

Design of new vaccines in the genomic and post-genomic era

Author

Serino, L, Seib, KL, Pizza, M

Published

2012

Book Title

Innovation in Vaccinology: From Design, Through to Delivery and Testing

DOI

[10.1007/978-94-007-4543-8_1](https://doi.org/10.1007/978-94-007-4543-8_1)

Rights statement

© 2012 Springer. This is the author-manuscript version of this paper. It is reproduced here in accordance with the copyright policy of the publisher. Please refer to the publisher's website for further information.

Downloaded from

<http://hdl.handle.net/10072/50248>

Link to published version

https://link.springer.com/chapter/10.1007%2F978-94-007-4543-8_1

Griffith Research Online

<https://research-repository.griffith.edu.au>

Design of new vaccines in the genomic and post-genomic era

Laura Serino, Kate L. Seib and Mariagrazia Pizza

Novartis Vaccines and Diagnostics, Research Centre, Via Fiorentina, 1, Siena, Italy

Abstract

During the twentieth century, the introduction of vaccines changed the history of mankind, eliminating most of the childhood diseases that used to cause millions of deaths. However, where conventional vaccinology approaches failed, many new and emerging infectious diseases remain a threat to health worldwide. The advent of whole-genome sequencing and innovation in bioinformatic tools radically changed the way to design and develop new and improved vaccines, starting from the genomic information of a single bacterial or viral isolate, with a process named reverse vaccinology. As the genomic era progressed, reverse vaccinology has evolved in combination with different approaches, such as transcriptomics, metabolomics, structural genomics, proteomics, and immunomics, contributing to the design of new and universal vaccines. Furthermore, the genomic information of the host is increasingly being used to aid understanding of the human immune response to vaccines.

Keywords

Reverse vaccinology, genomics, transcriptomics, metabolomics, structural genomics, proteomics, immunomics

Vaccines of the past

Since the time when Pasteur developed the first vaccine in the laboratory and introduced the terminology of vaccination into medicine (Artenstein and Grabenstein 2008), until recently, there have been two main approaches for vaccine development: attenuation or inactivation, and the production of recombinant subunits of the causative agent of a disease. With regard to attenuation, the first methods used involved heat, oxygenation, chemical agents, or aging, notably by Pasteur for rabies and anthrax vaccines. Another powerful approach for vaccine development involved the inactivation of the microorganism, with the discovery that whole organisms could be killed without losing its immunogenicity. This new strategy soon became the basis of vaccines for typhoid and cholera, and later for pertussis, influenza and hepatitis A.

Other approaches consisted of isolation of virulence factors from microorganisms, such as toxins or capsular polysaccharides. For example, for diphtheria and tetanus the exotoxins were formalin-inactivated (Plotkin 2003), or for *Haemophilus influenzae* type b (Hib) the polysaccharide capsule was extracted and used as a vaccine antigen. However, early observations demonstrated that the plain polysaccharide used as a *H. influenzae* vaccine antigen had poor immunogenicity. Following further studies with a variety of bacterial polysaccharides, the Hib polysaccharide was shown to be more immunogenic when covalently linked to a protein carrier, giving boosted responses characteristic of T-dependent memory (Kelly and Moxon 2004; Purcell et al. 2007).

Overall, conventional vaccinology approaches have led to the eradication of many infectious diseases (reviewed in Plotkin 2009). However, production of vaccines by conventional methods is very laborious and time-consuming and in many cases has failed to provide an effective and protective vaccine against lethal diseases. For these reasons, a faster and more reliable approach was needed in order to be able to quickly respond to old and emerging pathogens.

The Reverse Vaccinology approach: tackling serogroup B meningococcus

The first major innovation in vaccine design came with the advent of whole-genome sequencing and advances in bioinformatics. By “shotgun sequencing”, the first complete genome sequence of a free-living organism (*H. influenzae*) became available in 1995 (Fleischmann et al. 1995), marking the beginning of the “genomic era” and opening a new chapter in vaccine development. Starting from the genome information of a microorganism, all proteins encoded by the pathogen were available, making it possible to identify new potential vaccine candidates that conventional approaches had not discovered. This new approach was termed “Reverse Vaccinology” (Rappuoli 2000) (see Fig. 1). In the first step of this *in silico* analysis, using the appropriate combination of algorithms and the critical evaluation of the genome’s coding capacity, all the open reading frames (ORFs) are predicted. Since secreted or extracellular proteins are more accessible to antibodies than are intracellular proteins, they represent the ideal vaccine candidates, and therefore an initial surface localization criterion is applied for the selection process. Once putative targets are identified, the target genes are cloned, expressed and purified, and the recombinant proteins used to immunize mice. Then, post-immunization sera are analysed to confirm the computer-predicted surface localization of each polypeptide, and their ability to elicit a good and protective immune response is determined using both *in vitro* and *in vivo* assays.

The Reverse Vaccinology approach was applied for the first time to the bacterial pathogen *Neisseria meningitidis* serogroup B (MenB), a Gram-negative capsulated bacterium that causes approximately 50% of meningococcal cases of sepsis and meningitis with a mortality rate of 5–15% despite continuing advances in treatment. Although the use of vaccines based on the polysaccharide antigen has been successful for most of the species causing bacterial meningitis (*H. influenzae* type B, *Streptococcus pneumoniae* and *N. meningitidis* serogroups A, C, Y and W135), the same approach could not be applied to serogroup B meningococcus. This is because the MenB polysaccharide is a polymer of $\alpha(2-8)$ -linked N-acetylneuraminic acid (or polysialic acid) that is also present in glycoproteins of mammalian neural tissues and is therefore poorly immunogenic.

Thus, MenB was chosen as the first application of reverse vaccinology and demonstrated the power of genomic approaches for novel antigen identification (Pizza et al. 2000) (Fig. 1). Briefly, the complete genome sequence of an invasive isolate of *N. meningitidis* (strain MC58) was obtained and analyzed to identify suitable vaccine candidates. Computer analysis for prediction of surface-associated proteins led to the identification of 570 genes. 350 proteins were successfully cloned and expressed, and then purified and tested for localization, immunogenicity and protective efficacy. Of the 91 proteins found to be surface-exposed, 28 were able to induce complement-mediated bactericidal antibodies, providing a strong indication of the proteins ability to induce protective immunity. In addition, to evaluate the possibility that these antigens could confer protection also against heterologous strains, the proteins were evaluated for gene presence, phase variation and sequence conservation in a panel of genetically diverse MenB strains representative of the global diversity of the natural *N. meningitidis* population. From this analysis, five antigens were found to be conserved in sequence among the panel of strains tested and were able to elicit cross-bactericidal antibodies against distantly related isolates, demonstrating that they could confer general protection against serogroup B meningococcus. The five proteins have now been combined into one multicomponent vaccine, to increase the breadth of the vaccine coverage and avoid selection of escape mutants (Giuliani et al. 2006). This vaccine has been shown to be safe and induce a robust immune response in clinical trials (Toneatto et al. 2011) and was submitted for licensure to the European Medicines Agency in 2010. In less than two years, reverse vaccinology led to the discovery of numerous new vaccine candidates, paving the way for a safe and protective vaccine against meningococcus group B.

Evolution of Reverse Vaccinology: the pan-genome concept

While the genome sequence of a single strain reveals many aspects of the biology of a species, it fails to address how genetic variability drives pathogenesis within a bacterial species. This approach also limits genome-wide screens for vac-

cine candidates or for antimicrobial targets to a single strain. The availability of genome sequences for different isolates of a single species enables quantitative analyses of their genomic diversity through comparative genomic analysis (Fig. 2).

The limitation of the classical Reverse Vaccinology approach, based on the in silico analysis of a single genome, became evident when Reverse Vaccinology was applied to the Gram-positive microorganism *Streptococcus agalactiae* (also known as Group B Streptococcus (GBS)). This bacterium is one of the leading causes of bacterial sepsis, pneumonia and meningitis in neonates and is also an emerging cause of infection in the elderly population. GBS is an organism with a very high intra-species diversity, and comparative genomic hybridization analysis revealed that there was significant variation in gene content among different clinical isolates (Tettelin et al. 2002). This evidence supported the idea that one genome sequence was not enough to fully capture the diversity of the species and to enable identification of broadly protective vaccine candidates.

In order to develop a universal vaccine with broad coverage, a number of GBS strains belonging to the five major circulating serotypes were sequenced and compared allowing the definition of the species “pan-genome” (Tettelin et al. 2005). The pan-genome can be defined as the global gene repertoire pertaining to a species. In general, it can be divided into three parts: the core-genome, which includes the set of genes that are invariably present and conserved in all isolates; the ‘dispensable genome’, comprising genes present in some but not all strains; and the strain-specific genes, which are present only in a single isolate. Comparative genome analysis of eight clinical isolates of GBS showed that 1806 genes are shared by all strains, representing the ‘core genome’ that corresponds to approximately 80% of the average number of genes encoded in each strain. Whereas approximately 907 genes are present in one or more, but not all, strains (dispensable genome) (Tettelin et al. 2005). Moreover, computational predictions suggest that as more genomes are sequenced, more new genes will be found that belong to the pan-genome (Medini et al. 2005). In fact, the pan-genome is predicted to grow by about 33 new genes for every new strain that is sequenced.

Similarly to the approach used for the MenB vaccine, computer algorithms were used to select genes encoding putative surface-associated and secreted GBS proteins. Among the predicted surface-exposed proteins, 396 were core genes and 193 were variable genes. Of these 589 proteins, 312 were successfully cloned and expressed in *Escherichia coli*. Each antigen was tested in a mouse maternal immunization model to evaluate their ability to confer protection. The final candidate vaccine formulation comprises a combination of the four best protective antigens, which confers 59-100% protection against a panel of 12 GBS isolates, including the major serotypes responsible for disease (Maione et al. 2005).

Multi-genome sequencing has also been applied to different streptococcal species (Group A *Streptococcus* and *Streptococcus pneumoniae*), and has been shown to be a successful approach not only for vaccine antigen selection but also for the discovery of new important virulence factors such as streptococcal pili (Lauer et al. 2005; Mora et al. 2005; Barocchi et al. 2006).

Subtractive Reverse Vaccinology: the third phase

Comparative genome analysis of pathogenic strains has primarily concentrated on the comparison of pathogenic isolates of a species, aiming to identify antigens that could lead to maximum coverage in a universal vaccine. However, many bacterial species consist of commensal and pathogenic strains that both exist in the human host. The analysis of the genome of a nonpathogenic isolate, compared to a pathogenic isolate, could provide the information necessary for the identification of antigens that are critical in pathogenesis and responsible for establishing successful interactions with the host. Using a subtractive comparative genome analysis, genes conserved between pathogenic and nonpathogenic strains could be discarded, reducing the number of candidates and, consequently, reducing the time for development of a vaccine. In this regard, a new concept of reverse vaccinology, named subtractive reverse vaccinology, was applied for the first time to pathogenic and nonpathogenic *E. coli* species (Moriel et al. 2010; Brzuszkiewicz et al. 2006) (Fig. 2). *E. coli* is a commensal member of the gastro-

intestinal flora of most mammals. However, several facultative and obligate pathogenic variants exist which cause various types of intestinal or extraintestinal infections in man and animals. Comparative genome analysis of *E. coli* isolates revealed that genome evolution in these bacteria cannot be simply described by a “backbone and flexible gene pool” model, but must also be described by the repeated occurrence of insertions and deletions in certain parts of the genome (Brzuszkiewicz et al. 2006).

Recently, Moriel *et al.* (2010) described an innovative approach based on subtractive reverse vaccinology to develop a vaccine against extra-intestinal pathogenic *E. coli* (ExPEC) infections. As a first step, the neonatal meningitis-associated *E. coli* K1 strain was sequenced. This sequence was then compared to that of non-pathogenic *E. coli* strains, to minimize the potential impact of a vaccine on the equilibrium of the normal human intestinal flora. Of particular interest, this comparison revealed the presence of 19 genomic islands that were absent in the non-pathogenic genome. Many of these genomic islands contained the typical features of the pathogenicity islands (PAI), such as the presence of an integrase and tRNA genes and a different G+C% content compared to the core genome. Subsequently, by *in silico* analysis, specific antigens shared by ExPEC strains that were predicted to be surface associated or secreted were selected, whereas the presence and level of sequence similarity of these antigens in the non-pathogenic strains was used as exclusion criteria (hence the term subtractive). Among the predicted surface-exposed proteins, nine vaccine candidates were identified as being able to confer protection in a mouse model of sepsis. Interestingly, the gene encoding the most promising antigen was found both in pathogenic and non-pathogenic strains, and is located downstream of a type two secretion system (T2SS); however, in the non-pathogenic strains this T2SS is truncated, and consequently the antigen is expressed intracellularly, but not secreted. In contrast, in pathogenic strains, this protein is secreted into the supernatant, suggesting that it may play a role in virulence and interaction with host cells.

The subtractive reverse vaccinology approach used in this work lead to the identification of nine potential antigens present in ExPEC strains, but in some cas-

es these antigens are also present in the intestinal *E. coli* pathotypes, suggesting that they may be useful for a broadly cross-protective universal *E. coli* vaccine.

Vaccine design in the post-genomic era

With the development and rapid advancement of DNA sequencing technology, from the late 1990s up until now, the way in which vaccine candidates are identified has radically changed. Based on the availability of the entire genome sequences of an organism, new disciplines of molecular biology have emerged. These post-genomic technologies comprise a combination of transcriptomics, proteomics and immunomics (see Fig. 2). These strategies are generally referred to as “Functional Genomics”, and complement *in silico* antigen discovery, not only by directly examining the genetic content but also the transcription and expression profiles of pathogens. Structural genomics is an additional field that has emerged in the post-genomic era, which is increasingly being applied to vaccine development.

Transcriptomics

The study of global changes in bacterial gene expression is essential for understanding pathogenesis and survival in the host. For vaccine antigen discovery, it is important to know which genes are regulated *in vivo* during infection, as they could represent ideal protective vaccine candidates.

The complete set of transcripts of an organism, defined as the transcriptome, can be studied using a DNA microarray (or microchip), a powerful genomic technology (Brown and Botstein 1999; Cheung et al. 1999; Lipshutz et al. 1999; Lockhart and Winzeler 2000). DNA microarrays are particularly attractive in that DNA chips carrying the entire bacterial genome can be easily prepared and used for several applications including global gene expression profiling of the genes that are up- and down-regulated in a particular growth condition, genotyping and DNA sequencing. The first example where microarray-based transcriptional profiling was successfully used to identify potential vaccine candidates was in the case of *N. meningitidis* serogroup B (Grifantini et al. 2002). Transcriptional profil-

ing studies were performed under experimental conditions mimicking certain aspects of host–pathogen interactions, such as adherence to host epithelial cells and exposure to human serum. Of the twelve genes found to have increased expression during adhesion, five of these genes encoded proteins that were able to induce bactericidal antibodies. These novel antigens had not been identified by the reverse vaccinology approach described above, and four of these were detected on the bacterial surface only after adhesion to epithelial cells (Grifantini et al. 2002).

The continuing advances in next-generation sequencing have made it possible to use sequencing technology to characterize the transcriptome of an organism (Dhiman et al. 2009; Sorek and Cossart 2010). For example, RNA-seq has recently been used to map the transcriptional start sites and operons of *Helicobacter pylori* (Sharma et al. 2010). As this technology becomes more widely available it will be a valuable tool for antigen discovery and vaccine development.

Proteomics

Using proteomics approach the entire complement of proteins expressed by a cell (the proteome) can be defined, which is a valuable and rapid tool for antigen discovery (Grandi 2001). Advancements in proteomic approaches, such as 2D-PAGE coupled to mass spectrometry (MS), chromatographic techniques and protein arrays, have enabled researchers to explore this novel strategy in order to identify vaccine targets and proteins of therapeutic interest (Kaushik and Sehgal 2008). This approach has been used to provide insight into the composition and function of a specific subset of the proteome, such as the cell envelope of *Salmonella typhimurium* (Qi et al. 1996). Montigiani and colleagues used the approach of genomics combined with proteomics to characterize the surface proteins of *Chlamydia pneumoniae* (Montigiani et al. 2002).

For *Streptococcus pyogenes*, proteolytic enzymes were used to “shave” the bacterial cell under conditions that preserve cell viability, in order to specifically isolate bacterial surface proteins. The peptides released were then analyzed by MS and the peptide sequences identified using information from the published genome sequence data. More than 70 proteins were identified using this method, the ma-

majority of which were confirmed to be surface exposed by flow cytometry. Furthermore, most of the protective antigens described to date were identified in this analysis, plus one novel protective antigen (Rodriguez-Ortega et al. 2006). Other examples where proteomics has also been used to study bacterial pathogenesis and identify vaccine candidates include *Streptococcus agalactiae* (Hughes et al. 2002; Hughes et al. 2003) and *H. influenzae* (Langen et al. 2000; Thoren et al. 2002).

Immunomics

A combination of proteomic-based approaches and serological analysis (termed Immunoproteomics) can identify potential vaccine candidates and provide effective validation of these candidates. The set of proteins of a pathogen that are able to elicit a humoral immune response during the course of infection is called the “immunoproteome” or “antigenome”. Several methods have been developed to enable the high-throughput display of the pathogen’s proteome to the host immune system. SERPA (SERological Proteome Analysis) is a technology that combines proteomic-based approaches with serological analysis and it has been widely applied to antigen discovery and vaccine development. When applied to *Staphylococcus aureus*, this approach led to the identification of 15 highly immunogenic proteins, including known and novel vaccine candidates (Vytvytska et al. 2002).

Another immunomics approach is based on protein microarray technology, where individual proteins are spotted onto microarray chips which are then probed with sera to obtain immunodominant antigen profiles (Davies et al. 2005). This approach has been used to identify antigens in several pathogens, including *Francisella tularensis* (Sundaresh et al. 2007).

Structural genomics and structural vaccinology

High-throughput structural genomics projects have laid the groundwork to better understand the three-dimensional structures of the proteins produced by a species, in particular vaccine and drug targets (Stacy et al. 2011). Optimization of antigens can be achieved using structural biology to combine, expose and/or improve the immunogenicity of epitopes (Dormitzer et al. 2008). The ability to improve candi-

date antigens by rational design and structural vaccinology can provide increased protection against antigenically variable pathogens. For example, in the attempts to engineering conserved and protective antigens for a HIV vaccine, conserved epitopes of HIV antigens have been engineered to be immunogenic, while other non conserved but immunodominant antigens have been engineered to direct the immune system to well conserved epitopes (e.g. immune refocusing) (Burton 2010; Tobin et al. 2008).

Structural vaccinology has also been used to overcome the sequence variability of the serogroup B meningococcal vaccine antigen fHbp. A single antigen was engineering that was able to elicit broadly protective immunity against strains expressing different fHbp variants, by grafting multiple immunodominant epitopes onto a single fHbp backbone molecule (Scarselli et al. 2011). Similarly, a synthetic *S. agalactiae* pilus antigen was engineered that contains the protective domain of each of six variants of the pilus backbone protein, BP-2a (Nuccitelli et al. 2011).

Conclusions

Analysis of whole-genome sequences from a range of pathogens shows their diversity and adaptability to different environments. Valuable information can be obtained from complete genome sequences and this has revolutionized the approaches used for vaccine development. These new approaches start with the complete information of a bacterial genome and its gene products, with an aim to identify surface structures that could be used as vaccine antigens, as well as identify important factors involved in virulence. Moreover, the analysis of the transcriptome and proteome offers the opportunity to gain a better understanding of the biology of the pathogen as well as its interactions with the host immune system.

The encouraging results obtained using the classical Reverse Vaccinology approach with the pathogen *Meningococcus B* have opened the way to a revolution in the field of vaccine development. The traditional reverse vaccinology approach has since been complemented by multigenome comparisons and post genomic approaches, making it possible to address many infectious diseases for

which the development of effective vaccines has not been possible using conventional vaccine approaches.

The current “omics” revolution has provided researchers with advanced technologies, which are decreasing the time it takes to identify and design target antigens. Proteomics and immunoproteomics approaches are powerful tools for the identification of novel bacterial antigens and for the understanding of protein function, and their use is likely to increase in the years to come. Indeed, a complete understanding of protein function and the study of functional complexes between a given macromolecule and its effectors in the host will facilitate the rational design of and vaccines. Likewise, advances in structural vaccinology provide the possibility to improve candidate antigens by rational design leading to increased protection against antigenically variable pathogens. Finally, the increasing application of systems biology to vaccine development will help bridge the gaps in our current understanding of the human immune response to specific antigens and the changes that occur following vaccination (Buonaguro and Pulendran 2011).

The future success of vaccines will not only depend on the advancement of scientific platforms, but also the interdisciplinary involvement of researchers in different fields such as structural biology, physical chemistry, epidemiology, molecular immunology, and engineering. The success of vaccines in the future will be possible only through innovative ideas that will lead to fundamental breakthroughs in understanding of pathogens, the host immune system, and the interactions between these two factors.

References

- Artenstein AW, Grabenstein JD (2008) Smallpox vaccines for biodefense: need and feasibility. *Expert Rev Vaccines* 7 (8):1225-1237.
- Barocchi MA, Ries J, Zogaj X, Hemsley C, Albiger B, Kanth A, Dahlberg S, Fernebro J, Moschioni M, Massignani V, Hultenby K, Taddei AR, Beiter K, Wartha F, von Euler A, Covacci A, Holden DW, Normark S, Rappuoli R, Henriques-Normark B (2006) A pneumococcal pilus influences virulence and host inflammatory responses. *Proc Natl Acad Sci U S A* 103 (8):2857-2862.
- Brown PO, Botstein D (1999) Exploring the new world of the genome with DNA microarrays. *Nat Genet* 21 (1 Suppl):33-37.
- Bruszkiewicz E, Bruggemann H, Liesegang H, Emmerth M, Olschlager T, Nagy G, Albermann K, Wagner C, Buchrieser C, Emody L, Gottschalk G, Hacker J, Dobrindt U (2006) How to become a uropathogen: comparative genomic analysis of extraintestinal pathogenic *Escherichia coli* strains. *Proc Natl Acad Sci U S A* 103 (34):12879-12884.
- Buonaguro L, Pulendran B (2011) Immunogenomics and systems biology of vaccines. *Immunol Rev* 239 (1):197-208.
- Burton DR (2010) Scaffolding to build a rational vaccine design strategy. *Proc Natl Acad Sci U S A* 107 (42):17859-17860.
- Cheung VG, Morley M, Aguilar F, Massimi A, Kucherlapati R, Childs G (1999) Making and reading microarrays. *Nat Genet* 21 (1 Suppl):15-19.
- Davies DH, Liang X, Hernandez JE, Randall A, Hirst S, Mu Y, Romero KM, Nguyen TT, Kalantari-Dehaghi M, Crotty S, Baldi P, Villarreal LP, Felgner PL (2005) Profiling the humoral immune response to infection by using proteome microarrays: high-throughput vaccine and diagnostic antigen discovery. *Proc Natl Acad Sci U S A* 102 (3):547-552.
- Dhiman N, Smith DI, Poland GA (2009) Next-generation sequencing: a transformative tool for vaccinology. *Expert Rev Vaccines* 8 (8):963-967.
- Dormitzer PR, Ulmer JB, Rappuoli R (2008) Structure-based antigen design: a strategy for next generation vaccines. *Trends Biotechnol* 26 (12):659-667.

- Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269 (5223):496-512.
- Giuliani MM, Adu-Bobie J, Comanducci M, Arico B, Savino S, Santini L, Brunelli B, Bambini S, Biolchi A, Capecchi B, Cartocci E, Ciucchi L, Di Marcello F, Ferlicca F, Galli B, Luzzi E, Massignani V, Serruto D, Veggi D, Contorni M, Morandi M, Bartalesi A, Cinotti V, Mannucci D, Titta F, Ovidi E, Welsch JA, Granoff D, Rappuoli R, Pizza M (2006) A universal vaccine for serogroup B meningococcus. *Proc Natl Acad Sci U S A* 103 (29):10834-10839.
- Grandi G (2001) Antibacterial vaccine design using genomics and proteomics. *Trends Biotechnol* 19 (5):181-188.
- Grifantini R, Bartolini E, Muzzi A, Draghi M, Frigimelica E, Berger J, Ratti G, Petracca R, Galli G, Agnusdei M, Giuliani MM, Santini L, Brunelli B, Tettelin H, Rappuoli R, Randazzo F, Grandi G (2002) Previously unrecognized vaccine candidates against group B meningococcus identified by DNA microarrays. *Nature biotechnology* 20 (9):914-921.
- Hughes MJ, Moore JC, Lane JD, Wilson R, Pribul PK, Younes ZN, Dobson RJ, Everest P, Reason AJ, Redfern JM, Greer FM, Paxton T, Panico M, Morris HR, Feldman RG, Santangelo JD (2002) Identification of major outer surface proteins of *Streptococcus agalactiae*. *Infect Immun* 70 (3):1254-1259.
- Hughes MJ, Wilson R, Moore JC, Lane JD, Dobson RJ, Muckett P, Younes Z, Pribul P, Topping A, Feldman RG, Santangelo JD (2003) Novel protein vaccine candidates against Group B streptococcal infection identified using alkaline phosphatase fusions. *FEMS Microbiol Lett* 222 (2):263-271.
- Kaushik DK, Sehgal D (2008) Developing antibacterial vaccines in genomics and proteomics era. *Scand J Immunol* 67 (6):544-552.
- Kelly D, Moxon ER (2004) Is *Haemophilus influenzae* type b disease finished? *Adv Exp Med Biol* 549:221-229.

- Langen H, Takacs B, Evers S, Berndt P, Lahm HW, Wipf B, Gray C, Fountoulakis M (2000) Two-dimensional map of the proteome of *Haemophilus influenzae*. *Electrophoresis* 21 (2):411-429.
- Lauer P, Rinaudo CD, Soriani M, Margarit I, Maione D, Rosini R, Taddei AR, Mora M, Rappuoli R, Grandi G, Telford JL (2005) Genome analysis reveals pili in Group B *Streptococcus*. *Science* 309 (5731):105.
- Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ (1999) High density synthetic oligonucleotide arrays. *Nat Genet* 21 (1 Suppl):20-24.
- Lockhart DJ, Winzeler EA (2000) Genomics, gene expression and DNA arrays. *Nature* 405 (6788):827-836.
- Maione D, Margarit I, Rinaudo CD, Masignani V, Mora M, Scarselli M, Tettelin H, Brettoni C, Iacobini ET, Rosini R, D'Agostino N, Miorin L, Buccato S, Mariani M, Galli G, Nogarotto R, Dei VN, Vegni F, Fraser C, Mancuso G, Teti G, Madoff LC, Paoletti LC, Rappuoli R, Kasper DL, Telford JL, Grandi G (2005) Identification of a Universal Group B *Streptococcus* Vaccine by Multiple Genome Screen. *Science* 309 (5731):148-150.
- Medini D, Donati C, Tettelin H, Masignani V, Rappuoli R (2005) The microbial pan-genome. *Curr Opin Genet Dev* 15 (6):589-594.
- 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, Grandi G (2002) Genomic approach for analysis of surface proteins in *Chlamydia pneumoniae*. *Infect Immun* 70 (1):368-379.
- Mora M, Bensi G, Capo S, Falugi F, Zingaretti C, Manetti AG, Maggi T, Taddei AR, Grandi G, Telford JL (2005) Group A *Streptococcus* produce pilus-like structures containing protective antigens and Lancefield T antigens. *Proc Natl Acad Sci U S A* 102 (43):15641-15646.
- Moriel DG, Bertoldi I, Spagnuolo A, Marchi S, Rosini R, Nesta B, Pastorello I, Corea VA, Torricelli G, Cartocci E, Savino S, Scarselli M, Dobrindt U, Hacker J, Tettelin H, Tallon LJ, Sullivan S, Wieler LH, Ewers C, Pickard D, Dougan G, Fontana MR, Rappuoli R, Pizza M, Serino L (2010) Identification

of protective and broadly conserved vaccine antigens from the genome of extraintestinal pathogenic *Escherichia coli*. *Proc Natl Acad Sci U S A* 107 (20):9072-9077.

Nuccitelli A, Cozzi R, Gourlay LJ, Donnarumma D, Necchi F, Norais N, Telford JL, Rappuoli R, Bolognesi M, Maione D, Grandi G, Rinaudo CD (2011) Structure-based approach to rationally design a chimeric protein for an effective vaccine against Group B *Streptococcus* infections. *Proc Natl Acad Sci U S A* 108 (25):10278-10283.

Pizza M, Scarlato V, Maignani V, Giuliani MM, Arico B, Comanducci M, Jennings GT, Baldi L, Bartolini E, Capecchi B, Galeotti CL, Luzzi E, Manetti R, Marchetti E, Mora M, Nuti S, Ratti G, Santini L, Savino S, Scarselli M, Storni E, Zuo P, Broecker M, Hundt E, Knapp B, Blair E, Mason T, Tettelin H, Hood DW, Jeffries AC, Saunders NJ, Granoff DM, Venter JC, Moxon ER, Grandi G, Rappuoli R (2000) Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* 287 (5459):1816-1820.

Plotkin SA (2003) Vaccines, vaccination, and vaccinology. *J Infect Dis* 187 (9):1349-1359.

Plotkin SA (2009) Vaccines: the fourth century. *Clin Vaccine Immunol* 16 (12):1709-1719.

Purcell AW, McCluskey J, Rossjohn J (2007) More than one reason to rethink the use of peptides in vaccine design. *Nat Rev Drug Discov* 6 (5):404-414.

Qi SY, Moir A, O'Connor CD (1996) Proteome of *Salmonella typhimurium* SL1344: identification of novel abundant cell envelope proteins and assignment to a two-dimensional reference map. *J Bacteriol* 178 (16):5032-5038.

Rappuoli R (2000) Reverse vaccinology. *Curr Opin Microbiol* 3 (5):445-450.

Rodriguez-Ortega MJ, Norais N, Bensi G, Liberatori S, Capo S, Mora M, Scarselli M, Doro F, Ferrari G, Garaguso I, Maggi T, Neumann A, Covre A, Telford JL, Grandi G (2006) Characterization and identification of vaccine candidate

- proteins through analysis of the group A *Streptococcus* surface proteome. *Nature biotechnology* 24 (2):191-197.
- Scarselli M, Arico B, Brunelli B, Savino S, Di Marcello F, Palumbo E, Veggi D, Ciucchi L, Cartocci E, Bottomley MJ, Malito E, Lo Surdo P, Comanducci M, Giuliani MM, Cantini F, Dragonetti S, Colaprico A, Doro F, Giannetti P, Pallaoro M, Brogioni B, Tontini M, Hilleringmann M, Nardi-Dei V, Banci L, Pizza M, Rappuoli R (2011) Rational design of a meningococcal antigen inducing broad protective immunity. *Sci Transl Med* 3 (91):91ra62.
- Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeiss S, Sittka A, Chabas S, Reiche K, Hackermuller J, Reinhardt R, Stadler PF, Vogel J (2010) The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* 464 (7286):250-255.
- Sorek R, Cossart P (2010) Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. *Nat Rev Genet* 11 (1):9-16.
- Stacy R, Begley DW, Phan I, Staker BL, Van Voorhis WC, Varani G, Buchko GW, Stewart LJ, Myler PJ (2011) Structural genomics of infectious disease drug targets: the SSGCID. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 67 (Pt 9):979-984.
- Sundaresh S, Randall A, Unal B, Petersen JM, Belisle JT, Hartley MG, Duffield M, Titball RW, Davies DH, Felgner PL, Baldi P (2007) From protein microarrays to diagnostic antigen discovery: a study of the pathogen *Francisella tularensis*. *Bioinformatics* 23 (13):i508-518.
- Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, Deboy RT, Davidsen TM, Mora M, Scarselli M, Margarit y Ros I, Peterson JD, Hauser CR, Sundaram JP, Nelson WC, Madupu R, Brinkac LM, Dodson RJ, Rosovitz MJ, Sullivan SA, Daugherty SC, Haft DH, Selengut J, Gwinn ML, Zhou L, Zafar N, Khouri H, Radune D, Dimitrov G, Watkins K, O'Connor KJ, Smith S, Utterback TR, White O, Rubens CE, Grandi G, Madoff LC, Kasper DL, Telford JL, Wessels MR, Rappuoli R, Fraser CM (2005) Genome analysis of multiple pathogenic

- isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proc Natl Acad Sci U S A* 102 (39):13950-13955.
- Tettelin H, Massignani V, Cieslewicz MJ, Eisen JA, Peterson S, Wessels MR, Paulsen IT, Nelson KE, Margarit I, Read TD, Madoff LC, Wolf AM, Beanan MJ, Brinkac LM, Daugherty SC, DeBoy RT, Durkin AS, Kolonay JF, Madupu R, Lewis MR, Radune D, Fedorova NB, Scanlan D, Khouri H, Mulligan S, Carty HA, Cline RT, Van Aken SE, Gill J, Scarselli M, Mora M, Iacobini ET, Brettoni C, Galli G, Mariani M, Vegni F, Maione D, Rinaudo D, Rappuoli R, Telford JL, Kasper DL, Grandi G, Fraser CM (2002) Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V *Streptococcus agalactiae*. *Proc Natl Acad Sci U S A* 99 (19):12391-12396.
- Thoren K, Gustafsson E, Clevnert A, Larsson T, Bergstrom J, Nilsson CL (2002) Proteomic study of non-typable *Haemophilus influenzae*. *J Chromatogr B Analyt Technol Biomed Life Sci* 782 (1-2):219-226.
- Tobin GJ, Trujillo JD, Bushnell RV, Lin G, Chaudhuri AR, Long J, Barrera J, Pena L, Grubman MJ, Nara PL (2008) Deceptive imprinting and immune refocusing in vaccine design. *Vaccine* 26 (49):6189-6199.
- Toneatto D, Ismaili S, Ypma E, Vienken K, Oster P, Dull P (2011) The first use of an investigational multicomponent meningococcal serogroup B vaccine (4CMenB) in humans. *Hum Vaccin* 7 (6):646-653.
- Vytvytska O, Nagy E, Bluggel M, Meyer HE, Kurzbauer R, Huber LA, Klade CS (2002) Identification of vaccine candidate antigens of *Staphylococcus aureus* by serological proteome analysis. *Proteomics* 2 (5):580-590.

Figure Legends.

Figure 1. Overview of the Reverse Vaccinology approach, using Meningococcal serogroup B as an example. The genome sequence of the virulent MenB strain MC58 was used to identify open reading frames (ORFs) predicted to encode proteins that were surface exposed (i.e., secreted or located in the outer membrane). These proteins were then expressed and purified in *E. coli*. The recombinant proteins were then used to immunize mice to generate sera for further analysis. Antibodies generated in mice were then used to confirm the surface exposure of the candidate antigens by flow cytometry, and to identify proteins that induced bactericidal activity. The top candidates were then further investigated for their distribution and sequence conservation in a panel of strain representative of the disease strains circulating worldwide. This reverse vaccinology approach resulted in identification of several novel vaccine candidates for further preclinical study. Five candidates were chosen for clinical development: NadA, and fHbp-GNA2091 and NHBA-GNA1030 fusion proteins.

Figure 2. Overview of the approaches used for antigen discovery and vaccine development in the genomic and post-genomic era. Reverse Vaccinology started with the analysis of a single genome, then evolved to include pan-genomics, comparative genomics and subtractive genomics. Transcriptomics, proteomics, immunomics and structural biology are now commonly used in the post-genomic era as stand alone approaches as well as in combination with genomics to refine the process of antigen discovery and vaccine development.

REVERSE VACCINOLOGY

Meningococcal serogroup B Vaccine Development

ANTIGEN IDENTIFICATION

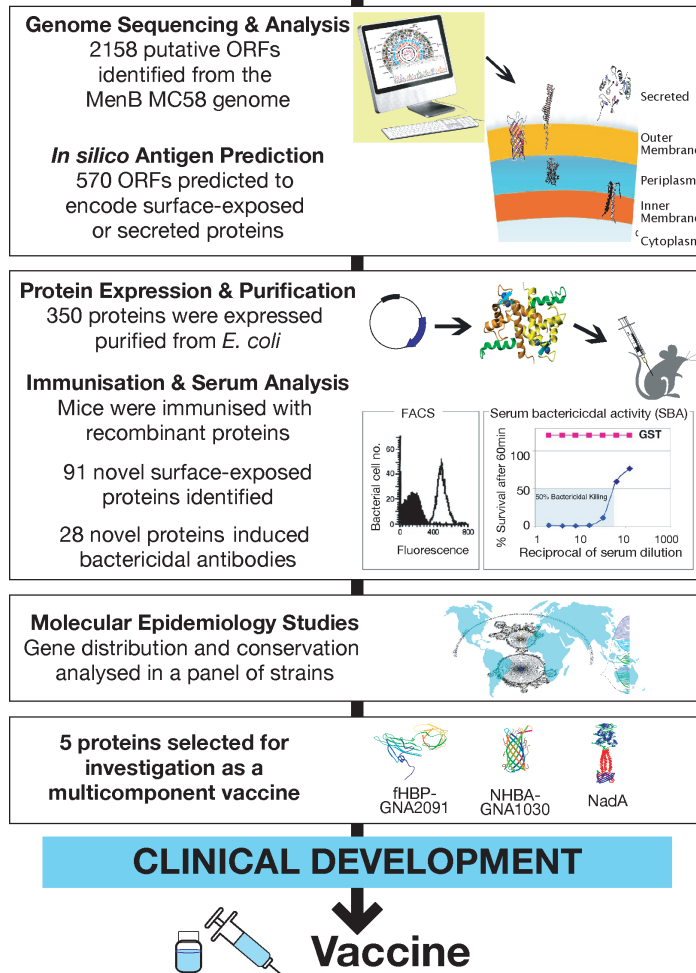


Figure 1

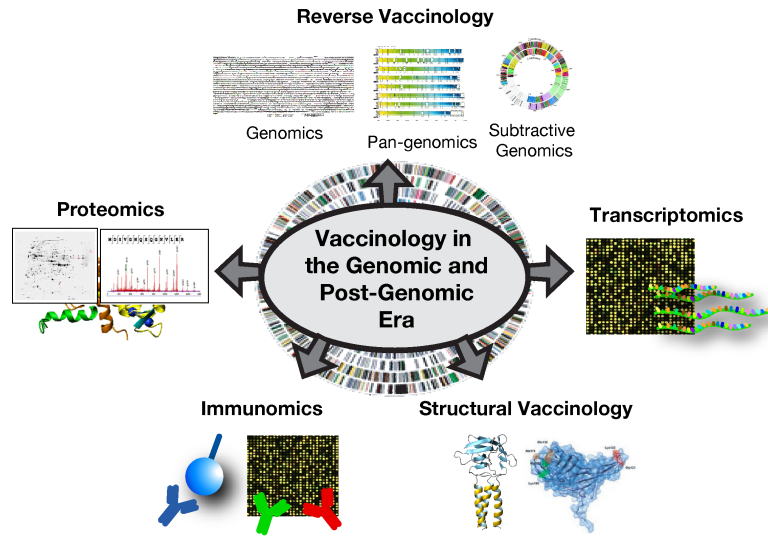


Figure 2