Abstract

Current vaccine development efforts are mainly focused on opportunistic pathogens that are frequent colonizers and have complex pathogenesis and interaction with the human host. Therefore, more sophisticated and comprehensive vaccine development approaches have to be considered than for strictly pathogenic bacteria with well-defined virulence mechanisms that typically rely on toxin production.

Multigenome analysis and genomic DNA-based screening approaches represent powerful strategies for identifying proteinaceous vaccine candidates. The two approaches we review here are the reverse vaccinology and the ANTIGENome technologies that have been applied for numerous human pathogens and resulted in clinical vaccine candidates. In both cases, the primary selections—that are based on in silico prediction or human antibody response, respectively—are complemented with a series of in vitro assays to preselect vaccine candidates for testing in animal models of efficacy to ultimately single out the vaccine antigens destined to move into development.

When applied to the same pathogen, the two approaches appear to identify overlapping pools of antigens that not completely superimpose, suggesting that the methods might complement each other. Importantly, the conclusions from the application of two technologies are similar: broadly protective antigens rarely exist, and combination of several protein antigens is necessary for the development of universal vaccines.
2.1 Introduction

Vaccine development against bacterial diseases has been the most successful against strictly pathogenic organisms when disease is typically caused by powerful toxins (e.g., anthrax, tetanus, pertussis, diphtheria). Neutralization of these secreted components is the cornerstone of protective immunity induced by vaccination. However, most of the human bacterial diseases are associated with microbes that do not rely on the production of single toxins but have a complex pathogenesis with redundant virulence mechanisms. The development of efficacious vaccines against these types of pathogens is largely based on the induction of antibodies with specificity for bacterial antigens that are accessible extracellularly and have high complement-fixing activity. This strategy stems from the fact that antigen–antibody complexes can activate the complement system which, in turn, can either directly kill pathogens (Gram-negative bacteria) or stimulate pathogen uptake by phagocytic cells.

The current vaccines that rely on antibody-mediated complement fixation contain capsular polysaccharide antigens (e.g., pneumococcus, *Haemophilus influenzae*, meningococci). Certain pathogens do not have capsules or, as it is the case for meningococcus serotype B, their sugar composition is too similar to that found in some human tissues. In both cases, development of polysaccharide-based vaccines is precluded. Furthermore, a given bacterial species can produce chemically distinct capsules, which are not immunologically cross-reactive. The number of capsule variants (serotypes) within the same species can be very high, as it is the case of *S. pneumoniae* (pneumococcus), for which more than 90 serotypes have been described so far. This high variability imposes the selection of a subset of capsular polysaccharides with the consequence that the derived vaccines are, first, not universal and, second, may favor serotype replacement and escape mechanism induced by immune pressure following mass immunization.

Therefore, novel approaches to vaccine development that target nonpolysaccharide antigens are needed. Proteins are generally more antigenic than polysaccharides and induce antibody subclasses with high complement-fixing capacity. Bacterial proteins that are accessible extracellularly can be considered viable vaccine antigens. Antibodies binding to bacterial surface proteins can mediate protection not only by inducing direct bactericidal activity or enhancing opsonophagocytic killing but also by blocking their function, often involved in virulence and inhibition of host protective responses.

The availability of complete genome sequences of bacterial pathogens has enabled the systematic search for protein antigens among all genome-encoded proteins. The aim of this chapter is to outline two approaches that have been developed and fine-tuned over the last 15 years in our laboratories and have proved to be particularly successful. More specifically, after a description of the approaches, we highlight their respective strengths and weaknesses, and we finally discuss differences and possible synergies.
2.2 Antigen Selection Based on Genome Sequence: Reverse Vaccinology

The rationale behind reverse vaccinology is discouragingly simple: assuming that for a given pathogen at least one protein antigen exists which induces a protective immune response, the identification of such antigen is guaranteed if one expresses all pathogen’s proteins and tests them in a reliable assay. Mandatory for the success of the approach is the availability of (1) a well-annotated genome sequence of the pathogen under investigation, (2) an efficient platform for heterologous protein expression starting from the PCR amplification of all annotated genes, and (3) a robust model which truly mimics human infection and/or immunological mechanisms that in humans correlate with protection. As discussed later in this chapter, frustrations and excitements of those who decide to embark in a reverse vaccinology-based approach will strongly depend upon how carefully the whole strategy is designed and, most importantly, the screening assay selected.

In the following section, three successful examples of reverse vaccinology are first described. Subsequently, the same examples are used to highlight the critical aspects of the approach and to provide general rules which, if followed, greatly enhance the chances to discover new efficacious vaccines.

2.2.1 Meningococcus B Vaccine

Although historically not the first example of reverse vaccinology (the approach was in fact conceptualized a few years before; Grandi 2001), the development of a protein-based meningococcus B (MenB) vaccine by Chiron (now Novartis Vaccines and Diagnostics) will surely survive history as the first successful application of this technology (Pizza et al. 2000). Indeed, the approach was so successful that the vaccine is now under registration in Europe for adolescents. In essence, the genome sequence of MC58 strain was determined by TIGR (Rockville, USA), and approximately 600 genes encoding surface-associated proteins were subsequently cloned in *E. coli*. About half (350) of the corresponding proteins were expressed, purified, and used to immunize groups of mice, thus producing immune sera for each of the proteins. Finally, the sera were used in a classical bactericidal assay, whereby an aliquot of each serum was incubated with MC58 bacteria and complement, and the capacity of each serum to kill MenB was established by colony counting. In total, 29 proteins were found to elicit bactericidal antibodies in mice. Subsequent extensive sequence analysis of the corresponding genes from hundreds of MenB isolates revealed that five of these proteins were sufficiently conserved to provide broad coverage when combined in a single adjuvanted formulation (Giuliani et al. 2006). The five proteins are all included in the vaccine under registration.
2.2.2 Group B Streptococcus Vaccine

Taking advantage of the experience gained from the MenB vaccine project, the reverse vaccinology approach for group B Streptococcus (GBS) started with the sequencing of eight clinical isolates belonging to different serotypes (Tettelin et al. 2005). In fact, the Chiron’s team realized that the genome variability within isolates of the same bacterial species requires the up-front selection of genes that, in addition of encoding surface-associated proteins, are also sufficiently conserved. Again, approximately 600 genes were cloned in E. coli, and 318 recombinant proteins were successfully purified. The selection of protective antigens was carried out using a quite cumbersome but highly reliable mouse model. According to this model, adult female mice are immunized with each protein and subsequently mated. Pups are finally challenged with a lethal dose of GBS 24–48 h after delivery. Should the antigen used for immunization elicit opsonophagocytic antibodies, such antibodies would be transferred from the mother to the pups which would survive the GBS challenge. The selection of the model was based on a large body of human data showing that newborns from mothers with elevated titers against GBS polysaccharides have a much lower risk of getting infected by GBS. Differently from what observed for MenB, the number of protective antigens discovered after the screening process was small: four antigens as opposed to the 29 MenB antigens (Maione et al. 2005). One of the protective antigen was already known (Brodeur et al. 2000), while the other three, originally annotated as “hypothetical/unknown,” turned out to be part of pilus-like structures never described before in pathogenic streptococci (Lauer et al. 2005). Such organelles were also subsequently found in group A Streptococcus (Mora et al. 2005) and Streptococcus pneumoniae (Barocchi et al. 2006). An extensive analysis of pilus distribution in GBS revealed the existence of three pilus variants and showed that all GBS isolates express at least one of the three variants. This explains why a three protein combination including one component from each pilus conferred remarkable cross-protection in mice against the challenge with a large panel of GBS isolates belonging to different serotypes (Margarit et al. 2009).

2.2.3 Chlamydia trachomatis Vaccine

Protection against both MenB and GBS requires the elicitation of functional antibodies capable of killing the pathogen in the presence of complement (and phagocytic cells in the case of GBS). A plausible question is whether reverse vaccinology is applicable also in the case where cell-mediated immunity is needed to prevent infection. The example recently published on a candidate vaccine against Chlamydia trachomatis (Finco et al. 2011) indicates that this is in fact the case. C. trachomatis is a human pathogen that because of its being an obligate intracellular bacterium requires γ-interferon-producing CD4+ T cells to be neutralized. The key role of CD4+ T cells in Chlamydia immunity was extensively demonstrated in mice by showing that passive transfer of CD4+ T cells from mice previously
exposed to *C. trachomatis* (when experimentally challenged with *C. trachomatis*, mice are infected but are capable of clearing the infection after approximately 4 weeks) protects naive mice from infection. On the basis of this experimental evidence, splenocytes from mice infected by Chlamydia were exposed to a large panel of highly purified *C. trachomatis* recombinant proteins in order to select those antigens which specifically stimulated γ-interferon production in CD4+ T cells. Twenty-one antigens were positive to this assay. Combinations of these antigens were used to immunize mice systemically in the presence of a Th1 adjuvant and two four-antigen combinations turned out to be remarkably protective.

### 2.2.4 Take-Home Lessons from 15-Year Experience in Reverse Vaccinology

It is now more than 15 years that Stephen Johnston and coworkers proposed the first “from-genome-to-vaccine” approach (Barry et al. 1995). The team constructed a plasmid library of the whole mycoplasma genome and used it in DNA immunization to select the genes which protect mice against mycoplasma challenge. Although the approach was not further pursued because of the inefficiency of DNA immunization when applied in high-throughput modalities, it represents a hallmark in vaccinology in that it set the basis of classical “reverse vaccinology” in which genes are not used directly in antigen screening but rather as template for the production of their encoding proteins. Reverse vaccinology has now been applied to a sufficiently large number of bacterial pathogens to allow a critical evaluation of the strategy and to propose a recipe to greatly enhance the chances of success. The “key ingredients” of the technology are discussed in detail below.

#### 2.2.4.1 Biological Assays

It may seem obvious, but starting a reverse vaccinology project in the absence of a robust and reliable assay to use for antigen screening is a scientific suicide. “Robust and reliable” means that the assay must select antigens on the basis of properties which ultimately correlate with protective immunity in humans. To underline the importance of the assay, suffice to say that the very first attempt to use reverse vaccinology was a failure. Lissolo and coworkers applied this approach for the first time to the discovery of a vaccine against *Helicobacter pylori*. Unfortunately, the mouse model of *H. pylori* infection they used is highly variable, and therefore, very inconsistent results in protective antigen selection were obtained. The “beauty” of MenB and GBS is that for both of them, protection in humans is mediated by bactericidal antibodies, and therefore, by using assays which screened antigens for their capacity to elicit functional antibodies, effective vaccines were developed.

#### 2.2.4.2 Strain Variability

The rapid advance of DNA sequence technologies has made available the genome sequences of several isolates of the same bacterial species. Comparative analysis of these genomes has revealed that many species have a “core genome” constituted by
genes highly conserved in all isolates and a “variable genome,” which include the genes that are present only in a fraction of genomes or whose sequence conservation is low (Tettelin et al. 2005). As it turns out, protective antigens are usually encoded by the “variable genome.” This has two important consequences in vaccinology. First, universal antigens (that is to say, antigens that elicit protection against all isolates of the same species) rarely exist. Second, in order to develop a universal vaccine, a combination of antigens, each inducing protection against a fraction of isolates, is needed. In the case of MenB, the Chiron scientists started from the genome sequence of a single isolate and subsequently looked at the conservation of the selected protective antigens by sequence analysis of different isolates. This approach has two limitations: first, it leads to unnecessary workload since nonconserved antigens, which would be discarded anyway, enter the screening process, the real bottleneck of reverse vaccinology (see below). Second, and most important, genes not present in the sequenced genome but sufficiently conserved in other isolates are missed.

Considering the ease with which bacterial genomes can be sequenced nowadays, it is highly recommended that a reverse vaccinology project starts with the selection and sequencing of a large panel (>100) of clinically relevant isolates. This allows a much more rational and precise selection of those genes to be subsequently expressed and screened for protective activity.

2.2.4.3 Filtering of Proteins to Be Screened in the Biological Assay

Although reverse vaccinology is a high-throughput approach which originates from the unbiased idea of testing all bacterial proteins for protection, practically speaking, only a fraction of the entire proteome enters the screening phase. The main reason for that is because the biological assay is the most time-consuming, laborious, and expensive step of the entire process and therefore proteins that according to selected criteria are unlikely to be relevant for protection are eliminated up-front. The need for this preselection becomes particularly evident if one considers the assay used for the identification of protective GBS antigens, assay which, considering the time needed for animal immunization, mating, pups delivery, challenge of newborn mice, and data collection, lasts approximately 60 days per antigen!

According to the classical reverse vaccinology, bioinformatics is used for protein preselection. Starting from the assumption that protective antigens belong to the categories of “surface-associated” and secreted proteins, the genome of the pathogen of interest is scanned using algorithms designed to predict this group of proteins. This filtering step usually brings the number of proteins which enters the screening pipeline down to 600–700 hundreds (approximately one-third of the total coding genes). Considering the time required completing the biological assay, this number is still very high.

Recently, two strategies have been developed for a more accurate selection of surface-exposed proteins, again, the proteins that are most relevant for the induction of protective antibodies. The first of these strategies has been designed for surface-exposed protein identification in Gram-positive bacteria. It consists of enzymatic
“shaving” of the bacterial surface with proteolytic enzymes under conditions that preserve the integrity and viability of bacterial cells. After digestion, the released peptides are separated from the “shaved” bacteria and subjected to mass spectrometry for protein identification. The approach has been successfully applied in surface protein identification of group A Streptococcus and group B Streptococcus (Rodríguez-Ortega et al. 2006; Doro et al. 2009). In Gram-negative bacteria, the “shaving” strategy is difficult to apply because of the relative fragility of the cells, which tend to die during protease treatment. For these bacteria, a different approach has been recently developed, which exploits their natural propensity to release outer membrane vesicles (OMVs). If specific mutations are selected, the amount of released OMVs, usually too minute for practical purposes, can substantially increase and reach values of several milligrams (in protein content) per liter of culture. Because of their small size (50–100 nm in diameter), OMVs can be easily separated from the bacterial cells by centrifugation and/or ultrafiltration and can therefore be subjected to mass spectrometry analysis for protein identification. The proteome characterization of OMVs from a Neisseria meningitidis group B isolate and from a pathogenic E. coli strain has been recently reported (Ferrari et al. 2006; Berlanda Scorza et al. 2008). The data show that OMVs are almost exclusively constituted by outer membrane proteins and few periplasmic proteins.

To exemplify the power of proteomic characterization of surface proteins in vaccine discovery, suffice it to say that our analysis of MenB OMVs led to the identification of approximately 40 integral membrane proteins and lipoproteins including the five proteins which constitute the MenB vaccine now ready to go to the market (Giuliani et al. 2006). Therefore, retrospectively, if 14 years ago we had applied the proteomic approach for our MenB vaccine discovery project, we would have identified our vaccine components by screening forty recombinant proteins rather than three hundred and fifty!

2.2.5 “Recipe” for Reverse Vaccinology

On the basis of what has been said above, a “vade mecum” for reverse vaccinology can be outlined which, if rigorously followed, should substantially enhance the probability of identifying the protective antigens against the pathogen of interest. The recommendations are given for selecting antigens eliciting protective antibody responses and can be summarized as follows:

1. Sequence the genomes of a sufficiently large panel of isolates—Isolates belonging to the same bacterial species can vary quite substantially in genome sequencing and proteome expression, and this is particularly true for those pathogens which choose humans as the only natural host. Therefore, it is highly recommended to sequence the genomes of several clinically relevant isolates in order to establish the overall level of conservation and be in the position to make a first bioinformatic selection on which proteins should be tested for protection in the animal models and which should be excluded. As already pointed out, single antigens with broad protective activity are rare, and it is
clear that highly effective vaccines can be developed only by combining multiple antigens. However, since for practical reasons the cocktail cannot include too many antigens, antigens with low degree of conservation (approx. \(<30\%\)) should be excluded a priori.

2. **Analyze surface and secreted proteomes from several clinically relevant isolates**—The original application of reverse vaccinology envisaged the bioinformatic selection of the proteins to be expressed and tested in the biological assay(s). It is strongly recommended to replace the in silico analysis with the proteomic characterization of surface and secreted proteins, starting from a sufficiently large panel of strains. If properly carried out, this approach leads to the identification of no more than 30–40 sufficiently conserved antigens, which, being surface-associated and/or secreted, constitute the pool of proteins with the highest probability of being protective. Therefore, these are the proteins to be expressed and tested in the biological models.

3. **Select the proper biological assay to be used for antigen screening**—The assay should be selected on the basis of the information available on the surrogates of protection in humans. In general, in vivo models are used according to which highly lethal doses of the pathogen are given to animals, and animals can survive the challenge only if bactericidal antibodies are induced. These are clearly “biased” models, which might exclude antigens which would work in humans through the elicitation of different protection mechanisms. However, there is no doubt that an antigen capable of inducing high titers of killing antibodies has a strong probability of becoming a good vaccine component. The biological assay should be set up using different clinical isolates. This is a particularly important step in that, at the end, the vaccine should provide broad coverage, and therefore, its preclinical efficacy should be tested against several clinically relevant isolates.

4. **Select the proper antigen combination**—The primary screening of recombinant antigens with the biological assay usually brings to the identification of a few antigens which are protective against a fraction of the strains tested in the models. The last step is to combine the most promising ones in different formulations and select the one which provides the best coverage.

### 2.3 B-Cell Antigen Selection Based on What Is Recognized by the Human Immune System During Disease

Antibodies are the most crucial effectors of protection against extracellular bacteria. Although not all of them are protective, their presence in serum and other body fluids constitutes a molecular imprint of the in vivo expression of the corresponding antigens. It is our philosophy at Intercell to use the human humoral response to bacteria as our guide to select protein antigens as potential novel vaccine candidates.

The important elements of the ANTIGENome technology are (1) the bacterial platform for the surface expression of peptides (Etz et al. 2001), (2) the
comprehensive and high complexity libraries generated by random genome fragmentation (Henics et al. 2003), (3) the purified immunoglobulins from well-characterized human donors, and (4) the availability of pathogen genome sequences. The set of antigens identified by screening genomic surface display libraries with purified human antibodies are referred as the ANTIGENome of a pathogen and typically consist of approximately 100 ORFs, that is, approximately 5% of the annotated genome-encoded bacterial proteins.

Affinity selection of libraries is carried out by streptavidin-coated magnetic beads and biotinylated human Igs and results in the selection of *E. coli* clones surface expressing epitope peptides. DNA sequence determination of the insert sequences of up to 10,000 clones/pathogens and their alignment to genome sequences delineates the corresponding genome-encoded genes. To preselect antigens of the ANTIGENomes for recombinant protein expression and in vivo testing, several in vitro assays are applied to determine (1) in vitro expression and surface location of selected antigens with epitope-specific immune sera, (2) immune reactivity of human sera of different donor groups with synthetic peptide epitopes, and (3) antigen conservation using genomic DNA from a broad range of clinical isolates. As a result of this comprehensive approach, the number of antigen candidates is reduced from >100 to 15–30 (Meinke et al. 2005).

The ANTIGENome technology has been applied to seventeen of the most important human pathogens: staphylococci (*S. aureus*, *S. epidermidis*), streptococci (*S. pneumoniae*, *S. pyogenes*, *S. agalactiae*), Helicobacter pylori, *Escherichia coli* (Enterotoxigenic *E. coli*, Enteroaggregative *E. coli*), *Shigella flexneri*, Chlamydia pneumoniae, Enterococcus faecalis, Klebsiella pneumoniae, nontypable *Haemophilus influenzae*, Moraxella catarrhalis, Borellia afzelii, Campylobacter jejuni, and Neisseria meningitides. The most comprehensive analyses of the ANTIGENome-derived antigens were done with the three important streptococcal species, *S. pneumonia* (Pneumococcus), *S. pyogenes* (GAS), and *S. agalactiae* (GBS).

### 2.3.1 Pneumococcus (*S. pneumoniae*) Antigens

The pneumococcal ANTIGENome consisted of 95 ORFs of the TIGR4 genome, and a dozen of these were already shown to be protective antigens at the time of our studies. One-third of the 30 preselected candidates used in recombinant forms for immunization of mice showed significant levels of protection in lethal sepsis models (Giefing et al. 2008 and unpublished data). Two lead vaccine candidates, PcsB (protein required for cell separation) and StkP (serine/threonine kinase protein), were selected based on their exceptional conservation among clinical isolates (>99.5% identity), cross-protectivity against different serotypes in lethal sepsis and pneumonia models and their important nonredundant functions in bacterial multiplication (Giefing et al. 2008). PcsB (N-terminal domain) was the most frequently selected pneumococcal antigen in our screens, and its homologues were also found in other ANTIGENomes (GBS, GAS, Enterococcus). The protein plays an
important role in peptidoglycan metabolism and cell separation. PcsB gene deletion mutant cells grow in clusters with aberrant divisional septum formation, overexpress other cell wall enzymes, such as the LysM domain-containing proteins, and alter the pattern of released/secreted proteins (Giefing-Kröll et al. 2011). The eukaryotic type serine/threonine kinase protein was less immunogenic, and epitopes were located only in its C-terminal PASTA (penicillin-binding protein and serine/threonine kinase associated) domains that were shown to be extracellular with the capacity to induce protective antibodies (Giefing et al. 2008, 2010). PcsB and StkP also induce T cells in humans, characterized mainly with a T_{H17} profile (Schmid et al. 2011). This seems to be especially relevant in the light of recent data suggesting an important role for T_{H17}-producing CD4^+ T cells in clearance of pneumococcal colonization at the mucosal surface in mice (Lu et al. 2008; Zhang et al. 2009). These two pneumococcal proteins are part of a protein-based combination vaccine, IC47 (together with PsaA that is not derived from the ANTIGENome) that completed the first-in-man clinical trial with good safety and immunogenicity profile.

A striking example for the validity of this approach is the pneumococcal PspA protein. Despite of its great variability (three protein families, six different clades), PspA was among the most frequently selected, therefore highly immunogenic, proteins of the pneumococcal ANTIGENome (Giefing et al. 2008). The vast majority of the selected PspA epitopes were mapped to the conserved proline-rich region at the C-terminus of the protein. Antibodies generated with the E. coli clones carrying this epitope recognized the native PspA on live pneumococcal cells based on flow cytometry-based surface staining and induced opsonophagocytic killing in vitro (Meinke et al. 2005). Recently, it was reported by the Briles laboratory that monoclonal antibodies specific for this proline-rich region were able to bind to the pneumococcal surface and were protective in animal models (Daniels et al. 2010). These data are especially important for a potential future PspA-based vaccine since the N-terminal variable coil-coiled region that was in the focus of intensive efforts toward a novel protein-based vaccine (also tested in clinical trials) raised concerns about human tissue cross-reactivity.

### 2.3.2 Group A Streptococcus (S. pyogenes) Antigens

The group A Streptococcus screens identified 95 antigen candidates annotated in the S. pyogenes SF370 genome including several of the previously published protective proteins, such as M1 protein, C5A peptidase, streptolysin O, exotoxin B, and SpeC (Fritzer et al. 2010). One-third of the 31 antigens preselected for animal testing do not possess predictable signatures for surface expression or secretion. Among these, four were proved to be protective in at least one of three animal models we applied, and three of these are hypothetical proteins without any annotated function. Only two of the seven LPXTG cell wall proteins selected for in vivo testing were protective—the ScpC (annotated as putative protease Spy0416) and the putative secreted 5’ nuclease Spy0872—however, with the most consistent
efficacy across the different models that entailed subcutaneous, intramuscular, or intranasal immunizations; alum or mucosal adjuvants (MALP or IC31); and intravenous or intranasal bacterial challenge at lethal doses with three different serotypes (M1, M20, and M106). Lately, ScpC has emerged as an important virulence factor with multiple functions: immune evasion due to its serine protease activity that inactivates the major chemoattractant IL-8, thereby impairing host neutrophil recruitment (Hidalgo-Grass et al. 2006; Fritzer et al. 2009) and invasion of human endothelial cells (Kaur et al. 2010). None of the nine antigens induced the same high level of protection as the homologous M protein, but the efficacy was comparable in case of heterologous challenge. Since the M protein exists in >100 different serotypes, compiling a vaccine even with the most relevant M types is challenging. The concern about the involvement of M protein-specific antibodies in poststreptococcal sequelae and the risk of inducing serotype shift and vaccine escape strongly support the efforts toward a non-M protein-based vaccine for the prevention of pharyngitis and postinfectious complications. Such a vaccine is likely to be multicomponent containing several immunogenic and conserved proteins. One of the major challenges of vaccine development against GAS is the lack of mucosal infection models that sufficiently mimic the human condition.

2.3.3 **Group B Streptococcus (S. agalactiae) Antigens as Targets for Protective Antibodies**

The group B Streptococcus program was tailored to the most relevant clinical problem associated with this pathogen and relied on antibodies from sera and cervical fluid samples of pregnant women with defined cervical and anorectal colonization status. It is the notion supported by immune surveillance studies that serum antibody levels of mothers are inversely related to susceptibility of neonates to invasive GBS disease. The surface display screens with IgG and IgA antibodies identified 168 ORFs in the *S. agalactiae* NEM316 genome, among those all the previously published protective antigens (for which the genes were present in this strain), such as the Sip, C5a peptidase, pilus proteins, the LrrG protein, PGK, and a homologue of the serine-rich repeat protein SAN1485 identified from COH1 (gbs1529) (Meinke et al. 2010). In vitro analyses preselected 10% of the ANTIGENome for further testing in murine lethal sepsis models with different GBS strains. These conserved and surface-expressed candidates were evaluated most extensively by passive immunization and protection studies using rabbit sera generated with the corresponding recombinant proteins. We identified four protective GBS antigens that were not shown to induce protection at the time of our studies. Two of these antigens, FbsA and BibA, are well-characterized LPXTG cell wall proteins and virulence factors and are among the most frequently selected GBS ORFs, together with other cell wall and secreted proteins. The two hypothetical protein candidates also contribute to virulence based on our studies with gene deletion mutants. Single-specificity rabbit sera or mAbs induced high level of, but strain-dependent, protection, while their combinations resulted in superior and
broad efficacy against all GBS strains tested. Monoclonal antibodies specific for FbsA and BibA were fully protective even as Fab fragments, suggesting that blocking the function of these proteins was the major mode of action (Senn et al. 2011). These features are supportive for developing immune prophylaxis of invasive GBS disease of prematurely born neonates (e.g., with human monoclonal antibodies) who receive low levels of antibodies by maternofetal transport and are characterized by not fully developed phagocytic and complement activity.

2.3.4 Conclusions from 12 Years of Experience with the ANTIGENome Platform

2.3.4.1 Type of Antigens Identified
The validity of the selection procedure is supported by the rediscovery of the majority of previously known well-characterized protective antigens within the ANTIGENomes of the different pathogens, such as the fibronectin binding protein; clumping factors A and B from *Staphylococcus aureus* (Etz et al. 2002); PspA, PspC, neuraminidases, and histidine triad proteins from *Streptococcus pneumoniae* (Giefing et al. 2008); Sip and C5a peptidase from group B Streptococcus (*Streptococcus agalactiae*) (Meinke et al. 2010); and also the M protein, streptolysin O, and C5a peptidase from group A Streptococcus (*Streptococcus pyogenes*) (Fritzer et al. 2010), mentioning only the best-known protective antigens.

Based on the cumulative data obtained for the ANTIGENomes of these species, nearly 50% of all antigens fell into four cellular role categories: cell wall, cellular processes, transport and binding proteins, and determinants of protein fate (mainly proteases) (Etz et al. 2002; Meinke et al. 2005; Giefing et al. 2008; Fritzer et al. 2010; Meinke et al. 2010). Consistent with the predominantly extracellular nature of the targeted pathogens, the large fraction of the identified antigens comprise cell surface, membrane-associated, or secreted proteins. This bias is even more pronounced when the analysis is applied to the most frequently selected epitopes and antigens. As the hit frequency of a particular screen reflects the antibody concentration of the applied serum, these results indicate that the majority of antibacterial antibodies in human sera are generated against extracellular bacterial components.

Importantly, we also detect proteins in the ANTIGENome that are not equipped with predictable signatures for extracellular localization but were reported by proteomic studies to be surface located or surface-secreted. The best examples are glycolytic enzymes, such as glyceraldehyde-3-phosphate dehydrogenase (GAPDH), enolase, and phosphoglycerate kinase (PGK), that were selected in multiple ANTIGENome screens from different pathogens. PGK from group B Streptococcus was shown to be surface located and protective in GBS animal models (Hughes et al. 2002). Enolase was associated with binding to host extracellular molecules, such as laminin and plasminogen from *S. aureus* (Carneiro et al. 2004) and from group A Streptococcus (Cork et al. 2009), respectively, and shown to induce protective antibodies in *Streptococcus suis* animal model (Feng et al. 2009).
2.3.4.2 Detection of In Vivo Expressed Antigens

It is well established that bacteria express different repertoires of proteins under in vitro culture conditions, due to missing in vivo signals, such as those coming from interaction with host cells and extracellular molecules, different inorganic milieu (e.g., low free iron), and restriction in nutrients. As the genomic DNA-based selection is not biased by the abundance or lack of protein expression of bacteria grown under in vitro conditions, we also detect immunogenic epitopes that are derived from proteins not expressed by in vitro grown bacteria. The danger of relying on in vitro assays to preselect antigens for animal testing is the lack of expression of under standard culture conditions. The challenge with the characterization of such antigens is the identification of the in vivo signal that drives their expression.

The most striking relevant examples for ANTIGENome-derived antigens are the staphylococcal haptoglobin receptor (HarA) (also known as IsdH) and IsdB proteins (originally described as LPXTGp5 and 6, respectively in Etz et al. 2002) that are regulated by iron availability and not detectable under traditional culture conditions (Dryla et al. 2003, 2007).

2.3.4.3 Selection of Conserved Antigens

It is the common strategy of microbes to evade the human immune system by changing the molecular targets of protective immune responses. To counteract this powerful escape mechanism, we use serum pools to dilute out the antibodies that are directed against variable epitopes and enrich for those that are recognized by the majority of donors. It is especially important for the development of subunit vaccines to select conserved antigens to minimize the risk of escape mutations induced by vaccination. Although we typically use only one well-characterized pathogen strain for genomic screening, detailed gene prevalence and conservation analysis are integrated into the procedure to select vaccine candidates. Antigens, for which genes are missing in more than 10% of clinical isolates or show great amino acid sequence variation (without the presence of a conserved domain), are not considered in our vaccine development programs.

2.3.4.4 The Challenge with Predictive Animal Models

The ultimate preclinical proof that a novel discovery platform is a valuable addition to the vaccine development efforts is its ability to identify novel vaccine candidates based on animal protection. The first ANTIGENome program was performed with the \textit{S. aureus} COL strain, and candidate antigens were evaluated for in vivo efficacy in partnership, by the vaccine research team at Merck & Co. As a result of their in vivo screening efforts, novel protective antigens were identified (Kuklin et al. 2006; Ebert et al. 2011). From these, IsdB was selected to be the sole antigen of a candidate vaccine (V710). Protection by IsdB immunization was also detected by independent research groups (Stranger-Jones et al. 2006; Kim et al. 2010). V710 successfully completed safety and immunogenicity studies performed in several populations (healthy adults, elderly, hemodialysis patients) with very promising results. Murine and human monoclonal antibodies generated with IsdB were shown
to have in vitro opsonophagocytic activity and efficacy in *S. aureus* infectious models (Brown et al. 2009; Ebert et al. 2010). Yet this single-component vaccine failed to reach predefined efficacy endpoint in a phase IIb study to prevent *S. aureus* infections in patients undergoing cardiac surgery.

Our strategy to address the weaknesses of animal models is to perform extensive testing in different models (sepsis, pneumonia, bacteremia) and to use several different strains preferentially clinical isolates. It has been a consistent finding that surface accessibility and in vivo efficacy of antigens are greatly influenced by strain variability (e.g., capsule) in spite of the presence of the corresponding genes. Therefore, it is our conclusion that protein-based vaccines need to be multicomponent to achieve broad protection to increase the likelihood of success of clinical trials.

### Conclusions

Both genome mining approaches allow the selection of vaccine antigens from the entire repertoire of proteins encoded by the pathogen genome. But how do the two technologies compare? Do they lead to the identification of overlapping antigens or distinct ones?

In order to rigorously address these questions, the two approaches should be carried out starting from the same strains, from which to make the expression libraries, using the same animal models and challenging the animals with the same challenge strains. This experimental comparison has not been done, and therefore, the comparison can only be made on a theoretical ground.

There is no doubt that the bacterial protein antigens that constitute vaccines currently on the market or in different development phases are surface or secreted proteins, which are both well expressed and immunogenic during natural infection. Therefore, it is highly likely that both the reverse vaccinology and the ANTIGENome approaches would have selected these protective antigens. By analogy, if the “level of expression” and “immunogenicity” rules apply also for pathogens, for which vaccines are still missing, the two technologies must perform similarly. To support this conclusion are the published and unpublished data on group A and group B streptococci for which the two approaches have been carried out independently and did select antigens in common. An even greater overlap is expected should the same experimental conditions were used. Differences in the protective antigen pools are expected in two instances. One of these is if a given protective antigen is poorly immunogenic during infection in the host in spite of the fact that it is sufficiently expressed either on the surface or in the extracellular milieu. Such an antigen would be missed by the ANTIGENome approach. However, at present, we are not aware of any antigen that induces protective antibody responses upon vaccination but is not immunogenic during infection. Another hypothetical situation is that an antigen can be well expressed only during infection, but not when the pathogen is grown under laboratory conditions. In this case, if the preselection of the proteins to enter the biological screening is carried out by proteomic analysis as recommended in the session “Recipe for Reverse
Vaccinology,” the protective antigen would be identified only using the ANTIGENome approach, provided that such antigen is upregulated also in the animal model used for the screening. This is because the proteomic analysis of the surface and secreted proteins is carried out after growing the pathogen under laboratory conditions. However, according to our experience, most of the in vivo upregulated antigens are sufficiently expressed also under laboratory conditions (sometimes in special media), and therefore, they should also be identified using the proposed “optimized” version of reverse vaccinology.

One of the major differences in the two approaches is that the comparative genome analysis also considers variable genes, while the ANTIGENome technology preferentially filters for highly conserved antigens. Importantly, the end result is similar in both cases: combination of protein antigens is necessary for viable, broadly protective vaccines. The best example for this conclusion is the antigen identification efforts for GBS. Multigenome prediction selected pilus proteins from the variable subgenomes, which resulted in a three-component vaccine representing the three genomic islands present in various GBS strains (Maione et al. 2005; Margarit et al. 2009). Pilus proteins were counterselected by the ANTIGENome screen during the primary prioritization due to their variability. Nonetheless, one single antigen was insufficient to induce broad protection, and a combination of ANTIGENome-derived antigens was necessary (Meinke et al. 2010; Senn et al. 2011). Similar conclusions were reached with other pathogens, and all the vaccine programs conducted by the two research groups result in multicomponent vaccines.

In conclusion, the several years of application of reverse vaccinology and ANTIGENome have demonstrated the robustness of both technologies and their power in identifying vaccine candidates. For the future, efforts should be dedicated to find ways to further restrict the number of antigens to be screened in the animal models and increase the predictive power of primary selection.

References


Development of Novel Vaccines
Skills, Knowledge and Translational Technologies
von Gabain, A.; Klade, C. (Eds.)
2012, XIV, 306 p. With online files/update., Hardcover
ISBN: 978-3-7091-0708-9