Rational Vaccine Design
The Power of Proteomics

Executive Summary

Traditionally the development of vaccines to protect against bacterial diseases has been empirical, relying on testing killed or weakened preparations of whole bacteria, or specific molecules from them (antigens), to see if they could generate an immune response in humans. However, recent developments in biology, such as the emergence of genomics (the study of the complete genetic material of an organism) and proteomics (the study of all the proteins and protein interactions of an organism), have raised the possibility that, in future, vaccines may be designed by more rational, predictive methods.

The discovery and development of these rational vaccines is likely to be accomplished through integrated proteomic strategies. Most proteomic studies are based on a technique for separating proteins called two-dimensional polyacrylamide gel electrophoresis (2D-PAGE). The 2D-PAGE technique, in combination with another technique called Western blotting, has been successfully applied in the discovery of potential vaccine candidates from three medically important bacteria: *Helicobacter pylori* (a cause of peptic ulcers), *Chlamydia trachomatis* (a cause of genital and eye infections) and *Borrelia garinii* (a cause of Lyme disease). In addition, new methods have been developed within the past two years that provide complementary information concerning protein production by bacteria. For example, two-dimensional semi-preparative electrophoresis has provided complementary information regarding membrane protein expression in a strain of *H. pylori*. Through another new technique, two-dimensional liquid chromatography-tandem mass spectrometry, the most comprehensive information to date regarding protein expression in yeast has been obtained. The latter technique may shortly become an important tool in vaccinology.

This review of the current state of bacterial proteomics as applied in vaccine development presents analytical techniques for protein separation, proteomics without gels, reverse vaccinology, and functional approaches to the identification of proteins in bacteria. These techniques may assume increasing importance for rational design of vaccines. However, because applied protemics is a very new field, it is more likely to have a long-term effect rather than make any short-term impact. Indeed, nearly all rationally designed bacterial vaccines are at the discovery or research stages of development and, given that they will need to undergo years of clinical trials, it does not seem likely that one will enter the marketplace in the immediate future.

The aim of vaccination is to generate immune responses that prevent bacterial infection. This is achieved by generating antibodies and immune system cells that destroy bacteria before they can cause an infection. Most vaccines available today were developed in the era prior to genome sequencing projects and have been developed empirically through testing of the immune response to bacterial virulence factors, such as proteins, that can be easily identified and purified. However, the elucidation of the complete genetic sequences (genomes) of several bacteria presents opportunities to develop new vaccines in a more rational manner. The genome of an organism predicts its entire repertoire of proteins, the proteome, within which potential vaccine candidates may be discovered. However, proteomics includes not only the catalogue of proteins expressed by an organism at a given time, but also includes quantitative information and the identification of proteins' interacting partners (carbohydrates, nucleic acids and other proteins). Thus, proteomic studies yield an enormous wealth of information.

The challenge of applying proteomics for vaccine development is to target relevant subsets of this information; i.e. identify the proteins most likely to be vaccine candidates, such as membrane proteins, toxins and adhesins. This review presents current and emerging technologies for bacterial proteomics and an evaluation of their applicability in vaccine design and development.

The Cornerstone of Proteomics: Two-Dimensional Electrophoresis

In order to characterize a proteome, it is first necessary to separate it into its individual proteins; only then can identification and characterization of the proteins and their interactions take place. Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) is by far the most popular protein-separation technique used in

The full text of this article was published in the American Journal of Pharmacogenomics 2002; 2(1): 59-65. Adis International. The original article was written by Dr Carol Nilsson from the Institute of Medical Biochemistry, Gothenberg University, Gothenberg, Sweden.

proteomics. There are several reasons for this, most importantly:

- the separating ability of 2D-PAGE is high, and a map of an organism's proteome can be easily obtained and compared with other 2D gels.
- the amounts of protein that can be separated are compatible with the amounts required for analysis by mass spectrometry or N-terminal sequencing, techniques that provide experimental data that can be used to help identify a protein.

In 2D-PAGE, proteins are initially separated on the basis of their different electrical charges, in a procedure called isoelectric focussing (IEF); this is the first dimension of separation. In the second dimension, the proteins are further separated on the basis of their sizes by "sieving" through a porous gel, in which smaller proteins travel faster than larger ones; this is termed sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE).

Through 2D-PAGE, thousands of unique proteins can be separated and detected by various means from the original complex mixture. By following the SDS-PAGE step with in-gel digestion of proteins, typically using the enzyme trypsin, fragments of protein (peptides) are generated and their physical size measured. Then, it is often possible to identify a protein by using bioinformatics software to compare the actual peptides generated with those generated *in silico* from theoretical digests of proteins in protein databases. If this approach does not conclusively identify the protein, sequence data from the peptides may be gathered by tandem mass spectrometry (MS/MS) and used to identify the protein.

The proteomic approach employing 2D-PAGE as a separation technique is well-suited to studies of protein expression because of its high sensitivity, high resolving power and reproducibility. There are, however, some important selective disadvantages to the technique that limit its usefulness for rational vaccine design. Many types of proteins are under-represented in 2D gels, including:

- hydrophobic proteins (i.e. membrane proteins)
- proteins present in only small amounts
- small proteins
- proteins with a net positive charge (highly basic proteins)

Unfortunately, as Murphy's Law would predict, several of these protein types are interesting to those dedicated to discovering new vaccines. For example, the outer membrane proteins of bacteria, also called adhesins, play a decisive role in initial adhesion to, and infection of, host cells. Thus, hypothetically, a vaccine based on these proteins may be very effective. Besides being membrane proteins, adhesins are also often present only in small amounts, further reducing their chances of being detected by 2D-PAGE. For instance, the Lewis^b-binding adhesin of *Helicobacter pylori* is produced at a level of only 500 copies per cell, an extraordinarily low concentration. Not surprisingly, this protein has never been detected by 2D-PAGE. Thus, the 2D-PAGE approach cannot be used as the only tool to investigate global protein expression; i.e. the lack of recovery of a given protein in a 2D gel does not necessarily mean that it is not produced. The largest number of proteins identified to date is 504 from a strain of *Haemophilus influenzae,* covering about one-third of this organism's predicted proteome. The study required pre-enrichment techniques to enhance the detection of certain types of proteins, such as those present in small amounts.

2D-PAGE: a useful but limited technology for identifying vaccine candidates

Despite the shortcomings of 2D gel separations, they have been applied with some success in the study of bacterial proteomes. However, studies have mostly been unsuccessful in identifying many proteins from the cell wall of bacteria probably due to the discrimination against these proteins by the 2D-PAGE separation technique. Unfortunately, these extracellular proteins are very likely to be useful vaccine candidates. Therefore, considerable efforts have been made to improve the recovery of hydrophobic proteins from 2D gels, but success has been limited.

Because of the limitations of 2D-PAGE, new technologies have been developed in the past few years that complement the results provided by gels. These technologies are designed to eliminate the systematic discrimination of important proteins that occurs by 2D-PAGE. They are based on using various types of liquid separations (e.g. liquid-phase IEF or chromatography), then identifying and characterizing the proteins by a technique called mass spectrometry. Unfortunately, the techniques possess a lower resolving power than 2D-PAGE and are generally not suitable for visualizing an entire proteome.

Complementary Proteomic Technologies

Two-dimensional semi-preparative electrophoresis

Two-dimensional semi-preparative electrophoresis is a separation technique that is based on the same

principles as analytical 2D-PAGE; i.e. separation of proteins by electrical charge (IEF) and molecular size (gel electrophoresis). In the first-dimensional separation, proteins are enriched up to 500-fold in liquid fractions by IEF. In the second dimension, IEF-enriched proteins are separated by SDS-PAGE then removed from the gel into liquid fractions, digested by enzymes and analyzed by mass spectrometry. The semi-preparative electrophoresis procedure has several advantages over analytical 2D-PAGE.

- It allows for high protein loads, and thus enables the detection of proteins that are naturally present only in small amounts.
- Because the proteins remain in a liquid solution, they are accessible to study by Western blotting and/or mass spectrometry after both the IEF- and size-separation steps.
- The procedure can be performed in about one day, making it more rapid than separation by 2D-PAGE.
- The separation technique does not discriminate against membrane proteins.

Two-dimensional semi-preparative electrophoresis has been applied in studies of *H. pylori* and *Escherichia coli*. Of the 40 proteins identified, more than one-third were cell envelope proteins. The Lewis^b-binding adhesin was among the proteins identified. In a study of a strain of *E. coli* that causes bladder infections, 30 proteins were identified, including a membrane protein.

Two-dimensional liquid chromatography-tandem mass spectrometry (2D-LC-MS/MS)

Because gel separations of proteins can be time-consuming, labor-intensive and may not detect interesting classes of proteins, including potential vaccine candidates, some investigators are inventing proteomic technologies that completely avoid the use of gels. One such method is 2D-liquid chromatography-tandem mass spectrometry (2D-LC-MS/MS). In this technique, proteins from an entire proteome or cellular subcompartment are cleaved into peptides and then separated by a two-dimensional liquid chromatography system which is linked to a mass spectrometer. The data collected by mass spectrometry is used to identify the proteins, via searching databases of known proteins.

This technique was recently applied to the yeast proteome. The investigation yielded 1484 proteins, including membrane proteins, proteins present in very small amounts and basic proteins. These are the very types of proteins that are often not detected by 2D-PAGE. In total, 131 membrane proteins were identified. Therefore, 2D-LC-MS/MS appears to be the most comprehensive technology to date for monitoring protein expression. However, a disadvantage of the technology is that no information on a protein's electrical charge or physical size can be acquired, because the intact protein is chopped up into peptides before separation and analysis. Although no applications of the 2D-LC-MS/MS technique have yet been described in vaccinology, it is likely to be valuable for proteomic studies of dangerous bacteria and their production of virulence proteins, which may be vaccine candidates.

Emerging Proteomic Technologies for Vaccine Development

Top-down proteomics

The measurement of peptides derived from proteins in order to obtain the identity of the protein, such as in 2D-LC-MS/MS, is often referred to as a 'bottom-up' approach. Conversely, it is also possible to perform mass spectrometry on intact proteins. This is called 'top-down' proteomics and it requires a special kind of mass spectrometer with the succinct name of Fourier transform ion cyclotron resonance mass spectrometer (FTICR). Using this top-down technology, both the physical size of the intact protein and its amino-acid sequence are obtained, which normally means easy identification of the protein by searching protein databases. Protein modifications can also be readily identified. The top-down proteomics technique has been applied to *Methanococcus jannaschii* and *Mycoplasma pneumoniae*. With the recent increase in the number of FTICR mass spectrometers installed worldwide, top-down proteomics looks set to become an important analytical technique in future studies of proteins isolated from dangerous bacteria.

Reverse vaccinology

In traditional vaccinology, development of vaccines to protect against bacterial diseases is empirical, relying on testing killed or weakened preparations of whole bacteria, or specific molecules from them (antigens), to see if they could generate an immune response in humans. However, in "reverse vaccinology", a bacterial genome is searched by bioinformatic software to identify genes that produce proteins present on the surface of the bacteria. Once identified, candidate proteins can be produced by recombinant DNA technology and tested for their ability to generate

immune responses. Obviously, in order to use this approach, the genome from the bacterium of interest must have been sequenced.

Using this approach, a putative vaccine candidate, membrane lipo-protein P6, was identified in *H. influenzae,* a major cause of respiratory tract infections. In addition, a large-scale effort to identify vaccine candidates in *Neisseria meningitidis*, the cause of the frequently fatal meningococcal disease, was recently described. This effort identified 570 genes encoding hypothetical proteins, of which 350 candidate proteins were produced in *E. coli* and tested for their ability to generate immune responses in mice.

Thus, it is clear that reverse vaccinology is a powerful and promising technique for rational design of vaccines against dangerous bacteria. However, a major drawback with the technology is that it requires industrial-scale expression of membrane proteins, which is a difficult and costly undertaking. Therefore, advances in production and purification of membrane proteins, currently an area of intense research, will be helpful in this respect.

The two-hybrid approach

In the real world, proteins do not exist on their own and it is interactions between them that are a key determinant of what the proteins actually do in living creatures. Therefore, new techniques to study protein interactions and functions on a proteome-wide scale are being developed. The determination of protein-interaction maps for entire bacteria can be achieved through implementation of a technique called a two-hybrid assay. Although not a new technique, it was only recently that the first two-hybrid study for a bacterial species was published. Using this system, it was possible to identify biological pathways and assign functions to *H. pylori* proteins, including proteins involved in chemical signalling, cell division and assembly of an important protein complex. The two-hybrid approach promises to provide complementary information to protein functions that might be difficult to obtain through studies involving protein identification only.

Identification of membrane lectins

The first step in bacterial infection is the adhesion of the bacteria to the host cells. This occurs through molecular interaction between bacterial membrane proteins (lectins) and carbohydrate compounds on the surface of host cells. Therefore, identification of relevant lectins is a high priority in the search for effective vaccines against infections. New analytical techniques have been developed to identify lectins with a known carbohydrate affinity by selectively enriching the proteins and identifying them against known genome sequences. This type of affinity enrichment technique is also compatible with mass spectrometry, which increases the sensitivity of analysis 1000-fold. This experimental design enables the identification of virulence proteins that are difficult or impossible to detect by 2D-PAGE.

Conclusions

During recent years, the complete genomes of approximately 40 bacteria have been sequenced. Technological advances have been made in the separation and identification of bacterial proteins against known genomic sequences, enabling the description of entire proteomes. Individual proteins may be selectively identified, produced and tested for their ability to generate immune responses. One of the separation techniques, analytical 2D-PAGE, is currently widely used in proteomic studies. The limitations of this technology, especially in the separation of proteins important in vaccinology, have forced the development of new technologies for the study of bacterial proteomes. Some of the newest analytical technologies, for instance 2D-LC-MS/MS and top-down proteomics, have not yet been used to specifically study candidate vaccine proteins. But it seems likely that they will become important technologies for rational design of vaccines. However, because no single technology can be claimed to provide comprehensive information about a given bacterial species, the future success of rational vaccine design is likely to depend on combinations of methods to study functional protein expression and interaction pathways in dangerous bacteria. ■