# The Genome Revolution in Vaccine Research

**Barbara Capecchi, Davide Serruto, Jeannette Adu-Bobie, Rino Rappuoli\* and Mariagrazia Pizza**

IRIS, Chiron S.r.l.,Via Fiorentina 1, 53100 Siena, Italy

## Abstract

**The conventional approach to vaccine development is based on dissection of the pathogen using biochemical, immunological and microbiological methods. Although successful in several cases, this approach has failed to provide a solution to prevent several major bacterial infections. The availability of complete genome sequences in combination with novel advanced technologies, such as bioinformatics, microarrays and proteomics, have revolutionized the approach to vaccine development and provided a new impulse to microbial research. The genomic revolution allows the design of vaccines starting from the prediction of all antigens *in silico*, independently of their abundance and without the need to grow the pathogen *in vitro*. This new genome-based approach, which we have named "Reverse Vaccinology", has been successfully applied for *Neisseria meningitidis* serogroup B for which conventional strategies have failed to provide an efficacious vaccine. The concept of "Reverse Vaccinology" can be easily applied to all the pathogens for which vaccines are not yet available and can be extended to parasites and viruses.**

## Introduction

Since its introduction 200 years ago, vaccination has prevented illness and death for millions of individuals every year. All available vaccines can be divided into two main categories: living and non-living vaccines. The first category uses live-attenuated organisms. These vaccines are able to induce a strong immune response, however, there is a possibility they can revert to the virulent form at any time. Differently, non-living vaccines are based on whole killed pathogens or components of them (subunit vaccines) as capsule, partially purified toxins or polysaccharides (Table 1). These vaccines are very efficacious and allowed the control and, in some cases, the eradication of very important infectious diseases such as smallpox and polio, at least in industrialised countries.

The advent of modern molecular biology and microbiology techniques, as well as the increasing knowledge of the pathogenesis of many bacteria, introduced the first important innovation in the vaccines field. This approach gave rise to the production of subunit vaccines based on specific protective antigens. Two significant examples are the recombinant vaccine against Hepatitis B virus, based on highly purified envelope protein

(Andre, 1990), and the acellular vaccine against whooping cough, based on three highly purified proteins of *Bordetella pertussis*, including a genetically detoxified toxin (Greco *et al.*, 1996; Pizza *et al.*, 1989).

All these conventional approaches to produce vaccines are based on the cultivation of the microorganism in laboratory conditions from which single components are isolated individually by using biochemical, microbiological and serological methods. Each antigen is produced in pure form either directly from the bacterium or using the recombinant DNA technology, and finally tested for its ability to induce an immune response. However, although successful in many cases, this approach presents several limitations. This approach needs to grow the pathogen *in vitro*, so it is not applicable to non-cultivable microorganisms, and in many cases the antigens expressed during infection are not produced in laboratory conditions. Moreover, the proteins that are most abundant and easily purified are not necessarily protective antigens and, in any case, only a few molecules can be isolated and tested simultaneously. In conclusion, this method can employ many years to identify a protective and useful antigen, and has failed to provide a vaccine against those pathogens that did not have obvious immunodominant protective antigens (*i.e.* capsule or toxins).

The genome era has completely changed the way to design vaccines. The availability of the complete genome of microorganisms combined with a novel advanced technology has introduced a new prospective in vaccine research. It is now possible to determine the complete genome sequence of a bacterial pathogen in a very few months at very low cost. In 1995, The Institute for Genomic Research (TIGR) published the first microbial complete genome sequence of *Haemophilus influenzae* (Fleischmann *et al.*, 1995). So far, approximately one hundred complete genomes of pathogens and non-pathogens are available, and more than 150 other microorganisms are being sequenced in various laboratories around the world (www.tigr.org/tdb).

As genome sequences become available, the ability to compare related bacteria, pathogens versus commensals of the same or related species and even bacteria with different or similar pathogenic profiles is providing basis for new concepts of bacterial pathogenesis. The commercial and medical importance of identifying genes that mediate microbial diseases make pathogens attractive candidates for genome-sequencing projects and the list of eubacterial genomes is now dominated by pathogenic species (Field *et al.*, 1999). The use of sophisticated bioinformatic tools is essential to unlock and interpret the huge amount of information contained in a microbial whole genome sequence (genomic mining). One of the most interesting applications of the genome analysis of pathogenic bacteria is the design of new vaccines and antimicrobial molecules. For the first time, the genome sequences represent an inclusive virtual catalogue of all the potential vaccine candidates from which it is possible

*For correspondence. Email rino_rappuoli@chiron.it.

Table 1. List of existing vaccines used in humans.
These data are derived from Levine *et al.*, 1997, as well as the web site www.niaid.nih.gov/dmid/vaccines/jordan20/.

| | Live-attenuated vaccines | Killed-inactivated vaccines | Subunit (acellular) vaccines | | | |
| --- | --- | --- | --- | --- | --- | --- |
| | | | Toxoid | Conjugate vaccines | Capsular polysaccharides | Recombinant proteins |
| Adenovirus | ✓ | | | | | |
| Anthrax | | ✓ | | | | |
| Cholera | ✓ | ✓ | | | | |
| Diphtheria | | | ✓ | | | |
| *Haemophilus influenzae* | | | | ✓ | | |
| Hepatitis A | | ✓ | | | | |
| Hepatitis B | | | | | | ✓ |
| Influenza virus | ✓ | ✓ | | | | |
| Japanese encephalitis | ✓ | ✓ | | | | |
| Lyme disease | | | | | | ✓ |
| Measles | ✓ | | | | | |
| Meningococcus A | | | | | ✓ | |
| Meningococcus C | | | | ✓ | ✓ | |
| Mumps | ✓ | | | | | |
| Pertussis | | ✓ | ✓ | | | ✓ |
| Plague | | ✓ | | | | |
| Pneumococcus | | | | ✓ | ✓ | |
| Polio | ✓ | ✓ | | | | |
| Rabies | ✓ | ✓ | | | | |
| Rotavirus | ✓ | | | | | |
| Rubella | ✓ | | | | | |
| Smallpox | ✓ | | | | | |
| Tetanus | | | ✓ | | | |
| Tickborne encephalitis | | ✓ | | | | |
| Tuberculosis | ✓ | | | | | |
| Typhoid | ✓ | ✓ | | | ✓ | |
| Varicella | ✓ | | | | | |
| Yellow fever | ✓ | | | | | |

to select the molecules that are likely more effective, regardless of their abundance, or if they are expressed *in vivo* or *in vitro* conditions. In fact, current computer programs allow to predict the function or the putative cellular localization of the newly-identified open reading frames (ORFs). This revolutionary approach to vaccine research has been named "Reverse Vaccinology" indicating that, differently from conventional vaccinology, the starting-point for vaccine design is the *in silico* analysis of the genome sequences and not longer the live bacterium (Rappuoli, 2000a). However, even if bioinformatics could provide many information about the new potential vaccine candidates, the *in silico* analysis must to be supported by experimental data.

Genomic information is also used in conjunction with powerful technologies, for istance, *in vivo* expression technology (IVET), signature tagged mutagenesis (STM), DNA microarrays and proteomics (two-dimensional gel electrophoresis and mass spectrometry) in order to select experimentally new virulence factors or novel surface-exposed antigens. All these fields, termed "Functional Genomics", provide us powerful tools to study the genome. Here we illustrate the application of functional genomics to vaccine research and in particular we describe the first example for which the reverse vaccinology approach has been applied to design a vaccine against the human pathogen *Neisseria meningitidis* serogroup B.

## Genomic technologies for vaccine candidates discovery

*Genome sequencing and* in silico *analysis*
DNA sequencing has become a completely automated, high throughput procedure. Basically, random overlapping libraries representing the complete genome in small fragments are prepared in plasmid vectors and the insert of each plasmid is sequenced. Random sequence is determined to an approximate 8-10 fold coverage of the genome. Sophisticated bioinformatic tools then compare the sequences and generate long sequence overlaps known as contigs (from contiguous sequence). After this procedure, the genome is usually found in a relatively small number of large contigs. Closure means filling in the gaps and involves libraries of larger fragments in lambda vectors and cloning of specific PCR products using primer sequences at the extremities of the contigs. The result is the complete genome in one continuous sequence.

Once the nucleotide sequencing is completed and the sequence assembled, potential ORFs are identified and used to search in the databases for homologies to genes having a known function in other organisms. In this way, functions are ascribed to ORFs and the sequence is annotated (Figure 1). Annotation can also be based on the predicted folding of putative proteins (Fischer and Eisenberg, 1997), even if this is still a relatively insensitive
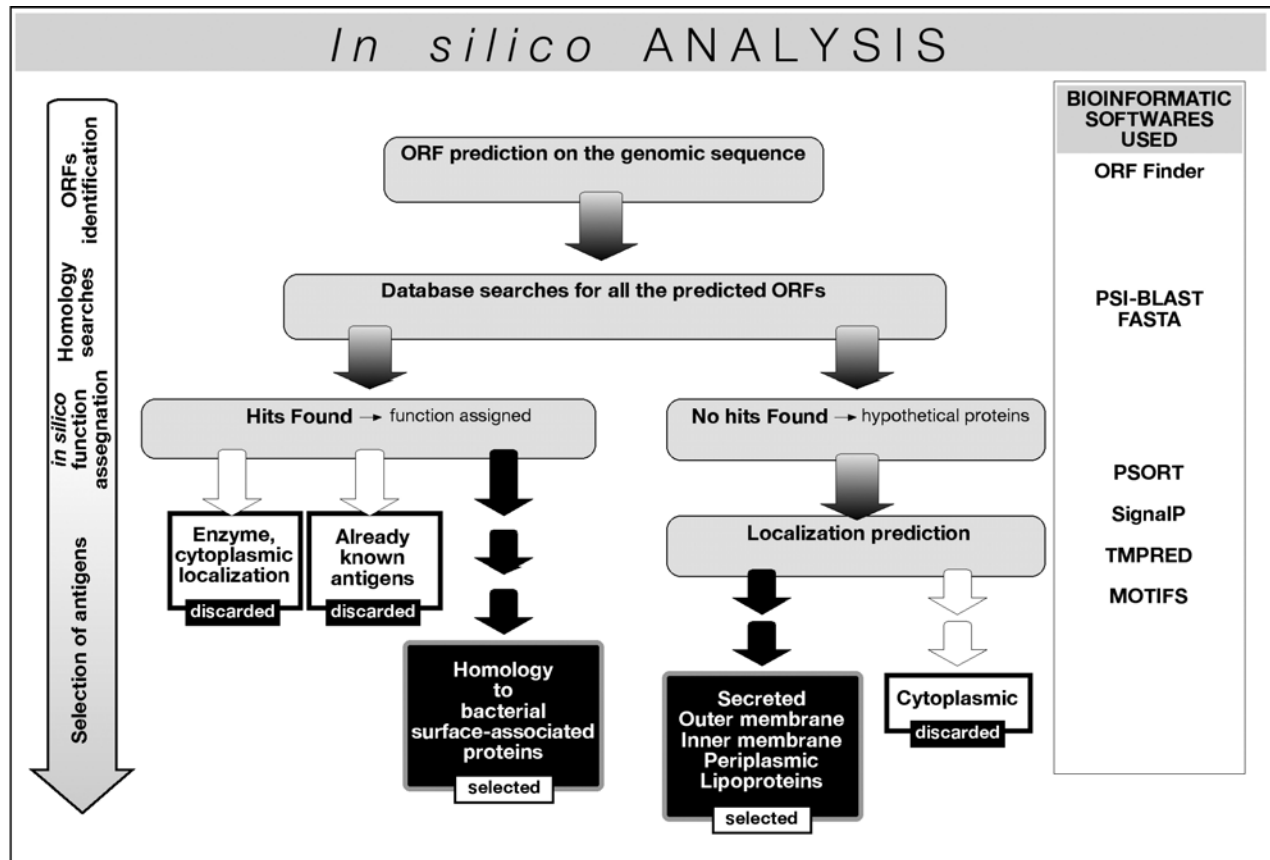
Figure 1. Flow chart showing criteria and bioinformatic softwares for the *in silico* genome analysis and selection of putative vaccine candidates.

method given the limited knowledge on protein structure. For highly conserved gene products involved in fundamental cellular process, the assignment of function is relatively secure. However, in same cases, information from database searches can be potentially misleading.

The computational software aims to assign a function to each gene. In the search for virulence genes, the most common approach is that of comparing predicted coding sequences with sequences present in databases using programs such as BLAST (Altschul *et al.*, 1997), which identify matches to known genes (Figure 1). A variety of software packages are currently accessible to assign functions to genes and predict key features such as cellular localization, topology, molecular weight, pI and solubility.

The primary condition for a bacterial protein to be considered as an antigen is its cellular localisation. In fact, proteins restricted to the cytosolic compartment are not good immunological targets, whereas surface-associated or secreted proteins are more accessible to antibody binding and are therefore considered as possible targets for the immune system.

Several computational methods are available to search for surface-associated or secreted proteins: PSORT is used for the prediction of protein sorting signals and localization sites in amino acid sequences (http://psort.nibb.ac.jp/); SignalP predicts the presence and location of signal peptide cleavage sites in amino acid sequences from different

organisms, Gram-positive and Gram-negative prokaryotes, and eukaryotes (http://www.cbs.dtu.dk/services/SignalP/); TMpred program makes a prediction of membrane-spanning regions and their orientation (http://www.ch.embnet.org/software/TMPRED_form.html). Other useful signatures for membrane-associated proteins is the presence of aromatic residues at the C-terminus of the proteins (Struyve *et al.*, 1991) and the Leu(Ala/Val)-Leu-Ala(Ser)-Gly(Ala)-Cys motif at the N-terminus of lipoproteins (Hayashi and Wu, 1990). Although much progress can be made *in silico*, the experimental approach is necessary to establish unambiguously the role of gene products in living bacteria.

*Studying gene expression in vivo: STM and IVET*
The search for virulence-associated factors of bacterial pathogens is considered to be a key step in vaccine discovery. A valuable way to study virulence factors is the use of *in vivo* experimental models. In the last five years, a variety of methods have been identified and described to isolate genes that are specifically induced during infection. These include STM and IVET (Chiang *et al.*, 1999; Shea *et al.*, 2000; Tang and Holden, 1999). STM, developed by David Holden, is an approach based on random mutagenesis to identify genes required for *in vivo* survival (Hensel *et al.*, 1995), and therefore, a powerful tool for identifying potential virulence factors. This negative
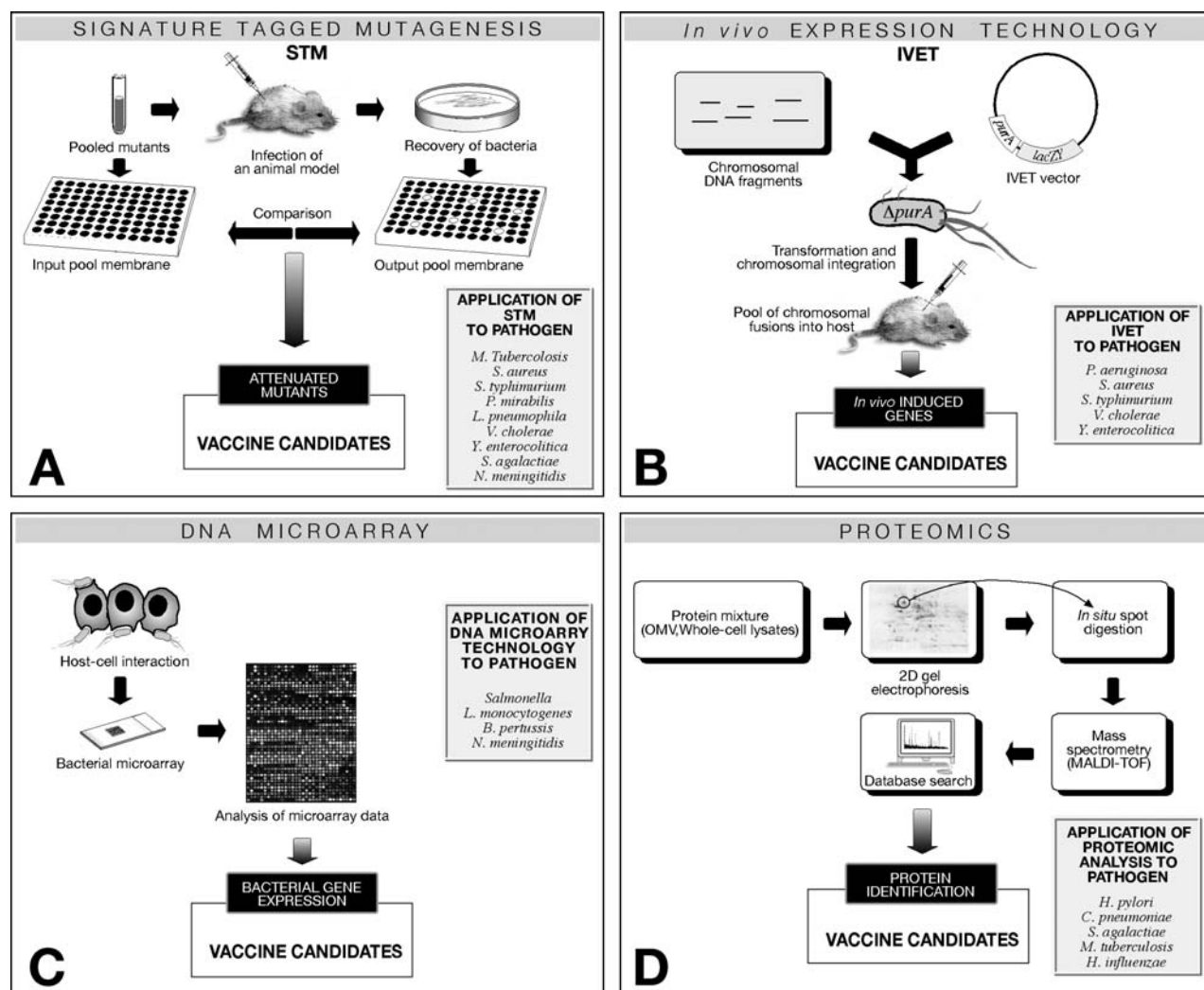
Figure 2. Principles of the functional genomics approaches applied to vaccines discovery and list of human pathogens for whose the different techniques have been used: STM (A), IVET (B), DNA microarray (C) and Proteomics (D).

selection technique uses a collection of transposons, each one modified by the incorporation of a different DNA sequence tag of approximately 40 bp. In this way, each mutant can be recognized by the specific sequence present in the integrated transposon. Tagged mutants strains are pooled together and inoculated into an animal. After infection is established, bacteria are recovered from the infected animals and plated. Mutants that are attenuated will not be recovered from the animals. Comparison of tags that are present in the inoculum but absent from the recovered bacteria identifies attenuated mutants (Figure 2A).

There are two advantages of this approach for the design of novel vaccines. On the one hand, the technique potentially allows the identification of attenuated mutants that fail to cause productive infection and hence may be used as live vaccines. On the other, proteins identified as being essential for infection or disease are likely to be good candidates for subunit vaccines.

STM has been utilized to discover virulence genes from a large variety of bacterial species including *Mycobacterium tuberculosis*, *Staphylococcus aureus*, *Salmonella typhimurium*, *Vibrio cholerae*, *Yersina enterocolitica* (Chiang *et al.*, 1999; Lehoux and Levesque, 2000; Perry, 1999; Shea *et al.*, 2000; Tang and Holden, 1999) and more recently *Streptococcus agalactiae* (Jones *et al.*, 2000) and *N. meningitidis* (Sun *et al.*, 2000). As an example, Sun *et al* used STM in conjunction with the two available complete genome sequences of meningococcus. Using an infant rat model of invasive infection, a library of 2850 insertional mutants of *N. meningitidis* was scored and 73 genes were identified that were essential for bacteraemia, many of which were of unknown function. In addition to eight known virulence genes (providing a validation of the method), 65 novel genes were found, none of which had previously been identified as essential to infection *in vivo*. Among these, 16 represented surface-expressed antigens that are now being evaluated as potential vaccine candidates.

IVET is a genetic system that positively selects bacterial genes that are specifically induced when bacteria

infect their host (Mahan *et al.*, 1993). The IVET strategy requires a bacterial strain carrying a mutation in a biosynthetic gene that attenuates growth *in vivo*, for example, a *purA* auxotroph. The biosynthetic function, essential for growth in the host, is provided by a promoterless *purA* gene, in which fragments obtained from a random library of the pathogen's chromosomal DNA supply the missing transcription elements. The positively selected fusions are then sequenced to identify *in vivo*-induced genes (Figure 2B). This IVET method necessitates the existence of an attenuating and complementable auxotrophy, which may not be available in all microbial systems. However, alternative systems have been proposed based on reporter genes encoding resistance to antibiotics (Mahan *et al.*, 1995), or encoding the green fluorescent protein (gfp). This latter method is known as differential fluorescence induction, DFI (Valdivia *et al.*, 1996; Valdivia and Falkow, 1997). IVET technology has been successfully utilized to identify several virulence genes in different human pathogen such as *Pseudomonas aeruginosa, S. typhimurium, Y. enterocolitica, V. cholerae* and *S. aureus* (Chiang *et al.*, 1999; Tang and Holden, 1999). It is important to point out that both IVET and STM technologies have been applied before the advent of the genomic era. In fact, the previous knowledge of the genome sequence is not strictly necessary for their application. Once the mutants are selected, the regions flanking the insertions are sequenced and the inactivated gene is identified. However, both technologies greatly benefit from the availability of genome sequences since the identification of the inactivated genes can be carried out rapidly by sequencing only few nucleotides upstream and downstream from the insertions.

*Microarray-based genomic technology and applications*
DNA microarray (or DNA microchips) is a recently developed genomic technology and is listed as one of ten breakthrough technologies in 1998, along with genomics. In the last few years, DNA microarray-based approaches have achieved rapid assent in a variety of fields to study the function of genes in the pathogenesis of infectious diseases. The theory and background of microarray technology have been described elsewhere by several excellent reviews (Brown and Botstein, 1999; Cheung *et al.*, 1999; Lipshutz *et al.*, 1999; Lockhart and Winzeler, 2000). DNA microarray technology appears to be particularly attractive for the analysis of organisms with relatively simple genomes, as in the case of bacteria. DNA chips carrying the entire bacterial genome can be easily prepared, allowing whole genome expression analysis. Bacterial microarrays promise to be particularly helpful in new virulence gene hunting. The complex interaction between host and pathogen is now being explored using microarrays (Cummings and Relman, 2000; Kato-Maeda *et al.*, 2001; Manger and Relman, 2000). Virulence gene expression can be monitored by growing the pathogens in the appropriate *in vivo* models (cell cultures and/or animals) and, after recovering the bacteria for RNA preparation and labelling, the gene activity is analysed and compared with the expression of the genes under *in vitro* conditions. By following the pattern of gene expression at different times,

it is possible to elucidate all of the host genes and those of the bacteria whose expression is modified (up- or down-regulated) during host-pathogen interaction (Figure 2C) (Rappuoli, 2000b). There are a few studies focused on host–pathogen interactions using DNA microarrays and they include the interactions between intestinal epithelial cells and *Salmonella* (Eckmann *et al.*, 2000), human promyelocytic cells and *Listeria monocytogenes* (Cohen *et al.*, 2000), and bronchial epithelial cells and *Bordetella pertussis* (Belcher *et al.*, 2000). However, these studies consider gene activation from the host perspective. Recently, the DNA microarrays technology has been applied to study the gene expression profile of the human pathogen *N. meningitidis* serogroup B during different stages of infection. The work by Grifantini *et al* (Grifantini *et al.*, 2002) represents an example of the use of DNA microarrays to study gene regulation during interaction with human epithelial cells from a bacterial perspective and an indication of how this technology can be exploited in vaccine design (see below). In an independent work, Kurz *et al* analysed the transcriptional changes in *N. meningitidis* in a model system of three key steps of meningococcal infection. RNA was isolated from meningococci incubated in human serum as well as adherent to human epithelial and endothelial cells. With this approach, authors found a wide range of surface proteins, which are induced under *in vivo* conditions. These antigens could represent novel candidates for a protein-based vaccine for the prevention of meningococcal disease (Kurz *et al.*, 2003). The whole-genome view of microarrays makes them promising tools to identify candidate genes as targets for vaccine or drug development. However, because the results of pathogen gene expression and of host-pathogen interactions are influenced by the model system used, such results must be interpreted cautiously. In addition, expression data have limitations because mRNA levels may not reflect protein levels, and expression of a protein may not always have a pathological consequence (Gygi *et al.*, 1999). Consequently, traditional biological, pathology and toxicity studies remain necessary.

Another use of whole genome DNA microarrays is to compare the genomes of related bacteria (complete genome hybridisation or CGH). DNA microarrays containing the genome of one strain can be hybridised with total genomic DNA from different strains or related bacteria for which genome sequence data do not exist thus permitting the identification of genes present in one strain and absent in another (Ochman and Jones, 2000; Ochman and Moran, 2001; Schoolnik, 2002). Recently, Tettelin *et al* (2002) used this technique to compare the genomes of 22 strains of *S. agalactiae* (Group B streptococcus), comprising examples of all nine known serotypes, with the genome of a serotype V isolate of which they had determined the complete genome sequence. The analysis revealed a number of regions of the genome that are highly variable and, more importantly, those genes common to all strains. This latter group contains the best candidates for a vaccine capable to induce cross serotype protection (Tettelin *et al.*, 2002).

*Proteomics in vaccine design*

With the availability of genomic sequences, the progress achieved in 2D-gel electrophoresis separation techniques and advances in mass spectrometry analysis means that it is now possible to separate, identify and catalogue the proteins expressed in a cell under several conditions. The entire set of proteins encoded by the genome has been defined as "proteome" (Grandi, 2001; Washburn and Yates, 2000). In proteome analysis, a protein mixture (*e.g.* outer membrane preparations or whole cell lysates) is first resolved in its individual components using separation procedures. Once separated, each protein undergoes digestion with a specific protease to generate discrete peptide fragments of which the molecular masses can be accurately evaluated by mass spectroscopy. The experimental result is then compared with theoretical results expected for the same specific degradation of all predicted proteins from the genome sequence. In this way, the protein can be unequivocally identified as the product of a specific gene (Figure 2D). Physical analysis of the proteome permits the identification of proteins actually expressed in a particular compartment or under different conditions of growth. Recently, this approach has been used to identify novel bacterial vaccine candidates against several human pathogens. As an example, we report here a few applications. Analysis of the proteome of the membrane compartment is a more direct way to identify potentially surface-exposed proteins and thus vaccine candidates. An analysis of the membrane compartment of *Helicobacter pylori* has identified a number of potential vaccine candidates currently awaiting testing in animal models (Chakravarti *et al.*, 2000). Proteome analysis of the outer surface proteins of the human pathogen *S. agalactiae* allowed the discovery of novel surface proteins. Sera, raised against some of these proteins, were protective in a neonatal-animal model system against a lethal dose of bacteria. Hence novel potential vaccine candidates against *S. agalactiae* were identified (Hughes *et al.*, 2002). Grandi and colleagues have recently combined genome mining and proteome technologies to identify surface-exposed antigens of *Chlamydia pneumoniae* (Montigiani *et al.*, 2002). The authors identified 157 putative surface-exposed proteins by *in silico* analysis of the pathogen genome. They then used recombinant forms of these proteins expressed in *E. coli* to raise antisera that were used to assess surface location by flow cytometry. Finally, 2D gel electrophoresis and mass spectroscopy were used to confirm the expression of the antigens in the elementary body phase of development. The result of this systematic genome–proteome combined approach represents the first successful effort to define surface protein organization of *C. pneumoniae* and opens the way to the selection of suitable components for a novel vaccine. Finally, proteome comparisons are likely to become progressively more important for the study of bacterial pathogenesis: comparison of virulent strains of a pathogen with non-virulent or commensal strains of the same bacteria should permit the identification of proteins involved in virulence (Jungblut *et al.*, 1999).

*Reverse Vaccinology: MenB*

The first example where the genome has been successful applied for the identification of potential vaccine candidates is the case of the human pathogen *N. meningitidis* serogroup B (MenB). In the last four decades efforts have not been sufficient to deliver an effective and universal vaccine against MenB, a major cause of meningitis and sepsis in children and young adults. Conjugate vaccines based on the capsular polysaccharide are available for pathogenic *N. meningitidis* serogroups A, C, Y and W135. In the case of MenB the capsular vaccine approach cannot be used as its capsule is chemically identical to α 2-8 linked polysialic acid present in human tissue, is poorly immunogenic, and is a potential cause of auto immunity.
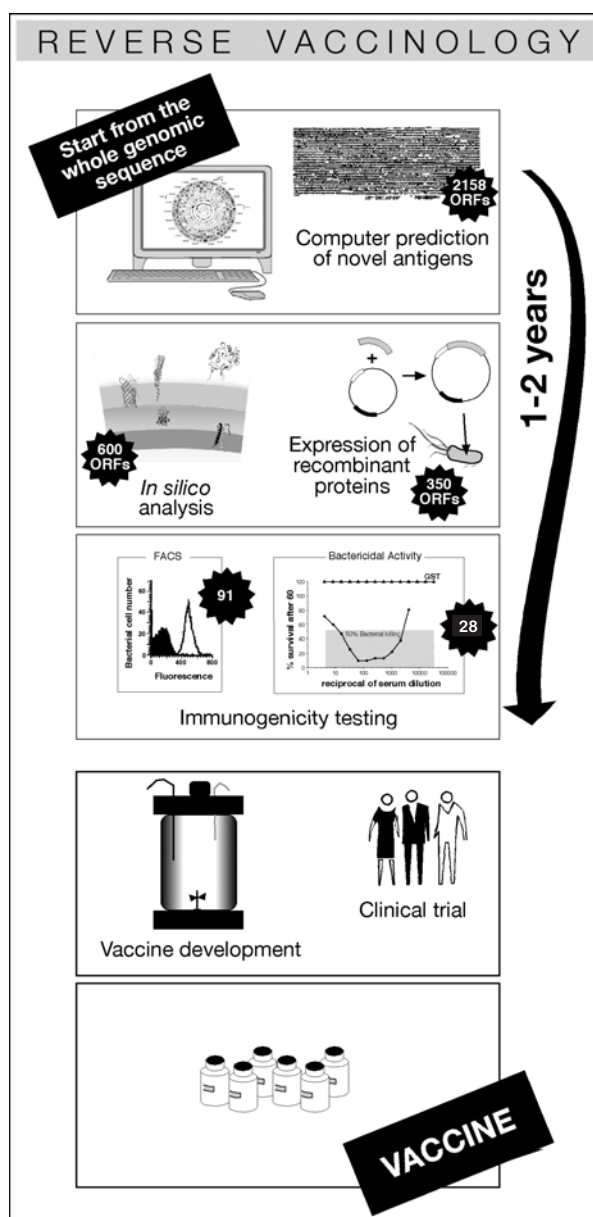


Figure 3. Reverse vaccinology applied to *Neisseria meningitidis* serogroup B. Schematic representation of the different steps from *in silico* analysis to vaccine development.

To identify novel vaccine candidates against MenB, Pizza *et al* applied reverse vaccinology (Pizza *et al.*, 2000). Briefly, the sequence of the virulent strain MC58 was determined by the shotgun strategy (Tettelin *et al.*, 2000) and while the sequencing was still in progress, the unassembled DNA fragments were analysed to identify the ORFs. Using computer prediction programs 600 surface-exposed or secreted antigens were identified. Of the predicted antigens, 350 were successfully expressed in *E. coli*, purified and used to immunize mice. The antisera obtained were tested by FACS and ELISA to evaluate the surface localisation of the antigens in MenB. In addition, the antisera were tested for bactericidal antibodies, a property known to correlate with protection in humans. Ninety-one novel surface-exposed antigens were identified, 28 of which induced bactericidal antibodies (Figure 3). Several of the antigens previously identified using conventional approaches showed strain variability or were only expressed in some strains, and most of them were effective only against the homologous strains. Therefore, the potential vaccine candidates identified were evaluated for degree of sequence variability among multiple isolates and serogroups of *N. meningitidis,* three strains of *N. gonorrhoeae*, as well as one strain each of *N. cinerea* and *N. lactamica*. The MenB antigens identified by the genome analysis are quite different from those identified using the conventional approaches. The majority of them are present in all strains tested and are conserved in sequence. Many of the newly-identified MenB antigens in fact include surface-exposed protein or lipoproteins with a globular structure and without membrane spanning domains and many of them are not abundant on the bacterial surface. Reverse vaccinology has identified more vaccine candidates as compared to those identified during the previous 40 years. These potential vaccine candidates, able to induce a broad strain coverage, are under evaluation and some of these antigens are likely to enter into development.

In a subsequent study Grifantini and co-workers (Grifantini *et al.*, 2002), using DNA microarray technology to follow gene regulation of MenB during interaction with human epithelial cells, identified 12 adhesion-induced surface antigens, five of which were able to induce bactericidal antibodies. The pioneering reverse vaccinology study by Pizza *et al* (Pizza *et al.*, 2000) did not identify these potential vaccine candidates. The study by Grifantini *et al* therefore shows that DNA microarray technology is able to identify vaccine candidates and complement other genome mining strategies. Grifantini *et al* observed that host-cell contact induced changes in the expression of 347 genes of which more than 30% encoded for proteins with unknown function. In addition, several of the upregulated genes encoded for peripherally located proteins, suggesting that cell contact promoted a reorganization of the cell membrane. The researchers confirmed this data by FACS analysis on adhering bacteria using the mouse antisera against 12 adhesion-induced proteins. The example of MenB provides the first demonstration of the potential and use of functional genomics to expand and accelerate the development of a vaccine.

In the last two years several groups have followed the footsteps of MenB, utilizing the approach of reverse vaccinology/functional genomics to identify new vaccine candidates (Adu-Bobie *et al.*, 2003). Pathogens that have been studied include *S. pneumoniae* (Wizemann *et al.*, 2001)*, S. aureus* (Etz *et al.*, 2002; Vytvytska *et al.*, 2002)*, C. pneumoniae* (Montigiani *et al.*, 2002)*, Porhyromonas gingivalis* (Ross *et al.*, 2001)*, Edwardsiella tarda* (Srinivasa Rao *et al.*, 2003), and *M. tuberculosis* (Betts, 2002)*.*

The availability of the genome allows the discovery of previously unknown and undescribed proteins. However, genome projects risk becoming an extensive collection of information. Therefore,  it is important not only to explore the potential as vaccine candidates, but also to evaluate their role and function in pathogenesis. Many of the antigens identified in MenB share homology with known virulence factors and here we describe the characterisation of several of the antigens identified, from the biochemical and functional point of view.

*App: a new adhesin of N. meningitidis*. NMB1985, named Adhesion and penetration protein (App) (Hadi *et al.*, 2001), is highly homologous to the Haemophilus adhesion and penetration protein (Hap) of *H. influenzae*. Both Hap and App belong to the autotransporter family that comprises proteins from Gram-negative bacteria characterized by a distinct mechanism of secretion. App protein is exported to the outer membrane, cleaved and released in the culture supernatant. App protein has been functionally characterized very recently, using *E. coli* as an expression system. These studies have demonstrated that App has a serine protease activity: mutation of S267 to Alanine abolishes processing and secretion of App, which remains cell-associated. Moreover, App is an adhesin able to bind to human epithelial but not to endothelial cells. Deletion of the *app* gene in a virulent serogroup B capsulated strain significantly reduces its adherence capability compared to the wild type strain. The results provide evidence that App is a new adhesin, which may play a role in *N. meningitidis* colonization of the nasopharyngeal mucosa (Serruto *et al.*, 2003).

*GNA1870: a novel lipoprotein as vaccine candidate*. GNA1870, a 26,9 kDa lipoprotein, has been detected on the surface of meningococci by FACS analysis and western blot on outer membrane vesicles. Analysis of the expression of different strains has shown that proteins are expressed at different levels in the different strains that have been therefore classified as high, intermediate and low expressors. The gene is present in all strains analysed and the sequence analysis has shown that it is present as three variants, namely variants 1, 2 and 3. Amino acid identity between variant 1 and variant 2 is 74.1%, between variant 1 and variant 3 is 62.8%, and 84.7% between variant 2 and variant 3. Sequences within each variant are well-conserved, the most distant showing 91.6%, 93.4% and 93.2% identity to their type strains, respectively. The protein is able to induce antibodies with high bactericidal activity against each strain carrying the same variant even when the amount of protein expressed is low. The activity is low

or absent against strains of the other variants. Furthermore, antibodies are able to induce protection in the infant rat model (Masignani et al., 2003).

*NadA: a new adhesin vaccine candidate.* NadA (<u>N</u>eisseria <u>ad</u>hesin <u>A</u>, NMB1994) is homologous to YadA of enteropathogenic *Yersinia*, (Hoiczyk *et al.*, 2000) and to UspA2 of *Moraxella catarrhalis*, (Lafontaine *et al.*, 2000). While sequence homology is limited to the carboxyl terminal region, an overall similarity can be observed at the level of the secondary structure (Comanducci *et al.*, 2002). The three proteins have an amino-terminal domain, an internal alpha-helix region with high coiled-coils probability and a carboxyl terminal membrane anchor domain. In meningococcus, NadA forms very stable high molecular weight oligomers anchored to the bacterial outer membrane. Furthermore, *nadA* gene is present in three out of four hypervirulent lineages clustering into three well-conserved alleles. NadA is able to bind to human epithelial cells *in vitro*. The predicted secondary structure and the ability of the purified protein to interact with the host cells, suggest that NadA could belong to a novel class of adhesins with YadA and UspA2. Finally, antibodies obtained against the recombinant protein, are able to induce a complement-mediated killing of both homologous and heterologous strains, suggesting that NadA could be a good candidate for a vaccine against meningococcus.

*GNA33: a murein lytic transglycosylase.* GNA33 (<u>G</u>enome-derived <u>N</u>eisseria <u>A</u>ntigen) is a lipoprotein homologous to a membrane-bound lytic transglycosylase (MltA) of *E. coli* (33% of identity). Biochemical characterization of the recombinant protein, expressed and purified from *E. coli*, showed that GNA33 is able to degrade both insoluble murein sacculi and unsubstituted glycan strands. The results confirmed that GNA33 is a lytic transglycosylase with muramidase activity (Jennings *et al.*, 2002). The recombinant protein elicits bactericidal antibody by mimicking a surface-exposed epitope of porin A (PorA) and confers passive protection against bacteremia in an infant rat model. GNA33 and PorA share a common short motif (QPT), localised in the loop 4 of PorA, that is necessary but not sufficient for binding (Granoff *et al.*, 2001).

*GNA992: a putative adhesin.* GNA992 is a surface-exposed protein able to induce bactericidal antibodies. The protein is homologous to Hsf and Hia (58% and 50%, respectively), two adhesins of *H. influenzae*. The high level of amino acid similarity of GNA992 with Hsf and Hia, and the similar structures of the three proteins, suggest that they could have a common role in the adhesion process. The three proteins have a conserved secondary structure and present a modular organization with a different number of repeats. These repeat units have a conserved motif that is present also in other adhesive molecules of *H. influenzae* as well as in human proteins belonging to the family of cell adhesion molecules (CAMs). The analogy between GNA992 and CAM could suggest a common mechanism of action; therefore, GNA992 could mediate the adhesion of meningococcus to human cells by mimicking the cell-cell interaction process (Scarselli *et al.*, 2001).

## Conclusions

Genomics has introduced a new paradigm in approaches to vaccine research and bacterial pathogenesis. To date, more than one hundred genome sequences are available (www.tigr.org/tdb). The recent advances in bioinformatics, experimental techniques as well as powerful technologies (proteomics and microarray), have unravelled valuable information. Moreover, several positive results, recently obtained with meningococcus and other microorganisms, show that functional genomics could be a solution to identify new vaccine candidates. In conclusion, the application of functional genomics to all those pathogens where conventional approaches have failed, could provide new impulse in vaccine research and development.

## References

Adu-Bobie, J., Capecchi, B., Serruto, D., Rappuoli, R., and Pizza, M. 2003. Two years into reverse vaccinology. Vaccine. 21: 605-610.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25: 3389-3402.

Andre, F.E. 1990. Overview of a 5-year clinical experience with a yeast-derived hepatitis B vaccine. Vaccine 8 Suppl: S74-78; discussion S79-80.

Belcher, C.E., Drenkow, J., Kehoe, B., Gingeras, T.R., McNamara, N., Lemjabbar, H., Basbaum, C., and Relman, D.A. 2000. The transcriptional responses of respiratory epithelial cells to Bordetella pertussis reveal host defensive and pathogen counter-defensive strategies. Proc. Natl. Acad. Sci. USA. 97: 13847-13852.

Betts, J.C. 2002. Transcriptomics and proteomics: tools for the identification of novel drug targets and vaccine candidates for tuberculosis. IUBMB Life. 53: 239-242.

Brown, P.O., and Botstein, D. 1999. Exploring the new world of the genome with DNA microarrays. Nat. Genet. 21: 33-37.

Chakravarti, D.N., Fiske, M.J., Fletcher, L.D., and Zagursky, R.J. 2000. Application of genomics and proteomics for identification of bacterial gene products as potential vaccine candidates. Vaccine. 19: 601-612.

Cheung, V.G., Morley, M., Aguilar, F., Massimi, A., Kucherlapati, R., and Childs, G. 1999. Making and reading microarrays. Nat. Genet. 21: 15-19.

Chiang, S.L., Mekalanos, J.J., and Holden, D.W. 1999. *In vivo* genetic analysis of bacterial virulence. Annu. Rev. Microbiol. 53: 129-154.

Cohen, P., Bouaboula, M., Bellis, M., Baron, V., Jbilo, O., Poinot-Chazel, C., Galiegue, S., Hadibi, E.H., and Casellas, P. 2000. Monitoring cellular responses to Listeria monocytogenes with oligonucleotide arrays. J. Biol. Chem. 275: 11181-11190.

Comanducci, M., Bambini, S., Brunelli, B., Adu-Bobie, J., Arico, B., Capecchi, B., Giuliani, M.M., Masignani, V., Santini, L., Savino, S., Granoff, D.M., Caugant, D.A., Pizza, M., Rappuoli, R., and Mora, M. 2002. NadA, a novel vaccine candidate of Neisseria meningitidis. J. Exp. Med. 195: 1445-1454.

Cummings, C.A., and Relman, D.A. 2000. Using DNA microarrays to study host-microbe interactions. Emerg. Infect. Dis. 6: 513-525.

Eckmann, L., Smith, J.R., Housley, M.P., Dwinell, M.B., and Kagnoff, M.F. 2000. Analysis by high density cDNA arrays of altered gene expression in human intestinal epithelial cells in response to infection with the invasive enteric bacteria Salmonella. J. Biol. Chem. 275: 14084-14094.

Etz, H., Minh, D.B., Henics, T., Dryla, A., Winkler, B., Triska, C., Boyd, A.P., Sollner, J., Schmidt, W., von Ahsen, U., Buschle, M., Gill, S.R., Kolonay, J., Khalak, H., Fraser, C.M., von Gabain, A., Nagy, E., and Meinke, A. 2002. Identification of in vivo expressed vaccine candidate antigens from Staphylococcus aureus. Proc. Natl. Acad. Sci. USA. 99: 6573-6578.

Field, D., Hood, D., and Moxon, R. 1999. Contribution of genomics to bacterial pathogenesis. Curr. Opin. Genet. Dev. 9: 700-703.

Fischer, D., and Eisenberg, D. 1997. Assigning folds to the proteins encoded by the genome of Mycoplasma genitalium. Proc. Natl. Acad. Sci. USA. 94: 11929-11934.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne J.D., Scott, J., Shirley, R., Liu, L-I., Glodek, A., Kelley, J.M., Weidman, J.F., Phillips, C.A., Spiggs, T., Hedblom, E., Cotton, M.D., Utterback, T.R., Hanna, M.C., Nguyen, D.T., Saudek, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann J.L., Geoghagen N.S.M., Gnehm C.L., McDonald, L.A., Small, K.V., Fraser, C.M., Smith H.O., and Venter, J.C. 1995. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science. 269: 496-512.

Grandi, G. 2001. Antibacterial vaccine design using genomics and proteomics. Trends Biotechnol. 19: 181-188.

Granoff, D.M., Moe, G.R., Giuliani, M.M., Adu-Bobie, J., Santini, L., Brunelli, B., Piccinetti, F., Zuno-Mitchell, P., Lee, S.S., Neri, P., Bracci, L., Lozzi, L., and Rappuoli, R. 2001. A novel mimetic antigen eliciting protective antibody to Neisseria meningitidis. J. Immunol. 167: 6487-6496.

Greco, D., Salmaso, S., Mastrantonio, P., Giuliano, M., Tozzi, A.E., Anemona, A., Ciofi degli Atti, M.L., Giammanco, A., Panei, P., Blackwelder, W.C., Klein, D.L., and Wassilak, S.G. 1996. A controlled trial of two acellular vaccines and one whole-cell vaccine against pertussis. Progetto Pertosse Working Group. N. Engl. J. Med. 334: 341-348.

Grifantini, R., Bartolini, E., Muzzi, A., Draghi, M., Frigimelica, E., Berger, J., Ratti, G., Petracca, R., Galli, G., Agnusdei, M., Giuliani, M.M., Santini, L., Brunelli, B., Tettelin, H., Rappuoli, R., Randazzo, F., and Grandi, G. 2002. Previously unrecognized vaccine candidates against group B meningococcus identified by DNA microarrays. Nat. Biotechnol. 20: 914-921.

Gygi, S.P., Rochon, Y., Franza, B.R., and Aebersold, R. 1999. Correlation between protein and mRNA abundance in yeast. Mol. Cell. Biol. 19: 1720-1730.

Hadi, H.A., Wooldridge, K.G., Robinson, K., and Ala'Aldeen, D.A. 2001. Identification and characterization of App: an immunogenic autotransporter protein of Neisseria meningitidis. Mol. Microbiol. 41: 611-623.

Hayashi, S., and Wu, H.C. 1990. Lipoproteins in bacteria. J. Bioenerg. Biomembr. 22: 451-471.

Hensel, M., Shea, J.E., Gleeson, C., Jones, M.D., Dalton, E., and Holden, D.W. 1995. Simultaneous identification of bacterial virulence genes by negative selection. Science. 269: 400-403.

Hoiczyk, E., Roggenkamp, A., Reichenbecher, M., Lupas, A., and Heesemann, J. 2000. Structure and sequence analysis of Yersinia YadA and Moraxella UspAs reveal a novel class of adhesins. Embo J. 19: 5989-5999.

Hughes, M.J., Moore, J.C., Lane, J.D., Wilson, R., Pribul, P.K., Younes, Z.N., Dobson, R.J., Everest, P., Reason, A.J., Redfern, J.M., Greer, F.M., Paxton, T., Panico, M., Morris, H.R., Feldman, R.G., and Santangelo, J.D. 2002. Identification of major outer surface proteins of Streptococcus agalactiae. Infect. Immun. 70: 1254-1259.

Jennings, G.T., Savino, S., Marchetti, E., Arico, B., Kast, T., Baldi, L., Ursinus, A., Holtje, J.V., Nicholas, R.A., Rappuoli, R., and Grandi, G. 2002. GNA33 from Neisseria meningitidis serogroup B encodes a membrane-bound lytic transglycosylase (MltA). Eur. J. Biochem. 269: 3722-3731.

Jones, A.L., Knoll, K.M., and Rubens, C.E. 2000. Identification of Streptococcus agalactiae virulence genes in the neonatal rat sepsis model using signature-tagged mutagenesis. Mol. Microbiol. 37: 1444-1455.

Jungblut, P.R., Schaible, U.E., Mollenkopf, H.J., Zimny-Arndt, U., Raupach, B., Mattow, J., Halada, P., Lamer, S., Hagens, K., and Kaufmann, S.H. 1999. Comparative proteome analysis of Mycobacterium tuberculosis and Mycobacterium bovis BCG strains: towards functional genomics of microbial pathogens. Mol. Microbiol. 33: 1103-1117.

Kato-Maeda, M., Gao, Q., and Small, P.M. 2001. Microarray analysis of pathogens and their interaction with hosts. Cell. Microbiol. 3: 713-719.

Kurz, S., Hubner, C., Aepinus, C., Theiss, S., Guckenberger, M., Panzner, U., Weber, J., Frosch, M., and Dietrich, G. 2003. Transcriptome-based antigen identification for Neisseria meningitidis. Vaccine. 21: 768-775.

Lafontaine, E.R., Cope, L.D., Aebi, C., Latimer, J.L., McCracken, G.H., Jr., and Hansen, E.J. 2000. The UspA1 protein and a second type of UspA2 protein mediate adherence of Moraxella catarrhalis to human epithelial cells in vitro. J. Bacteriol. 182: 1364-1373.

Lehoux, D.E., and Levesque, R.C. 2000. Detection of genes essential in specific niches by signature-tagged mutagenesis. Curr. Opin. Biotechnol. 11: 434-439.

Levine M.M., Woodrow G.C., Kaper J.B., and Cobon G.S. 1997. New generation vaccines. Second edition, revised and expanded. Marcel Dekker, Inc. New York.

Lipshutz, R.J., Fodor, S.P., Gingeras, T.R., and Lockhart, D.J. 1999. High density synthetic oligonucleotide arrays. Nat. Genet. 21: 20-24.

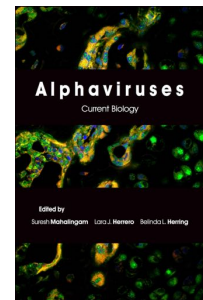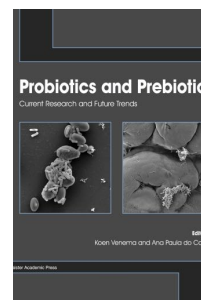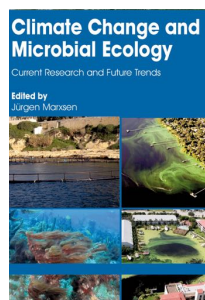Lockhart, D.J., and Winzeler, E.A. 2000. Genomics, gene expression and DNA arrays. Nature. 405: 827-836.

Mahan, M.J., Slauch, J.M., and Mekalanos, J.J. 1993. Selection of bacterial virulence genes that are specifically induced in host tissues. Science. 259: 686-688.

Mahan, M.J., Tobias, J.W., Slauch, J.M., Hanna, P.C., Collier, R.J., and Mekalanos, J.J. 1995. Antibiotic-based selection for bacterial genes that are specifically induced during infection of a host. Proc. Natl. Acad. Sci. USA. 92: 669-673.

Manger, I.D., and Relman, D.A. 2000. How the host 'sees' pathogens: global gene expression responses to infection. Curr. Opin. Immunol. 12: 215-218.

Masignani, V., Comanducci, M., Giuliani, M.M., Bambini, S., Adu-Bobie, J., Arico, B., Brunelli, B., Pieri, A., Santini, L., Savino, S., Serruto, D., Litt, D., Kroll, S., Welsch, J.A., Granoff, D.M., Rappuoli, R., and Pizza, M. 2003. Vaccination against Neisseria meningitidis using three variants of the lipoprotein GNA1870. J. Exp. Med. 197: 789-799.

Montigiani, S., Falugi, F., Scarselli, M., Finco, O., Petracca, R., Galli, G., Mariani, M., Manetti, R., Agnusdei, M., Cevenini, R., Donati, M., Nogarotto, R., Norais, N., Garaguso, I., Nuti, S., Saletti, G., Rosa, D., Ratti, G., and Grandi, G. 2002. Genomic approach for analysis of surface proteins in Chlamydia pneumoniae. Infect. Immun. 70: 368-379.

Ochman, H., and Jones, I.B. 2000. Evolutionary dynamics of full genome content in Escherichia coli. Embo J. 19: 6637-6643.

Ochman, H., and Moran, N.A. 2001. Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. Science. 292: 1096-1099.

Perry, R.D. 1999. Signature-tagged mutagenesis and the hunt for virulence factors. Trends Microbiol. 7: 385-388; discussion 388-389.

Pizza, M., Covacci, A., Bartoloni, A., Perugini, M., Nencioni, L., De Magistris, M.T., Villa, L., Nucci, D., Manetti, R., Bugnoli, M., Giovannoni, F., Olivieri, R., Barbieri, J.T., Sato, H., and Rappuoli, R. 1989. Mutants of pertussis toxin suitable for vaccine development. Science. 246: 497-500.

Pizza, M., Scarlato, V., Masignani, V., Giuliani, M.M., Arico, B., Comanducci, M., Jennings, G.T., Baldi, L., Bartolini, E., Capecchi, B., Galeotti, C.L., Luzzi, E., Manetti, R., Marchetti, E., Mora, M., Nuti, S., Ratti, G., Santini, L., Savino, S., Scarselli, M., Storni, E., Zuo, P., Broeker, M., Hundt, E., Knapp, B., Blair, E., Mason, T., Tettelin, H., Hood, D.W., Jeffries, A.C., Saunders, N.J., Granoff, D.M., Venter, J.C., Moxon, E.R., Grandi, G., and Rappuoli, R. 2000. Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. Science. 287: 1816-1820.

Rappuoli, R. 2000a. Reverse vaccinology. Curr. Opin. Microbiol. 3: 445-450.

Rappuoli, R. 2000b. Pushing the limits of cellular microbiology: microarrays to study bacteria-host cell intimate contacts. Proc. Natl. Acad. Sci. USA. 97: 13467-13469.

Ross, B.C., Czajkowski, L., Hocking, D., Margetts, M., Webb, E., Rothel, L., Patterson, M., Agius, C., Camuglia, S., Reynolds, E., Littlejohn, T., Gaeta, B., Ng, A., Kuczek, E.S., Mattick, J.S., Gearing, D., and Barr, I.G. 2001. Identification of vaccine candidate antigens from a genomic analysis of Porphyromonas gingivalis. Vaccine. 19: 4135-4142.

Scarselli, M., Rappuoli, R., and Scarlato, V. 2001. A common conserved amino acid motif module shared by bacterial and intercellular adhesins: bacterial adherence mimicking cell cell recognition? Microbiology. 147: 250-252.

Schoolnik, G.K. 2002. Functional and comparative genomics of pathogenic bacteria. Curr. Opin. Microbiol. 5: 20-26.

Serruto, D., Adu-Bobie, J., Scarselli, M., Veggi, D., Pizza, M., Rappuoli, R., and Aricò, B. 2003. Neisseria meningitidis App, a new adhesin with autocatalytic serine-protease activity. Mol. Microbiol. In press.

Shea, J.E., Santangelo, J.D., and Feldman, R.G. 2000. Signature-tagged mutagenesis in the identification of virulence genes in pathogens. Curr. Opin. Microbiol. 3: 451-458.

Srinivasa Rao, P.S., Lim, T.M., and Leung, K.Y. 2003. Functional genomics approach to the identification of virulence genes involved in Edwardsiella tarda pathogenesis. Infect. Immun. 71: 1343-1351.

Struyve, M., Moons, M., and Tommassen, J. 1991. Carboxy-terminal phenylalanine is essential for the correct assembly of a bacterial outer membrane protein. J. Mol. Biol. 218: 141-148.

Sun, Y.H., Bakshi, S., Chalmers, R., and Tang, C.M. 2000. Functional genomics of Neisseria meningitidis pathogenesis. Nat. Med. 6: 1269-1273.

Tang, C., and Holden, D. 1999. Pathogen virulence genes—implications for vaccines and drug therapy. Br. Med. Bull. 55: 387-400.

Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C., Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Peden, J.F., Dodson, R.J., Nelson, W.C., Gwinn, M.L., DeBoy, R., Peterson, J.D., Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D., Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blair, E., Cittone, H., Clark, E.B., Cotton, M.D., Utterback, T.R., Khouri, H., Qin, H., Vamathevan, J., Gill, J., Scarlato, V., Masignani, V., Pizza, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R., Rappuoli, R., and Venter, J.C. 2000. Complete genome sequence of Neisseria meningitidis serogroup B strain MC58. Science. 287: 1809-1815.

Tettelin, H., Masignani, V., Cieslewicz, M.J., Eisen, J.A., Peterson, S., Wessels, M.R., Paulsen, I.T., Nelson, K.E., Margarit, I., Read, T.D., Madoff, L.C., Wolf, A.M., Beanan, M.J., Brinkac, L.M., Daugherty, S.C., DeBoy, R.T., Durkin, A.S., Kolonay, J.F., Madupu, R., Lewis, M.R., Radune, D., Fedorova, N.B., Scanlan, D., Khouri, H., Mulligan, S., Carty, H.A., Cline, R.T., Van Aken, S.E., Gill, J., Scarselli, M., Mora, M., Iacobini, E.T., Brettoni, C., Galli, G., Mariani, M., Vegni, F., Maione, D., Rinaudo, D., Rappuoli, R., Telford, J.L., Kasper, D.L., Grandi, G., and Fraser, C.M. (2002) Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V Streptococcus agalactiae. Proc. Natl. Acad. Sci. USA. 99: 12391-12396.

Valdivia, R.H., Hromockyj, A.E., Monack, D.,

Ramakrishnan, L., and Falkow, S. 1996. Applications for green fluorescent protein (GFP) in the study of host-pathogen interactions. Gene. 173: 47-52.

Valdivia, R.H., and Falkow, S. 1997 Fluorescence-based isolation of bacterial genes expressed within host cells. Science. 277: 2007-2011.

Vytvytska, O., Nagy, E., Bluggel, M., Meyer, H.E., Kurzbauer, R., Huber, L.A., and Klade, C.S. 2002. Identification of vaccine candidate antigens of Staphylococcus aureus by serological proteome analysis. Proteomics. 2: 580-590.

Washburn, M.P., and Yates, J.R.3rd. 2000. Analysis of the microbial proteome. Curr. Opin. Microbiol. 3: 292-297.

Wizemann, T.M., Heinrichs, J.H., Adamou, J.E., Erwin, A.L., Kunsch, C., Choi, G.H., Barash, S.C., Rosen, C.A., Masure, H.R., Tuomanen, E., Gayle, A., Brewah, Y.A., Walsh, W., Barren, P., Lathigra, R., Hanson, M., Langermann, S., Johnson, S., and Koenig, S. 2001. Use of a whole genome approach to identify vaccine molecules affording protection against Streptococcus pneumoniae infection. Infect. Immun. 69: 1593-1598.

# Further Reading