EXPERT REVIEWS

Proteomic technology in the design of new effective antibacterial vaccines

Expert Rev. Proteomics 6(3), 315–330 (2009)

Elżbieta Katarzyna Jagusztyn-Krynicka†, Michał Dadlez, Anna Grabowska* and Paula Roszczenko*

†*Author for correspondence*
Department of Bacterial Genetics, Faculty of Biology, University of Warsaw, Miecznikowa 1, 02-096 Warsaw, Poland
Tel.: +48 225 541 216
Fax: +48 225 541 402
kjkryn@biol.uw.edu.pl

**Both authors contributed equally to this article.*

Infectious diseases still remain the main cause of human premature deaths, especially in developing countries. Vaccines constitute the most cost-effective tool for prophylaxis of infectious diseases. Elucidation of the complete genomes of many bacterial pathogens has provided a new blueprint for the search of novel vaccine candidates. At the same time, it was a turning point in the development of transcriptomics and proteomics. This article concentrates on the proteomic contribution to vaccinology, pointing out relationships between genomic, transcriptomic and proteomic approaches and describing how they complement one another. It also highlights the recent proteomic techniques applied to antigen identification, their capabilities and limitations, as well as the strategies that are taken to overcome technical difficulties and to refine applied methods. Finally, some recent experimental data concerning the proteomic/immunoproteomic influence on identification of vaccine candidates to prevent human infections caused by *Streptococcus* spp., as well as by a major bioterrorist agent, *Bacillus anthracis* is presented.

Keywords: antigen • *Bacillus anthracis* • immunoproteomic • proteomic • *Streptococcus* spp. • subunit vaccine • vaccinology

Infectious diseases remain a significant global socioeconomic and public-health problem, having a great impact on the development of many nations. According to data reported by the WHO, infectious diseases are globally responsible for more than 15 million premature deaths every year. The most cost-effective strategy to decrease the human mortality and morbidity due to infectious diseases is implementation of effective immunoprophylactic methods. Despite the enormous progress that has been made recently in our understanding of the molecular bases of bacterial pathogenesis, we still do not have effective and safe vaccines against many infectious diseases. On the other hand, some effective vaccines are not readily available, mainly due to economic reasons. Thus, the fundamental scientific need presently is to improve existing vaccines or create new ones.

Existing active antibacterial vaccines can be classified into three main categories: killed, attenuated and subunit vaccines. The former two classes are based on the whole cells of the pathogen, while subunit vaccines based on selected antigens and/or their respective genes are recognized as the safest type, although their efficacy is lower than those comprised of whole attenuated cells. The bottleneck in the development of effective subunit vaccines is the choice of the antigens. A proper candidate for immunization must possess a wide range of different properties: extracytoplasmic localization, abundant presence in the cell, conservation among different pathogen serotypes/genotypes. Additionally, the gene encoding relevant to vaccination antigen has to be expressed *in vivo* during infection, when the pathogen is present within host organism.

For more than 100 years, diverse strategies have been applied to select candidates for subunit vaccine construction. In the beginning of 20th Century, bacterial virulence factors were identified, mainly by biochemical approaches. In the second half of 20th Century, progress in the field of vaccinology was possible owing to application of new recombinant DNA technologies. Another step forward in microbiology and medicine took place at the end of 20th Century in 1995, when the first bacterial genome (*Haemophilus influenzae*) was sequenced. Presently, the genetic information of approximately 2000 bacterial strains is available, including many strains of major human pathogens [201,202]. Development of rapid sequencing methods combined with novel bioinformatics

 10.1586/EPR.09.47 ISSN 1478-9450

strategies has revolutionized the study of bacterial pathogenesis and facilitated the development of a new strategy for identifying of novel antigens, termed 'reversed vaccinology' [1]. Sequencing and comparing multiple bacterial genomes documented the enormous genetic variability within single species and raised questions concerning the species gene pool. In the case of some pathogens, every sequenced genome augments the species gene pool by at least several dozen new genes. Thus, this kind of bacterial genome, known as 'open pangenome', is comprised of core genes, present in every strain, and dispensable genes, present in more than one genome, as well as genes unique for a single isolate [2–4]. It was proven that, for some pathogens possessing open pangenomes, only multiple genome/proteome screening and analysis of dispensable genes can guarantee successful protective antigen identification [5].

Proteomics, large-scale analysis of cellular proteins, is a powerful tool for studying protein identification, localization, modifications, function and possible interactions or complexes they can form. Proteome databases of diverse pathogenic microorganisms, including *Mycobacterium tuberculosis*, *Helicobacter pylori*, *Salmonella enterica*, *Bacillus anthracis* and many others, were established by different research groups. The information accessible via the internet significantly supports the research community working on identification candidates for the next generation of vaccines [6–9]. For representative examples of the impact of postgenomic strategies on vaccinology, see recent reviews on *H. pylori* [10–13] or *Neisseria meningitidis* [14].

New technical achievements of proteomics & immunoproteomics

Proteomics has achieved tremendous progress during the past few years. The progress could be observed at all steps of the proteomic analysis: sample preparation protocols, peptide/protein separation methods, mass spectrometry (MS) data collection, data analysis and interpretation. All these aspects are reviewed frequently [15–19]. Of many possible areas of proteomic activity, two seem to be of most interest in immunoproteomics. The first is protein profiling, the identification of a set of proteins of interest, along with their post-translational modifications, such as the extracellular proteome consisting of membrane protein domains or secreted proteins (secretome) in immunoproteomics [20]. This requires specialized approaches of sample preparation to be worked out [21–23], for instance, specific biotinylation of cell surface proteins to improve their further affinity enrichment [24]. The second area is relative quantitation, which allows the comparison of the contents of the subproteome of interest in different biological contexts.

Strategies for protein identification

The identification of a protein that is at the heart of proteomic technology always assumes the measurements of molecular masses of protein fragments. In this way, for a given protein, a unique set of protein-specific numbers is obtained in MS measurement, instead of a single number – an intact protein molecular mass. When this set of numbers is compared with protein sequence databases, fragmented *in silico* and a statistically significant match is found, reliable protein identification is obtained.

Two strategies of the stepwise protein fragmentation coupled to MS analysis are known. In the so-called bottom-up strategy, proteins are first digested with a specific protease (most often trypsin) and the masses of these peptide fragments are measured, giving the so-called peptide mass fingerprint (PMF) of the protein. The PMF itself could be specific enough to identify a protein but it is difficult to obtain a statistically significant match in PMF-only-based identification, and the usage of this method is discouraged [25]. Instead, a second step of fragmentation is introduced inside the mass spectrometer (gas-phase fragmentation), referred to as MS-MS. In MS-MS, the breakdown of covalent bonds in the molecule (e.g., a tryptic peptide) is induced, and masses of resulting fragments of a peptide are measured and compared with sequence databases. A set of protein-unique measured numbers is now much expanded and the identification is much more reliable. Owing to easy access to hybrid spectrometers equipped with the MS-MS capability, the fragmentation-based identification is now routine. In the second strategy (top-down approach), the initial proteolytic fragmentation is omitted, masses of intact proteins are measured first and MS-MS is the only fragmentation applied to obtain information on protein fragment masses. The top-down strategy has an obvious advantage as it retains the link between the intact protein and the full spectrum of its post-translational modifications (PTMs), whereas in the bottom-up approach, the modifications can be identified only at the peptide level. Thus, in the bottom-up strategy, a significant part of information on PTMs is lost. However, since the sensitivity of mass spectrometers decreases with molecular mass measured, the MS-MS of proteins is much less efficient than for peptides, and the top-down approach is successful only in special cases, mainly for small proteins [26,27]. New fragmentation methods bring promise for improvements in this field [28].

Analytical steps

Step one: separation

In a routine proteomic experiment, the protein or a set of proteins is first digested into fragments. Molecular masses of these peptides (parent ion masses) and their MS-MS fragments (daughter ion masses) have to be measured. At some point of the increasing complexity of the analyzed protein set, the MS measurements cannot yield good results without prior separation of peptides. A typical separation is the reversed-phase liquid chromatography (LC) in nanoversion, where the column diameter is decreased to 75 µm and flow to 0.2 µl/min. This leads to a very efficient LC-MS coupling, where thousands of species can be analyzed by MS in an automated run. Recent advances in the LC technology prove that the peak capacity (i.e., number of species that can be resolved) of the LC columns can still be increased and still more-complex peptide mixtures could be resolved in a single run. In parallel, alternative separation techniques are applied to support MS analyses. Besides classic 1D and 2D polyacrylamide gel electrophoresis gels, molecular sieves or ion-exchange chromatography, recent new ideas have expanded the spectrum of available separation techniques. The most interesting seem to be isoelectrofocusing of peptides [19] either in gel [29] or in

solution [30], or capillary electrophoresis [31] and ion mobility spectroscopy (IMS) [32]. IMS allows for peptide ion distribution in the gas phase, which separates them according to their collision cross-section, which is correlated with molecular shape. The plethora of available separation techniques allows one to build multidimensional separation strategies [33] by combining several, preferably orthogonal, separation steps prior to MS analyses. This allows one to carry out proteomic experiments on very complex protein sets, whole proteomes or subproteomes, with reasonable protein coverage.

Step two: ionization

After separation, the molecular masses of the peptides and their decomposition fragments have to be measured in the mass spectrometer. For this aim, the molecules have to be transferred to the gas phase and ionized (charged). Two main methods, MALDI [34] and electrospray ionization (ESI) [35], are used to achieve this goal in proteomics and little has changed in this respect since the end of the 1980s, when invention of the methods marked the beginning of MS-based proteomics. ESI allows the transfer directly from solution (also aqueous), and is well suited to the biological milieu; MALDI requires embedding the analyte into a paracrystalline matrix. Thus, ESI will couple well with LC separation but, when LC is not necessary, MALDI becomes preferable due to a shorter time of measurement. While of comparable sensitivity in low attomolar range for small peptides, for complex mixtures where separation is a must, MALDI has disadvantages, which make it less useful. These are the necessity to include a fraction-collection step, interference of matrix signals below 500 Da in the spectrum and a poor reproducibility of spectra due to uneven dispersion of analyte in the matrix. However, recent improvements in fragmentation quality of MALDI-based instruments (TOF–TOF technology [36]) revived interest in utilization of the main MALDI advantage – the speed of measurement. New matrices also increase sensitivity and improve peptide recovery in proteomic analyses [37].

Interestingly, both ionization methods give access to nonoverlapping fractions of the proteome, and their results are complementary [38]. New variants of ESI have recently been worked out, which may find application in proteomics: electrosonic spray ionization (ESSI) [39] and desorption of the analyte from the surface by electrospray ionization (DESI) [40].

Step three: mass measurement

Ionized molecules in the analyzer compartment of the spectrometer undergo separation according to their molecular mass, or more precisely according to mass divided by charge (m/z), which allows one to measure their molecular masses. Several measurement parameters are crucial in proteomics, including measurement accuracy, resolving power, sensitivity and dynamic range. Several available variants of spectrometers are often compared with respect to their performance (see Table 1 in [15]), and such comparisons allow one to assess the suitability of a given spectrometer for a given purpose. However, with the evolution of the technology, a constant improvement of the overall performance can still be observed. Recent highlights are the two types of MS spectrometers (Fourier-transform ion cyclotron resonance [41] and Orbitrap [42]) that outperform other types at least in resolving power and accuracy, proving that the masses can routinely be measured with the accuracy of single parts per million (1 ppm accuracy is measuring 1000-Da mass with precision of 0.001 Da). Such incremental improvements allow for gradual progress in proteomic analyses of complex protein sets – an ultimate goal of proteomics.

For a required step of gas-phase fragmentation, new developments also promise progress. Instead of classic collision-induced dissociation (CID), where peptide decomposition is induced by collisions with gas atoms (helium or argon) in the collision cell of the spectrometer, the decomposition may be induced by electrons (electron capture detector [43], electron transfer dissociation [44]) or laser radiation (infrared multiphoton dissociation [45]). When combined with the Orbitrap spectrometer, electron transfer dissociation fragmentation proved to complement well with the CID results in global analyses of yeast proteome and outperformed it for highly charged or phosphorylated peptides [46]. These new fragmentation methods are beginning to be included in commercially available spectrometers and will certainly become routine option in proteomics.

Step four: protein identification by a database search

Results

Measured parent and daughter ion masses from MS of peptides provide the basis for protein identification. The identification is usually carried out by comparison with protein-sequence databases, for which all entries were converted to the lists of expected peptide fragment masses calculated by *in silico* fragmentation [47]. At this stage, the bioinformatic part of MS data analysis is initiated [48]. In recent years, multiple software tools were created for this purpose, and many papers review these developments [49–51]. In the most frequently used software platforms, MASCOT [52] and SEQUEST [53], a scoring system is used to assess the reliability of the results with threshold values that have to be exceeded for the protein hit/identification to be accepted as significant. However, more and more often this analysis is supplemented by an additional step that involves the direct calculation of the false-discovery rate (FDR) value corresponding to a given score, providing a more strict measure of statistical significance of the result. The FDR value may be calculated, for instance [54], by repeating the search against a new database (decoy database), expanded by a large set of random sequences – the number of hits on random sequences allows one to calculate the level of false-positive results in the search against the real database. Since the procedures of database search are automated and large datasets are being analyzed, the verification of the results on the raw data level is rarely conducted, and sound statistical analysis is a must to avoid multiplication of false-positive results. This concern led to formulation of guidelines that should be carefully obeyed [55,56]. The bioinformatic stage of proteomic identification analysis is the step that most requires an experienced operator in the otherwise routine procedure, which is supposed to yield a reliable list of proteins present in the sample.

Quantitative comparisons: differential proteomics

The identification of protein contents of a sample of interest, discussed previously, is a necessary step preceding relative quantitation experiments. In relative quantitation, the levels of proteins are compared between the sample of interest and a control sample (pairwise comparison) or in a series of samples representing different conditions. In the classic approach, the intensities of protein spots on 2D-PAGE gels provide the relative measure of protein quantities [57]. 2D gels suffer from many disadvantages: poor automation, limited dynamic range, loss of some protein classes (e.g., membrane and extreme pI), poor reproducibility and a technically demanding and difficult process of high cost, low throughput and long analysis time. Since gels retain the link between the identified protein and its modifications, which is lost after trypsin digestion, this approach is still superior if specifically modified protein forms have to be detected and compared (e.g., PTMs or proteolytic variants). However, in many cases, standard 2D gel quantitation is more and more often substituted by LC-MS-based quantitation [58].

In MS-only methods, the protein-level measurements are based on comparing the amplitudes of MS signals of peptides originating from proteins in samples versus controls. In these experiments, the absolute concentrations of proteins are not measured. Absolute concentration measurements of a limited set of proteins with the use of MS are also possible with the use of an isotopically labeled internal standard added to the sample in a known quantity (absolute quantification strategy [59]). In isotopic labeling, chemically identical moieties of different isotopic composition (i.e., different molecular mass) are introduced into peptides from the sample and standard/control. Thus, MS spectra allow one to differentiate the signals originating from peptides from the sample and peptides from the internal standard/control. Isotopic labeling is also used for relative quantitation and allows one to mix the sample with the control before MS-based quantitation. Isotopic labeling is, however, not a must; MS-based relative quantitation can also be carried out in a label-free fashion in LC-MS experiments. Knowing the retention time of a given peptide and its molecular mass, its signals can be traced across the set of MS data collected separately for each sample in the series [60,61]. This approach is, however, demanding on data analysis, where the peptide signal localization in a series of complex spectra is prone to error and data-normalization procedures are necessary. At a cost of an additional step of isotopic labeling, these difficulties are overcome in relative quantitation methods involving isotopic labeling [62]. Isotopic labels can be introduced *in vivo* by metabolic labeling of proteins with stable isotopically labeled amino acids (SILAC) [63] and *in vitro* by an enzyme-catalyzed reaction coupled with trypsin digestion (^{18}O labeling) [64] or by chemical reaction with a labeling reagent, for instance, isotope-coded affinity tag (ICAT) [65] or isobaric tag for relative and absolute quantitation (iTRAQ) [66]. SILAC is most suitable for cell culture studies. The main advantage of iTRAQ, which has become the most popular labeling method, is that it allows for parallel protein identification and quantitation of proteins in up to eight samples in a single experiment. Its disadvantages are relatively high cost and incompatibility with sensitive ion-trap MS analyzers. MS-based relative quantitation methods require appropriate software tools and statistical analysis tools, which are now being worked out [67,68].

The proteomic methods mentioned previuosly are commonly named 'shotgun proteomics', as they aim at identification and quantitation of all detectable species in the sample. However, the complexity of proteomes or subproteomes under study exceeds the capabilities of even the most sophisticated proteomic systems. Shotgun analysis focuses on thousands of peptides from the most abundant proteins but the less abundant, and possibly more important, proteins may escape attention. Therefore, more hypothesis-driven approaches have recently been advocated [17], in which only a subset of expected peptide signals, coming from a small number of peptides characteristic for a given protein (known as 'proteotypic peptides') is measured, and the remaining are ignored. This simplifies the analysis and allows one to use dedicated MS methods of high sensitivity and selectivity, such as multiple-reaction monitoring (MRM) [69], which gives access to quantitative data for less-abundant proteins of interest, or their modifications, selectively chosen from the extremely complex background of samples containing thousands of other proteins.

Immunoproteomics, a novel strategy combining standard proteomics with serological screening (serological proteome analysis), is currently the method of choice for identification of new immunogenes of protective values [70]. Serological screening confirms protein *in vivo* expression and its immunogenicity. A single experiment can expand the list of immunogenic proteins of an investigated microorganism. Those vaccine candidates proposed by different research groups usually overlap but rarely completely coincide. Observed discrepancies result from enormous genetic diversity of isolates of the same species, as well as from variability of the sera used for the screening [71,72]. There are two main strategies for conducting immunoproteomic studies, differing in the methodology of protein preparation: MS or protein-chip technology. They both have advantages and shortcomings. The MS is based on pathogen proteins produced *in vitro* by bacteria growing on synthetic media; thus, this strategy is not able to identify immunogenic proteins expressed only *in vivo*, during infection. The protein microarray strategy permits investigation of immunogenicity of proteins expressed *in vivo*, as well as those produced under laboratory conditions. However, the microarray technology encounters the problem of gaining the natural conformation of immobilized proteins, as attachment to the chip may affect their 3D structure and in consequence their interaction with antibodies [73–75]. However, the protein-chip technology permits screening for immune responses, not only against proteins but also against small, synthesized, *in situ* peptides containing epitopes. The procedure might be the first step for construction of the epitope-derived vaccines [76]. The recently developed technology of quantitative proteomics combined with immunoproteomics may provide insights into shift of the pathogen immunome during infection [77,78].

Although recent studies indicated that including more antigens might increase vaccine efficacy, the majority of subunit vaccine prototypes still contain not more than one or two antigens [79,80]. Immunoproteomics provides a reliable way for discovery of novel vaccine candidates; particularly low-abundant proteins or those with untypical properties (see next section). However, in the case of some pathogens, identification of antigens of protective and/or therapeutic value requires screening proteomes derived from many clinical isolates with sera taken from patients with diverse disease outcomes [11,81,82].

In many cases, standard immunoproteomic experiments have indicated several hypothetical proteins that might be candidates for vaccine development. Understanding their role in pathogenesis and deciphering their biological functions are of great importance for further selection. One method of achieving the goal is the identification and characterization of protein complexes. The merits and limitations of proteomic analysis of protein complexes by MS and protein-chip technology are discussed in detail by Zhou and Veenstra, and Bertone and Snyder [73,83].

In conclusion, it should be pointed out that even if proteomic/immunoproteomic experiments conducted by various research groups do not always provide compatible datasets, the comprehensive comparison of multiple results enables generation of a list of potential candidates for effective vaccine construction.

Proteomics is complementary & generally compatible with genomics & transcriptomics

Proteomics, together with genomics and transcriptomics, are emerging complementary approaches that have led, in combination with bioinformatics tools, to enormous progress in basic as well as applied biology. Other than vaccine development, applied aspects of proteomics are reviewed in [84–87]. In contrast to information provided by genome analysis, which is rather invariable, the information originated in transcriptomics and proteomics is considered to be dynamic. Proteomics is the most informative one, able to define different isoforms of the proteins and post-translationally modified proteins. Rapid growth of proteomics is evidently grounded on the progress made in the genomics field, where interpretation of the proteomic analysis is based on data created by annotation of the sequenced genomes. In return, proteomics is often employed for validation of bacterial genome annotations [88–90]. Additionally, analysis of the surface proteomes allows verification of the membrane protein localization predicted with Prediction of Protein Sorting Signals and Localization Sites in Amino Acid Sequences (PSORT) [91].

Proteomics versus genomics

Reverse vaccinology, a procedure based on bioinformatic analysis of the genomic sequence, is having a great impact on vaccine development. Despite numerous criteria that are applied for antigen selection, a large number of candidates is usually designated, and, consequently, their usefulness has to be checked experimentally. Although many strategies of gene cloning and recombinant protein purification have been discovered recently, this is still not a trivial work [92–94]. Usually, only a portion of selected pathogen genes are successfully cloned and expressed in *Escherichia coli*. In the first genome-derived vaccinology experiment concerning *N. meningitidis* (MenB), out of 600 genes encoding putative surface-exposed or secreted proteins, only 350 were cloned and expressed in *E. coli* [95]. Using proteomics and/or immunoproteomics as a supporting approach ensures the reduction of the number of vaccine candidates, indicating the most abundant and highly seropostive proteins for the future studies.

Experimental analysis of bacterial subproteomes often identifies novel immunogenic proteins that cannot be predicted by bioinformatic means. Several proteomic studies indicated extracytoplasmic proteins lacking any structural features that could be marked out by currently available algorithms for protein localization prediction. These unexpected results might be due to imperfect cell-fractionation techniques that cause cross-contamination. On the other hand, our knowledge concerning the mechanism of protein secretion/transport is incomplete. Examples include elongation factor Tu, ribosomal proteins, enolase, glyceraldehyde-3-phosphate dehydrogenase and phosphoglycerate kinase [96–99]. The secretion of elongation factor Tu and enolase of *M. tuberculosis* and *Steptococcus pneumoniae*, was experimentally proven [100,101]. Severin *et al.* identified *Streptococcus pyogenes* surface-exposed proteins by MS analysis of the peptides released from bacterial cells using trypsin digestion, a method that should not be burdened with cross-contamination. Among the 79 identified proteins, 21 are predicted by bioinformatics analysis to be cytoplasmic [102].

Proteomics versus transcriptomics

An understanding of the dynamics of gene expression during infection is crucial for selecting effective vaccine candidates. It is generally accepted that an antigen expressed or upregulated during infection and relevant to pathogenesis is probably a more effective antigen. Many analyses of the level of gene expression under different conditions on the proteome level have also been recently performed. Examples include *M. tuberculosis* [103–107], *Campylobacter jejuni* [108], *Francisella tularensis* [109], *H. pylori* [110], *B. anthracis* [111], *Pseudomonas aeruginosa* [112,113], *Salmonella enterica* sv. Typhimurium [114], *S. pneumoniae* [115].

In the majority of proteomic experiments, proteomes are derived from bacterial cells growing on synthetic media under conditions imitating those encountered by pathogens during infection. The presented data should be treated with a certain measure of criticism as the generated conditions may not precisely reflect those met by pathogen inside the host organism. Although the changes in the proteomes driven by environmental stimuli have been analyzed for many pathogenic bacteria, there are few examples of direct comparison of the transcriptome and proteome derived from the same bacterial culture. Generally, the changes in transcript level due to different growth conditions correspond with the abundance of relative gene product. However, in some analyses, a significant discrepancy between transcriptomic and proteomic data was observed, implying the role of post-transcriptional and post-translational mechanisms

in determining the protein profile changes in response to various growth conditions. Thus, the transcriptome analysis aimed at identification of new vaccine candidates should be verified by proteome analysis [116,117].

Host response to infection: crucial point for effective vaccine construction

Comparisons of the protein profile of bacteria growing *in vitro* to the proteome obtained from bacterial cells growing in cell culture, as well as comparison of the proteomes of the same species isolates obtained from patients with different disease outcomes, reveals significant differences in the abundance of some proteins [78,118–120].

The knowledge of the host cell's response to pathogenic microorganisms is important for understanding the disease process. It also helps decide what type of immune response should be induced by immunization to get the desired protective effect. Jenner and Young presented a comprehensive review of host transcriptional-response to different pathogens. They collected data from 32 published papers describing 77 different host–pathogen interactions [121]. The cellular response to infection on the proteomic level has, so far, been evaluated for a small number of microorganisms or their products [122]. *H. pylori* is a pathogenic microorganism that has been studied extensively in this respect to understand the mechanism of gastric cancer induction and to search for new diagnostic markers and vaccine candidates [98,123–126]. To avoid misinterpretation of results, researchers have to be aware that most of the cell lines used are of a tumor origin and, subsequently, their protein profiles are similar but not identical to protein profiles of the cells present in the host tissues [127].

In conclusion, proteomics, a high-throughput analysis, allows not only identification of a large number of proteins that are candidates for vaccine construction but also facilitates understanding of their role in pathogenesis by estimating protein levels during different steps of infection and analyzing the influence of the pathogen's protein on host physiology. Proteomics ensures verification of data obtained by the transcriptomic approach, since the transcriptome and proteome are not always coordinately regulated. With respect to reverse vaccinology, immunoproteomics enables one to narrow the search to a smaller number of vaccine candidates.

Sample preparation: bottleneck of prokaryotic proteomics

Classical proteomics combines 2D electrophoresis (2DE) and MS for protein identification. The major limitation of this strategy arises from the fact that the number of selected proteins is underestimated and the generated list includes mostly abundant proteins. This limitation concerns mainly hydrophobic membrane proteins, which are difficult to isolate and to separate by 2DE. Therefore, enrichment of particular protein categories and subsequent characterization of subproteomes is necessary. Advances in various MS-based technologies, mainly the improvement of non-gel methods using LC-MS, permits one, at least to some extent, to overcome these problems, although gel-based proteomics are still the most frequently used techniques in investigation of pathogenic bacteria. The preparation of protein extracts containing many proteins in a form that would be soluble and compatible with further analysis is still a critical starting point for proteomic analysis. Proteins desirable for subunit vaccine generation are largely surface, membrane or cell wall-located and secreted proteins. Thus, major research effort is directed toward improving membrane protein isolation methods, working out effective procedures for removing proteins from the bacterial surface and enriching the amounts of low-abundance proteins.

Outer-membrane proteins (OMPs) of Gram-negative pathogenic bacteria play an important role in the virulence process, being involved in adherence, eukaryotic cell invasion and interaction with components of the host immune system, as well as in active transport of many vital metabolites and iron. Thus, OMPs are considered as potential candidates for effective vaccine construction. However analysis of this bacterial subproteome is difficult, mainly due to the hydrophobic nature of proteins. Recent improvements of agents used for solubilization of membrane proteins, especially the introduction of new classes of detergents, helps overcome this limitation. OMPs are generally solubilized in buffer containing 3-([3-cholamidopropyl]dimethylammonio)-1-propanesulfonate, Triton X-100 or n-dodecyl-β-D-maltopyranoside and supplemented with surfactant amidosulfobetain-14 [128,129]. Some nonionic detergents (e.g., lauryldimethylaminoxide [LDAO] and octyl-β-D-glucopyranoside [OG]) are exceptionally useful for obtaining intact membrane protein complexes [130]. OMPs of some pathogenic Gram-negative microorganisms have been analyzed by immunoproteomic strategies with sera obtained from patients or immunized animals. Every experiment confirmed the location and immunogenicity of previously characterized antigens and detected novel ones [131–137].

The Gram-positive cell wall consists of a thick homogenous peptidoglycan layer lying outside the inner membrane. In addition to the cell wall, many Gram-positive bacterial species have polysaccharide capsules. Both of these features prevent release of proteins external to cytoplasmic membrane, a problem that is technically difficult to overcome. The classical way of releasing cell surface-associated proteins of Gram-positive bacteria encompasses mechanical cell treatment combined with the digestion the peptidoglycan-associated proteins with lysozyme, mutanolysin and/or lysostaphin in the presence of an osmotic protective agent, but these methods are inadequate. Thus, many laboratories are presently elaborating new protocols for obtaining high-yield cell envelope proteins and for precise determining their localization in the cell envelope [91,99,102,138].

Much work has been carried out to improve the extraction method for *M. tuberculosis* membrane proteins, which are hydrophobic and low in abundance. Generally, the strategies relied on cell membrane and cell wall fractionation and the use of different washing techniques and different detergents for membrane protein extraction, followed by 2DE MS identification [139,140]. Recently, Målen *et al.* showed that combining Triton X-114

extraction with LC-MS is an efficient method for *Mycobacterium bovis* BCG membrane proteins identification that allows skipping of the cell envelope fractionation step [141].

It is generally accepted that some of the secreted bacterial components recognized by host immune system could be protective antigens. To address this issue, the secretomes of some bacterial pathogens were elucidated and the immunogenicity of the proteins was analyzed. These experiments demand specific methods of cell growing to restrict contamination with proteins, which are present in the medium or released from lysed bacterial cells. Many strategies, mainly trichloroacetic acid or ammonium sulphate precipitation of the proteins from filtered culture supernatant, have been employed to obtain protein samples susceptible to proteomic/immunoproteomic analysis [111,142–147].

One important aspect of vaccine research is also elaboration of a reproducible protocol for vaccine preparation. MS techniques can be used to compare vaccine preparations generated by different technologies. One example is the analysis of outer-membrane vesicle (OMV) protein composition [148–150]. The comprehensive proteomics analyses of OMVs derived from *E. coli* K-12, *N. meningitidis* and *Legionella pneumophila* have been recently performed [149,151]. As OMVs are enriched in OMPs, these organelles were an objective of vaccine research. Knowledge of protein composition of OMVs and reproducibility of their preparation has a great value for the development of vaccine of broad protection range. Toward this end, Ferrari *et al.* conducted proteomic comparison of two types OMVs derived from New Zealand MenB strain, potential candidates for anti-MenB vaccine: detergent-derived OMVs (DOMVs) obtained by physical/chemical bacterial cell treatment and OMVs released by the *gna33* mutant strain lacking lytic transglycosidase (m-OMVs) [152,153]. The proteomic analysis showed that the two types of OMVs differ substantially in their protein composition. While DOMVs contain many cytoplasmic and inner membrane proteins, m-OMVs (released by the *gna33* mutant) contain mainly outer membrane proteins. Additionally, the authors documented that immunization with m-OMVs elicits bactericidal antibodies with broader cross-protective activities than DOMVs.

Two comprehensive reviews on bioinformatic and proteomic tools widely applied to understand functioning of the *E. coli* cell envelope-associated proteome have been published recently [154,155]. The application of bioinformatic prediction tools, in combination with proteomic gel-free and gel-based strategies toward other Gram-negative pathogenic microorganisms, will provide insight into their cell envelope proteomes. This knowledge will be of great value for the future research regarding vaccine development.

Examples of the proteomics impact on vaccine development

B. anthracis

B. anthracis, the etiological agent of anthrax, is a spore-forming Gram-positive bacterium. The disease starts when spores enter the host organism and are taken up by macrophages. *B. anthracis* produces a tripartite toxin (encoded by pXO1 plasmid), composed of subunits: protective antigen (PA), lethal factor (LF) and edema factor. Anthrax is mainly an animal disease but also can infect humans with life-threatening sequelae. Although the cases of human disease are rather rare and restricted to specific at-risk groups, *B. anthracis* is considered to be a potential bioterrorism agent and is rated category A by the CDC and National Institute of Allergy and Infectious Diseases (NIAID). A routinely used veterinary vaccine (Sterne vaccine) is based on a capsule-negative, toxin-positive strain. Live vaccines, owing to their high reactogenicities, can be used only for veterinary purposes. The currently used human vaccine (anthrax vaccine adsorbed [AVA]; BioThrax®) is an aluminum hydroxide-adsorbed, formalin-treated culture supernatant of an avirulent *B. anthracis* strain. Even though AVA vaccine is safe and effective, the schedule of immunization is extremely complicated and requires six subcutaneous injections followed by annual boosters [156]. As PA is highly immunogenic, recent research has focused on evaluation of the efficacy of diverse recombinant PAs delivered as pure proteins, as DNA vaccine or by different vaccine-delivery systems [150,157,158].

The availability of the *B. anthracis* genome sequence [159], together with previously determined nucleotide sequences of two virulence plasmids [160,161], generated possibilities to speed up the development of an improved anthrax vaccine. Several studies documented that a live noncapsulated anti-*B. anthracis* vaccine has a greater protective efficacy than the AVA vaccine, probably due to unidentified anthrax vegetative or spore antigen/s that augments the immune response. Coupling of genomic and proteomic techniques facilitate searching for novel proteins, potential enhancers of AVA efficiency. The comprehensive analysis of the *B. anthracis* proteome and immunome, including chromosome- and plasmid-encoded proteins, has been conducted recently by a research group from the Israel Institute for Biological Research. Exhaustive and carefully performed *in silico* inspection of the genome and plasmid ORFs permitted selection of 197 genes (161 located in the chromosomes and 36 located in the plasmid) whose products were classified as novel candidates for subunit vaccine construction. Serological analysis with sera obtained by animal immunization with *B. anthracis* of different plasmid profiles allowed the identification of 43 (out of 109 tested) seropositive proteins. As expected, most of them appeared to be cell wall anchored or involved in cell wall biogenesis or function [98,162,163]. At the same time, the research group examined secretomes of virulent and plasmid-cured avirulent *B. anthracis* strains growing in minimal medium under high CO_2 tension, conditions that the pathogen encounters during infection. The comparative proteomic study led to recognition of dozen of proteins relevant to vaccine design. Finally, 58 immunogenic proteins were identified. SERPA methodology allowed the researchers to determine the strength of protein immunogenicity. Finally, the authors chose eight genes for a protection study (*pagA* was employed as a control). Although all investigated vaccine prototypes induced a specific humoral immune response, in contrast to *pagA* immunization, none revealed any protection effect [72,164]. Although the routinely administrated anthrax immunization in subcutaneous injections consists of proteins,

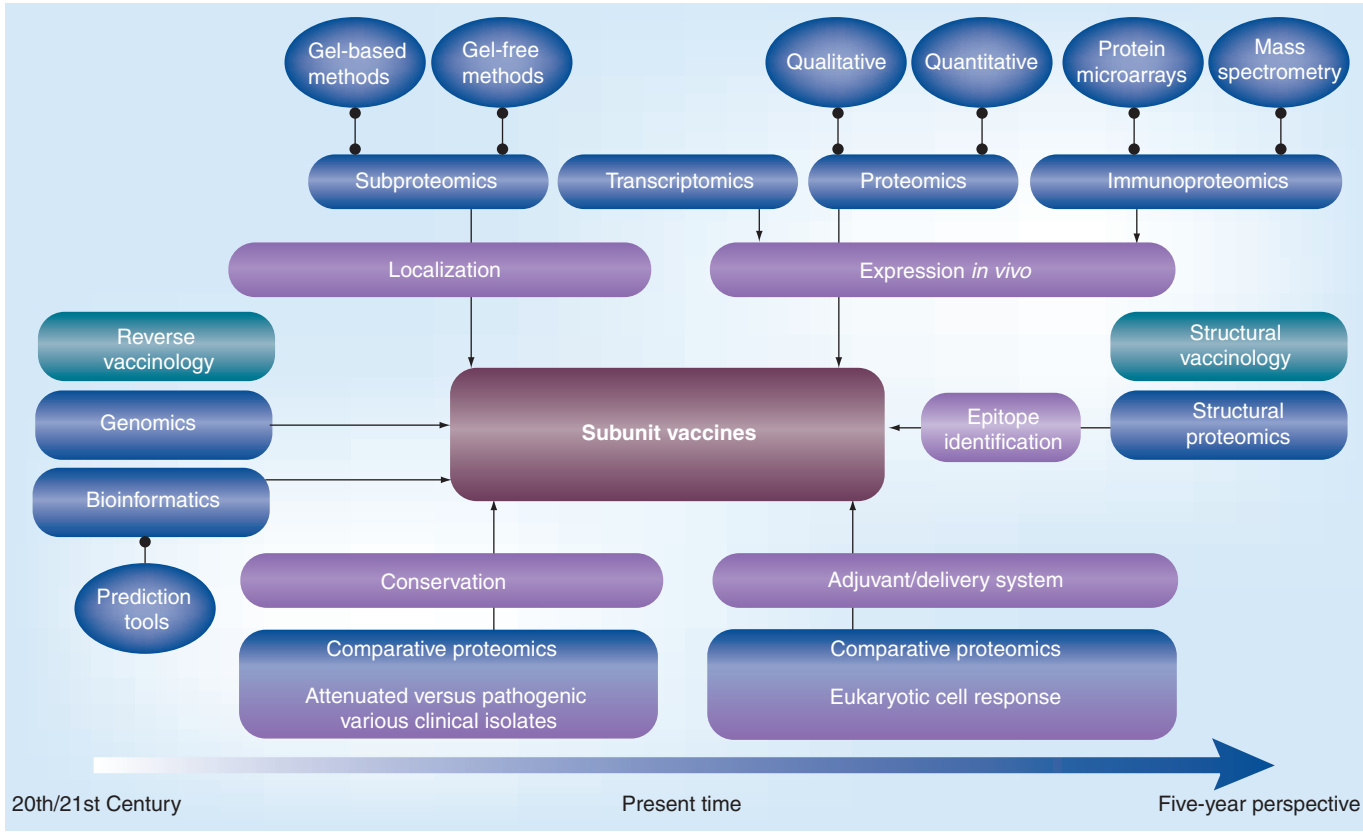

Figure 1. Impact of proteomics on the vaccine development.

Chitlaru *et al.* documented that PA delivered as the *pagA* gene present on a eukaryotic plasmid also induced protective immunity. Thus, the reason why immunization with all selected genes turned out to be ineffective remains unclear. One can speculate that for diverse immunogenes to be effective requires a different means of delivery and that DNA vaccination directs the immune response toward a Th1 response, even though the authors observed some level of the specific antibodies. The work clearly demonstrated the enormous effort that is required to find novel protective antigens.

Several studies showed that spore administration contributes to anthrax vaccine efficacy [165]. Based on this observation DelVecchio *et al.* conducted comparative proteomic analysis of spore proteins of the *Bacillus cereus* group of bacteria, including virulent and avirulent *B. anthracis* strains. Part of the experiment, aimed at identification of novel anthrax vaccine candidates, included screening of the spore proteomes with polyclonal antisera from a human following infection with cutaneous anthrax. In total, 15 highly immunogenic *B. anthracis* spore proteins were identified. Two of these (alanine racemase and neutral protease) appeared to be unique for *B. anthracis* and the immunogenicity of one other protein was reported [166]. It is intriguing that only half of the immunoreactive proteins found by DelVecchio were also identified by proteome serological screening performed by Chitlaru *et al.* [72]. In another study of *B. anthracis* spore immunome, Kudva *et al.* identified 69 immunoreactive proteins by screening an expression library of putative spore surface proteins with human sera from adults immunized with the AVA vaccine. The two proteomic studies gave rather inconsistent results, as only two proteins were found in common for both analyses [167]. This clearly demonstrates that the obtained results depend on the applied technology, such as sera used for the screening assay or the immunization procedure. Thus, the interpretation of results should put emphasis on the applied methodology.

Streptococcus *spp.*

S. pneumoniae is a microorganism responsible for a wide spectrum of diseases, ranging from localized noninvasive infection (middle-ear infection and sinusitis) to invasive, life-threatening diseases, such as pneumonia, meningitis and bacteremia. At present, 23-valent polysaccharide (PSV23) and 7-valent protein polysaccharide (PSV7) conjugate pneumococcal vaccines are available to prevent invasive diseases in adults and children, respectively. One constraint of all PSV vaccines is the fact that they only protect against strains of which capsular types are included in the vaccines; 91 various pneumococcal serotypes have been identified. What is even more troublesome is that large-scale use of the capsular-polysaccharide vaccines will allow strains not included in vaccines to replace the strains whose capsular polysaccharides were used in the vaccines as a cause of pneumococcal diseases. The potential shift in prevailing serotypes can result in the lack of effective available vaccines [168,169].

The aforementioned results inspired a search for an alternative protein pneumococcal vaccine comprising conserved immunogenic proteins, which would elicit protection regardless of the

infecting strain's serotype. Most of the analyzed proteins were selected based on their role in the pathogenicity and their surface localization. The vaccine prototypes were tested individually or jointly in different combinations using various murine models and different delivery strategies. However, none of them induced the protection level comparable to that elicited by polysaccharide protein-conjugated vaccine [170–175].

Thus, the need for a pneumococcal proteome survey aimed at identification of novel protective antigens still exists. Morsczeck *et al.* conducted analysis of *S. pneumoniae* cell envelope proteins isolated by mutanolysin digestion and elution of cell wall-associated proteins at high pH. More than 270 cell wall-associated proteins were identified using three different methods based on MS. A total of 37 polypeptides were identified with all three strategies, and five of them (putative hydrolase, lipoate-protein ligase LplI, ATP-dependent CLP protease, autoinducer-2 production protein, conserved hypothetical protein and RpoD) were selected for further studies. Genes encoding the chosen proteins were highly conserved among 40 different *S. pneumoniae* serotypes and showed expression during animal infection. Vaccination of mice with two candidate proteins (Lpl and ClpP) resulted in lower colony-forming unit titer after infection with invasive *S. pneumoniae* strains, but significantly prolonged survival of immunized animals was not noticed [99].

Recent conducted comprehensive work resulted in identification of two novel *S. pneumoniae* proteins (PcsB, a protein required for cell wall separation, and StkP, a serine/threonine kinase) as candidates for the prevention of infections caused by diverse microorganism serotypes. The used antigenome preselection strategy was based on screening of the display library expressing 15–150 amino acids fragments of the pathogen proteome with sera obtained from patients. A novel approach, which leads to the defining of the 'antigenome' of pathogenic bacteria, combines the advantages of genomics and serological antigen identification. The antigenome technology provides a subset of all proteins, which are expressed *in vivo* and induce humoral responses in the host. Final selection is performed using many assays and various criteria [176,177,203].

S. pyogenes (a group A *Streptococcus* [GAS]) is a Gram-positive human pathogen that causes a wide range of diseases from minor infections, such as pharyngitis and impetigo, to severe invasive illnesses, such as necrotizing fasciitis and streptococcal toxic shock-like syndrome. Additionally, 3–4% of untreated susceptible individuals infected by *S. pyogenes* develop rheumatic fever, an autoimmune disease. The most serious manifestation of rheumatic fever is carditis, which finally might result in chronic rheumatic heart disease. The disease is a consequence of autoimmune reactions between antibodies and/or T cells and host proteins, such as myosin, tropomyosin and kreatin [178–180]. Recently, a novel mass spectral and proteomic approach was used as a tool to detect protein diagnosis biomarkers to distinguish invasive and noninvasive GAS strains [181].

A safe and effective vaccine against GAS has not yet been developed. For many years surface-associated M-protein, the major virulence factor of *S. pyogenes* was under investigation as a candidate for vaccine antigen, but the highly variable N-terminal region of M-protein and elicitation of antibodies that can cross-react with human proteins means that whole M-protein is not a valuable candidate for vaccine antigen. However, many experiments documented that C- or N-terminal peptides of M-protein delivered alone or in different combinations can elicit protective opsonic antibodies against GAS isolates [179,182–184,204].

The search for novel protective antigens was performed by different research groups with various global strategies. The *in silico* approach, conducted by McMillan *et al.,* indicated 17 proteins of known or putative function and 11 without known homologues [185]. Rodriguez-Ortega *et al.* developed a method that allows fast and selective identification of surface proteins. The procedure was based on proteolytic enzymes (trypsin or proteinase K) which 'shave' the surface-exposed proteins. The released peptides were identified by MS-MS. Of the 72 surface-exposed identified proteins, 95% were previously predicted to be cell wall-anchored proteins, lipoproteins, transmembrane or secreted proteins, and that confirmed effectiveness of the approach. However, among the 14 identified proteins expressed as recombinant proteins and tested in mice, only one (Spy0416) elicited high protection levels [91]. In spite of using a similar experimental approach, Severin *et al.* identified a comparable but not identical set of proteins. Only 33 proteins overlapped in both studies. The set of identified proteins was influenced by the cell phase of growth. To confirm the proteomic-predicted surface localization, 16 out of the 79 proteins were expressed in *E. coli*. Whole-cell ELISA tests proved surface localization of 12 out of 16 proteins. However, only three of them – hypothetical proteins (Spy0843 and Spy0130) and a putative cell envelope proteinase (Spy0416) – induced high antibody titers when tested in an animal model. Thus, they could be considered as potential vaccine candidates [102].

Expert commentary & five-year view

The identification of new protective antigens is essential for development of effective vaccines against infectious diseases. Despite being a young field, proteomics has left its mark on vaccine research. The availability of genome sequences of many bacterial pathogens, sometimes many isolates of the same species, generates tremendous opportunity to accelerate vaccine research. Many improvements of classical 2DE MS technology have led to more efficient identification of new antigens, as well as to precise analysis of their role in pathogenesis. Proteomic strategies replace technologies widely used towards the end of last century, such as *in vivo* expression technology or signature-tag mutagenesis. Global strategies used to compare proteomes of virulent and avirulent strains, or proteomes/immunomes of different clinical isolates, led to long lists of potential candidates for vaccine construction. Comparison of numerous proteomes of the same species and their analysis using sera taken from patients with different disease outcomes is of great importance, especially in case of pathogenic microorganisms containing open pan-genomes where, sometimes, the number of dispensable genes exceeds the number of core genes.

However, due to various reasons pointed out in this article, the sets of potential protective antigens identified by different research groups are overlapping but not identical, and only a small number of them revealed protective effects when analyzed using different animal models. Additionally, not every result obtained in animal's models can be translated to humans. This means that gathering of basic research data by proteomic strategies, animal experiments and immunological studies should be thoroughly analyzed to select proteins for further investigations and that only clinical trials can provide reliable answers concerning their efficacy. Thus, one of the primary challenges of proteomic approaches aimed at effective vaccine construction is the integration of large quantities of data and their comprehensive analysis. To ensure success in combating infectious diseases by immunization, a multidisciplinary approach combining basic research, including proteomics, with computational biology, as well as molecular epidemiology, is needed. To achieve fast progress in the field of vaccinology, close international cooperation between academic research groups and commercial partners is indispensable to optimize the chance of success. With this respect, generation of proteome databases that are accessible via the internet was the first step to providing easy exchange of basic information. A representative example of this approach is the work on integration of proteomic data, presented by Schmidt *et al.* [186]. However, even when international consortia are involved in the initiatives, the introduction of new effective vaccines into the market is a long-term undertaking. It is estimated that licensing a new effective vaccine against tuberculosis by the year 2015 will require at least 20 new vaccine prototypes entering into Phase I safety clinical trials [187,205].

The current leading-edge research is moving toward structural proteomics; understanding 3D structures of the selected antigens using x-ray crystallography, NMR or modeling. This will guarantee progress in creating fusion protein molecules, multicomponent subunit vaccines or polyepitope-containing vaccines [91,188–192].

In the future, more focus will probably be applied to employ proteomics as an experimental tool to study protein function, especially protein–protein interactions. Learning the roles of immunogenic proteins of unknown function should facilitate rational vaccine design. The level of many virulence factors is affected by small RNAs, which mainly regulate the translation process or mRNA stability. It was documented that small RNAs are global regulatory molecules and a single small RNA is able to bind several different targets and modulate the level of various proteins. Applying transcriptomic and proteomic analysis to identify mRNAs/proteins influenced by small RNAs might speed up the vaccine research process [193–195]. In this aspect, it is worthwhile to mention the work performed by Wilson *et al.* that determined changes of *S. enterica* sv. Typhimurium transcriptome and proteome in response to a space-flight environment. The experiment revealed the alteration of the conserved RNA-binding protein Hfq regulon and showed that Hfq plays a central regulatory role in gene expression under the space-created conditions [196,197]. We can also expect development of new global strategies known as infectomics, which were initiated by the human genome-sequencing project. Gaining knowledge about the host cell response to infection will not only clear up several aspects of host–pathogen interaction, but also will explain why some of us are more susceptible to infection than others, and what kind of genetic polymorphisms predispose some individuals to specific disease outcomes. This information might be important when choosing a target population for vaccination, especially when disease outcomes are influenced by the host genotype and/or when patients develop postinfection sequelae.

Acknowledgements

We thank Jeff Hansen for critical reading of the manuscript.

Financial & competing interests disclosure

This work was supported by a grant from the Ministry of Science and Higher Education (No. N401 183 31/3968). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Key issues

- Integration of different techniques of global bacterial cell investigation and analysis of the interactions between host cells and pathogens will accelerate the development of subunit vaccines in combating bacterial diseases.
- Proteomics often verifies genome annotations.
- Proteomic technologies enable the discovery of extracytoplasmic proteins that can not be predicted by *in silico* strategies.
- Novel quantitative proteomic strategies are extremely useful to examine changes in gene expression on proteome level during infection.
- Immunoproteomics, which combines proteomic protein identification with immunological screening, is the method of choice for discovering new protective antigens.
- Although proteomic/immunoproteomic technologies shorten the list of proteins, potential candidates for vaccine construction provided by the reverse vaccinology strategy, all still have to be tested separately or jointly for immunogenicity and protection.
- Lists of potential protective antigens provided by various research groups are different but overlapping. The observed discrepancies are mainly due to differences in used experimental strategies and also due to genetic diversity of clinical isolates.
- One of the significant shortcomings of proteomics is the lack of reproducible and reliable methods of extracytoplasmic protein isolation.
- Proteomic/immunoproteomic data simultaneously convey information about protein immunogenicity and localization as well as gene expression during infection.
- Proteomic analysis of eukaryotic cell response to infection is of great value for new vaccine construction.

References

Papers of special note have been highlighted as:
• of interest

1 Serruto D, Rappuoli R. Post-genomic vaccine development. *FEBS Lett.* 580(12), 2985–2992 (2006).

2 Medini D, Donati C, Tettelin H, Masignani V, Rappuoli R. The microbial pan-genome. *Curr. Opin. Genet. Dev.* 15(6), 589–594 (2005).

3 Mora M, Donati C, Medini D, Covacci A, Rappuoli R. Microbial genomes and vaccine design: refinements to the classical reverse vaccinology approach. *Curr. Opin. Microbiol.* 9(5), 532–536 (2006).

4 Tettelin H, Masignani V, Cieslewicz MJ *et al.* Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proc. Natl Acad. Sci. USA* 102(39), 13950–13955 (2005).

5 Maione D, Margarit I, Rinaudo CD *et al.* Identification of a universal Group B streptococcus vaccine by multiple genome screen. *Science* 309(5731), 148–150 (2005).

6 Nesvizhskii AI. Protein identification by tandem mass spectrometry and sequence database searching. *Methods Mol. Biol.* 367, 87–119 (2007).

7 Pleissner KP, Eifert T, Buettner S *et al.* Web-accessible proteome databases for microbial research. *Proteomics* 4(5), 1305–1313 (2004).

8 Pleissner KP, Eifert T, Jungblut PR. A European pathogenic microorganism proteome database: construction and maintenance. *Comp. Funct. Genomics* 3(2), 97–100 (2002).

9 Zhang C, Crasta O, Cammer S *et al.* An emerging cyberinfrastructure for biodefense pathogen and pathogen–host data. *Nucleic Acids Res.* 36(Database issue), D884–D891 (2008).

10 Aebischer T, Schmitt A, Walduck AK, Meyer TF. *Helicobacter pylori* vaccine development: facing the challenge. *Int. J. Med. Microbiol.* 295(5), 343–353 (2005).

11 Bernardini G, Braconi D, Lusini P, Santucci A. *Helicobacter pylori*: immunoproteomics related to different pathologies. *Expert Rev. Proteomics* 4(5), 679–689 (2007).

12 Jagusztyn-Krynicka EK, Godlewska R. New approaches for *Helicobacter* vaccine development – difficulties and progress. *Pol. J. Microbiol.* 57(1), 3–9 (2008).

13 Kabir S. The current status of *Helicobacter pylori* vaccines: a review. *Helicobacter* 12(2), 89–102 (2007).

14 Bernardini G, Braconi D, Martelli P, Santucci A. Postgenomics of *Neisseria meningitidis* for vaccines development. *Expert Rev. Proteomics* 4(5), 667–677 (2007).

15 Domon B, Aebersold R. Mass spectrometry and protein analysis. *Science* 312(5771), 212–217 (2006).

16 Smith RD. Future directions for electrospray ionization for biological analysis using mass spectrometry. *Biotechniques* 41(2), 147–148 (2006).

17 Malmstrom J, Lee H, Aebersold R. Advances in proteomic workflows for systems biology. *Curr. Opin. Biotechnol.* 18(4), 378–384 (2007).

18 Doerr A. Focus on mass spectrometry in proteomics. *Nat. Methods* DOI:10.1038/nmeth1007-781 (2007) (Epub ahead of print).

19 Canas B, Lopez-Ferrer D, Ramos-Fernandez A, Camafeita E, Calvo E. Mass spectrometry technologies for proteomics. *Brief. Funct. Genomic. Proteomic.* 4(4), 295–320 (2006).

20 Gulcicek EE, Colangelo CM, McMurray W *et al.* Proteomics and the analysis of proteomic data: an overview of current protein-profiling technologies. *Curr. Protoc. Bioinformatics* Chapter 13, Unit 13.1 (2005).

21 Poetsch A, Wolters D. Bacterial membrane proteomics. *Proteomics* 8(19), 4100–4122 (2008).

• **Reviews various aspects of bacterial membranome studies presenting some technical achievements that allow the difficulties encountered during membrane protein extraction to be overcome.**

22 Lu B, McClatchy DB, Kim JY, Yates JR 3rd. Strategies for shotgun identification of integral membrane proteins by tandem mass spectrometry. *Proteomics* 8(19), 3947–3955 (2008).

23 Tan S, Tan HT, Chung MC. Membrane proteins and membrane proteomics. *Proteomics* 8(19), 3924–3932 (2008).

24 Elia G. Biotinylation reagents for the study of cell surface proteins. *Proteomics* 8(19), 4012–4024 (2008).

25 Carr S, Aebersold R, Baldwin M *et al.* The need for guidelines in publication of peptide and protein identification data: Working Group on Publication Guidelines for Peptide and Protein Identification Data. *Mol. Cell. Proteomics* 3(6), 531–533 (2004).

26 Whitelegge JP, Zabrouskov V, Halgand F *et al.* Protein-sequence polymorphisms and post-translational modifications in proteins from human saliva using top-down fourier-transform ion cyclotron resonance mass spectrometry. *Int. J. Mass Spectrom.* 268(2–3), 190–197 (2007).

27 Han X, Jin M, Breuker K, McLafferty FW. Extending top-down mass spectrometry to proteins with masses greater than 200 kilodaltons. *Science* 314(5796), 109–112 (2006).

28 McAlister GC, Berggren WT, Griep-Raming J *et al.* A proteomics grade electron transfer dissociation-enabled hybrid linear ion trap-orbitrap mass spectrometer. *J. Proteome Res.* 7(8), 3127–3136 (2008).

29 Cargile BJ, Bundy JL, Freeman TW, Stephenson JL Jr. Gel based isoelectric focusing of peptides and the utility of isoelectric point in protein identification. *J. Proteome Res.* 3(1), 112–119 (2004).

30 Malmstrom J, Lee H, Nesvizhskii AI *et al.* Optimized peptide separation and identification for mass spectrometry based proteomics via free-flow electrophoresis. *J. Proteome Res.* 5(9), 2241–2249 (2006).

31 Herrero M, Ibanez E, Cifuentes A. Capillary electrophoresis–electrospray–mass spectrometry in peptide analysis and peptidomics. *Electrophoresis* 29(10), 2148–2160 (2008).

32 Ruotolo BT, Gillig KJ, Stone EG, Russell DH. Peak capacity of ion mobility mass spectrometry: separation of peptides in helium buffer gas. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 782(1–2), 385–392 (2002).

33 Motoyama A, Yates JR 3rd. Multidimensional LC separations in shotgun proteomics. *Anal. Chem.* 80(19), 7187–7193 (2008).

34 Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* 60(20), 2299–2301 (1988).

35 Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246(4926), 64–71 (1989).

36 Medzihradszky KF, Campbell JM, Baldwin MA *et al.* The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. *Anal. Chem.* 72(3), 552–558 (2000).

37 Jaskolla TW, Lehmann WD, Karas M. 4-chloro-α-cyanocinnamic acid is an advanced, rationally designed MALDI matrix. *Proc. Natl Acad. Sci. USA* 105(34), 12200–12205 (2008).

38 Bodnar WM, Blackburn RK, Krise JM, Moseley MA. Exploiting the complementary nature of LC/MALDI/MS/MS and LC/ESI/MS/MS for increased proteome coverage. *J. Am. Soc. Mass Spectrom*. 14(9), 971–979 (2003).

39 Takats Z, Wiseman JM, Gologan B, Cooks RG. Electrosonic spray ionization. A gentle technique for generating folded proteins and protein complexes in the gas phase and for studying ion-molecule reactions at atmospheric pressure. *Anal. Chem*. 76(14), 4050–4058 (2004).

40 Takats Z, Wiseman JM, Gologan B, Cooks RG. Mass spectrometry sampling under ambient conditions with desorption electrospray ionization. *Science* 306(5695), 471–473 (2004).

41 Marshall AG, Hendrickson CL, Jackson GS. Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom. Rev*. 17(1), 1–35 (1998).

42 Scigelova M, Makarov A. Orbitrap mass analyzer – overview and applications in proteomics. *Proteomics* 6(Suppl. 2), 16–21 (2006).

43 Zubarev RA. Electron capture dissociation LC/MS/MS for bottom-up proteomics. *Methods Mol. Biol*. 492, 413–416 (2009).

44 Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl Acad. Sci. USA* 101(26), 9528–9533 (2004).

45 Crowe MC, Brodbelt JS. Infrared multiphoton dissociation (IRMPD) and collisionally activated dissociation of peptides in a quadrupole ion trap with selective IRMPD of phosphopeptides. *J. Am. Soc. Mass Spectrom*. 15(11), 1581–1592 (2004).

46 Good DM, Wirtala M, McAlister GC, Coon JJ. Performance characteristics of electron transfer dissociation mass spectrometry. *Mol. Cell. Proteomics* 6(11), 1942–1951 (2007).

47 Baldwin MA. Protein identification by mass spectrometry: issues to be considered. *Mol. Cell. Proteomics* 3(1), 1–9 (2004).

48 Fenyo D, Beavis RC. Informatics and data management in proteomics. *Trends Biotechnol*. 20(12 Suppl.), S35–S38 (2002).

49 Fenyo D. Identifying the proteome: software tools. *Curr. Opin. Biotechnol.* 11(4), 391–395 (2000).

50 Kapp E, Schutz F. Overview of tandem mass spectrometry (MS/MS) database search algorithms. *Curr. Protoc. Protein Sci*. Chapter 25, Unit 25.2 (2007).

51 Deutsch EW, Lam H, Aebersold R. Data analysis and bioinformatics tools for tandem mass spectrometry in proteomics. *Physiol. Genomics* 33(1), 18–25 (2008).

52 Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20(18), 3551–3567 (1999).

53 Eng JK, McCormack AL, Yates JR. An approach to correlate tandem mass spectral data of peptides with amino acids in a protein database. *J. Am. Soc. Mass Spectrom*. 5(976), 976–989 (1994).

54 Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* 4(3), 207–214 (2007).

55 Bradshaw RA, Burlingame AL, Carr S, Aebersold R. Reporting protein identification data: the next generation of guidelines. *Mol. Cell. Proteomics* 5(5), 787–788 (2006).

56 Taylor CF, Binz PA, Aebersold R *et al.* Guidelines for reporting the use of mass spectrometry in proteomics. *Nat. Biotechnol*. 26(8), 860–861 (2008).

57 Issaq H, Veenstra T. Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE): advances and perspectives. *Biotechniques* 44(5), 697–698, 700 (2008).

58 Ong SE, Mann M. Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol*. 1(5), 252–262 (2005).

59 Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl Acad. Sci. USA* 100(12), 6940–6945 (2003).

60 Chelius D, Bondarenko PV. Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *J. Proteome Res*. 1(4), 317–323 (2002).

61 Pasa-Tolic L, Masselon C, Barry RC, Shen Y, Smith RD. Proteomic analyses using an accurate mass and time tag strategy. *Biotechniques* 37(4), 621–624, 626–633, 636 passim (2004).

62 Pan S, Aebersold R. Quantitative proteomics by stable isotope labeling and mass spectrometry. *Methods Mol. Biol*. 367, 209–218 (2007).

63 Gruhler S, Kratchmarova I. Stable isotope labeling by amino acids in cell culture (SILAC). *Methods Mol. Biol*. 424, 101–111 (2008).

64 Mason CJ, Therneau TM, Eckel-Passow JE *et al.* A method for automatically interpreting mass spectra of ^{18}O-labeled isotopic clusters. *Mol. Cell. Proteomics* 6(2), 305–318 (2007).

65 Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol*. 17(10), 994–999 (1999).

66 Aggarwal K, Choe LH, Lee KH. Shotgun proteomics using the iTRAQ isobaric tags. *Brief. Funct. Genomic. Proteomic*. 5(2), 112–120 (2006).

67 Mueller LN, Brusniak MY, Mani DR, Aebersold R. An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data. *J. Proteome Res*. 7(1), 51–61 (2008).

68 Park SK, Venable JD, Xu T, Yates JR 3rd. A quantitative analysis software tool for mass spectrometry-based proteomics. *Nat. Methods* 5(4), 319–322 (2008).

69 Anderson L, Hunter CL. Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol. Cell. Proteomics* 5(4), 573–588 (2006).

70 Falisse-Poirrier N, Ruelle V, ElMoualij B *et al.* Advances in immunoproteomics for serological characterization of microbial antigens. *J. Microbiol. Methods* 67(3), 593–596 (2006).

71 Bumann D, Jungblut PR, Meyer TF. *Helicobacter pylori* vaccine development based on combined subproteome analysis. *Proteomics* 4(10), 2843–2848 (2004).

- **Comparative and comprehensive analysis of *Helicobacter pylori* immunogens identified by numerous research groups that resulted in a list of 15 antigens – candidates for vaccine development.**

72 Chitlaru T, Gat O, Grosfeld H, Inbar I, Gozlan Y, Shafferman A. Identification of *in vivo*-expressed immunogenic proteins by serological proteome analysis of the *Bacillus anthracis* secretome. *Infect. Immun*. 75(6), 2841–2852 (2007).

73 Bertone P, Snyder M. Advances in functional protein microarray technology. *FEBS J*. 272(21), 5400–5411 (2005).

- **Clear review highlighting the great potential of the protein microarray technology in basic and applied biology.**

74 Eyles JE, Unal B, Hartley MG *et al.* Immunodominant *Francisella tularensis* antigens identified using proteome microarray. Crown Copyright 2007 Dstl. *Proteomics* 7(13), 2172–2183 (2007).

75 Hu S, Li Y, Liu G *et al.* A protein chip approach for high-throughput antigen identification and characterization. *Proteomics* 7(13), 2151–2161 (2007).

76 Glokler J, Angenendt P. Protein and antibody microarray technology. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci*. 797(1–2), 229–240 (2003).

77 Wu HJ, Wang AH, Jennings MP. Discovery of virulence factors of pathogenic bacteria. *Curr. Opin. Chem. Biol*. 12(1), 93–101 (2008).

78 Xia Q, Hendrickson EL, Wang T, Lamont RJ, Leigh JA, Hackett M. Protein abundance ratios for global studies of prokaryotes. *Proteomics* 7(16), 2904–2919 (2007).

• **Presents recently developed technologies of quantitative proteomics, which allows precise comparison of protein quantity under different conditions. Combined with immunoproteomics, these strategies may reveal the shift of pathogen immunome during infection.**

79 Giuliani MM, Adu-Bobie J, Comanducci M *et al.* A universal vaccine for serogroup B meningococcus. *Proc. Natl Acad. Sci. USA* 103(29), 10834–10839 (2006).

80 Malfertheiner P, Schultze V, Rosenkranz B *et al.* Safety and immunogenicity of an intramuscular *Helicobacter pylori* vaccine in noninfected volunteers: a Phase I study. *Gastroenterology* 135(3), 787–795 (2008).

81 Mini R, Bernardini G, Salzano AM *et al.* Comparative proteomics and immunoproteomics of Helicobacter pylori related to different gastric pathologies. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci*. 833(1), 63–79 (2006).

82 Pheiffer C, Betts JC, Flynn HR, Lukey PT, van Helden P. Protein expression by a Beijing strain differs from that of another clinical isolate and *Mycobacterium tuberculosis* H37Rv. *Microbiology* 151(Pt 4), 1139–1150 (2005).

83 Zhou M, Veenstra TD. Proteomic analysis of protein complexes. *Proteomics* 7(16), 2688–2697 (2007).

84 Lee SY, Jeoung D. The reverse proteomics for identification of tumor antigens. *J. Microbiol. Biotechnol.* 17(6), 879–890 (2007).

85 Matharoo-Ball B, Ball G, Rees R. Clinical proteomics: discovery of cancer biomarkers using mass spectrometry and bioinformatics approaches – a prostate cancer perspective. *Vaccine* 25(Suppl. 2), B110–B121 (2007).

86 Simpson RJ, Bernhard OK, Greening DW, Moritz RL. Proteomics-driven cancer biomarker discovery: looking to the future. *Curr. Opin. Chem. Biol*. 12(1), 72–77 (2008).

87 Sleno L, Emili A. Proteomic methods for drug target discovery. *Curr. Opin. Chem. Biol*. 12(1), 46–54 (2008).

88 de Souza GA, Malen H, Softeland T *et al.* High accuracy mass spectrometry analysis as a tool to verify and improve gene annotation using *Mycobacterium tuberculosis* as an example. *BMC Genomics* 9, 316 (2008).

89 Jungblut PR, Muller EC, Mattow J, Kaufmann SH. Proteomics reveals open reading frames in *Mycobacterium tuberculosis* H37Rv not predicted by genomics. *Infect. Immun*. 69(9), 5905–5907 (2001).

90 Maillet I, Berndt P, Malo C *et al.* From the genome sequence to the proteome and back: evaluation of *E. coli* genome annotation with a 2-D gel-based proteomics approach. *Proteomics* 7(7), 1097–1106 (2007).

91 Rodriguez-Ortega MJ, Norais N, Bensi G *et al.* Characterization and identification of vaccine candidate proteins through analysis of the group A *Streptococcus* surface proteome. *Nat. Biotechnol*. 24(2), 191–197 (2006).

92 Ding XZ, Paulsen IT, Bhattacharjee AK, Nikolich MP, Myers G, Hoover DL. A high efficiency cloning and expression system for proteomic analysis. *Proteomics* 6(14), 4038–4046 (2006).

93 Guerrier L, D'Autreaux B, Atanassov C, Khoder G, Boschetti E. Evaluation of a standardized method of protein purification and identification after discovery by mass spectrometry. *J. Proteomics* 71(3), 368–378 (2008).

94 McCluskey AJ, Poon GM, Gariepy J. A rapid and universal tandem-purification strategy for recombinant proteins. *Protein Sci*. 16(12), 2726–2732 (2007).

95 Pizza M, Scarlato V, Masignani V *et al.* Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* 287(5459), 1816–1820 (2000).

96 Backert S, Kwok T, Schmid M *et al.* Subproteomes of soluble and structure-bound *Helicobacter pylori* proteins analyzed by two-dimensional gel electrophoresis and mass spectrometry. *Proteomics* 5(5), 1331–1345 (2005).

97 Covert BA, Spencer JS, Orme IM, Belisle JT. The application of proteomics in defining the T cell antigens of *Mycobacterium tuberculosis*. *Proteomics* 1(4), 574–586 (2001).

98 Gat O, Grosfeld H, Ariel N *et al.* Search for *Bacillus anthracis* potential vaccine candidates by a functional genomic-serologic screen. *Infect. Immun*. 74(7), 3987–4001 (2006).

99 Morsczeck C, Prokhorova T, Sigh J *et al.* *Streptococcus pneumoniae*: proteomics of surface proteins for vaccine development. *Clin. Microbiol. Infect*. 14(1), 74–81 (2008).

100 Bergmann S, Wild D, Diekmann O *et al.* Identification of a novel plasmin(ogen)-binding motif in surface displayed α-enolase of *Streptococcus pneumoniae*. *Mol. Microbiol*. 49(2), 411–423 (2003).

101 Xolalpa W, Vallecillo AJ, Lara M *et al.* Identification of novel bacterial plasminogen-binding proteins in the human pathogen *Mycobacterium tuberculosis*. *Proteomics* 7(18), 3332–3341 (2007).

102 Severin A, Nickbarg E, Wooters J *et al.* Proteomic analysis and identification of *Streptococcus pyogenes* surface-associated proteins. *J. Bacteriol*. 189(5), 1514–1522 (2007).

• **First study in which innovative method (trypsin digestion) was applied to release surface-exposed proteins for MS analysis.**

103 Betts JC, Lukey PT, Robb LC, McAdamRA, Duncan K. Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. *Mol. Microbiol*. 43(3), 717–731 (2002).

104 Cho SH, Goodlett D, Franzblau S. ICAT-based comparative proteomic analysis of non-replicating persistent *Mycobacterium tuberculosis*. *Tuberculosis (Edinb.)* 86(6), 445–460 (2006).

105 Florczyk MA, McCue LA, Stack RF, Hauer CR, McDonough KA. Identification and characterization of mycobacterial proteins differentially expressed under standing and shaking culture conditions, including Rv2623 from a novel class of putative ATP-binding proteins. *Infect. Immun*. 69(9), 5777–5785 (2001).

106 Mattow J, Siejak F, Hagens K *et al.* Proteins unique to intraphagosomally grown *Mycobacterium tuberculosis*. *Proteomics* 6(8), 2485–2494 (2006).

107 Starck J, Kallenius G, Marklund BI, Andersson DI, Akerlund T. Comparative proteome analysis of *Mycobacterium tuberculosis* grown under aerobic and anaerobic conditions. *Microbiology* 150(Pt 11), 3821–3829 (2004).

108 Seal BS, Hiett KL, Kuntz RL *et al.* Proteomic analyses of a robust *versus* a poor chicken gastrointestinal colonizing isolate of *Campylobacter jejuni*. *J. Proteome Res.* 6(12), 4582–4591 (2007).

109 Titball RW, Petrosino JF. *Francisella tularensis* genomics and proteomics. *Ann. NY Acad. Sci.* 1105, 98–121 (2007).

110 Chuang MH, Wu MS, Lin JT, Chiou SH. Proteomic analysis of proteins expressed by *Helicobacter pylori* under oxidative stress. *Proteomics* 5(15), 3895–3901 (2005).

111 Walz A, Mujer CV, Connolly JP *et al.* *Bacillus anthracis* secretome time course under host-simulated conditions and identification of immunogenic proteins. *Proteome Sci*. 5, 11 (2007).

112 Guina T, Wu M, Miller SI *et al.* Proteomic analysis of *Pseudomonas aeruginosa* grown under magnesium limitation. *J. Am. Soc. Mass Spectrom*. 14(7), 742–751 (2003).

113 Upritchard HG, Cordwell SJ, Lamont IL. Immunoproteomics to examine cystic fibrosis host interactions with extracellular *Pseudomonas aeruginosa* proteins. *Infect. Immun*. 76(10), 4624–4632 (2008).

114 Manes NP, Gustin JK, Rue J *et al.* Targeted protein degradation by *Salmonella* under phagosome-mimicking culture conditions investigated using comparative peptidomics. *Mol. Cell. Proteomics* 6(4), 717–727 (2007).

115 Nanduri B, Shah P, Ramkumar M *et al.* Quantitative analysis of *Streptococcus pneumoniae* TIGR4 response to *in vitro* iron restriction by 2-D LC ESI MS/MS. *Proteomics* 8(10), 2104–2114 (2008).

116 Chaussee MA, Dmitriev AV, Callegari EA, Chaussee MS. Growth phase-associated changes in the transcriptome and proteome of *Streptococcus pyogenes*. *Arch. Microbiol*. 189(1), 27–41 (2008).

117 Choi YW, Park SA, Lee HW, Kim DS, Lee NG. Analysis of growth phase-dependent proteome profiles reveals differential regulation of mRNA and protein in *Helicobacter pylori*. *Proteomics* 8(13), 2665–2675 (2008).

118 Adkins JN, Mottaz HM, Norbeck AD *et al.* Analysis of the *Salmonella typhimurium* proteome through environmental response toward infectious conditions. *Mol. Cell. Proteomics* 5(8), 1450–1461 (2006).

119 Shi L, Adkins JN, Coleman JR *et al.* Proteomic analysis of *Salmonella enterica* serovar typhimurium isolated from RAW 264.7 macrophages: identification of novel protein that contributes to the replication of serovar typhimurium inside macrophages. *J. Biol. Chem*. 281(39), 29131–29140 (2006).

120 Sriramulu DD, Nimtz M, Romling U. Proteome analysis reveals adaptation of *Pseudomonas aeruginosa* to the cystic fibrosis lung environment. *Proteomics* 5(14), 3712–3721 (2005).

121 Jenner RG, Young RA. Insights into host responses against pathogens from transcriptional profiling. *Nat. Rev. Microbiol*. 3(4), 281–294 (2005).

122 Coiras M, Camafeita E, Lopez-Huertas MR, Calvo E, Lopez JA, Alcami J. Application of proteomics technology for analyzing the interactions between host cells and intracellular infectious agents. *Proteomics* 8(4), 852–873 (2008).

• **Outstanding presentation of the impact of proteomics on understanding eukaryotic cell response to infection on proteome level. The knowledge of the host cell's response to pathogenic microorganisms is of great importance for the generation of new effective prophylactic and therapeutic strategies.**

123 Backert S, Gressmann H, Kwok T *et al.* Gene expression and protein profiling of AGS gastric epithelial cells upon infection with *Helicobacter pylori*. *Proteomics* 5(15), 3902–3918 (2005).

124 Chan CH, Ko CC, Chang JG *et al.* Subcellular and functional proteomic analysis of the cellular responses induced by *Helicobacter pylori*. *Mol. Cell. Proteomics* 5(4), 702–713 (2006).

125 Galindo CL, Fadl AA, Sha J, Pillai L, Gutierrez C Jr, Chopra AK. Microarray and proteomics analyses of human intestinal epithelial cells treated with the Aeromonas hydrophila cytotoxic enterotoxin. *Infect. Immun*. 73(5), 2628–2643 (2005).

126 Ge Z, Zhu YL, Zhong X, Yu JK, Zheng S. Discovering differential protein expression caused by CagA-induced ERK pathway activation in AGS cells using the SELDI-ProteinChip platform. *World J. Gastroenterol*. 14(4), 554–562 (2008).

127 Lenaerts K, Bouwman FG, Lamers WH, Renes J, Mariman EC. Comparative proteomic analysis of cell lines and scrapings of the human intestinal epithelium. *BMC Genomics* 8, 91 (2007).

• **First work showing that the protein profiles of the cell lines routinely used in *in vitro* experiments are not identical to protein profiles of the cells present in the host tissues.**

128 Khemiri A, Galland A, Vaudry D *et al.* Outer-membrane proteomic maps and surface-exposed proteins of *Legionella pneumophila* using cellular fractionation and fluorescent labelling. *Anal. Bioanal. Chem*. 390(7), 1861–1871 (2008).

129 Cordwell SJ, Len AC, Touma RG *et al.* Identification of membrane-associated proteins from *Campylobacter jejuni* strains using complementary proteomics technologies. *Proteomics* 8(1), 122–139 (2008).

130 Hooker BS, Bigelow DJ, Lin CT. Methods for mapping of interaction networks involving membrane proteins. *Biochem. Biophys. Res. Commun*. 363(3), 457–461 (2007).

131 Carlsohn E, Nystrom J, Karlsson H, Svennerholm AM, Nilsson CL. Characterization of the outer membrane protein profile from disease-related *Helicobacter pylori* isolates by subcellular fractionation and nano-LC FT-ICR MS analysis. *J. Proteome Res*. 5(11), 3197–3204 (2006).

132 Alteri CJ, Mobley HL. Quantitative profile of the uropathogenic *Escherichia coli* outer membrane proteome during growth in human urine. *Infect. Immun*. 75(6), 2679–2688 (2007).

133 Hagan EC, Mobley HL. Uropathogenic *Escherichia coli* outer membrane antigens expressed during urinary tract infection. *Infect. Immun*. 75(8), 3941–3949 (2007).

134 Kurupati P, Teh BK, Kumarasinghe G, Poh CL. Identification of vaccine candidate antigens of an ESBL producing *Klebsiella pneumoniae* clinical strain by immunoproteome analysis. *Proteomics* 6(3), 836–844 (2006).

135 Peng X, Ye X, Wang S. Identification of novel immunogenic proteins of *Shigella flexneri* 2a by proteomic methodologies. *Vaccine* 22(21–22), 2750–2756 (2004).

136 Ying T, Wang H, Li M *et al.* Immunoproteomics of outer membrane proteins and extracellular proteins of *Shigella flexneri* 2a 2457T. *Proteomics* 5(18), 4777–4793 (2005).

137 Ying TY, Wang JJ, Wang HL *et al.* Immunoproteomics of membrane proteins of *Shigella flexneri* 2a 2457T. *World J. Gastroenterol*. 11(43), 6880–6883 (2005).

138 Tjalsma H, Lambooy L, Hermans PW, Swinkels DW. Shedding & shaving: disclosure of proteomic expressions on a bacterial face. *Proteomics* 8(7), 1415–1428 (2008).

139 Mattow J, Siejak F, Hagens K *et al.* An improved strategy for selective and efficient enrichment of integral plasma membrane proteins of mycobacteria. *Proteomics* 7(10), 1687–1701 (2007).

140 Zheng J, Wei C, Leng W *et al.* Membrane subproteomic analysis of *Mycobacterium bovis* bacillus Calmette-Guerin. *Proteomics* 7(21), 3919–3931 (2007).

141 Malen H, Berven FS, Softeland T *et al.* Membrane and membrane-associated proteins in Triton X-114 extracts of *Mycobacterium bovis* BCG identified using a combination of gel-based and gel-free fractionation strategies. *Proteomics* 8(9), 1859–1870 (2008).

142 Bumann D, Aksu S, Wendland M *et al.* Proteome analysis of secreted proteins of the gastric pathogen *Helicobacter pylori*. *Infect. Immun*. 70(7), 3396–3403 (2002).

143 Chaussee MA, McDowell EJ, Chaussee MS. Proteomic analysis of proteins secreted by *Streptococcus pyogenes*. *Methods Mol. Biol.* 431, 15–24 (2008).

144 Hathout Y. Approaches to the study of the cell secretome. *Expert Rev. Proteomics* 4(2), 239–248 (2007).

145 Lamonica JM, Wagner M, Eschenbrenner M *et al.* Comparative secretome analyses of three *Bacillus anthracis* strains with variant plasmid contents. *Infect. Immun*. 73(6), 3646–3658 (2005).

146 Malen H, Berven FS, Fladmark KE, Wiker HG. Comprehensive analysis of exported proteins from *Mycobacterium tuberculosis* H37Rv. *Proteomics* 7(10), 1702–1718 (2007).

147 Smith TG, Lim JM, Weinberg MV, Wells L, Hoover TR. Direct analysis of the extracellular proteome from two strains of *Helicobacter pylori*. *Proteomics* 7(13), 2240–2245 (2007).

148 Kuehn MJ, Kesty NC. Bacterial outer membrane vesicles and the host–pathogen interaction. *Genes Dev*. 19(22), 2645–2655 (2005).

149 Lee EY, Bang JY, Park GW *et al.* Global proteomic profiling of native outer membrane vesicles derived from *Escherichia coli*. *Proteomics* 7(17), 3143–3153 (2007).

150 Little SF. Anthrax vaccines: a development update. *BioDrugs* 19(4), 233–245 (2005).

- **Description of the work of several laboratory groups searching for alternative effective and well-tolerated vaccines against anthrax.**

151 Galka F, Wai SN, Kusch H *et al.* Proteomic characterization of the whole secretome of *Legionella pneumophila* and functional analysis of outer membrane vesicles. *Infect. Immun*. 76(5), 1825–1836 (2008).

152 Ferrari G, Garaguso I, Adu-Bobie J *et al.* Outer membrane vesicles from group B *Neisseria meningitidis* δ gna33 mutant: proteomic and immunological comparison with detergent-derived outer membrane vesicles. *Proteomics* 6(6), 1856–1866 (2006).

153 Vipond C, Suker J, Jones C, Tang C, Feavers IM, Wheeler JX. Proteomic analysis of a meningococcal outer membrane vesicle vaccine prepared from the group B strain NZ98/254. *Proteomics* 6(11), 3400–3413 (2006).

154 Diaz-Mejia JJ, Babu M, Emili A. Computational and experimental approaches to chart the *Escherichia coli* cell-envelope-associated proteome and interactome. *FEMS Microbiol. Rev*. 33(1), 66–97 (2009).

155 Weiner JH, Li L. Proteome of the *Escherichia coli* envelope and technological challenges in membrane proteome analysis. *Biochim. Biophys. Acta* 1778(9), 1698–1713 (2008).

156 Grabenstein JD. Vaccines: countering anthrax: vaccines and immunoglobulins. *Clin. Infect. Dis*. 46(1), 129–136 (2008).

157 Leppla SH, Robbins JB, Schneerson R, Shiloach J. Development of an improved vaccine for anthrax. *J. Clin. Invest*. 110(2), 141–144 (2002).

158 Wang JY, Roehrl MH. Anthrax vaccine design: strategies to achieve comprehensive protection against spore, bacillus, and toxin. *Med. Immunol*. 4(1), 4 (2005).

159 Read TD, Peterson SN, Tourasse N *et al.* The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. *Nature* 423(6935), 81–86 (2003).

160 Okinaka RT, Cloud K, Hampton O *et al.* Sequence and organization of pXO1, the large *Bacillus anthracis* plasmid harboring the anthrax toxin genes. *J. Bacteriol.* 181(20), 6509–6515 (1999).

161 Pannucci J, Okinaka RT, Williams E, Sabin R, Ticknor LO, Kuske CR. DNA sequence conservation between the *Bacillus anthracis* pXO2 plasmid and genomic sequence from closely related bacteria. *BMC Genomics* 3(1), 34 (2002).

162 Ariel N, Zvi A, Grosfeld H *et al.* Search for potential vaccine candidate open reading frames in the *Bacillus anthracis* virulence plasmid pXO1: *in silico* and *in vitro* screening. *Infect. Immun*. 70(12), 6817–6827 (2002).

163 Ariel N, Zvi A, Makarova KS *et al.* Genome-based bioinformatic selection of chromosomal *Bacillus anthracis* putative vaccine candidates coupled with proteomic identification of surface-associated antigens. *Infect. Immun*. 71(8), 4563–4579 (2003).

164 Chitlaru T, Gat O, Gozlan Y, Ariel N, Shafferman A. Differential proteomic analysis of the *Bacillus anthracis* secretome: distinct plasmid and chromosome CO2-dependent cross talk mechanisms modulate extracellular proteolytic activities. *J. Bacteriol.* 188(10), 3551–3571 (2006).

165 Brossier F, Levy M, Mock M. Anthrax spores make an essential contribution to vaccine efficacy. *Infect. Immun.* 70(2), 661–664 (2002).

166 Delvecchio VG, Connolly JP, Alefantis TG *et al.* Proteomic profiling and identification of immunodominant spore antigens of *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*. *Appl. Environ. Microbiol*. 72(9), 6355–6363 (2006).

167 Kudva IT, Griffin RW, Garren JM, Calderwood SB, John M. Identification of a protein subset of the anthrax spore immunome in humans immunized with the anthrax vaccine adsorbed preparation. *Infect. Immun*. 73(9), 5685–5696 (2005).

168 Destefano F, Pfeifer D, Nohynek H. Safety profile of pneumococcal conjugate vaccines: systematic review of pre- and post-licensure data. *Bull. World Health Organ*. 86(5), 373–380 (2008).

169 23-valent pneumococcal polysaccharide vaccine. WHO position paper. *Wkly Epidemiol. Rec*. 83(42), 373–384 (2008).

170 Cao J, Chen D, Xu W *et al.* Enhanced protection against pneumococcal infection elicited by immunization with the combination of PspA, PspC, and ClpP. *Vaccine* 25(27), 4996–5005 (2007).

171 Csordas FC, Perciani CT, Darrieux M *et al.* Protection induced by pneumococcal surface protein A (PspA) is enhanced by conjugation to a *Streptococcus pneumoniae* capsular polysaccharide. *Vaccine* 26(23), 2925–2929 (2008).

172 Curtiss R 3rd, Zhang X, Wanda S *et al.* Induction of host immune response using *Salmonella*-vector vaccine. In: *Virulence Mechanisms of Bacterial Pathogens.* Teb KA (Ed.). ASM Press, Washington, DC, USA, 297–313 (2007).

173 Hamel J, Charland N, Pineau I *et al.* Prevention of pneumococcal disease in mice immunized with conserved surface-accessible proteins. *Infect. Immun*. 72(5), 2659–2670 (2004).

174 Kwon HY, Ogunniyi AD, Choi MH, Pyo SN, Rhee DK, Paton JC. The ClpP protease of *Streptococcus pneumoniae* modulates virulence gene expression and protects against fatal pneumococcal challenge. *Infect. Immun*. 72(10), 5646–5653 (2004).

175 Ogunniyi AD, Grabowicz M, Briles DE, Cook J, Paton JC. Development of a vaccine against invasive pneumococcal disease based on combinations of virulence proteins of *Streptococcus pneumoniae*. *Infect. Immun*. 75(1), 350–357 (2007).

176 Giefing C, Meinke AL, Hanner M *et al.* Discovery of a novel class of highly conserved vaccine antigens using genomic scale antigenic fingerprinting of pneumococcus with human antibodies. *J. Exp. Med*. 205(1), 117–131 (2008).

177 Meinke A, Henics T, Hanner M, Minh DB, Nagy E. Antigenome technology: a novel approach for the selection of bacterial vaccine candidate antigens. *Vaccine* 23(17–18), 2035–2041 (2005).

178 Cunningham MW. Autoimmunity and molecular mimicry in the pathogenesis of post-streptococcal heart disease. *Front. Biosci*. 8, s533–543 (2003).

179 Guilherme L, Fae KC, Higa F *et al.* Towards a vaccine against rheumatic fever. *Clin. Dev. Immunol.* 13(2–4), 125–132 (2006).

180 Musser JM, DeLeo FR. Toward a genome-wide systems biology analysis of host-pathogen interactions in group A *Streptococcus*. *Am. J. Pathol*. 167(6), 1461–1472 (2005).

181 Moura H, Woolfitt AR, Carvalho MG *et al.* MALDI-TOF mass spectrometry as a tool for differentiation of invasive and noninvasive *Streptococcus pyogenes* isolates. *FEMS Immunol. Med. Microbiol*. 53(3), 333–342 (2008).

182 Batzloff MR, Pandey M, Olive C, Good MF. Advances in potential M-protein peptide-based vaccines for preventing rheumatic fever and rheumatic heart disease. *Immunol. Res*. 35(3), 233–248 (2006).

183 Olive C. Progress in M-protein-based subunit vaccines to prevent rheumatic fever and rheumatic heart disease. *Curr. Opin. Mol. Ther*. 9(1), 25–34 (2007).

184 Shaila MS, Nayak R, Prakash SS *et al.* Comparative *in silico* analysis of two vaccine candidates for group A *Streptococcus* predicts that they both may have similar safety profiles. *Vaccine* 25(18), 3567–3573 (2007).

185 McMillan DJ, Batzloff MR, Browning CL *et al.* Identification and assessment of new vaccine candidates for group A streptococcal infections. *Vaccine* 22(21–22), 2783–2790 (2004).

186 Schmidt F, Schmid M, Thiede B, Pleissner KP, Bohme M, Jungblut PR. Assembling proteomics data as a prerequisite for the analysis of large scale experiments. *Chem. Cent. J*. 3(1), 2 (2009).

187 Young D, Dye C. The development and impact of tuberculosis vaccines. *Cell* 124(4), 683–687 (2006).

188 Andersen P, Doherty TM. TB subunit vaccines – putting the pieces together. *Microbes Infect*. 7(5–6), 911–921 (2005).

189 Gayathri P, Balaram H, Murthy MR. Structural biology of plasmodial proteins. *Curr. Opin. Struct. Biol.* 17(6), 744–754 (2007).

190 McMurry J, Sbai H, Gennaro ML, Carter EJ, Martin W, De Groot AS. Analyzing *Mycobacterium tuberculosis* proteomes for candidate vaccine epitopes. *Tuberculosis (Edinb)* 85(1–2), 95–105 (2005).

191 Sable SB, Verma I, Khuller GK. Multicomponent antituberculous subunit vaccine based on immunodominant antigens of *Mycobacterium tuberculosis*. *Vaccine* 23(32), 4175–4184 (2005).

192 Vani J, Shaila MS, Chandra NR, Nayak R. A combined immuno-informatics and structure-based modeling approach for prediction of T cell epitopes of secretory proteins of *Mycobacterium tuberculosis*. *Microbes Infect*. 8(3), 738–746 (2006).

193 Livny J, Waldor MK. Identification of small RNAs in diverse bacterial species. *Curr. Opin. Microbiol*. 10(2), 96–101 (2007).

194 Romby P, Vandenesch F, Wagner EG. The role of RNAs in the regulation of virulence-gene expression. *Curr. Opin. Microbiol.* 9(2), 229–236 (2006).

195 Pichon C, Felden B. Proteins that interact with bacterial small RNA regulators. *FEMS Microbiol. Rev*. 31(5), 614–625 (2007).

196 Brennan RG, Link TM. Hfq structure, function and ligand binding. *Curr. Opin. Microbiol*. 10(2), 125–133 (2007).

197 Wilson JW, Ott CM, Honer zu Bentrup K *et al.* Space flight alters bacterial gene expression and virulence and reveals a role for global regulator Hfq. *Proc. Natl Acad. Sci. USA* 104(41), 16299–16304 (2007).

Websites

201 J. Craig Venter Institute www.tigr.org/tigr-script/CMRHomePage.spl

202 GOLD: Genomes online database V. 2.0 www.genomesonline.org/

203 Drugs Information Online. Clinical trial results www.drugs.com/clinical_trials

204 Group A streptococcal vaccine development: current status and issues of relevance to less developed countries (WHO) www.who.int/child_adolescent_health/documents/ivb_05_14/en/index.html

205 Stop TB Partnership and WHO. 2006. The Global Plan to Stop TB, 2006–2015, WHO/HTM/STB/2006.35 www.who.int/tb/publications/global_plan_to_stop_tb/en/index.html

Affiliations

- Elżbieta Katarzyna Jagusztyn-Krynicka
Full Professor, Department of Bacterial Genetics, Faculty of Biology, University of Warsaw, Miecznikowa 1, 02–096 Warsaw, Poland
Tel.: +48 225 541 216
Fax: +48 225 541 402
kjkryn@biol.uw.edu.pl
- Michał Dadlez
Full Professor, Institute of Biochemistry and Biophysics, Polish Academy of Science, Pawińskiego 5a, 02–106 Warsaw; Institute of Genetics and Biotechnology, Biology Department, University of Warsaw, Pawińskiego 5a, 02–106 Warsaw, Poland
Tel.: +48 225 923 471
michald@ibb.waw.pl
- Anna Grabowska
PhD Student, Department of Bacterial Genetics, Faculty of Biology, University of Warsaw, Miecznikowa 1, 02–096 Warsaw, Poland
Tel.: +48 225 541 411
Fax: +48 225 541 402
aniagra@biol.uw.edu.pl
- Paula Roszczenko
PhD Student, Interdepartmental Interdisciplinary PhD program in the Natural Sciences, Department of Bacterial Genetics, Faculty of Biology, University of Warsaw, Miecznikowa 1, 02–096 Warsaw, Poland
Tel.: +48 225 541 411
Fax: +48 225 541 402
p.roszczenko@biol.uw.edu.pl