health. *N. gonorrhoeae* is rapidly evolving into a superbug and no effective vaccine exists to prevent gonococcal infections. Untreated or inadequately treated gonorrhea can lead to severe sequelae, including pelvic inflammatory disease and infertility in women, epididymitis in men, and sight-threatening conjunctivitis in infants born to infected mothers. Therefore, there is an immediate need for accelerated research toward the identification of molecular targets for development of drugs with new mechanisms of action and preventive vaccine(s). Global proteomic approaches are ideally suited to guide these studies. Recent quantitative proteomics (SILAC, iTRAQ, and ICAT) have illuminated the pathways utilized by *N. gonorrhoeae* to adapt to different lifestyles and micro-ecological niches within the host, while comparative 2D SDS-PAGE analysis has been used to elucidate spectinomycin resistance mechanisms. Further, high-throughput examinations of cell envelopes and naturally released membrane vesicles have unveiled the ubiquitous and differentially expressed proteins between temporally and geographically diverse *N. gonorrhoeae* isolates. This review will focus on these different approaches, emphasizing the role of proteomics in the search for vaccine candidates. Although our knowledge of *N. gonorrhoeae* has been expanded, still far less is known about this bacterium than the closely related *N. meningitidis*, where genomics- and proteomics-driven studies have led to the successful development of vaccines.

Keywords: *Neisseria gonorrhoeae*, gonorrhea, proteomics, molecular targets, vaccine, drugs, antibiotic resistance, surveillance

INTRODUCTION

Gonorrhea is an ancient human disease, with references to its symptoms found in the Old Testament of the Bible (Leviticus 15:1–3). For almost 700 years, it has been known as "the clap," a likely reference to the old Le Clapiers district of Paris where prostitutes were housed. This sexually transmitted disease remains a global scourge today, causing an estimated 106 million new cases worldwide each year WHO (2012). In the United States, gonorrhea is the second most common sexually transmitted disease, with over 300,000 new cases, primarily affecting those between

20 and 24 years of age, reported to the Centers for Disease Control and Prevention (CDC) annually CDC (2013). The Gram-negative diplococcus *Neisseria gonorrhoeae*, the gonococcus (GC), is the sole cause of gonorrhea. In men, infections typically present as profuse, localized inflammatory response of the urethra (i.e., urethritis). In contrast, gonorrhea remains asymptomatic in 50–80% of infected women (Farley et al., 2003; WHO, 2011). Untreated or inadequately treated gonococcal infections can have severe consequences including epididymitis in men and pelvic inflammatory disease and inflammation of the uterus, ovaries and fallopian tubes in women, which can lead to infertility. Neonatal health is also detrimentally affected by GC infection, as this pathogen can cause a sight-threatening conjunctivitis in infants born to infected mothers (Creighton, 2011). Additionally, infection with GC increases the risk of HIV transmission (Tapsall, 2005). Further compounding the difficulty in treating gonorrheal infections, through a number of point mutations, as well as horizontally acquired genes, GC has gained resistance to nearly all antibiotics currently in use (Tapsall et al., 2010; Unemo and Shafer, 2014). The CDC now recommends a combination of ceftriaxone with either doxycycline or azithromycin for empirical gonorrhea treatment (CDC, 2012b); however, treatment failures with ceftriaxone have been verified in Japan, Australia, Sweden, and Slovenia (reviewed in Unemo, 2015). Additionally, GC demonstrates remarkable heterogeneity and strain-to-strain variability, which represent a significant challenge in vaccine development (Zhu et al., 2011; Jerse et al., 2014).

Immediate action is critically needed against gonorrhea before it becomes completely untreatable. In response to this dire possibility, the World Health Organization (WHO) published the "Global Action Plan to Control the Spread and Impact of Antimicrobial Resistance in *Neisseria gonorrhoeae*" (WHO, 2012), and the CDC (CDC, 2012a), as well as the European Centre for Disease Prevention and Control (ECDC; ECDC, 2012) proposed region-specific response plans. Overall, these proposals stressed the importance of implementing holistic action against gonorrhea, which would encompass early prevention, diagnosis, contact tracing, treatment, and surveillance of antimicrobial resistance and treatment failures.

An ideal method of gonorrhea prevention would be the development of a protective vaccine(s). Indeed, according to model simulations, gonococcal prevalence could be reduced by at least 90% after 20 years if all 13-year olds were given a non-waning vaccine with 50% efficacy or a vaccine with 100% efficacy that wanes after 7.5 years. Further, even with a non-waning vaccine of 20% efficacy, as much as a 40% decrease in prevalence could be anticipated (Craig et al., 2015). In line with the WHO, CDC, and ECDC call for action, it is a prerequisite that the anti-gonorrhea vaccine and new drug development be made a priority. Different proteomic approaches are exceedingly valuable to accompany the progress of effective new therapeutic interventions by identifying vaccine and drug targets. Herein, we guide through these different approaches in the treasure hunt for countermeasures against gonorrhea, emphasizing the role of proteomics in the search for GC vaccine candidates (**Figure 1**).

HOW DOES A HEALTH-OBSESSED RESEARCHER EVALUATE AN ENERGY BAR?

WITH PROTEOMICS

As genomic approaches, whole-genome sequencing in particular, have become relatively inexpensive and increasingly high-throughput, with short turn around times and great resolution in the past few years, they have grown increasingly useful in basic research and in clinical diagnosis. The Broad Institute has recently released the whole genome sequences of 14 GC clinical isolates in collaboration with the *Neisseria gonorrhoeae* Group to facilitate research into pathogenesis and genetic determinants of disease states (*Neisseria gonorrhoeae* Group Sequencing Project, Broad Institute of Harvard and MIT[1]). Additionally, multilocus sequence typing (MLST), which characterizes isolates based on internal fragments of housekeeping gene alleles, has been used to cluster GC patient isolates based on phenotype (Ilina et al., 2010). A database of MLST data for *Neisseria* species has been established[2], further facilitating MLST identification and genotypic grouping of GC isolates. Genomic-derived methodologies have identified GC iron-responsive genes (Ducey et al., 2005), the anaerobic stimulon (Isabella and Clark, 2011), as well as gene expression patterns during infection of the lower female genital tract (McClure et al., 2015). In clinical applications, the proliferation of genomic datasets has promoted the development of nucleic acid amplification tests (NAATs), which allow for rapid identification of GC in patient samples without the need for culture (Low et al., 2014).

Ultimately, however, genomics is unable to capture the complete biological complexity present. A combination of different proteomic approaches can greatly complement genomic-acquired data by examining the GC protein population in biofilms or upon exposure to relevant host stimuli (Wu et al., 2010; Phillips et al., 2012), proteins associated with drug resistance (Nabu et al., 2014), proteome expression patterns during infection, post-translational modifications, or by providing information about proteins' subcellular location, structures and protein–protein interactions. The knowledge gained from proteomic studies can be useful for identifying GC in clinical samples (Gudlavalleti et al., 2008; Carannante et al., 2015), evaluating antibiotic resistance, and discovering potential vaccine and drug candidate proteins (Zielke et al., 2014, 2015).
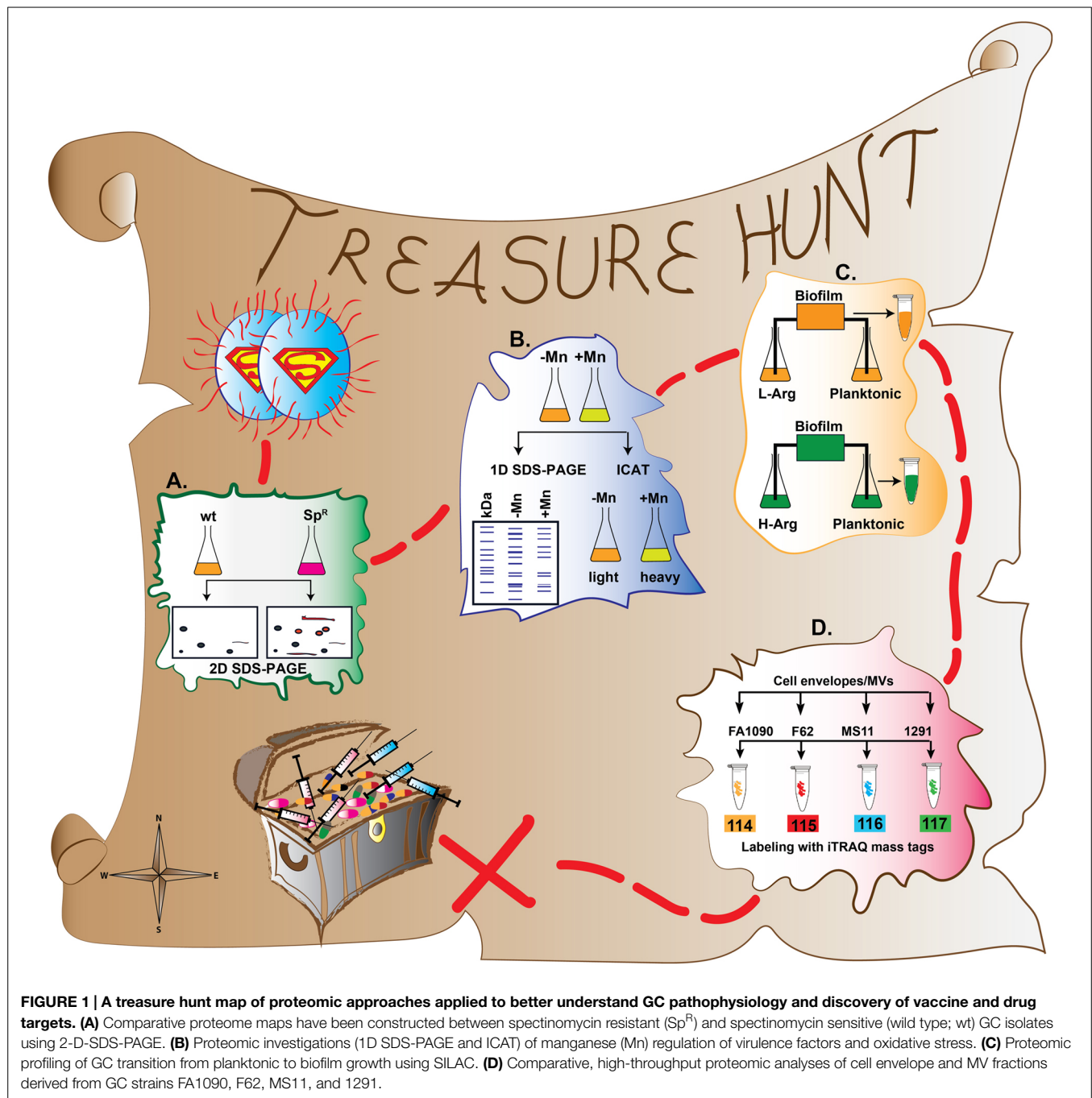
WHEN DO YOU NOT WANT TO RECEIVE APPLAUSE?

WHEN YOU IDENTIFY "THE CLAP"

Rapidly identifying GC in clinical isolates is vital to initiate treatment as quickly as possible to prevent the severe consequences of untreated gonorrhea, as well as to limit the spread of antimicrobial resistant strains. NAATs, developed

---

[1]http://www.broadinstitute.org/
[2]http://pubmlst.org/neisseria/

FIGURE 1 | A treasure hunt map of proteomic approaches applied to better understand GC pathophysiology and discovery of vaccine and drug targets. **(A)** Comparative proteome maps have been constructed between spectinomycin resistant (Sp^R) and spectinomycin sensitive (wild type; wt) GC isolates using 2-D-SDS-PAGE. **(B)** Proteomic investigations (1D SDS-PAGE and ICAT) of manganese (Mn) regulation of virulence factors and oxidative stress. **(C)** Proteomic profiling of GC transition from planktonic to biofilm growth using SILAC. **(D)** Comparative, high-throughput proteomic analyses of cell envelope and MV fractions derived from GC strains FA1090, F62, MS11, and 1291.

with the use of genomic data, are more rapid and sensitive than culture, which has resulted in an increase in the number of infections detected (Unemo and Shafer, 2014; Wind et al., 2015). Current commercially available gonococcal NAATs are unable to satisfactorily measure antimicrobial resistance, undermining the surveillance of antimicrobial resistance trends. Because of this limitation, the ECDC has recommended that all specimens positive for GC by NAAT are subsequently cultured to monitor antimicrobial resistance trends (ECDC, 2012). However, laboratory-developed NAATs have been utilized to identify known genetic antimicrobial resistance determinants

against several classes of antibiotics as well as detecting the crucial mutations conferring resistance to extended-spectrum cephalosporins in superbug strains H041 and F89 (reviewed, by Unemo, 2015).

As antibiotic resistance determinants continuously evolve, the development of new detection tests is required (Unemo, 2015). Proteomics can provide sensitive, accurate, rapid, and cost-effective methods of GC identification and determination of antimicrobial resistance patterns in clinical samples. Proteomic identification of bacteria is primarily performed with data generated by matrix-assisted laser desorption ionization

time-of-flight mass spectrometry (MALDI-TOF-MS). Unique and representative biomarker ions can be established from intact cell MALDI-TOF-MS analysis (van Baar, 2000; Amiri-Eliasi and Fenselau, 2001; Fenselau and Demirev, 2001; Fagerquist et al., 2010; Murray, 2010; Niyompanich et al., 2014). The most important advantages in direct bacterial profiling by means of MALDI-TOF-MS are: (1) the requirement for only a small amount of biological material, (2) the possibility of examining intact cells without preceding extraction and separation, (3) a fast and straightforward procedure, and (4) high specificity in species differentiation (Ilina et al., 2009).

One of the first examples of direct GC profiling with MALDI-TOF-MS used surface enhanced MALDI-TOF-MS to analyze over 350 GC strains and closely related species (Schmid et al., 2005). These comparisons enabled the design of multilayer artificial neural networks that revealed 20 ion peak descriptors of positive, negative and secondary nature that were supreme for GC identification (over 96% efficiency, a sensitivity of 95.7% and a specificity of 97.1%). Another study used atmospheric pressure MALDI-TOF to determine that a putative DNA binding protein from *N. meningitidis*; its homolog in GC, DbhA; and acyl carrier proteins in each species could be used as protein biomarkers for identifying pathogenic *Neisseria* (Gudlavalleti et al., 2008). Although this approach does not provide a method to distinguish between *N. meningitidis* and *N. gonorrhoeae*, it can act as a base upon which to build techniques for identifying pathogenic *Neisseria* in clinical samples. The successful application of MALDI-TOF-MS for GC identification has also been recently demonstrated on 92 out of 93 isolates of gonococci collected from 2007 to 2012 as part of the European Gonococcal Antimicrobial Surveillance Programme (Carannante et al., 2015).

Together, these studies highlight the potential of proteomic approaches to rapidly and correctly identify GC in various clinical isolates, which, if implemented on a larger scale, will promote rapid initiation of treatment while still allowing antimicrobial susceptibility testing to be performed.

WHAT DID THE BACTERIA CALL THEIR GUERILLA WARFARE UNIT?

THE SPECTINOMYCIN RESISTANCE

In a survey of antimicrobial resistance in Southeast Asia, the WHO Global Gonococcal Antimicrobial Surveillance Program found that between 0.6 and 10.5 percent of isolates demonstrated spectinomycin resistance (Bala et al., 2013). Spectinomycin directly interacts with 16S rRNA and inhibits the elongation factor G (EF-G)-catalyzed translocation of the peptidyl-tRNA from the A site to the P site during protein synthesis (Bilgin et al., 1990; Ramakrishnan and White, 1992). Not surprisingly, spectinomycin resistance determinants traditionally involve mutations in 16S rRNA (Maness et al., 1974; Galimand et al., 2000). However, a deletion of amino acid 25 and a K26E amino acid alteration (*E. coli* numbering) in the ribosomal protein S5 is a newly identified mechanism associated with high-level spectinomycin resistance in GC (Unemo et al., 2013). Overall, however, the *in vitro* susceptibility to this aminocyclitol compound is remarkably

high worldwide and this antibiotic is an effective alternative for treatment of anogenital gonorrhea, in particular for multidrug resistant cases (Unemo, 2015).

The proteomic signatures of GC spectinomycin resistance as well as cellular responses to spectinomycin treatment were assessed through qualitative and relative quantitative proteomics using densitometry analysis of proteins separated by two-dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis (2D SDS-PAGE) and protein identification with MALDI-TOF-MS (Otto et al., 2012; Nabu et al., 2014). Comparative proteome maps have been constructed between a spectinomycin resistant (Spec^R) clinical isolate and a spectinomycin sensitive (Spec^S) reference strain (ATCC 49226; **Figure 1A**). When the two strains were not exposed to spectinomycin, their protein profiles were largely the same, with EF-Tu and EF-Ts, cysteine synthase, and the septum site-determining protein MinD upregulated in the Spec^R isolate. Additionally, MinD, oxidoreductase, and hypothetical protein NGO1873 were shifted to a more acidic isoelectric point, and the ribosomal protein S6 was shifted to a more basic pI in the resistant isolate. Finally, ABC transporter substrate-binding protein showed decreased expression and alcohol dehydrogenase was not detected in the Spec^R strain. Interestingly, in the presence of spectinomycin, although dihydrolipoamide dehydrogenase, peroxiredoxin, an outer membrane protein Rmp, and the 50S ribosomal protein L7/L12 were upregulated in both strains, none of the proteins were expressed as highly in the resistant strain as in the wild type. Overall, the spectinomycin treatment of the Spec^S GC resulted in alterations in the abundance of proteins involved in energy metabolism and detoxification, as well as cell envelope proteins (Rmp and ABC transporter substrate-binding protein).

Based on the obtained proteomic comparisons, a mode of spectinomycin action on GC is proposed but not experimentally validated (Nabu et al., 2014). Briefly, in the presence of sub-minimal inhibitory concentrations, the drug destabilizes the outer membrane, recruiting Rmp to maintain the integrity of the cell envelope. Spectinomycin accumulates in the periplasm, generating a concentration imbalance between the inside and outside of the cell. In response, significant changes in energy metabolism including an increase in NADH production and oxidation through the electron transport chain occur. At the same time, additional proton motive force allows more drug molecules to enter the cell. Increased NADH oxidation leads to an increase in reactive oxygen species, which are detoxified by the increase in peroxiredoxin and glutamate dehydrogenase production. Higher levels of L7/L12 allow the cell to overcome the inhibition of ribosomal translocation imposed by spectinomycin. In the Spec^R strain, amino acid ATP binding cassette transporter substrate-binding protein levels are decreased even in the absence of spectinomycin, possibly affecting uptake of the drug. Expression of enolase is decreased, which may result in increased levels of 2P-D-glycerate and 3P-D-glycerate. Together with an increase in cysteine synthetase expression, this may improve the cell's defense against reactive oxygen species. Finally, the already high levels of EF-Tu and –Ts, together with the upregulation of L7/L12, may assist with protein translation in the presence of spectinomycin (Nabu et al., 2014).

Future research may be able to exploit some of the suggested secondary pathways and increase the efficacy or reduce the resistance potential of future antimicrobial compounds. Also apparent in this study are some of the shortcomings of 2D SDS-PAGE, which has limited utility when dealing with low abundance and/or membrane proteins.

HOW DID THE HEALTH-OBSESSED RESEARCHER COMPARE TWO PROTEIN SHAKES?

WITH QUANTITATIVE PROTEOMICS

In contrast to qualitative proteomics, quantitative proteomic approaches allow for (absolute or relative) quantification of proteins on a global scale. The object of quantitative proteomics is to identify specific alterations between control samples and particular experimental conditions (e.g., healthy versus diseased state). In addition, quantitative proteomic profiling may focus on a specific subset of proteins (subproteome), where for instance bacterial whole cell lysates are subjected to fractionation to enrich for cell envelope proteins. Very often the proteins of interest are relatively low in abundance; therefore it is critical to utilize appropriate combinations of pre-fractionation techniques such as different kinds of liquid chromatography (affinity, reversed phase, size-exclusion, or ion exchange) to reduce the complexity of analyzed samples, in addition to employing sensitive mass spectrometry instruments. Different approaches have been developed to facilitate quantitative proteome profiling studies involving stable isotope labeling, such as ICAT, ICPL, SILAC, iTRAQ, TMT, and IPTL. In addition, label-free statistical methodologies (MRM, SWATH) and absolute quantification methods by mass spectrometry (AQUA strategy) have become available (reviewed in Chahrour et al., 2015).

Quantitative proteomic approaches, involving different stable isotope labels and summarized in **Table 1**, have been used to investigate how GC responds to oxidative stress (Wu et al., 2010), transitions from planktonic to biofilm growth (Phillips et al., 2012), and adapts the composition of the cell envelope in response to different environmental cues encountered in micro-ecological niches within the host (Zielke et al., 2015). In addition, iTRAQ technology has also revealed the dynamic subproteomes of cell envelopes and naturally released membrane vesicles in four different GC isolates (Zielke et al., 2014).

WHAT HAPPENS WHEN TWO OXEN GO ON THEIR FIRST DATE?

OXIDATIVE STRESS

In a typical inflammatory response to GC, neutrophils are recruited to the site of infection by chemokines, such as IL-8, released by infected mucosal surfaces (Criss et al., 2009). When activated neutrophils phagocytize GC, the production of reactive oxygen species (ROS) is either stimulated or inhibited, depending on GC expression of opacity-associated proteins

(Opa). Cells expressing Opa protein (Opa$^+$) ligate to human carcinoembryonic antigen-related cell adhesion molecules (CEACAMs), specifically CEACAM3, on the surface of the neutrophil, triggering phagocytosis, NADPH subunit assembly, and degranulation (Schmitter et al., 2007; Sarantis and Gray-Owen, 2012). Additionally, Opa$_{57}$ protein ligated to CEACAM3 amplifies the inflammatory response by activating nuclear factor (NF)-κB and increasing phosphorylation of the p38 kinase (Sintsova et al., 2014).

This cascade of ROS production puts the GC cells under tremendous oxidative stress. However, viable GC has been shown to survive and replicate within neutrophils, even after NADPH oxidase activation (Simons et al., 2005), indicating that GC is able to protect itself against oxidative stress. This pathogen can defend against superoxide radicals (Tseng et al., 2001) and hydrogen peroxide (Seib et al., 2004) in a manganese (Mn) dependent manner. A study combining transcriptomic and qualitative- and quantitative-proteomic approaches examined the protective mechanism utilized by GC in the presence of Mn (Wu et al., 2010). One-dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis (1D SDS-PAGE) paired with one-dimensional liquid chromatography-tandem mass spectrometry (1D LC-MS/MS) identified 46 proteins expressed only in the presence of Mn. Notably, bacterioferritin, azurin (Laz), and iron-superoxide dismutase (SodB)—proteins involved in defense against oxidative stress—were expressed in bacteria cultured in the presence of Mn, but were not detected in those cultured without Mn. When cultured in the absence of Mn, GC was found to express an outer membrane methionine sulfate reductase that is also involved in defense against superoxide and hydrogen peroxide.

Further, a quantitative proteomic investigation using Isotope-Coded Affinity Tag (ICAT) labeling was used to determine which proteins were differentially regulated by the presence or absence of Mn (Wu et al., 2010). For ICAT analysis, two protein samples are labeled with reactive groups (biotinylated iodoacetamide or acrylamide derivatives) that specifically react with the sulphydryl groups of denatured peptides' cysteine side chains (Gygi et al., 1999). One sample is labeled with a light isotope, while the other sample is labeled with a heavy isotope. The samples are then combined and analyzed by mass spectrometry (**Figure 1B**; **Table 1**). The protein populations can then be quantified by comparing the ratio of heavy to light proteins (Gygi et al., 1999; Colangelo and Williams, 2006; Chahrour et al., 2015).

In this study, ICAT labeling, coupled with MS/MS, revealed numerous proteins that were downregulated more than 1.5-fold in the presence of Mn, including PilT (an ATPase involved in pilus disassembly), OmpR, a 64 KDa outer membrane protein (OMP P64k), and peroxiredoxin, which reduces and detoxifies peroxides (Seib et al., 2006). Additionally, Mn affected the levels of pilin, superoxide dismutase, and pyrophosphatase without causing a corresponding change in the transcript level, indicating that these proteins are likely to be regulated post-transcriptionally by the presence of Mn (Wu et al., 2010).

Taken together, the results of this study suggest that, in the presence of Mn, GC upregulates the expression of iron storage proteins that protect against oxidative damage. Concomitantly, the bacterium downregulates pyrophosphatase

TABLE 1 | Quantitative Proteomic Approaches utilized to studying GC.

Proteomic Approach	Principles	Use in GC	References
Isotope coded affinity tag (ICAT)	- Post-harvest labeling - Heavy or light isotopic reactive groups label cysteine residues in a protein sample. - Two samples are pooled and analyzed by MS. - Relative quantification of proteins by comparing heavy: light ratio.	Response to oxidative stress	Wu et al. (2010)
Stable isotope labeling by amino acids in cell culture (SILAC)	- Metabolic labeling - Growth medium is supplemented with either light (unlabeled) or heavy (labeled) amino acid. - Two samples are pooled and analyzed by MS. - Relative quantification of proteins by comparing heavy: light ratio	Shift from planktonic to biofilm growth	Phillips et al. (2012)
isobaric Tagging for Relative and Absolute Quantification (iTRAQ)	- Post-harvest labeling - Proteolytically digested proteins are tagged on N-termini and lysine side chains with isobaric mass labels. - Two to eight samples are pooled and analyzed by MS. - Relative or absolute quantification of proteins by measuring intensity of reporter ion peak.	Cell envelopes and membrane vesicles Comparison of cell envelope profiles upon environmental cues encountered in the host	Zielke et al. (2014); Zielke et al. (2015)

(Ppa) and polyphosphate kinase (Ppk). Ppa hydrolyzes inorganic pyrophosphate (PP_i) into two molecules of orthophosphate (P_i), while Ppk synthesizes polyP from P_i. When these two proteins are downregulated, PP_i accumulates in the cell and chelates Mn. In a Mn(II)-PP_i complex, Mn is able to defend against ROS non-enzymatically (Wu et al., 2010).

The information gleaned from these proteomic experiments could suggest a method to combat gonorrhea whereby the downregulation of protective proteins required only in the absence of Mn is maintained, but the alteration of the levels of proteins required for protection from oxidative stress in the presence of Mn is blocked. Importantly, this study highlights the utility of proteomic approaches to investigate biological responses involving post-transcriptional regulation that genomic methods alone cannot discover. It is important to keep in mind, however, that ICAT has two major drawbacks: (1) proteins that do not have any cysteine residues will be eliminated from this analysis, and (2) the release of biotinylated peptides from the streptavidin column is not quantitative for low-abundance peptides (Chahrour et al., 2015).

WHAT IS A MICROBIOLOGIST'S FAVORITE KIND OF MOVIE?

A BIOFILM

Bacteria often shift from planktonic (free living bacteria) growth to a biofilm community, where bacteria grow in close proximity to each other, protected by an extracellular polymer composed of polysaccharides, proteins, nucleic acids, and lipids (Flemming and Wingender, 2010). GC has been shown to form biofilms *in vitro* (Greiner et al., 2005), and an examination of primary cervical epithelial cells from cervical biopsy samples revealed biofilm growth in culture positive gonorrhea cases (Steichen et al., 2008). Biofilms exacerbate antibiotic resistance by providing a protective barrier against antimicrobial action, and biofilms formed by GC are thought to contribute to asymptomatic infections (Steichen et al., 2008).

To understand the mechanisms underlying biofilm formation, a quantitative proteomic study examined the proteome changes GC undergoes in the transition from planktonic to biofilm growth using stable isotope labeling by amino acids in cell culture (SILAC; **Figure 1C**; **Table 1**; Phillips et al., 2012). One of the important advantages of SILAC over other stable isotope labeling methods is that the label is integrated into the peptide at early stages of experimentation when the sample is metabolically active. Thus, possible variability due to sample preparation and purification losses are eliminated (Chahrour et al., 2015).

In this analysis, planktonic cells of GC strain 1291, which is an arginine auxotroph, were grown with labeled $^{13}C_6$-Arg, and biofilm cells were grown with unlabeled Arg in a continuous-flow apparatus. Extracted protein samples from each bacterial population were combined and subjected to MALDI-TOF mass spectrometry analysis. Overall, this global analysis identified 757 proteins, 152 of which were significantly differentially expressed. In particular, GC cultured in a biofilm exhibited 73 upregulated- and 54 downregulated-proteins when compared to planktonic growth. The results of this study indicated that the bacteria upregulate proteins to respond to an oxygen-limited environment, including cytochrome c oxidase subunit III CcoP and nitrite reductase AniA. To cope with restricted nutrient availability in the biofilm, bacterial metabolism is shifted to increase sugar fermentation and tricarboxylic acid (TCA) cycle enzymes. The composition of the outer membrane is also altered during growth in a biofilm, with increased levels of 9 proteins including OpaB and OpaD (Phillips et al., 2012), both of which have been shown to adhere to and damage fallopian tube mucosa (Dekker et al., 1990). Among the downregulated proteins in the biofilm were proteins involved in energy metabolism, protein fate and synthesis, and transport and binding proteins, specifically iron complex outer membrane receptor protein (FetA) as well as transferrin-binding protein B and A (TbpB and TbpA, respectively).

A direct comparison of the transcriptome (Falsetta et al., 2009) and proteome expression profiles of GC biofilms showed a very poor correlation with only seven overlapping hits including AniA, OpaB, cytochrome C peroxidase CcpR, putative

dihydrolipoamide dehydrogenase, putative cysteine synthase/cystathionine beta-synthase, hypothetical protein NGO0905, and putative ABC transporter NGO1494 (Phillips et al., 2012).

This study gives insight into the adaptations necessary for GC to establish long-term infections and emphasizes the utility of proteomic approaches to examine these adaptations. In addition, the identified upregulated outer membrane proteins may be utilized as biomarkers for gonorrhea diagnostics.

WHAT BOOK-LIKE CANDIDATE WAS NOT AT THE LAST PRESIDENTIAL DEBATE?

A NOVEL VACCINE CANDIDATE

Perhaps one of the most exciting uses of proteomic approaches is in the search for new ways to combat multidrug resistant GC. We are applying a proteomics-driven reverse vaccinology approach to identify vaccine candidate proteins against gonorrhea (Zielke et al., 2014, 2015). Reverse vaccinology searches for possible vaccine candidate proteins using different genomics and proteomics methodologies and has already been successfully applied to different pathogenic bacteria including *N. meningitidis* serogroup B (Heckels and Williams, 2010; Adamczyk-Poplawska et al., 2011; Seib et al., 2012; Delany et al., 2013; Heinson et al., 2015).

No vaccine against GC currently exists, although research has been ongoing for decades. Two attempted vaccines, comprised of killed whole cells and purified pilin protein, failed in clinical trials over 13 years ago (Zhu et al., 2011). Since that time, very little research into gonorrheal vaccines has occurred, mainly due to the highly variable targeted surface proteins. Because GC is a strict human pathogen, research was also hampered by the lack of a suitable small animal model for gonorrheal infection (Jerse et al., 2014). Fortunately, a mouse model of female infection was developed, in which female mice are treated with 17-β-estradiol when they are in the diestrus stage of the estrus cycle. The mice are also treated with an antibiotic cocktail of streptomycin sulfate, vancomycin HCl, and trimethoprim sulfate to prevent overgrowth of commensal vaginal bacteria while under the influence of estradiol. 2 days after estradiol treatment, GC is introduced intravaginally. Using this model, GC can be recovered an average of 12.2 days post-inoculation with 10^6 colony forming units (Jerse, 1999). A further advancement in the mouse model has been the development of transgenic mice that express human CEACAM proteins, providing a closer reproduction of conditions encountered in the human host (Jerse et al., 2014). The availability of a mouse model has greatly facilitated vaccine research. The immune response to infection, as well as resistance to subsequent infections after inoculation with an experimental vaccine can be closely monitored and investigated with the genetic tools available for studying mice (Zhu et al., 2011). However, to fully utilize this model for vaccine research, suitable candidate proteins must be identified—a goal for which proteomic approaches are ideally suited.

During the development of the MenB vaccine, out of nearly 600 candidates selected by reverse vaccinology, 350 recombinant proteins were successfully expressed in *Escherichia coli* and evaluated for their surface exposure. A total of 28 among them elicited bactericidal antibodies against Group B meningococci *in vitro*. Finally, the neisserial heparin-binding antigen NHBA, factor H-binding protein fHbp, as well as the neisserial adhesin NadA were chosen as part of the MenB vaccine (Seib et al., 2012; Delany et al., 2013; Jerse et al., 2014). In contrast, only 12 different candidates are being evaluated as potential gonorrhea vaccine antigens (Jerse et al., 2014). Therefore, a more far-reaching effort is required to make a gonorrhea vaccine a reality.

Of particular interest for vaccine development and identification of new drug targets are proteins localized to the bacterial cell envelope and membrane vesicles (MVs)—spherical outpouchings of the cell envelope—as they interact directly or indirectly with host tissues; play roles in pathogenesis, antibiotic resistance, and biofilm formation; and participate in general physiological processes. Surprisingly few studies addressed GC cell envelope composition (Yoo et al., 2007; Phillips et al., 2012; Zielke et al., 2014). Also, despite studies reporting the release of MVs and their different morphological forms (spherical, lobed, and tubular) in GC from the early 1970s, only a few reports focused on elucidating their components (Swanson et al., 1971; Dorward et al., 1989; Pettit and Judd, 1992a,b; Falsetta et al., 2011; Zielke et al., 2014).

To begin the systematic mining of GC cell envelope and MVs for the discovery of vaccine and drug candidates, we first used the PSORTb 3.0.2 (Gardy et al., 2005) bioinformatics predictions and analyzed the subcellular localization of all ORFs in the completed genome sequence of strains FA1090 (Gen Bank accession number AE004969) and NCCP11945 (Gen Bank accession number CP001050), as well as the draft genome sequences of 14 different GC strains (downloaded from the Broad Institute website[3]). These studies revealed that, on average, about 50 of the 2,000 ORFs present in the GC genome encode outer membrane proteins (Zielke and Sikora, 2014). However, the subcellular location could not be predicted for about 30% of all ORFs. This analysis demonstrated that there is still much to learn about GC cell envelope composition and opened up exciting prospects for applying proteomics for the discovery of vaccine targets.

As proteomic investigations of membrane proteins are technically challenging, we chose to apply gel-free quantitative proteomic approaches including isobaric tagging for relative and absolute quantification (iTRAQ, **Table 1**) combined with multidimensional liquid chromatography and tandem mass spectrometry (2D-LC/MS/MS) to examine cell envelopes and naturally released MVs (Zielke et al., 2014). Four GC strains: FA1090, F62, MS11, and 1291 were cultured in liquid media under standard growth conditions and their cell envelopes and MVs were harvested in mid-logarithmic phase of growth. iTRAQ quantification was performed by labeling proteins isolated from subproteome fractions of each strain of interest with one of four isobaric tags (N-hydroxysuccinimide ester-activated compounds) that react to free amine groups on the N-termini and lysine side chains of proteins with high efficiency (Ross et al., 2004). Each of

---

[3]http://www.broadinstitute.org/annotation/genome/neisseria_gonorrhoeae/MultiHome.html

the four tags contains a reporter ion of a unique mass (**Figure 1D**; **Table 1**). When the samples are combined and subjected to mass spectrometry, the reporter ions are released from the labeled peptide. After the release of the reporter ions, all of the identical peptides in a sample will result in identical mass spectra, and the abundance of the peptide in the four multiplexed samples can be quantified by the relative intensity of the corresponding reporter ion peak (Ross et al., 2004; Wiese et al., 2007). The advantage of using iTRAQ is that it can be easily multiplexed and up to 8 different samples can be simultaneously analyzed within a single experiment. Additionally, as the iTRAQ tags react with all primary amine functional groups of peptides, nearly all peptides are labeled and information about not only their abundance but also their modification(s) can be acquired (Chahrour et al., 2015).

Our proteomic profiling of cell envelopes and native MVs revealed 533 and 168 common proteins, respectively, in analyzed GC strains. A total of 22 differentially abundant proteins were discovered including hitherto unknown proteins. Among those proteins that displayed similar abundance in four GC strains, we identified 305 and 46 cell envelope-and MVs-associated proteins, respectively. In addition, 34 proteins were found in both cell envelopes and MVs with eleven of them differentially regulated (Zielke et al., 2014). A few of these differentially expressed proteins included cytoplasmic proteins, an observation that was confirmed by a subsequent, independent proteomic study of MVs (Perez-Cruz et al., 2015).

The ubiquitous outer membrane proteins identified included GC homologs of the outer membrane β-barrel assembly (Bam) protein complex (Ricci and Silhavy, 2012), including BamA, BamD, and BamE; lipopolysaccharide transport protein LptD; and TamA (NGO1956) and TamB (NGO1955), two proteins thought to work cooperatively to assist in the assembly of autotransporter proteins (Heinz et al., 2015). Numerous uncharacterized proteins were also ubiquitously expressed, including NGO1344, NGO1985, NGO2111, NGO2121, and NGO2139 (Zielke et al., 2014).

We further examined LptD, NGO1344, NGO1985, NGO2111, NGO2121, NGO2139, and TamB by constructing conditional- (LptD) or complete- knockout strains for each protein. These proteins were chosen because they contain domains predicted to function in maintaining cell envelope homeostasis. LptD expression in GC was placed under the control of an isopropyl β-D-1-thiogalactopyranoside (IPTG)-inducible promoter. This strain was unable to grow when streaked onto media lacking IPTG, while it grew robustly on plates supplemented with IPTG. Further, after 3 h of culturing in liquid media without IPTG, the bacteria ceased to grow. By 5 h, bacterial viability decreased dramatically (nearly 13-fold) compared to the LptD-expressing strain. These experiments indicated that LptD is likely essential for GC viability. To test whether NGO1344, NGO1985, NGO2111, NGO2121, NGO2139, and TamB play functions in the integrity of the GC cell envelope, the individual clean deletion mutants were constructed in strain FA1090 and spotted on plates supplemented with various compounds. Although the loss of these proteins had no effect on bacterial growth under permissive conditions, the loss of NGO1985 resulted in a severe growth defect in the presence of bile salts, polymyxin B, Tween 20, SDS, urea, and

chloramphenicol. Further, these phenotypes could be completely reversed by complementation with an IPTG-inducible version of the *ngo1985* gene. Additionally, NGO2121 exhibited reduced growth in the presence of bile salts and polymyxin B. These proteins, identified by quantitative proteomic approaches, appear to provide an important function in maintaining cell membrane integrity and, as such, are promising targets for development of new therapeutic interventions against GC (Zielke et al., 2014).

To continue this research endeavor, we went on to determine the ubiquitously and specifically expressed cell envelope proteins of GC FA1090 challenged with host-relevant environmental stimuli: oxygen availability, iron deprivation, and the presence of human serum (Zielke et al., 2015). A myriad of novel proteins have been identified. Our initial characterization of five novel vaccine candidates that were ubiquitously expressed under these different growth conditions demonstrated that BamA, LptD, TamA, NGO2054, and NGO2139 were surface exposed and produced bactericidal antibodies that cross-reacted with a panel of diverse GC isolates. These promising results strongly suggest that the proteomics-driven approach will provide a foundation for the development of anti-GC vaccine(s), which would be the ideal way to prevent gonorrhea. Finally, to promote the utilization of the newly identified proteins and the knowledge of the GC subproteome dynamics among the scientific community, our entire data sets from all these investigations were made publicly available via the ProteomeXchange Consortium[4], the PRIDE partner repository ProteomeXchange with the identifiers PXD000549 and PXD001944.

FUTURE DIRECTIONS

Although proteomic approaches have revealed a multitude of information on the physiology of GC, to formulate an effective vaccine, more information needs to be gathered about the way its surface proteins interact with host cells during an infection. Ideally, the proteome of clinical samples that have never been subcultured should be examined to determine which proteins are expressed during different stages of infection. This approach presents a significant technical challenge to overcome, as sufficient bacterial material must be collected for a quantitative proteomic analysis. In our investigation of the cell envelope and MV proteome, we collected material from 1 liter of culture (Zielke et al., 2014). Although other proteomic studies have used less material, including 5 milliliters of culture (Wu et al., 2010) or bacteria harvested from 20 plates (Anonsen et al., 2012; Perez-Cruz et al., 2015) collecting this amount of sample from a patient is not feasible; more sensitive MS analyzers or alternate methods of sample enrichment will be required. These studies will give insight into the proteomic adaptations the bacteria undergo to establish and maintain infection. Information collected from proteomic studies of clinical samples and host tissue culture can help further drive vaccine development and have the potential to aid in the discovery of stably expressed protein targets of antimicrobial agents with novel modes of action. Integration of multiple approaches, including public access to on-line raw data

---

[4]http://www.proteomexchange.org

is essential if there is to be a sense of participation across the biomedical research community.

An interesting analysis technique was recently pioneered by Altindis et al. (2015). Termed "protectome analysis," their technique searches for vaccine candidate proteins in proteomic datasets by identifying proteins with structural or functional features in common with proteins known to provide protection. This analysis tool could be used in combination with other proteomic studies to immediately identify proteins expressed during infection with the potential to provide protection against reinfection.

As more information is deposited into proteomic databases, the utility of proteomic approaches to identify GC in clinical samples will increase. For example, in cases where the molecular determinant of antimicrobial resistance is known, proteomic approaches have the potential to immediately recognize the protein modification(s) that result in antimicrobial resistance, a feat that is not possible with current NAAT identification techniques.

Finally, proteomic investigations of multidrug resistance strains can reveal the mode of drug action, as well as the pathway(s) the bacterium uses to resist multiple antibiotics. One mystery that proteomic approaches may be able to solve is the identity of "Factor X," an unknown determinant of penicillin and cephalosporin resistance that is non-transformable and therefore is difficult to study with typical genetic methods (Unemo and Shafer, 2014).

CONCLUSION

Developing novel vaccines or antimicrobial agents is critical in the face of growing antibiotic resistance, and global and quantitative proteomic approaches have begun to reveal potential targets in the fight against GC. Proteomic approaches are ideal for the discovery of vaccine candidate proteins, as well as protein targets for the development of novel antimicrobial agents. Qualitative proteomic studies revealed the GC defense response to spectinomycin, while quantitative proteomics have demonstrated bacterial adaptations to conditions encountered in the host, including oxidative stress, anoxia, iron deprivation, and the presence of human serum. Proteomics have also been recently adapted to identify GC in clinical samples, which can expedite treatment. Importantly for vaccine development, stably expressed proteins have been identified through high-throughput examinations of cell envelopes and naturally released MVs. These proteomic studies will act as starting points for studies into structural vaccinology, protein–protein interactions, and GC physiology, and have already given new insights into ways to combat this important, difficult to treat pathogen.

ACKNOWLEDGMENTS

REFERENCES

Adamczyk-Poplawska, M., Markowicz, S., and Jagusztyn-Krynicka, E. K. (2011). Proteomics for development of vaccine. *J. Proteom.* 74, 2596–2616. doi: 10.1016/j.jprot.2011.01.019

Altindis, E., Cozzi, R., Di Palo, B., Necchi, F., Mishra, R. P., Fontana, M. R., et al. (2015). Protectome analysis: a new selective bioinformatics tool for bacterial vaccine candidate discovery. *Mol. Cell. Proteom.* 14, 418–429. doi: 10.1074/mcp.M114.039362

Amiri-Eliasi, B., and Fenselau, C. (2001). Characterization of protein biomarkers desorbed by MALDI from whole fungal cells. *Anal. Chem.* 73, 5228–5231. doi: 10.1021/ac010651t

Anonsen, J. H., Egge-Jacobsen, W., Aas, F. E., Borud, B., Koomey, M., and Vik, A. (2012). Novel protein substrates of the phospho-form modification system in *Neisseria gonorrhoeae* and their connection to O-linked protein glycosylation. *Infect. Immun.* 80, 22–30. doi: 10.1128/IAI.05920-11

Bala, M., Kakran, M., Singh, V., Sood, S., Ramesh, V., and Members of WHO GASP SEAR Network. (2013). Monitoring antimicrobial resistance in *Neisseria gonorrhoeae* in selected countries of the WHO South-East Asia Region between 2009 and 2012: a retrospective analysis. *Sex Transm. Infect.* 89(Suppl. 4), iv28–iv35. doi: 10.1136/sextrans-2012-050904

Bilgin, N., Richter, A. A., Ehrenberg, M., Dahlberg, A. E., and Kurland, C. G. (1990). Ribosomal RNA and protein mutants resistant to spectinomycin. *EMBO J.* 9, 735–739.

Carannante, A., De Carolis, E., Vacca, P., Vella, A., Vocale, C., De Francesco, M. A., et al. (2015). Evaluation of matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) for identification and clustering of *Neisseria gonorrhoeae*. *BMC Microbiol.* 15:142. doi: 10.1186/s12866-015-0480-y

CDC. (2012a). Cephalosporin-resistant *Neisseria gonorrhoeae* public health response plan. Available at: http://www.cdc.gov/std/treatment/Ceph-R-ResponsePlanJuly30-2012.pdf (accessed May 5, 2015).

CDC. (2012b). Update to CDC's Sexually transmitted diseases treatment guidelines, 2010: oral cephalosporins no longer a recommended treatment for gonococcal infections. *MMWR Morb. Mortal. Wkly. Rep.* 61, 590–594. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22874837 (accessed August 24, 2015).

CDC. (2013). *2013 STD Surveillance Gonorrhea*. Available at: http://www.cdc.gov/std/stats13/gonorrhea.htm (accessed August 24, 2015).

Chahrour, O., Cobice, D., and Malone, J. (2015). Stable isotope labelling methods in mass spectrometry-based quantitative proteomics. *J. Pharm. Biomed. Anal.* 113, 2–20. doi: 10.1016/j.jpba.2015.04.013

Colangelo, C. M., and Williams, K. R. (2006). Isotope-coded affinity tags for protein quantification. *Methods Mol. Biol.* 328, 151–158. doi: 10.1385/1-59745-026-x:151

Craig, A. P., Gray, R. T., Edwards, J. L., Apicella, M. A., Jennings, M. P., Wilson, D. P., et al. (2015). The potential impact of vaccination on the prevalence of gonorrhea. *Vaccine* 33, 4520–4525. doi: 10.1016/j.vaccine.2015.07.015

Creighton, S. (2011). Gonorrhoea. *BMJ Clin. Evid*. 1604, 1–11.

Criss, A. K., Katz, B. Z., and Seifert, H. S. (2009). Resistance of *Neisseria gonorrhoeae* to non-oxidative killing by adherent human polymorphonuclear leucocytes. *Cell. Microbiol.* 11, 1074–1087. doi: 10.1111/j.1462-5822.2009.01308.x

Dekker, N. P., Lammel, C. J., Mandrell, R. E., and Brooks, G. F. (1990). Opa (protein II) influences gonococcal organization in colonies, surface appearance, size and attachment to human fallopian tube tissues. *Microb. Pathog.* 9, 19–31. doi: 10.1016/0882-4010(90)90037-Q

Delany, I., Rappuoli, R., and Seib, K. L. (2013). Vaccines, reverse vaccinology, and bacterial pathogenesis. *Cold Spring Harb. Perspect. Med.* 3, a012476. doi: 10.1101/cshperspect.a012476

Dorward, D. W., Garon, C. F., and Judd, R. C. (1989). Export and intercellular transfer of DNA via membrane blebs of *Neisseria gonorrhoeae*. *J. Bacteriol.* 171, 2499–2505.

Ducey, T. F., Carson, M. B., Orvis, J., Stintzi, A. P., and Dyer, D. W. (2005). Identification of the iron-responsive genes of *Neisseria gonorrhoeae* by microarray analysis in defined medium. *J. Bacteriol.* 187, 4865–4874. doi: 10.1128/JB.187.14.4865-4874.2005

ECDC. (2012). Response plan to control and manage the threat of multidrug-resistant gonorrhoea in Europe. Available at: www.ecdc.europa.eu/en/publications/ (accessed May 5, 2015).

Fagerquist, C. K., Garbus, B. R., Miller, W. G., Williams, K. E., Yee, E., Bates, A. H., et al. (2010). Rapid identification of protein biomarkers of *Escherichia coli* O157:H7 by matrix-assisted laser desorption ionization-time-of-flight-time-of-flight mass spectrometry and top-down proteomics. *Anal. Chem.* 82, 2717–2725. doi: 10.1021/ac902455d

Falsetta, M. L., Bair, T. B., Ku, S. C., Vanden Hoven, R. N., Steichen, C. T., McEwan, A. G., et al. (2009). Transcriptional profiling identifies the metabolic phenotype of gonococcal biofilms. *Infect. Immun.* 77, 3522–3532. doi: 10.1128/IAI.00036-09

Falsetta, M. L., Steichen, C. T., McEwan, A. G., Cho, C., Ketterer, M., Shao, J., et al. (2011). The composition and metabolic phenotype of *Neisseria gonorrhoeae* biofilms. *Front. Microbiol.* 2:75. doi: 10.3389/fmicb.2011.00075

Farley, T. A., Cohen, D. A., and Elkins, W. (2003). Asymptomatic sexually transmitted diseases: the case for screening. *Prev. Med.* 36, 502–509. doi: 10.1016/S0091-7435(02)00058-0

Fenselau, C., and Demirev, P. A. (2001). Characterization of intact microorganisms by MALDI mass spectrometry. *Mass Spectrom. Rev.* 20, 157–171. doi: 10.1002/mas.10004

Flemming, H. C., and Wingender, J. (2010). The biofilm matrix. *Nat. Rev. Microbiol.* 8, 623–633. doi: 10.1038/nrmicro2415

Galimand, M., Gerbaud, G., and Courvalin, P. (2000). Spectinomycin resistance in *Neisseria* spp. due to mutations in 16S rRNA. *Antimicrob. Agents Chemother.* 44, 1365–1366. doi: 10.1128/AAC.44.5.1365-1366.2000

Gardy, J. L., Laird, M. R., Chen, F., Rey, S., Walsh, C. J., Ester, M., et al. (2005). PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics* 21, 617–623. doi: 10.1093/bioinformatics/bti057

Greiner, L. L., Edwards, J. L., Shao, J., Rabinak, C., Entz, D., and Apicella, M. A. (2005). Biofilm formation by *Neisseria gonorrhoeae*. *Infect. Immun.* 73, 1964–1970. doi: 10.1128/IAI.73.4.1964-1970.2005

Gudlavalleti, S. K., Sundaram, A. K., Razumovski, J., and Doroshenko, V. (2008). Application of atmospheric pressure matrix-assisted laser desorption/ionization mass spectrometry for rapid identification of Neisseria species. *J. Biomol. Tech.* 19, 200–204.

Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* 17, 994–999. doi: 10.1038/13690

Heckels, J. E., and Williams, J. N. (2010). The influence of genomics and proteomics on the development of potential vaccines against meningococcal infection. *Genom. Med.* 2, 43. doi: 10.1186/gm164

Heinson, A. I., Woelk, C. H., and Newell, M. L. (2015). The promise of reverse vaccinology. *Int. Health* 7, 85–89. doi: 10.1093/inthealth/ihv002

Heinz, E., Selkrig, J., Belousoff, M. J., and Lithgow, T. (2015). Evolution of the translocation and assembly module (TAM). *Genome Biol. Evol.* 7, 1628–1643. doi: 10.1093/gbe/evv097

Ilina, E. N., Borovskaya, A. D., Malakhova, M. M., Vereshchagin, V. A., Kubanova, A. A., Kruglov, A. N., et al. (2009). Direct bacterial profiling by matrix-assisted laser desorption-ionization time-of-flight mass spectrometry for identification of pathogenic *Neisseria*. *J. Mol. Diagn.* 11, 75–86. doi: 10.2353/jmoldx.2009.080079

Ilina, E. N., Oparina, N. Y., Shitikov, E. A., Borovskaya, A. D., and Govorun, V. M. (2010). Molecular surveillance of clinical *Neisseria gonorrhoeae* isolates in Russia. *J. Clin. Microbiol.* 48, 3681–3689. doi: 10.1128/JCM.00565-10

Isabella, V. M., and Clark, V. L. (2011). Deep sequencing-based analysis of the anaerobic stimulon in *Neisseria gonorrhoeae*. *BMC Genom.* 12:51. doi: 10.1186/1471-2164-12-51

Jerse, A. E. (1999). Experimental gonococcal genital tract infection and opacity protein expression in estradiol-treated mice. *Infect. Immun.* 67, 5699–5708.

Jerse, A. E., Bash, M. C., and Russell, M. W. (2014). Vaccines against gonorrhea: current status and future challenges. *Vaccine* 32, 1579–1587. doi: 10.1016/j.vaccine.2013.08.067

Low, N., Unemo, M., Skov Jensen, J., Breuer, J., and Stephenson, J.M. (2014). Molecular diagnostics for gonorrhoea: implications for antimicrobial resistance and the threat of untreatable gonorrhoea. *PLoS Med.* 11:e1001598. doi: 10.1371/journal.pmed.1001598

Maness, M. J., Foster, G. C., and Sparling, P. F. (1974). Ribosomal resistance to streptomycin and spectinomycin in *Neisseria gonorrhoeae*. *J. Bacteriol.* 120, 1293–1299.

McClure, R., Nudel, K., Massari, P., Tjaden, B., Su, X., Rice, P. A., et al. (2015). The Gonococcal Transcriptome during Infection of the Lower Genital Tract in Women. *PLoS ONE* 10:e0133982. doi: 10.1371/journal.pone.0133982

Murray, P. R. (2010). Matrix-assisted laser desorption ionization time-of-flight mass spectrometry: usefulness for taxonomy and epidemiology. *Clin. Microbiol. Infect.* 16, 1626–1630. doi: 10.1111/j.1469-0691.2010.03364.x

Nabu, S., Lawung, R., Isarankura-Na-Ayudhya, P., Isarankura-Na-Ayudhya, C., Roytrakul, S., and Prachayasittikul, V. (2014). Reference map and comparative proteomic analysis of *Neisseria gonorrhoeae* displaying high resistance against spectinomycin. *J. Med. Microbiol.* 63, 371–385. doi: 10.1099/jmm.0.067595-0

Niyompanich, S., Jaresitthikunchai, J., Srisanga, K., Roytrakul, S., and Tungpradabkul, S. (2014). Source-identifying biomarker ions between environmental and clinical *Burkholderia pseudomallei* using whole-cell matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). *PLoS ONE* 9:e99160. doi: 10.1371/journal.pone.0099160

Otto, A., Bernhardt, J., Hecker, M., and Becher, D. (2012). Global relative and absolute quantitation in microbial proteomics. *Curr. Opin. Microbiol.* 15, 364–372. doi: 10.1016/j.mib.2012.02.005

Perez-Cruz, C., Delgado, L., Lopez-Iglesias, C., and Mercade, E. (2015). Outer-inner membrane vesicles naturally secreted by gram-negative pathogenic bacteria. *PLoS ONE* 10:e0116896. doi: 10.1371/journal.pone.0116896

Pettit, R. K., and Judd, R. C. (1992a). Characterization of naturally elaborated blebs from serum-susceptible and serum-resistant strains of *Neisseria gonorrhoeae*. *Mol. Microbiol.* 6, 723–728. doi: 10.1111/j.1365-2958.1992.tb01521.x

Pettit, R. K., and Judd, R. C. (1992b). The interaction of naturally elaborated blebs from serum-susceptible and serum-resistant strains of *Neisseria gonorrhoeae* with normal human serum. *Mol. Microbiol.* 6, 729–734. doi: 10.1111/j.1365-2958.1992.tb01522.x

Phillips, N. J., Steichen, C. T., Schilling, B., Post, D. M., Niles, R. K., Bair, T. B., et al. (2012). Proteomic analysis of *Neisseria gonorrhoeae* biofilms shows shift to anaerobic respiration and changes in nutrient transport and outer membrane proteins. *PLoS ONE* 7:e38303. doi: 10.1371/journal.pone.0038303

Ramakrishnan, V., and White, S. W. (1992). The structure of ribosomal protein S5 reveals sites of interaction with 16S rRNA. *Nature* 358, 768–771. doi: 10.1038/358768a0

Ricci, D. P., and Silhavy, T. J. (2012). The Bam machine: a molecular cooper. *Biochim. Biophys. Acta* 1818, 1067–1084. doi: 10.1016/j.bbamem.2011.08.020

Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., et al. (2004). Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* 3, 1154–1169. doi: 10.1074/mcp.M400129-MCP200

Sarantis, H., and Gray-Owen, S. D. (2012). Defining the roles of human carcinoembryonic antigen-related cellular adhesion molecules during neutrophil responses to *Neisseria gonorrhoeae*. *Infect. Immun.* 80, 345–358. doi: 10.1128/IAI.05702-11

Schmid, O., Ball, G., Lancashire, L., Culak, R., and Shah, H. (2005). New approaches to identification of bacterial pathogens by surface enhanced laser desorption/ionization time of flight mass spectrometry in concert with artificial neural networks, with special reference to *Neisseria gonorrhoeae*. *J. Med. Microbiol.* 54, 1205–1211. doi: 10.1099/jmm.0.46223-0

Schmitter, T., Pils, S., Weibel, S., Agerer, F., Peterson, L., Buntru, A., et al. (2007). Opa proteins of pathogenic neisseriae initiate Src kinase-dependent or lipid raft-mediated uptake via distinct human carcinoembryonic antigen-related cell adhesion molecule isoforms. *Infect. Immun.* 75, 4116–4126. doi: 10.1128/IAI.01835-06

Seib, K. L., Tseng, H. J., McEwan, A. G., Apicella, M. A., and Jennings, M. P. (2004). Defenses against oxidative stress in *Neisseria gonorrhoeae* and Neisseria meningitidis: distinctive systems for different lifestyles. *J. Infect. Dis.* 190, 136–147. doi: 10.1086/421299

Seib, K. L., Wu, H. J., Kidd, S. P., Apicella, M. A., Jennings, M. P., and McEwan, A. G. (2006). Defenses against oxidative stress in *Neisseria gonorrhoeae*: a system tailored for a challenging environment. *Microbiol. Mol. Biol. Rev.* 70, 344–361. doi: 10.1128/MMBR.00044-05

Seib, K. L., Zhao, X., and Rappuoli, R. (2012). Developing vaccines in the era of genomics: a decade of reverse vaccinology. *Clin. Microbiol. Infect.* 18(Suppl. 5), 109–116. doi: 10.1111/j.1469-0691.2012.03939.x

Simons, M. P., Nauseef, W. M., and Apicella, M. A. (2005). Interactions of *Neisseria gonorrhoeae* with adherent polymorphonuclear leukocytes. *Infect. Immun.* 73, 1971–1977. doi: 10.1128/IAI.73.4.1971-1977.2005

Sintsova, A., Sarantis, H., Islam, E. A., Sun, C. X., Amin, M., Chan, C. H., et al. (2014). Global analysis of neutrophil responses to *Neisseria gonorrhoeae* reveals a self-propagating inflammatory program. *PLoS Pathog.* 10:e1004341. doi: 10.1371/journal.ppat.1004341

Steichen, C. T., Shao, J. Q., Ketterer, M. R., and Apicella, M. A. (2008). Gonococcal cervicitis: a role for biofilm in pathogenesis. *J. Infect. Dis.* 198, 1856–1861. doi: 10.1086/593336

Swanson, J., Kraus, S. J., and Gotschlich, E. C. (1971). Studies on gonococcus infection. I. Pili and zones of adhesion: their relation to gonococcal growth patterns. *J. Exp. Med.* 134, 886–906. doi: 10.1084/jem.134.4.886

Tapsall, J. W. (2005). Antibiotic resistance in *Neisseria gonorrhoeae*. *Clin. Infect. Dis.* 41(Suppl. 4), S263–S268. doi: 10.1086/430787

Tapsall, J. W., Ray, S., and Limnios, A. (2010). Characteristics and population dynamics of mosaic penA allele-containing *Neisseria gonorrhoeae* isolates collected in Sydney, Australia, in 2007–2008. *Antimicrob. Agents Chemother.* 54, 554–556. doi: 10.1128/AAC.01108-09

Tseng, H. J., Srikhanta, Y., McEwan, A. G., and Jennings, M. P. (2001). Accumulation of manganese in *Neisseria gonorrhoeae* correlates with resistance to oxidative killing by superoxide anion and is independent of superoxide dismutase activity. *Mol. Microbiol.* 40, 1175–1186. doi: 10.1046/j.1365-2958.2001.02460.x

Unemo, M. (2015). Current and future antimicrobial treatment of gonorrhoea—the rapidly evolving *Neisseria gonorrhoeae* continues to challenge. *BMC Infect. Dis.* 15:364. doi: 10.1186/s12879-015-1029-2

Unemo, M., Golparian, D., Skogen, V., Olsen, A. O., Moi, H., Syversen, G., et al. (2013). *Neisseria gonorrhoeae* strain with high-level resistance to spectinomycin due to a novel resistance mechanism (mutated ribosomal protein S5) verified in Norway. *Antimicrob. Agents Chemother.* 57, 1057–1061. doi: 10.1128/AAC.01775-12

Unemo, M., and Shafer, W. M. (2014). Antimicrobial resistance in *Neisseria gonorrhoeae* in the 21st century: past, evolution, and future. *Clin. Microbiol. Rev.* 27, 587–613. doi: 10.1128/CMR.00010-14

van Baar, B. L. (2000). Characterisation of bacteria by matrix-assisted laser desorption/ionisation and electrospray mass spectrometry. *FEMS Microbiol. Rev.* 24, 193–219. doi: 10.1016/S0168-6445(99)00036-4

Wiese, S., Reidegeld, K. A., Meyer, H. E., and Warscheid, B. (2007). Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. *Proteomics* 7, 340–350. doi: 10.1002/pmic.200600422

Wind, C. M., de Vries, H. J., Schim van der Loeff, M. F., Unemo, M., and van Dam, A. P. (2015). Successful combination of nucleic acid amplification test diagnostics and targeted deferred *Neisseria gonorrhoeae* culture. *J. Clin. Microbiol.* 53, 1884–1890. doi: 10.1128/JCM.00369-15

WHO. (2011). Prevalence and incidence of selected sexually transmitted infections. *Chlamydia trachomatis, Neisseria gonorrhoeae*, syphilis, and *Trichomonas vaginalis*. Methods, and results used by WHO to generate 2005 Estimates. 36. Available at: http://www.who.int/reproductivehealth/publications/rtis/9789241502450/en/ (accessed August 27, 2015).

WHO. (2012). Global action plan to control the spread and impact of antimicrobial resistance in *Neisseria gonorrhoeae*. Available at: http://whqlibdoc.who.int/publications/2012/9789241503501_eng.pdf (accessed May 5, 2015)

Wu, H. J., Seib, K. L., Srikhanta, Y. N., Edwards, J., Kidd, S. P., Maguire, T. L., et al. (2010). Manganese regulation of virulence factors and oxidative stress resistance in *Neisseria gonorrhoeae*. *J. Proteom.* 73, 899–916. doi: 10.1016/j.jprot.2009.12.001

Yoo, J. S., Seong, W. K., Kim, T. S., Park, Y. K., Oh, H. B., and Yoo, C. K. (2007). Comparative proteome analysis of the outer membrane proteins of *in vitro*-induced multi-drug resistant *Neisseria gonorrhoeae*. *Microbiol. Immunol.* 51, 1171–1177. doi: 10.1111/j.1348-0421.2007.tb04012.x

Zhu, W., Chen, C. J., Thomas, C. E., Anderson, J. E., Jerse, A. E., and Sparling, P. F. (2011). Vaccines for gonorrhea: can we rise to the challenge? *Front. Microbiol.* 2:124. doi: 10.3389/fmicb.2011.00124

Zielke, R. A., and Sikora, A. E. (2014). Isolation of cell envelopes and naturally released membrane vesicles of *Neisseria gonorrhoeae*. *Curr. Protoc. Microbiol.* 34, 4A 3 1–4A 3 17. doi: 10.1002/9780471729259.mc04a03s34

Zielke, R. A., Wierzbicki, I. H., Jerse, A. E., Gafken, P. R., Soge, O. O., Holmes, K. K., et al. (2015). Proteomics-driven Reverse Vaccinology for Gonorrhea. Available at: www.abstractsonline.com/Plan/ViewAbstract.aspx?mID=3729&sKey=fa3e7acf-5429-438d-84f7-832c1e0ed4b4&cKey=c67c0637-58c5-4ea2-9872-869ccc94ef34&mKey=224bad71-94ea-4fa5-8df3-f4087bdc3625

Zielke, R. A., Wierzbicki, I. H., Weber, J. V., Gafken, P. R., and Sikora, A. E. (2014). Quantitative proteomics of the *Neisseria gonorrhoeae* cell envelope and membrane vesicles for the discovery of potential therapeutic targets. *Mol. Cell Proteom.* 13, 1299–1317. doi: 10.1074/mcp.M113.029538

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Proteomics As a Tool for Studying Bacterial Virulence and Antimicrobial Resistance

Francisco J. Pérez-Llarena and Germán Bou*

Servicio de Microbiología-INIBIC, Complejo Hospitalario Universitario A Coruña, A Coruña, Spain

Proteomic studies have improved our understanding of the microbial world. The most recent advances in this field have helped us to explore aspects beyond genomics. For example, by studying proteins and their regulation, researchers now understand how some pathogenic bacteria have adapted to the lethal actions of antibiotics. Proteomics has also advanced our knowledge of mechanisms of bacterial virulence and some important aspects of how bacteria interact with human cells and, thus, of the pathogenesis of infectious diseases. This review article addresses these issues in some of the most important human pathogens. It also reports some applications of Matrix-Assisted Laser Desorption/Ionization-Time-Of-Flight (MALDI-TOF) mass spectrometry that may be important for the diagnosis of bacterial resistance in clinical laboratories in the future. The reported advances will enable new diagnostic and therapeutic strategies to be developed in the fight against some of the most lethal bacteria affecting humans.

Keywords: resistance, diagnostic, antibiotic, virulence, proteomics, bacteria

INTRODUCTION

Bacterial diseases continue to be a major cause of death throughout the world as a result of the emergence of new infectious agents, increased transmission due to human migration, and the development of antibiotic resistance (Beceiro et al., 2013). Novel antibiotics and therapies are urgently needed to control these infections, together with new rapid and reliable diagnostic techniques for characterizing resistant strains. Matrix-Assisted Laser Desorption/Ionization-Time-Of-Flight (MALDI-TOF) mass spectrometry techniques may be useful for this task. The present review considers some recent advances in the diagnosis of microbial resistance. Examination of protein profiles by proteomic analysis has become an essential tool for studying the basic mechanisms of bacterial resistance and virulence. This has led to a better understanding of the biology of pathogens that cannot be investigated by reductionist or even genomic studies (e.g., global post-translational protein modifications, subcellular protein location, and protein turnover rates, which are also considered in this review; Chao and Hansmeier, 2012).

Abbreviations: ATCC, American Type Culture Collection; c-di-GMP, cyclic diguanylate; DIGE, difference gel electrophoresis; ICAT, isotope-coded affinity tags; iTRAQ, isobaric tags for relative and absolute quantification; LC–MS, liquid chromatography–mass spectrometry; MIC, minimal inhibitory concentration; MRSA, methicillin-resistant *Staphylococcus aureus*; MS, mass spectrometry; MSSA, methicillin-sensitive *Staphylococcus aureus*; Omp, outer membrane proteins; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; SILAC, stable isotope labeling by amino acids in cell culture; SRM, selected reaction monitoring; TMT, tandem mass tags.

Proteomic techniques are continuously being developed and there is now a great diversity of methods and applications available (Van Oudenhove and Devreese, 2013; Otto et al., 2014). The improved sensitivity of mass spectrometers, together with upgraded sample preparation and protein fractionation technologies, has enabled a more complete study of proteomes. In the past, most quantitative proteomic studies were performed by 2D-PAGE, particularly after the incorporation of immobilized pH gradients and of the difference fluorescent labeling method (DIGE). The gel free method then became more popular and metabolic labeling was included in quantitative analysis. In this labeling method, cells are grown under appropriate conditions in media supplied with either a light or heavy stable isotope of a nutrient, commonly a nitrogen source or an amino acid. The so-called stable isotope labeling by amino acids in cell culture (SILAC) strategy now prevails (Ong et al., 2002). As a substitute to metabolic labeling, differential analysis can be carried out by chemical labeling. Isotope-coded affinity tags (ICAT) covalently label cysteines in extracted proteins with either a light or heavy (deuterium containing) conformation of the ICAT indicator (Gygi et al., 1999). Today, the use of multiple isobaric tags is prevailing among chemical isotope labeling methods. Some products such as isobaric tags for relative and absolute quantification (iTRAQ) and tandem mass tags (TMT) are commercially available (Thompson et al., 2003; Ross et al., 2004). A similar strategy, known as isotope-coded protein label (ICPL), in which both N-termini and lysine side chains are labeled, has been developed for use at the protein level (Schmidt et al., 2005). Recent progress in the development of high throughput and automation of liquid chromatography–mass spectrometry (LC–MS) instruments and in particular progress toward the use of novel algorithms to handle LC–MS data is facilitating the application of quantitative proteomics using label-free strategies. The methods, in which synthetic proteotypic peptides labeled with stable isotopes are used as the ideal internal standard, are becoming particularly popular. Thus, these internal standard peptides are used to measure the total quantity of individual proteins of interest after digestion by selected reaction monitoring (SRM)-MS measurements (Gerber et al., 2003). The various advantages and disadvantages of the different protocols are summarized in **Table 1** (Tiwari and Tiwari, 2014).

The potential application of proteomics to bacterial pathogen research is huge. However, this review will only consider some of these applications and will specifically address bacteria (rather than fungi). The basic knowledge concerning resistance and virulence obtained recently through proteomic technology should be useful for developing new diagnostic and therapeutic applications for the treatment of infectious diseases.

PROTEOMICS FOR STUDYING INTERACTIONS BETWEEN BACTERIA AND ANTIBIOTICS

General Aspects

Antibiotic resistance has become a serious problem in the past two decades and an understanding of the mechanisms of the process is necessary to extend the life of current antibiotics and to enable discovery of novel targets (Lee C. R. et al., 2015). We will describe how antibiotic resistance in bacteria is studied by proteomic techniques. In proteomics, analysis of the total protein content of cells is usually carried out after sublethal contact between selected antibiotics and a pair of isogenic resistant and susceptible bacterial strains. In some cases, intermediate phenotypes are also studied. Resistant strains may be clinical strains or strains obtained *in vitro* (Park et al., 2016). The proteomic response is usually specific to each antibiotic, but proteins involved in energy and nitrogen metabolism, protein and nucleic acid synthesis, glucan biosynthesis, and stress response are often affected (Park et al., 2016). The findings of proteomic studies are commonly confirmed by genomic and/or transcriptomic analysis of the strains and in some cases also by study of the response of strains in which relevant genes are inactivated by gene replacement technology (Lima et al., 2013).

We will describe essential studies and novel findings as well as some proteomic studies involving the main bacterial antibiotic families.

Antibiotics Targeting the Cell Wall
Beta-lactams

Resistance to beta-lactam antibiotics is one of the types of resistance most commonly studied by proteomics methods (Lima et al., 2013). The beta-lactams antibiotics (e.g., penicillin, cephalosporin, carbapenens, monobactam, and beta-lactamase inhibitors) may disturb the synthesis and/or stability of the cell envelope, thus disrupting cell-wall biogenesis and leading to loss of selective permeability and osmotic integrity, finally causing bacterial cell death (Waxman and Strominger, 1983). The main mechanism of resistance to beta-lactam antibiotics is the presence of antibiotic hydrolyzing proteins, known as beta-lactamases (Pérez-Llarena and Bou, 2009). Other important mechanisms include the imbalance in transport proteins such as efflux pumps and porins and alteration in the penicillin binding protein targets (Poole, 2004). The increased use of antibiotics has generally led to the prevalence of some important resistance strains such as penicillin resistant *Streptococcus pneumoniae*, methicillin resistant *Staphylococcus aureus*, and extended spectrum beta-lactamase (ESBL), and cabapenemase-producing *Enterobacteriaceae*, *Pseudomonas aeruginosa,* and *Acinetobacter baummanni* (Boucher et al., 2009).

One of the earliest proteomic studies was an investigation of ampicillin resistant *Pseudomonas aeruginosa*, in which novel porins involved in resistance were discovered (Peng et al., 2005). Study of resistance to piperacillin/tazobactam in *Escherichia coli* has revealed reduced expression of porin OmpX and increased expression of TolC (Dos Santos et al., 2010). In the case of the penicillin-tolerant Gram-positive *Streptococcus pyogenes*, overexpression of murein metabolism proteins and general alteration of bacterial physiology were observed (Chaussee et al., 2006).

Alterations in cell physiology and overexpression of catalase and superoxide dismutase were detected in methicillin resistant *S. aureus*. The role of alanine dehydrogenase was indicated as being important in antibiotic resistance (Monteiro et al., 2012).

TABLE 1 | Different quantitative proteomic approaches with the associated advantages and disadvantages.

Proteomic tools	Advantage	Disadvantage
GEL-BASED METHODS		
2DE	(1) Simple (2) Robust (3) Suitable for MS analysis	(1) Involves large amount of sample (2) Low throughput (3) Poor recovery of hydrophobic proteins (4) High inter-gel variability
2-DIGE	(1) Multiplexing (2) Better quantitation (3) Minimal gel to gel variation	(1) Expensive Cy dyes (2) Poor recovery of hydrophobic proteins (3) Difficulty in separating low molecular weight compounds
GEL-FREE METHODS		
SILAC	(1) High throughput (2) Robust (3) Sensitive and simple	(1) Only suitable for tissue culture models (2) Costly reagents (3) Not applicable to tissue sample
ICAT	(1) Selectively isolates peptide (2) Compatible with any amount of protein (3) Complexity of the peptide mixture is reduced	(1) Cannot identify proteins with less than eight cysteines (2) Large ICAT label ($\approx$500 Da) (3) Post-translational modification information is frequently lost
iTRAQ	(1) Applicable to versatile samples (2) Multiplexing (3) Better quantitation	(1) Involves high amount of sample (2) Incomplete labeling (3) Expensive reagents
ICPL	(1) High-throughput quantitative proteome profiling on a global scale (2) Able to detect post-translational modifications and protein isoforms (3) Applicable to protein samples such as tissue extracts or body fluids	(1) Isotopic effect of deuterated tags interferes with retention time of the peptides labeled during LC
Label-free	(1) Involves less amount of sample (2) Higher proteome coverage (3) Avoids labeling	(1) High throughput instrumentation (2) Not suitable for low abundant proteins (3) Incomplete digestion may introduce error (4) Multiplexed analysis not possible in one experiment
SRM	(1) Highly sensitive, quantitatively accurate and highly reproducible (2) Protein detection is relatively rapid and straightforward (3) Enables detection of non-abundant (>10 ng/ml) proteins (4) Quantification of post-translational modifications	(1) Limited broad scale application because of difficulty in generating high-quality SRM assay (2) Sensitivity is not comparable to immunological assays (3) Detection and quantification of non-abundant proteins (i.e., ~10 ng/ml or less)

Obtained from Tiwari and Tiwari (2014) and reprinted with permission from the publisher.

Adaptation to oxacillin in *S. aureus* has recently been investigated (Solis et al., 2014). These authors concluded that proteins involved in capsule formation, peptidoglycan biosynthesis, and wall remodeling are regulated in response to antibiotics. Spectral counting-based label-free quantitative proteomics has been applied to study global responses in methicillin-resistant *Staphylococcus aureus* (MRSA) and methicillin susceptible *S. aureus* treated with subinhibitory doses of oxacillin (Liu et al., 2014). Beta-lactamase and penicillin-binding protein 2a were uniquely upregulated in oxacillin-treated MRSA (**Table 2**). Analysis of the inner membrane fraction of carbapenem resistant *A. baumanni* has shown an association with beta-lactamase AmpC and OXA-51 production as well as metabolic enzymes, elongation factor Tu, and ribosomal proteins (Tiwari et al., 2012; Tiwari and Tiwari, 2014).

Glycopeptides

The glycopeptide vancomycin acts by inhibiting peptidoglycan synthesis. It binds to the DAla-DAla terminus of the nascent peptidoglycan, thus blocking its correct synthesis.

In *Enterococcus* spp., a substitution of the DAla residue from peptidoglycan termini by D-lactose or D-Serine has been detected as the main mechanism of resistance to vancomycin. In *S. aureus*, a more complex scenario involving different enzymes and gene clusters implicated in vancomycin resistance has been proposed. Some resistant strains such as vancomycin resistant *Staphylococcus aureus* (VRSA) and vancomycin resistant *enterococci* (VRE) are of serious clinical concern (Lima et al., 2013).

In the first proteomic study of vancomycin resistant *Enterococcus faecalis*, Wang et al. (2010) examined a reference strain (V583) and a clinical isolate (V309) in the presence and absence of vancomycin. These authors found that proteins involved in vancomycin resistance functions, virulence factors, stress, metabolism, translation, and conjunction were regulated. Proteomic profiles of vancomycin resistant *E. faecium* SU18 strain treated and not treated with vancomycin have recently been obtained (Ramos et al., 2015). Fourteen proteins were differentially expressed in SU18. Proteins involved in the vancomycin resistance mechanisms were upregulated in the

TABLE 2 | Pathway enrichment study by Database for Annotation, Visualization, and Integrated Discovery (DAVID) of the differentially expressed proteins in oxacillin-treated MRSA and MSSA compared with their untreated controls.

Pathway	Oxacillin treated MRSA Control	Oxacillin treated MSSA Control
Alanine, aspartate, and glutamate metabolism	–	
Beta-lactam resistance	+	
Peptidoglycan biosynthesis	+	+
Pantothenate and CoA biosynthesis	+	+
Aminoacyl-tRNA biosynthesis		–
Pentose and glucuronate interconversions		–
ABC transporters		–
Porphyrin and chlorophyll metabolism		–
Two component system		+
Nitrogen metabolism		+
Oxidative phosphorylation		+
RNA degradation		+
Nicotinate and nicotinamide metabolism		+
Arginine and proline metabolism		+
Ribosome		+
Pyrimidine metabolism		+
Purine metabolism		+
Nucleotide excision repair		+
Pyruvate metabolism		+
Citrate cycle (TCA cycle)		+

Obtained from Liu et al. (2014) and reprinted with permission from the publisher. (+) Upregulated; (–) Downregulated.

presence of vancomycin, while metabolism-related proteins were downregulated, leading to compensatory effects. Differential expression of proteins has been observed in vancomycin resistant *S. aureus*, thus distinguishing vancomycin intermediate (VISA) type strain Mu50 and vancomycin resistant strains (Drummelsmith et al., 2007). More recently, the proteomic profile of a group of heterogeneous vancomycin-intermediate *Staphylococcus aureus* (hVISA) was compared with that of vancomycin susceptible *S. aureus* (Chen et al., 2013). The study initially detected five upregulated proteins in hVISA, although only one was definitely confirmed by real time quantitative reverse transcription PCR (qRT-PCR): the protein encoded by the isaA gene involved in cell wall biogenesis.

Antibiotics Targeting Protein Synthesis
Chloramphenicol

Chloramphenicol acts by binding to the 50 S ribosome subunit. Three mechanisms of resistance to chloramphenicol are known: reduced membrane permeability, mutation of the 50S ribosomal subunit, and production of chloramphenicol acetyltransferase (Civljak et al., 2014). Li et al. (2007) observed differential expression of 10 membrane proteins, including TolC, OmpC, OmpW, and OmpT, in chloramphenicol resistant *E. coli*. A more recent study demonstrated overexpression of two different efflux

pumps in chloramphenicol resistant strains of *Burkholderia thailandensis* obtained in the laboratory (Biot et al., 2011).

Linezolid

In the case of linezolid, an oxazolidinone that binds to the 23S rRNA (Ribosomal ribonucleic acid), different resistance mechanisms have emerged, including increased expression of ABC transporters, mutations in 23S rRNA, mutations in ribosomal proteins L3 and L4, and mutations in an RNA methyltransferase (Lee C. R. et al., 2015). Proteomic and transcriptomic screening of linezolid revealed a possible increase in the metabolism and transport of carbohydrates in some linezolid-resistant *S. pneumoniae* mutants (Feng et al., 2011). Several glycolytic proteins were overexpressed in the resistant strains, along with other enzymes and transporters involved in sugar metabolism.

Tetracyclines

The tetracycline family of antibiotics inhibits aminoacyl tRNA binding to the mRNA-ribosome complex. Cells become resistant to tetracycline via at least three mechanisms: enzymatic inactivation of tetracycline, efflux, and ribosomal protection (Falagas et al., 2015).

Yun et al. (2008) used 2-DE/MS-MS (Two-dimensional tandem mass spectrometry) and 1-DE/LC/MS-MS (One-dimensional Liquid chromatography tandem mass spectrometry) techniques to study the surface proteome of *A. baumannii* DU202 (which is highly resistant to tetracycline) after treatment with subminimal inhibitory concentrations (subMIC) of tetracycline. These authors observed that OmpA38, CarO, OmpW, and other Omps were increasingly secreted on exposure to tetracycline. This indicates an important role for Omps in overcoming antibiotic-induced stress.

A parallel proteomic approach has been applied in the analysis of sarcosine-insoluble outer membrane fraction of *P. aeruginosa* responding to ampicillin, kanamycin, and tetracycline resistance. The effects on expression levels were variable and expression of OprG decreased as expression of OprF, OprI, Omp, and MexA increased (Peng et al., 2005). In an interesting study, Lin et al. (2014) labeled the differential proteome of *E. coli* K12 BW25113 in response to chlortetracycline stress with isobaric tags and applied quantitative proteomics technology for relative and absolute quantitation of the labeling. Crucial metabolic pathways such as the tricarboxylic acid cycle, pyruvate metabolism, and glycolysis/gluconeogenesis fluctuated greatly. The ribosome protein complexes contributing to the translation process were generally elevated under conditions of chlortetracycline stress, which is known to be a compensatory mechanism caused by the action of chlortetracycline on the ribosome.

Aminoglycosides

The aminoglycoside antibiotic family interrupts protein synthesis by blocking the small 16S subunit of the bacterial ribosome. Three mechanisms of aminoglycoside resistance are known: reduced uptake or decreased cell permeability, alterations at the ribosomal binding sites, and production of aminoglycoside modifying enzymes (Jackson et al., 2013).

In an initial study, a subproteomic approach was used to characterize the outer membrane proteins of streptomycin resistant *E. coli*. TolC, OmpT, and LamB were upregulated and FadL, OmpW and a location-unknown protein Dps were downregulated in the streptomycin-resistant *E. coli* strain (Li et al., 2008). Another study analyzed and compared the protein profile of whole cell extracts from *Mycobacterium tuberculosis* clinical isolates susceptible and resistant to streptomycin. On comparing 2DE patterns, nine proteins were consistently overexpressed in streptomycin resistant isolates and were identified. Moreover, *in silico* docking analysis revealed significant interactions between these proteins and streptomycin (Sharma et al., 2010). In a native/SDS-PAGE based proteomic study, low levels of NarG and NarH, two components of respiratory nitrate reductase (Nar), were observed in streptomycin, gentamicin, ceftazidime, tetracycline, and nalidixic acid-resistant *E. coli* strains (Ma et al., 2013).

Nabu et al. (2014) compared the protein expression profiles of a high-level spectinomycin-resistant (clinical isolate) and a susceptible (reference strain) *Neisseria gonorrhoeae* after treatment with subminimal inhibitory concentrations (subMICs) of spectinomycin. The 50S ribosomal protein L7/L12, an essential component for ribosomal translocation, was over-expressed in both strains, indicating that compensatory mechanisms may work in response to antibiotics that inhibit protein synthesis. Proteomics techniques have been used to establish the effects of gentamicin on the proteomes of aerobic and oxygen-limited *E. coli* (Al-Majdoub et al., 2013). Ribosomal proteins L1, L9, L10, and S2 were upregulated under both conditions, and the authors postulated that these are candidate drug targets for the development of synergistic combinations with gentamicin.

Kanamycin and amikacin resistant isolates of *M. tuberculosis* have been studied by proteomic analysis. Twelve proteins, two of which are of unknown function, were upregulated in both antibiotic-resistant isolates. All of the proteins were cellular proteins. Kanamycin and amikacin interacted correctly with the proteins according to molecular docking studies. Kumar et al. (2013) suggested that two of them were putative iron regulation/metabolism related proteins, thereby indicating the role of iron in conferring resistance to second-line drugs. In a recent proteomic and western blotting study of the *E. coli* K-12 outer membrane (OM) proteins involved in kanamycin resistance, Li H. et al. (2015) observed upregulation of some OM proteins such as Tolc, TsX, and OstA, and downregulation of MipA, OmpA, FadL, and OmpW OM proteins in the kanamycin resistant *E. coli* K-12 strain. These authors concluded that MipA is a novel OM protein implicated in antibiotic resistance.

Macrolides

Macrolide antibiotics act by binding reversibly to the P site on the subunit 23 S of the bacterial ribosome. The primary means of bacterial resistance to macrolides is by post-transcriptional methylation of the 23S bacterial ribosomal RNA. Two other rarely observed types of acquired resistance include the production of drug-inactivating enzymes (esterases or kinases) and the production of active ATP (Adenosine triphosphate)-dependent efflux proteins that transport the drug outside of the cell (Cornick and Bentley, 2012).

In an early study, Cash et al. (1999) examined the proteins synthesized by erythromycin-susceptible and erythromycin-resistant *S. pneumoniae*. These authors used peptide mass mapping to identify a 38,500 Dalton protein upregulated in resistant strains as glyceraldehyde-3-phosphate dehydrogenase (GAPDH). Considering this as the possible cause of resistance against erythromycin, the authors proposed an increment in energy production for the efflux system.

In comparative proteomic analysis of isolated sarcosine-insoluble outer membrane protein (OMP) fractions from clarithromycin-susceptible and resistant *Helicobacter pylori* strains, Smiley et al. (2013) showed that iron-regulated membrane protein, UreaseB, EF-Tu, and a putative OMP were downregulated; the HopT (BabB) transmembrane protein, HofC and OMP31 were upregulated in clarithromycin-resistant *H. pylori*, revealing that alteration of the outer membrane protein profile may be a novel mechanism involved in clarithromycin resistance in *H. pylori*.

Antibiotics Targeting DNA or RNA Synthesis

Metronidazole

Metronidazole is an antibiotic of the nitroimidazole class that inhibits nucleic acid synthesis by disrupting the DNA of microbial cells. Multiple mechanisms or resistance have been described in *Bacteroides fragilis* and *H. pylori* (Chong et al., 2014)

After analysis of the protein profiles of a derivative of *H. pylori* strain 26695, which is resistant to moderate levels of metronidazole, McAtee et al. (2001) proposed that the ability of the mutant strain to increase various isoforms of alkylhydroperoxide reductase during exposure to metronidazole is critically important in producing the resistant phenotype. In a metronidazole-resistant strain derived from *B. fragilis* ATCC 25285, the proteomic changes affected a wide range of metabolic proteins, including lactate dehydrogenase (upregulated) and flavodoxin (downregulated), which may be involved in electron transfer reactions: the enzymatic activity of the pyruvate-ferredoxin oxidoreductase (PorA) complex was also found to be impaired (Diniz et al., 2004).

Alterations in the metabolic pathway involving pyruvate-ferredoxin oxidoreductase were also found in a multidisciplinary analysis of a non-toxigenic *Clostridium difficile* strain with stable resistance to metronidazole (Moura et al., 2014).

Another recent proteomic analysis revealed regulation of DNA repair proteins, putative nitroreductases and the ferric uptake regulator (Fur) in a NAP1 *C. difficile* clinical isolate resistant to metronidazole (Chong et al., 2014).

Rifampicin

Rifampicin inhibits bacterial DNA-dependent RNA synthesis by inhibiting bacterial DNA-dependent RNA polymerase. Resistance to rifampicin arises from mutations that alter residues of the rifampicin binding site on RNA polymerase, resulting in decreased affinity for rifampicin. Resistant mutations map to the

rpoB gene, encoding RNA polymerase beta subunit (Goldstein, 2014).

Neri et al. (2010) found that 23 proteins were differentially expressed in two rifampicin resistant and one susceptible meningococcus. Increased expression of some of the proteins involved in the major metabolic pathways, mainly pyruvate catabolism and the tricarboxylic acid cycle, was observed. Decreased expression of proteins involved in gene regulation and in polypeptide folding was also observed. Sandalakis et al. (2012) analyzed rifampicin resistance in a rifampicin resistant strain of *Brucella abortus* 2308 developed *in vitro*. The resistant strain contained 39 differentially regulated proteins, most of which are involved in various metabolic pathways. The authors suggested that, apart from mutations in the rpoB gene, rifampicin resistance in *Brucella* mainly involves excitation of several metabolic processes and possible use of the already existing secretion mechanisms at a more efficient level.

Quinolones

The fluoroquinolones are the most commonly used family of quinolones in clinical settings. First and second generation fluoroquinolones selectively inhibit the topoisomerase II ligase domain. Third and fourth generation fluoroquinolones are more selective for the topoisomerase IV ligase domain, and thus have enhanced Gram-positive coverage. Three mechanisms of resistance to quinolones are known: efflux pumps; proteins that can bind to DNA gyrase, protecting it from the action of quinolones; and mutations at key sites in DNA gyrase or topoisomerase IV, which decrease their binding affinity and thus decrease the effectiveness of the drug (Redgrave et al., 2014). Two early proteomic studies considered the effects of fluoroquinolones on *P. aeruginosa* and *Salmonella enterica*. In the first of these, expression of a probable ATP-binding component of ATP binding cassette (ABC) transporter was observed in ciprofloxacin-intermediate and resistant strains, but not in the sensitive strain (Zhou et al., 2006). The authors of the second study, involving the *S. enterica* multiple antibiotic resistant strain, suggested that the increased expression of AcrAB/TolC was associated with resistance, while increases in e.g., F1F0-ATP synthase and Imp were a response to fluoroquinolone (Coldham et al., 2006). In a proteomic study of nalidixic acid (NA) resistance in *E. coli*, Lin et al. (2008) observed upregulation of TolC, OmpT, OmpC, and OmpW and downregulation of FadL in resistant strains. In a broader search for mechanisms at the protein level that confer resistance to fluoroquinolones, Vranakis et al. (2011) compared the proteomes of fluoroquinolone-susceptible *Coxiella burnetii* and fluoroquinolone-resistant samples of the bacterium (developed *in vitro*). The study revealed differential expression of 15 bacterial proteins involved in different cellular processes, suggesting that the mechanism of antibiotic resistance in the bacterium is a multifaceted process.

Antibiotics Targeting Cell Membranes
Daptomycin

Daptomycin is a lipopeptide that interacts with phosphatidylglycerol in the bacterial membrane, leading to the formation of holes that leak intracellular ions. Until now,

specific genetic determinants of the daptomycin-resistant strains remain to be elucidated (Lee C. R. et al., 2015). Fischer et al. (2011) studied an isogenic daptomycin-susceptible and daptomycin-resistant pair of *S. aureus* strains (616 and 701) by using comparative proteomics of 616 vs. 701, revealing different concentrations of proteins in various functional categories, including cell wall-associated targets and biofilm formation proteins.

Colistin

Colistin is an antimicrobial peptide that interacts with the bacterial outer membrane, by displacing bacterial counter ions in the lipopolysaccharide (LPS). Hydrophobic/hydrophilic regions interact with the cytoplasmic membrane just like a detergent, solubilizing the membrane. The most common mechanisms of resistance to colistin are modifications to LPS (Bialvaei and Samadi Kafil, 2015).

Fernández-Reyes et al. (2009) induced resistance in the susceptible *A. baumannii* ATCC 19606 by growing the strain under increasing colistin pressure. These researchers then carried out 2-D difference gel electrophoresis (DIGE) experiments and identified 35 proteins that were expressed differently in the two phenotypes. Most of the proteins appearing were downregulated in the colistin-resistant strain. These include outer membrane (OM) proteins, chaperones, protein biosynthesis factors, and metabolic enzymes, indicating an important loss of biological fitness in the resistant phenotype. In a later study, Pournaras et al. (2014) sequentially collected samples from two colistin-susceptible/colistin-resistant (Col(s)/Col(r)) pairs of *A. baumannii* strains assigned to international clone 2, which is prevalent worldwide after extended exposure to colistin. These researchers observed underexpression of the protein CsuA/B and C from the csu operon in Col(r) isolate Ab347, relative to its Col(s) counterpart Ab299.

The biofilm efficiency of the *A. baumannii* 19606 type strain depends on the formation of pili, cell-surface appendages assembled via the CsuAB-A-B-C-D-E chaperone–usher secretion system, according to Tomaras et al. (2008). Chaperone usher (CU) pili are linear polymers made of subunits capable of either self-polymerization or assembly with other subunits. The biogenesis of CU fibers need a periplasmic chaperone and outer membrane assembly platform named the usher. Among the three chaperone systems, the archaic systems are associated with bacteria that cause some of the most serious diseases in humans, animals, and plants. This pilus is formed from four subunits, namely CsuA/B, CsuA, CsuB, and CsuE, and is assembled using the CsuC-CsuD chaperone-usher secretion machinery (Pakharukova et al., 2015).

In addition, the Col(r) isolate Ab347 underexpressed aconitase B and some enzymes implicated in the oxidative stress response, such as KatE catalase, superoxide dismutase, and alkyl hydroperoxide reductase. This possibly suggests a limited response to reactive oxygen species (ROS) and, therefore, impaired colistin-mediated cell death by means of hydroxyl radical production (**Table 3**).

The low intracellular c-di-GMP level in dispersed cells of a *P. aeruginosa* strain coincided with increased expression of proteins

TABLE 3 | Proteins involved in antibiotic resistance and virulence differentially expressed in Colr strain Ab347.

Accession no.	No. of peptides[a]	Score[b]	ANOVA[c] (P)	Fold change[d]	Description
ABYAL1639/katE	39	2919	1.78×10^{-13}	−35.39	Catalase hydroperoxidase II
ABYAL3026	4	207	5.00×10^{-15}	−28.34	Putative porin protein associated with imipenem resistance, CarO
ABYAL2670	6	549	3.27×10^{-13}	−27.38	Putative protein CsuA/B
ABYAL 2667	3	130	1.24×10^{-6}	−14.64	Putative protein CsuC
ABYAL2566/acnB	9	410	1.64×10^{-9}	−3.13	Fragment of aconitate hydratase (part3)
ABYAL2567/acnB	7	332	1.43×10^{-7}	−2.46	Fragment aconitate hydratase 2 (part 2)
ABYAL3715/sodC	7	376	1.41×10^{-4}	−2.15	Superoxide dismutase precursor (Cu−Zn)
ABYAL1416/ahpC	14	881	1.81×10^{-8}	−2.13	Alkyl hydroperoxide reductase (detoxification of hydroperoxides)
ABYAL2568/acnB	7	291	3.17×10^{-6}	−2.12	Fragment of aconitate hydratase 2 (part 1)
ABYAL1790/bla_{OXA-66}	3	104	2.17×10^{-7}	−2.03	Carbapenem-hydrolyzing oxacillinase Oxa-66
ABYAL0009	9	578	1.08×10^{-9}	+3.15	Putative RND-type efflux pump involved in aminoglycoside resistance (AdeT)

[a]Number of peptides for identification and quantification.
[b]Confidence score for identification by Mascot software.
[c]ANOVA, analysis of variance.
[d]Negative values indicate underexpression in Col^r strain Ab347.
Obtained from Pournaras et al. (2014) and reprinted with permission from the publisher.

required for the virulence and development of antimicrobial peptide resistance in *P. aeruginosa* (Chua et al., 2013), and *P. aeruginosa* cells with low c-di-GMP levels were consequently found to be more resistant to colistin than *P. aeruginosa* cells with high c-di-GMP levels.

Chopra et al. (2013) analyzed the proteome of two strains of *A. baummanii*: the multidrug resistant *A. baummanii* strain BAA-1605 and the drug sensitive strain ATCC 1798. The analysis was performed by by iTRAQ labeling and online 2D LC/MS/MS for peptide/protein identification. A significant number of proteins were overexpressed at least twofold in the multidrug resistant strains including drug-, antibiotic-, and heavy metal-resistance proteins, porins, lipoproteins, stress-related proteins, membrane transporters, proteins important for acquisition of foreign DNA, cell-wall, and exopolysaccharide-related proteins, biofilm-related proteins, metabolic proteins, and many with no annotated function. However, the porin OmpW, less abundant in carbapenem- and colistin-resistant *A. baumannii* strains, was overexpressed by three times in BAA-1605.

Conclusions

Greater knowledge of the specific mechanisms involved in bacterial resistance is needed to improve the treatment of diseases caused by infectious bacteria and thus control the survival of recalcitrant populations. This interesting area of research is complicated by the fact that multiple, super-imposed and/or balancing resistance mechanisms coexist in the same bacterial species.

Most proteomic analysis of antibiotic resistance can be classified into two broad groups: comparison of resistant and susceptible bacteria, and bacterial responses to the presence of antibiotics (**Table 4**). In general, in the first type of studies the least abundant proteins are related to secretion and metabolism, specifically OmpW, which is a known bacteriocin receptor. The most commonly expressed proteins are those involved in cell wall biogenesis, known resistance mechanisms, polysaccharide metabolism, and transport. The most frequently cited of these is Tolc, an outer membrane channel that participates in several efflux systems. In the second case, analysis of the bacterial response to antibiotic challenge revealed that the proteins most commonly affected are chaperone proteins and proteins involved in stress response, amino acid metabolism and energy metabolism. Some proteins involved in amino acid and energy metabolism are overexpressed while others are underexpressed, indicating the problem of using broad functional classifications and also the complexity of antibiotic response (**Table 4**). Some of the most revealing analysis coupled both previously described approaches, to investigate the response of both susceptible and resistant strains to antibiotic exposure. Multivariable approaches, i.e., those combining several strains, growth conditions, concentrations and types of antibiotics, and time points, may contribute to a better understanding of bacterial pathways and systems relevant to resistance.

Some unique mechanisms or groups of proteins are expressed at low or indiscernible levels in the unchallenged resistant organisms, but are regulated in response to antibiotic exposure. These represent potentially useful targets in new therapies against resistance.

ANALYSIS OF ANTIMICROBIAL RESISTANCE FOR DIAGNOSTIC PURPOSES

General Aspects

The main purpose of a clinical microbiology laboratory is to provide reliable information as quickly as possible about the etiological agents responsible for infections and their sensitivity to antibiotics. The information obtained should allow enable selection of the most appropriate antimicrobial therapy to improve the care provided to the patient while at the same time

TABLE 4 | Representative proteomic techniques for studying antibiotic resistance.

Proteomic technique (s)	Antibiotic	Physiological effects	Experimental approach	Pathogen (s)	References
Two dimensional gel electrophoresis (2-DE) and tandem mass spectrometry	Penicillin	Growth phase, stress, and fatty acid biosynthesis (FAB) proteins expression altered	Wild type and *rgg* isogenic mutant with or without exposure to penicillin	*Streptococcus pyogenes*	Chaussee et al., 2006
2D-fluorescence difference gel electrophoresis (2D-DIGE)	Piperacillin/tazobactam	Bacterial virulence, antibiotic resistance, DNA protection, and multidrug efflux pump expression associated with resistance	Laboratory derived resistant strain and susceptible wild type	*Escherichia coli*	Dos Santos et al., 2010
iTRAQ	Oxacillin	LytR-CPsA-PsR (LCP) proteins, capsule, peptidoglycan biosynthesis, cell wall remodeling and urease proteins associated with oxacillin resistance	Methicillin resistant strain compared to isogenic oxacillin adapted strain	*Staphylococcus aureus*	Solis et al., 2014
2D-DIGE	Carbapenem	Beta-lactamases, energy, and protein production enzymes are upregulated; OmpW and surface antigen downregulated	Carbapenem resistant strains and wild type strain	*Acinetobacter baumannii*	Tiwari et al., 2012
2-DE and LC-MS/MS	Vancomycin	Vancomycin resistance proteins upregulated; metabolism-related proteins downregulated	Vancomycin resistant strain with or without exposure to vancomycin	*Enterococcus faecium*	Ramos et al., 2015
SDS-PAGE electrophoresis and LC-MS/MS	Chloramphenicol	Overexpression of efflux pump systems associated with resistance	Laboratory derived resistant strain and susceptible wild type	*Burkholderia thailandensis*	Biot et al., 2011
2-DE and iTRAQ	Linezolid	Metabolism and transport of carbohydrates involved in resistance to linezolid	Linezolid resistant and wild type strains	*Streptococcus pneumoniae*	Feng et al., 2011
2-DE/MS-MS and 1-DE /LC/MS-MS	Tetracycline	Outer membrane proteins decreased expression in membrane and increased secretion	Response to sub-minimal inhibitory concentrations of tetracycline	*Acinetobacter baumannii*	Yun et al., 2008
2-DE and LC-MS/MS	Kanamycin	Outer membrane protein expression altered. Identification of novel membrane MipA protein involved in antibiotic resistance	Laboratory derived resistant and susceptible strains	*Escherichia coli*	Li H. et al., 2015
2-DE and LC-MS/MS	Erythromycin	Glyceraldehyde-3-phosphate dehydrogenase upregulation in resistant strain	Susceptible and resistant strains	*Streptococcus pneumoniae*	Cash et al., 1999
iTRAQ and 2D-LC-MS/MS	Metronidazole	RecA, ferric uptake regulator (Fur), putative nitroreductases and altered expression of stress-related proteins	Susceptible and resistant strains before and after treatment with antibiotic	*Clostridium difficile*	Chong et al., 2014
2-DE and LC-MS/MS	Fluoroquinolone (ciprofloxacin)	Overexpression of ATP-binding component of ATP binding cassette (ABC)	Laboratory derived resistant and susceptible strains	*Pseudomonas aeruginosa*	Zhou et al., 2006
2-DE and LC-MS/MS	Rifampicin	Alterations in several metabolic processes and secretion mechanisms	Laboratory derived resistant and susceptible strains	*Brucella abortus*	Sandalakis et al., 2012
iTRAQ and IPG-isoelectric focusing with LC-MS	Daptomycin	Differences in biofilm formation proteins, cell wall-associated targets	Isogenic daptomycin susceptible and resistant strain pair	*Staphylococcus aureus*	Fischer et al., 2011
2-DE and LC-MS/MS	Colistin	Outer membrane proteins, chaperones, protein biosynthesis factors and metabolic enzymes downregulated associated with loss of biological fitness	Laboratory derived resistant and susceptible strains	*Acinetobacter baumannii*	Fernández-Reyes et al., 2009

contributing to better control of resistance and also to decreasing expenditure on antibiotics (DeMarco and Ford, 2013).

Recent findings indicate the potential usefulness of mass spectrometry techniques, in particular MALDI-TOF MS, to identify specific mechanisms of resistance. These techniques are less expensive and produce results more quickly than the currently used methods. The applications of MALDI-TOF MS for diagnosis of antibiotic resistance have been discussed in two recent excellent reviews (Hrabák et al., 2013; Kostrzewa et al., 2013). These methods are at the initial stages of evaluation

and some are only useful for centers of excellence and research laboratories.

Here, we will try to summarize and classify methods available with some valuable examples and providing the latest contributions. MALDI-TOF MS applications in detection of antimicrobial resistance can be classified depending on the type of target and the methodology into the following approaches.

Identification of the Entire Cell Profile

Identification of the entire cell profile involves establishing differences in the whole protein spectra profile of susceptible and resistant strains. This approach has been used to identify methicillin-resistant *S. aureus* (MRSA), with the first attempts being carried out by Edwards-Jones et al. (2000) at Manchester University. MRSA and MSSA strains were subsequently successfully distinguished using a protein Chip array (SELDI-TOF, Surface-enhanced laser desorption/ionization; Shah et al., 2011). Retrospective typing of an MRSA outbreak showed that it is possible to differentiate unrelated MSSA, MRSA and borderline resistant *S. aureus* (BORSA) strains.

MALDI-TOF MS has been also used to detect vancomycin-resistant enterococci. The method was used to detect vanB positive *Enterococcus faecium* with a high sensitivity and specificity (Griffin et al., 2012), and vanA positive *E. faecium* has recently been detected (Nakano et al., 2014). Similar findings have recently been reported for an important Gram-negative anaerobic pathogen, *B. fragilis*. Two groups (division I and II), division II harboring the gene cfiA, encoding a potent metallo-beta-lactamase have been differentiated by a specific peak in their MALDI-TOF profile spectra (Nagy et al., 2011; Wybo et al., 2011).

MALDI-TOF MS techniques have also been used by different researchers to differentiate beta-lactamase genes carrying strains of *Enterobacteriaceae* and *P. aeruginosa*. Dubska et al. (2011) used SELDI-TOF MS, but were unable to clearly differentiate different beta-lactamases. However, Camara and Hays (2007) detected a peak corresponding to a beta-lactamase in *E. coli* strains. Overall the results suggest that the routine detection of beta-lactamase producing strains is not yet possible in clinical laboratories.

Identification of an Antibiotic and Product of Hydrolysis

This promising approach has been reported for the important clinical carbapenemases and extended-spectrum beta-lactamases. The mass shift that occurs during the addition of a water residue in the beta-lactam molecule after enzymatic hydrolysis can be detected by MALDI-TOF MS. The beta-lactamase negative strains do not modify the molecular weight of the beta-lactam (Hrabák et al., 2013; Kostrzewa et al., 2013). Moreover, the procedure enables quantitative analysis that is useful for direct comparison with MIC-values, and it also provides rapid resolution. Furthermore, the method can be improved by using beta-lactamase inhibitors to identify specific types of beta-lactamase and it is applicable to positive blood cultures. The main limitation is the interference with other resistance mechanisms such as porins and efflux pumps when

intact cells are used; however, this can be overcome by using different lysis methods. Analysis of the raw spectra can also be difficult for unexperienced microbiologists.

The first two studies involving direct carbapenemase detection were both published in 2011 (Burckhardt and Zimmermann, 2011; Hrabák et al., 2011). The prevailing approach consists of the incubation of a pellet from a bacterial culture, usually grown overnight, with the beta-lactam antibiotic. After incubation of the culture for 1–3 h at 35°C, the supernatant is analyzed by MALDI-TOF MS (**Figure 1**). The addition of NH_4HCO_3 increased the sensitivity of detection of OXA-48 (Papagiannitsis et al., 2015; Studentova et al., 2015). The method has also been extended to OXA-51 and SIM-1 carbapenemases (Lee et al., 2013). MALDI-TOF MS can be used detect OXA- and GES-carbapenemases (Chong et al., 2015). The method has been used to detect carbapenemase resistance associated with cifA encoded metallo-beta-lactamase in blood samples (Johansson et al., 2014). The capacity of these techniques for detecting carbapenemases has recently been enhanced by different research groups (Álvarez-Buylla et al., 2013; Peaper et al., 2013; Wang et al., 2013). The method has been shown to be useful in detecting ESBL-producing *Enterobacteriaceae* from positive blood cultures, with 99% of sensitivity within a maximum of 150 min, with cefotaxime and ceftazime antibiotics (Oviaño et al., 2014).

Some MALDI-TOF MS based methods can also be used to detect modifications that occur in resistance to aminoglycosides, such as inactivation by acetyltransferases (Kostrzewa et al., 2013), with poor reproducibility and rRNA methyltransferase (Hrabák et al., 2013). Overall these data suggest that MALDI-TOF can be used to detect aminoglycoside resistance, although these methods are generally limited to research laboratories.

Detection of Resistance Proteins within the Cell

In this case, MALDI-TOF MS can be used to help detect some microbial biomarkers (essentially proteins or their fragments, obtained after trypsin digestion) that confer resistance to the pathogen.

For example, methicillin-resistant *S. aureus* positive for agr (accessory gene regulator) and harboring the class A mec complex was identified by detection of the small peptide called PSM-mec in whole cells (Josten et al., 2014).

Detection of antibiotic resistance markers, such as peptide fragments, in clinical strains of *E. coli* associated with some beta-lactamases (CTX-M1, CMY-2, VIM, and TEM) together with kanR and aminoglycoside modifying enzyme was obtained by periplasm extraction followed by nano-LC separation (Hart et al., 2015) Capillary-electrophoresis mass spectrometry is useful for detecting OXA-48 and KPC carbapenemases in multi-drug resistant Gram-negative bacteria (Fleurbaaij et al., 2014; **Figure 2**), and detection of beta-lactamase proteins has been improved by the use of MALDI-TOF MS methods rather than PCR detection of the corresponding genes (Trip et al., 2015).
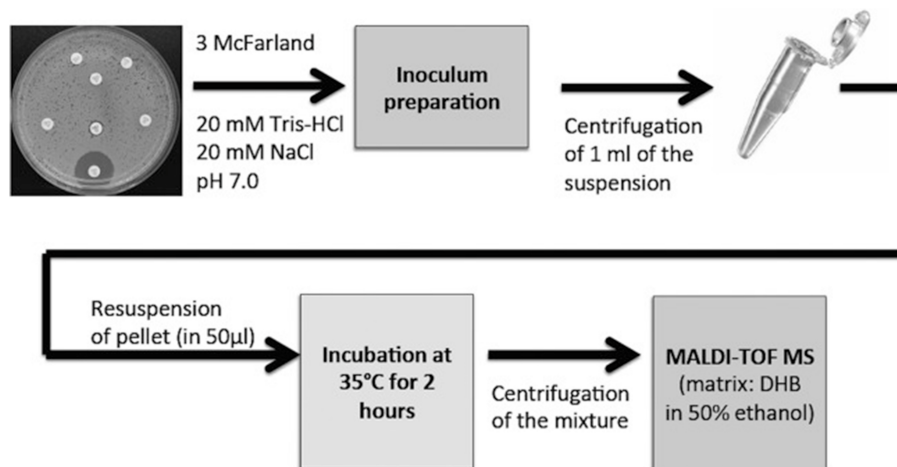
FIGURE 1 | Basic protocol for detecting beta-lactam hydrolyzed products by MALDI-TOF MS. 3 McFarland is a measure of optical density of the bacterial culture. Obtained from Hrabák et al. (2013) and reprinted with permission from the publisher.
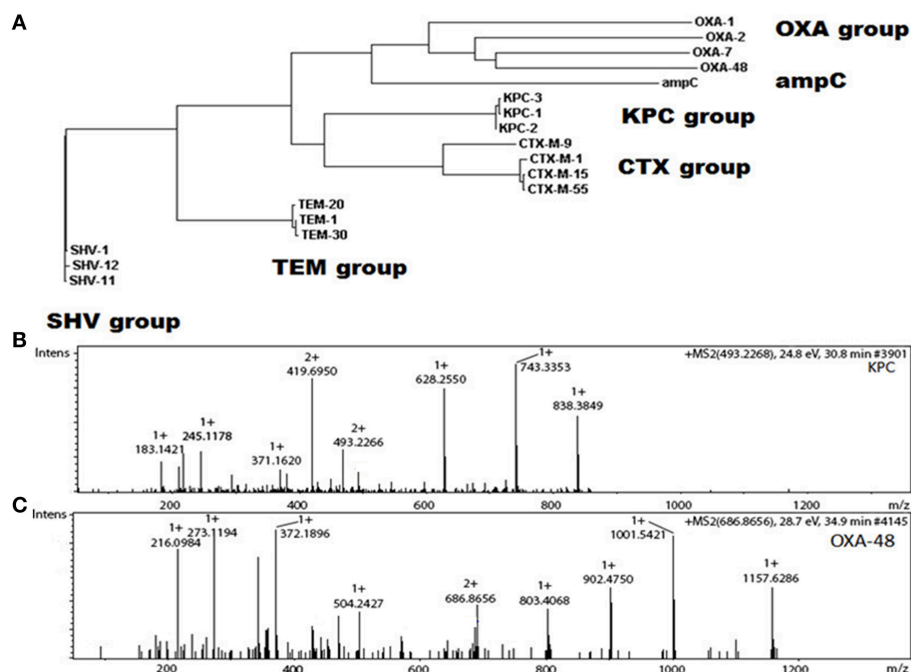

FIGURE 2 | Differentiation between different β-lactamase subgroups and species. (A) Phylogenetic tree based on a sequence alignment of different β-lactamases from different functional subgroups. For some classes, the primary structures are very similar (e.g., KPC and CTX) while for others, individual sequences vary considerably (e.g., OXA). **(B)** Tandem mass spectrum for a unique tryptic peptide from the KPC-group of carbapenemases (FPLCSSFK). **(C)** Tandem mass spectrum for a unique tryptic peptide from OXA-48 (SQGVVVLWNENK). Obtained and adapted from Fleurbaaij et al. (2014) and reprinted with permission from the publisher.

Cell Wall Analysis

The cell wall is the target of most antibiotics and is a barrier to other antibiotics that act in the cytosol. Some components of the outer membrane of Gram-negative bacteria such as porins, efflux pumps, and lipopolysaccharides have been quantified by MALDI-TOF MS techniques to distinguish between resistant and sensitive strains (Peng et al., 2005; Imperi et al., 2009). For example, a MALDI-TOF MS based method proved better than SDS-PAGE for characterizing the Ompk36 porin in *K. pneumoniae* (Cai et al., 2012). Changes in the structure of the lipopolysaccharide lipid A that occur during emergence of resistance to colistin can be

detected by MALDI-TOF MS in *A. baumanni* (Beceiro et al., 2011).

Discovery of Mutations within Resistance Genes through Mini-Sequencing

MALDI-TOF MS methods have been used to analyse DNA sequencing (Pusch et al., 2002). The technique is based on a primer extension assay for only a few bases, and the molecular weight shift is then detected by MALDI-TOF MS. The technique is limited to DNA molecules smaller than 40 bp (**Figure 3**).

Although, it has been used to detect mutations for a large and diverse number of resistance mechanism (Hrabák et al., 2013), this approach is time-consuming and does not confer any advantages over standard sequencing protocols.

PCR/electrospray ionization Mass spectrometry (PCR/ESI MS), which involves mini-sequencing of larger PCR products (110–450 bp), has been successfully used to identify pathogens and has also proved useful for detecting some bla_{KPC} beta-lactamase producing Enterobacteria (Endimiani et al., 2010). This technique has been commercialized via development of a fully automated system (PLEX-ID, Abbot Biosciencies).

Stable Isotope Labeling and Monitoring of Cell Growth

The technique has been used for monitoring resistance to fluconazole (Marinach et al., 2009) and caspofungin (De Carolis et al., 2012) in fungi. Cells harvested from solid cultures were incubated with different concentrations of antibiotic for 24 h and the total peak profile was determined; however, no improvement in time-to-result was achieved, despite the good correlation between MIC-values and different spectra. A simplified version of this approach that facilitates discrimination of susceptible and

FIGURE 3 | Simplified procedure of minisequencing linked to MALDI-TOF MS detection. Obtained from Hrabák et al. (2013) and reprinted with permission from the publisher.

resistant isolates of *Candida albicans* after incubation for 3 h in the presence of "breakpoint" level drug concentrations of the caspofungin has been described (Vella et al., 2013).

Some studies have attempted to detect bacterial resistance by incorporating stable isotope labeled nutrients during growth in the presence of antibiotics (Demirev et al., 2013). The peak profiles were compared with those of intact microorganisms grown in unlabeled media without the drug. Some biomarkers were detected by characteristic mass shifts provided by a suitable isotope. A similar approach has recently been used with $^{13}C_6^{15}N_2$-L-lysine as the isotope for detection of methicillin-resistant *S. aureus* (Sparbier et al., 2013). Stable isotope labeling by amino acids in cell culture (SILAC) technology with the same isotope have been used to distinguish resistant and susceptible strains of *P. aeruginosa* to meropenem, tobramycin, and ciprofloxacin (Jung et al., 2014).

Conclusions

Although, the use of MALDI-TOF MS in clinical microbiology has been successfully applied in identifying bacteria, detection of antibiotic resistance, essential in bacterial diagnostics, is still at an early stage of development. Some promising achievements are the detection of proteins that hydrolyze antibiotics by means of direct detection of antibiotic modifications. The most remarkable example is the detection of beta-lactamases, particularly carbapenemases, by a method that has been validated and routinely used in clinical and reference laboratories. However, detection of resistance mechanism elements (particularly Qnr proteins and mutant PBP and beta-lactamases) for improved diagnosis of multiresistant bacteria must be improved and validated, despite some reports about the detection of vancomycin resistant *Enterococcus* spp. and discrimination between MRSA and MSSA strains.

In summary, several of the cited methods require further validation, simplification and automation. MALDI TOF-MS analysis may prove to be too labor intensive for routine use in clinical laboratories. Other challenges include the better study of the interaction between proteomic results and MIC-values for used in prescribing appropriate treatment and monitoring infection. Methods of testing multiple antibiotics are urgently required in order to provide a complete picture of antibiotic resistance. Standard susceptibility testing will probably coexist with new proteomic methods, as resistance to antibiotic is a complex process.

PROTEOMICS FOR STUDYING BACTERIAL VIRULENCE

General Aspects

Proteomic techniques are becoming important tools for investigating microbial pathogenesis. Applications include identification of virulence factors and study of the response of both host and pathogen to infection.

Research was initially limited to the classical proteomic analysis of protein contents of bacterial cultures, usually involving comparison of avirulent and virulent strains or conditions that simulate stresses that bacteria encounter during
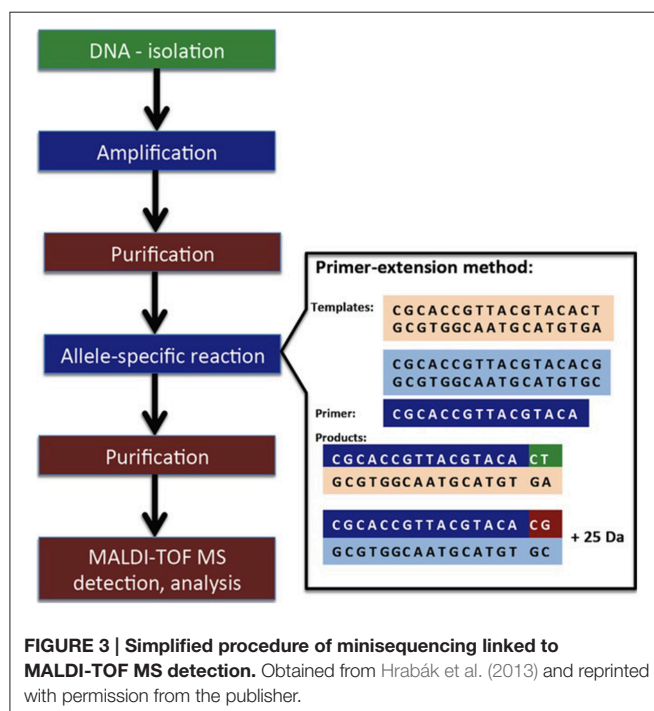
the infection process, such as acid stress, low oxygen content, high osmolarity, and other conditions of temperature, presence or limited availability of iron and presence of urine or plasma. Co-cultivation with host cell cultures and the direct isolation of bacteria from samples were not achieved until a later stage, because of the technical challenge that arose in measuring bacterial proteins against the overwhelming background of host proteins. Different approaches have been used to solve these problems, usually involving different separation techniques such as differential centrifugation, cell sorting, and immunomagnetic separation. Some review articles have already considered these early studies (Bhavsar et al., 2010; Cash, 2011; Yang et al., 2015), and we therefore focus on recent studies that we consider important in the study of bacterial pathogenesis by proteomics.

Bacterial Pathogenesis

Wang et al. (2014) recently compared the proteome profile of the *S. enterica* subsp. *enterica* serovar *Typhimurium* and *S. typhi,* which are responsible for, respectively, gastroenteritis and typhoid fever types. These researchers first detected a group of proteins with serovar-specific expression, which can be used as new biomarkers for identifying clinical serotypes. They also found that expression of flagella and chemotaxis proteins was lower in *S. typhi* than in *S. typhimurium.* Finally, the expression of core genes, which were involved in metabolism and transport of carbohydrates and amino acids, differed in the two serovars.

Proteomics, microbial genetics, competitive infections, and computational approaches have previously been used to obtain a comprehensive overview of *Salmonella* nutrition and growth in a mouse typhoid fever model (Steeb et al., 2013). In a study of the virulence of *H. pylori*, Vitoriano et al. (2011) compared the proteomes of strains of the bacterium isolated from children with peptic ulcer disease and from children with non-ulcer dyspepsia. In addition to the presence of genes clearly encoding virulence factor, the pediatric ulcerogenic strains presented a proteome profile defined by changes of motility-associated proteins (involved in higher motility), antioxidant proteins (involved in better resistance to inflammation), and proteins implicated in key functions in the metabolism of glucose, amino acids, and urea (probably advantageous for confronting changes in nutrient availability).

More recently, Ansong et al. (2013a) performed a temporal multi-omic analysis, at physiologically relevant temperatures, of *Yersinia pestis* (YP), the causative agent of plague with a high mortality rate, and *Yersinia pseudotuberculosis* (YPT), an enteric pathogen with a modest mortality rate. Gene and protein expression levels of conserved major virulence factors were higher in YP than in YPT, including the Yop virulon and the pH6 antigen. The global transcriptome and proteome responses of YP and YPT revealed conserved post-transcriptional control of metabolism and the translational machinery including the modulation of glutamate levels in *Yersiniae*.

Madeira et al. (2013) compared the proteomic profiles of three clonal isolates of *Burkholderia cenocepacia* obtained from a cystic fibrosis patient between onset of infection and before

death from cepacia syndrome 3.5 years later. They found that 52 proteins were similarly altered in both late-stage isolates, relative to the first isolate, which suggests an important role for metabolic reprogramming in the virulence potential and persistence of *B. cenocepaci*a, particularly in regard to bacterial adaptation to microaerophilic conditions. The content of the virulence determinant AidA was also higher in the two late-stage isolates.
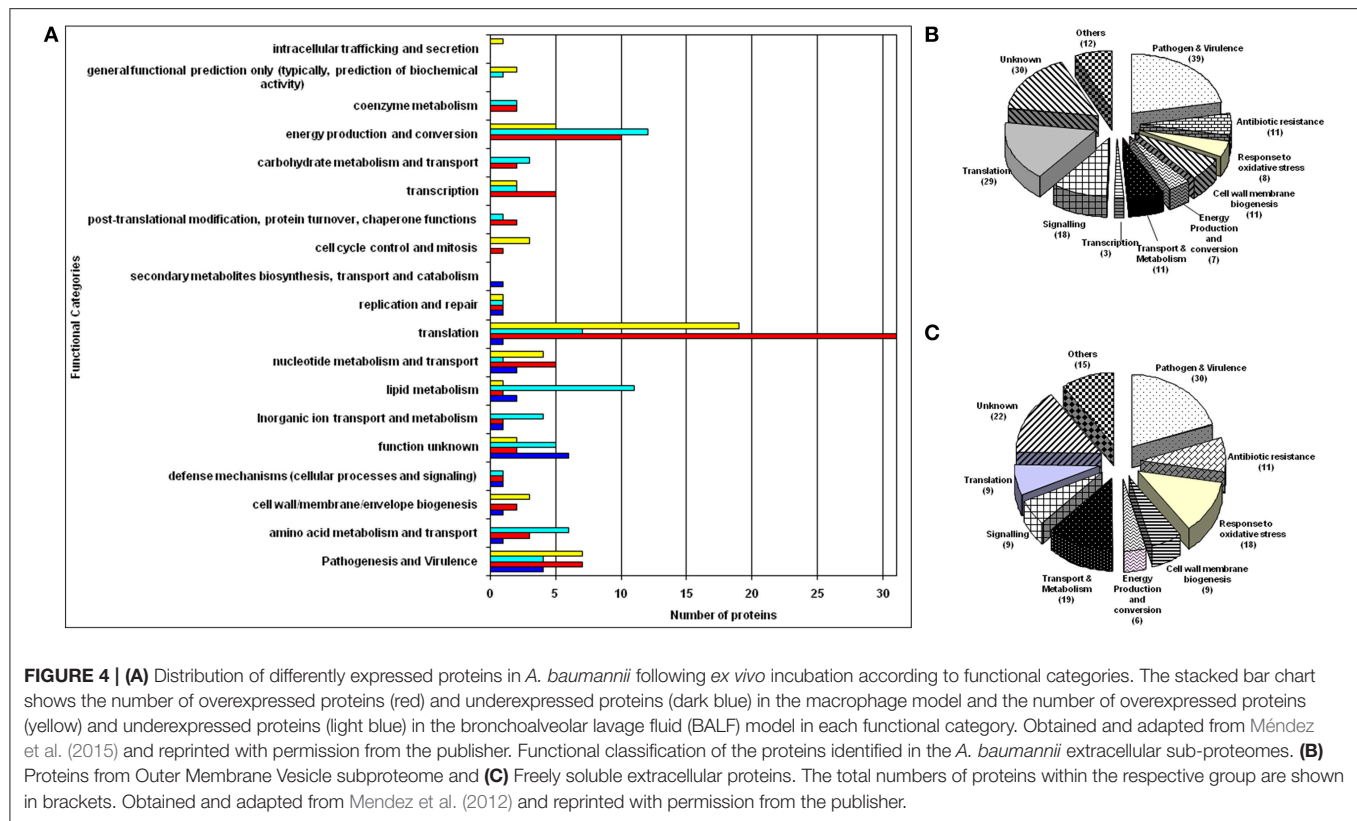
In an original approach, Provenzano et al. (2013) analyzed the metaproteome of microbial communities from endodontic infections associated with acute apical abscesses and asymptomatic apical periodontal lesions. They found several proteins related to pathogenicity and resistance/survival in endodontic samples, including proteins involved in adhesion, biofilm formation, and antibiotic resistance, stress proteins, exotoxins, invasins, proteases, and endopeptidases (mostly in abscesses), and an archaeal protein linked to methane production. Most of the human proteins detected were involved in cellular processes and metabolism, as well as immune defense.

In a study involving quantitative proteomic analysis, Liu et al. (2012) showed that on entry into host cells, *Campylobacter jejuni* undergoes a significant metabolic downshift. They also observed reprogramming of respiration in intracellular *C. jejuni*, favoring respiration of fumarate.

Pieper et al. (2013) performed global proteomic analysis of *Shigella flexneri* strain 2457T in association with three distinct growth environments: in broth (*in vitro*), in epithelial cell cytoplasm (intracellular), and coculture with extracellular epithelial cells. The intracellular bacteria showed elevated protein expression in invasion and cell-to-cell spread determinants, including IpA, Mxi, and Ics proteins, compared with *in vitro* and extracellular bacteria. The intracellular environment was also characterized by changes in iron stress and carbon metabolism protein expression levels, which may indicate its important role in the transition of *S. flexneri* from the extra to the intracellular milieu.

Another elegant study identified the pathogenic mechanisms of *Chlamydia trachomatis* after entry into host cells and the formation of a membrane-bound compartment (the inclusion) accompanied by secretion of inclusion membrane proteins (Incs). Mirrashidi et al. (2015) used affinity purification-mass spectroscopy (AP-MS) to detect Inc-human interactions for 38/58 Incs involved in host processes of intracellular life cycles, including retromer components as sorting nexins. Inc targets and overlapping of viral proteins were detected, suggesting common pathogenic mechanisms among obligate intracellular microbes.

During a very recent study with the infectious agent *A. baumannii*, we obtained an *ex vivo* protein expression profiling for pneumonia (Méndez et al., 2015). We characterized the proteome of *A. baumannii* in the presence of bronchoalveolar lavage fluid from infected rats (to simulate conditions in the respiratory tract) and in the presence RAW 264.7 cells (control conditions). We observed alterations in cell wall synthesis and identified two upregulated virulence-associated proteins with >15 peptides/protein in both *ex vivo* models (OmpA and YjjK), which suggests that these proteins are fundamental for pathogenesis and virulence in the airways (**Figure 4A**).

FIGURE 4 | (A) Distribution of differently expressed proteins in *A. baumannii* following *ex vivo* incubation according to functional categories. The stacked bar chart shows the number of overexpressed proteins (red) and underexpressed proteins (dark blue) in the macrophage model and the number of overexpressed proteins (yellow) and underexpressed proteins (light blue) in the bronchoalveolar lavage fluid (BALF) model in each functional category. Obtained and adapted from Méndez et al. (2015) and reprinted with permission from the publisher. Functional classification of the proteins identified in the *A. baumannii* extracellular sub-proteomes. **(B)** Proteins from Outer Membrane Vesicle subproteome and **(C)** Freely soluble extracellular proteins. The total numbers of proteins within the respective group are shown in brackets. Obtained and adapted from Mendez et al. (2012) and reprinted with permission from the publisher.

Studies of Microbial Biofilm

At least 65% of all human infectious diseases are known to be associated with the biofilm form of microorganisms. The most notable and clinically relevant property of biofilms is the greater resistance to antimicrobials than shown by their planktonic counterparts (Seneviratne et al., 2012). Some proteomic studies have revealed an association between resistance mechanisms and biofilm state, the subject of a recent review article (Seneviratne et al., 2012). Our research group has been able to identify a putative resistance-nodulation-cell division type efflux pump (RND pump), which increased in *A. baumannii* biofilms. Overexpression of this type of efflux pump in *A. baumannii* confers resistance to aminoglycosides, as well as to other drugs, including fluoroquinolones, tetracyclines, chloramphenicol, erythromycin, trimethoprim, and ethidium bromide (Cabral et al., 2011). In another study recently carried out in our laboratory, *A. baumannii* clinical strain AbH12O-A2 was able to survive long periods of desiccation as a result of the presence of cells in a dormant state, via mechanisms affecting control of cell cycling, DNA coiling, transcriptional and translational regulation, protein stabilization, antimicrobial resistance, and toxin synthesis; a few surviving cells embedded in a biofilm matrix were able to resume growth and restore the original population in appropriate environmental conditions (Gayoso et al., 2014).

Relationships between biofilm state and virulence factors have also been discovered recently. High-resolution 2-dimensional gel electrophoresis was used to obtain the proteome profiles of spiral *H. pylori* and early biofilm (Shao et al., 2013). Differential protein spots were identified and associated with flagellar movement, bacterial virulence, signal transduction, and regulation.

MALDI-TOF MS/MS analysis of the secretome of single and mixed biofilms of *P. aeruginosa* and *C. albicans* carried out at different times (Purschke et al., 2012) identified 16 proteins overexpressed or exclusively expressed in *P. aeruginosa* when interacting with *C. albicans* including virulence factors such as exotoxin A and iron acquisition systems.

Another strategy was used to study the biofilm expression profile in a clinical strain of *S. pneumoniae* serotype 14. A decrease in the expression of enzymes related to the glycolytic pathway, as well as proteins involved in translation, transcription, and virulence was detected during biofilm growth by the proteomics method (Allan et al., 2014).

Outer Membrane Vesicles and their Impact on Virulence

Outer membrane vesicles (OMVs) are nanoscale structures secreted by bacteria and that can carry nucleic acids, proteins, and small metabolites. They mediate intracellular communication and play a role in virulence. One of the early proteomic studies of the protein fraction of the outer membrane vesicles was carried out in our laboratory. This study focused on two main protein fractions of the extracellular proteome: proteins exported by outer membrane vesicles (OMVs) and freely soluble

extracellular proteins (FSEPs) present in the culture medium of a highly invasive, multidrug-resistant strain of *A. baumannii* (clone AbH12O-A2; Mendez et al., 2012). Of the OMV proteins, 39 were associated with pathogenesis and virulence, including proteins associated with attachment to host cells (e.g., CsuE, CsuB, CsuA/B) and specialized secretion systems for delivery of virulence factors (e.g., P-pilus assembly and FilF), whereas the FSEP fraction possessed extracellular enzymes with degradative activity, such as alkaline metalloprotease. Among the FSEP, we have also detected at least 18 proteins with a known role in the oxidative stress response (e.g., catalase, thioredoxin, oxidoreductase, superoxide dismutase; **Figures 4B,C**). These findings are supported by those of a recent study of other clinical *A. baummanni* strains in which it was discovered that *A. baumanni* strain A38 possesses an outer membrane vesicle containing virulence factors including Omp38, EpsA, Ptk, GroEL, hemagglutinin-like protein, and FilF (Li Z. T. et al., 2015). The high content of virulence factors in outer membrane vesicles has been demonstrated in other Gram-negative bacteria such as clinical strains of *E. coli* and *Stenotrophomonas maltophila* (Devos et al., 2015; Wurpel et al., 2015). Indeed, this phenomenon is also conserved in some Gram-positive pathogenic strains such as *M. tuberculosis* and *S. pneumoniae* (Olaya-Abril et al., 2014; Lee J. et al., 2015).

However, as pointed out by several authors (Koning et al., 2013; Pérez-Cruz et al., 2013), OMVs may be composed by inner membrane fractions and can thus be considered outer-inner membrane vesicles. These structures may thus contain cytosolic fractions. OMV-related virulence can be considered in the context of this novel concept.

Post-Translational Protein Modifications and their Impact on Virulence

Bacteria contain different types of post-translational modifications, many of which are commonly present in eukaryotes: phosphorylation, acetylation, glycosylation, ubiquitination, and glutathionylation (Soufi et al., 2012). Relationships between these types of post-translational modifications and virulence have been established in recent years. One of the first studies with *Pseudomonas* species revealed that phosphoproteins involved in motility, transport and pathogenicity pathways in *P. aeruginosa* are critical for survival and virulence (Ravichandran et al., 2009).

Around the same time, the phosphoproteome of *K. pneumoniae* NTUH-K2044 was analyzed by a shotgun approach, and 117 unique phosphopeptides were identified along with 93 *in vivo* phosphorylated sites corresponding to 81 proteins (Lin et al., 2009). Interestingly, three of these were found to be distributed in proteins of the cps locus and the authors of the study speculated that they were involved in the converging signal transduction of capsule biosynthesis, which has been related to virulence. Involvement of phosphorylated proteins in virulence was also hypothesized after proteomic analysis of *S. pneumoniae* and *Mycoplasma pneumoniae* (Schmidl et al., 2010; Sun et al., 2010).

Interestingly, qualitative comparison between the Ser/Thr/Tyr phosphoproteomes of two *A. baumannii* strains (a reference strain and a highly invasive multidrug-resistant clinical isolate) led to the identification of phosphoproteins with a role in pathogenicity and also involved in drug resistance (Soares et al., 2014).

Other types of post-translational modifications have recently been discovered by proteomic methods. In a well-designed study, Iwashkiw et al. (2012) showed that *A. baumannii* ATCC 17978 possesses an O-glycosylation system responsible for the glycosylation of multiple proteins. These authors identified seven *A. baumannii* glycoproteins, of yet unknown function, by 2D-DIGE and mass spectrometry. A glycosylation-deficient strain was generated by homologous recombination. This strain did not show any growth defects, but exhibited a severely diminished capacity to generate biofilms. Disruption of the glycosylation machinery also resulted in reduced virulence in two infection models (*Dictyostelium discoideum* and *Galleria mellonella*) and reduced *in vivo* fitness in a mouse model of peritoneal sepsis. In a study involving proteomic and glycoproteomic analysis, Scott et al. (2014) found higher N-linked glycosylation in *Campylobacter jejuni* NCTC11168 O clinical strain than in the laboratory-derived strain.

Another post-translational modification was described by Xie et al. (2015) who investigated proteome lysine acetylation profiling in *M. tuberculosis* by using a combination of anti-acetyl lysine antibody-based immunoaffinity enrichment and high-resolution mass spectrometry. These authors identified 1128 acetylation sites on 658 acetylated *M. tuberculosis* proteins, several of which, e.g., isocitrate lyase, were involved in the persistence, virulence and antibiotic resistance. In an interesting study, Ansong et al. (2013b) identified different uses for the protein S-thiolation forms S-glutathionylation and S-cysteinylation in response to infection-like conditions and basal conditions in *S. typhimurium* and supported by analysis of protein structure and gene deletion.

Conclusions

The use of proteomic analysis to study the interactions between bacterial pathogenesis and host is at a very early stage. Nevertheless, we have described the use of proteomic techniques to investigate many aspects of bacterial pathogenesis: assignment of a virulence factor reservoir of a single pathogen, analysis of the interaction between host and pathogen response during the infection, identification of interactions between several virulence factors and host cell components and even the identification of biochemical action of virulence factors (**Table 5**).

The limit of sensitivity has prevented a more detailed proteomic study of *in vivo* models of infection, particularly in humans, in comparison with tissue culture models. Although, current methods of analysis are proving very useful, they must be improved further in order to advance our knowledge in this field. A multidisciplinary approach including methods such as immmunoassays, histological methods, and electron microscopy should be undertaken for an integrated understanding of bacterial pathogenesis. Monitoring of the temporal and spatial

TABLE 5 | Representative proteomic techniques for studying virulence determinants.

Proteomic technique (s)	Physiological effects	Experimental approach	Pathogen (s)	References
SDS-PAGE electrophoresis and LC-MS/MS	Metabolic downshift, reprogramming of respiration to fumarate	Isolation of bacteria from intracellular culture	*Campylobacter jejuni*	Liu et al., 2012
LTQ-Orbitrap Velos and LC-QTOF analysis	Identification of proteins involved in adhesion, biofilm formation, antibiotic resistance, stress proteins, exotoxins, invasions, proteases, and endopeptidases	Metaproteome analysis	Endodontic bacteria	Provenzano et al., 2013
2-D DIGE	Metabolic reprogramming to microaerophilic conditions, upregulation of vilrulence determinant AidA, and iron uptake proteins	Study of clonal isolates in different stages during cystic fibrosis infection	*Burkholderia cenocpacia*	Madeira et al., 2013
2-DE and LC-MS/MS	Yop virulon and pH 6 antigen overexpressed in *Yersinia pestis*. Modulation of glutamate levels.	Proteomic analysis at physiologically relevant temperatures	*Yersinia pestis* and *Yersinia pseudotuberculosis*	Ansong et al., 2013a
2-DE and LC-MS/MS	Ulcerogenic strains associated with virulence factors, motility proteins, antioxidant proteins, and metabolism of glucose, amino acids, and higher expression of urea proteins	Comparision of clinical isolates with peptic ulcer disease with non-ulcer dyspepsia	*Helicobacter pylori*	Vitoriano et al., 2011
iBAQ label free	Reprogramming of metabolism	Bacteria isolated from mouse typhoid fever model	*Salmonella enterica*	Steeb et al., 2013
SILAC	Flagella and chemotaxis proteins downregulated in *Salmonella typhi*. Identification of strain associated biomarkers	*In vitro* cultures of both strains	*Salmonella typhimurium* and *Salmonella typhi*	Wang et al., 2014
Affinity purification-mass spectroscopy (AP-MS)	Identification of interactions of inclusion membrane proteins with sorting nexins, components of the retromer.	Human targets of bacterial inclusion membrane proteins	*Chlamydia trachomatis*	Mirrashidi et al., 2015
iTRAQ and LC-MS/MS	Alterations in cell wall synthesis and upregulation of virulence proteins OmpA and Yjjk	Response to host airways "*ex vivo*" models	*Acinetobacter baumannii*	Méndez et al., 2015
2-DE and LC-MS/MS	Biofilm associated proteins related to flagellar movement, bacterial virulence, and signal transduction and regulation	Biofilm analysis	*Helicobacter pylori*	Shao et al., 2013
iTRAQ	Decrease in proteins involved in glycolytic, translation, transcription, and virulence. Increase in proteins, pyruvate, carbohydrate, and arginine metabolism	Biofilm analysis	*Streptococcus pneumoniae*	Allan et al., 2014
2-DE and LC-MS/MS	Intracellular bacteria increased expression of invasion and cell-to-cell spread proteins. Mixed-acid fermentation pathway proteins and iron stress proteins upregulated in intracellular bacteria	Comparison of *in vitro*, intracellular and extracellular cultures	*Shigella flexneri*	Pieper et al., 2013

action of virulence factors may also contribute to advances in this field.

The relationships between biofilm formation, outer membrane vesicles, and virulence have been highlighted during this review.

Identification of variations in post-translational changes in the bacteria and/or in the host proteomes after infection is also important for studying interactions between signaling cascades.

A detailed global picture of the interaction between host and pathogen may provide opportunities for identifying new targets for antimicrobial programs and vaccine development.

PRACTICAL APPLICATIONS

We have reported the huge progress in diagnostics of microbial resistance and basic research in resistance and virulence achieved in recent years. Many of these advances have been transferred to protocols for practical diagnosis and will potentially lead to the development of new therapies against pathogenic bacteria. Such examples include the discovery of 34 unknown proteins of similar abundance found in both cell envelopes and membrane vesicles fractions in four *N. gonorrhoeae* strains (Zielke et al., 2014). Depletion of one of these, a homolog of an outer membrane protein LptD, was found to cause loss of viability

in *N. gonorrhoeae.* These authors also identified another six predicted outer membrane proteins of unknown function. Loss of NGO1985, in particular, resulted in dramatically decreased viability of *N. gonorrhoeae* on treatment with detergents, polymyxin B and chloramphenicol, suggesting that this protein functions in maintaining the cell envelope barrier to permeability and may represent a new therapeutic target (Zielke et al., 2014).

Proteomic studies involving metabolism of virulent or resistant strains have been a source of potential therapeutic targets. For example, in a study investigating determinants of the biofilm condition of *A. baumannii,* Cabral et al. (2011) found that histidine metabolism (like Urocanase) was implicated in biofilm formation, as confirmed by gene disruption experiments. The authors proposed a model in which novel proteins are suggested for the first time as targets for preventing the formation of *A. baumannii* biofilms.

Vaccination has been revealed as an important strategy for reducing the incidence of infectious diseases. Immunoproteomics permits the characterization of putative vaccine candidates by combining proteomic analysis during infection together with serum identification of immunoreactive antigens. A recent review described examples of vaccine candidates that have been identified by immunoproteomics and have successfully protected animals against challenge when tested in immunization studies (Dennehy and McClean, 2012).

An exopolysaccharide and a protein-based biofilm expressed by two clinical strains of *S. aureus* have recently been studied by proteomic methods. Moreover, the number of bacterial cells inside a biofilm and in the surrounding tissue decreased after immunization with the biofilm matrix exoproteome in an *in vivo* model with a biofilm linked to infection (Gil et al., 2014).

CONCLUDING REMARKS

The method used for MS-based identification of proteins is constantly improving and has allowed progress from qualitative determination of proteins (i.e., presence or absence) toward quantitative studies of the proteome. However, the method still has many limitations. The inability to determine many proteins present at low concentrations, together with small number of proteins in the dataset may lead to errors in interpretation (e.g., associating complete biochemical pathways with virulence or resistance mechanisms). Comprehensive interpretation is also very difficult when the genome has not yet been obtained. This could be solved by using the highest resolution bottom-up approach and appropriate use of the knowledge base to improve the design of new studies and by the inclusion of specific statistical strategies to identify significant proteins. Next-generation sequencing should be linked with proteomic techniques. The combined use of different methods will cut down on the number of experiments required and improve the chances of identifying effective targets.

Basic aspects of bacterial growth and physiology are often ignored. Collaboration between a broad range of specialist including biochemists, biologists, and statisticians will be necessary, especially once metabolites can be synthesized. These metabolites must be assayed by metabolomic methods. Recent reviews of the relationship between metabolism and antibiotic resistance revealed global regulators that simultaneously alter virulence, metabolism and antibiotic resistance (Martínez and Rojo, 2011).

It can be difficult to coordinate and interpret the increasing volume of high resolution MS data obtained in the field of bacterial proteomics. Thus, an outstanding role for microbial bioinformatics is expected and more software packages for MS analysis must be shared. Attempts could also be made to merge several previous studies and select common patterns that ultimately enable better design of future experiments.

On the other hand, the multifactorial nature of bacterial infection must be investigated using several approaches, in a systems biology strategy. The coordinated use of different techniques including genomics, transcriptomics, and metabolomics, together with good standard proteomic methods will improve the capacity for detecting bacterial resistance, determination of resistance mechanisms and the understanding of virulence response in bacteria and host. This will hopefully help in the discovery of new biomarkers, vaccines, and drugs for combating bacterial infection worldwide.

In conclusion, in this review we summarize some of the applications of proteomic technology for diagnosing infections as well as performing basic studies for understanding antimicrobial resistance and virulence. This will enable discovery of new targets for novel and innovative antibiotics as well as the implementation of new therapies in the near future.

AUTHOR CONTRIBUTIONS

FP, revised literature and wrote; GB, wrote and contributed with critical analysis.

FUNDING

REFERENCES

Al-Majdoub, Z. M., Owoseni, A., Gaskell, S. J., and Barber, J. (2013). Effects of gentamicin on the proteomes of aerobic and oxygen-limited *Escherichia coli*. *J. Med. Chem.* 11, 2904–2910. doi: 10.1021/jm301858u

Allan, R. N., Skipp, P., Jefferies, J., Clarke, S. C., Faust, S. N., Hall-Stoodley, L., et al. (2014). Pronounced metabolic changes in adaptation to biofilm growth by *Streptococcus pneumoniae*. *PLoS ONE* 9:e107015. doi: 10.1371/journal.pone.0107015

Álvarez-Buylla, A., Picazo, J. J., and Culebras, E. (2013). Optimized method for Acinetobacter species carbapenemase detection and identification by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J. Clin. Microbiol.* 51, 1589–1592. doi: 10.1128/JCM.00181-13

Ansong, C., Schrimpe-Rutledge, A. C., Mitchell, H. D., Chauhan, S., Jones, M. B., Kim, Y. M., et al. (2013a). A multi-omic systems approach to elucidating Yersinia virulence mechanisms. *Mol. Biosyst.* 27, 44–54. doi: 10.1039/C2MB25287B

Ansong, C., Wu, S., Meng, D., Liu, X., Brewer, H. M., Deatherage Kaiser, B. L., et al. (2013b). Top-down proteomics reveals a unique protein S-thiolation switch in *Salmonella Typhimurium* in response to infection-like conditions. *Proc. Natl. Acad. Sci. U.S.A.* 110, 10153–10158. doi: 10.1073/pnas.1221210110

Beceiro, A., Llobet, E., Aranda, J., Bengoechea, J. A., Doumith, M., Hornsey, M., et al. (2011). Phosphoethanolamine modification of lipid A in colistin-resistant variants of *Acinetobacter baumannii* mediated by the pmrAB two-component regulatory system. *Antimicrob. Agents Chemother.* 55, 3370–3379. doi: 10.1128/AAC.00079-11

Beceiro, A., Tomás, M., and Bou, G. (2013). Antimicrobial resistance and virulence: a successful or deleterious association in the bacterial world? *Clin. Microbiol. Rev.* 26, 185–230. doi: 10.1128/CMR.00059-12

Bhavsar, A. P., Auweter, S. D., and Finlay, B. B. (2010). Proteomics as a probe of microbial pathogenesis and its molecular boundaries. *Future Microbiol.* 5, 2532–2565. doi: 10.2217/fmb.09.114

Bialvaei, A. Z., and Samadi Kafil, H. (2015). Colistin, mechanisms and prevalence of resistance. *Curr. Med. Res. Opin.* 4, 707–721. doi: 10.1185/03007995.2015.1018989

Biot, F. V., Valade, E., Garnotel, E., Chevalier, J., Villard, C., Thibault, F. M., et al. (2011). Involvement of the efflux pumps in chloramphenicol selected strains of *Burkholderia thailandensis*: proteomic and mechanistic evidence. *PLoS ONE* 6:e16892. doi: 10.1371/journal.pone.0016892

Boucher, H. W., Talbot, G. H., Bradley, J. S., Edwards, J. E., Gilbert, D., Rice, L. B., et al. (2009). Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America. *Clin. Infect. Dis.* 48, 1–12. doi: 10.1086/595011

Burckhardt, I., and Zimmermann, S. (2011). Using matrix-assisted laser desorption ionization–time of flight mass spectrometry to detect carbapenem resistance within 1 to 2.5 hours. *J. Clin. Microbiol.* 49, 3321–3324. doi: 10.1128/JCM.00287-11

Cabral, M. P., Soares, N. C., Aranda, J., Parreira, J. R., Rumbo, C., Poza, M., et al. (2011). Proteomic and functional analyses reveal a unique lifestyle for *Acinetobacter baumannii* biofilms and a key role for histidine metabolism. *J. Proteome Res.* 10, 3399–3417. doi: 10.1021/pr101299j

Cai, J. C., Hu, Y. Y., Zhang, R., Zhou, H. W., and Chen, G-X. (2012). Detection of Ompk36 porin loss in *Klebsiella* spp. by matrix-assisted laser desorption ionization–time of flight mass spectrometry. *J. Clin. Microbiol.* 50, 2179–2182. doi: 10.1128/JCM.00503-12

Camara, J. E., and Hays, F. A. (2007). Discrimination between wild-type and ampicillin-resistant *Escherichia coli* by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Anal. Bioanal. Chem.* 389, 1633–1638. doi: 10.1007/s00216-007-1558-7

Cash, P. (2011). Investigating pathogen biology at the level of the proteome. *Proteomics* 11, 3190–3202. doi: 10.1002/pmic.201100029

Cash, P., Argo, E., Ford, L., Lawrie, L., and McKenzie, H. (1999). A proteomic analysis of erythromycin resistance in *Streptococcus pneumoniae*. *Electrophoresis* 20, 2259–2268.

Chao, T. C., and Hansmeier, N. (2012). The current state of microbial proteomics: where we are and where we want to go. *Proteomics* 12, 638–650. doi: 10.1002/pmic.201100381

Chaussee, M. A., McDowell, E. J., Rieck, L. D., Callegari, E. A., and Chaussee, M. S. (2006). Proteomic analysis of a penicillin tolerant rgg mutant strain of *Streptococcus pyogenes*. *J. Antimicrob. Chemother.* 58, 752–759. doi: 10.1093/jac/dkl319

Chen, H., Liu, Y., Zhao, C., Xiao, D., Zhang, J., Zhang, F., et al. (2013). Comparative proteomics-based identification of genes associated with glycopeptide resistance in clinically derived heterogeneous vancomycin-intermediate *Staphylococcus aureus* strains. *PLoS ONE* 8:e66880. doi: 10.1371/journal.pone.0066880

Chong, P. M., Lynch, T., McCorrister, S., Kibsey, P., Miller, M., Gravel, D., et al. (2014). Proteomic analysis of a NAP1 *Clostridium difficile* clinical isolate resistant to metronidazole. *PLoS ONE.* 6:e82622. doi: 10.1371/journal.pone.0082622

Chong, P. M., McCorrister, S. J., Unger, M. S., Boyd, D. A., Mulvey, M. R., and Westmacott, G. R. (2015). MALDI-TOF MS detection of carbapenemase activity in clinical isolates of *Enterobacteriaceae* spp., *Pseudomonas aeruginosa*, and Acinetobacter *baumannii* compared against the Carba-NP assay. *J. Microbiol. Methods* 111, 21–23. doi: 10.1016/j.mimet.2015.01.024

Chopra, S., Ramkissoon, K., and Anderson, D. C. (2013). A systematic quantitative proteomic examination of multidrug resistance in *Acinetobacter baumannii*. *J. Proteomics* 12, 17–39. doi: 10.1016/j.jprot.2013.03.008

Chua, S. L., Tan, S. Y., Rybtke, M. T., Chen, Y., Rice, S. A., Kjelleberg, S., et al. (2013). Bis-(3′-5′)-cyclic dimeric GMP regulates antimicrobial peptide resistance in *Pseudomonas aeruginosa*. *Antimicrob. Agents Chemother.* 57, 2066–2075. doi: 10.1128/AAC.02499-12

Civljak, R., Giannella, M., Di Bella, S., and Petrosillo, N. (2014). Could chloramphenicol be used against ESKAPE pathogens? A review of *in vitro* data in the literature from the 21st century. *Expert Rev. Anti. Infect. Ther.* 12, 249–264. doi: 10.1586/14787210.2014.878647

Coldham, N. G., Randall, L. P., Piddock, L. J., and Woodward, M. J. (2006). Effect of fluoroquinolone exposure on the proteome of *Salmonella enterica* serovar Typhimurium. *J. Antimicrob. Chemother.* 58, 1145–1153. doi: 10.1093/jac/dkl413

Cornick, J. E., and Bentley, S. D. (2012). *Streptococcus pneumoniae*: the evolution of antimicrobial resistance to beta-lactams, fluoroquinolones and macrolides. *Microbes Infect.* 14, 573–583. doi: 10.1016/j.micinf.2012.01.012

De Carolis, E., Vella, A., Florio, A. R., Posteraro, P., Perlin, D. S., Sanguinetti, M., et al. (2012). Use of matrix-assisted laser desorption ionization-time of flight mass spectrometry for caspofungin susceptibility testing of Candida and Aspergillus species. *J. Clin. Microbiol.* 50, 2479–2483. doi: 10.1128/JCM.00224-12

DeMarco, M. L., and Ford, B. A. (2013). Beyond identification: emerging and future uses for MALDI-TOF mass spectrometry in the clinical microbiology laboratory. *Clin. Lab. Med.* 33, 611–628. doi: 10.1016/j.cll.2013.03.013

Demirev, P. A., Hagan, N. S., Antoine, M. D., Lin, J. S., and Feldman, A. B. (2013). Establishing drug resistance in microorganisms by mass spectrometry. *Am. Soc. Mass Spectrom.* 24, 1194–1201. doi: 10.1007/s13361-013-0609-x

Dennehy, R., and McClean, S. (2012). Immunoproteomics: the key to discovery of new vaccine antigens against bacterial respiratory infections. *Curr. Protein Pept. Sci.* 13, 807–815. doi: 10.2174/138920312804871184

Devos, S., Van Oudenhove, L., Stremersch, S., Van Putte, W., De Rycke, R., Van Driessche, G., et al. (2015). The effect of imipenem and diffusible signaling factors on the secretion of outer membrane vesicles and associated Ax21 proteins in *Stenotrophomonas maltophilia*. *Front. Microbiol.* 14:298. doi: 10.3389/fmicb.2015.00298

Diniz, C. G., Farias, L. M., Carvalho, M. A., Rocha, E. R., and Smith, C. J. (2004). Differential gene expression in a *Bacteroides fragilis* metronidazole-resistant mutant. *J. Antimicrob. Chemother.* 54, 100–108. doi: 10.1093/jac/dkh256

Dos Santos, K. V., Diniz, C. G., Veloso Lde, C., de Andrade, H. M., Giusta Mda, S., Pires Sda, F., et al. (2010). Proteomic analysis of *Escherichia coli* with experimentally induced resistance to piperacillin/tazobactam. *Res. Microbiol.* 161, 268–275. doi: 10.1016/j.resmic.2010.03.006

Drummelsmith, J., Winstall, E., Bergeron, M. G., Poirier, G. G., and Ouellette, M. (2007). Comparative proteomics analyses reveal a potential biomarker for the detection of vancomycin intermediate *Staphylococcus aureus* strains. *J. Proteome Res.* 6, 4690–4702. doi: 10.1021/pr070521m

Dubska, L., Pilatova, K., Dolejska, M., Bortlicek, Z., Frostova, T., Literak, I., et al. (2011). Surface-enhanced laser desorption ionization/timeof-flight (SELDI-TOF) mass spectrometry (MS) as a phenotypic method for

rapid identification of antibiotic resistance. *Anaerobe* 17, 444–447. doi: 10.1016/j.anaerobe.2011.05.008

Edwards-Jones, V., Claydon, M. A., Evason, D. J., Walker, J., Fox, A. J., and Gordon, D. B. (2000). Rapid discrimination between methicillin-sensitive and methicillin-resistant *Staphylococcus aureus* by intact cell mass spectrometry. *J. Med. Microbiol.* 49, 295–300. doi: 10.1099/0022-1317-49-3-295

Endimiani, A., Hujer, K. M., Hujer, A. M., Sampath, R., Ecker, D. J., and Bonomo, R. A. (2010). Rapid identification of blaKPC-possessing Enterobacteriaceae by PCR/electrospray ionization-mass spectrometry. *J. Antimicrob. Chemother.* 65, 1833–1834. doi: 10.1093/jac/dkq207

Falagas, M. E., Vardakas, K. Z., Kapaskelis, A., Triarides, N. A., and Roussos, N. S. (2015). Tetracyclines for multidrug-resistant *Acinetobacter baumannii* infections. *Int. J. Antimicrob. Agents.* 45, 455–460. doi: 10.1016/j.ijantimicag.2014.12.031

Feng, J., Billal, D. S., Lupien, A., Racine, G., Winstall, E., Légaré, D., et al. (2011). Proteomic and transcriptomic analysis of linezolid resistance in *Streptococcus pneumoniae*. *J. Proteome Res.* 7, 4439–4452. doi: 10.1021/pr200221s

Fernández-Reyes, M., Rodríguez-Falcón, M., Chiva, C., Pachón, J., Andreu, D., and Rivas, L. (2009). The cost of resistance to colistin in *Acinetobacter baumannii*: a proteomic perspective. *Proteomics* 9, 1632–1645. doi: 10.1002/pmic.200800434

Fischer, A., Yang, S. J., Bayer, A. S., Vaezzadeh, A. R., Herzig, S., Stenz, L., et al. (2011). Daptomycin resistance mechanisms in clinically derived *Staphylococcus aureus* strains assessed by a combined transcriptomics and proteomics approach. *J. Antimicrob. Chemother.* 66, 1696–1711. doi: 10.1093/jac/dkr195

Fleurbaaij, F., Heemskerk, A. A., Russcher, A., Klychnikov, O. I., Deelder, A. M., Mayboroda, O. A., et al. (2014). Capillary-electrophoresis mass spectrometry for the detection of carbapenemases in (multi-)drug-resistant Gram-negative bacteria. *Anal. Chem.* 86, 9154–9161. doi: 10.1021/ac502049p

Gayoso, C. M., Mateos, J., Méndez, J. A., Fernández-Puente, P., Rumbo, C., Tomás, M., et al. (2014). Molecular mechanisms involved in the response to desiccation stress and persistence in *Acinetobacter baumannii*. *J. Proteome Res.* 13, 460–476. doi: 10.1021/pr400603f

Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., and Gygi, S. P. (2003). Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. U.S.A.* 100, 6940–6945. doi: 10.1073/pnas.0832254100

Gil, C., Solano, C., Burgui, S., Latasa, C., García, B., Toledo-Arana, A., et al. (2014). Biofilm matrix exoproteins induce a protective immune response against *Staphylococcus aureus* biofilm infection. *Infect. Immun.* 82, 1017–1029. doi: 10.1128/IAI.01419-13

Goldstein, B. P. (2014). Resistance to rifampicin: a review. *J. Antibiot.* 67, 625–630. doi: 10.1038/ja.2014.107

Griffin, P. M., Price, G. R., Schooneveldt, J. M., Schlebusch, S., Tilse, M. H., Urbanski, T., et al. (2012). Use of matrix-assisted laser desorption ionization–time of flight mass spectrometry to identify vancomycin-resistant enterococci and investigate the epidemiology of an outbreak. *J. Clin. Microbiol.* 50, 2918–2931. doi: 10.1128/JCM.01000-12

Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotopecoded affinity tags. *Nature Biotechnol.* 17, 994–999.

Hart, P. J., Wey, E., McHugh, T. D., Balakrishnan, I., and Belgacem, O. (2015). A method for the detection of antibiotic resistance markers in clinical strains of *Escherichia coli* using MALDI mass spectrometry. *J. Microbiol. Methods.* 111, 1–8. doi: 10.1016/j.mimet.2015.01.020

Hrabák, J., Chudáčková, E., and Walková, R. (2013). Matrix-assisted laser desorption ionization-time of flight (maldi-tof) mass spectrometry for detection of antibiotic resistance mechanisms: from research to routine diagnosis. *Clin. Microbiol. Rev.* 26, 103–114. doi: 10.1128/CMR.00058-12

Hrabák, J., Walková, R., Studentová, V., Chudáčková, E., and Bergerová, T. (2011). Carbapenemase activity detection by matrix-assisted laser desorption–ionization time of flight mass spectrometry. *J. Clin. Microbiol.* 49, 3222–3227. doi: 10.1128/JCM.00984-11

Imperi, F., Ciccosanti, F., Perdomo, A. B., Tiburzi, F., Mancone, C., Alonzi, T., et al. (2009). Analysis of the periplasmic proteome of *Pseudomonas aeruginosa*, a metabolically versatile opportunistic pathogen. *Proteomics* 9, 1901–1915. doi: 10.1002/pmic.200800618

Iwashkiw, J. A., Seper, A., Weber, B. S., Scott, N. E., Vinogradov, E., Stratilo, C., et al. (2012). Identification of a general O-linked protein glycosylation system

in *Acinetobacter baumannii* and its role in virulence and biofilm formation. *PLoS Pathog.* 8:e1002758. doi: 10.1371/journal.ppat.1002758

Jackson, J., Chen, C., and Buising, K. (2013). Aminoglycosides: how should we use them in the 21st century? *Curr. Opin. Infect. Dis.* 26, 516–525. doi: 10.1097/QCO.0000000000000012

Johansson, Å., Nagy, E., and Sóki, J. (2014). Instant screening and verification of carbapenemase activity in *Bacteroides fragilis* in positive blood culture, using matrix-assisted laser desorption ionization–time of flight mass spectrometry. *J. Med. Microbiol.* 63, 1105–1110. doi: 10.1099/jmm.0.075465-0

Josten, M., Dischinger, J., Szekat, C., Reif, M., Al-Sabti, N., Sahl, H. G., et al. (2014). Identification of agr-positive methicillin-resistant *Staphylococcus aureus* harbouring the class A mec complex by MALDI-TOF mass spectrometry. *Int. J. Med. Microbiol.* 304, 1018–1023. doi: 10.1016/j.ijmm.2014.07.005

Jung, J. S., Eberl, T., Sparbier, K., Lange, C., Kostrzewa, M., Schubert, S., et al. (2014). Rapid detection of antibiotic resistance based on mass spectrometry and stable isotopes. *Eur. J. Clin. Microbiol. Infect. Dis.* 33, 949–955. doi: 10.1007/s10096-013-2031-5

Koning, R. I., de Breij, A., Oostergetel, G. T., Nibbering, P. H., Koster, A. J., and Dijkshoorn, L. (2013). Cryo-electron tomography analysis of membrane vesicles from *Acinetobacter baumannii* ATCC19606 T. *Res. Microbiol.* 164, 397–405. doi: 10.1016/j.resmic.2013.02.007

Kostrzewa, M., Sparbier, K., Maier, T., and Schubert, S. (2013). MALDI-TOF MS: an upcoming tool for rapid detection of antibiotic resistance in microorganisms. *Proteomics Clin. Appl.* 7, 767–778. doi: 10.1002/prca.201300042

Kumar, B., Sharma, D., Sharma, P., Katoch, V. M., Venkatesan, K., and Bisht, D. (2013). Proteomic analysis of *Mycobacterium tuberculosis* isolates resistant to kanamycin and amikacin. *J. Proteomics* 6, 68–77. doi: 10.1016/j.jprot.2013.08.025

Lee, C. R., Lee, J. H., Park, K. S., Jeong, B. C., and Lee, S. H. (2015). Quantitative proteomic view associated with resistance to clinically important antibiotics in Gram-positive bacteria: a systematic review. *Front. Microbiol.* 11:828. doi: 10.3389/fmicb.2015.00828

Lee, J., Kim, S. H., Choi, D. S., Lee, J. S., Kim, D. K., Go, G., et al. (2015). Proteomic analysis of extracellular vesicles derived from *Mycobacterium tuberculosis*. *Proteomics* 15, 3331–3337. doi: 10.1002/pmic.201500037

Lee, W., Chung, H. S., Lee, Y., Yong, D., Jeong, S. H., Lee, K., et al. (2013). Comparison of matrix-assisted laser desorption ionization-time-of-flight mass spectrometry assay with conventional methods for detection of IMP-6, VIM-2, NDM-1, SIM-1, KPC-1, OXA-23, and OXA-51 carbapenemase-producing *Acinetobacter* spp., *Pseudomonas aeruginosa*, and *Klebsiella pneumoniae*. *Diagn. Microbiol. Infect. Dis.* 77, 227–230. doi: 10.1016/j.diagmicrobio.2013.07.005

Li, H., Lin, X. M., Wang, S. Y., and Peng, X. X. (2007). Identification and antibody-therapeutic targeting of chloramphenicol-resistant outer membrane proteins in *Escherichia coli*. *J. Proteome Res.* 6, 3628–3636. doi: 10.1021/pr070307y

Li, H., Wang, B. C., Xu, W. J., Lin, X. M., and Peng, X. X. (2008). Identification and network of outer membrane proteins regulating streptomycin resistance in *Escherichia coli*. *J. Proteome Res.* 7, 4040–4049. doi: 10.1021/pr800310y

Li, H., Zhang, D. F., Lin, X. M., and Peng, X. X. (2015). Outer membrane proteomics of kanamycin-resistant *Escherichia coli* identified MipA as a novel antibiotic resistance-related protein. *FEMS Microbiol Lett.* 362:fnv074. doi: 10.1093/femsle/fnv074

Li, Z. T., Zhang, R. L., Bi, X. G., Xu, L., Fan, M., Xie, D., et al. (2015). Outer membrane vesicles isolated from two clinical *Acinetobacter baumannii* strains exhibit different toxicity and proteome characteristics. *Microb. Pathog.* 81, 46–52. doi: 10.1016/j.micpath.2015.03.009

Lima, T. B., Pinto, M. F. S., Ribeiro, S. M., de Lima, L. A., Viana, J. C., Júnior, N. G., et al. (2013). Bacterial resistance mechanism: what proteomics can elucidate. *FASEB J.* 27, 1291–1303. doi: 10.1096/fj.12-221127

Lin, M. H., Hsu, T. L., Lin, S. Y., Pan, Y. J., Jan, J. T., Wang, J. T., et al. (2009). Phosphoproteomics of *Klebsiella pneumoniae* NTUH-K2044 reveals a tight link between tyrosine phosphorylation and virulence. *Mol. Cell. Proteomics* 8, 2613–2623. doi: 10.1074/mcp.M900276-MCP200

Lin, X., Kang, L., Li, H., and Peng, X. (2014). Fluctuation of multiple metabolic pathways is required for *Escherichia coli* in response to chlortetracycline stress. *Mol. Biosyst.* 10, 901–908. doi: 10.1039/c3mb70522f

Lin, X. M., Li, H., Wang, C., and Peng, X. X. (2008). Proteomic analysis of nalidixic acid resistance in *Escherichia coli*: identification and functional characterization of OM proteins. *J. Proteome Res.* 7, 2399–23405, 1223–1233. doi: 10.1021/pr800073c

Liu, X., Gao, B., Novik, V., and Galán, J. E. (2012). Quantitative proteomics of intracellular *Campylobacter jejuni* reveals metabolic reprogramming. *PLoS Pathog* 8:e1002562. doi: 10.1371/journal.ppat.1002562

Liu, X., Hu, Y., Pai, P. J., Chen, D., and Lam, H. (2014). Label-free quantitative proteomics analysis of antibiotic response in *Staphylococcus aureus* to oxacillin. *J. Proteome Res.* 13, 1223–1233. doi: 10.1021/pr400669d

Ma, Y., Guo, C., Li, H., and Peng, X. X. (2013). Low abundance of respiratory nitrate reductase is essential for *Escherichia coli* in resistance to aminoglycoside and cephalosporin. *J. Proteomics* 11, 78–88. doi: 10.1016/j.jprot.2013.05.019

Madeira, A., dos Santos, S. C., Santos, P. M., Coutinho, C. P., Tyrrell, J., McClean, S., et al. (2013). Proteomic profiling of *Burkholderia cenocepacia* clonal isolates with different virulence potential retrieved from a cystic fibrosis patient during chronic lung infection. *PLoS ONE* 8:e83065. doi: 10.1371/journal.pone.0083065

Marinach, C., Alanio, A., Palous, M., Kwasek, S., Fekkar, A., Brossas, J. Y., et al. (2009). MALDITOF MS-based drug susceptibility testing of pathogens: the example of *Candida albicans* and fluconazole. *Proteomics* 9, 4627–4631. doi: 10.1002/pmic.200900152

Martínez, J. L., and Rojo, F. (2011). Metabolic regulation of antibiotic resistance. *FEMS Microbiol. Rev.* 35, 768–789. doi: 10.1111/j.1574-6976.2011.00282.x

McAtee, C. P., Hoffman, P. S., and Berg, D. E. (2001). Identification of differentially regulated proteins in metronidozole resistant *Helicobacter pylori* by proteome techniques. *Proteomics* 1, 516–521. doi: 10.1002/1615-9861(200104)1:4<516::AID-PROT516>3.0.CO;2-O

Méndez, J. A., Mateos, J., Beceiro, A., Lopez, M., Tomás, M., Poza, M., et al. (2015). Quantitative proteomic analysis of host-pathogen interactions: a study of *Acinetobacter baumannii* responses to host airways. *BMC Genomics* 16:422. doi: 10.1186/s12864-015-1608-z

Mendez, J. A., Soares, N. C., Mateos, J., Gayoso, C., Rumbo, C., Aranda, J., et al. (2012). Extracellular proteome of a highly invasive multidrug-resistant clinical strain of *Acinetobacter baumannii*. *J. Proteome Res.* 7, 5678–5694. doi: 10.1021/pr300496c

Mirrashidi, K. M., Elwell, C. A., Verschueren, E., Johnson, J. R., Frando, A., Von Dollen, J., et al. (2015). Global mapping of the Inc-human interactome reveals that retromer restricts Chlamydia infection. *Cell Host Microbe* 18, 109–121. doi: 10.1016/j.chom.2015.06.004

Monteiro, R., Vitorino, R., Domingues, P., Radhouani, H., Carvalho, C., Poeta, P., et al. (2012). Proteome of a methicillin-resistant *Staphylococcus aureus* clinical strain of sequence type ST398. *J. Proteomics.* 75, 2892–2915. doi: 10.1016/j.jprot.2011.12.036

Moura, I., Monot, M., Tani, C., Spigaglia, P., Barbanti, F., Norais, N., et al. (2014). Multidisciplinary analysis of a nontoxigenic *Clostridium difficile* strain with stable resistance to metronidazole. *Antimicrob. Agents Chemother.* 58, 4957–4960. doi: 10.1128/AAC.02350-14

Nabu, S., Lawung, R., Isarankura-Na-Ayudhya, P., Isarankura-Na-Ayudhya, C., Roytrakul, S., and Prachayasittikul, V. (2014). Reference map and comparative proteomic analysis of *Neisseria gonorrhoeae* displaying high resistance against spectinomycin. *J. Med. Microbiol.* 63, 371–385. doi: 10.1099/jmm.0.067595-0

Nagy, E., Becker, S., Sóki, J., Urbán, E., and Kostrzewa, M. (2011). Differentiation of division I (cfiA-negative) and division II (cfiApositive) *Bacteroides fragilis* strains bymatrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *J. Med. Microbiol.* 60, 1584–1590. doi: 10.1099/jmm.0.031336-0

Nakano, S., Matsumura, Y., Kato, K., Yunoki, T., Hotta, G., Noguchi, T., et al. (2014). Differentiation of vanA-positive *Enterococcus faecium* from vanA-negative *E. faecium* by matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Int. J. Antimicrob. Agents* 44, 256–259. doi: 10.1016/j.ijantimicag.2014.05.006

Neri, A., Mignogna, G., Fazio, C., Giorgi, A., Schininà, M. E., and Stefanelli, P. (2010). *Neisseria meningitidis* rifampicin resistant strains: analysis of protein differentially expressed. *BMC Microbiol.* 24:246. doi: 10.1186/1471-2180-10-246

Olaya-Abril, A., Prados-Rosales, R., McConnell, M. J., Martín-Peña, R., González-Reyes, J. A., Jiménez-Munguía, I., et al. (2014). Characterization of protective extracellular membrane-derived vesicles produced by *Streptococcus pneumoniae*. *J. Proteomics* 25, 46–60. doi: 10.1016/j.jprot.2014.04.023

Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., et al. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* 1, 376–386. doi: 10.1074/mcp.M200025-MCP200

Otto, A., Becher, D., and Schmidt, F. (2014). Quantitative proteomics in the field of microbiology. *Proteomics* 14, 547–565. doi: 10.1002/pmic.201300403

Oviaño, M., Fernández, B., Fernández, A., Barba, M. J., Mouriño, C., and Bou, G. (2014). Rapid detection of enterobacteriaceae producing extended spectrum beta-lactamases directly from positive blood cultures by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Clin. Microbiol. Infect.* 20, 1146–1157. doi: 10.1111/1469-0691.12729

Pakharukova, N., Garnett, J. A., Tuittila, M., Paavilainen, S., Diallo, M., Xu, Y., et al. (2015). Structural insight into archaic and alternative chaperone usher pathways reveals a novel mechanism of pilus biogenesis. *PLoS Pathog.* 11:e1005269. doi: 10.1371/journal.ppat.1005269

Papagiannitsis, C. C., Študentová, V., Izdebski, R., Oikonomou, O., Pfeifer, Y., Petinaki, E., et al. (2015). Matrix-assisted laser desorption ionization-time of flight mass spectrometry meropenem hydrolysis assay with NH_4HCO_3, a reliable tool for direct detection of carbapenemase activity. *J. Clin. Microbiol.* 53, 1731–1735. doi: 10.1128/JCM.03094-14

Park, A. J., Krieger, J. R., and Khursigara, C. M. (2016). Survival proteomes: the emerging prototype of antimicrobial resistance. *FEMS Microbiol. Rev.* 40:fuv051. doi: 10.1093/femsre/fuv051

Peaper, D. R., Kulkarni, M. V., Tichy, A. N., Jarvis, M., Murray, T. S., and Hodsdon, M. E. (2013). Rapid detection of carbapenemase activity through monitoring ertapenem hydrolysis in Enterobacteriaceae with LC-MS/MS. *Bioanalysis* 5, 147–157. doi: 10.4155/bio.12.310

Peng, X., Xu, C., Ren, H., Lin, X., Wu, L., and Wang, S. (2005). Proteomic analysis of the sarcosine-insoluble outer membrane fraction of *Pseudomonas aeruginosa* responding to ampicilin, kanamycin, and tetracycline resistance. *J. Proteome Res.* 4, 2257–2265. doi: 10.1021/pr050159g

Pérez-Cruz, C., Carrión, O., Delgado, L., Martinez, G., López-Iglesias, C., and Mercade, E. (2013). New type of outer membrane vesicle produced by the Gram-negative bacterium *Shewanella vesiculosa* M7T: implications for DNA content. *Appl. Environ. Microbiol.* 79, 1874–1881. doi: 10.1128/AEM.03657-12

Pérez-Llarena, F. J., and Bou, G. (2009). Beta-lactamase inhibitors: the story so far. *Curr. Med. Chem.* 16, 3740–3765. doi: 10.2174/092986709789104957

Pieper, R., Fisher, C. R., Suh, M. J., Huang, S. T., Parmar, P., and Payne, S. M. (2013). Analysis of the proteome of intracellular *Shigella flexneri* reveals pathways important for intracellular growth. *Infect. Immun.* 81, 4635–4648. doi: 10.1128/IAI.00975-13

Poole, K. (2004). Resistance to beta-lactam antibiotics. *Cell. Mol. Life Sci.* 61, 2200–2223. doi: 10.1007/s00018-004-4060-9

Pournaras, S., Poulou, A., Dafopoulou, K., Chabane, Y. N., Kristo, I., Makris, D., et al. (2014).Growth retardation, reduced invasiveness, and impaired colistin-mediated cell death associated with colistin resistance development in *Acinetobacter baumannii*. *Antimicrob. Agents Chemother.* 58, 828–832. doi: 10.1128/AAC.01439-13

Provenzano, J. C., Siqueira, J. F. Jr., Rôças, I. N., Domingues, R. R., Paes Leme, A. F., and Silva, M. R. (2013). Metaproteome analysis of endodontic infections in association with different clinical conditions. *PLoS ONE* 15:e76108. doi: 10.1371/journal.pone.0076108

Purschke, F. G., Hiller, E., Trick, I., and Rupp, S. (2012). Flexible survival strategies of *Pseudomonas aeruginosa* in biofilms result in increased fitness compared with *Candida albicans*. *Mol. Cell. Proteomics* 11, 1652–1669. doi: 10.1074/mcp.M112.017673

Pusch, W., Wurmbach, J. H., Thiele, H., and Kostrzewa, M. (2002). MALDI-TOF mass spectrometry-based SNP genotyping. *Pharmacogenomics* 3, 537–548. doi: 10.1517/14622416.3.4.537

Ramos, S., Chafsey, I., Silva, N., Hébraud, M., Santos, H., Capelo-Martinez, J. L., et al. (2015). Effect of vancomycin on the proteome of the multiresistant *Enterococcus faecium* SU18 strain. *J. Proteomics* 15, 378–387. doi: 10.1016/j.jprot.2014.10.012

Ravichandran, A., Sugiyama, N., Tomita, M., Swarup, S., and Ishihama, Y. (2009). Ser/Thr/Tyr phosphoproteome analysis of pathogenic and non-pathogenic Pseudomonas species. *Proteomics* 9, 2764–2775. doi: 10.1002/pmic.200800655

Redgrave, L. S., Sutton, S. B., Webber, M. A., and Piddock, L. J. (2014). Fluoroquinolone resistance: mechanisms, impact on bacteria, and

role in evolutionary success. *Trends Microbiol.* 22, 438–445. doi: 10.1016/j.tim.2014.04.007

Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., et al. (2004). Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* 3, 1154–1169. doi: 10.1074/mcp.M400129-MCP200

Sandalakis, V., Psaroulaki, A., De Bock, P. J., Christidou, A., Gevaert, K., Tsiotis, G., et al. (2012). Investigation of rifampicin resistance mechanisms in Brucella abortus using MS-driven comparative proteomics. *J. Proteome Res.* 6, 2374–2385. doi: 10.1021/pr201122w

Schmidl, S. R., Gronau, K., Pietack, N., Hecker, M., Becher, D., and Stülke, J. (2010). The phosphoproteome of the minimal bacterium *Mycoplasma pneumoniae*: analysis of the complete known Ser/Thr kinome suggests the existence of novel kinases. *Mol. Cell. Proteomics* 9, 1228–1242. doi: 10.1074/mcp.M900267-MCP200

Schmidt, A., Kellermann, J., and Lottspeich, F. (2005). A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics* 5, 4–15. doi: 10.1002/pmic.200400873

Scott, N. E., Marzook, N. B., Cain, J. A., Solis, N., Thaysen-Andersen, M., Djordjevic, S. P., et al. (2014). Comparative proteomics and glycoproteomics reveal increased N-linked glycosylation and relaxed sequon specificity in *Campylobacter jejuni* NCTC11168 O. *J. Proteome Res.* 7, 5136–5150. doi: 10.1021/pr5005554

Seneviratne, C. J., Wang, Y., Jin, L., Wong, S. S., Herath, T. D., and Samaranayake, L. P. (2012). Unraveling the resistance of microbial biofilms:has proteomics been helpful? *Proteomics* 12, 651–665. doi: 10.1002/pmic.201100356

Shah, H. N., Rajakaruna, L., Ball, G., Misra, R., Al-Shahib, A., Fang, M., et al. (2011). Tracing the transition of methicillin resistance in sub-populations of *Staphylococcus aureus*, using SELDI-TOF mass spectrometry and artificial neural network analysis. *Syst. Appl. Microbiol.* 34, 81–86. doi: 10.1016/j.syapm.2010.11.002

Shao, C., Sun, Y., Wang, N., Yu, H., Zhou, Y., Chen, C., et al. (2013). Changes of proteome components of *Helicobacter pylori* biofilms induced by serum starvation. *J. Mol. Med. Rep.* 8, 1761–1766. doi: 10.3892/mmr.2013.1712

Sharma, P., Kumar, B., Gupta, Y., Singhal, N., Katoch, V. M., Venkatesan, K., et al. (2010). Proteomic analysis of streptomycin resistant and sensitive clinical isolates of *Mycobacterium tuberculosis*. *Proteome Sci.* 18:59. doi: 10.1186/1477-5956-8-59

Smiley, R., Bailey, J., Sethuraman, M., Posecion, N., and Showkat Ali, M. (2013). Comparative proteomics analysis of sarcosine insoluble outer membrane proteins from clarithromycin resistant and sensitive strains of *Helicobacter pylori*. *J. Microbiol.* 51, 612–618. doi: 10.1007/s12275-013-3029-5

Soares, N. C., Spät, P., Méndez, J. A., Nakedi, K., Aranda, J., and Bou, G. (2014). Ser/Thr/Tyr phosphoproteome characterization of *Acinetobacter baumannii*: comparison between a reference strain and a highly invasive multidrug-resistant clinical isolate. *J. Proteomics* 6, 113–114. doi: 10.1016/j.jprot.2014.03.009

Solis, N., Parker, B. L., Kwong, S. M., Robinson, G., Firth, N., and Cordwell, S. J. (2014). *Staphylococcus aureus* surface proteins involved in adaptation to oxacillin identified using a novel cell shaving approach. *J. Proteome Res.* 6, 2954–2972. doi: 10.1021/pr500107p

Soufi, B., Soares, N. C., Ravikumar, V., and Macek, B. (2012). Proteomics reveals evidence of cross-talk between protein modifications in bacteria: focus on acetylation and phosphorylation. *Curr. Opin. Microbiol.* 15, 357–363. doi: 10.1016/j.mib.2012.05.003

Sparbier, K., Lange, C., Jung, J., Wieser, A., Schubert, S., and Kostrzewa, M. (2013). MALDI biotyper-based rapid resistance detection by stable-isotope labeling. *J. Clin. Microbiol.* 51, 3741–3748. doi: 10.1128/JCM.01536-13

Steeb, B., Claudi, B., Burton, N. A., Tienz, P., Schmidt, A., Farhan, H., et al. (2013). Parallel exploitation of diverse host nutrients enhances Salmonella virulence. *PLoS Pathog.* 9:e1003301. doi: 10.1371/journal.ppat.1003301

Studentova, V., Papagiannitsis, C. C., Izdebski, R., Pfeifer, Y., Chudackova, E., Bergerova, T., et al. (2015). Detection of OXA-48-type carbapenemase-producing Enterobacteriaceae in diagnostic laboratories can be enhanced by addition of bicarbonates to cultivation media or reaction buffers. *Folia Microbiol.* 60, 119–129. doi: 10.1007/s12223-014-0349-8

Sun, X., Ge, F., Xiao, C. L., Yin, X. F., Ge, R., Zhang, L. H., et al. (2010). Phosphoproteomic analysis reveals the multiple roles of phosphorylation in

pathogenic bacterium *Streptococcus pneumoniae*. *J. Proteome Res.* 9, 275–282. doi: 10.1021/pr900612v

Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., et al. (2003). Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* 75, 1895–1904. doi: 10.1021/ac0262560

Tiwari, V., and Tiwari, M. (2014). Quantitative proteomics to study carbapenem resistance in *Acinetobacter baumannii*. *Front. Microbiol.* 5:512. doi: 10.3389/fmicb.2014.00512

Tiwari, V., Vashistt, J., Kapil, A., and Moganty, R. R. (2012). Comparative proteomics of inner membrane fraction from carbapenem-resistant *Acinetobacter baumannii* with a reference strain. *PLoS ONE* 7:e39451. doi: 10.1371/journal.pone.0039451

Tomaras, A. P., Flagler, M. J., Dorsey, C. W., Gaddy, J. A., and Actis, L. A. (2008). Characterization of a two-component regulatory system from *Acinetobacter baumannii* that controls biofilm formation and cellular morphology. *Microbiology* 154, 3398–3409. doi: 10.1099/mic.0.2008/019471-0

Trip, H., Mende, K., Majchrzykiewicz-Koehorst, J. A., Sedee, N. J., Hulst, A. G., Jansen, H. J., et al. (2015). Simultaneous identification of multiple β-lactamases in *Acinetobacter baumannii* in relation to carbapenem and ceftazidime resistance, using liquid chromatography-tandem mass spectrometry. *J. Clin. Microbiol.* 53, 1927–1930. doi: 10.1128/JCM.00620-15

Van Oudenhove, L., and Devreese, B. (2013). A review on recent developments in mass spectrometry instrumentation and quantitative tools advancing bacterial proteomics. *Appl. Microbiol. Biotechnol.* 97, 4749–4762. doi: 10.1007/s00253-013-4897-7

Vella, A., De Carolis, E., Vaccaro, L., Posteraro, P., Perlin, D. S., Kostrzewa, M., et al. (2013). Rapid antifungal susceptibility testing by matrix-assisted laser desorption ionization time-of-flight mass spectrometry analysis. *J. Clin. Microbiol.* 51, 2964–2969. doi: 10.1128/JCM.00903-13

Vitoriano, I., Saraiva-Pava, K. D., Rocha-Gonçalves, A., Santos, A., Lopes, A. I., Oleastro, M., et al. (2011). Ulcerogenic *Helicobacter pylori* strains isolated from children: a contribution to get insight into the virulence of the bacteria. *PLoS ONE* 6:e26265. doi: 10.1371/journal.pone.0026265

Vranakis, I., De Bock, P. J., Papadioti, A., Tselentis, Y., Gevaert, K., Tsiotis, G., et al. (2011). Identification of potentially involved proteins in levofloxacin resistance mechanisms in *Coxiella burnetii*. *J. Proteome Res.* 10, 756–762. doi: 10.1021/pr100906v

Wang, L., Han, C., Sui, W., Wang, M., and Lu, X. (2013). MALDI-TOF MS applied to indirect carbapenemase detection: a validated procedure to clearly distinguish between carbapenemase-positive and carbapenemase-negative bacterial strains. *Anal. Bioanal. Chem.* 405, 5259–5266. doi: 10.1007/s00216-013-6913-2

Wang, X., He, X., Jiang, Z., Wang, J., Chen, X., Liu, D., et al. (2010). Proteomic analysis of the *Enterococcus faecalis* V583 strain and clinical isolate V309 under vancomycin treatment. *J. Proteome Res.* 9, 1772–1785. doi: 10.1021/pr901216e

Wang, Y., Huang, K. Y., and Huo, Y. (2014). Proteomic comparison between Salmonella Typhimurium and Salmonella Typhi. *J. Microbiol.* 52, 71–76. doi: 10.1007/s12275-014-3204-3

Waxman, D. J., and Strominger, J. L. (1983). Penicillin-binding proteins and the mechanism of action of beta-lactam antibiotics. *Annu. Rev. Biochem.* 52, 825–869. doi: 10.1146/annurev.bi.52.070183.004141

Wurpel, D. J., Moriel, D. G., Totsika, M., Easton, D. M., and Schembri, M. A. (2015). Comparative analysis of the uropathogenic *Escherichia coli* surface proteome by tandem mass-spectrometry of artificially induced outer membrane vesicles. *J. Proteomics* 6, 93–106. doi: 10.1016/j.jprot.2014.12.005

Wybo, I., De Bel, A., Soetens, O., Echahidi, F., Vandoorslaer, K., Van Cauwenbergh, M., et al. (2011). Differentiation of cfiA-negative and cfiA-positive *Bacteroides fragilis* isolates by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J. Clin. Microbiol.* 49, 1961–1964. doi: 10.1128/JCM.02321-10

Xie, L., Wang, X., Zeng, J., Zhou, M., Duan, X., Li, Q., et al. (2015). Proteome-wide lysine acetylation profiling of the human pathogen *Mycobacterium tuberculosis*. *Int. J. Biochem. Cell Biol.* 59, 193–202. doi: 10.1016/j.biocel.2014.11.010

Yang, Y., Hu, M., Yu, K., Zeng, X., and Liu, X. (2015). Mass spectrometry-based proteomic approaches to study pathogenic bacteria-host interactions. *Protein Cell* 6, 265–274. doi: 10.1007/s13238-015-0136-6

Yun, S. H., Choi, C. W., Park, S. H., Lee, J. C., Leem, S. H., Choi, J. S., et al. (2008). Proteomic analysis of outer membrane proteins from *Acinetobacter baumannii* DU202 in tetracycline stress condition. *J. Microbiol.* 46, 720–727. doi: 10.1007/s12275-008-0202-3

Zhou, J., Hao, D., Wang, X., Liu, T., He, C., Xie, F., et al. (2006). An important role of a "probable ATP-binding component of ABC transporter" during the process of *Pseudomonas aeruginosa* resistance to fluoroquinolone. *Proteomics* 6, 2495–2503. doi: 10.1002/pmic.200501354

Zielke, R. A., Wierzbicki, I. H., Weber, J. V., Gafken, P. R., and Sikora, A. E. (2014). Quantitative proteomics of the *Neisseria gonorrhoeae* cell envelope and membrane vesicles for the discovery of potential therapeutic targets. *Mol. Cell. Proteomics* 13, 1299–12317. doi: 10.1074/mcp.M113.029538

Mass Spectrometry-Based Bacterial Proteomics: Focus on Dermatologic Microbial Pathogens

Youcef Soufi[1] and Boumediene Soufi[2]*

[1] *College of Medicine, University of Manitoba, Winnipeg, MB, Canada,* [2] *Independent Academic Scholar, Magdeburg, Germany*

The composition of human skin acts as a natural habitat for various bacterial species that function in a commensal and symbiotic fashion. In a healthy individual, bacterial flora serves to protect the host. Under certain conditions such as minor trauma, impaired host immunity, or environmental factors, the risk of developing skin infections is increased. Although a large majority of bacterial associated skin infections are common, a portion can potentially manifest into clinically significant morbidity. For example, Gram-positive species that typically reside on the skin such as *Staphylococcus* and *Streptococcus* can cause numerous epidermal (impetigo, ecthyma) and dermal (cellulitis, necrotizing fasciitis, erysipelas) skin infections. Moreover, the increasing incidence of bacterial antibiotic resistance represents a serious challenge to modern medicine and threatens the health care system. Therefore, it is critical to develop tools and strategies that can allow us to better elucidate the nature and mechanism of bacterial virulence. To this end, mass spectrometry (MS)-based proteomics has been revolutionizing biomedical research, and has positively impacted the microbiology field. Advances in MS technologies have paved the way for numerous bacterial proteomes and their respective post translational modifications (PTMs) to be accurately identified and quantified in a high throughput and robust fashion. This technological platform offers critical information with regards to signal transduction, adherence, and microbial–host interactions associated with bacterial pathogenesis. This mini-review serves to highlight the current progress proteomics has contributed toward the understanding of bacteria that are associated with skin related diseases, infections, and antibiotic resistance.

Keywords: mass spectrometry, proteomics, pathogenic bacteria, skin disease, dermatology, bacterial resistance

INTRODUCTION

Skin has a primary role to act as a physical barrier in order to protect the body from temperature variations, microorganisms, or toxic substrates. While *in utero* skin is sterile, after birth it becomes rapidly colonized with numerous microorganisms which aids toward protecting the body (Capone et al., 2011). Bacterially derived dermatologic conditions can range from relatively benign conditions (assuming non-immunocompromised patients) such as cellulitis, erysipelas, folliculitis, or impetigo to serious clinical morbidity, such as necrotizing fasciitis (Stulberg et al., 2002). Underlying medical conditions such as diabetes mellitus or AIDS are examples where

commensal bacteria can invade the skin resulting in infections that can manifest from mild to the life threatening (Swartz, 2004).

In recent years there have been growing concerns with respect to the challenges faced in the emergent inability to provide effective therapeutic interventions for patients with pathologies of bacterial etiology, due to increasing rates of antibiotic resistance. Considering the ease of transmission with respect to resistance genes amongst mucosal and epidermal flora, special consideration should be given to dermatologic related bacterial conditions (Espersen, 1998).

Mass spectrometry (MS)-based proteomics has the capability to study proteins and their interactions in order to better understand dysregulations in infection disorders (List et al., 2008), reveal antibiotic resistance mechanisms (Lee et al., 2015) and significant new targets for future drug discovery. Current MS-based proteomics technologies have advanced to the point where they are amenable to any biological system. With regards to bacterial organisms, they are particularly attractive models to apply proteomics based approaches due to their smaller proteomes and modifications compared to eukaryotes allowing for comprehensive proteome coverage (Stekhoven et al., 2014).

SKIN MICROBIOME

The skin is host to many different bacterial species, fungi, viruses and mites. Most of the bacteria fall into four phyla: Actinobacteria, Firmicutes, Bacteroidetes, and Proteobacteria from which *Propionibacterium*, *Corynebacterium*, *Staphylococcus*, *Micrococcus*, *Streptococcus*, and *Brevibacterium* are highly abundant (van Rensburg et al., 2015) Gram-positive bacteria, since Gram-negative members typically cannot cope with the relatively dry environment of healthy human skin (Del Rosso and Leyden, 2007). In certain physiological conditions where the skin becomes moist, some Gram-negative bacteria can colonize such as *Acinetobacter* sp. (Del Rosso and Leyden, 2007). The diverse community of the skin, known as the skin microbiome, is defined by host physiology, host genotype, immune system, environment, and lifestyle (Grice and Segre, 2011). However, most of these microorganisms are harmless and many bacterial species act in a commensal fashion by which each species benefits through the exchange of nutrients as well as protect the host from pathogens without negatively affecting each other (Mankowska-Wierzbicka et al., 2015). Knowledge toward the bacterial microbiome is mainly derived from the conventional culture-based approaches, although in recent years DNA sequencing technologies combined with bioinformatic analysis and metagenomics approaches have allowed comprehensive examination of microbial communities (Kuczynski et al., 2012), yet uncertainties remain as to what defines a "normal" microbiome (Backhed et al., 2012).

The underlying contribution of the microbiome in the clinical picture of skin disorders is clear from the perspective of antibacterial treatment, however, the molecular dynamics between the microbiome and host remains largely unknown. Disturbances in homeostasis between microbiome and host can manifest into skin disorders such as atopic dermatitis (AD) or psoriasis (Kuo et al., 2013). Both disorders are connected with dysregulation of the skin immune response. However, while AD lesions are characterized by low level, psoriatic lesions are characterized by high level of antimicrobial peptide production (Nomura et al., 2003). Additionally, AD lesions are regularly infected with microbial pathogens.

The bacterial genus of *Staphylococcus* sp. is composed of approximately 40 different species both commensal (unable to produce virulence factor coagulase) and pathogenic, and play an important role in skin health and pathology (Del Rosso and Leyden, 2007). For example, *Staphylococcus aureus* is a medically relevant bacterial pathogen that is associated with the increasing rate of antibiotic resistance mechanisms as well as systemic and cutaneous infections (Becker et al., 2007), including AD (Hanifin and Rogge, 1977). However, it was previously thought that *S. aureus* causes AD alone, a recent study showed that the skin microbiome composition has a temporal change and it depends on disease flares and treatment. In the active state of the disease, the abundance of *S. aureus* and skin commensal *S. epidermidis* was increased, while *Streptococcus, Propionibacterium,* and *Corynebacterium* were increased following therapy (Kong et al., 2012). Increased abundance of *S. epidermidis* could reflect a microbial response to overgrowth of *S. aureus,* since it can selectively inhibit *S. aureus* growth (Iwase et al., 2010; Zeeuwen et al., 2013). Moreover, a recent study identified the presence of a distinct microbiome capable of direct communication with the host at the sub-epidermal compartments of the skin, an area previously thought to be sterile (Nakatsuji et al., 2013). While these types of studies employing different techniques have characterized the bacterial species that make up the skin microbiome, proteomics is required in order to identify proteins and their respective pathways involved during these interactions, which could directly contribute to the better understanding of the complex interplay between microbiome and host.

SHOTGUN MS-BASED PROTEOMICS

Due to the progress made toward state of the art next generation sequencing methodologies applied to generate fully annotated genomes, the number of fully sequenced microbial genomes has increased dramatically (Ribeiro et al., 2012). This has enabled state of the art MS-based proteomics technologies to rapidly advance in order to study microbial models and their communities in a robust and systematic fashion.
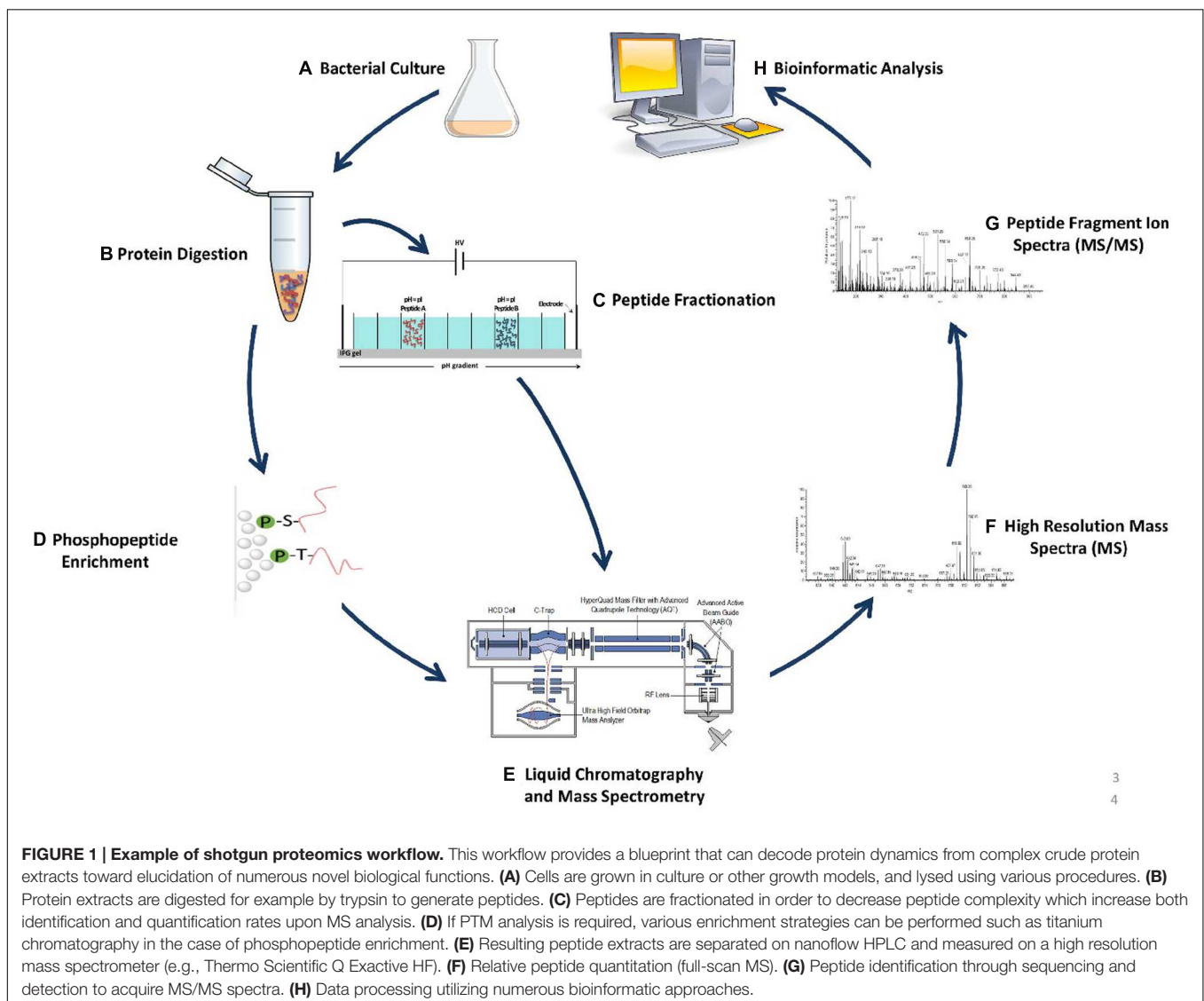
Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), followed by MS analysis has served as the main proteomics method of choice in the past and has been utilized in many skin associated bacterial pathogen studies (Kohler et al., 2005; Becher et al., 2009; Francois et al., 2014). In this technique, proteins are separated according to their molecular weight and isoelectric point migrating and accumulating as protein spots on the gel matrix. Although this approach can resolve multiples of thousands of proteins, it does carry some disadvantages especially with respect toward the identification of low abundant proteins, proteins with extremely high and low molecular weights, and

is rather impractical with respect to large scale site specific analysis of post translational modifications (PTMs) such as phosphorylation.

The inherent limitations associated with 2D-PAGE MS lead to the development of gel-free MS-based approaches better known as "shotgun" proteomics. Advances made in shotgun based proteomics approaches is largely due to technological improvements made in high performance mass spectrometers. Since most biological samples consist of very complex peptide mixtures, mass spectrometers must be capable of ensuring a deep analytical coverage while maintaining a high level of robustness, sensitivity, and measurement accuracy (Mann and Kelleher, 2008). Newer generations of hybrid mass spectrometers such as but not limited to the LTQ-Orbitrap (Hu et al., 2005), LTQ Orbitrap Velos (Olsen et al., 2009), Q-Exactive HF (Scheltema et al., 2014), and Orbitrap Fusion (Erickson et al., 2015) can achieve precise mass accuracy at very high acquisition speeds, and resolution which allows for a complete

sampling of complex peptide extracts providing comprehensive proteome coverage. Combined with high performance liquid chromatography (HPLC) technologies, these LC-MS workflows also referred to as gel-free shotgun proteomics allow for the quantification and identification of entire proteomes as well as their protein modifications across different biological samples. An overview of a typical shotgun MS-based experimental workflow is illustrated in **Figure 1**.

For example, a similar approach as described above was applied to the bacterium *Propionibacterium acnes* (Bek-Thomsen et al., 2014). *P. acnes* is known to be associated with the inflammatory condition of the sebaceous human follicles known as acne vulgaris (Williams et al., 2012). The authors utilized proteomics in order to identify both human and *P. acnes* proteins using sebaceous follicular casts. To this end, many proteins involved in wound healing, inflammation, and tissue formation were identified. The most abundant *P. acnes* proteins were CAMP factors, and surface exposed dermatan sulfate adhesins.

FIGURE 1 | Example of shotgun proteomics workflow. This workflow provides a blueprint that can decode protein dynamics from complex crude protein extracts toward elucidation of numerous novel biological functions. **(A)** Cells are grown in culture or other growth models, and lysed using various procedures. **(B)** Protein extracts are digested for example by trypsin to generate peptides. **(C)** Peptides are fractionated in order to decrease peptide complexity which increase both identification and quantification rates upon MS analysis. **(D)** If PTM analysis is required, various enrichment strategies can be performed such as titanium chromatography in the case of phosphopeptide enrichment. **(E)** Resulting peptide extracts are separated on nanoflow HPLC and measured on a high resolution mass spectrometer (e.g., Thermo Scientific Q Exactive HF). **(F)** Relative peptide quantitation (full-scan MS). **(G)** Peptide identification through sequencing and detection to acquire MS/MS spectra. **(H)** Data processing utilizing numerous bioinformatic approaches.

This study is the first of its kind in terms of analyzing the proteomes of both humans and bacteria on sebaceous follicular casts and demonstrates the importance of employing shotgun based proteomics workflows (Bek-Thomsen et al., 2014).

A recent study employed a combinatorial approach utilizing 16SrRNA sequencing with ultra-performance liquid chromatography/quadrupole time of flight (UPLC-TOF) MS toward the generation of a 3-dimensional topographical map of microbes and their molecules of origin (peptides, metabolites) distributed on the surface of the human skin (Bouslimani et al., 2015). This information was utilized to identify microbial species present and correlated with the chemical environment of the skin and serves as a powerful approach toward elucidating how the microbiome interacts and subsequently modifies different areas of the human skin in a species specific fashion.

SHOTGUN MS-BASED QUANTITATIVE PROTEOMICS

The complete identification of a proteome and its respective PTMs provides a better understanding toward biological function and regulation. However, it is equally important to quantify dynamics of proteins relative to each other under different biological conditions, disease states, and perturbations. This information can be obtained through the use of quantitative proteomics of which numerous approaches have been developed that are compatible with shotgun MS-based proteomics strategies.

One common approach is to introduce a stable non-radioactive isotope label either chemically or metabolically on the peptide or protein level. Through this technique, the relative intensities of peptides are measured, via MS thus identifying and quantifying regulated proteins from different samples through the obtained mass spectra. Examples of such techniques include but they are not limited to Stable Isotope Labeling by Amino acids in Cell culture (SILAC; Ong et al., 2002) or ^{15}N labeling (Hempel et al., 2010; Soufi et al., 2015), both of which have been successfully implemented in a variety of bacteria (Soufi et al., 2010, 2015; Hempel et al., 2011; Soares et al., 2013; Misra et al., 2014; Soufi and Macek, 2014; Boysen et al., 2015). Furthermore, an innovative MS-based technique known as cell type-specific labeling using amino acid precursors (CTAP) has the ability to continuously label the proteome of individual cell types actively growing in a co-culture environment, allowing for the elucidation of the "cell-of-origin" of proteins in multicellular environments (Gauthier et al., 2013). Although this method was demonstrated in eukaryotic models, the application toward bacteria especially those in multispecies environments, such as human microbiomes, biofilms, etc. could significantly contribute toward the better understanding of the dynamics between different bacterial species inhabiting the same environment.

Although metabolic labeling approaches such as SILAC are considered by many to be the most accurate method toward global protein quantitation (Ong et al., 2002), they may be difficult to implement due to technical challenges with incorporation of the chemical or metabolic labeling approach,

or challenges posed in complex biological models such as those involving the skin. Therefore, due to rapid advances made in LC-MS, label free quantitation (LFQ) approaches are now possible and routinely performed (Neilson et al., 2011). In this approach, different peptide samples are measured via MS and compared either by the total number of sequenced (MS/MS) spectra or the total extracted ion currents under controlled conditions.

The LFQ methodology was successfully applied to study the role of S. aureus in patients with ectodermal dysplasia and AD (Burian et al., 2015). Patients suffering from AD are highly correlated with the presence of opportunistic S. aureus infections due to having a reduced immune response, and have been shown to contain a lower amount of peptides originating from the natural antimicrobial dermcidin found in sweat leading to decreased levels of antimicrobial activity (Schittek, 2011). The authors proved through proteomics analysis of the secretome of diseased vs. healthy patients that a similar mechanism of reduced sweat derived dermcidin also exists in patients with ectodermal dysplasia, making these individuals highly prone to acquiring S. aureus infections (Burian et al., 2015).

Quantitative proteomics techniques can be utilized to estimate the absolute molar amount or concentration of a particular protein per cell. This information is highly relevant especially within the clinical context (Rodriguez-Suarez and Whetton, 2013). Many variants of this technique exist and all involve spiking in a known concentration of an internal protein standard into the sample, followed by MS analysis, and comparing resulting sample peptide measurements to the internal standard. Techniques include FLEXIQuant (Singh et al., 2009), absolute quantification using protein epitope signature tags (PrEST; Zeiler et al., 2012), intensity based absolute quantification (iBAQ; Schwanhausser et al., 2011), and absolute quantification (AQUA; Gerber et al., 2003). These methods can be applied to bacteria and serve as an important tool toward understanding the underlying mechanisms during bacterial virulence and antibiotic resistance.

POST TRANSLATIONAL MODIFICATIONS (PTMS)

Various PTMs on bacterial proteins such as phosphorylation, acetylation, methylation, and deamidation serve as an efficient means of controlling signal transduction, virulence and regulatory processes. PTMs represent a significant process in the life cycle of bacteria and can modulate key virulence factors and are attractive targets for novel therapies.

Detecting these PTMs in bacteria poses a technical challenge due to the fact that they are difficult to discover as these modifications typically exist at low levels of abundance. To circumvent this issue, specific enrichment strategies targeting certain PTMs can be utilized in order to decrease peptide complexity thereby increasing the likeliness of detection and subsequent characterization. For example, immunoaffinity enrichment is typically employed to select for lysine acetylated peptides (Rardin et al., 2013). Moreover, similar enrichment strategies are employed to capture phosphorylation events on serine, threonine, and tyrosine (S/T/Y) amino acid residues.

These PTMs were once thought to exist solely in eukaryotes, however, MS-based proteomics combined with phosphopeptide enrichment strategies such as titanium dioxide chromatography (TiO$_2$; Pinkse et al., 2004), immunoprecipitation (Rush et al., 2005) or immobilized metal ion affinity chromatography (IMAC; Villen and Gygi, 2008) have now established S/T/Y phosphorylation as a frequent and important PTM among different bacterial species (Macek et al., 2007, 2008; Soufi et al., 2008; Lin et al., 2009; Sun et al., 2010; Manteca et al., 2011; Misra et al., 2011; Cousin et al., 2013).

A recent shotgun based high resolution LC-MS/MS approach involving the enrichment of surface proteins using "trypsin shaving" was applied to the opportunistic pathogen *S. aureus* toward the identification of hydroxymethylation on aspargine and glutamine amino acid residues in an attempt to identify the presence and potential regulatory importance of this PTM on surface proteins which are known to assist in the colonization and invasion of the host cell (Waridel et al., 2012). The authors reported a total of 15 proteins (mostly surface proteins) that contained hydroxymethylation modifications and could play a role in virulence factor modulation.

Pseudomonas aeruginosa, an opportunistic bacterial pathogen, is associated with immunocompromised patients and nosocomial infections. *P. aeruginosa* infections can enter the body through the skin, and can be life threatening due to their ability to develop antibiotic resistance (Ouidir et al., 2015). A study involving *P. aeruginosa,* employed an effective combinatorial method of immunoaffinity assays, complex peptide fractionation, and shotgun proteomics toward the identification and characterization of lysine acetylation, a reversible PTM that has recently been implicated as an important regulatory mechanism in many bacteria (Yu et al., 2008; Kim et al., 2013; Zhang et al., 2013; Liao et al., 2014). This approach lead to the identification of 320 acetylated proteins in a wide variety of functional classes. The study also identified novel lysine acetylation events in virulence factors known to assist in host immune response evasion such as chitin binding protein, serine protease, exotoxin A, and hemolysin which potentially implies that lysine acetylation events in *P. aeruginosa* plays a role in mechanisms involving virulence (Ouidir et al., 2015).

Cysteine phosphorylation in *S. aureus* was shown to assist in the regulation of bacterial virulence and vancomycin resistance (Sun et al., 2012). Utilizing high resolution MS, the authors elucidated in a site specific fashion, that cysteine phosphorylation events occurred in various proteins many of which are global regulators that control important biological processes (Sun et al., 2012). Moreover, the eukaryotic-like Ser/Thr kinase and phosphatase pair Stk1/Stp1 was found to regulate cysteine phosphorylation in many Gram-positive bacteria providing an important piece of information toward the underlying regulatory mechanism of these events (Sun et al., 2012).

CONCLUSION AND FUTURE OUTLOOK

A wide range of dermatological microbial associated diseases presents current and future challenges to health care providers. While in the last decade metagenomics data provided a higher level of understanding of the microbial skin environment, understanding signal transduction, virulence, regulatory processes, and dynamics between different bacterial species is essential in order to improve the overall standards and quality of patient care and treatment. MS-based proteomics has the capability to provide this knowledge. To this end, a ground-breaking endeavor with proteomics at the forefront is required in order to elucidate the specific mechanisms involved in skin infections, bacterial resistance as well as the complex microbiome and host relationship.

AUTHOR CONTRIBUTIONS

YS wrote the manuscript. BS assisted with writing of the manuscript and generated the figure.

ACKNOWLEDGMENT

REFERENCES

Backhed, F., Fraser, C. M., Ringel, Y., Sanders, M. E., Sartor, R. B., Sherman, P. M., et al. (2012). Defining a healthy human gut microbiome: current concepts, future directions, and clinical applications. *Cell Host Microbe* 12, 611–622. doi: 10.1016/j.chom.2012.10.012

Becher, D., Hempel, K., Sievers, S., Zuhlke, D., Pane-Farre, J., Otto, A., et al. (2009). A proteomic view of an important human pathogen–towards the quantification of the entire *Staphylococcus aureus* proteome. *PLoS ONE* 4:e8176. doi: 10.1371/journal.pone.0008176

Becker, K., Bierbaum, G., Von Eiff, C., Engelmann, S., Gotz, F., Hacker, J., et al. (2007). Understanding the physiology and adaptation of staphylococci: a post-genomic approach. *Int. J. Med. Microbiol.* 297, 483–501. doi: 10.1016/j.ijmm.2007.04.004

Bek-Thomsen, M., Lomholt, H. B., Scavenius, C., Enghild, J. J., and Bruggemann, H. (2014). Proteome analysis of human sebaceous follicle infundibula extracted from healthy and acne-affected skin. *PLoS ONE* 9:e107908. doi: 10.1371/journal.pone.0107908

Bouslimani, A., Porto, C., Rath, C. M., Wang, M., Guo, Y., Gonzalez, A., et al. (2015). Molecular cartography of the human skin surface in 3D. *Proc. Natl. Acad. Sci. U.S.A.* 112, E2120–E2129. doi: 10.1073/pnas.1424409112

Boysen, A., Borch, J., Krogh, T. J., Hjerno, K., and Moller-Jensen, J. (2015). SILAC-based comparative analysis of pathogenic *Escherichia coli* secretomes. *J. Microbiol. Methods* 116, 66–79. doi: 10.1016/j.mimet.2015.06.015

Burian, M., Velic, A., Matic, K., Gunther, S., Kraft, B., Gonser, L., et al. (2015). Quantitative proteomics of the human skin secretome reveal a reduction in immune defense mediators in ectodermal dysplasia patients. *J. Invest. Dermatol.* 135, 759–767. doi: 10.1038/jid.2014.462

Capone, K. A., Dowd, S. E., Stamatas, G. N., and Nikolovski, J. (2011). Diversity of the human skin microbiome early in life. *J. Invest. Dermatol.* 131, 2026–2032. doi: 10.1038/jid.2011.168

Cousin, C., Derouiche, A., Shi, L., Pagot, Y., Poncet, S., and Mijakovic, I. (2013). Protein-serine/threonine/tyrosine kinases in bacterial signaling and regulation. *FEMS Microbiol. Lett.* 346, 11–19. doi: 10.1111/1574-6968.12189

Del Rosso, J. Q., and Leyden, J. J. (2007). Status report on antibiotic resistance: implications for the dermatologist. *Dermatol. Clin.* 25, 127–132. doi: 10.1016/j.det.2007.01.001

Erickson, B. K., Jedrychowski, M. P., Mcalister, G. C., Everley, R. A., Kunz, R., and Gygi, S. P. (2015). Evaluating multiplexed quantitative phosphopeptide analysis on a hybrid quadrupole mass filter/linear ion trap/orbitrap mass spectrometer. *Anal. Chem.* 87, 1241–1249. doi: 10.1021/ac503934f

Espersen, F. (1998). Resistance to antibiotics used in dermatological practice. *Br. J. Dermatol.* 139(Suppl. 53), 4–8. doi: 10.1046/j.1365-2133.1998.1390s3004.x

Francois, P., Scherl, A., Hochstrasser, D., and Schrenzel, J. (2014). Proteomic approach to investigate pathogenicity and metabolism of methicillin-resistant *Staphylococcus aureus*. *Methods Mol. Biol.* 1085, 231–250. doi: 10.1007/978-1-62703-664-1_14

Gauthier, N. P., Soufi, B., Walkowicz, W. E., Pedicord, V. A., Mavrakis, K. J., Macek, B., et al. (2013). Cell-selective labeling using amino acid precursors for proteomic studies of multicellular environments. *Nat. Methods* 10, 768–773. doi: 10.1038/nmeth.2529

Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., and Gygi, S. P. (2003). Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. U.S.A.* 100, 6940–6945. doi: 10.1073/pnas.0832254100

Grice, E. A., and Segre, J. A. (2011). The skin microbiome. *Nat. Rev. Microbiol.* 9, 244–253. doi: 10.1038/nrmicro2537

Hanifin, J. M., and Rogge, J. L. (1977). Staphylococcal infections in patients with atopic dermatitis. *Arch. Dermatol.* 113, 1383–1386. doi: 10.1001/archderm.1977.01640100061009

Hempel, K., Herbst, F. A., Moche, M., Hecker, M., and Becher, D. (2011). 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.* 10, 1657–1666. doi: 10.1021/pr1009838

Hempel, K., Pane-Farre, J., Otto, A., Sievers, S., Hecker, M., and Becher, D. (2010). Quantitative cell surface proteome profiling for SigB-dependent protein expression in the human pathogen *Staphylococcus aureus* via biotinylation approach. *J. Proteome Res.* 9, 1579–1590. doi: 10.1021/pr901143a

Hu, Q., Noll, R. J., Li, H., Makarov, A., Hardman, M., and Graham Cooks, R. (2005). The Orbitrap: a new mass spectrometer. *J. Mass Spectrom.* 40, 430–443. doi: 10.1002/jms.856

Iwase, T., Uehara, Y., Shinji, H., Tajima, A., Seo, H., Takada, K., et al. (2010). *Staphylococcus epidermidis* Esp inhibits *Staphylococcus aureus* biofilm formation and nasal colonization. *Nature* 465, 346–349. doi: 10.1038/nature09074

Kim, D., Yu, B. J., Kim, J. A., Lee, Y. J., Choi, S. G., Kang, S., et al. (2013). The acetylproteome of gram-positive model bacterium *Bacillus subtilis*. *Proteomics* 13, 1726–1736. doi: 10.1002/pmic.201200001

Kohler, C., Wolff, S., Albrecht, D., Fuchs, S., Becher, D., Buttner, K., et al. (2005). Proteome analyses of *Staphylococcus aureus* in growing and non-growing cells: a physiological approach. *Int. J. Med. Microbiol.* 295, 547–565. doi: 10.1016/j.ijmm.2005.08.002

Kong, H. H., Oh, J., Deming, C., Conlan, S., Grice, E. A., Beatson, M. A., et al. (2012). Temporal shifts in the skin microbiome associated with disease flares and treatment in children with atopic dermatitis. *Genome Res.* 22, 850–859. doi: 10.1101/gr.131029.111

Kuczynski, J., Lauber, C. L., Walters, W. A., Parfrey, L. W., Clemente, J. C., Gevers, D., et al. (2012). Experimental and analytical tools for studying the human microbiome. *Nat. Rev. Genet.* 13, 47–58. doi: 10.1038/nrg3129

Kuo, I. H., Yoshida, T., De Benedetto, A., and Beck, L. A. (2013). The cutaneous innate immune response in patients with atopic dermatitis. *J. Allergy Clin. Immunol.* 131, 266–278. doi: 10.1016/j.jaci.2012.12.1563

Lee, C. R., Lee, J. H., Park, K. S., Jeong, B. C., and Lee, S. H. (2015). Quantitative proteomic view associated with resistance to clinically important antibiotics in gram-positive bacteria: a systematic review. *Front. Microbiol.* 6:828. doi: 10.3389/fmicb.2015.00828

Liao, G., Xie, L., Li, X., Cheng, Z., and Xie, J. (2014). Unexpected extensive lysine acetylation in the trump-card antibiotic producer *Streptomyces roseosporus*

revealed by proteome-wide profiling. *J. Proteomics* 106, 260–269. doi: 10.1016/j.jprot.2014.04.017

Lin, M. H., Hsu, T. L., Lin, S. Y., Pan, Y. J., Jan, J. T., Wang, J. T., et al. (2009). Phosphoproteomics of *Klebsiella pneumoniae* NTUH-K2044 reveals a tight link between tyrosine phosphorylation and virulence. *Mol. Cell. Proteomics* 8, 2613–2623. doi: 10.1074/mcp.M900276-MCP200

List, E. O., Berryman, D. E., Bower, B., Sackmann-Sala, L., Gosney, E., Ding, J., et al. (2008). The use of proteomics to study infectious diseases. *Infect. Disord. Drug Targets* 8, 31–45. doi: 10.2174/187152608784139640

Macek, B., Gnad, F., Soufi, B., Kumar, C., Olsen, J. V., Mijakovic, I., et al. (2008). Phosphoproteome analysis of *E. coli* reveals evolutionary conservation of bacterial Ser/Thr/Tyr phosphorylation. *Mol. Cell. Proteomics* 7, 299–307. doi: 10.1074/mcp.M700311-MCP200

Macek, B., Mijakovic, I., Olsen, J. V., Gnad, F., Kumar, C., Jensen, P. R., et al. (2007). The serine/threonine/tyrosine phosphoproteome of the model bacterium *Bacillus subtilis*. *Mol. Cell. Proteomics* 6, 697–707. doi: 10.1074/mcp.M600464-MCP200

Mankowska-Wierzbicka, D., Karczewski, J., Dobrowolska-Zachwieja, A., and Adamski, Z. (2015). The microbiome and dermatological diseases. *Postepy Hig. Med. Dosw.* 69, 978–985.

Mann, M., and Kelleher, N. L. (2008). Precision proteomics: the case for high resolution and high mass accuracy. *Proc. Natl. Acad. Sci. U.S.A.* 105, 18132–18138. doi: 10.1073/pnas.0800788105

Manteca, A., Ye, J., Sanchez, J., and Jensen, O. N. (2011). Phosphoproteome analysis of *Streptomyces* development reveals extensive protein phosphorylation accompanying bacterial differentiation. *J. Proteome Res.* 10, 5481–5492. doi: 10.1021/pr200762y

Misra, S. K., Milohanic, E., Ake, F., Mijakovic, I., Deutscher, J., Monnet, V., et al. (2011). Analysis of the serine/threonine/tyrosine phosphoproteome of the pathogenic bacterium *Listeria monocytogenes* reveals phosphorylated proteins related to virulence. *Proteomics* 11, 4155–4165. doi: 10.1002/pmic.201100259

Misra, S. K., Moussan Desiree Ake, F., Wu, Z., Milohanic, E., Cao, T. N., Cossart, P., et al. (2014). Quantitative proteome analyses identify PrfA-responsive proteins and phosphoproteins in *Listeria monocytogenes*. *J. Proteome Res.* 13, 6046–6057. doi: 10.1021/pr500929u

Nakatsuji, T., Chiang, H. I., Jiang, S. B., Nagarajan, H., Zengler, K., and Gallo, R. L. (2013). The microbiome extends to subepidermal compartments of normal skin. *Nat. Commun.* 4:1431. doi: 10.1038/ncomms2441

Neilson, K. A., Ali, N. A., Muralidharan, S., Mirzaei, M., Mariani, M., Assadourian, G., et al. (2011). Less label, more free: approaches in label-free quantitative mass spectrometry. *Proteomics* 11, 535–553. doi: 10.1002/pmic.201000553

Nomura, I., Gao, B., Boguniewicz, M., Darst, M. A., Travers, J. B., and Leung, D. Y. (2003). Distinct patterns of gene expression in the skin lesions of atopic dermatitis and psoriasis: a gene microarray analysis. *J. Allergy Clin. Immunol.* 112, 1195–1202. doi: 10.1016/j.jaci.2003.08.049

Olsen, J. V., Schwartz, J. C., Griep-Raming, J., Nielsen, M. L., Damoc, E., Denisov, E., et al. (2009). A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. *Mol. Cell. Proteomics* 8, 2759–2769. doi: 10.1074/mcp.M900375-MCP200

Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., et al. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* 1, 376–386. doi: 10.1074/mcp.M200025-MCP200

Ouidir, T., Cosette, P., Jouenne, T., and Hardouin, J. (2015). Proteomic profiling of lysine acetylation in *Pseudomonas aeruginosa* reveals the diversity of acetylated proteins. *Proteomics* 15, 2152–2157. doi: 10.1002/pmic.201500056

Pinkse, M. W., Uitto, P. M., Hilhorst, M. J., Ooms, B., and Heck, A. J. (2004). Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal. Chem.* 76, 3935–3943. doi: 10.1021/ac0498617

Rardin, M. J., Held, J. M., and Gibson, B. W. (2013). Targeted quantitation of acetylated lysine peptides by selected reaction monitoring mass spectrometry. *Methods Mol. Biol.* 1077, 121–131. doi: 10.1007/978-1-62703-637-5_8

Ribeiro, F. J., Przybylski, D., Yin, S., Sharpe, T., Gnerre, S., Abouelleil, A., et al. (2012). Finished bacterial genomes from shotgun sequence data. *Genome Res.* 22, 2270–2277. doi: 10.1101/gr.141515.112

Rodriguez-Suarez, E., and Whetton, A. D. (2013). The application of quantification techniques in proteomics for biomedical research. *Mass Spectrom. Rev.* 32, 1–26. doi: 10.1002/mas.21347

Rush, J., Moritz, A., Lee, K. A., Guo, A., Goss, V. L., Spek, E. J., et al. (2005). Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nat. Biotechnol.* 23, 94–101. doi: 10.1038/nbt1046

Scheltema, R. A., Hauschild, J. P., Lange, O., Hornburg, D., Denisov, E., Damoc, E., et al. (2014). The Q exactive HF, a benchtop mass spectrometer with a pre-filter, high-performance quadrupole and an ultra-high-field Orbitrap analyzer. *Mol. Cell. Proteomics* 13, 3698–3708. doi: 10.1074/mcp.M114.043489

Schittek, B. (2011). The antimicrobial skin barrier in patients with atopic dermatitis. *Curr. Probl. Dermatol.* 41, 54–67. doi: 10.1159/000323296

Schwanhausser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., et al. (2011). Global quantification of mammalian gene expression control. *Nature* 473, 337–342. doi: 10.1038/nature10098

Singh, S., Springer, M., Steen, J., Kirschner, M. W., and Steen, H. (2009). FLEXIQuant: a novel tool for the absolute quantification of proteins, and the simultaneous identification and quantification of potentially modified peptides. *J. Proteome Res.* 8, 2201–2210. doi: 10.1021/pr800654s

Soares, N. C., Spat, P., Krug, K., and Macek, B. (2013). Global dynamics of the *Escherichia coli* proteome and phosphoproteome during growth in minimal medium. *J. Proteome Res.* 12, 2611–2621. doi: 10.1021/pr3011843

Soufi, B., Gnad, F., Jensen, P. R., Petranovic, D., Mann, M., Mijakovic, I., et al. (2008). The Ser/Thr/Tyr phosphoproteome of *Lactococcus lactis* IL1403 reveals multiply phosphorylated proteins. *Proteomics* 8, 3486–3493. doi: 10.1002/pmic.200800069

Soufi, B., Krug, K., Harst, A., and Macek, B. (2015). Characterization of the *E. coli* proteome and its modifications during growth and ethanol stress. *Front. Microbiol.* 6:103. doi: 10.3389/fmicb.2015.00103

Soufi, B., Kumar, C., Gnad, F., Mann, M., Mijakovic, I., and Macek, B. (2010). Stable isotope labeling by amino acids in cell culture (SILAC) applied to quantitative proteomics of *Bacillus subtilis*. *J. Proteome Res.* 9, 3638–3646. doi: 10.1021/pr100150w

Soufi, B., and Macek, B. (2014). Stable isotope labeling by amino acids applied to bacterial cell culture. *Methods Mol. Biol.* 1188, 9–22. doi: 10.1007/978-1-4939-1142-4_2

Stekhoven, D. J., Omasits, U., Quebatte, M., Dehio, C., and Ahrens, C. H. (2014). Proteome-wide identification of predominant subcellular protein localizations in a bacterial model organism. *J. Proteomics* 99, 123–137. doi: 10.1016/j.jprot.2014.01.015

Stulberg, D. L., Penrod, M. A., and Blatny, R. A. (2002). Common bacterial skin infections. *Am. Fam. Physician* 66, 119–124.

Sun, F., Ding, Y., Ji, Q., Liang, Z., Deng, X., Wong, C. C., et al. (2012). Protein cysteine phosphorylation of SarA/MgrA family transcriptional regulators mediates bacterial virulence and antibiotic resistance. *Proc. Natl. Acad. Sci. U.S.A.* 109, 15461–15466. doi: 10.1073/pnas.1205952109

Sun, X., Ge, F., Xiao, C. L., Yin, X. F., Ge, R., Zhang, L. H., et al. (2010). Phosphoproteomic analysis reveals the multiple roles of phosphorylation in pathogenic bacterium *Streptococcus pneumoniae*. *J. Proteome Res.* 9, 275–282. doi: 10.1021/pr900612v

Swartz, M. N. (2004). Clinical practice. Cellulitis. *N. Engl. J. Med.* 350, 904–912. doi: 10.1056/NEJMcp031807

van Rensburg, J. J., Lin, H., Gao, X., Toh, E., Fortney, K. R., Ellinger, S., et al. (2015). The human skin microbiome associates with the outcome of and is influenced by bacterial infection. *MBio* 6:e1315-15. doi: 10.1128/mBio.01315-15

Villen, J., and Gygi, S. P. (2008). The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nat. Protoc.* 3, 1630–1638. doi: 10.1038/nprot.2008.150

Waridel, P., Ythier, M., Gfeller, A., Moreillon, P., and Quadroni, M. (2012). Evidence for a new post-translational modification in *Staphylococcus aureus*: hydroxymethylation of asparagine and glutamine. *J. Proteomics* 75, 1742–1751. doi: 10.1016/j.jprot.2011.12.014

Williams, H. C., Dellavalle, R. P., and Garner, S. (2012). Acne vulgaris. *Lancet* 379, 361–372. doi: 10.1016/S0140-6736(11)60321-8

Yu, B. J., Kim, J. A., Moon, J. H., Ryu, S. E., and Pan, J. G. (2008). The diversity of lysine-acetylated proteins in *Escherichia coli*. *J. Microbiol. Biotechnol.* 18, 1529–1536.

Zeeuwen, P. L., Kleerebezem, M., Timmerman, H. M., and Schalkwijk, J. (2013). Microbiome and skin diseases. *Curr. Opin. Allergy Clin. Immunol.* 13, 514–520. doi: 10.1097/ACI.0b013e328364ebeb

Zeiler, M., Straube, W. L., Lundberg, E., Uhlen, M., and Mann, M. (2012). A protein epitope signature tag (PrEST) library allows SILAC-based absolute quantification and multiplexed determination of protein copy numbers in cell lines. *Mol. Cell. Proteomics* 11, O111009613. doi: 10.1074/mcp.O111.009613

Zhang, K., Zheng, S., Yang, J. S., Chen, Y., and Cheng, Z. (2013). Comprehensive profiling of protein lysine acetylation in *Escherichia coli*. *J. Proteome Res.* 12, 844–851. doi: 10.1021/pr300912q

CrossMark

Ionic Liquids as Unforeseen Assets to Fight Life-Threatening Mycotic Diseases

Diego O. Hartmann, Marija Petkovic and Cristina Silva Pereira *

Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Oeiras, Portugal

Ionic liquids discovery has celebrated 100 years. They consist solely of ions, one of which is typically organic and asymmetrical. Remarkable physical and chemical properties stirred their use as alternative solvents in many chemical processes. The recent demonstration of their occurrence in nature might boost their interest in biological sciences. In the search of mechanistic understandings of ionic liquids' ecotoxicological impacts in fungi, we have analyzed the proteome, transcriptome, and metabolome responses to this chemical stress. Data illuminated new hypotheses that altered our research path – exploit ionic liquids as tools for the discovery of pathways and metabolites that may impact fungal development and pathogenicity. As we get closer to solve the primary effects of each ionic liquid family and their specific gene targets, the vision of developing antifungal ionic liquids and/or materials, by taking advantage of elegant progresses in this field, might become a reality. *Task-designed* formulations may improve the performance of conventional antifungal drugs, build functional coatings for reducing allergens production, or aid in the recovery of antifungal plant polymers. The frontier research in this cross-disciplinary field may provide us unforeseen means to address the global concern of mycotic diseases. Pathogenic and opportunistic fungi are responsible for numerous infections, killing annually nearly 1.5 million immunocompromised individuals worldwide, a similar rate to malaria or tuberculosis. This perspective will review our major findings and current hypotheses, contextualizing how they might bring us closer to efficient strategies to prevent and fight mycotic diseases.

Keywords: ionic liquids, filamentous fungi, fungal infections, proteomics, stress response, natural compounds discovery, antifungal drugs

Ionic liquids, which consist entirely of ionic species, are conventionally defined as salts that are liquid below 100°C. Their history started in 1914 when the physical properties of ethylammonium nitrate were first reported (Plechkova and Seddon, 2008). Nonetheless, only in the last decades the term *ionic liquid* emerged and a new scientific area arose. Their generic – yet not universal – properties include features such as negligible vapor pressure, conventional non-flammability and excellent solvation potential (Endres and Zein El Abedin, 2006), which do not occur concurrently either in molecular compounds or in crystalline salts. These properties boosted the interest of chemists and chemical engineers, and were the basis for the classification of ionic liquids as *green* solvents (Earle and Seddon, 2000). Their potential was further emphasized with the insight that numerous structural variations can be obtained via relatively simple synthesis, categorizing ionic

liquids as *designer solvents* (Ranke et al., 2007). Such subtle structural variations in the composing ions allow fine-tuning their physical and chemical properties to promote them as *task-designed solvents*.

These liquid salts have been widely investigated and several hundred are already chemically well characterized. By now, numerous applications of ionic liquids have been proposed, impacting diverse relevant areas, such as Catalysis, Separations, Materials, Sustainable Energies, Biorefineries, Renewable Fuels, and Chemicals, just to name a few. Some remarkable examples include ionic liquids in membranes for gas separation (Tomé et al., 2014), and for the extraction of disease biomarkers and antibodies (Taha et al., 2015). Unfortunately, only a few applications have successfully reached the industrial panorama (Plechkova and Seddon, 2008), e.g., BASIL^{TM} and cellulose reshaping (BASF, Germany), dye sensitizing solar cells (G24i Power, UK) and – certainly one of the most fascinating – the use of mercury grabbing ionic liquids to clean natural gas streams (Abai et al., 2015), commercialized as HycaPure Hg by Clariant (Switzerland).

As a scientific topic, ionic liquids are likely among the most intensely researched ones, especially in the Chemistry, Physics, and Materials disciplines. Over 10000 publications on ionic liquids can be found in the *Web of Knowledge* but little more than 10% intersect with the Life Sciences. So the question that rises from all the above is *why should any environmentalist, biologist or biochemist care about ionic liquids at first?* Perhaps the demonstration that ionic liquids might occur in nature sounds appealing. In a recent study, it was suggested that an ionic liquid is formed during the confrontation of two ant species, *Nylanderia fulva* and *Solenopsis invicta*, as a form of defense mechanism of the former against the venom of the latter ant species (Chen et al., 2014). In line with this idea, it was shown that certain metabolites abundant in plants become liquid when mixed together (Choi et al., 2011). These so-called *natural deep eutectic solvents*, with properties strongly resembling those of ionic liquids, would ensure cellular processes involving water-insoluble compounds. The likelihood of naturally occurring ionic liquids creates a new paradigm – they are not exclusively man-made chemicals – and fosters a new boost of interest in their research.

Ionic liquids and life sciences intertwined for the first time, however, to answer a very fundamental need: understanding ionic liquids' environmental impact. Our initial motivation, similar to tens of other research laboratories worldwide, was partially due to the fact that, despite being classified as *green* solvents, these organic salts comprise a disparate group of compounds that are not all intrinsically green. Many have been shown to be toxic and recalcitrant to biodegradation. Data collected so far has been compiled in a series of elegant and comprehensive reviews on their environmental impact (Petkovic et al., 2011) and biodegradability (Coleman and Gathergood, 2010). The large majority of early studies on ionic liquids' toxicity aimed at defining their inhibitory concentrations to very distinct organisms, essentially as to guide chemical research efforts toward more sustainable formulations. These studies have ascertained that different testing models exhibit fairly diverse susceptibilities to ionic liquids, however, often suggesting a similar mechanism of toxicity or cytotoxicity, i.e., plasma membrane permeabilization and oxidative stress (Yu et al., 2009; Petkovic et al., 2011). Unsurprisingly, the chemical nature of the ions rules their specific molecular and/or cellular mode of action. Most toxicity assays, if not all, were based on aqueous systems, in which the composing ions were fully solubilized in water. By accepting this principle, the biological effect of an ionic liquid should consider the individual contributions of its ions. Among the most common cations, the aromatic ones appear to be more toxic than the alicyclic or the quaternary ammonium (Stolte et al., 2007a,b). Nevertheless, the prevailing idea around ionic liquid's mechanism of toxicity is that, in either of the composing ions, the length of the alkyl chain is directly correlated with lipophilicity and permeabilization of biological membranes, leading to cell death (Zhao et al., 2007). This seems valid only for lipophilic cations, since our most recent data showed that permeabilization by long chain anions (i.e., alkanoates) is hindered by negative charges in the membrane outer surface (Hartmann et al., 2015). These were great news since we have been for long pursuing the use of cholinium alkanoates as novel biocompatible solvents for plant polyesters (Garcia et al., 2010). This idea has been nurturing our research in identifying *task-designed* ionic liquids for the hydrolysis of structural polymers in plant cell walls, i.e., cutin and suberin. Our goal was to preserve the native properties of the polyesters, particularly their function as barriers to microbial pathogens (Ranathunge et al., 2011). In some ways unpredictably, an ionic liquid – cholinium hexanoate – provided us the right means for that. It plays the dual role of solvent and catalyst, promoting the specific cleavage of particular ester bonds of suberin (Ferreira et al., 2014). This ensures the partial preservation of its tridimensional structure, hence the spontaneous formation of films with potentially broad antimicrobial properties (Garcia et al., 2014).

Notwithstanding significant progresses in the field of ionic liquids' toxicity, our curiosity did not allow us to stop there. We wanted to seek for better mechanistic understandings of how these allegedly man-made chemicals would impact living organisms at a cellular and molecular level. Our front-runner candidates for study organisms were undoubtedly filamentous fungi. Fungi, which are unique and remarkable eukaryotic organisms, act as key colonizers of the soil and ensure major ecosystem functions, including the mitigation of hazardous chemicals (Harms et al., 2011; Varela et al., 2015). Moreover, these organisms are well known as proficient producers of enzymes and metabolites of great biotechnological and pharmacological interest. Several studies revealed that fungal strains commonly found in soil can resist high concentrations of ionic liquids (Petkovic et al., 2009; Singer et al., 2011; Simpson, 2012). In particular our study also demonstrated that all the tested compounds could completely alter the fungal metabolic footprint, i.e., the diversity of diffusible small molecules (Petkovic et al., 2009). These promising and stimulating findings constituted a foundation for our subsequent research efforts.

Looking for a holistic view of the impact of these organic ions in fungal metabolism, we decided to perform a proteomic analysis of model filamentous fungi exposed to ionic liquids

(Martins et al., 2013). We specifically selected cholinium chloride and 1-ethyl-3-methylimidazolium chloride, which carry cations currently attracting most academic and industrial interest. These compounds have been previously observed to display very distinct antifungal activities and biodegradability potential (Coleman and Gathergood, 2010; Petkovic et al., 2011). *Aspergillus nidulans* and *Neurospora crassa* – prime model fungal systems for genetic, cellular, and biochemical research – are very dissimilar when accounting for their halo-tolerance (Gunde-Cimerman et al., 2009) and secondary metabolite producing capacity (Khaldi et al., 2010; Inglis et al., 2013). The differential proteome showed that several critical biological processes and pathways were affected by either cation, reflected in the accumulation of numerous stress-responsive proteins and osmolytes, and in the alteration of developmental programs in both fungi. Encouragingly, in this study we observed the accumulation of proteins likely involved in the biosynthesis of non-proteinogenic amino acids in *N. crassa* in the presence of either cation. These rare amino acids are found in secondary metabolites with potent biological activity, e.g., neoefrapeptins and acretocins (Degenkolb et al., 2007; Degenkolb and Bruckner, 2008). Another promising example is the case of *A. nidulans*, which genome has nearly 70 genes coding for multi-domain enzymes likely involved in secondary metabolite biosynthesis (Inglis et al., 2013). Through whole-genome profiling, our recent research efforts revealed that, upon exposure to certain organic ions, this fungus up-regulated a series of secondary metabolism backbone genes (Petkovic and Silva Pereira, 2012). This resulted in a differential metabolic profile that conceals small compounds with biological activities of high pharmacological value (unpublished data). These promising findings open new perspectives on ionic liquids' potential in the discovery of natural compounds.

The large amount of data that emerge from proteomics or transcriptomics analyses can provide fundamental information on very specific scientific questions. As an excellent example, the resistance of the bacterium *Enterobacter lignolyticus* to 1-ethyl-3-methylimidazolium chloride was in part unraveled using whole genome profiling (Khudyakov et al., 2012). This study arose from the question of how this solvent – able to efficiently dissolve cellulose from plant biomass – could impact biological and fermentation processes. The authors showed that bacteria partially circumvent the toxicity of the cation by increasing membrane transporters and the concentrations of osmolytes. These findings further inspired the design of biofuel cells where ionic liquids are employed for biopolymer dissolution (Ruegg et al., 2014).

As we gather more data from global analyses of the impact of these organic ions, we move deeper into exploring these compounds as tools to solve fundamental questions in fungal biology. Our group rapidly advanced from a rather simplistic view of morphological alterations perceived microscopically (Petkovic et al., 2012) to evaluate, at a gene expression level, membrane and cell wall damage induced by ionic liquids (Hartmann and Silva Pereira, 2013). Fungi can alter the composition of their membranes, regulating its fluidity to overcome adverse environments. The membrane fluidity, which

is inversely related to its resistance to permeabilizing compounds, is essentially controlled by the levels of ergosterol and by the balance between saturated and unsaturated fatty acids. The fungal cell wall, on its turn, is responsible for maintaining cell shape, counteracting the turgor pressure and protecting the plasma membrane. Upon damage to the cell wall, fungi respond by activating several genes involved in its biosynthesis, creating conditions that allow them to re-establish its integrity, through the so-called cell wall integrity pathway. This salvage mechanism, better understood in the yeast *Saccharomyces cerevisiae*, remains poorly characterized in filamentous fungi (Fujioka et al., 2007; Valiante et al., 2015). We have demonstrated that some ionic liquids can cause membrane and cell wall damage in *A. nidulans*, most likely activating an alternative cell wall integrity pathway, yet to be characterized (Hartmann and Silva Pereira, 2013). More intriguing is the fact that these organic ions can also activate sphingolipid biosynthesis, leading to the differential accumulation of intermediates, including unknown species (Hartmann and Silva Pereira, 2015). These molecules may participate in the stress response of *A. nidulans*, including the activation of the cell wall integrity pathway. These are noteworthy results, not only for the prospect of unraveling a cross-talk mechanism between the cell wall integrity pathway and sphingolipids biosynthesis, but also because both pathways have for long been considered to be excellent candidate targets for the development of new antifungal agents.

Conventional antifungals, which target, directly or indirectly, the fungal plasma membrane or cell wall, are limited to just a few classes (*viz.* azoles, echinocandins and polyenes) (Odds et al., 2003). New generations of the classical antifungal drugs, as well as non-conventional agents and targets are already available, such as flucytosine and sordarins, which act by inhibiting DNA and protein synthesis, respectively, (Odds et al., 2003). However, clinical development and implementation of new drugs is notoriously long. Hence, the current challenge is to better understand the biology of filamentous fungi, aiming at the discovery of novel targets and the development of new effective drugs and antifungal strategies (Ostrosky-Zeichner et al., 2010; Denning and Bromley, 2015). We hope to make further evident how our cross-disciplinary research will provide means to address these global concerns. As long-term perspective, we seek to deepen our knowledge on fungal biology by exploring organic ions as the right stimuli for deciphering key cellular and molecular processes. We now rely on proteomic tools and, more specifically, phosphoproteomics, to attain deeper insights on the potential elements of the cell wall integrity pathway, as a foundation to solve the puzzling roles of sphingolipids in filamentous fungi.

Although debatable, the intriguing application of ionic liquids in pharmaceuticals development – often mentioned as *third evolution of ionic liquids* – has produced so far notable improvement of drugs solubility, delivery and biological activity through their conversion to a salt form (Hough et al., 2007). This seems a rather interesting prospect, especially when applied to the salt form of the antifungal drug amphotericin B to overcome its low solubility (Petkovic et al., 2015). Nevertheless, the fundamental question we are trying to address is how

ionicity impacts the drug primary mode of action. Another path being investigated by us is the use of ionic liquids to reduce the negative impact of pathogens as *Aspergillus fumigatus* by targeting allergen production. There are nearly twenty fully described allergen peptides in this fungus and as many predicted ones (Kurup, 2005; Fedorova et al., 2008). Transcriptomic data suggested that exposure of fungi to certain ionic liquids can strongly reduce the expression of genes coding for putative allergenic peptides (unpublished data). The current challenge is to identify, supported by immunoproteomics, organic ions that strongly interfere with the biosynthesis of allergenic peptides in *A. fumigatus*. This constitutes another elegant example of cell biochemistry manipulation using chemical stimuli and may inspire the use of ionic liquids for developing novel antifungal materials/coatings.

Life-threatening fungal infections present an uprising burden that affects millions of individuals, with more than 2 million invasive fungal infections reported every year worldwide (Brown et al., 2012). Fungi constitute a high risk to immunocompromised individuals of all ages, such as HIV/AIDS, cancer, transplant, and diabetes patients, which represent a significant percentage of the world population. The healthcare costs are enormous, estimated to billions of dollars per year on antifungal drugs only. Mortality rates often exceed 50% even with the current treatment options. This reality is aggravated when considering that the available therapies are sometimes inadequate, as many resistant strains (Anderson, 2005) and emerging fungal pathogens (Fisher et al., 2012) are now being discovered at a regular basis. The identification of new potential risk groups, from asthma sufferers to gastric ulcer patients, further emphasizes the need for efficient

antifungal drugs (van Woerden et al., 2013). Our vision is to produce valuable far-reaching insights to advance on the identification and development of novel antifungal strategies to, ultimately, fight fungal pathogenicity.

AUTHOR CONTRIBUTIONS

CP conceived and written the first draft; DH and MP contributed in the acquisition, analysis, and interpretation of the data included. All authors revised the manuscript critically for important intellectual content and gave final approval of the version to be submitted.

FUNDING

ACKNOWLEDGMENT

REFERENCES

Abai, M., Atkins, M. P., Hassan, A., Holbrey, J. D., Kuah, Y., Nockemann, P., et al. (2015). An ionic liquid process for mercury removal from natural gas. *Dalton Trans.* 44, 8617–8624. doi: 10.1039/c4dt03273j

Anderson, J. B. (2005). Evolution of antifungal-drug resistance: mechanisms and pathogen fitness. *Nat. Rev. Microbiol.* 3, 547–556. doi: 10.1038/nrmicrol179

Brown, G. D., Denning, D. W., Gow, N. A. R., Levitz, S. M., Netea, M. G., and White, T. C. (2012). Hidden killers: human fungal infections. *Sci. Transl. Med.* 4:165rv113. doi: 10.1126/scitranslmed.3004404

Chen, L., Mullen, G. E., Le Roch, M., Cassity, C. G., Gouault, N., Fadamiro, H. Y., et al. (2014). On the formation of a protic ionic liquid in nature. *Angew. Chem. Int. Ed. Engl.* 53, 11762–11765. doi: 10.1002/anie.201404402

Choi, Y. H., van Spronsen, J., Dai, Y. T., Verberne, M., Hollmann, F., Arends, I. W. C. E., et al. (2011). Are natural deep eutectic solvents the missing link in understanding cellular metabolism and physiology? *Plant Physiol.* 156, 1701–1705. doi: 10.1104/pp.111.178426

Coleman, D., and Gathergood, N. (2010). Biodegradation studies of ionic liquids. *Chem. Soc. Rev.* 39, 600–637. doi: 10.1039/b817717c

Degenkolb, T., and Bruckner, H. (2008). Peptaibiomics: towards a myriad of bioactive peptides containing C-alpha-dialkylaimino acids? *Chem. Biodivers.* 5, 1817–1843. doi: 10.1002/cbdv.200890171

Degenkolb, T., Kirschbaum, J., and Bruckner, H. (2007). New sequences, constituents, and producers of peptaibiotics: an updated review. *Chem. Biodivers.* 4, 1052–1067. doi: 10.1002/cbdv.200790096

Denning, D. W., and Bromley, M. J. (2015). How to bolster the antifungal pipeline. *Science* 347, 1414–1416. doi: 10.1126/science.aaa6097

Earle, M. J., and Seddon, K. R. (2000). Ionic liquids. Green solvents for the future. *Pure Appl. Chem.* 72, 1391–1398. doi: 10.1351/pac200072071391

Endres, F., and Zein El Abedin, S. (2006). Air and water stable ionic liquids in physical chemistry. *Phys. Chem. Chem. Phys.* 8, 2101–2116. doi: 10.1039/b600519p

Fedorova, N. D., Khaldi, N., Joardar, V. S., Maiti, R., Amedeo, P., Anderson, M. J., et al. (2008). Genomic islands in the pathogenic filamentous fungus *Aspergillus fumigatus*. *PLoS Genet.* 4:e1000046. doi: 10.1371/journal.pgen.1000046

Ferreira, R., Garcia, H., Sousa, A. F., Guerreiro, M., Duarte, F. J. S., Freire, C. S. R., et al. (2014). Unveiling the dual role of the cholinium hexanoate ionic liquid as solvent and catalyst in suberin depolymerisation. *RSC Adv.* 4, 2993–3002. doi: 10.1039/c3ra45910a

Fisher, M. C., Henk, D. A., Briggs, C. J., Brownstein, J. S., Madoff, L. C., McCraw, S. L., et al. (2012). Emerging fungal threats to animal, plant and ecosystem health. *Nature* 484, 186–194. doi: 10.1038/nature10947

Fujioka, T., Mizutani, O., Furukawa, K., Sato, N., Yoshimi, A., Yamagata, Y., et al. (2007). MpkA-dependent and -independent cell wall integrity signaling in *Aspergillus nidulans*. *Eukaryot. Cell* 6, 1497–1510. doi: 10.1128/EC.00281-06

Garcia, H., Ferreira, R., Martins, C., Sousa, A. F., Freire, C. S. R., Silvestre, A. J. D., et al. (2014). Ex situ reconstitution of the plant biopolyester suberin as a film. *Biomacromolecules* 15, 1806–1813. doi: 10.1021/bm500201s

Garcia, H., Ferreira, R., Petkovic, M., Ferguson, J. L., Leitão, M. C., Gunaratne, H. Q. N., et al. (2010). Dissolution of cork biopolymers by biocompatible ionic liquids. *Green Chem.* 12, 367–369. doi: 10.1039/B922553F

Gunde-Cimerman, N., Ramos, J., and Plemenitaš, A. (2009). Halotolerant and halophilic fungi. *Mycol. Res.* 113, 1231–1241. doi: 10.1016/j.mycres.2009.09.002

Harms, H., Schlosser, D., and Wick, L. Y. (2011). Untapped potential: exploiting fungi in bioremediation of hazardous chemicals. *Nat. Rev. Microbiol.* 9, 177–192. doi: 10.1038/nrmicro2519

Hartmann, D. O., Shimizu, K., Siopa, F., Leitão, M. C., Afonso, C. A. M., Canongia Lopes, J. N., et al. (2015). Plasma membrane permeabilisation by ionic liquids: a matter of charge. *Green Chem.* 17, 4587–4598. doi: 10.1039/c5gc01472g

Hartmann, D. O., and Silva Pereira, C. (2013). A molecular analysis of the toxicity of alkyltributylphosphonium chlorides in *Aspergillus nidulans*. *New J. Chem.* 37, 1569–1577. doi: 10.1039/C3NJ00167A

Hartmann, D. O., and Silva Pereira, C. (2015). "Toxicity of ionic liquids: past, present and future," in *Ionic Liquids in Lipid Processing and Analysis: Opportunities and Challenges*, eds X. Xu, Z. Guo, and L.-Z. Cheong (Urbana: AOCS Press), 403–421.

Hough, W. L., Smiglak, M., Rodriguez, H., Swatloski, R. P., Spear, S. K., Daly, D. T., et al. (2007). The third evolution of ionic liquids: active pharmaceutical ingredients. *New J. Chem.* 31, 1429–1436. doi: 10.1039/b706677p

Inglis, D. O., Binkley, J., Skrzypek, M. S., Arnaud, M. B., Cerqueira, G. C., Shah, P., et al. (2013). Comprehensive annotation of secondary metabolite biosynthetic genes and gene clusters of *Aspergillus nidulans*, *A. fumigatus*, *A. niger* and *A. oryzae*. *BMC Microbiol.* 13:91. doi: 10.1186/1471-2180-13-91

Khaldi, N., Seifuddin, F. T., Turner, G., Haft, D., Nierman, W. C., Wolfe, K. H., et al. (2010). SMURF: genomic mapping of fungal secondary metabolite clusters. *Fungal Genet. Biol.* 47, 736–741. doi: 10.1016/j.fgb.2010.06.003

Khudyakov, J. I., D'haeseleer, P., Borglin, S. E., DeAngelis, K. M., Woo, H., Lindquist, E. A., et al. (2012). Global transcriptome response to ionic liquid by a tropical rain forest soil bacterium, *Enterobacter* lignolyticus. *Proc. Natl. Acad. Sci. U.S.A.* 109, E2173–E2182. doi: 10.1073/pnas.1112750109

Kurup, V. P. (2005). Aspergillus antigens: which are important? *Med. Mycol.* 43, 189–196. doi: 10.1080/13693780500046763

Martins, I., Hartmann, D. O., Alves, P. C., Planchon, S., Renaut, J., Leitao, C., et al. (2013). Proteomic alterations induced by ionic liquids in *Aspergillus nidulans* and *Neurospora crassa*. *J. Proteomics* 94C, 262–278. doi: 10.1016/j.jprot.2013.09.015

Odds, F. C., Brown, A. J., and Gow, N. A. (2003). Antifungal agents: mechanisms of action. *Trends Microbiol.* 11, 272–279. doi: 10.1016/S0966-842X(03)00117-3

Ostrosky-Zeichner, L., Casadevall, A., Galgiani, J. N., Odds, F. C., and Rex, J. H. (2010). An insight into the antifungal pipeline: selected new molecules and beyond. *Nat. Rev. Drug Discov.* 9, 719–727. doi: 10.1038/nrd3074

Petkovic, M., Ferguson, J. L., Bohn, A., Trindade, J., Martins, I., Carvalho, M. B., et al. (2009). Exploring fungal activity in the presence of ionic liquids. *Green Chem.* 11, 889–894. doi: 10.1039/B823225C

Petkovic, M., Hartmann, D. O., Adamová, G., Seddon, K. R., Rebelo, L. P. N., and Silva Pereira, C. (2012). Unravelling the mechanism of toxicity of alkyltributylphosphonium chlorides in *Aspergillus nidulans* conidia. *New J. Chem.* 36, 56–63. doi: 10.1039/c1nj20470j

Petkovic, M., Hartmann, D. O., and Silva Pereira, C. (2015). Poster presentations. Modulating antifungal activity of amphotericin B through its conversion to a salt form. *Mycoses* 58:203. doi: 10.1111/myc.12380

Petkovic, M., Seddon, K. R., Rebelo, L. P. N., and Silva Pereira, C. (2011). Ionic liquids: a pathway to environmental acceptability. *Chem. Soc. Rev.* 40, 1383–1403. doi: 10.1039/c004968a

Petkovic, M., and Silva Pereira, C. (2012). "Pioneering biological processes in the presence of ionic liquids: the potential of filamentous fungi," in *Ionic Liquids UnCOILed: Critical Expert Overviews*, eds N. V. Plechkova and K. R. Seddon (Hoboken, NJ: John Wiley & Sons, Inc.), 283–303.

Plechkova, N. V., and Seddon, K. R. (2008). Applications of ionic liquids in the chemical industry. *Chem. Soc. Rev.* 37, 123–150. doi: 10.1039/B006677j

Ranathunge, K., Schreiber, L., and Franke, R. (2011). Suberin research in the genomics era—New interest for an old polymer. *Plant Sci.* 180, 399–413. doi: 10.1016/j.plantsci.2010.11.003

Ranke, J., Stolte, S., Stormann, R., Arning, J., and Jastorff, B. (2007). Design of sustainable chemical products - the example of ionic liquids. *Chem. Rev.* 107, 2183–2206. doi: 10.1021/cr050942s

Ruegg, T. L., Kim, E.-M., Simmons, B. A., Keasling, J. D., Singer, S. W., Soon Lee, T., et al. (2014). An auto-inducible mechanism for ionic liquid resistance in microbial biofuel production. *Nat. Commun.* 5:1. doi: 10.1038/ncomms4490

Simpson, T. J. (2012). Genetic and biosynthetic studies of the fungal prenylated xanthone shamixanthone and related metabolites in *Aspergillus* spp. revisited. *ChemBioChem* 13, 1680–1688. doi: 10.1002/cbic.201200014

Singer, S. W., Reddy, A. P., Gladden, J. M., Guo, H., Hazen, T. C., Simmons, B. A., et al. (2011). Enrichment, isolation and characterization of fungi tolerant to 1-ethyl-3-methylimidazolium acetate. *J. Appl. Microbiol.* 110, 1023–1031. doi: 10.1111/j.1365-2672.2011.04959.x

Stolte, S., Arning, J., Bottin-Weber, U., Müller, A., Pitner, W. R., Welz-Biermann, U., et al. (2007a). Effects of different head groups and functionalised side chains on the cytotoxicity of ionic liquids. *Green Chem.* 9, 760–767. doi: 10.1039/b615326g

Stolte, S., Matzke, M., Arning, J., Böschen, A., Pitner, W. R., Welz-Biermann, U., et al. (2007b). Effects of different head groups and functionalised side chains on the aquatic toxicity of ionic liquids. *Green Chem.* 9, 1170–1179. doi: 10.1039/B711119c

Taha, M., Almeida, M. R., Silva, F. A., Domingues, P., Ventura, S. P., Coutinho, J. A., et al. (2015). Novel biocompatible and self-buffering ionic liquids for biopharmaceutical applications. *Chem. Euro. J.* 21, 4781–4788. doi: 10.1002/chem.201405693

Tomé, L. C., Patinha, D. J., Ferreira, R., Garcia, H., Silva Pereira, C., Freire, C. S., et al. (2014). Cholinium-based supported ionic liquid membranes: a sustainable route for carbon dioxide separation. *ChemSusChem* 7, 110–113. doi: 10.1002/cssc.201300613

Valiante, V., Macheleidt, J., Föge, M., and Brakhage, A. A. (2015). The *Aspergillus fumigatus* cell wall integrity signaling pathway: drug target, compensatory pathways, and virulence. *Front. Microbiol.* 6:325. doi: 10.3389/fmicb.2015.00325

van Woerden, H. C., Gregory, C., Brown, R., Marchesi, J. R., Hoogendoorn, B., and Matthews, I. P. (2013). Differences in fungi present in induced sputum samples from asthma patients and non-atopic controls: a community based case control study. *BMC Infect. Dis.* 13:69. doi: 10.1186/1471-2334-13-69

Varela, A., Martins, C., Núñez, O., Martins, I., Houbraken, J. A. M. P., Martins, T. M., et al. (2015). Understanding fungal functional biodiversity during the mitigation of environmentally dispersed pentachlorophenol in cork oak forest soils. *Environ. Microbiol.* 17, 2922–2934. doi: 10.1111/1462-2920.12837

Yu, M., Wang, S.-H., Luo, Y.-R., Han, Y.-W., Li, X.-Y., Zhang, B.-J., et al. (2009). Effects of the 1-alkyl-3-methylimidazolium bromide ionic liquids on the antioxidant defense system of Daphnia magna. *Ecotoxicol. Environ. Saf.* 72, 1798–1804. doi: 10.1016/j.ecoenv.2009.05.002

Zhao, D., Liao, Y., and Zhang, Z. (2007). Toxicity of ionic liquids. *CLEAN – Soil Air Water* 35, 42–48. doi: 10.1002/clen.200600015

Elucidating Host–Pathogen Interactions Based on Post-Translational Modifications Using Proteomics Approaches

Vaishnavi Ravikumar [1], Carsten Jers [2] and Ivan Mijakovic [1,3]*

[1] Systems and Synthetic Biology Division, Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden, [2] Department of Systems Biology, Technical University of Denmark, Lyngby, Denmark, [3] Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Hørsholm, Denmark

Microbes with the capability to survive in the host tissue and efficiently subvert its innate immune responses can cause various health hazards. There is an inherent need to understand microbial infection patterns and mechanisms in order to develop efficient therapeutics. Microbial pathogens display host specificity through a complex network of molecular interactions that aid their survival and propagation. Co-infection states further lead to complications by increasing the microbial burden and risk factors. Quantitative proteomics based approaches and post-translational modification analysis can be efficiently applied to gain an insight into the molecular mechanisms involved. The measurement of the proteome and post-translationally modified proteome dynamics using mass spectrometry, results in a wide array of information, such as significant changes in protein expression, protein abundance, the modification status, the site occupancy level, interactors, functional significance of key players, potential drug targets, etc. This mini review discusses the potential of proteomics to investigate the involvement of post-translational modifications in bacterial pathogenesis and host–pathogen interactions.

Keywords: co-cultures, pathogenesis, bacteria, proteomics, post-translational modifications

INTRODUCTION

Phylogenetically diverse microorganisms such as bacteria, viruses or protozoans commonly pose as "pathogens." They have the capacity to survive in the host tissue, adapt to and suppress the host innate immune responses, thus constantly challenging human health and welfare. The normal microbial flora that inhabits specific regions in the host organism is essential for the proper functioning of the host. In contrast, pathogens elicit specific unnatural responses in the host system. Over the years, the field of infection biology has thus generated much interest, especially with the current emergence of various multi-drug resistant strains. The infection cycle of a pathogen comprises mainly of three stages, namely introduction of the pathogen into the host cell or tissue, establishment of the pathogen within the host and dissemination within the host. Despite continuous improvements in the techniques developed to identify and elucidate infection patterns, there is a pressing demand for more efficient advancements to interpret the molecular cues during infection. Mass spectrometry based proteomics, through the years, has been instrumental in deciphering the molecular mechanisms underlying host–pathogen interactions (Yang et al., 2015). Development of

high-resolution instruments and improvements in experimental techniques has helped expand the scope of infection biology.

The delicate balance that exists between the host immune responses and the pathogen can be tipped toward the latter by the co-existence of multiple microbial species capable of colonizing the same host niche. Bacteria have been identified in a wide range of symbiotic associations such as pathogenic, mutualistic or as commensals. The state of "natural selection" enforces interspecies specific interactions that differ significantly from the behavior of the individual species. Bacterial co-cultures or mixed populations find numerous benefits (1) they are commonly employed in the biotechnology industry such as in waste water treatment (Tijhuis et al., 1994), biogas plants (Klocke et al., 2007), production of vitamin C (Ma et al., 2014); (2) they form an important part of the human gut microbiota (Putignani et al., 2014). Adversely, symbiotic interactions can also be unfavorable, resulting in host invasion and damage (Chowdhury et al., 2010; Kluge et al., 2012). Thus systematic analysis of bacteria in co-culture conditions is essential to obtain an understanding of microbial behavior under conditions of infection. Routine strategies employed by pathogens to defeat host immune responses generally include adherence to host cell surfaces, followed by invasion and subsequent production and secretion of enterotoxins along with exhibiting host molecular mimicry mechanisms (Zhang et al., 2005). Protein secretion systems are thus pivotal in infection (Tseng et al., 2009). Targeting any one of these systems could help curb transmission of the disease or elimination of the pathogen. The following sections in this review describe the implications of various post-translational modifications (PTMs) on bacterial virulence and the proteomics technologies available to investigate such events along with the associated technical challenges.

POST-TRANSLATIONAL MODIFICATIONS

Protein PTMs are known to play diverse roles in the cellular milieu. Bacterial virulence is orchestrated by a multitude of PTMs. Some of the most common PTMs critical for the infection process are reviewed here.

Phosphorylation

Phosphorylation is a well-characterized and ubiquitous PTM that is crucial for most functions occurring in a bacterial cell. Apart from regulating metabolic pathways, phosphorylation has also been seen to be involved in various virulence mechanisms. The existence of Hanks-type Ser/Thr kinases in bacteria has generated a huge interest in the field of infection biology. Pathogenic strains such as *Streptococcus* spp., *Pseudomonas aeruginosa*, *Mycobacteria*, *Yersinia* spp., to name a few, employ Ser/Thr kinase-mediated host–pathogen interactions to mediate diverse cellular networks required for adhesion to and invasion of the host. While the exact mechanism of infection is not yet well understood, the mode of infection has been speculated to follow three basic modes: (1) phosphorylation of host proteins, (2) disruption of host defense mechanisms due to kinase activity, and lastly (3) essential role of Ser/Thr kinases by unrealized processes (Canova and Molle, 2014). The global phosphoproteome of a number of pathogenic bacteria such as *Corynebacterium glutamicum* (Bendt

et al., 2003), *Campylobacter jejuni* (Voisin et al., 2007), *Klebsiella pneumoniae* (Lin et al., 2009), *Mycobacterium tuberculosis* (Prisic et al., 2010), *Streptomyces coelicolor* (Parker et al., 2010), to name a few, have been analyzed. Changes that occur in the host phosphoproteome upon bacterial infection have also been investigated (Schmutz et al., 2013; Scholz et al., 2015). Dynamics of the bacterial phosphoproteome at the time of infection and post infection would be an interesting avenue to embark upon.

Secretion systems play a vital role during pathogenesis in a large number of bacteria. *Yersinia enterocolitica* and other species encode for a protein kinase A, YopO, which is secreted into the host via the type III secretion system. This kinase helps resist phagocytosis by macrophages via disruption of host cytoskeletal elements (Juris et al., 2000; Grosdent et al., 2002). This kinase was also reported to phosphorylate actin and otubain, resulting in inhibition of phagocytosis (Juris et al., 2006). Mutation of the kinase domain has been shown to reduce lethality during infection (Galyov et al., 1993; Wiley et al., 2006). Similarly, Stk1 from *Staphylococcus aureus* has been shown to phosphorylate numerous host substrates involved in cell cycle signaling or apoptotic pathways (Miller et al., 2010). SteC of *Salmonella enterica* serovar Typhimurium, like YopO, induces reorganization of actin filaments in the host on infection (Odendall et al., 2012). Host immune cell responses depend on the proper functioning of the NF-κB pathway. *Legionella pneumophila* LegK acts as an inflammatory agent and interferes with the NF-κB pathway (Ge et al., 2009). Likewise, protein kinases NleH1 and NleH2 from enteropathogenic *Escherichia coli* work by inhibiting the transcription factor, NF-κB (Royan et al., 2010). Phosphorylation of the central core of type II fatty acid synthase (FASH) in *Mycobacterium tuberculosis*, catalyzed by the kinase KasB, governs the physiopathology of tuberculosis (Vilcheze et al., 2014). Apart from phosphorylation by Ser/Thr Hanks-type kinases, bacterial two-component systems involving phosphorylation on histidine-aspartate residues (TCS) form a major adaptive mechanism in pathogenic strains. For example, CovRS (the control of virulence regulator/sensor kinase) in the human pathogen group A *Streptococcus* is fundamental for virulence (Horstmann et al., 2015). Interestingly, cysteine protein phosphorylation events are also reported to mediate bacterial virulence. The SarA/MarA staphylococcal accessary regulator A, part of the family of global transcriptional regulators (MgrA), is phosphorylated/dephosphorylated by the *Staphylococcus aureus* kinase/phosphatase pair Stk1-Stp1 and speculated to play a crucial role in shifting the intracellular redox balance, contributing to virulence (Sun et al., 2012).

Cognate to kinase activity is the activity of phosphatases, making phosphorylation a reversible and tightly regulated PTM. In many organisms, protein phosphatases act as essential virulence determinants, thus playing a central role in infection and dissemination. YopH tyrosine phosphatase from *Yersinia* is involved in the dephosphorylation of the focal adhesion complexes and essential for antiphagocytosis (Persson et al., 1999). SptP tyrosine phosphatase of *Salmonella typhimurium* was observed to be required for virulence in murine models (Kaniga et al., 1996). The phosphothreonine lyase protein, OspF, from *Shigella flexneri* irreversibly phosphorylates members of

the MAPK and ERK pathway, subsequently affecting the innate immune system (Reiterer et al., 2011). Dephosphorylation of tyrosine kinases such as the BY-kinase Wzc-ca from *E. coli* K12 by Wzb causes increased capsular polysaccharide formation which can further act as a poor immunogen (Whitmore and Lamont, 2012; Hansen et al., 2013).

Acylation

Acetylation can be used as a mechanism to modulate phosphorylation-based signaling. *Yersinia* species use a serine/threonine acetyltransferase, YopJ, to interfere with host MAPK kinase signaling by acetylating serine and threonine residues in the activation loop thereby preventing phosphorylation-dependent activation (Mukherjee et al., 2006). YopJ homologs are widely distributed in both mammal and plant pathogens suggesting that inhibition of host kinases by serine/threonine acetylation could be a common strategy for bacterial pathogens (Lewis et al., 2011). Lysine acetylation and succinylation have in recent years been shown to be abundant modifications in bacteria, and there are some indications that they could be important for bacterial pathogenesis. In *E. coli*, the transcription factor RcsB that controls colanic acid capsule synthesis, is acetylated on lysine thereby reducing its DNA binding activity (Thao et al., 2010). A global study indicated that lysine acetylation is involved in regulation of cell wall fatty acids synthesis in *M. tuberculosis*, which in turn is implicated in pathogenicity (Liu et al., 2014). In addition, a lysine deacetylase (MRA_1161) mutant exhibits a defect in biofilm formation (Liu et al., 2014). KasA, a protein involved in biofilm formation, is modified by another type of acylation namely lysine succinylation. Further, a number of proteins involved in antibiotic resistance are succinylated (Xie et al., 2015). With only few functional studies, it is still unclear to what extent lysine acetylation and succinylation contribute to bacterial virulence.

Ubiquitination

Ubiquitination is an important PTM in Eukarya and regulates several processes including key cell defense systems. Ubiquitin is a small polypeptide (78 amino acids) that can be covalently linked to primarily lysine. Ubiquitination requires the activities of an E1 activating enzyme, an E2 conjugating enzyme and an E3 ligase (Ashida et al., 2014). Ubiquitin contains seven lysines and can itself be ubiquitinated leading to formation of polyubiquitin chains with various linkages that in turn dictates biological function (Ashida et al., 2014). Bacterial pathogens have developed several ways of targeting the host ubiquitin system including the use of Eukaryotic-like and novel E3 ligases as well as de-ubiquitinating enzymes. The *S. flexneri* E3 ligase IpaH9.8 reduces the NF-κB-mediated inflammatory response by polyubiquitination of NEMO, a protein involved in NF-κB activation, thereby targeting it for degradation (Ashida et al., 2010). Additionally, IpaH9.8 appears to modulate gene expression via ubiquitination of the splicing factor U2AF[35] (Okuda et al., 2005; Seyedarabi et al., 2010). In *S. enterica* infection, ubiquitinated protein aggregates are formed near the *Salmonella*-containing vacuole which targets it for autophagic degradation. This is countered by SseL, a deubiquitinase, that

deubiquitinates the protein aggregates (Mesquita et al., 2012). The pathogens can also use the host ubiquitination system to modify their own proteins. To control the timing of effector protein-mediated functions, *S. enterica* SopE, that targets RHO-GTPases leading to membrane ruffling, is ubiquitinated and degraded earlier than the protein SptP that prevents membrane ruffling after invasion (Kubori and Galan, 2003). Another *S. enterica* effector protein, the phosphatase SopB, can be mono-ubiquitinated at six positions and this in turn modulates its cellular location and enzyme activity (Knodler et al., 2009).

AMPylation

AMPylation is the covalent attachment of AMP to a threonine or tyrosine residue. *Vibrio parahaemolyticus* effector VopS uses ATP to covalently modify Rho-GTPases with AMP on threonine. This is turn affects its interaction with downstream signaling proteins leading to inhibition of actin assembly (Yarbrough et al., 2009). To exploit host cell vesicle transport, *L. pneumophila* SidM activates the small GTPase Rab1 by AMPylation on a tyrosine, and when Rab1 is no longer needed, it is de-AMPylated by SidD, and subsequently targeted for degradation by polyubiquitination (Neunuebel et al., 2011). Host GTPases are also targeted by another novel, reversible PTM, namely phosphocholination. *L. pneumophila* effectors AnkX and Lpg0696 phosphocholinate and de-phosphocholinate Rab1 and Rab35, respectively (Mukherjee et al., 2011; Tan et al., 2011).

Alkylation

Protein alkylation is the addition of alkyl groups on specific amino acids, notably methyl on arginine and lysine (methylation) and the lipids farnesyl or geranylgeranyl isoprenyl on cysteine (prenylation). Histone proteins are regulated by lysine methylation, and this is also a target for bacterial pathogens. The *Bacillus anthracis* protein BaSET trimethylates histone H1 on lysine and thereby reduces activation of NF-κB response elements (Mujtaba et al., 2013). Prenylation confers hydrophobicity to its substrate proteins and target them to membranes. This is employed by *L. pneumophila* to assure correct cellular localization of its effector AnkB. Host-mediated farnesylation of AnkB, targets it to the cytosolic face of the *Legionella*-containing vacuole and this in turn is essential for its function (Price et al., 2010).

Eliminylation

Shigella flexneri effector protein OspF, a phosphothreonine lyase, interferes with host signaling by irreversibly dephosphorylating MAP kinases. In this PTM termed eliminylation, not only the phosphate but also the hydroxyl group of threonine is removed, thereby preventing any future phosphorylation. It irreversibly inactivates the kinase (Li et al., 2007).

Glycosylation

Protein glycosylation is now a well-established modification in bacteria. O-linked glycosylation of flagellar proteins has been linked to virulence in pathogenic bacteria. In the opportunistic pathogen *Burkholderia cenocepacia*, glycosylation of the flagellar protein FliC reduces binding to the host TLR5 receptor, thereby weakening the immune response. Additionally, a

non-glycosylated FliC mutant exhibits a defect in biofilm formation (Hanuszkiewicz et al., 2014). *Helicobacter pylori* flagellar proteins are also glycosylated, and this is essential for virulence (Schirm et al., 2003). In a systematic study of *H. pylori* glycoproteins, 26 proteins with diverse roles in pathogenesis were identified, arguably indicating a broader array of mechanisms for glycosylation-based virulence (Champasa et al., 2013).

POTENTIAL OF PROTEOMICS

Assessment of host–pathogen interactions is beneficial from a surveillance and diagnostic perspective. Pursuing the proteomic signatures of pathogenic bacteria during infection states could reveal dynamic changes that occur in the proteome or PTM proteome, which are often induced due to the selection pressure occurring during the interaction with the host. High-resolution mass spectrometry based proteomics has the ability to bridge the gap between genomics and the physiological mechanisms of behavior (Dove, 1999). Global and quantitative analysis of the pathogen proteome provides an in depth understanding of the molecular events that are differentially regulated during the onset of infection. Dynamic changes occurring at the proteome level, quantitative post-translational modification analysis are some of the basic and widely used applications of the shotgun proteomics technology. The high-throughput "omics" technology also allows for identification of macromolecular components important for specific cell-to-cell communications (Chait, 2006), for monitoring metabolic shifts and revealing metabolic pathways involved and/or activated during specific stages of infection. The conventional shot-gun proteomics or the bottom-up approach involves protein extraction from the organism of interest followed by digestion with a suitable endoprotease and subsequent analysis by LC-MS/MS. Entire bacterial proteomes and factors that play a key role in certain virulence pathways can be identified and analyzed conveniently by employing this approach (Tracz et al., 2013; Alvarez Hayes et al., 2015). Conversely, top-down approaches that involve the analysis of intact proteins have also been adopted to characterize bacterial proteomes (Ansong et al., 2013). Additionally, while challenging, top-down proteomics can be employed to study PTMs such as phosphorylation, acetylation, etc, occurring in the bacterial proteome as well as the host proteome upon infection (Kelleher, 2004; Zhang and Ge, 2011).

Microbial proteomics additionally aids in detection of biomarkers which could possibly act as targets for drug based therapy (Guest et al., 2013). Proteomics studies conducted on the effects of drug dosage are also crucial for acquiring information about the real-time *in vivo* scenario, thus assisting in the development of antimicrobial drugs. Targeted proteomics is also a powerful method for focusing on a specific subset of proteins involved in infection and resistance mechanisms, allowing for mechanistic questions to be answered.

CURRENT TECHNOLOGY

A limited number of non-proteomics based methodologies are currently employed for the investigation of population dynamics of binary mixed cultures (Kluge et al., 2012), namely, Fluorescence

In Situ Hybridization (FISH; Rogers et al., 2000), Real-Time PCR (Higuchi et al., 1993), Flow Cytometry (Müller et al., 1995) and Terminal Restriction Fragment Length Polymorphism (T-RFLP; Schmidt et al., 2007). Metaproteome analysis of mixed cultures using 2-DE followed by LC-ESI-MS has been utilized to identify diverse microbial communities in the environment (Benndorf et al., 2007; Kluge et al., 2012). High-resolution mass spectrometry based approaches are routinely employed to study bacterial infection models or bacterial co-cultures. Global subcellular protein profiling has been applied to reveal subcellular distributions of proteins, subsequently used to reconstruct functional networks and to identify new targets responsible for pathogenicity (Mawuenyega et al., 2005). *De novo* sequencing has been applied for peptide/protein identification when the DNA sequence coverage of a particular organism has been unavailable or incomplete (Wilmes and Bond, 2004). Centrifugation followed by detergent solubilization (Fernandez-Arenas et al., 2007); ImmunoMagnetic Separation (IMS) using anti-IgG-coated Dynabeads[TM] in combination with antisera (Twine et al., 2006); or Fluorescence-Activated Cell Sorting (FACS; Becker et al., 2006) based approaches are routinely used for separation of bacterial cells from host cells. Pulse-chase experiments using ^{35}S-labeled methionine or cysteine can also be employed to monitor changes occurring at the level of protein synthesis or protein turn-over (Schmidt and Volker, 2011). Quantification of proteome dynamics post infection can be done by label free approaches [for example, Luo et al. (2014) identified 2125 phosphopeptides and quantified 253 and 344 up-regulated phosphorylation events from PRRSV-infected pulmonary alveolar macrophages 12 and 36 h postinoculation respectively by employing label free approaches (Luo et al., 2014)] or by other label based relative quantitation methods such as iTRAQ (isobaric tagging for relative and absolute quantitation) or SILAC (stable isotope labeling with amino acids in cell culture; Shui et al., 2009). In the paper by Shui et al. (2009), a total of 1286 proteins were identified from murine macrophages during *M. tuberculosis* infection, of which 463 were identified by both SILAC and iTRAQ labeling strategies. Targeted proteomics, that forms a part of the newer generation proteomics approaches, can also be applied to study specific changes occurring in the host or bacterial cell by applying multiple reaction monitoring mass spectrometry (MRM-MS). Lange et al. (2008) employ multiple reaction monitoring to understand the dynamics of virulence factors from the Gram-positive bacterium *Streptococcus pyogenes*. Similarly, Karlsson et al. (2012) use selected reaction monitoring mass spectrometry (SRM-MS) to decipher the *S. pyogenes* proteome. Mass-spectrometry based imaging techniques form a fresh avenue of approach for studying differences in microbial communities (Wilmes and Bond, 2009).

TECHNICAL CHALLENGES

In the field of infection biology, the characterization of microbes existing as interacting communities is extremely challenging but essential (Kluge et al., 2012). Genome is a static entity, while the proteome is dynamic, creating a higher order of

complexity in whole proteome analysis in comparison to the genome. The presence of PTMs further increases the functional diversity of an organism's proteome. Furthermore, only a fraction of a given organism's proteome is modified at any given time point, making it harder to detect or quantify certain events. Limitations with respect to instrumentation can also decrease the mass range of detection, hence losing out on critical biological information. Another major limiting factor is the absence of a complete genome sequence for many relevant bacterial species in complex environmental samples. As a consequence, limitations in the analysis of protein modifications could occur mainly due to the partial sequence coverage. When investigating bacterial proteomes under conditions of infection, extra care must be taken to ensure that the ratio of abundance of the bacterial to eukaryal proteins is not disproportional, as the bacterial proteome can easily be drowned in the noise of the highly abundant eukaryal proteins. Furthermore, the complete recoverability of the low number of bacterial cells under such conditions is difficult. This is a challenge for the PTM analysis, as generally a high amount of protein is required for enrichment of modified peptides. Additionally, mammalian-based infection studies are limited by ethical constraints, and a delicate balance must be struck to gain mechanistic understanding in a true *in vivo* setup.

PERSPECTIVES

Developments in experimental methodologies and instrumentation have made it possible to efficiently investigate host–pathogen interactions. However, there is still a necessity for further advances in instrumentation, technology and improvements in data processivity to be able to (a) obtain a complete or sufficiently high coverage of the pathogen proteome starting with low input material; (b) conveniently identify and differentiate between organism-specific proteomes to be able to monitor the species-specific changes occurring; and (c) routinely monitor alterations occurring at the post-translation level to reveal the pathways relevant for drug targeting. Proteomics based mass-spectrometry is a powerful tool that helps obtain a functional understanding of microbial co-operativity, interaction, evolution, physiological changes and cellular functioning. Quantitative proteomic profiling of host–pathogen interactions and discovery mode proteomics, in combination with bioinformatics, have been shown as an effective approach to illuminate key factors intrinsic to host invasion and disease propagation. Additionally, data-dependent analysis or targeted approaches show a high degree of promise and are now being applied for biomarker discovery and clinical applications. The potential for definite target selection and validation should be further exploited to reveal basic solutions to important biological issues. Furthermore, proteomics analysis of PTMs from the bacterial perspective would be an interesting avenue to venture into. Apart from a few global phosphoproteome analyses, currently there are not many reports on other post-translationally modified pathogenic bacterial proteomes. Resolving some of the above-described technical challenges could aid in unraveling physiological pathways affected by PTMs that help in virulence.

REFERENCES

Alvarez Hayes, J., Lamberti, Y., Surmann, K., Schmidt, F., Volker, U., and Rodriguez, M. E. (2015). Shotgun proteome analysis of *Bordetella pertussis* reveals a distinct influence of iron availability on the bacterial metabolism, virulence, and defense response. *Proteomics* 15, 2258–2266. doi: 10.1002/pmic.201400512

Ansong, C., Wu, S., Meng, D., Liu, X., Brewer, H. M., Deatherage Kaiser, B. L., et al. (2013). Top-down proteomics reveals a unique protein S-thiolation switch in *Salmonella typhimurium* in response to infection-like conditions. *Proc. Natl. Acad. Sci. U.S.A.* 110, 10153–10158. doi: 10.1073/pnas.1221210110

Ashida, H., Kim, M., and Sasakawa, C. (2014). Exploitation of the host ubiquitin system by human bacterial pathogens. *Nat. Rev. Microbiol.* 12, 399–413. doi: 10.1038/nrmicro3259

Ashida, H., Kim, M., Schmidt-Supprian, M., Ma, A., Ogawa, M., and Sasakawa, C. (2010). A bacterial E3 ubiquitin ligase IpaH9.8 targets NEMO/IKKgamma to dampen the host NF-kappaB-mediated inflammatory response. *Nat. Cell Biol.* 12, 66–73; sup 61–69. doi: 10.1038/ncb2006

Becker, D., Selbach, M., Rollenhagen, C., Ballmaier, M., Meyer, T. F., Mann, M., et al. (2006). Robust Salmonella metabolism limits possibilities for new antimicrobials. *Nature* 440, 303–307. doi: 10.1038/nature04616

Bendt, A. K., Burkovski, A., Schaffer, S., Bott, M., Farwick, M., and Hermann, T. (2003). Towards a phosphoproteome map of *Corynebacterium glutamicum*. *Proteomics* 3, 1637–1646. doi: 10.1002/pmic.200300494

Benndorf, D., Balcke, G. U., Harms, H., and von Bergen, M. (2007). Functional metaproteome analysis of protein extracts from contaminated soil and groundwater. *ISME J.* 1, 224–234. doi: 10.1038/ismej.2007.39

Canova, M. J., and Molle, V. (2014). Bacterial serine/threonine protein kinases in host–pathogen interactions. *J. Biol. Chem.* 289, 9473–9479. doi: 10.1074/jbc.R113.529917

Chait, B. T. (2006). Chemistry. Mass spectrometry: bottom-up or top-down? *Science* 314, 65–66. doi: 10.1126/science.1133987

Champasa, K., Longwell, S. A., Eldridge, A. M., Stemmler, E. A., and Dube, D. H. (2013). Targeted identification of glycosylated proteins in the gastric pathogen *Helicobacter pylori* (Hp). *Mol. Cell. Proteomics* 12, 2568–2586. doi: 10.1074/mcp.M113.029561

Chowdhury, F., Begum, Y. A., Alam, M. M., Khan, A. I., Ahmed, T., Bhuiyan, M. S., et al. (2010). Concomitant enterotoxigenic *Escherichia coli* infection induces increased immune responses to *Vibrio cholerae* O1 antigens in patients with cholera in Bangladesh. *Infect. Immun.* 78, 2117–2124. doi: 10.1128/IAI.01426-09

Dove, A. (1999). Proteomics: translating genomics into products? *Nat. Biotechnol.* 17, 233–236. doi: 10.1038/6972

Fernandez-Arenas, E., Cabezon, V., Bermejo, C., Arroyo, J., Nombela, C., Diez-Orejas, R., et al. (2007). Integrated proteomics and genomics strategies bring new insight into *Candida albicans* response upon macrophage interaction. *Mol. Cell. Proteomics* 6, 460–478. doi: 10.1074/mcp.M600210-MCP200

Galyov, E. E., Hakansson, S., Forsberg, A., and Wolf-Watz, H. (1993). A secreted protein kinase of *Yersinia pseudotuberculosis* is an indispensable virulence determinant. *Nature* 361, 730–732. doi: 10.1038/361730a0

Ge, J., Xu, H., Li, T., Zhou, Y., Zhang, Z., Li, S., et al. (2009). A Legionella type IV effector activates the NF-kappaB pathway by phosphorylating the IkappaB family of inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* 106, 13725–13730. doi: 10.1073/pnas.0907200106

Grosdent, N., Maridonneau-Parini, I., Sory, M. P., and Cornelis, G. R. (2002). Role of Yops and adhesins in resistance of *Yersinia enterocolitica* to phagocytosis. *Infect. Immun.* 70, 4165–4176. doi: 10.1128/IAI.70.8.4165-4176.2002

Guest, P. C., Gottschalk, M. G., and Bahn, S. (2013). Proteomics: improving biomarker translation to modern medicine? *Genome Med.* 5, 17. doi: 10.1186/gm421

Hansen, A. M., Chaerkady, R., Sharma, J., Diaz-Mejia, J. J., Tyagi, N., Renuse, S., et al. (2013). The *Escherichia coli* phosphotyrosine proteome relates to core pathways and virulence. *PLoS Pathog.* 9:e1003403. doi: 10.1371/journal.ppat.1003403

Hanuszkiewicz, A., Pittock, P., Humphries, F., Moll, H., Rosales, A. R., Molinaro, A., et al. (2014). Identification of the flagellin glycosylation system in *Burkholderia cenocepacia* and the contribution of glycosylated flagellin to evasion of

human innate immune responses. *J. Biol. Chem.* 289, 19231–19244. doi: 10.1074/jbc.M114.562603

Higuchi, R., Fockler, C., Dollinger, G., and Watson, R. (1993). Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Biotechnology (N Y)* 11, 1026–1030.

Horstmann, N., Sahasrabhojane, P., Saldana, M., Ajami, N. J., Flores, A. R., Sumby, P., et al. (2015). Characterization of the effect of the histidine kinase CovS on response regulator phosphorylation in group A *Streptococcus*. *Infect. Immun.* 83, 1068–1077. doi: 10.1128/IAI.02659-14

Juris, S. J., Rudolph, A. E., Huddler, D., Orth, K., and Dixon, J. E. (2000). A distinctive role for the *Yersinia protein* kinase: actin binding, kinase activation, and cytoskeleton disruption. *Proc. Natl. Acad. Sci. U.S.A.* 97, 9431–9436. doi: 10.1073/pnas.170281997

Juris, S. J., Shah, K., Shokat, K., Dixon, J. E., and Vacratsis, P. O. (2006). Identification of otubain 1 as a novel substrate for the *Yersinia protein* kinase using chemical genetics and mass spectrometry. *FEBS Lett.* 580, 179–183. doi: 10.1016/j.febslet.2005.11.071

Kaniga, K., Uralil, J., Bliska, J. B., and Galan, J. E. (1996). A secreted protein tyrosine phosphatase with modular effector domains in the bacterial pathogen *Salmonella typhimurium*. *Mol. Microbiol.* 21, 633–641.

Karlsson, C., Malmstrom, L., Aebersold, R., and Malmstrom, J. (2012). Proteome-wide selected reaction monitoring assays for the human pathogen *Streptococcus pyogenes*. *Nat. Commun.* 3, 1301. doi: 10.1038/ncomms2297

Kelleher, N. L. (2004). Top-down proteomics. *Anal. Chem.* 76, 197A–203A. doi: 10.1021/ac0415657

Klocke, M., Mahnert, P., Mundt, K., Souidi, K., and Linke, B. (2007). Microbial community analysis of a biogas-producing completely stirred tank reactor fed continuously with fodder beet silage as mono-substrate. *Syst. Appl. Microbiol.* 30, 139–151. doi: 10.1016/j.syapm.2006.03.007

Kluge, S., Hoffmann, M., Benndorf, D., Rapp, E., and Reichl, U. (2012). Proteomic tracking and analysis of a bacterial mixed culture. *Proteomics* 12, 1893–1901. doi: 10.1002/pmic.201100362

Knodler, L. A., Winfree, S., Drecktrah, D., Ireland, R., and Steele-Mortimer, O. (2009). Ubiquitination of the bacterial inositol phosphatase, SopB, regulates its biological activity at the plasma membrane. *Cell. Microbiol.* 11, 1652–1670. doi: 10.1111/j.1462-5822.2009.01356.x

Kubori, T., and Galan, J. E. (2003). Temporal regulation of salmonella virulence effector function by proteasome-dependent protein degradation. *Cell* 115, 333–342. doi: 10.1016/S0092-8674(03)00849-3

Lange, V., Malmstrom, J. A., Didion, J., King, N. L., Johansson, B. P., Schafer, J., et al. (2008). Targeted quantitative analysis of *Streptococcus pyogenes* virulence factors by multiple reaction monitoring. *Mol. Cell. Proteomics* 7, 1489–1500. doi: 10.1074/mcp.M800032-MCP200

Lewis, J. D., Lee, A., Ma, W., Zhou, H., Guttman, D. S., and Desveaux, D. (2011). The YopJ superfamily in plant-associated bacteria. *Mol. Plant Pathol.* 12, 928–937. doi: 10.1111/j.1364-3703.2011.00719.x

Li, H., Xu, H., Zhou, Y., Zhang, J., Long, C., Li, S., et al. (2007). The phosphothreonine lyase activity of a bacterial type III effector family. *Science* 315, 1000–1003. doi: 10.1126/science.1138960

Lin, M. H., Hsu, T. L., Lin, S. Y., Pan, Y. J., Jan, J. T., Wang, J. T., et al. (2009). Phosphoproteomics of *Klebsiella pneumoniae* NTUH-K2044 reveals a tight link between tyrosine phosphorylation and virulence. *Mol. Cell. Proteomics* 8, 2613–2623. doi: 10.1074/mcp.M900276-MCP200

Liu, F., Yang, M., Wang, X., Yang, S., Gu, J., Zhou, J., et al. (2014). Acetylome analysis reveals diverse functions of lysine acetylation in *Mycobacterium tuberculosis*. *Mol. Cell Proteomics* 13, 3352–3366. doi: 10.1074/mcp.M114.041962

Luo, R., Fang, L., Jin, H., Wang, D., An, K., Xu, N., et al. (2014). Label-free quantitative phosphoproteomic analysis reveals differentially regulated proteins and pathway in PRRSV-infected pulmonary alveolar macrophages. *J. Proteome Res.* 13, 1270–1280. doi: 10.1021/pr400852d

Ma, Q., Zou, Y., Lv, Y., Song, H., and Yuan, Y. J. (2014). Comparative proteomic analysis of experimental evolution of the *Bacillus cereus–Ketogulonicigenium vulgare* co-culture. *PLoS ONE* 9:e91789. doi: 10.1371/journal.pone.0091789

Mawuenyega, K. G., Forst, C. V., Dobos, K. M., Belisle, J. T., Chen, J., Bradbury, E. M., et al. (2005). *Mycobacterium tuberculosis* functional network analysis by global subcellular protein profiling. *Mol. Biol. Cell* 16, 396–404. doi: 10.1091/mbc.E04-04-0329

Mesquita, F. S., Thomas, M., Sachse, M., Santos, A. J., Figueira, R., and Holden, D. W. (2012). The *Salmonella deubiquitinase* SseL inhibits selective autophagy of cytosolic aggregates. *PLoS Pathog.* 8:e1002743. doi: 10.1371/journal.ppat.1002743

Miller, M., Donat, S., Rakette, S., Stehle, T., Kouwen, T. R., Diks, S. H., et al. (2010). Staphylococcal PknB as the first prokaryotic representative of the proline-directed kinases. *PLoS ONE* 5:e9057. doi: 10.1371/journal.pone.0009057

Mujtaba, S., Winer, B. Y., Jaganathan, A., Patel, J., Sgobba, M., Schuch, R., et al. (2013). Anthrax SET protein: a potential virulence determinant that epigenetically represses NF-kappaB activation in infected macrophages. *J. Biol. Chem.* 288, 23458–23472. doi: 10.1074/jbc.M113.467696

Mukherjee, S., Keitany, G., Li, Y., Wang, Y., Ball, H. L., Goldsmith, E. J., et al. (2006). *Yersinia* YopJ acetylates and inhibits kinase activation by blocking phosphorylation. *Science* 312, 1211–1214. doi: 10.1126/science.1126867

Mukherjee, S., Liu, X., Arasaki, K., McDonough, J., Galan, J. E., and Roy, C. R. (2011). Modulation of Rab GTPase function by a protein phosphocholine transferase. *Nature* 477, 103–106. doi: 10.1038/nature10335

Müller, S., Lösche, A., Bley, T., and Scheper, T. (1995). A flow cytometric approach for characterization and differentiation of bacteria during microbial processes. *Appl. Microbiol. Biotechnol.* 43, 93–101. doi: 10.1007/BF00170629

Neunuebel, M. R., Chen, Y., Gaspar, A. H., Backlund, P. S., Jr., Yergey, A., and Machner, M. P. (2011). De-AMPylation of the small GTPase Rab1 by the pathogen *Legionella pneumophila*. *Science* 333, 453–456. doi: 10.1126/science.1207193

Odendall, C., Rolhion, N., Forster, A., Poh, J., Lamont, D. J., Liu, M., et al. (2012). The Salmonella kinase SteC targets the MAP kinase MEK to regulate the host actin cytoskeleton. *Cell Host Microbe* 12, 657–668. doi: 10.1016/j.chom.2012.09.011

Okuda, J., Toyotome, T., Kataoka, N., Ohno, M., Abe, H., Shimura, Y., et al. (2005). Shigella effector IpaH9.8 binds to a splicing factor U2AF(35) to modulate host immune responses. *Biochem. Biophys. Res. Commun.* 333, 531–539. doi: 10.1016/j.bbrc.2005.05.145

Parker, J. L., Jones, A. M., Serazetdinova, L., Saalbach, G., Bibb, M. J., and Naldrett, M. J. (2010). Analysis of the phosphoproteome of the multicellular bacterium *Streptomyces coelicolor* A3(2) by protein/peptide fractionation, phosphopeptide enrichment and high-accuracy mass spectrometry. *Proteomics* 10, 2486–2497. doi: 10.1002/pmic.201000090

Persson, C., Nordfelth, R., Andersson, K., Forsberg, A., Wolf-Watz, H., and Fallman, M. (1999). Localization of the *Yersinia* PTPase to focal complexes is an important virulence mechanism. *Mol. Microbiol.* 33, 828–838.

Price, C. T., Al-Quadan, T., Santic, M., Jones, S. C., and Abu Kwaik, Y. (2010). Exploitation of conserved eukaryotic host cell farnesylation machinery by an F-box effector of *Legionella pneumophila*. *J. Exp. Med.* 207, 1713–1726. doi: 10.1084/jem.20100771

Prisic, S., Dankwa, S., Schwartz, D., Chou, M. F., Locasale, J. W., Kang, C. M., et al. (2010). Extensive phosphorylation with overlapping specificity by *Mycobacterium tuberculosis* serine/threonine protein kinases. *Proc. Natl. Acad. Sci. U.S.A.* 107, 7521–7526. doi: 10.1073/pnas.0913482107

Putignani, L., Del Chierico, F., Petrucca, A., Vernocchi, P., and Dallapiccola, B. (2014). The human gut microbiota: a dynamic interplay with the host from birth to senescence settled during childhood. *Pediatr. Res.* 76, 2–10. doi: 10.1038/pr.2014.49

Reiterer, V., Grossniklaus, L., Tschon, T., Kasper, C. A., Sorg, I., and Arrieumerlou, C. (2011). *Shigella flexneri* type III secreted effector OspF reveals new crosstalks of proinflammatory signaling pathways during bacterial infection. *Cell. Signal.* 23, 1188–1196. doi: 10.1016/j.cellsig.2011.03.006

Rogers, J. B., DuTeau, N. M., and Reardon, K. F. (2000). Use of 16S-rRNA to investigate microbial population dynamics during biodegradation of toluene and phenol by a binary culture. *Biotechnol. Bioeng.* 70, 436–445. doi: 10.1002/1097-0290(20001120)70:4<436::AID-BIT9>3.0.CO;2-6

Royan, S. V., Jones, R. M., Koutsouris, A., Roxas, J. L., Falzari, K., Weflen, A. W., et al. (2010). Enteropathogenic *E. coli* non-LEE encoded effectors NleH1 and NleH2 attenuate NF-kappaB activation. *Mol. Microbiol.* 78, 1232–1245. doi: 10.1111/j.1365-2958.2010.07400.x

Schirm, M., Soo, E. C., Aubry, A. J., Austin, J., Thibault, P., and Logan, S. M. (2003). Structural, genetic and functional characterization of the flagellin glycosylation process in *Helicobacter pylori*. *Mol. Microbiol.* 48, 1579–1592. doi: 10.1046/j.1365-2958.2003.03527.x

Schmidt, F., and Volker, U. (2011). Proteome analysis of host–pathogen interactions: investigation of pathogen responses to the host cell environment. *Proteomics* 11, 3203–3211. doi: 10.1002/pmic.201100158

Schmidt, J. K., König, B., and Reichl, U. (2007). Characterization of a three bacteria mixed culture in a chemostat: evaluation and application of a quantitative terminal-restriction fragment length polymorphism (T-RFLP) analysis for absolute and species specific cell enumeration. *Biotechnol. Bioeng.* 96, 738–756. doi: 10.1002/bit.21147

Schmutz, C., Ahrne, E., Kasper, C. A., Tschon, T., Sorg, I., Dreier, R. F., et al. (2013). Systems-level overview of host protein phosphorylation during *Shigella flexneri* infection revealed by phosphoproteomics. *Mol. Cell Proteomics* 12, 2952–2968. doi: 10.1074/mcp.M113.029918

Scholz, R., Imami, K., Scott, N. E., Trimble, W. S., Foster, L. J., and Finlay, B. B. (2015). Novel host proteins and signaling pathways in enteropathogenic *E. coli* pathogenesis identified by global phosphoproteome analysis. *Mol. Cell Proteomics* 14, 1927–1945. doi: 10.1074/mcp.M114.046847

Seyedarabi, A., Sullivan, J. A., Sasakawa, C., and Pickersgill, R. W. (2010). A disulfide driven domain swap switches off the activity of Shigella IpaH9.8 E3 ligase. *FEBS Lett.* 584, 4163–4168. doi: 10.1016/j.febslet.2010.09.006

Shui, W., Gilmore, S. A., Sheu, L., Liu, J., Keasling, J. D., and Bertozzi, C. R. (2009). Quantitative proteomic profiling of host–pathogen interactions: the macrophage response to *Mycobacterium tuberculosis* lipids. *J. Proteome Res.* 8, 282–289. doi: 10.1021/pr800422e

Sun, F., Ding, Y., Ji, Q., Liang, Z., Deng, X., Wong, C. C., et al. (2012). Protein cysteine phosphorylation of SarA/MgrA family transcriptional regulators mediates bacterial virulence and antibiotic resistance. *Proc. Natl. Acad. Sci. U.S.A.* 109, 15461–15466. doi: 10.1073/pnas.1205952109

Tan, Y., Arnold, R. J., and Luo, Z. Q. (2011). *Legionella pneumophila* regulates the small GTPase Rab1 activity by reversible phosphorylcholination. *Proc. Natl. Acad. Sci. U.S.A.* 108, 21212–21217. doi: 10.1073/pnas.1114023109

Thao, S., Chen, C. S., Zhu, H., and Escalante-Semerena, J. C. (2010). Nepsilon-lysine acetylation of a bacterial transcription factor inhibits Its DNA-binding activity. *PLoS ONE* 5:e15123. doi: 10.1371/journal.pone.0015123

Tijhuis, L., Rekswinkel, E., Vanloosdrecht, M. C. M., and Heijnen, J. J. (1994). Dynamics of population and biofilm structure in the biofilm airlift suspension reactor for carbon and nitrogen removal. *Water Sci. Technol.* 29, 377.

Tracz, D. M., McCorrister, S. J., Chong, P. M., Lee, D. M., Corbett, C. R., and Westmacott, G. R. (2013). A simple shotgun proteomics method for rapid bacterial identification. *J. Microbiol. Methods* 94, 54–57. doi: 10.1016/j.mimet.2013.04.008

Tseng, T. T., Tyler, B. M., and Setubal, J. C. (2009). Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *BMC Microbiol.* 9(Suppl. 1), S2. doi: 10.1186/1471-2180-9-S1-S2

Twine, S. M., Mykytczuk, N. C., Petit, M. D., Shen, H., Sjostedt, A., Wayne Conlan, J., et al. (2006). *In vivo* proteomic analysis of the intracellular bacterial pathogen, *Francisella tularensis*, isolated from mouse spleen. *Biochem. Biophys. Res. Commun.* 345, 1621–1633. doi: 10.1016/j.bbrc.2006.05.070

Vilcheze, C., Molle, V., Carrere-Kremer, S., Leiba, J., Mourey, L., Shenai, S., et al. (2014). Phosphorylation of KasB regulates virulence and acid-fastness in *Mycobacterium tuberculosis*. *PLoS Pathog.* 10:e1004115. doi: 10.1371/journal.ppat.1004115

Voisin, S., Watson, D. C., Tessier, L., Ding, W., Foote, S., Bhatia, S., et al. (2007). The cytoplasmic phosphoproteome of the Gram-negative bacterium *Campylobacter jejuni*: evidence for modification by unidentified protein kinases. *Proteomics* 7, 4338–4348. doi: 10.1002/pmic.200700483

Whitmore, S. E., and Lamont, R. J. (2012). Tyrosine phosphorylation and bacterial virulence. *Int. J. Oral. Sci.* 4, 1–6. doi: 10.1038/ijos.2012.6

Wiley, D. J., Nordfeldth, R., Rosenzweig, J., DaFonseca, C. J., Gustin, R., Wolf-Watz, H., et al. (2006). The Ser/Thr kinase activity of the *Yersinia protein* kinase A (YpkA) is necessary for full virulence in the mouse, mollifying phagocytes, and disrupting the eukaryotic cytoskeleton. *Microb. Pathog.* 40, 234–243. doi: 10.1016/j.micpath.2006.02.001

Wilmes, P., and Bond, P. L. (2004). The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms. *Environ. Microbiol.* 6, 911–920. doi: 10.1111/j.1462-2920.2004.00687.x

Wilmes, P., and Bond, P. L. (2009). Microbial community proteomics: elucidating the catalysts and metabolic mechanisms that drive the Earth's biogeochemical cycles. *Curr. Opin. Microbiol.* 12, 310–317. doi: 10.1016/j.mib.2009.03.004

Xie, L., Liu, W., Li, Q., Chen, S., Xu, M., Huang, Q., et al. (2015). First succinyl-proteome profiling of extensively drug-resistant *Mycobacterium tuberculosis* revealed involvement of succinylation in cellular physiology. *J. Proteome Res.* 14, 107–119. doi: 10.1021/pr500859a

Yang, Y., Hu, M., Yu, K., Zeng, X., and Liu, X. (2015). Mass spectrometry-based proteomic approaches to study pathogenic bacteria-host interactions. *Protein Cell* 6, 265–274. doi: 10.1007/s13238-015-0136-6

Yarbrough, M. L., Li, Y., Kinch, L. N., Grishin, N. V., Ball, H. L., and Orth, K. (2009). AMPylation of Rho GTPases by Vibrio VopS disrupts effector binding and downstream signaling. *Science* 323, 269–272. doi: 10.1126/science.1166382

Zhang, C. G., Chromy, B. A., and McCutchen-Maloney, S. L. (2005). Host–pathogen interactions: a proteomic view. *Expert. Rev. Proteomics* 2, 187–202. doi: 10.1586/14789450.2.2.187

Zhang, H., and Ge, Y. (2011). Comprehensive analysis of protein modifications by top-down mass spectrometry. *Circ. Cardiovasc Genet.* 4, 711. doi: 10.1161/CIRCGENETICS.110.957829

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Mass Spectrometry Offers Insight into the Role of Ser/Thr/Tyr Phosphorylation in the Mycobacteria

Bridget Calder, Claudia Albeldas, Jonathan M. Blackburn and Nelson C. Soares**

Applied and Chemical Proteomics Group, Medical Biochemistry Division, Faculty of Health Sciences, Institute of Infectious Diseases and Molecular Medicine, University of Cape Town, Cape Town, South Africa

Phosphorylation is a post translational modification which can rapidly regulate biochemical pathways by altering protein function, and has been associated with pathogenicity in bacteria. Once engulfed by host macrophages, pathogenic bacteria are exposed to harsh conditions and must respond rapidly in order to survive. The causative agent of TB, *Mycobacterium tuberculosis*, is unusual amongst the bacteria because it can survive within the host macrophage for decades in a latent state, demonstrating a remarkable capacity to successfully evade the host immune response. This ability may be mediated in part by regulatory mechanisms such as ser/thr/tyr phosphorylation. Mass spectrometry-based proteomics has afforded us the capacity to identify hundreds of phosphorylation sites in the bacterial proteome, allowing for comparative phosphoproteomic studies in the mycobacteria. There remains an urgent need to validate the reported phosphosites, and to elucidate their biological function in the context of pathogenicity. However, given the sheer number of putative phosphorylation events in the mycobacterial proteome, and the technical difficulty of assigning biological function to a phosphorylation event, it will not be trivial to do so. There are currently six published phosphoproteomic investigations of a member of mycobacteria. Here, we combine the datasets from these studies in order to identify commonly detected phosphopeptides and phosphosites in order to present high confidence candidates for further validation. By applying modern mass spectrometry-based techniques to improve our understanding of phosphorylation and other PTMs in pathogenic bacteria, we may identify candidates for therapeutic intervention.

Keywords: post-translational modification, phosphorylation, mycobacteria, tuberculosis, proteomics, phosphoproteomics, mass spectrometry, virulence

INTRODUCTION

Proteins are the bioactive molecule in the cell, and contribute to survival, growth, and reproduction by interacting with each other, and with metabolites, lipids, nucleic acids and carbohydrates, and catalyzing biological reactions (Nørregaard Jensen, 2004). Protein biosynthesis and degradation are tightly regulated by complex biochemical systems in response to the changing needs of the cell. However, there is an additional mechanism which allows cells to respond rapidly and efficiently to the external and internal conditions. Post translational modification (PTM) by selective covalent processing of proteins—by proteolytic cleavage or the addition of a modifying

group—can drastically alter the properties of a protein (Mann and Jensen, 2003; Mijakovic, 2010; Stülke, 2010). Post translational modifications add a layer of complexity to both bacterial and eukaryotic mechanisms of adaptation to the surrounding environment.

Modern mass spectrometry (MS) has enabled us to perform high throughput analysis of PTMs. Since modified proteins usually occur at low abundance, an enrichment process is typically carried out for a specific PTM prior to MS analysis (Semanjski and Macek, 2016). This results in increased resolution, sensitivity and fragmentation (Cain et al., 2014) and has contributed to the identification and localization of phosphosites in many species, including pathogenic bacteria. Once thought to be found only in eukaryotes, the discovery of Hanks-type family of kinases in bacteria indicates that a complex bacterial phosphorylation-mediated signaling system exists (Bakal and Davies, 2000). These serine/threonine protein kinases (STPKs) add a phosphate group to a serine/threonine, while BY-kinases add phosphate groups to Tyrosine residues. PTMs in bacteria have been linked to pathogenicity, virulence, resistance and persistence, and are vital for survival (Ge and Shan, 2011; Van Els et al., 2014). MS-based phosphoproteomic analysis has been carried out in a number of bacterial species, including *Bacillus subtilis* (Macek et al., 2007), *Escherichia coli* (Macek et al., 2008; Soares et al., 2013), *Streptococcus pneumonia* (Sun et al., 2009) *Listeria monocytogenes* (Misra et al., 2011), *Acinetobacter baumanii* (Soares et al., 2014), and *Mycobacterium tuberculosis*.

The Role of Phosphorylation in *M. tuberculosis*

The bacillus *M. tuberculosis* is the causative agent of tuberculosis (TB), a leading global health crisis which has claimed millions of lives by continuing to evade clinical intervention. This is largely due to its ability to lay dormant for many years in the body, resurfacing if the hosts' immune system becomes compromised (Gengenbacher and Kaufmann, 2012). When the bacilli enter the human lung, they are ingested by alveolar macrophage cells of the human immune system. The macrophages respond by becoming acidic and exposing the pathogen to lytic enzymes, oxygenated lipids, fatty acids, and reactive oxygen and nitrogen intermediates (Schnappinger et al., 2003). In order to survive in the adverse environment of the macrophage, *M. tuberculosis* needs to react swiftly, which is possible through the regulatory mechanisms afforded by PTMs (Cain et al., 2014). Mycobacterial STPK phosphorylation has been long associated with pathogenicity (Sherman and Grundner, 2014), which has driven efforts to improve our understanding of the role of phosphorylation in *M. tuberculosis* (Prisic and Husson, 2014). Currently, there are six published manuscripts describing the phosphoproteome of a member of the Mycobacteria—*M. tuberculosis* H37Rv (Prisic et al., 2010; Kusebauch et al., 2014), *M. smegmatis,* and *M. bovis* BCG (Nakedi et al., 2015; Zheng et al., 2015), a clinical isolate of *M. tuberculosis* Beijing lineage (Fortuin et al., 2015) and a ΔpknE deletion mutant strain of *M. tuberculosis* (Parandhaman et al., 2014b). While these studies are discussed in more detail below, a summary of the methods used and relevant results for each of them is presented in **Table 1**. A major difficulty in comparing

phosphoproteomic studies is that we cannot compare those phosphosites that were uniquely identified by each study, because it is impossible to determine whether that uniqueness is as a result of biological differences or the stochastic nature of discovery-driven MS-based proteomics. The only available study in the mycobacteria where those uniquely detected phosphosites are comparable is Nakedi et al. (2015), because the study compared two strains under the same experimental conditions. In this case, a relevant conclusion drawn by the authors is that the phosphosite patterns detected in *M. smegmatis* and *M. bovis* BCG are often species specific and these phosphorylation events are commonly occurring on entirely different peptides.

The identification of PTMs which contribute to the pathogenicity of *M. tuberculosis* by enhancing its ability to survive in the macrophage is of great interest to the medical community, as these represent attractive candidates for therapeutic intervention. This review will focus specifically on the use of MS-based techniques which have been used to identify phosphorylated proteins in the mycobacteria, with particular focus on the identification of phosphorylation sites which may contribute to pathogenicity and are conserved across pathogenic strains of mycobacteria.

COMPARATIVE PHOSPHOPROTEOMIC ANALYSIS OF MYCOBACTERIA

Prisic et al. (2010) were the first to present a global view of the phosphoproteome of *M. tuberculosis* H37Rv. They used in-gel tryptic digest to proteolytically prepare samples of H37Rv lysate grown under conditions of NO stress, oxidative stress, hypoxia and using glucose, or acetate as a carbon source. A total of 152 samples were analyzed on an LTQ mass spectrometer following enrichment for phosphopeptides using titanium dioxide beads. In this manner, the authors detected a total of 506 phosphosites on 301 proteins and identified a dominant motif for *M. tuberculosis* STPKs, which was validated using synthetic peptides. This initial investigation reported 40% Ser: 60% Thr phosphorylation, and no Tyr phospho-sites—at the time of publication there was no conclusive molecular evidence for Tyr phosphorylation in *M. tuberculosis*, although there was a long-established association between Tyr phosphorylation and pathogenicity in other bacteria. (Ilan et al., 1999). Kusebauch et al. (2014) reported Tyr phosphorylation in *M. tuberculosis* for the first time, after establishing that the known *M. tuberculosis* STPKs have the capacity to phosphorylate Tyr, and then by carrying out LC MS/MS analysis on *M. tuberculosis* culture lysate enriched for phosphopeptides. In this manner, they detected 30 high-confidence Tyr phospho-sites on 17 *M. tuberculosis* proteins, contributing to a Ser: Thr: Tyr ratio of 34:62:4%. Intriguingly, these authors also found an additional 35 Tyr phosphorylation sites in publically accessible MS data from previously published proteomic studies of *M. tuberculosis*, where the assumption that there was no Tyr phosphorylation in *M. tuberculosis* had led the authors to overlook them. Subsequent MS based descriptions of the *M. tuberculosis* phosphoproteome have identified similar proportions of Tyr phosphorylation sites,

TABLE 1 | An overview of the methods used and results generated in the six currently available published studies investigating phosphorylation in the mycobacteria using mass spectrometry.

References	Strain/species	Fractionation method	Enrichment method	# Proteins identified	# Sites	%S	%T	%Y	Major conclusions
Prisic et al., 2010	*M. tuberculosis* H37Rv	SDS PAGE	Titanium dioxide beads	301	516	40	60	-	A broad range of proteins are phosphorylated on S/T residues in *M. tuberculosis*, and may contribute to virulence.
Kusebauch et al., 2014	*M. tuberculosis* H37Rv	None	IMAC with PHOS-select iron affinity gel (Sigma) and then Titan-sphere Phos-TiO kit (GL Sciences Inc.)	232		32	64	4	Tyr phosphorylation does occur in *M. tuberculosis*, a Tyr phosphosite of PknB in is important in regulating growth and therefore may affect survival in stress conditions.
Parandhaman et al., 2014b	*M. tuberculosis* H37Rv, *M. tuberculosis*△PknE mutant	2D SDS PAGE	None	68	N/A	N/A	N/A	N/A	Proteins which are affected by the △PknE mutation under NO stress are hypothesized to contribute to pathogenicity of *M. tuberculosis*.
Nakedi et al., 2015	*M. bovis* BCG	None	TiO2 beads	203	289	35	61.6	3.1	Differences between fast and slow growing mycobacteria may be related to phosphorylation, with the ability to respond rapidly to stress associated with metabolic cost and slow growth.
	M. smegmatis	None	TiO2 beads	76	106	39.47	57.02	3.51	
Fortuin et al., 2015	*M. tuberculosis* clinical isolate	Strong cation exchange	TiO2 beads	214	414	38	59	3	Novel phosphorylation events were identified in this clinical strain, expanding our understanding of phosphorylation in *M. tuberculosis*.
Zheng et al., 2015	*M. bovis* BCG	SDS PAGE or offline ACN fractionation	TiO2 beads	398	659	39.5	48.7	11.8	Gel-based and gel-free phosphoproteomic analysis of *M. tuberculosis* may identify new pharmacological targets for drug development.

some of which may be of particular importance in establishing virulence.

Recently, Nakedi et al. (2015) investigated the impact of protein phosphorylation in growth-related functions by measuring differential phosphorylation between two mycobacterial species during exponential growth phase—the fast growing, non-pathogenic, soil dwelling *Mycobacterium smegmatis* and the slow growing, attenuated strain of *Mycobacterium bovis* (BCG). BCG had consistently higher phosphorylation levels, with 289 phosphosites on 203 proteins, compared to *M. smegmatis* with 106 phosphosites found on 76 proteins. The phosphoproteins which were uniquely found in BCG were generally involved in cell growth and stress response. Even under optimal growth conditions, BCG appears to have a high level of phosphorylated stress response proteins which suggests the capability for quick on/off responses to stressors
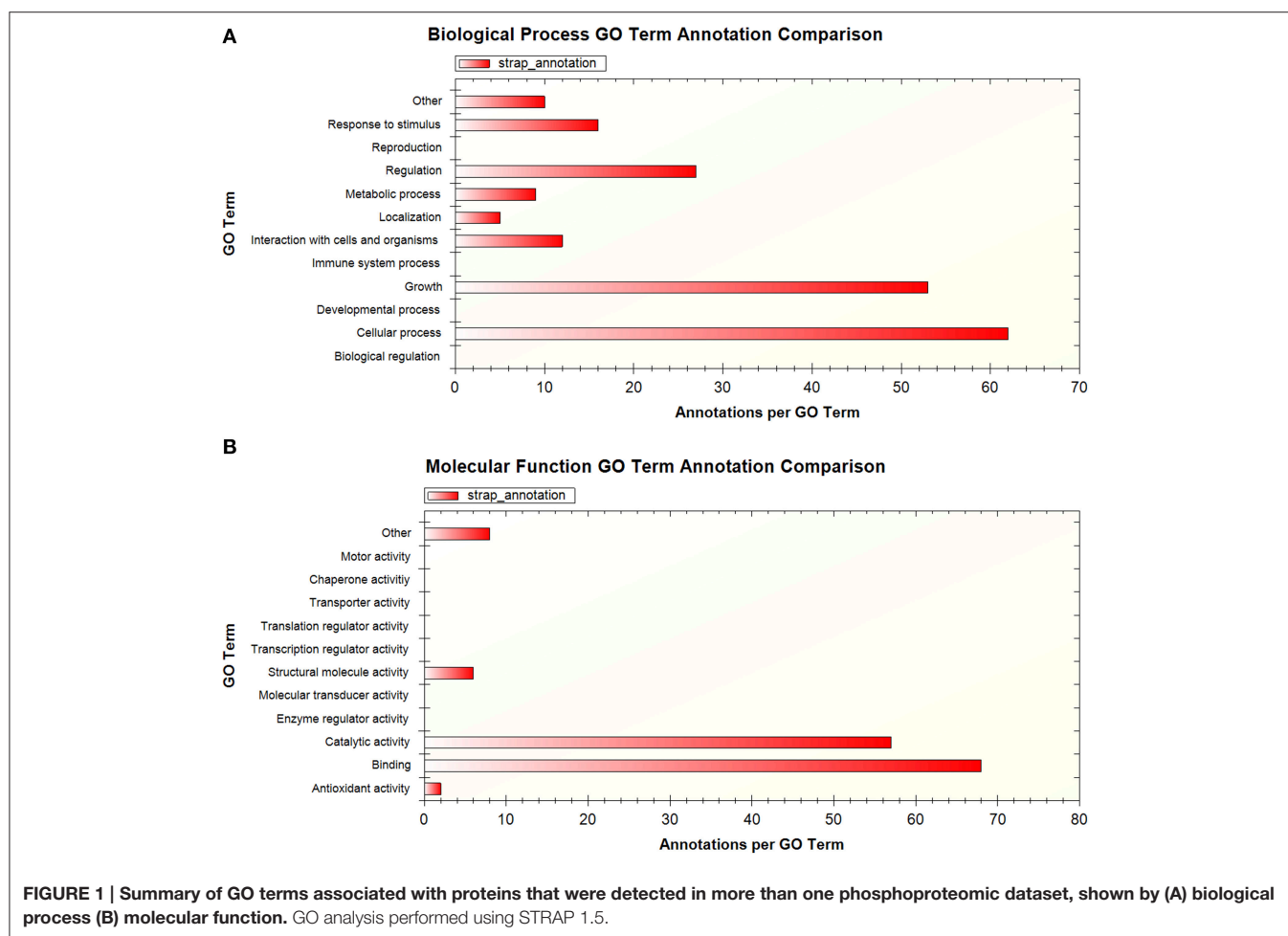
within the host, which ultimately allows the bacteria to respond rapidly and survive more effectively. The potential adaptive advantage for the pathogen may result in a fitness cost of slower growth. In a follow-up phosphoproteomic study, Zheng et al. (2015) found 659 phosphosites on 398 proteins in BCG harvested during stationary phase. The majority (40.1%) of identified phosphoproteins in this case were involved in regulation of metabolism. Again, these findings indicate that phosphorylation plays an important role in the slower metabolism of BCG, which may ultimately increase the capability for persistence within the host, and may be beneficial to the bacteria when faced with drug treatment (Evangelopoulos and McHugh, 2015).

A phosphoproteomic investigation of a hyper-virulent Beijing strain of *M. tuberculosis* by Fortuin et al. (2015) reported the identification of 414 phosphosites on 214 proteins. Of these, 252 were novel phosphosites which had not been identified in

TABLE 2 | Previously reported virulence factors in the *M. tuberculosis* complex which have been identified in more than one phosphoproteomic dataset, and their putative role in virulence.

Accession	Protein names	Phosphopeptides/sites	Putative role in virulence	References
P9WI73	Serine/threonine-protein kinase PknG	**S**GPG**T**QPADAQTA**T**SA**T**VR	Unclear, possibly growth and metabolism	Nguyen et al., 2005
P9WI77	Serine/threonine-protein kinase PknE	LPVPSTHPV**S**PGTR	MAPK signaling, survival	Parandhaman et al., 2014a
P9WI79	Serine/threonine-protein kinase PknD	WSPGDSA**T**VAGPLAADSR	Unclear, plays a role in TB infection of the CNS	Be et al., 2012
P9WIE5	Catalase-peroxidase (CP) (Peroxidase/catalase) KatG	DAITSGIEVVWTNTPTK	Antioxidant	Ng et al., 2004
P9WJB7	Nucleoid-associated protein EspR	AHGLP**S**AAQQK	Transcriptional regulation	Blasco et al., 2012
P9WMH1	Iron-dependent repressor IdeR	MNELVD**T**TEMYLR	Iron dependent regulation	Banerjee et al., 2011
I6Y748	Membrane protein	DPP**T**DPNLR	Unknown	n/a

The commonly detected phosphopeptide is indicated in bold, along with phosphosites that have been confirmed in more than one study.

FIGURE 1 | Summary of GO terms associated with proteins that were detected in more than one phosphoproteomic dataset, shown by (A) biological process (B) molecular function. GO analysis performed using STRAP 1.5.

previous phosphoproteome research on the H37Rv strain. Since the capability for complex signaling is directly related to the number of phosphorylation events, it could be inferred that an increasing number of phosphorylation events may play a role in differentiating virulence in different *M. tuberculosis* strains. This is supported by findings that a highly virulent, drug resistant

strain of *Acinetobacter baumannii* has almost double the number of phosphosites in comparison to the reference strain (Soares et al., 2014). It is also interesting to note that phosphorylation may be linked to drug resistance, which is of concern given the emergence of drug resistant strains of TB (Evangelopoulos and McHugh, 2015). Although the mechanism by which a

bacterium might accumulate additional phosphosites is not fully understood, it is of interest that Fortuin et al. (2015) identified phosphorylated forms of 9 out of the 11 STPKs encoded by the *M. tuberculosis* genome in this hypervirulent Beijing strain (Prisic and Husson, 2014), while Prisic et al. (2010) only identified four in H37Rv.

Although our understanding of the specific activity of Mycobacterial STPKs is largely incomplete, some substrates and their downstream functions have been associated with specific kinases (Supplementary Table 1). While mass spectrometry is capable of identifying hundreds of phospho-substrates, more laborious methods are necessary to associate these substrates with their corresponding kinases. The challenges inherent in mass-spectrometry-based kinase substrate identification are discussed in more detail by Sherman and Grundner (2014). Parandhaman et al. (2014b) made use of a ΔPknE deletion mutant of *M. tuberculosis* to identify PknE substrates during NO stress in *M. tuberculosis*, and identified 68 phosphoproteins by combining 2D PAGE MS with phospho-serine and phospho-threonine specific antibodies. The candidate PknE substrates identified in this manner may play a role in dormancy within the macrophage, which may have implications for virulence, once again highlighting the importance of phosphorylation in allowing the bacterium to respond effectively to the host environment. Given this information, and somewhat surprisingly, there is evidence that the PknE gene is non-essential for growth of *M. tuberculosis* in culture (Sassetti et al., 2003). Indeed, in culture-based models it seems that only three of the STPKs available to *M. tuberculosis* are individually essential for growth: PknA and PknB (Wehenkel et al., 2008); and PknG, which is essential for survival of *M. tuberculosis* within the macrophage. Walburger et al. (2004) demonstrated that a strain of *M. bovis* BCG carrying an inactivated PknG gene showed no differences in growth or cell morphology in liquid medium to the wild type, whereas the mutant was unable to survive when inside a macrophage because the bacteria were no longer able to prevent lysosomal fusion. In many pathogenic bacteria, but particularly in *M. tuberculosis*, we are increasingly aware of the complexity of the host/pathogen interaction during disease progression, while being limited to culture or animal based models in our attempts to understand it[1]. The fact that *M. tuberculosis* has evolved to exist in the intracellular space is highlighted by the metabolic changes which are observed in intracellular *M. tuberculosis* (Lee et al., 2013), but we have yet to explore the consequences of phosphorylation for *M. tuberculosis in vivo*. It is conceivable that intracellular STPKs have different activity compared to culture-based systems, and high-throughput phosphoproteomic analysis of intracellular *M. tuberculosis* would allow us to better understand the role of phosphorylation in the host pathogen interaction.

Identifying Functional Phosphosites for Further Characterization

The aim of discovery-based proteomics has been to catalog as many proteins as possible in a sample, with the intent being to better understand a biological condition or response. However,

the volumes of data generated in this manner have not necessarily achieved a deeper understanding of the biological systems in question. This problem is equally as confounding, if not more so, in phosphoproteomic investigations, since it is difficult to attribute biological significance to a discreet phosphorylation event. Many of the detected phosphorylation events reported by these large-scale studies require validation before we can begin to determine their function. To this end, meta-analysis of the available data may provide some insight for future investigations. Although the above-mentioned mycobacterial phosphoproteomic studies were conducted using different MS-based methodologies, and in different strains of mycobacteria, there are discreet phosphorylation events that were commonly observed in several datasets, and are therefore unlikely to be random. A phosphopeptide which is observed in more than one study, or across several mycobacterial strains, may therefore be the best starting point for further investigation in *M. tuberculosis*. To this end, here we present a compilation of commonly detected phosphopeptides/sites which may thus be of use for future functional phosphoproteomic investigations in the mycobacteria (Supplementary Table 2).

A total of 194 phosphopeptides representing 148 proteins were found to be represented by two or more sets of data according to the available published supplementary information. To facilitate comparison at the protein level, the peptide sequences were matched to H37Rv protein identifiers using the Protein Information Resource (PIR) batch peptide match tool against the Uniprot H37Rv proteome (Chen et al., 2013). GO analysis of the proteins corresponding to these shared phosphopeptides using STRAP 1.5 (Bhatia et al., 2009) revealed that the associated GO terms relate to broad regulatory functions, such as the regulation of cell metabolism, cellular processes, and growth (**Figures 1A,B**). While this is expected given the established role of phosphorylation in metabolic regulation in bacteria (Kochanowski et al., 2015), this is also potentially noteworthy as we are only beginning to unravel the importance of metabolic regulation for intracellular *M. tuberculosis* and how this relates to virulence (Eisenreich et al., 2010). Of particular interest are incidences where the more virulent clinical strain is phosphorylated on a different residue compared to the other strains, as in the example of the AEASIETPTPVQSQR peptide of TatA, a Sec-independent protein translocase protein, which was phosphorylated on T7 and/or T9 in *M. bovis* BCG and *M. tuberculosis* H37Rv but on S4 in the clinical strain. The functional significance of this difference remains unknown and should be validated and investigated, particularly in light of the contribution of the Tat pathway to virulence in *M. tuberculosis* (Feltcher et al., 2010). It should be noted that the utility of GO analysis in mycobacteria is limited by the availability of GO annotations and other resources such as KEGG pathway representation, for which coverage is generally poor.

Differential phosphorylation of the STPKs themselves can alter their enzymatic activity, and is another possible mechanism for altered pathogenicity in *M. tuberculosis* (Chopra et al., 2003; Durán et al., 2005). Supplementary Table 2 highlights that phosphorylation is commonly detected in PknA, B, D, E, G, and H in these mycobacteria; however the differences in localized

---

[1]MISSING:pmid:22320122. MISSING:pmid:22320122. (2016).

phosphosites between the virulent and less virulent strains are more pronounced in peptides corresponding to PknA, D, and G. Having already established the importance of PknA (Singh et al., 2006) and G (Walburger et al., 2004) for survival within the macrophage, the biological significance of these differences in phosphorylation and how they contribute virulence should now be ascertained. Forrellad et al. (2013) published a summary of the known virulence factors in the *M. tuberculosis* complex and their putative contribution to virulence. Cross referencing their table of results to ours identified proteins which are known virulence factors which are also commonly detected in these phosphoproteomic datasets. These proteins and their putative role in virulence in the *M. tuberculosis* complex are presented in **Table 2**. Included in these are some previously mentioned STPKs, as well as the proteins KatG, EspR, and IdeR.

CURRENT STATE OF PHOSPHOPROTEOMICS

Currently, the mycobacterial phosphoproteome has been qualitatively described, which has identified many high confidence phosphosites. However, a complete description of the phosphoproteome should include quantitative information for the specific phosphosite in question as well as for the phosphorylated protein. In addition, most of the available phosphoproteomic data is for bacterial culture under a single condition, and yet we recognize that phosphorylation is dynamic and can change rapidly (Macek et al., 2009). In order to address this, it is important to incorporate multiple time points and experimental conditions in future phosphoproteomic investigations. These objectives are specifically challenging for mass spectrometry-based proteomics, and despite promising technological advances the mycobacterial phosphoproteome has not yet been quantitatively assessed (de la Fuente van Bentem et al., 2008). There are many available quantitative proteomic tools which are applicable in bacterial phosphoproteomics, which have been discussed in detail elsewhere (Jers et al., 2008; Macek et al., 2009), however, not all of these are suitable for the mycobacteria. SILAC, for example, has been used successfully to quantify proteins in the bacterium *Bacillus subtilis* (Ravikumar et al., 2014), but is not currently possible in *M. tuberculosis* because the available lysine deficient mutants are not viable for SILAC. Alternative quantitative methods include label-free quantitation, which has the benefit of cost effectiveness allowing for multiple time points/conditions to be explored, or dimethyl labeling, which is a promising alternative to iTRAQ and other label-based techniques (Lau et al., 2014). iTRAQ has been successfully used to quantify the phosphorylation of arginine in *B. subtilis* (Schmidt et al., 2014), and therefore is suitable for measuring S/T/Y phosphorylation in bacterial systems, but is limited by the number of samples that can be analyzed concurrently and the high cost of the iTRAQ reagents. Validation of phosphoproteomic data using targeted proteomics along with specialized analysis software such as Skyline is an exciting prospect, as these powerful tools have the capability to

validate specific phosphosites as well as providing quantitative information. The absolute quantitation of modified peptides is possible through a combination of heavy-labeled AQUA peptides and selected reaction monitoring (SRM) MS, although the successful use of this strategy has not yet been reported in bacteria (Kirkpatrick et al., 2005).

RECOMMENDATIONS FOR FUTURE RESEARCH

The field of phosphoproteomics would benefit greatly from an integrated, easily accessible database containing all available information, which would facilitate meta-analysis of PTMs between multiple datasets. Furthermore, such a database for all known PTMs would allow for more in-depth analysis of the cross-talk between bacterial PTMs and how this may relate to pathogenicity, which has been discussed in a review by Soufi et al. (2012). Such a resource could provide a standardized format for reporting PTMs, as well as the opportunity to automatically update database accession numbers for modified proteins— which would in turn facilitate more meaningful comparison between different species or strains. Significant challenges exist prior to the development of such a database, including the lack of standardized reporting in currently published studies. However, the capacity to detect PTMs using MS will only increase with improving technology and the interpretation and management of the data thus generated is therefore crucial if we are to translate this into meaningful clinical applications. Through better understanding of the function of regulatory PTMs in *M. tuberculosis*, we may reveal the means to control or cure it.

AUTHOR CONTRIBUTIONS

CA Wrote the original draft of this paper, read, and edited final draft. BC Edited draft submitted by CA, compiled data from phosphoproteomic reviews and performed meta-data analysis, put together final draft of paper. NS Supervised revisions on paper, provided editing support, and directed the topic to be covered by the review. JB Provided supervision and guidance throughout and offered editing support.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2016.00141

REFERENCES

Bakal, C. J., and Davies, J. E. (2000). No longer an exclusive club: eukaryotic signalling domains in bacteria. *Trends Cell Biol.* 10, 32–38. doi: 10.1016/S0962-8924(99)01681-5

Banerjee, S., Farhana, A., Ehtesham, N. Z., and Hasnain, S. E. (2011). Iron acquisition, assimilation and regulation in mycobacteria. *Infect Genet. Evol.* 11, 825–838. doi: 10.1016/j.meegid.2011.02.016

Be, N. A., Bishai, W. R., and Jain, S. K. (2012). Role of *Mycobacterium tuberculosis* pknD in the pathogenesis of central nervous system tuberculosis. *BMC Microbiol.* 12:7. doi: 10.1186/1471-2180-12-7

Bhatia, V. N., Perlman, D. H., Costello, C. E., and McComb, M. E. (2009). Software tool for researching annotations of proteins: open-source protein annotation software with data visualization. *Anal. Chem.* 81, 9819–9823. doi: 10.1021/ac901335x

Blasco, B., Chen, J. M., Hartkoorn, R., Sala, C., Uplekar, S., Rougemont, J., et al. (2012). Virulence regulator EspR of *Mycobacterium tuberculosis* is a nucleoid-associated protein. *PLoS Pathog.* 8:e1002621. doi: 10.1371/journal.ppat.1002621

Cain, J. A., Solis, N., and Cordwell, S. J. (2014). Beyond gene expression: the impact of protein post-translational modifications in bacteria. *J. Proteomics* 97, 265–286. doi: 10.1016/j.jprot.2013.08.012

Chen, C., Li, Z., Huang, H., Suzek, B. E., Wu, C. H., and UniProt Consortium (2013). A fast peptide match service for uniprot knowledgebase. *Bioinformatics* 29, 2808–2809. doi: 10.1093/bioinformatics/btt484

Chopra, P., Singh, B., Singh, R., Vohra, R., Koul, A., Meena, L. S., et al. (2003). Phosphoprotein phosphatase of *Mycobacterium tuberculosis* dephosphorylates serine-threonine kinases PknA and PknB. *Biochem. Biophys. Res. Commun.* 311, 112–120. doi: 10.1016/j.bbrc.2003.09.173

de la Fuente van Bentem, S., Mentzen, W. I., de la Fuente, A., and Hirt, H. (2008). Towards functional phosphoproteomics by mapping differential phosphorylation events in signaling networks. *Proteomics* 8, 4453–4465. doi: 10.1002/pmic.200800175

Durán, R., Villarino, A., Bellinzoni, M., Wehenkel, A., Fernandez, P., Boitel, B., et al. (2005). Conserved autophosphorylation pattern in activation loops and juxtamembrane regions of *Mycobacterium tuberculosis* Ser/Thr protein kinases. *Biochem. Biophys. Res. Commun.* 333, 858–867. doi: 10.1016/j.bbrc.2005.05.173

Eisenreich, W., Dandekar, T., Heesemann, J., and Goebel, W. (2010). Carbon metabolism of intracellular bacterial pathogens and possible links to virulence. *Nat. Rev. Microbiol.* 8, 401–412. doi: 10.1038/nrmicro2351

Evangelopoulos, D., and McHugh, T. D. (2015). Improving the tuberculosis drug development pipeline. *Chem. Biol. Drug Des.* 86, 951–960. doi: 10.1111/cbdd.12549

Feltcher, M. E., Sullivan, J. T., and Braunstein, M. (2010). Protein export systems of *Mycobacterium tuberculosis*: novel targets for drug development? *Future Microbiol.* 5, 1581–1597. doi: 10.2217/fmb.10.112

Forrellad, M. A., Klepp, L. I., Gioffré, A., Sabio y Garcia, J., Morbidoni, H. R., de la Paz Santangelo, M., et al. (2013). Virulence factors of the *Mycobacterium tuberculosis* complex. *Virulence* 4, 3–66. doi: 10.4161/viru.22329

Fortuin, S., Tomazella, G. G., Nagaraj, N., Sampson, S. L., van Pittius, N. C. G., Soares, N. C., et al. (2015). Phosphoproteomics analysis of a clinical *Mycobacterium tuberculosis* Beijing isolate: expanding the mycobacterial phosphoproteome catalog. *Front. Microbiol.* 6:6. doi: 10.3389/fmicb.2015.00006

Ge, R., and Shan, W. (2011). Bacterial phosphoproteomic analysis reveals the correlation between protein phosphorylation and bacterial pathogenicity. *Genomics. Proteomics Bioinformatics* 9, 119–127. doi: 10.1016/S1672-0229(11)60015-6

Gengenbacher, M., and Kaufmann, S. H. (2012). *Mycobacterium tuberculosis*: success through dormancy. *FEMS Microbiol. Rev.* 36, 514–532. doi: 10.1111/j.1574-6976.2012.00331.x

Ilan, O., Bloch, Y., Frankel, G., Ullrich, H., Geider, K., and Rosenshine, I. (1999). Protein tyrosine kinases in bacterial pathogens are associated with virulence and production of exopolysaccharide. *EMBO J.* 18, 3241–3248. doi: 10.1093/emboj/18.12.3241

Jers, C., Soufi, B., Grangeasse, C., Deutscher, J., and Mijakovic, I. (2008). Phosphoproteomics in bacteria: towards a systemic understanding of bacterial phosphorylation networks. *Expert Rev. Proteomics* 5, 619–627. doi: 10.1586/14789450.5.4.619

Kirkpatrick, D. S., Gerber, S. A., and Gygi, S. P. (2005). The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications. *Methods* 35, 265–273. doi: 10.1016/j.ymeth.2004.08.018

Kochanowski, K., Sauer, U., and Noor, E. (2015). Posttranslational regulation of microbial metabolism. *Curr. Opin. Microbiol.* 27, 10–17. doi: 10.1016/j.mib.2015.05.007

Kusebauch, U., Ortega, C., Ollodart, A., Rogers, R. S., Sherman, D. R., Moritz, R. L., et al. (2014). *Mycobacterium tuberculosis* supports protein tyrosine phosphorylation. *Proc. Natl. Acad. Sci. U.S.A.* 111, 9265–9270. doi: 10.1073/pnas.1323894111

Lau, H.-T., Suh, H. W., Golkowski, M. G., and Ong, S.-E. (2014). Comparing SILAC- and stable isotope dimethyl-labeling approaches for quantitative proteomics. *J. Proteome Res.* 13, 4164–4174. doi: 10.1021/pr500630a

Lee, W., VanderVen, B. C., Fahey, R. J., and Russell, D. G. (2013). Intracellular *Mycobacterium tuberculosis* exploits host-derived fatty acids to limit metabolic stress. *J. Biol. Chem.* 288, 6788–6800. doi: 10.1074/jbc.M112.445056

Macek, B., Gnad, F., Soufi, B., Kumar, C., Olsen, J. V., Mijakovic, I., et al. (2008). Phosphoproteome analysis of *E. coli* reveals evolutionary conservation of bacterial Ser/Thr/Tyr phosphorylation. *Mol. Cell. Proteomics* 7, 299–307. doi: 10.1074/mcp.M700311-MCP200

Macek, B., Mann, M., and Olsen, J. V. (2009). Global and site-specific quantitative phosphoproteomics: principles and applications. *Annu. Rev. Pharmacol. Toxicol.* 49, 199–221. doi: 10.1146/annurev.pharmtox.011008.145606

Macek, B., Mijakovic, I., Olsen, J. V., Gnad, F., Kumar, C., Jensen, P. R., et al. (2007). The serine/threonine/tyrosine phosphoproteome of the model bacterium *Bacillus subtilis*. *Mol. Cell. Proteomics* 6, 697–707. doi: 10.1074/mcp.M600464-MCP200

Mann, M., and Jensen, O. N. (2003). Proteomic analysis of post-translational modifications. *Nat. Biotechnol.* 21, 255–261. doi: 10.1038/nbt0303-255

Mijakovic, I. (2010). Protein phosphorylation in bacteria. *Issues* 5, 21–25.

Misra, S. K., Milohanic, E., Aké, F., Mijakovic, I., Deutscher, J., Monnet, V., et al. (2011). Analysis of the serine/threonine/tyrosine phosphoproteome of the pathogenic bacterium Listeria monocytogenes reveals phosphorylated proteins related to virulence. *Proteomics* 11, 4155–4165. doi: 10.1002/pmic.201100259

Nakedi, K. C., Nel, A. J., Garnett, S., Blackburn, J. M., and Soares, N. C. (2015). Comparative Ser/Thr/Tyr phosphoproteomics between two mycobacterial species: the fast growing *Mycobacterium smegmatis* and the slow growing *Mycobacterium bovis* BCG. *Front. Microbiol.* 6:237. doi: 10.3389/fmicb.2015.00237

Ng, V. H., Cox, J. S., Sousa, A. O., MacMicking, J. D., and McKinney, J. D. (2004). Role of KatG catalase-peroxidase in mycobacterial pathogenesis: countering the phagocyte oxidative burst. *Mol. Microbiol.* 52, 1291–1302. doi: 10.1111/j.1365-2958.2004.04078.x

Nguyen, L., Walburger, A., Houben, E., Koul, A., Muller, S., Morbitzer, M., et al. (2005). Role of protein kinase G in growth and glutamine metabolism of *Mycobacterium bovis* BCG. *J. Bacteriol.* 187, 5852–5856. doi: 10.1128/JB.187.16.5852-5856.2005

Nørregaard Jensen, O. (2004). Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Curr. Opin. Chem. Biol.* 8, 33–41. doi: 10.1016/j.cbpa.2003.12.009

Parandhaman, D. K., Hanna, L. E., and Narayanan, S. (2014a). PknE, a serine/threonine protein kinase of *Mycobacterium tuberculosis* initiates survival crosstalk that also impacts HIV coinfection. *PLoS ONE* 9:e83541. doi: 10.1371/journal.pone.0083541

Parandhaman, D. K., Sharma, P., Bisht, D., and Narayanan, S. (2014b). Proteome and phosphoproteome analysis of the serine/threonine protein kinase E mutant of *Mycobacterium tuberculosis*. *Life Sci.* 109, 116–126. doi: 10.1016/j.lfs.2014.06.013

Prisic, S., Dankwa, S., Schwartz, D., Chou, M. F., Locasale, J. W., Kang, C.-M., et al. (2010). Extensive phosphorylation with overlapping specificity by *Mycobacterium tuberculosis* serine/threonine protein kinases. *Proc. Natl. Acad. Sci. U.S.A.* 107, 7521–7526. doi: 10.1073/pnas.0913482107

Prisic, S., and Husson, R. N. (2014). *Mycobacterium tuberculosis* serine/threonine protein kinases. *Microbiol. Spectr.* 2. doi: 10.1128/microbiolspec.MGM2-0006-2013

Ravikumar, V., Shi, L., Krug, K., Derouiche, A., Jers, C., Cousin, C., et al. (2014). Quantitative phosphoproteome analysis of *Bacillus subtilis* reveals novel

substrates of the kinase PrkC and phosphatase PrpC. *Mol. Cell. Proteomics* 13, 1965–1978. doi: 10.1074/mcp.M113.035949

Sassetti, C. M., Boyd, D. H., and Rubin, E. J. (2003). Genes required for mycobacterial growth defined by high density mutagenesis. *Mol. Microbiol.* 48, 77–84. doi: 10.1046/j.1365-2958.2003.03425.x

Schmidt, A., Trentini, D. B., Spiess, S., Fuhrmann, J., Ammerer, G., Mechtler, K., et al. (2014). Quantitative phosphoproteomics reveals the role of protein arginine phosphorylation in the bacterial stress response. *Mol. Cell. Proteomics* 13, 537–550. doi: 10.1074/mcp.M113.032292

Schnappinger, D., Ehrt, S., Voskuil, M. I., Liu, Y., Mangan, J. A., Monahan, I. M., et al. (2003). Transcriptional adaptation of *Mycobacterium tuberculosis* within macrophages insights into the phagosomal environment. *J. Exp. Med.* 198, 693–704. doi: 10.1084/jem.20030846

Semanjski, M., and Macek, B. (2016). Shotgun proteomics of bacterial pathogens: advances, challenges and clinical implications. *Expert Rev. Proteomics* 13, 139–156. doi: 10.1586/14789450.2016.1132168

Sherman, D. R., and Grundner, C. (2014). Agents of change-concepts in *Mycobacterium tuberculosis* Ser/Thr/Tyr phosphosignalling. *Mol. Microbiol.* 94, 231–241. doi: 10.1111/mmi.12747

Singh, A., Singh, Y., Pine, R., Shi, L., Chandra, R., and Drlica, K. (2006). Protein kinase I of *Mycobacterium tuberculosis*: cellular localization and expression during infection of macrophage-like cells. *Tuberculosis (Edinb)* 86, 28–33. doi: 10.1016/j.tube.2005.04.002

Soares, N. C., Spät, P., Krug, K., and, Macek, B. (2013). Global dynamics of the *Escherichia coli* proteome and phosphoproteome during growth in minimal medium. *J. Proteome Res.* 12, 2611–2621. doi: 10.1021/pr3011843

Soares, N. C., Spät, P., Méndez, J. A., Nakedi, K., Aranda, J., and Bou, G. (2014). Ser/Thr/Tyr phosphoproteome characterization of Acinetobacter baumannii: comparison between a reference strain and a highly invasive multidrug-resistant clinical isolate. *J. Proteomics* 102, 113–124. doi: 10.1016/j.jprot.2014.03.009

Soufi, B., Soares, N. C., Ravikumar, V., and Macek, B. (2012). Proteomics reveals evidence of cross-talk between protein modifications in bacteria: focus on acetylation and phosphorylation. *Curr. Opin. Microbiol.* 15, 357–363. doi: 10.1016/j.mib.2012.05.003

Stülke, J. (2010). More than just activity control: phosphorylation may control all aspects of a protein's properties. *Mol. Microbiol.* 77, 273–275. doi: 10.1111/j.1365-2958.2010.07228.x

Sun, X., Ge, F., Xiao, C.-L., Yin, X.-F., Ge, R., Zhang, L.-H., et al. (2009). Phosphoproteomic analysis reveals the multiple roles of phosphorylation in pathogenic bacterium *Streptococcus pneumoniae*. *J. Proteome Res.* 9, 275–282. doi: 10.1021/pr900612v

Van Els, C. A. C. M., Corbière, V., Smits, K., van Gaans-van den Brink, J. A. M., Poelen, M. C. M., Mascart, F., et al. (2014). Toward Understanding the Essence of Post-Translational Modifications for the *Mycobacterium tuberculosis* Immunoproteome. *Front. Immunol.* 5:361. doi: 10.3389/fimmu.2014.00361

Walburger, A., Koul, A., Ferrari, G., Nguyen, L., Prescianotto-Baschong, C., Huygen, K., et al. (2004). Protein kinase G from pathogenic mycobacteria promotes survival within macrophages. *Science* 304, 1800–1804. doi: 10.1126/science.1099384

Wehenkel, A., Bellinzoni, M., Graña, M., Duran, R., Villarino, A., Fernandez, P., et al. (2008). Mycobacterial Ser/Thr protein kinases and phosphatases: physiological roles and therapeutic potential. *Biochim. Biophys. Acta* 1784, 193–202. doi: 10.1016/j.bbapap.2007.08.006

Zheng, J., Liu, L., Liu, B., and Jin, Q. (2015). Phosphoproteomic analysis of bacillus Calmette-Guérin using gel-based and gel-free approaches. *J. Proteomics* 126, 189–199. doi: 10.1016/j.jprot.2015.06.003

Mass Spectrometry Targeted Assays as a Tool to Improve Our Understanding of Post-translational Modifications in Pathogenic Bacteria

Nelson C. Soares * and *Jonathan M. Blackburn*

Division of Chemical and Systems Biology, Department of Integrative Biomedical Sciences, Faculty of Health Sciences, Institute of Infectious Diseases and Molecular Medicine, University of Cape Town, Cape Town, South Africa

ADVANCES IN MASS SPECTROMETRY-BASED PROTEOMICS IMPACT THE ANALYSIS OF BACTERIAL PTMS

Post translational modifications (PTMs) play a vital role in maintaining protein function, regulation, signaling, and many other important cellular processes (Anonsen et al., 2012). For several years, protein PTMs were thought to be unique to eukaryotes; only recently it has been demonstrated that bacteria also contain a wide variety of PTMs including those previous described in eukaryotic cells, for example Ser/Thr/Tyr phosphorylation, lysine-acetylation, glycosylation, ubiquitination, and glutathionylation, as reviewed (Macek and Mijakovic, 2011; Soufi et al., 2012; Michard and Doublet, 2015; Ravikumar et al., 2015). PTMs have been linked to numerous essential bacterial cellular events, including cell division, morphology, and more recently they have also been associated with: Pathogenicity (Michard and Doublet, 2015; Ravikumar et al., 2015); virulence (Lin et al., 2009; Calder et al., 2015; Fortuin et al., 2015); and drug resistance (Soares et al., 2014; Lai et al., 2016). The employment of powerful new proteomics platforms has contributed to significant insight in to the regulatory functions of bacterial PTMs. More than a decade ago, the first attempts to catalog bacterial phosphoproteomes were performed using 2D gel electrophoresis-based (2-DE) approaches, first combined with ^{33}P radiolabelling and later with phosphoprotein-specific dyes (Cortay et al., 1986; Bendt et al., 2003; Lévine et al., 2006). However, such studies were limited by a number of technical challenges, including those inherent to gel based approaches (Pietrograde et al., 2003; Monteoliva and Albar, 2004). Consequently, at the time few phosphoproteins and even fewer phosphorylation sites were reported. As the field of proteomics progressed from 2-DE gel based methodologies to modern liquid chromatography mass-spectrometry (LC-MS/MS) based workflows, significant advances in the analysis of bacterial PTMs were then achieved. The introduction of modern mass spectrometry (MS), specifically the inclusion of Orbitrap mass spectrometers, brought new momentum to the field. In 2006, Mann and colleagues proposed a gel-free approach to explore the phosphoproteome of eukaryotic cells in depth. The study reported the identification of more than 6600 phosphorylation sites (Olsen et al., 2006). Their workflow involved the enzymatic digestion of complex protein extracts into peptides and then phosphopeptides were enriched by a combination of strong cation exchange and TiO$_2$ chromatography. Macek et al. (2007) used similar technology to describe a site–specific, *in vivo* phosphoproteome of a gram positive model organism, *Bacillus subtilis*. The study represented a main breakthrough in the bacterial phosphoproteomic field: the number of reported *B. subtilis* phosphoproteins went from 16 phosphorylation sites on eight proteins to 103 unique phosphopetides on 78 proteins (Macek et al., 2007). This order of magnitude jump in the field

then inspired many other subsequent studies that successfully applied variants of the protocol described by Macek et al. (2007), in particular those that employed phosphopetide enrichment via immobilized metal affinity chromatography (IMAC; Thingholm and Jensen, 2009) to perform high throughput analysis of different bacterial phosphoproteomes (Ravichandran et al., 2009; Qu et al., 2013). Today, several bacterial phosphoproteomes have been successfully characterized including some important human pathogenic bacteria: *Pseudomonas aeruginosa* (Ravichandran et al., 2009); *Klebsiella pneumoniae* (Lin et al., 2009); mycobacterium spp. (Prisic et al., 2010; Fortuin et al., 2015; Nakedi et al., 2015; Calder et al., 2016); *Helicobacter pylori* (Ge et al., 2011), streptomyces spp. (Manteca et al., 2011); *Acinectobacter baummnii* (Soares et al., 2014; Lai et al., 2016); *Staphylococcus aureus* (Basell et al., 2014); and Enteropathogenic *Escherichia coli* (Scholz et al., 2015). The development of new high throughput mass spectrometric instrumentation (Michalski et al., 2011b, 2012; Scheltema et al., 2014; Erickson et al., 2015) together with continuous optimization of LC-MS/MS shotgun proteomics workflows (Michalski et al., 2011a; Kelstrup et al., 2012; Pirmoradian et al., 2013) resulted in increased resolution, sensitivity, and fragmentation and has contributed to the successful identification and localisation of phosphosites. Such approaches have also now allowed for the identification of other bacterial PTMs such as lysine acetylation (Zhang et al., 2009; Weinert et al., 2013; Wu et al., 2013; Liu et al., 2014; Ouidir et al., 2015; Xie et al., 2015), ubiquitination (Valkevich et al., 2014), and glycosylation (Scott et al., 2011; Anonsen et al., 2012; Smith et al., 2014). It is therefore well established that large scale LC-MS/MS shotgun proteomics is the technique of choice for identifying bacterial post-translationally modified peptides/proteins.

When quantifying the state of a modified peptide/protein under a certain experimental condition(s) and/or during different cellular processes, it is important to determine its regulatory function and ultimately link this to a specific molecular mechanism. Recent advances in quantitative LC-MS/MS based proteomics have made possible the reliable quantification of thousands of phosphorylation peptides in eukaryotic cells (Pan et al., 2008) and, likewise, a number of studies have successfully applied different quantification methods to describe the dynamics of bacterial phosphoproteomes under different cellular states or experimental conditions. For instance, stable isotope labeling by amino acids in cell culture (SILAC) has recently been applied to establish the dynamics of both the proteome and phosphoproteome of *E. coli* during five different growth phases in liquid cell culture (Soares et al., 2013). In that study, 76 Ser/Thr/Tyr phosphorylation events were quantified in all growth phases. Importantly the use of SILAC enabled the measurement of median occupancies of phosphorylation sites, which, unlike eukaryotic cells (Pan et al., 2008), were generally low (<12%) (Soares et al., 2013). SILAC has been successfully used on a number of occasions to determine the varying levels of Ser/Thr/Tyr phosphorylation events. However, the issues associated with SILAC when applied to bacteria are well documented (Soufi and Macek, 2014) and include, in particular, cases such as mycobacteria where there is no available viable, lysine-deficient mutant. Alternatively, chemical labeling,

such as peptide demethylation labeling (Boersema et al., 2009), has been proven a convenient and efficient method to obtain differential quantification of bacterial PTMs (Spat et al., 2015; De Keijzer et al., 2016) and has gained increasing popularity. Additionally, an always attractive quantitative approach is label free quantification (LFQ), which has been applied on a number of occasions to further investigate the different PTMs levels in bacteria. Within this context, although phospho-enrichment strategies combined with LFQ analyses look promising, it has been pointed out (Rosenberg et al., 2015) that low occupancy among bacterial phosphorylation events may compound possible inconsistences amongst biological replicates.

MS based detection of PTMs is normally based on a mass shift introduced by the site-specific modification, however these are often sub-stoichiometric and usually occur at low abundance, thus an enrichment process (such as those mentioned above) is typically carried out for a specific PTM prior to subsequent MS analysis. The enrichment steps are however, for many, a major drawback in bacterial PTM analyses, since the enrichment step typically requires large amounts of starting material including protein concentrations in the order of 5–9 mg (Soufi et al., 2008; Soares et al., 2013, 2014; Spat et al., 2015). While this is can be easily achieved when using liquid cell culture, this may represent a major technical challenge for other studies, such those looking at host pathogen interactions at the site of infection, where it is virtually impossible to isolate the large number of bacterial cells or the physical amount of protein needed in those protocols. Additionally, the multiple rounds of enrichment steps remain a source of possible technical variations that can bias downstream quantitative analyses. Ultimately, the efficiency and reproducibility of the enrichment protocol across multiple samples is determinant both in terms of coverage as well as in terms of accurate quantification.

MS BASED TARGETED METHODOLOGIES WILL ALLOW DETECTION AND QUANTIFICATION OF LOW ABUNDANT BACTERIAL PTMS

Emerging targeted MS methodologies, namely selected reaction monitoring (SRM), and parallel reaction monitoring (PRM), appear as perhaps the latest breakthrough within the field. Specifically, targeted mass spectrometry offers unequaled capability to characterize and quantify a specific set of proteins/peptides reproducibility, in any biological sample. In this hypothesis-driven approach (Picotti et al., 2013), only a small number of peptides are used as surrogate markers for the protein of interest, which are selectively measured in predefined m/z ranges and retention time windows. SRM (often called MRM, multiple reaction monitoring) is a targeted approach to quantitate pre-selected peptides by monitoring specific precursor-to-product ion transitions in a triple-quadrupole mass spectrometer (Picotti et al., 2010, 2013). While SRM measurements use a low resolution MS instrument, PRMs alternatively benefit from the capabilities of high resolution quadrupole and Orbitrap analysers (Peterson et al., 2012). The

increased quality of hybrid quadrupole-Orbitrap analyses has consequently improved the quality of the target methods, mainly in terms of accuracy at the level of fragment, which enables the fundamental distinction of target compounds from the undesired background ions and enables a significantly lowered detection limit. In mammalian systems, SRM assays have been successfully applied in a large scale targeted proteomics experiment of a phosphorylation network. In the first of its kind study, Wolf-Yadlin et al. (2007) applied SRM to monitor the dynamics over time of tyrosine phosphorylation events in EGFR signaling after EGF stimulation. The authors used a combination of iTRAQ peptide labeling and phosphotyrosine peptide immune-purification coupled with IMAC to enrich for phosphotyrosine containing peptides (Wolf-Yadlin et al., 2007). More recently, Adachi et al. (2016) described the application of a large-scale phosphoproteome analysis and SRM-based quantification to develop a strategy for the systematic discovery and validation of biomarkers. Their two step approach described uses a typical IMAC based phosphoenrichment coupled to shotgun MS-based proteomic analysis in order to identify differentially modulate phosphopeptides; identified candidates are then validated by SRM analysis (Adachi et al., 2016). In another recent, well-designed application, Altelaar and co-workers suggested a strategy to monitor signal transduction pathways: By combining targeted quantitative proteomics with high selective phosphopeptide enrichment, the authors monitored the phosphorylation dynamics of the PI3K-mTOR and MAPK signaling network (De Graaf et al., 2015). Interestingly, in order to increase the success rate of the phosphopeptide SRM assays the authors focused on phosphopeptides reported previously by the group and from publicly available shotgun proteomics experimental data. The rationale for this approach is that phosphorylation alters the local charge distribution, which may interfere with proteolytic cleavage, thus previous knowledge of the phosphosite would enable a more efficient SRM assay. The study used synthetic stable isotope standard phosphopeptides for increased accuracy in quantification (De Graaf et al., 2015). Despite all this, measuring PTMs events through SRM can be challenging and in most cases involves time consuming assay optimization and reliance on synthetic peptide standards (Lawrence et al., 2016), which can be somewhat discouraging in practice. However, recently Lawrence et al. have proposed the use of PRMs as a straightforward and efficient means for large-scale targeted phosphoproteomic analysis. Here, instead of using databases of previously reported human phosphopeptide sequences, the authors report a highly comprehensive shotgun global phosphoproteomic analysis that resulted in the identification of more than 7.5 million phosphopeptide spectral matches corresponding to 109,611 phosphorylation sites (Lawrence et al., 2016). Subsequently, the authors used the generated information, including the respective phosphopeptides retention times, to develop a plug-and-play phosphoproteomic system, based on sets of targeted, label-free phosphopeptide PRM assays in order to interrogate the dynamics of the IGF1/AKT signaling pathway. Importantly, the study demonstrated that without sample fractionation and or phosphopeptide enrichment (Lawrence et al., 2016), PRM

label-free quantification assays is a rapid assay for measuring virtually any known phosphorylation event in the human species.

Targeted methodologies have been successfully applied to investigate aspects of the pathogenic bacteria proteome (Lange et al., 2008; Karlsson et al., 2012; Schubert and Aebersold, 2015; Schubert et al., 2015; Peters et al., 2016). A noteworthy highlight is that Ruedi Aebersold and colleagues have recently generated a library of targeted SRM assays for ∼97% of the 4012 annotated *Mycobacterium tuberculosis* proteins and were able to reproducibly quantify ∼72% of the theoretical *M. tuberculosis* proteome in single unfractionated runs on a triple quadrupole MS (Schubert et al., 2013). As such, the generated *M. tuberculosis* proteome library represents a valuable experimental resource that now in theory enables researchers to interrogate and quantify essentially the entire proteome of *M. tuberculosis* in a single experiment. This allows for greater understanding of, for example, differences in expressed pathogenicity and virulence between clinical isolates. However, in order to further investigate mycobacterial or any other bacterial PTMs in a large scale, targeted and quantitative manner, it will be necessary first to follow a workflow similar to that used in mammalian cells as suggested by Lawrence et al. (2016). A strategy based on previous knowledge of the post translationally modified proteins/peptide is sensitive enough to detect low abundance, modified peptides. To date, much of the work on pathogenic bacteria using MS-based proteomics has focused on cataloging, and in some cases quantifying, PTMs that occur under *in vitro* conditions. Although informative, such reports come with the caveat of the studied conditions often being detached from the unique microenvironment that pathogenic bacteria are exposed to during host infection and/ or during colonization (e.g., inside the macrophage). Thus, in order to gain meaningful insights into the dynamics and role of different bacterial PTMs during host-pathogen interactions, it is crucial that researchers in the field join efforts to once again take advantage of recent advances in mass spectrometry instrumentation and develop targeted, quantitative PTM workflows that enable detection and accurate quantification of low-abundance, modified analytes, ideally circumventing the need for additional PTM enrichment steps in the process.

AUTHOR CONTRIBUTIONS

NS wrote the original draft of this paper read, and edit final draft. JB critically discussed the content of the manuscript and provided editing support.

FUNDING

REFERENCES

Adachi, J., Narumi, R., and Tomonaga, T. (2016). "Targeted phosphoproteome analysis using selected/multiple reaction monitoring (SRM/MRM)," in *Proteomics in Systems Biology: Methods and Protocols*, ed J. Reinders (New York, NY: Springer), 87–100.

Anonsen, J. H., Vik, Å., Egge-Jacobsen, W., and Koomey, M. (2012). An extended spectrum of target proteins and modification sites in the general o-linked protein glycosylation system in *Neisseria gonorrhoeae*. *J. Proteome Res.* 11, 5781–5793. doi: 10.1021/pr300584x

Basell, K., Otto, A., Junker, S., Zuhlke, D., Rappen, G. M., Schmidt, S., et al. (2014). The phosphoproteome and its physiological dynamics in *Staphylococcus aureus*. *Int. J. Med. Microbiol.* 304, 121–132. doi: 10.1016/j.ijmm.2013.11.020

Bendt, A. K., Burkovski, A., Schaffer, S., Bott, M., Farwick, M., and Hermann, T. (2003). Towards a phosphoproteome map of *Corynebacterium glutamicum*. *Proteomics* 3, 1637–1646. doi: 10.1002/pmic.200300494

Boersema, P. J., Raijmakers, R., Lemeer, S., Mohammed, S., and Heck, A. J. R. (2009). Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat. Protoc.* 4, 484–494. doi: 10.1038/nprot.2009.21

Calder, B., Albeldas, C., Blackburn, J. M., and Soares, N. C. (2016). Mass spectrometry offers insight into the role of Ser/Thr/Tyr phosphorylation in the mycobacteria. *Front. Microbiol.* 7:141. doi: 10.3389/fmicb.2016.00141

Calder, B., Soares, N. C., De Kock, E., and Blackburn, J. M. (2015). Mycobacterial proteomics: analysis of expressed proteomes and post-translational modifications to identify candidate virulence factors. *Expert Rev. Proteomics* 12, 21–35. doi: 10.1586/14789450.2015.1007046

Cortay, J.-C., Rieul, C., Duclos, B., and Cozzone, A. J. (1986). Characterization of the phosphoproteins of *Escherichia coli* cells by electrophoretic analysis. *Eur. J. Biochem.* 159, 227–237. doi: 10.1111/j.1432-1033.1986.tb09858.x

De Graaf, E. L., Kaplon, J., Mohammed, S., Vereijken, L. A. M., Duarte, D. P., Redondo Gallego, L., et al. (2015). Signal transduction reaction monitoring deciphers site-specific PI3K-mTOR/MAPK pathway dynamics in oncogene-induced senescence. *J. Proteome Res.* 14, 2906–2914. doi: 10.1021/acs.jproteome.5b00236

De Keijzer, J., Mulder, A., De Beer, J., De Ru, A. H., Van Veelen, P. A., and Van Soolingen, D. (2016). Mechanisms of phenotypic rifampicin tolerance in *Mycobacterium tuberculosis* Beijing genotype strain B0/W148 revealed by proteomics. *J. Proteome Res.* 15, 1194–1204. doi: 10.1021/acs.jproteome.5b01073

Erickson, B. K., Jedrychowski, M. P., McAlister, G. C., Everley, R. A., Kunz, R., and Gygi, S. P. (2015). Evaluating multiplexed quantitative phosphopeptide analysis on a hybrid quadrupole mass filter/linear ion trap/orbitrap mass spectrometer. *Anal. Chem.* 87, 1241–1249. doi: 10.1021/ac503934f

Fortuin, S., Tomazella, G. G., Nagaraj, N., Sampson, S. L., Gey Van Pittius, N. C., Soares, N. C., et al. (2015). Phosphoproteomics analysis of a clinical *Mycobacterium tuberculosis* Beijing isolate: expanding the mycobacterial phosphoproteome catalog. *Front. Microbiol.* 6:6. doi: 10.3389/fmicb.2015.00006

Ge, R., Sun, X., Xiao, C., Yin, X., Shan, W., Chen, Z., et al. (2011). Phosphoproteome analysis of the pathogenic bacterium *Helicobacter pylori* reveals over-representation of tyrosine phosphorylation and multiply phosphorylated proteins. *Proteomics* 11, 1449–1461. doi: 10.1002/pmic.201000649

Karlsson, C., Malmström, L., Aebersold, R., and Malmström, J. (2012). Proteome-wide selected reaction monitoring assays for the human pathogen Streptococcus pyogenes. *Nat. Commun.* 3, 1301. doi: 10.1038/ncomms2297

Kelstrup, C. D., Young, C., Lavallee, R., Nielsen, M. L., and Olsen, J. V. (2012). Optimized fast and sensitive acquisition methods for shotgun proteomics on a quadrupole orbitrap mass spectrometer. *J. Proteome Res.* 11, 3487–3497. doi: 10.1021/pr3000249

Lai, J.-H., Yang, J.-T., Chern, J., Chen, T.-L., Wu, W.-L., Liao, J.-H., et al. (2016). Comparative phosphoproteomics reveals the role of AmpC β-lactamase phosphorylation in the clinical imipenem-resistant strain *Acinetobacter baumannii* SK17. *Mol. Cell. Proteomics* 15, 12–25. doi: 10.1074/mcp.M115.051052

Lange, V., Malmström, J. A., Didion, J., King, N. L., Johansson, B. P., Schäfer, J., et al. (2008). Targeted quantitative analysis of Streptococcus pyogenes virulence factors by multiple reaction monitoring. *Mol. Cell. Proteomics* 7, 1489–1500. doi: 10.1074/mcp.M800032-MCP200

Lawrence, R. T., Searle, B. C., Llovet, A., and Villen, J. (2016). Plug-and-play analysis of the human phosphoproteome by targeted high-resolution mass spectrometry. *Nat. Methods* 13, 431–434. doi: 10.1038/nmeth.3811

Lévine, A., Vannier, F., Absalon, C., Kuhn, L., Jackson, P., Scrivener, E., et al. (2006). Analysis of the dynamic *Bacillus subtilis* Ser/Thr/Tyr phosphoproteome implicated in a wide variety of cellular processes. *Proteomics* 6, 2157–2173. doi: 10.1002/pmic.200500352

Lin, M.-H., Hsu, T.-L., Lin, S.-Y., Pan, Y.-J., Jan, J.-T., Wang, J.-T., et al. (2009). Phosphoproteomics of *Klebsiella pneumoniae* NTUH-K2044 reveals a tight link between tyrosine phosphorylation and virulence. *Mol. Cell. Proteomics* 8, 2613–2623. doi: 10.1074/mcp.M900276-MCP200

Liu, F., Yang, M., Wang, X., Yang, S., Gu, J., Zhou, J., et al. (2014). Acetylome analysis reveals diverse functions of lysine acetylation in *Mycobacterium tuberculosis*. *Mol. Cell. Proteomics* 13, 3352–3366. doi: 10.1074/mcp.M114.041962

Macek, B., and Mijakovic, I. (2011). Site-specific analysis of bacterial phosphoproteomes. *Proteomics* 11, 3002–3011. doi: 10.1002/pmic.201100012

Macek, B., Mijakovic, I., Olsen, J. V., Gnad, F., Kumar, C., Jensen, P. R., et al. (2007). The serine/threonine/tyrosine phosphoproteome of the model bacterium *Bacillus subtilis*. *Mol. Cell. Proteomics* 6, 697–707. doi: 10.1074/mcp.M600464-MCP200

Manteca, A., Ye, J., Sanchez, J., and Jensen, O. N. (2011). Phosphoproteome analysis of Streptomyces development reveals extensive protein phosphorylation accompanying bacterial differentiation. *J. Proteome Res.* 10, 5481–5492. doi: 10.1021/pr200762y

Michalski, A., Cox, J., and Mann, M. (2011a). More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *J. Proteome Res.* 10, 1785–1793. doi: 10.1021/pr101060v

Michalski, A., Damoc, E., Hauschild, J.-P., Lange, O., Wieghaus, A., Makarov, A., et al. (2011b). Mass spectrometry-based proteomics using Q exactive, a high-performance benchtop quadrupole orbitrap mass spectrometer. *Mol. Cell. Proteomics* 10:M111.011015. doi: 10.1074/mcp.m111.011015

Michalski, A., Damoc, E., Lange, O., Denisov, E., Nolting, D., Müller, M., et al. (2012). Ultra high resolution linear ion trap orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes. *Mol. Cell. Proteomics* 11:O111.013698. doi: 10.1074/mcp.O111.013698

Michard, C., and Doublet, P. (2015). Post-translational modifications are key players of the *Legionella pneumophila* infection strategy. *Front. Microbiol.* 6:87. doi: 10.3389/fmicb.2015.00087

Monteoliva, L., and Albar, J. P. (2004). Differential proteomics: an overview of gel and non-gel based approaches. *Brief. Funct. Genomic. Proteomic.* 3, 220–239. doi: 10.1093/bfgp/3.3.220

Nakedi, K. C., Nel, A. J., Garnett, S., Blackburn, J. M., and Soares, N. C. (2015). Comparative Ser/Thr/Tyr phosphoproteomics between two mycobacterial species: the fast growing *Mycobacterium smegmatis* and slow growing *Mycobacterium bovis* BCG. *Front. Microbiol.* 6:237. doi: 10.3389/fmicb.2015.00237

Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., et al. (2006). Global, *in vivo*, and site-specific phosphorylation dynamics in signaling networks. *Cell* 127, 635–648. doi: 10.1016/j.cell.2006.09.026

Ouidir, T., Cosette, P., Jouenne, T., and Hardouin, J. (2015). Proteomic profiling of lysine acetylation in *Pseudomonas aeruginosa* reveals the diversity of acetylated proteins. *Proteomics* 15, 2152–2157. doi: 10.1002/pmic.201500056

Pan, C., Gnad, F., Olsen, J. V., and Mann, M. (2008). Quantitative phosphoproteome analysis of a mouse liver cell line reveals specificity of phosphatase inhibitors. *Proteomics* 8, 4534–4546. doi: 10.1002/pmic.200800105

Peters, J., Calder, B., Gonnelli, G., Degroeve, S., Rajaonarifara, E., Mulder, N., et al. (2016). Identification of quantitative proteomic differences between *Mycobacterium tuberculosis* lineages with altered virulence. *Front. Microbiol.* 7:813. doi: 10.3389/fmicb.2016.00813

Peterson, A. C., Russell, J. D., Bailey, D. J., Westphall, M. S., and Coon, J. J. (2012). Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Mol. Cell. Proteomics* 11, 1475–1488. doi: 10.1074/mcp.O112.020131

Picotti, P., Bodenmiller, B., and Aebersold, R. (2013). Proteomics meets the scientific method. *Nat. Methods* 10, 24–27. doi: 10.1038/nmeth.2291

Picotti, P., Rinner, O., Stallmach, R., Dautel, F., Farrah, T., Domon, B., et al. (2010). High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. *Nat. Methods* 7, 43–46. doi: 10.1038/nmeth.1408

Pietrograne, M. C., Marchetti, N., Dondi, F., and Righetti, P. G. (2003). Spot overlapping in two-dimensional polyacrylamide gel electrophoresis maps: relevance to proteomics. *Electrophoresis* 24, 217–224. doi: 10.1002/elps.200390018

Pirmoradian, M., Budamgunta, H., Chingin, K., Zhang, B., Astorga-Wells, J., and Zubarev, R. A. (2013). Rapid and deep human proteome analysis by single-dimension shotgun proteomics. *Mol. Cell. Proteomics* 12, 3330–3338. doi: 10.1074/mcp.O113.028787

Prisic, S., Dankwa, S., Schwartz, D., Chou, M. F., Locasale, J. W., Kang, C.-M., et al., (2010). Extensive phosphorylation with overlapping specificity by *Mycobacterium tuberculosis* serine/threonine protein kinase. *Proc. Natl. Acad. Sci. U.S.A.* 107, 7521–7526. doi: 10.1073/pnas.0913482107

Qu, Y., Wu, S., Zhao, R., Zink, E., Orton, D. J., Moore, R. J., et al. (2013). Automated immobilized metal affinity chromatography system for enrichment of *Escherichia coli* phosphoproteome. *Electrophoresis* 34, 1619–1626. doi: 10.1002/elps.201200628

Ravichandran, A., Sugiyama, N., Tomita, M., Swarup, S., and Ishihama, Y. (2009). Ser/Thr/Tyr phosphoproteome analysis of pathogenic and non-pathogenic Pseudomonas species. *Proteomics* 9, 2764–2775. doi: 10.1002/pmic.200800655

Ravikumar, V., Jers, C., and Mijakovic, I. (2015). Elucidating host-pathogen interactions based on post-translational modifications using proteomics approaches. *Front. Microbiol.* 6:1313. doi: 10.3389/fmicb.2015.01312

Rosenberg, A., Soufi, B., Ravikumar, V., Soares, N. C., Krug, K., Smith, Y., et al. (2015). Phosphoproteome dynamics mediate revival of bacterial spores. *BMC Biol.* 13:76. doi: 10.1186/s12915-015-0184-7

Scheltema, R. A., Hauschild, J.-P., Lange, O., Hornburg, D., Denisov, E., Damoc, E., et al. (2014). The Q exactive HF, a benchtop mass spectrometer with a pre-filter, high-performance quadrupole and an ultra-high-field orbitrap analyzer. *Mol. Cell. Proteomics* 13, 3698–3708. doi: 10.1074/mcp.M114.043489

Scholz, R., Imami, K., Scott, N. E., Trimble, W. S., Foster, L. J., and Finlay, B. B. (2015). Novel host proteins and signaling pathways in enteropathogenic *E. coli* pathogenesis identified by global phosphoproteome analysis. *Mol. Cell. Proteomics* 14, 1927–1945. doi: 10.1074/mcp.M114.046847

Schubert, O. T., and Aebersold, R. (2015). "Microbial proteome profiling and systems biology: applications to *Mycobacterium tuberculosis*," in *Prokaryotic Systems Biology,* eds P. J. N. Krogan and P. M. Babu (Cham: Springer International Publishing), 235–254.

Schubert, O. T., Ludwig, C., Kogadeeva, M., Zimmermann, M., Rosenberger, G., Gengenbacher, M., et al. (2015). Absolute proteome composition and dynamics during dormancy and resuscitation of *Mycobacterium tuberculosis*. *Cell Host Microbe* 18, 96–108. doi: 10.1016/j.chom.2015.06.001

Schubert, O. T., Mouritsen, J., Ludwig, C., Rost, H. L., Rosenberger, G., Arthur, P. K., et al. (2013). The Mtb proteome library: a resource of assays to quantify the complete proteome of *Mycobacterium tuberculosis*. *Cell Host Microbe* 13, 602–612. doi: 10.1016/j.chom.2013.04.008

Scott, N. E., Parker, B. L., Connolly, A. M., Paulech, J., Edwards, A. V. G., Crossett, B., et al. (2011). Simultaneous glycan-peptide characterization using hydrophilic interaction chromatography and parallel fragmentation by CID, higher energy collisional dissociation, and electron transfer dissociation MS applied to the N-linked glycoproteome of *Campylobacter jejuni*. *Mol. Cell. Proteomics* 10, M000031-MCP201. doi: 10.1074/mcp.M000031-MCP201

Smith, G. T., Sweredoski, M. J., and Hess, S. (2014). O-linked glycosylation sites profiling in *Mycobacterium tuberculosis* culture filtrate proteins. *J. Proteomics* 97, 296–306. doi: 10.1016/j.jprot.2013.05.011

Soares, N. C., Spat, P., Krug, K., and Macek, B. (2013). Global dynamics of the *Escherichia coli* proteome and phosphoproteome during growth in minimal medium. *J. Proteome Res.* 12, 2611–2621. doi: 10.1021/pr3011843

Soares, N. C., Spat, P., Mendez, J. A., Nakedi, K., Aranda, J., and Bou, G. (2014). Ser/Thr/Tyr phosphoproteome characterization of *Acinetobacter baumannii*: comparison between a reference strain and a highly invasive multidrug-resistant clinical isolate. *J. Proteomics* 102, 113–124. doi: 10.1016/j.jprot.2014.03.009

Soufi, B., Gnad, F., Jensen, P. R., Petranovic, D., Mann, M., Mijakovic, I., et al. (2008). The Ser/Thr/Tyr phosphoproteome of *Lactococcus lactis* IL1403 reveals multiply phosphorylated proteins. *Proteomics* 8, 3486–3493. doi: 10.1002/pmic.200800069

Soufi, B., and Macek, B. (2014). "Stable isotope labeling by amino acids applied to bacterial cell culture," in *Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC): Methods and Protocols*, ed B. Warscheid (New York, NY: Springer), 9–22.

Soufi, B., Soares, N. C., Ravikumar, V., and Macek, B. (2012). Proteomics reveals evidence of cross-talk between protein modifications in bacteria: focus on acetylation and phosphorylation. *Curr. Opin. Microbiol.* 15, 357–363. doi: 10.1016/j.mib.2012.05.003

Spat, P., Macek, B., and Forchhammer, K. (2015). Phosphoproteome of the cyanobacterium Synechocystis sp. PCC 6803 and its dynamics during nitrogen starvation. *Front. Microbiol.* 6:248. doi: 10.3389/fmicb.2015.00248

Thingholm, T. E., and Jensen, O. N. (2009). "Enrichment and characterization of phosphopeptides by immobilized metal affinity chromatography (IMAC) and mass spectrometry," in *Phospho-Proteomics: Methods and Protocols,* ed M. D. Graauw (Totowa, NJ: Humana Press), 47–56.

Valkevich, E. M., Sanchez, N. A., Ge, Y., and Strieter, E. R. (2014). Middle-down mass spectrometry enables characterization of branched ubiquitin chains. *Biochemistry* 53, 4979–4989. doi: 10.1021/bi5006305

Weinert, B. T., Iesmantavicius, V., Wagner, S. A., Scholz, C., Gummesson, B., Beli, P., et al. (2013). Acetyl-phosphate is a critical determinant of lysine acetylation in *E. coli*. *Mol. Cell* 51, 265–272. doi: 10.1016/j.molcel.2013.06.003

Wolf-Yadlin, A., Hautaniemi, S., Lauffenburger, D. A., and White, F. M. (2007). Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proc. Natl. Acad. Sci. U.S.A.* 104, 5860–5865. doi: 10.1073/pnas.0608638104

Wu, X., Vellaichamy, A., Wang, D., Zamdborg, L., Kelleher, N. L., Huber, S. C., et al. (2013). Differential lysine acetylation profiles of *Erwinia amylovora* strains revealed by proteomics. *J. Proteomics* 79, 60–71. doi: 10.1016/j.jprot.2012.12.001

Xie, L., Wang, X., Zeng, J., Zhou, M., Duan, X., Li, Q., et al. (2015). Proteome-wide lysine acetylation profiling of the human pathogen *Mycobacterium tuberculosis*. *Int. J. Biochem. Cell Biol.* 59, 193–202. doi: 10.1016/j.biocel.2014.11.010

Zhang, J., Sprung, R., Pei, J., Tan, X., Kim, S., Zhu, H., et al. (2009). Lysine acetylation is a highly abundant and evolutionarily conserved modification in *Escherichia coli*. *Mol. Cell. Proteomics* 8, 215–225. doi: 10.1074/mcp.M800187-MCP200

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read, for greatest visibility

COLLABORATIVE PEER-REVIEW
Designed to be rigorous – yet also collaborative, fair and constructive

FAST PUBLICATION
Average 85 days from submission to publication (across all journals)

COPYRIGHT TO AUTHORS
No limit to article distribution and re-use

TRANSPARENT
Editors and reviewers acknowledged by name on published articles

SUPPORT
By our Swiss-based editorial team

IMPACT METRICS
Advanced metrics track your article's impact

GLOBAL SPREAD
5'100'000+ monthly article views and downloads

LOOP RESEARCH NETWORK
Our network increases readership for your article

Find us on