Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

Biologia Cellulare, Molecolare e Industriale: Biologia Funzionale dei Sistemi Cellulari e Molecolari

Ciclo XXIII

Settore/i scientifico-disciplinare/i di afferenza: BIO-11

MetaVaccinology: A new Vaccine Discovery Tool

Presentata da: Emrah Altındiş

Coordinatore Dottorato

Prof. Scarlato Vincenzo

Relatore

Dott. sa Sabrina Liberatori

Prof. Scarlato Vincenzo

Esame finale anno 2011

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Emrah Altındiş

Signature:

ABSTRACT

METAVACCINOLOGY: A NEW VACCINE DISCOVERY TOOL

Emrah Altındiş

Dr. Sabrina Liberatori
Prof. Dr. Vincenzo Scarlato
March 2011, 162 pages

In the last decade, the reverse vaccinology approach shifted the paradigm of vaccine discovery from conventional culture-based methods to high-throughput genome-based approaches for the development of recombinant protein-based vaccines against pathogenic bacteria. Besides reaching its main goal of identifying new vaccine candidates, this new procedure produced also a huge amount of molecular knowledge related to them. In the present work, we explored this knowledge in a species-independent way and we performed a systematic *in silico* molecular analysis of more than 100 protective antigens, looking at their sequence similarity, domain composition and protein architecture in order to identify possible common molecular features. This meta-analysis revealed that, beside a low sequence similarity, most of the known bacterial protective antigens shared structural/functional Pfam domains as well as specific protein architectures. Based on this, we formulated the hypothesis that the occurrence of these molecular signatures can be predictive of possible protective properties of other proteins in different bacterial species. We tested this hypothesis in *Streptococcus agalactiae* and identified four new protective antigens. Moreover, in order to provide a second proof of the concept for our approach, we used *Staphyloccus aureus* as a

second pathogen and identified five new protective antigens. This new knowledge-driven selection process, named MetaVaccinology, represents the first *in silico* vaccine discovery tool based on conserved and predictive molecular and structural features of bacterial protective antigens and not dependent upon the prediction of their sub-cellular localization.

Key words: MetaVaccinology, *Streptoccus agalactiae*, *Staphylococcus aureus*, 5'nucleotidases, vaccinology, genomics, proteomics

To my dear parents,

Zeynep Çemrek Altındiş

Cengiz Altındiş

To my dear grandfather:

Ahmet Çemrek

And to our ahparig:

Hrant Dink

ACKNOWLEDGEMENTS

My PhD has been a new and long period of my life that was more than 3.5 years spent in Italy. First of all, I would like to express my sincerest appreciation to my supervisor Dr. Sabrina Liberatori for her guidance, supervision and endless patience given during this long process. It was a pleasure to work with a brillant scientist like Sabrina who is always curious and has a deep knowledge on different aspects of microbiology and molecular biology.

Secondly, I would also like to thank Prof. Dr. Vincenzo Scarlato for his kindness, guidance and encouragement. I also thank to my labmates Christina Schluepen, Beate Bruske, Benedetta Lombardi for their friendship, smiles and their help whenever I needed. I am also grateful to my friend Roberta Cozzi for her invaluable help during the cloning and expression experiments of GBS proteins. My special thanks also go to Domenico Maione, Fabio Bagnoli for their help in planning the experiments for GBS and *Staphylococcus aureus*. Thanks to Sara Marchi for her role in purification step of the proteins and to Fabiana Falugi for cloning of *Staphylococcus* proteins and I acknowledge to Marco Soriani, Benedetta di Palo and Miriam Christina Richter for their help with Xcellingence system.

I would like to express my indebtedness to Dr. Rino Rappuoli, Dr. Guido Grandi, Dr. Nathalie Norais and Dr. Nissum Mikkel for their support and contribution in different levels of my project. Very special thanks to the members of our function, Massimiliano Biagini, Chiara Tani, Danilo Donnarumma, Agnese Faleri, Francesco Doro, Francesco Berlanda, Maria Stella, Laura Ciucchi and Paolo Mariotti for their friendship and support. Thanks to Elisabetta

Soldaini and Francesca Mancini for their help and thanks to Novartis Vaccines animal facility and “kitchen” workers for their invaluable input to this work.

Living abroad is not so easy for a person who always misses and thinks his country, family, friends, darling, culture and language. So thanks to all my friends who made life much easier for me in Italy, especially thanks to Giulia Grosso, Daniele Cavicchioli, Antonella Iocca and Erdem Sönmez, My deepest thanks to Gül Bayram that has brought the smell of the red roses to my life in Siena and the life is much more beautiful with her presence.

My special gratitude is extended to my precious family, to my mother Zeynep Çemrek ALTINDİŞ, my father Cengiz ALTINDİŞ, to my brother Burak ALTINDİŞ and at last but not least to my grandfather Ahmet Çemrek for their endless love, kind understanding, patience and endless support during my studies.

The famous phiolosopher Heraclitus (from Ephesus-one hour to my city Izmir) said (*πάντα χωρεῖ καὶ οὐδὲν μένει*” καὶ "*δὶς ἐς τὸν αὐτὸν ποταμὸν οὐκ ἂν ἐμβαίης")* "Everything changes and nothing remains still .... and ... you cannot step twice into the same stream. It is very clear that the battle in between the microbes and human kind will never end since the evolution never stops and the pathogens are more “intelligent” than we think! Thanks to all scientists that fight in their laboratories in order to prevent infectious diseases and to save the lifes of the people.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1. 1 Brief History of Vaccinology

In the beliefs of ancient peoples, diseases were inflicted on mankind by intangible and capricious deities as punishment for ill-defined transgressions. Fear of destruction by disease became an effective tool used by rulers and politicians to instill terrors, which would prove useful in controlling human behavior in the long climb from early tribal to "civilized" existence. However, some, who were the forerunner of modern science, did discover microbial life forms, the relationships of environment to disease, and the fact that there was no second occurrence following certain clinically definable illnesses. Such heretical concepts revealed that man himself, rather than devils and demons, were the source of pestilence (1).

It was common knowledge that survivors of smallpox became immune to the disease. The most successful way of combating smallpox before the discovery of vaccination was inoculation. The word is derived from the Latin *inoculare,* meaning "to graft." Inoculation referred to the subcutaneous instillation of smallpox virus into nonimmune individuals. In 1670, Circassian traders introduced variolation to the Turkish "Ottoman" Empire. Variolation came to Europe at the beginning of the 18th

century with the arrival of travelers from Istanbul. It was the continued advocacy of the English aristocrat Lady Mary Wortley Montague that was responsible for the introduction of

variolation in England. In 1717, Lady Montague's husband, Edward Wortley Montague, was appointed ambassador to the Sublime Porte. A few weeks after their arrival in Istanbul, Lady Montague wrote to her friend about the method of variolation used at the Ottoman court. Lady Montague was so determined to prevent the ravages of smallpox that she ordered the embassy surgeon, Charles Maitland, to inoculate her 5-year-old son. The inoculation procedure was performed in March 1718. Upon their return to London in April 1721, Lady Montague had Charles Maitland inoculate her 4-year-old daughter in the presence of physicians of the royal court. After these first professional variolation procedures, word of the practice spread to several members of the royal family (2).

Figure 1: The first smallpox vaccination. Smallpox is stemmed (3).

The modern science of vaccinology took off on 14 May 1796 when Edward Jenner inoculated James Phipps, with the vaccinia virus obtained from a young woman who had been accidentally infected by a cow. Jenner describes this key experiment with the following words:

"The more accurately to observe the progress of the infection, I selected a healthy boy about eight years old for the purpose of inoculation with the cowpox. The matter was taken from the suppurated sore on the hand of a dairy Maid who was infected by her master's Cows, and it was inserted on the 14th May 1796 into the arms of the Boy by means of two superficial incisions each about three quarters of an inch long. .During the whole of [the ninth day after this] he was perceptibly indisposed and had a restless night; but, on the following day he was perfectly well. On the 1st of July following this, the Boy was inoculated with Matter immediately taken from a smallpox Pustule. Several punctures and slight incisions were made in both his arms, and the matter was well rubbed into them, but no disease followed (4).

The history of modern vaccination as a deliberate endeavor began in the laboratory of Louis Pasteur. His aphorism that ***'chance favors the prepared mind'*** was never more aptly illustrated than by his own discovery of attenuation. Pasteur was on vacation in the summer of 1881, and returned in the autumn to studies of chicken cholera, caused by what we call today *Pasteurella multocida*. A culture left on the bench during the summer was inoculated into chickens but did not cause disease. Pasteur then made a fresh culture and inoculated the same chickens, whether through parsimony or purpose we do not know. In any case, the chickens were resistant to the fresh challenge, and Pasteur realized that the aged culture had rendered them immune. From these observations Pasteur constructed the hypothesis that pathogens

could be *attenuated* by exposure to environmental insults such as high temperature, oxygen and chemicals. His ensuing work on anthrax and rabies confirmed the hypothesis. In the next century, Calmette and Guérin used passage in artificial media to attenuate *Mycobacterium bovis*, and Theiler used passage in mice and chick embryos to attenuate yellow fever virus (5).

Figure 2: Two important figures of modern vaccionology: Robert Koch and Louis Pasteur.

That was apparent during the first era of modern vaccinology, from the late nineteenth century through the 1930s: a long era of grand expectations. From the point of the 1890s, there was every reason to believe that a remarkable series of vaccine innovations would follow the scientific breakthroughs of Robert Koch, Louis Pasteur, and others. For the first time in human history, the sources of disease could be analyzed in a systematic, scientific manner. Thanks to Pasteur, society now had a vaccine effective against rabies, and other scientists followed with killed-organism vaccines for protection against cholera, typhoid, and plague.

Armed with a better understanding of immunology and with a new serum antitoxin that was effective against diphtheria, leading scientists and physicians at the turn of the century can be excused their sense of vigorous optimism. After all, they were not along in their hubris many prominent political leaders, pundits, and tycoons joined them in expressing wild expectations about progress in the years ahead. There were as well experimental vaccines, not all of which were effective and safe, against diphtheria, pertussis, tuberculosis (BCG), tetanus, yellow fever, and typhus (Rickettsia) (6).

After the applications of killed and attenuated vaccines, the vaccine revolution continued to progress by the invention of cell culture, recombinant vaccines and polysaccharide technology. Today, iimmunization is a proven tool for controlling and even eradicating diseases. Table 1 shows the list of the approved vaccines yet. An immunization campaign carried out by the World Health Organization (WHO) from 1967 to 1977 eradicated the natural occurrence of smallpox. When the programme began, the disease still threatened 60% of the world's population and killed every fourth victim. Eradication of poliomyelitis is within reach. Since the launch by WHO and its partners of the Global Polio Eradication Initiative in 1988, infections have fallen by 99%, and some five million people have escaped paralysis. Between 1999 and 2003, measles deaths dropped worldwide by almost 40%, and some regions have set a target of eliminating the disease. Maternal and neonatal tetanus will soon be eliminated in 14 of 57 high-risk countries (7). In the next chapter, we will go on with the influence of genomics revolution to vaccine development.

Table 1. Dates of introduction of commonly used vaccines (8)

	Vaccine	Date		Vaccine	Date
1	Smallpox	1796	**22**	Pneumococcus	1976
2	Rabies	1885	**23**	Acellular P (Pa)	1981
3	Cholera	1896	**24**	Hepatitis B (HB)	1981
4	Typhoid	1896	**25**	Varicella (V)	1984
5	Plague	1896	**26**	rDNA HB	1986
6	Diphtheria (D)	1923	**27**	H. influenzae b (Hib)	1988
7	Pertussis (Pw)	1926	**28**	Hepatitis A (HA)	1991
8	Tetanus (T)	1927	**29**	DTPwIPVHib	1993
9	Tuberculosis (BCG)	1927	**30**	DTPa	1994
10	Yellow fever	1935	**31**	DTPwHB	1996
11	Influenza	1936	**32**	HBHA	1996
12	Polio (IPV)	1955	**33**	DTPaHib	1997
13	DTPw	1957	**34**	DTPaIPVHib	1997
14	Polio (OPV)	1958	**35**	Lyme	1998
15	DTIPV	1961	**36**	Rotavirus	1998
16	Measles (M)	1963	**37**	Dtpa	1999
17	DTPIPV	1966	**38**	HATy	1999
18	Mumps (M)	1967	**39**	DTPaHBIPV	2000
19	Rubella (R)	1969	**40**	DTPaHBIPVHib	2000
20	MMR	1971	**41**	MCCVb (MenC conjugate vaccine)	2000

21	Meningococcus	1972	**42**	PCVa(conjugate vaccine)	2000

1.2 A new generation: Subunit Vaccines

After attenuated and killed type of vaccines, a new generation introduced as vaccines against diphtheria and tetanus in the 1920s is a much more sophisticated product, a purified bacterial component. In both these cases, it is a protein toxin previously demonstrated to be an essential cause of the disease. For the vaccine, the toxin is chemically modified to yield the non-toxic toxoid. Both these vaccines have performed extremely well in terms of both safety and efficacy, demonstrating that the theory behind them was correct. Purified single component vaccines have many attractions: the immune stimulus is maximally directed to the molecule relevant for protection and additional components that could cause adverse reactions or other, unwanted but unknown problems are avoided (9).

This significant knowledge, that a protein and/or a subunit of a pathogen is enough to stimulate a specific immune response then used in order to make vaccines against pathogens that cannot be grown or can be grown only with difficulty *in vitro* pose a special problem to vaccine development. Molecular biology and genetic engineering have had a dramatic effect on the field of vaccinology, although many of the important advances have not yet made it into the market. The first success story in this area was the development of the hepatitis B vaccine, which was licensed in 1986. The surface protein of hepatitis B virus is expressed from a DNA plasmid in yeast cells, purified and adsorbed on alum for injection (10).

As with diphtheria, the new generation of vaccines against pertussis (whooping cough) was made from a toxin that had been deactivated with formaldehyde. Rino Rappuoli and his colleagues achieved to make this subunit vaccine against *Bordetella pertussis*, the etiologic agent of whooping cough. Dr. Rappuoli describes this process as: "I cloned and sequenced the gene for pertussis toxin and did what Pappenheimer had done 15 years before with diphtheria" (11). But this time they used site-directed mutagenesis to specifically alter amino acids in the active site of the toxin. The result was a nontoxic molecule that made a potent vaccine. The pertussis vaccine also established a new generation of so-called acellular vaccines, which, unlike older vaccines, did not contain cells or cell fragments (12).

Both of Hepatitis B and acellular pertussis vaccines are widely used in all over the world. Hepatitis B vaccine is 95% effective in preventing HBV infection and its chronic consequences, and is the first vaccine against a major human cancer. The vaccine has an outstanding record of safety and effectiveness. Over one billion doses of hepatitis B vaccine have been used worldwide. In many countries where 8% to 15% of children used to become chronically infected with HBV, vaccination has reduced the rate of chronic infection to less than 1% among immunized children (13).

1.3 Genomics and Vaccinology

Louis Pasteur, who developed the first vaccine against rabies, established in 1881 the basic paradigm for vaccine development, which included the isolation, inactivation and injection of the causative microorganism. These basic principles have guided vaccine development during the twentieth century. All existing vaccines are based on killed or live-attenuated microorganisms or subunits purified from the microorganism such as toxins detoxified by chemical treatment, purified antigens or polysaccharide conjugated to proteins (Table 1). Vaccines produced following Pasteur's principles allowed the control and, in some cases, the eradication of many important infectious diseases. Despite several successes, the Pasteur's approach to vaccine development took a long time to generate vaccines against those pathogens for which the solution was feasible, but failed to produce vaccines for those bacteria and parasites that do not have obvious immunodominant protective antigens or for as yet uncultivable microorganisms (14). 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* (15). As of December 2010, 1,283 bacterial genomes are completely sequence and more than 5433 are ongoing (http://www.genomesonline.org/cgi-bin/GOLD/bin/gold.cgi).

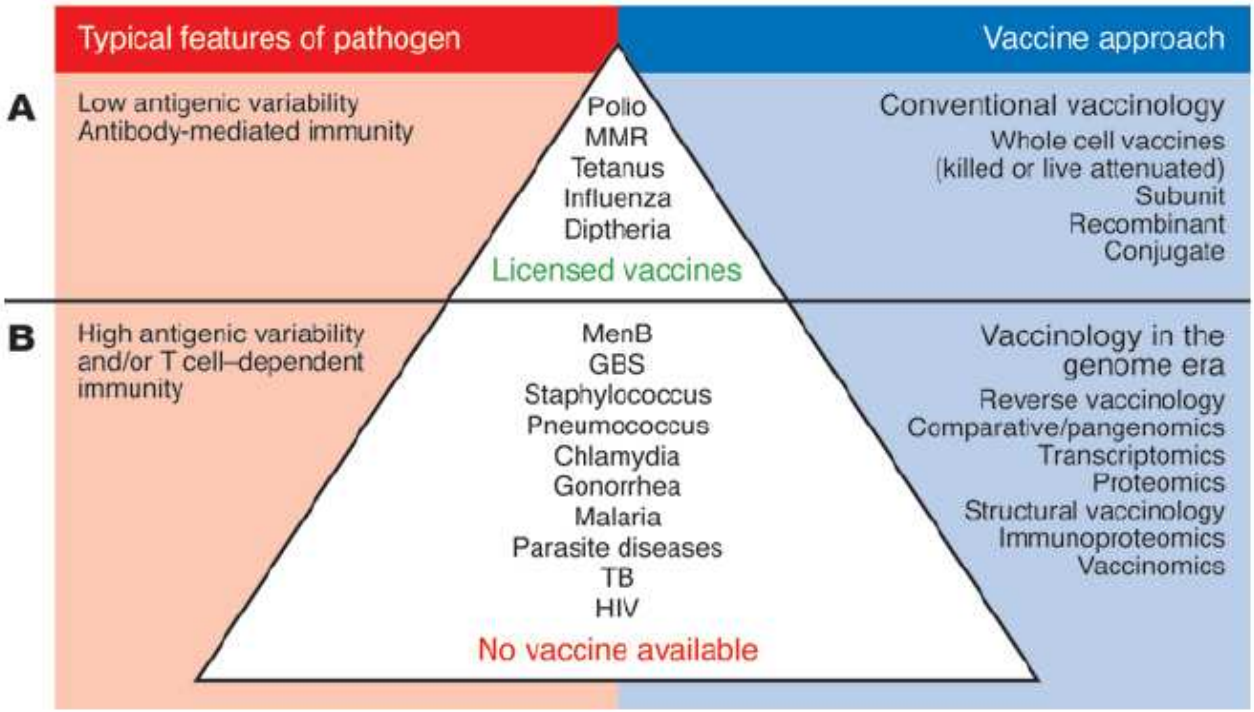


Figure 3. Schematic overview of conventional vaccinology versus vaccinology in the genome era. (A) Most licensed vaccines target pathogens that have low antigenic variability and pathogens for which protection depends on antibody-mediated immunity. These vaccines have typically been developed using conventional vaccinology. (B) Several pathogens are shown for which no vaccine is available, due to either their high antigenic variability and/or the need to induce T cell–dependent immunity to elicit protection. New approaches are being applied to vaccine development for these pathogens in the genome era. Vaccines/diseases shown in the figure are selected examples of each category and are not a complete list (16).

The application of genome analysis to vaccine development, a concept termed "reverse vaccinology," initiated a positive feedback loop in terms of the development and application of novel approaches to the field of vaccinology. As a result, it is becoming possible to systematically examine almost every aspect of a pathogen and its interactions with the host immune system in the search for vaccine candidates. Reverse vaccinology applied to the genome of a pathogen aims to identify the complete repertoire of antigens that an organism is capable of expressing on its surface. Transcriptomics and proteomics enable the investigation of the array of antigens actually expressed by a pathogen under specified conditions, by examining the mRNA and protein of the organism, respectively. Analysis can also focus on the subset of proteins that are surface exposed (surface proteome) or the subset of genes that are functionally important for infection (functional genomics). Newer fields of study are focused on elucidating the set of antigens that interact with the host immune system and the mechanisms involved in these interactions (immunomics), the structural epitopes of immunogenic antigens (structural vaccinology), and the way in which individual host immune systems respond to a vaccine (vaccinomics) (16). While each of these approaches has limitations, they have all emerged as powerful tools in vaccine development.

The approach referred to as 'reverse vaccinology' uses the genome sequences of viral, bacterial or parasitic pathogens of interest rather than the cells as starting material for the identification of novel antigens, whose activity should be subsequently confirmed by experimental biology. In general, the aim is the identification of genes potentially encoding pathogenicity factors and secreted or membrane-associated proteins. Specific algorithms suitable for the *in silico* identification of novel surface-exposed and, thus, antibody

accessible proteins mediating a protective response are used (17). The first example of a successful application of the reverse vaccinology approach was provided by Pizza and coworkers in collaboration with The Institute for Genomic Research.

Pathogen	Disease	Approach
Neisseria meningitidis serogroup B	Bacterial meningitis and septicemia	Reverse vaccinology Microarray Proteomics
Streptococcus pneumoniae	Bacterial pneumonia, sepsis, sinusitis, otitis media and bacterial meningitis	Classical or comparative reverse vaccinology Proteomics
Bacillus anthracis	Anthrax	Reverse vaccinology CGH microarray Microarray Proteomics and immunoproteomics
Staphylococcus aureus	Variety of infections, including 'pelvic syndrome', rapidly progressive pneumonia, ocular infections, septic thrombophlebitis	CGH microarray Immunoproteomics
Porphyromonas gingivalis	Periodontitis	Reverse vaccinology
Mycobacterium tuberculosis	Tuberculosis	Reverse vaccinology
Helicobacter pylori	Ulcer, atrophic gastritis, adenocarcinoma, lymphoma	Reverse vaccinology Immunoproteomics
Streptococcus agalactiae (GBS)	Bacterial sepsis, pneumonia, meningitis	Classical or comparative reverse vaccinology
Streptococcus pyogenes (GAS)	Many systemic invasive infections including necrotizing fasciitis, myositis, pneumonia, sepsis, arthritis	Genome-wide analysis Proteomics (surface proteome)
Chlamydia pneumoniae	Pneumonia, meningitis, middle era infections	Reverse vaccinology and proteomics

Figure 4: Examples of different post genomics approaches in the development of vaccines against some bacterial pathogens, and the status of the corresponding vaccine development (18)

1.3.1 The first Steps of Reverse Vaccinology

Neisseria meningitidis is a major cause of meningitis and sepsis in children and young adults. While polysaccharidebased vaccines are available for A, C, Y and W135 serogroups, conventional vaccinology has failed for serogroup B. Group B meningococcus (MenB) represents the first example to which reverse vaccinology has been applied. MenB complete genome from strain MC58 was obtained by the random shotgun strategy (19). While the sequencing was still in progress, the MenB genome was screened, using several softwares, in order to select putative ORF coding for surface-exposed or secreted proteins. Among the 2158 putative open reading frames (ORFs) annotated, 600 ORFs were selected on the basis of these criteria. The selected 600 ORFs were amplified from meningococcus by PCR, and cloned into *Escherichia coli* in order to express each gene as His-tag or GST fusion protein. Out of these 600 putative ORFs, 350 were successfully expressed, purified and used to immunize mice. Screening of immune sera was performed by Western blot on meningococcus total cell lysates and outer membrane vesicles to verify whether the protein was really expressed in meningococcus and to determine its subcellular localization. The surface-exposure of each antigen was then confirmed by fluorensce-activated cell sorter (FACS) analysis and ELISA on whole cell bacteria. Finally, sera were tested in bactericidal assay, an assay which is known to correlate with the protection in humans. Ninety-one proteins were found to be surface-exposed, 29 of them were able to induce bactericidal antibodies (20).

These 29 antigens selected by reverse vaccinology were prioritized based on their ability to induce broad protection. The three top antigens that met the prioritization criteria were GNA2132, GNA1870, and NadA. Two additional antigens, GNA1030 and GNA2091, were

selected because they also induced protective immunity but only in some of the assays. The results of the multicomponent vaccine were very promising, the vaccine covers 78% of the strains in the basic formulation using aluminum hydroxide as adjuvant, and coverage can be increased to 95% of the strains by using other adjuvants. The progress compared with vaccines available so far is gigantic; in fact, OMV vaccines, which are the best vaccines against MenB developed during the last 30 years, are shown to cover only 20% of the MenB strains (21). This work also shows that universal protein-based vaccines can be developed against those encapsulated bacteria that are usually targeted by conjugate vaccines.

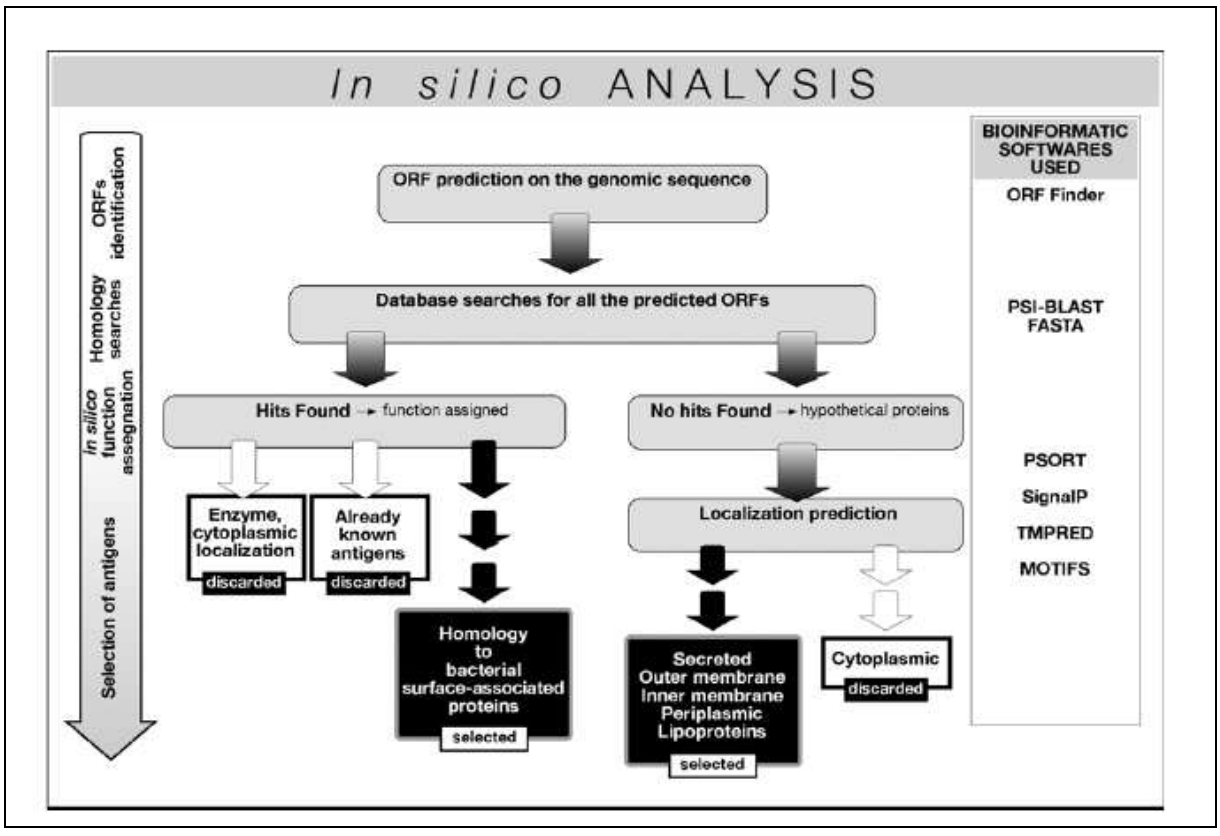


Figure 5. Flow chart showing criteria and bioinformatic softwares for the *in silico* genome analysis and selection of putative vaccine candidates (15).

1.3.2 Pan-Genome and Reverse Vaccinology

Tettelin *et al.* introduced the concept of pan-genome, which was defined as the global gene repertoire pertaining to a given species (22). The pan-genome can be defined as the global gene repertoire pertaining to a species. In general, it can be divided in three parts: the core-genome, which includes the set of genes invariably present and conserved in all the isolates; the 'dispensable genome', comprising genes present in some but not all the strains, and the strain-specific genes, which are present only in one single isolate. The analysis performed on GBS genomes indicated that the different isolates have an estimated core-genome containing 1806 genes, whereas each single genome contained between 2000 and 2400 genes. In other words, each strain contains a relatively large number of dispensable genes (200–600) that are missing in at least one of the other strain genomes. The most interesting finding is that when a new genome sequence is added to the pool of the others, an estimated number of 33 new strain specific genes, which are exclusively present in that genome, are added. Consequently, the pool of genes comprising the coregenome, dispensable genome and all the strain-specific genes, globally defined as the pan-genome, represents an open entity (open pan-genome) that is continuously increasing in size (24).

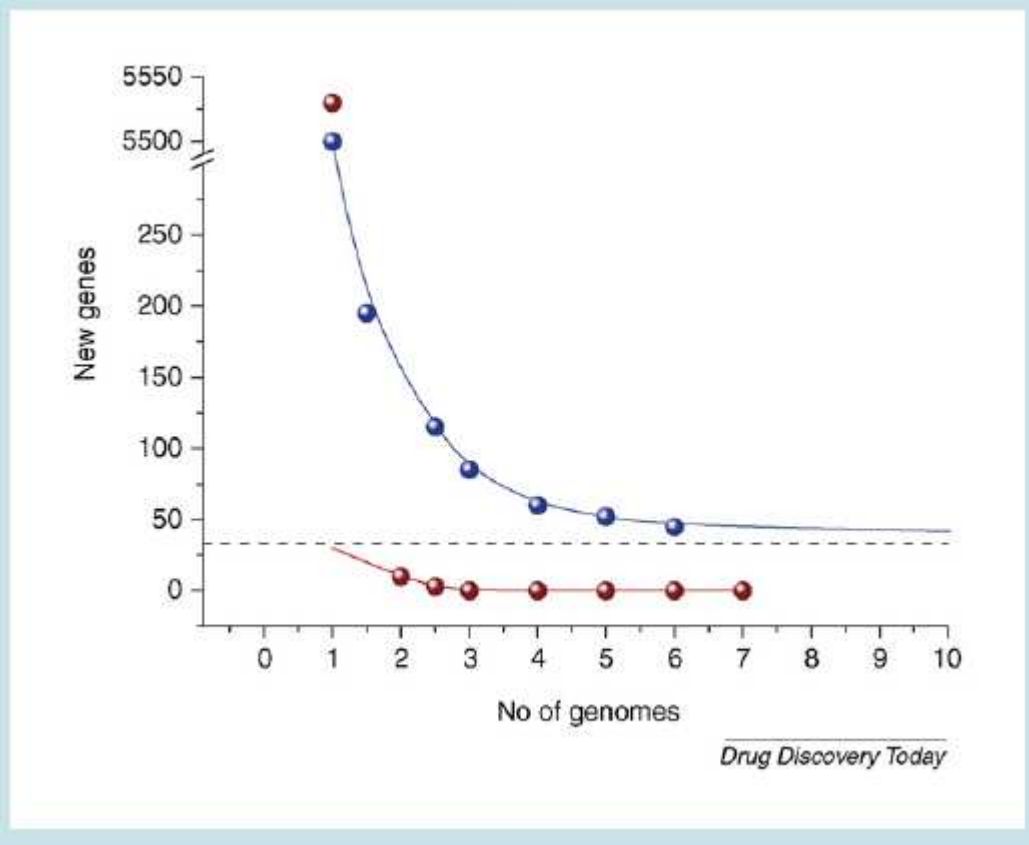


Figure 6. The pan-genome. The number of specific genes is plotted as a function of the number *n* of strains sequentially added. For GBS (blue line), the extrapolated average number of strain-specific genes, 33, is shown as a dashed line. For *Bacillus anthracis* (red line), the curve reaches zero after addition of the fourth genome. No new genes will be discovered after this threshold (24).

Besides the study of the diversity inside a species, one possible application of the pan-genome in vaccinology (pan-genomic reverse vaccinology) is the identification of novel vaccine candidates and targets for antimicrobials. Maione and colleagues performed the first application of the pan-genome to vaccines to design a universal vaccine against GBS (25). By computational algorithms 589 surface-associated proteins are predicted, 396 of which were

core genes and those remaining were genes absent in at least one strain. Selected potential antigens were expressed as recombinant proteins, purified and tested for protection against GBS, and four were found to elicit protective immunity in an animal model. Among these antigens, only one was part of the core genome; however, it was not able to confer global protection, hence the final vaccine formulation should include a combination of the four antigens (25). The GBS example has demonstrated that multiple genome sequences of each species are important to cover the diversity of many pathogens

1.3.3 Substractive Reverse Vaccinology for *Escherichia coli*

The most recent paper of Reverse Vaccinology is published on extraintestinal pathogenic *Escherichia coli* (ExPEC) which is a common cause of disease in both mammals and birds. Moriel *et al.* have determined the genome sequence of ExPEC IHE3034 (ST95) isolated from a case of neonatal meningitis and compared this to available genome sequences of other ExPEC strains and a few nonpathogenic *E. coli*. To identify potential vaccine candidates against ExPEC, they applied a "subtractive reverse vaccinology" approach. Briefly, antigens predicted to be surface associated or secreted and with no more than three transmembrane domains were selected by bioinformatic analysis of the IHE3034, 536, and CFT073 genomes. The presence (and the level of similarity) of these antigens in the nonpathogenic strains MG1655, DH10B, and W3110 were used as exclusion criteria. By this approach, they were able to identify 230 potential antigens, which were then expressed as His-tagged proteins, purified and tested for protection in a sepsis mouse model. Of these, 220 were successfully purified, 69 as soluble and 151 as insoluble proteins. Nine antigens were protective in a

mouse challenge model. Some of them were also present in other pathogenic non-ExPEC strains, suggesting that a broadly protective *E. coli* vaccine may be possible (26).

1.3.4 ANTIGENome Technology

In 2000, Etz *et al* published the first article of ANTIGENome technology. They developed a new procedure to identify commonly recognized antigens, and provide a comprehensive *in vivo* antigenic profile of *Staphylococcus aureus*. Very briefly, *S. aureus* peptides were displayed on the surface of *Escherichia coli* via fusion to one of two outer membrane proteins (LamB and FhuA) and probed with sera selected for high Ab titer and opsonic activity. A total of 60 antigenic proteins were identified, most of which are located or predicted to be located on the surface of the bacterium or secreted. The authors claim that the identification of these antigens and their reactivity with individual sera from patients and healthy individuals could facilitate the selection of promising vaccine candidates for further evaluation.(27). The authors argue that the ANTIGENome technology does not directly rely on genome annotation and, thus, has the potential to select proteins that are not predicted by ORF-finding algorithms (28).

ANTIGENome approach was secondly applied to *Streptococcus pneumoniae* by using display libraries expressing 15–150 amino acid fragments of the pathogen's proteome. Serum antibodies of exposed, but not infected, individuals and convalescing patients identified the ANTIGENome of pneumococcus consisting of 140 antigens, many of them surface exposed. Based on several *in vitro* assays, 18 novel candidates were preselected for animal studies, and 4 of them showed significant protection against lethal sepsis.. A vaccine containing two of

four protective antigens, PcsB and StkP, is planned for the prevention of infections caused by all serotypes of pneumococcus in the elderly and in children (29).

Very recently, the same group published their ANTIGENome approach for *Streptococcus pyogenes*. For immune selection, human serum antibodies obtained from patients who recovered from common *S. pyogenes* infections and healthy, noncolonized parents of small children, were used. These studies led to the discovery of eight novel antigens in addition to Spy0416/ScpC, all of which are highly conserved among GAS clinical isolates and provide significant protection in murine challenge models (30).

1.4 Proteomics and Vaccinology

The term proteome, in analogy to the term genome, was coined to describe the complete set of proteins that an organism has produced under a defined set of conditions. The genome is static since it represents the blueprint for all cellular properties that a cell is able to develop. In contrast, the proteome is highly dynamic and much more complex than the genome. It is critical for survival that the protein composition of a cell is constantly adjusted to meet the challenges of changing environmental conditions (31). Already in 1975, the powerful method of two-dimensional-polyacrylamide gel electrophoresis (2D PAGE) was introduced that allowed one to separate highly complex cellular protein extracts into individual proteins on a single gel based on two properties of the proteins the isoelectric point (pI) and the molecular weight (MW) (32). Based on a well-annotated genomic sequence, it became possible to introduce large-scale mass spectrometry (MS) techniques to identify virtually every protein detected on a 2D gel. The increase in throughput, the partial automation, and the higher

reproducibility of 2D-PAGE analysis recently made it a very attractive tool to study cellular functions on a molecular level.

Proteomics approaches to pathogens may have different targets: (i) Characterization of submicrobial proteomes (for example, secreted proteins, surface proteins and immunogenic proteins), (ii) comparative analysis of different strains and physiological states, (iii) identification of proteins related to pathogenicity, (iv) identification of proteins involved in host–pathogen interactions and (v) evaluation of mechanisms of action of antimicrobials are the most important ones (33). Proteomic approaches to vaccine candidate selection are able to go beyond several limitations of bioinformatics tools, which rely on homology searches, in predicting whether or not a protein is surface-exposed (34).

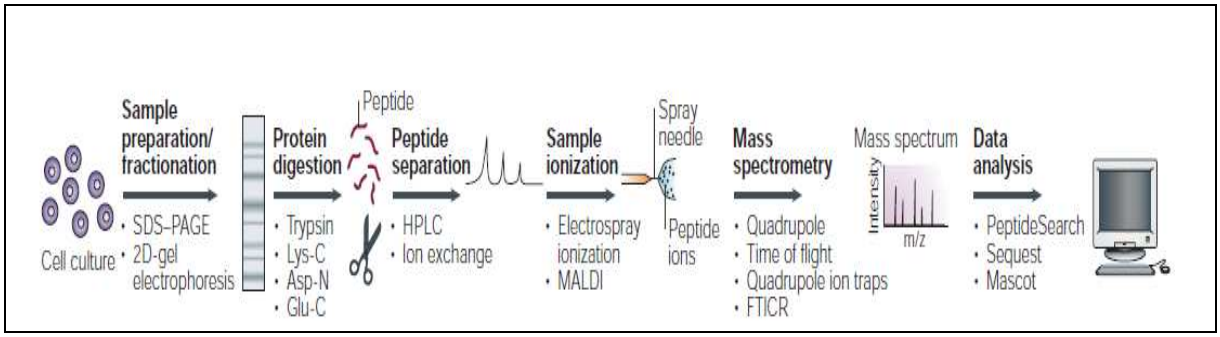


Figure 6. The mass-spectrometry/proteomic experiment. A protein population is prepared from a biological source — for example, a cell culture — and the last step in protein purification is often SDS–PAGE. The gel lane that is obtained is cut into several slices, which are then in-gel digested. The generated peptide mixture is separated on- or off-line using single or multiple dimensions of peptide separation. Peptides are then ionized by electrospray ionization (depicted) or matrix-assisted laser desorption/ionization (MALDI) and can be analysed by various different mass spectrometers. Finally, the peptide-sequencing data that are obtained from the mass spectra are searched against protein databases using one of a

number of database-searching programmes (2D, two-dimensional; FTICR, Fourier-transform ion cyclotron resonance; HPLC, high-performance liquid chromatography) (35).

1.4.1 Immunoproteomics

Much of information about immunogenic components of a bacterial pathogen can be derived from proteomics coupled to Western blotting, namely immunoproteomics. The aim of the immunoproteomics studies is to identify the immunogenic antigens of a pathogen by using sera of human or mice. Immunproteomics has been used to identify novel bacterial vaccine candidates against several human pathogens, examples including *Helicobacter pylori, Staphylococcus aureus*, *Bacillus anthracis*, *Shigella flexneri*, *Francisella tularensis*, *Corynebacterium diphtheriae*, *Streptococcus pyogenes*, *Chlamydia pneumoniae*, *Neisseria meningitidis* and *Bordetella pertussis* (36).

During my master thesis (2004-2007, Middle East Technical University, Ankara), I had used immunoproteomics in order to study *B. pertussis*, the causative agent of highly communicable respiratory infection whooping cough (http://etd.lib.metu.edu.tr/upload/12608320/index.pdf). In a few words, the total soluble proteins extracted from two *B. pertussis* strains, Tohama I and the local isolate Saadet, two strains used for vaccine production in Turkey, were separated by two-dimensional gel electrophoresis and analyzed by Western blotting for their reactivity with the antisera obtained from the mice immunized with inactivated whole cells as well as those collected from the mice challenged intraperitoneally with live cells of each strain. Of a total of 25 immunogenic proteins identified, in which 21 were shown to be the novel antigens for *B. pertussis*. This was the first immunoproteomics study of the *Bordetella* and has

provided us a deeper idea about the pathogen.

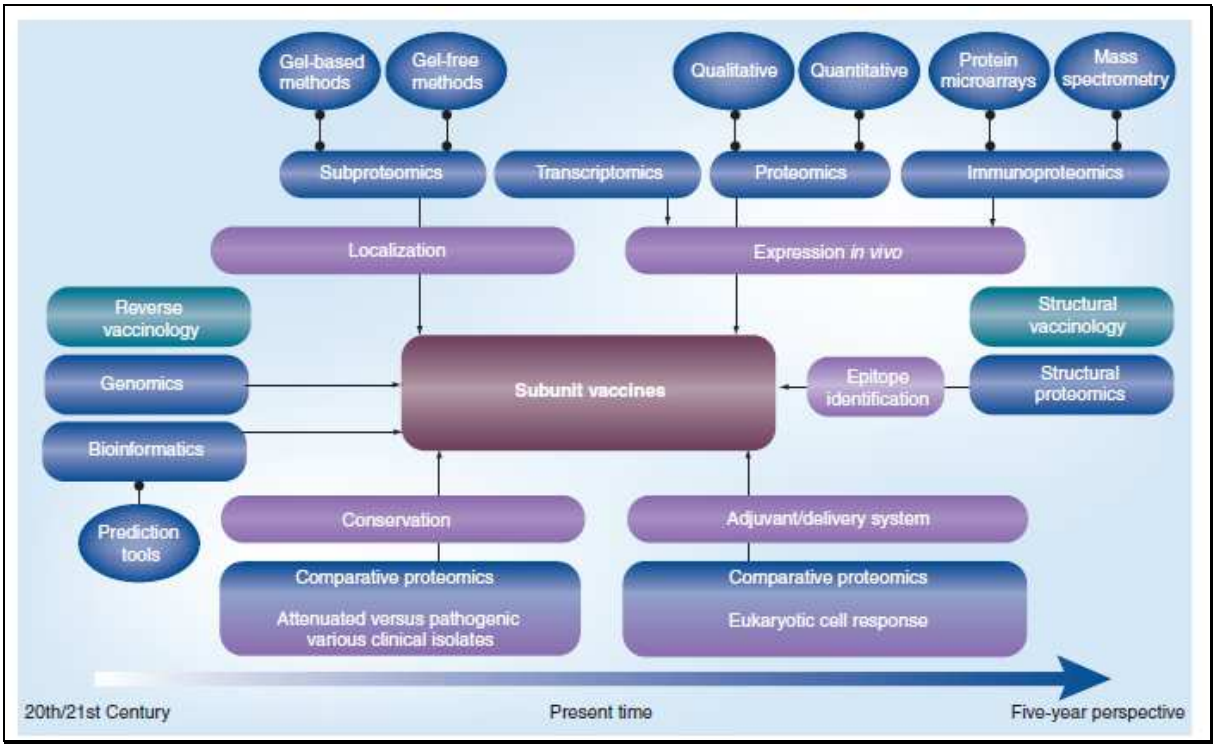


Figure 7: Impact of proteomics on vaccine development (37)

1.4.2 Surfome Analysis as a Vaccine Discovery Tool

Bacterial surface proteins are essential compartments of the pathogens. They have fundamental roles in interaction with the host and environment hence they are the main virulence factors that involve in adhesion and invasion of the host cells. Moreover, because surface proteins are likely to interact with the host immune system, they may become components of effective vaccines. There are three main methods currently in practice to identify surface proteins: (i) prediction by genome analysis using algorithms such as PSORT, (ii) separation of membrane and cell wall fractions from the cytoplasmic fraction and then

identification of proteins by two-dimensional (2D)-electrophoresis or 2D-chromatography coupled to mass spectrometry and (iii) definition of membrane proteins by using one of the two methods described above and then confirmation of surface localization by producing polyclonal antibodies against the recombinant forms of each predicted protein and by assaying antibody binding to whole bacterial cells. All these methods are used widely but they are all extremely labor intensive and/or not fully reliable and is not quantitative (38).

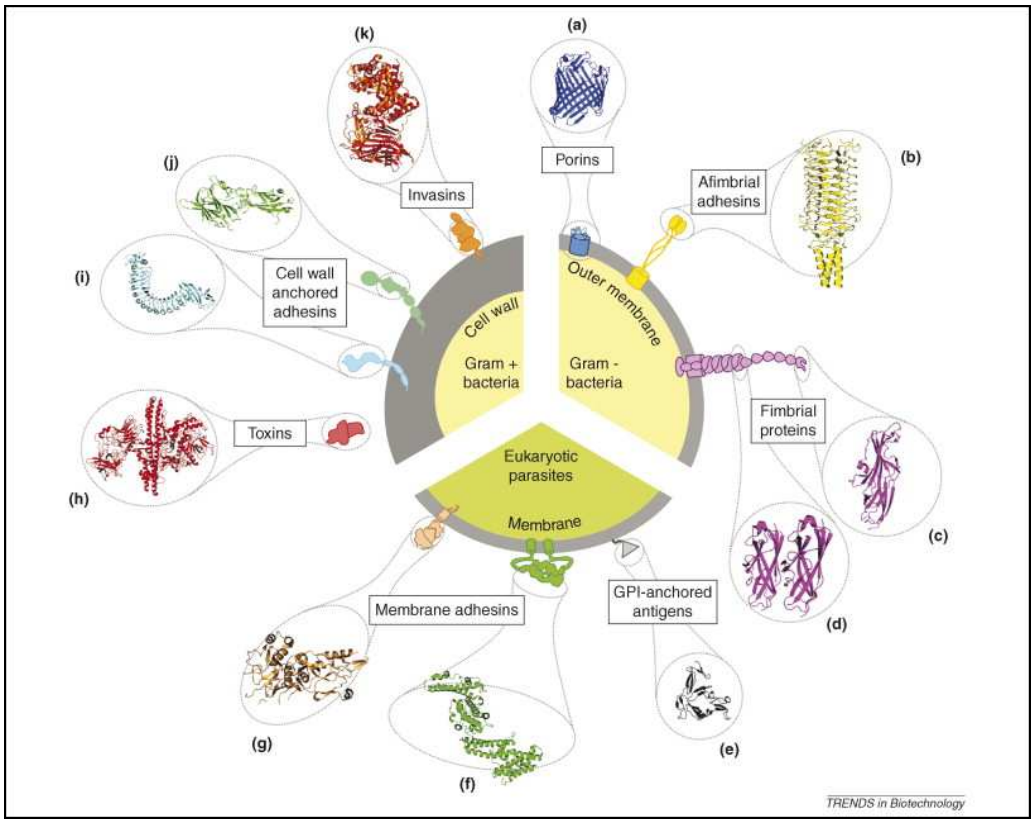


Figure 8. Structures of surface-exposed and secreted proteins relevant to the pathogen–host interface. Represented examples can be specific or common to different pathogen classes. The following indicates the antigen's name, pathogen's species and PDB codes, respectively, for each structure drawn: **(a)** OmpF, Escherichia coli, 1gfn; **(b)** YadA, Yersinia pestis, 1p9h; **(c)** PapG, Escherichia coli, 1j8s; **(d)** PapE-PapK, Escherichia coli, 1n12; **(e)** Psv25, Plasmodium vivax, 1z27; **(f)** EBA127, Plasmodium falciparum, 1zrl; **(g)**

AMA-1, Plasmodium vivax, 1w8k; **(h)** BoNT-B, Clostridium botulinum, 1epw; **(i)** InlA, Lysteria monocytogenes, 1O6T; **(j)** SdrG, Staphylococcus epidermidis, 1r19; **(k)** SpnHl, Streptococcus pneumoniae, 1egu. (39)

Recently, Rodriguez-Ortega *et al.* from Novartis Vaccines, described a new approach that allows the rapid and selective identification of bacterial surface-exposed proteins, the pool of proteins which are entirely or partially exposed on the outside of bacterial cells. The method uses proteolytic enzymes to 'shave' the bacterial surface and the peptides generated are separated from the whole cells and identified by mass spectrometry. The approach described provided the most extensive and detailed map of the surface-exposed antigens of a GAS isolate to date. A relevant result of this work was the demonstration that comprehensive characterization of surface-exposed proteins can lead to new vaccine candidate discovery. Among the 14 identified surface proteins tested, one protein, Spy0416, conferred high protection levels. This was a remarkable result, considering the small number of protective antigens that have been identified to date (38).

Secondly, our group used this approach in order to analyze the surfome of Group B Streptococcus, COH1 strain and to identify new vaccine candidates. We confirmed previous data showing that whole viable bacterial cell treatment with proteases followed by the identification of released peptides by mass spectrometry is the method of choice for the rapid and reliable identification of vaccine candidates in Gram-positive bacteria. When applied to the Group B *Streptococcus* COH1 strain, 43 surface-associated proteins were identified, including all the protective antigens described in the literature as well as a new protective

antigen, the cell wall-anchored protein SAN_1485 belonging to the serine-rich repeat protein family (40).

In the present study, the surface digestion methodology was exploited in order to analyze expression of new protective antigens on the surfome of GBS.

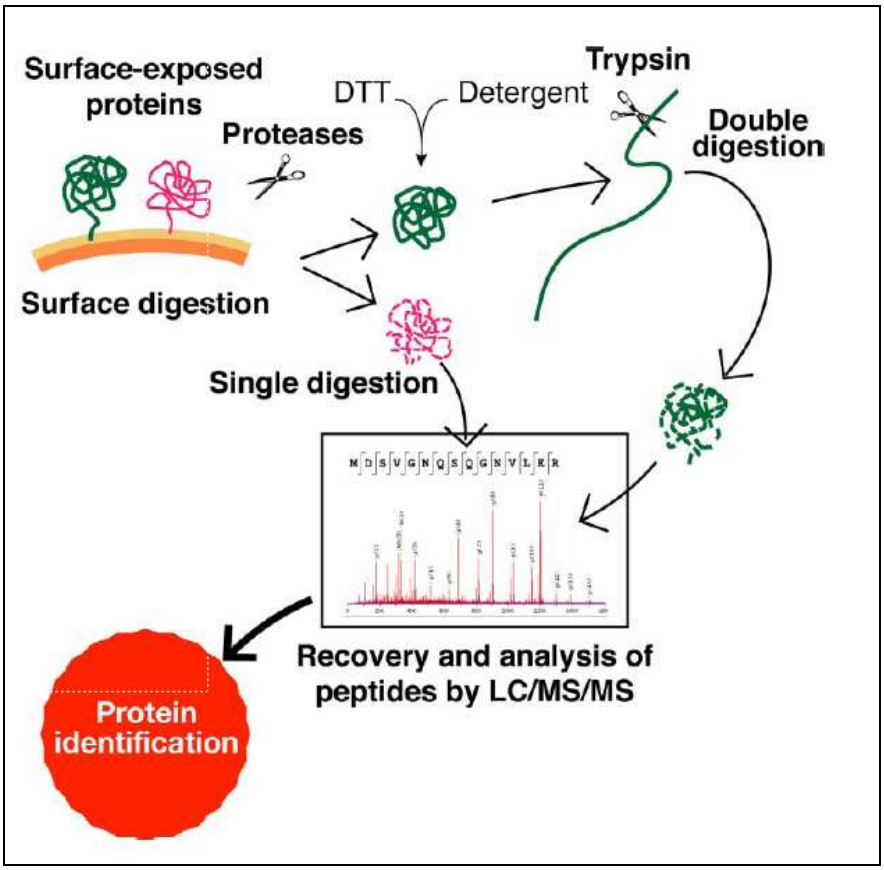


Figure 9. Representation of the proteomics strategy used to identify surface-exposed proteins (40).

1.5 Protein Domains

Traditionally, scientists use sequence similarity searches to compare a query sequence to those of known function, but this method has its limitations and relies on the quality of existing data. Alternative methods for protein sequence classification use protein signatures. A number of different databases developing protein signatures diagnostic for known protein families or domains have arisen (41).

Protein domains are compact regions of a protein's structure that often convey some distinct function. Domain architecture, or order of domains in a protein, is frequently considered as a fundamental level of protein functional complexity. The prevalence of proteins with more than two domains and the recurrent appearance of the same domain in non-homologues proteins show that functional domains are reused when creating new proteins. Because of this, domains have been likened to Lego bricks that can be recombined in various ways to build proteins with completely new functions. Hence, one way to study evolution of protein function and structure is by looking at the evolution of protein domain composition. The average length of a protein domain is approx. 120 amino acids, so changes in domain architecture are underlined by large alterations at the gene level (42). Examples of molecular mechanisms that can direct these rearrangements are gene fusion and fission, exon shuffling through intronic recombination, alternative gene splicing and retropositioning

Ernest Rutherford, who is known as the father of nuclear physics famously and contemptuously said, 'All science is either physics or stamp collecting'. But 'stamp collecting' or classification is of central importance in science. The advent of high

throughput sequencing and bioinformatics has enabled the classification of the proteins through the identification of sequences similarities they contain. These similarities are often characteristic of shared protein domains, which can be considered as the common currency of protein structure and function (43). Pfam is a large collection of protein domains and families. Its scientific goal is to provide a complete and accurate classification of protein families and domains. The Pfam database is accessible via the Web and available in several different downloadable formats (http://pfam.sanger.ac.uk). Currently Pfam matches 72% of known protein sequences, but for proteins with known structure Pfam matches 95%, which represents the likely upper bound.

Function prediction is filled with potential pitfalls such as considerable sequence divergence, non-equivalent functions of homologues and non-identical multi-domain architectures. Detecting non-enzymatic regulatory domains is essential to predict a protein's cellular role, binding partners and subcellular localisation. Such domains are usually divergent in sequence and occur in contrasting multi-domain contexts. This leads to difficulties unravelling the evolution and function of multi-domain proteins. These problems are addressed by the SMART Web tool as a database for signalling domains. SMART (a Simple Modular Architecture Research Tool) allows the identification and annotation of genetically mobile domains and the analysis of domain architectures (http://SMART.embl-heidelberg.de). More than 400 domain families found in signaling, extra-cellular and chromatin-associated proteins are detectable. These domains are extensively annotated with respect to phyletic distributions, functional class, tertiary structures and functionally important residues (44).

Both of these two bioinformatics tools, SMART and Pfam, are used extensively for the present study in order to analyze the protein architecture of known protective antigens.

1.6. Model Pathogens for the MetaVaccinology Approach

1.6.1 Streptococcus agalactiae

Invasive bacterial infections and the ensuing severe inflammatory response remains a significant cause of morbidity and mortality in human newborns and adults. Group B Streptococci (GBS), or *Streptococcus agalactiae,* is the most common cause of life-threatening bacterial infection in human newborns. These bacteria are Gram-positive, β-hemolytic, chainforming cocci that are normal residents of the vaginal flora in 25% of healthy women. Transmission of GBS from colonized mothers to the newborn can occur *in utero* owing to ascending infection or during birth when the neonate aspirates contaminated amniotic/vaginal fluids. Affected newborns include preterm, low birth weight and full-term infants GBS disease in newborns is classified as early-onset disease (EOD) or late-onset disease(LOD), depending on the age of the infant at the time of disease manifestation. Maternal colonization is a prerequisite for EOD, and infection presents in infants within a few hours to days of life (≤7 days of age). EOD manifests as respiratory failure and pneumonia that rapidly progresses into bacteremia and septic shock syndrome. In contrast, LOD is characterized by bloodstream infection, with a high risk of progression to meningitis. LOD can present in infants up to several months in age (7–90 days) (45).

Studies demonstrate that pregnant women, who have opsonically active levels of antibodies to GBS, are unlikely to deliver babies that suffer from GBS infections. For example, in one

study, 88% of babies were protected against GBS Ia if the maternal anti-GBS Ia antibody was greater than 5 μg/ml. Unfortunately, the majority of women do not have adequate levels of such antibody (66% UK and 88% USA). This knowledge is the basis of the vaccination idea (46). Human isolates of GBS express a capsular polysaccharide (CPS), a major virulence factor that helps the microorganism evade host defence mechanisms. Isolates of GBS can be divided into nine CPS serotypes (Ia Ib, II, III, IV, V, VI, VII and VIII) each antigenically and structurally unique. In the 1930s, Rebecca Lancefield established that protection against GBS infection in mice could be achieved using CPS-specific polyclonal rabbit serum. CPS-tetanus toxoid conjugate vaccines effective against all nine currently identified GBS serotypes have been prepared and were shown to induce functionally active CPS-specific IgG. Clinical trials of conjugate vaccines prepared with purified CPS types Ia, Ib, II, III and V have demonstrated that these preparations are safe and immunogenic. Not unexpectedly, these preparations do not offer protection against other GBS serotypes, such as type VIII, prevalent in other regions of the world. As an alternative to CPS-based vaccines, a number of groups have explored the development of vaccines based on antigenic proteins. So far, however, the protein antigens investigated have been restricted to particular serotypes and no complete cross-serotype protection has been achieved (47). To address this need, efforts are ongoing to develop a universally effective vaccine for GBS disease that exploits the recently acquired genomic sequences of GBS strains, and to then use this information to identify new candidate antigens of global relevance. As deeply mentioned above, GBS is the one of the first organism that Reverse Vaccinology approach is used, by using pan-genome information.

Vaccine target	Advantages/ approach	Limitations
Capsular carbohydrate		
Unmodified polysaccharide vaccine (type III serotype)	Phase I trials indicated that the vaccine was safe and well tolerated[67]	Only 60% of the recipients showed an immune response; Requirement to improve immunogenicity of the CPS
Conjugate polysaccharide vaccine	Type III serotype: increase in immunogenicity when coupled to an immunogenic protein (tetanus toxoid (TT)); Conjugate vaccine with all nine currently identified GBS serotypes (Ia, Ib, II, III, IV, V, VI, VII and VIII) prepared and tested preclinically[11,68,108,109]	Capsular conjugate vaccines of this type need to be multivalent in order to provide sufficient coverage against prevalent serotypes
Conjugate bivalent polysaccharide vaccine	Bivalent vaccine (GBS type II-TT and type III-TT) combined and administered; Well tolerated	Further testing is warranted to investigate immune interference when more than two GBS CPS conjugate vaccines are simultaneously administered[10]
Conjugate multivalent polysaccharide vaccine	Proposed that effective GBS vaccine in the United States includes five major serotypes (Ia, Ib, II, III and V); It is anticipated that multivalent vaccines will include each conjugate vaccine prepared separately[10]	Formulation of a GBS conjugate vaccine for use in the United States might not be effective in other regions[110]
Proteins		
C5a peptidase	Present on all strains and serotypes of GBS; Little or no antigenic variability; Capable of inducing antibodies that are opsonically active[73]; Immunization induces serotype-independent protection	Progress as a potential vaccine is unknown
β-Component of the C protein	Elicits protective immunity in animal models[111]	This protein is only present in a minority of strains that cause infection (~20%)
LmbP	Expressed by most GBS strains	Progress as a potential vaccine is unknown[84]
Sip	Present on all GBS strains; Induces protective antibodies; Recombinant SIP protein protected mice infected with numerous GBS strains[72]	Biological function is not well understood; No recent reports of progress towards the development of a vaccine[14,83,84]
LrrG	Highly conserved protein antigen that induces protection[85]	Progress as a potential vaccine is unknown

Figure 10: Current status of GBS vaccine research and development, CPS, capsular polysaccharide; LmbP, laminin binding protein; Sip, surface immunogenic protein (47).

1.6.2 *Staphylococcus aureus*

Staphylococcus aureus is an important gram positive bacterial pathogen that causes skin and soft-tissue infections as well as life-threatening bacteraemias with metastatic complications, such as pneumonia, endocarditis, septic arthritis and osteomyelitis (48). Methicillin-resistant *S. aureus* (MRSA) isolates are resistant to all available penicillins and other β-lactam antimicrobial drugs. They were once confined largely to hospitals, other health care environments, and patients frequenting these facilities. Since the mid-1990s, however, there has been an explosion in the number of MRSA infections reported for populations lacking risk factors for exposure to the health care system. This increase has been associated with the recognition of new MRSA strains, often called community-associated MRSA (CA-MRSA) strains that have been responsible for a large proportion of the increased disease burden observed in the last decade. These CA-MRSA strains appear to have rapidly disseminated among the general population in most areas of the United States and affect patients with and without exposure to the health care environment (49).

The pathogenicity of *S. aureus* is particularly complex, involving numerous bacterial products as well as elaborated regulation pathways. *S. aureus* is able to produce a wide range of toxins showing a deleterious effect on cell integrity and functions. Most of these factors (e.g., toxic shock syndrome toxin-1, exfoliatin toxins A and B, Panton-Valentine leukocidin, enterotoxins, and hemolysins) contribute to the virulence of clinical isolates in the context of acute infections (50). In addition to these excreted compounds, *S. aureus* is able to produce several cell wall–associated proteins allowing interactions with host plasma or extracellular

proteins, such as fibronectin, fibrinogen, collagen, vitronectin, laminin, and bone sialoprotein.

No immunological therapy and/or prophylaxis for *S. aureus* infections is available, but it might be possible as (i) previous infective exposure to the pathogen results in subsequent less severe infections; (ii) patients with higher anti-staphylococcal antibody levels are less likely to contract staphylococcal infections and (iii) a combined killed-whole bacteria plus toxoid vaccine showed 50–70% protection against bovine mastitis. Historically, a variety of whole bacteria and toxoid anti-staphylococcal preparations were used in clinical and veterinary trials, with little benefit and common adverse reactions. An ideal vaccine against *S. aureus* would aim to prevent bacterial adherence, promote phagocytic killing and/or neutralize toxic exoproteins, and be optimally directed to one or more well-characterized antigenic targets expressed during infection (51). We used *S. aureus* as a second model organism for MetaVaccinology approach in order to identify new vaccine candidate against this life-threatening pathogen.

1.7 Aim of the study

The Reverse Vaccinology approach, defined at the beginning of this century, opened the way to Genomics applied to vaccine research, by shifting the paradigm of vaccine discovery from conventional culture-based methods to high-throughput genome-based approaches for the development of recombinant protein-based vaccines. These are comprehensive, large-scale approaches, where vaccine candidates are selected according to their *in silico* predicted subcellular localization and then screened in relevant animal models. These genome-based approaches have been successfully applied to a range of bacteria, including *Neisseria meningitidis* serogroup B (21), *Streptococcus agalactiae* (25), *Chlamydia trachomatis* (52), *Chalmydia pneumoniae* (53), *Bacillus anthracis* (54), *Bacillus anthracis* 55), *Brucella melitensis* (56) and, very recently to extraintestinal pathogenic *Escherichia coli* (26) addressing challenging human pathogens for which conventional methods failed to identify a vaccine.

Besides reaching their main goal of identifying good vaccine candidates, these efforts produced also a huge amount of molecular knowledge related to the basic biology and the pathogenesis mechanisms of many bacteria, examples being breakthrough discoveries of pili in pathogenic *Streptococci* (57) and of innate immunity subverting systems in staphylococci and meningococci (58). Nevertheless, the question of which are the elemental molecular properties that can be used to *in silico* predict the few bacterial proteins able of eliciting a protective immune response, and thus to be used as effective vaccine components, among the complete bacterial proteome remains a challenge. Moreover, these approaches still require labor intensive activities in the wet lab to find out, among the large number of secreted and

surface exposed proteins, those very few which are protective. For these reasons, *in silico* methods for the selection of possible new vaccine candidates are still an attractive topic in vaccine discovery processes.

The aim of the present study is to analyze the structural and molecular properties of these known bacterial protective antigens in order to (i) investigate possible common molecular features among these bacterial protective antigens and (ii) extract possible predictive rules leading to a simplification in the vaccine discovery process based on the *in silico* prediction of possible new protective antigens. Group B *Streptococcus* and *S. aureus* are used as model pathogens to develop this new vaccine discovery tool.

2. MATERIAL AND METHODS

2.1 Meta-analysis of bacterial protective antigens

A comprehensive list of bacterial antigens described as potential vaccine candidates was created from literature data and from on line available databases, *e.g.* the VIOLIN web site (http://www.violinet.org , 59). 115 different protective antigens from 23 bacteria, both gram negative and gram positive, and the corresponding proteins sequences were considered for this process. A systematic analysis is carried out for this dataset using different bioinformatics tools, looking for conserved molecular features at the following molecular levels: primary sequence (BLAST, http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins), domain occurrence (Pfam, http://pfam.sanger.ac.uk/) protein architecture (SMART, http://smart.embl-heidelberg.de/smart/set_mode.cgi?GENOMIC=1) and 3D structure (PDB and Pfam Clans). The analysis of the alignments within the protective antigens was carried out by using ClustalW (http://www.ebi.ac.uk/clustalw/).

2.2 Bacterial Surface Digestion of *Streptococcus agalactiae*

8 different *Streptococcus agalactiae* strains, belonging to the most important disease-causing serotypes Ia (515 and A909), Ib (H36B), II (18RS21), III (COH1 and NEM316), and V (2603 and CJB111) were plated o/n in blood agar separately (Trypticase™ Soy Agar II with 5% sheep blood, Becton Dickinson & Co., Heidelberg, Germany). The bacteria colonies were grown at 37°C in 200 ml of Todd-Hewitt broth (THB) in the presence of 5% CO_2 until an OD_{600} of 0.3 was reached. Bacteria were harvested by centrifugation at 3,500 x *g* for 10

min at 4°C, and washed twice with PBS. Cells were resuspended in 800 µL of PBS containing 40% sucrose (pH 7.4 for trypsin or papain digestions and pH 6.0 for proteinase K digestion). Digestions were carried out with 10 µg trypsin (Promega, Madison, U.S.A) or 5 µg proteinase K (Sigma, St. Louis, USA) for 30 min at 37°C.. Bacterial cells were then spun down at 3,500 x *g* for 10 min at 4°C and the supernatants were filtered through 0.22 µm pore-size filters (Millex, Millipore, Beford, U.S.A). Protease reactions were stopped with formic acid at 0.1% final concentration. Before analysis, PBS and sucrose were removed by off-line desalting procedure using OASIS cartridges (Waters, Milford, USA) following producer's protocol. Desalted peptides were concentrated with a Centrivap Concentrator (Labconco, Kansas City, U.S.A), and kept at -20°C until further analysis.

2.2. 1 Bacterial Surface Double Digestion

Streptococcus agalactiae strains were cultured and surface-digested as described above. Digestion supernatants were then denatured and reduced with Rapigest® (Waters) and 5 mM DTT at 100°C respectively for 10 min and an additional o/n proteolytic step with 2 µg trypsin (Promega) at 37°C was performed. The second digestion reaction was stopped with formic acid at 0.1% final concentration. The peptide mixtures were then desalted and concentrated as described above and stored at -20°C until further analysis.

2.2.2 Protein Identification By Nano-LC/MS/MS

Peptides were separated by nano-LC on a NanoAcquity UPLC system (Waters) connected to a Q-ToF Premier Electro Spray Ionization (ESI) mass spectrometer equipped with a nanospray source (Waters). Samples were loaded onto a NanoAcquity 1.7µm BEH130 C_{18}

column (75µm X 25mm, Waters), through a NanoAcquity 5µm Symmetry® C_{18} trap column (180µm X 20mm, Waters). Peptides were eluted with a 120-min gradient of 2–40% of 98% acetonitrile, 0.1% formic acid solution at a flow rate of 250 nL/min. The eluted peptides were subjected to an automated data-dependent acquisition, using the MassLynx software, version 4.1 (Waters), where a MS survey scan was used to automatically select multicharged peptides over the *m/z* ratio range of 300–2,000 for further MS/MS fragmentation. Up to eight different components were subjected to MS/MS fragmentation at the same time. For all samples, a second nano-LC-MS/MS analysis was carried out for the selective fragmentation of mono-charged peptide species.

After data acquisition, individual MS/MS spectra were combined, smoothed and centroided using ProteinLynx, version 3.5 (Waters) to obtain the peak list file. The Mascot Daemon application (Matrixscience Ltd., London, UK) was used for the automatic submission of data files to an in-house licensed version of MASCOT, version 2.2.1, running on a local server. Protein identification was achieved by searching in a locally created database containing protein sequence data derived from the eight completely sequenced GBS strains.The MASCOT search parameters were set to (i) 1 as number of allowed missed cleavages (only for trypsin digestion), (ii) 0.3Da as peptide tolerance, and (iii) 0.3Da as MS/MS tolerance. Only significant hits were considered, as defined by the MASCOT scoring and probability system. The score thresholds for acceptance of peptide identification were ≥18 for trypsin digestion or ≥36 for proteinase K and papain digestions.

2.3 Selection of MetaV candidates

2.3.1 MetaV Antigens for *GBS*

UniProt was used for obtaining the sequence information of the GBS genomes (http://www.uniprot.org/). The corresponding protein sequence information from the UniProt knowledgebase was scanned for the occurrence of the MetaV molecular features and occurance of the multiple internal repeats (60). Prospero was used to scan the GBS genomes for ORFs with more than two internal repeats combined to PSORTb to predict their possible subcellular localization, and ten multiple internal repeat-containing proteins were added to the list. GBS specific ORFs were obtained from the genomic analysis carried out by Tettelin *et al.* (61).

2.3.2 Selection of MetaV candidate antigens for *S. aureus*

The MetaV list for *S. aureus* is prepared as described above for GBS. Briefly, UniProt was used for obtaining the sequence information of the NCTC 8325 strain of *Staphylococcus aureus* (http://www.uniprot.org/). The corresponding protein sequence information from the UniProt knowledgebase was scanned for the occurrence of the MetaV molecular features and occurance of the multiple internal repeats. Prospero was used to scan the NCTC 8325 genome for ORFs with more than two internal repeats combined to PSORTb to predict their possible subcellular localization, and ten multiple internal repeat-containing proteins were added to the list.

2.5 Cloning, Expression and Purification of Selected GBS Proteins

2603 V/R strain was used as source of DNA for amplification of selected MetaVaccinology candidates. Genes coding for the protein SAG_0954 was cloned as C-terminal His-tag fusion protein and then expressed and purified as already reported (25). PCR primers were designed to amplify gene without predicted signal peptide coding sequences for all proteins. PCR fragments were cloned by using the Polymerase Incomplete Primer Extension (PIPE) method, developed by GNF (Genomics Institute of the Novartis Research Foundation, San Diego, CA, USA). All cloning operations were achieved by transforming HK100 competent cells with PCR products (I-PCR) immediately following amplification mixed with the V-PCR of SpeedET vector (N-term 6xHis tag) (62) . Protein expression was achieved maintainig the cultures at 25°C for 4h after the induction using arabinose 0.2 % final concentration. Procedures for protein purification were as described in**,** briefly cells were harvested by centrifugation and lysed in "B-PER buffer" (Pierce) containing lysozyme 1mg\ml, DNAse 0.5 mg\ml and COMPLETE inhibitors cocktail, (Roche). The cell-lysate was clarified by centrifugation and applied onto His-Trap HP column (Armesham Biosciences) pre-equilibrated in buffer containing 10mM imidazole. Protein elution was performed using an imidazole gradient (25). Protein concentration was estimated using BCA assay (Pierce). SAG_0954, SAG_0416, SAG_0771 were kindly provided by GBS project of Novartis Vaccines.

Table 2: PCR primers designed to amplify corresponding gene

Gene ID	Sequence of Primers
SAG_1386	5'-CTGTACTTCCAGGGCTTTATAGTATTTTATACTTCAAATAGA-3' (forward) 5'-AATTAAGTCGCGTTAATTTACAACCACAGGATCGCCTGGATT-3' (reverse)
SAG_1333	5'-CTGTACTTCCAGGGCGACCAAGTCGGTGTCCAAGTTATAGG-3' (forward), 5'-AATTAAGTCGCGTTAAGTACCAATTTTAGCTTCTGTTACTTG-3' (reverse).
SAG_0907	5'-CTGTACTTCCAGGGCCAAGAACACAAAAATTCTCATCATATT-3' (forward), 5'-AATTAAGTCGCGTTAATGGTGATGATGACCTACATGTGC-3' (reverse).
SAG_2148	5'- CTGTACTTCCAGGGCTTACCACTTTCAGTAAGCGCAGCA-3' (forward), 5'- AATTAAGTCGCGTTATTAATACCAGCCGTTACTATTCCAAA-3' (reverse).
SAG_1350	5'- CTGTACTTCCAGGGCACAAGTCCTGTTTTTGCGGATC-3' (forward), 5'- AATTAAGTCGCGTTATAAACCATTTTCAATAGGTTCTTGAG-3' (reverse).
SAG_0017	5'-CTGTACTTCCAGGGCAACGCTGATGACTTTGACTCGAAAATTG-3'(forward), 5'- AATTAAGTCGCGTTAAGTAGCTGTAGCTGTAGTTGTAGC-3' (reverse).

2.4 Cloning, Expression and Purification of Selected *Staphylococcus aureus* Proteins

NCTC 8325 strain was used as source of DNA for amplification of selected MetaVaccinology candidates. All the proteins are cloned, expressed and purified according to procedures described above for the GBS proteins.

2.5 Active Maternal Immunization for GBS proteins

A maternal immunization/neonatal pup challenge model of GBS infection was used to verify the protective efficacy of the MetaV candidates as previously described in (16). In brief, CD-1 female mice (6-8 weeks old) were immunized before breeding, on days 1, 21 and 35. The mice received either PBS or 20 µg of protein per dose. Mice were bred 2-7 days after the last immunization. Within 48 h of birth, pups were injected intraperitoneally with 50 µl of GBS COH1 culture corresponding to a LD_{90}. Challenge inocula were prepared starting from frozen cultures diluted to the appropriate concentration with THB. Survival of pups was monitored for 2 days after challenge. Protection was calculated as (percentage deadControl minus percentage deadVaccine) divided by percentage deadControl multiplied by 100. The 515 (GBS-01) and COH1 (GBS-02) hypervirulent strains were used for challenging the mice.

2.6 S. aureus Immunization Experiments

CD-1 female mice (5 week old) were intraperitonally immunized with 20 µg protein+ALUM, on days 1 and 14, Then on 24^{th} day, mice are intraperitonally challenged with 100 µl of *S.aureus* Newman culture corresponding to $5x10^8$ CFU/mice. Survival of mice was monitored for fifteen days. SAOUHSC_00427, SAOUHSC_01949, SAOUHSC_02979, SAOUHSC_02576 and a protective *S. aureus* antigen (as a positive control) were used for the immunization schema against *S. aureus* (SA-1). The same immunization schema was used to test proteins SAOUHSC_00356, SAOUHSC_00400, SAOUHSC_00248, SAOUHSC_00256, SAOUHSC_00994, SAOUHSC_00392 (SA-2)

2.7. Functional Characterization of SAG_1333

2.7.1 xCELLigence System with carcinomic human alveolar basal epithelial cells (A549)

The xCELLigence measurement was performed with the xCELLigence System from Roche that monitors cellular events in real time without the incorporation of labels (http://www.roche-applied-science.com/sis/xcelligence/ezhome.html). The System measures electrical impedance across interdigitated micro-electrodes integrated in the bottom of tissue culture E-Plates. The impedance measurement provides quantitative information about the biological status of the cells, including cell number, viability, and morphology.

Prior to seeding the cells in the E-Plates, the concentration of the cell-suspension was determined by Invitrogen™ Countess® Automated Cell Counter. $5*10^4$ A549 cells were seeded in 14 of 16 wells. The E-Plates were filled with 50µl Medium (Invitrogen™ RPMI 1640 Medium 1X, liquid, 2 % FBS (Fetal Bovine Serum), 1% Antibiotics) and 50µl cell suspension. The E-Plates were incubated at room temperature for 30 min and then at 37°C for 24h for cells to reach confluence in the wells. After 24h incubation, the supernatant was removed in the E-plates and 50µl of the reactant is added. (The level of concentrations is listed on Table 3 below). The behavior of the cells was monitored by the xCELLigence System for the next 24h at 37°C.

Table 3: Concentrations of compounds 50µl of reactants were added when the cells have reached confluence. The reactants were diluted in PBS. α :a hemolysin from GAS β: 137mM NaCl 2.7mM KCl 4.3mM $NaHPO_4$ 1.47mM KH_2PO_4 γ: A non-toxic protein of *S.aureus* as control

	Compounds	Concentrations		
Positive control	Pore forming toxin $^{\alpha}$	20µl/ml		
Negative Control	PBS$^{\beta}$	1:1000	1:100	1:10
Negative Control	Non toxic protein$^{\gamma}$	25µl/ml		
	N-term of SAG_1333	25µl/ml	10µl/ml	
	Full length SAG_1333	25µl/ml	10µl/ml	5 µl/ml

In the wells with non toxic protein, SAG_1333 and the N-term of SAG_1333, different concentrations of AMP were added:

AMP	0mM	0.5mM	5mM

2.7.2. xCELLigence with mouse leukemic monocyte macrophage cell (Raw cells)

The same procedure used for A549 cell line was applied to Raw cells. Table 4 shows the compounds and concentrations used for the experiments. All experiments were repeated at least two times for both cell lines.

Table 4: Concentrations of compounds 50µl of reactants were added when the cells have reached confluence. The reactants were diluted in PBS. α: α heamolysin from GAS β: 137mM NaCl 2.7mM KCl 4.3mM $NaHPO_4$ 1.47mM KH_2PO_4 γ: A non-toxic protein of *S.aureus* as control

	Compounds	Concentrations		
Positive control	Pore forming toxin $^{\alpha}$	20µl/ml		
Negative Control	PBS$^{\beta}$	1:1000	1:100	1:10
Negative Control	Non toxic protein$^{\gamma}$	25 µl/ml		

N-term of SAG_1333	25µl/ml	10µl/ml			
Full length SAG_1333	100µl/ml	50µl/ml	25µl/ml	10µl/ml	5 µl/ml

AMP with different concentrations were added:

AMP	0mM	0.1mM	1mM	2.5mM	5mM

3. RESULTS AND DISCUSSION

3.1 Main Question: What makes an antigen protective?

Today, we have many genomic tools to investigate the microbial world as summarized in the introduction. Moreover, all these different instruments open new insights for vaccinology science but the main question remains constant: what makes an antigen protective and why? Could we select this set of protective antigens by just using *in-silico* approaches? Reverse Vaccinology made a step forward to reply this question and caused a paradigm shift by selecting all surface associated and secreted proteins as potential vaccine candidates. But even this progressive approach selects hundreds of proteins to be tested in *in-vivo* models and the percentage of protective antigens within all set is very low.

In the present study, we tried to answer this simple but basic question: are there any recurring/common molecular features within all identified protective antigens? If there are, could we use these recurring signatures as predictive features to identify new vaccine candidates? According to our knowledge, this is the first time that this question is answered, and yes…

3.2 Systematic Analysis of Known Protective Antigens

In order to answer this fundamental question; a comprehensive list of bacterial antigens described as potential vaccine candidates, based on *in vivo* animal models and/or on *in vitro* assays, was created from literature data and from available on line databases, *e.g.* the VIOLIN web site (59). We were able to collect information from 115 different protective antigens from 23 bacterial pathogens, of those 13 are gram negative and 10 are gram positive. In particular, *Helicobacter pylori, Escherichia coli, Haemophilus influenzae Neisseria meningitidis serogroup B, Pasteurella haemolytica, Pseudomonas aeruginosa , Yersinia pestis, Campylobacter jejuni, Brucella abortus, Brucella burgdorferi, Brucella melitensis, Bordetella pertussis, Pseudomonas aeruginosa, Streptococcus pneumoniae*, *Streptococcus pyogenes, Streptococcus agalactiae, Staphylococcus aureus Mycobacterium tuberculosis, Bacillus anthracis, Listeria monocytogenes, Clostridium difficile, Clostridium tetani* and lastly *Clostridium botulinum* were used to construct our data set.

The corresponding proteins sequences of all these 115 different antigens became our total protective antigens dataset. We carried out a systematic analysis of this dataset using different bioinformatic tools, looking for conserved molecular features at the following molecular levels of each protein: primary sequence (BLAST), domain occurrence (PFam), protein architecture (SMART) and 3D structure (PDB and Pfam Clans).

The results of this meta-analysis revealed that protective antigens have recurring functional and/or structural units, which were in most cases associated to either specific domain from the Pfam databases or to a conserved protein architecture organized in a variable number of

multiple internal repeats. The rest of the protective antigens not showing these properties were found to be either species-specific or genus-specific ORFs by genomic analysis (Figure 11).

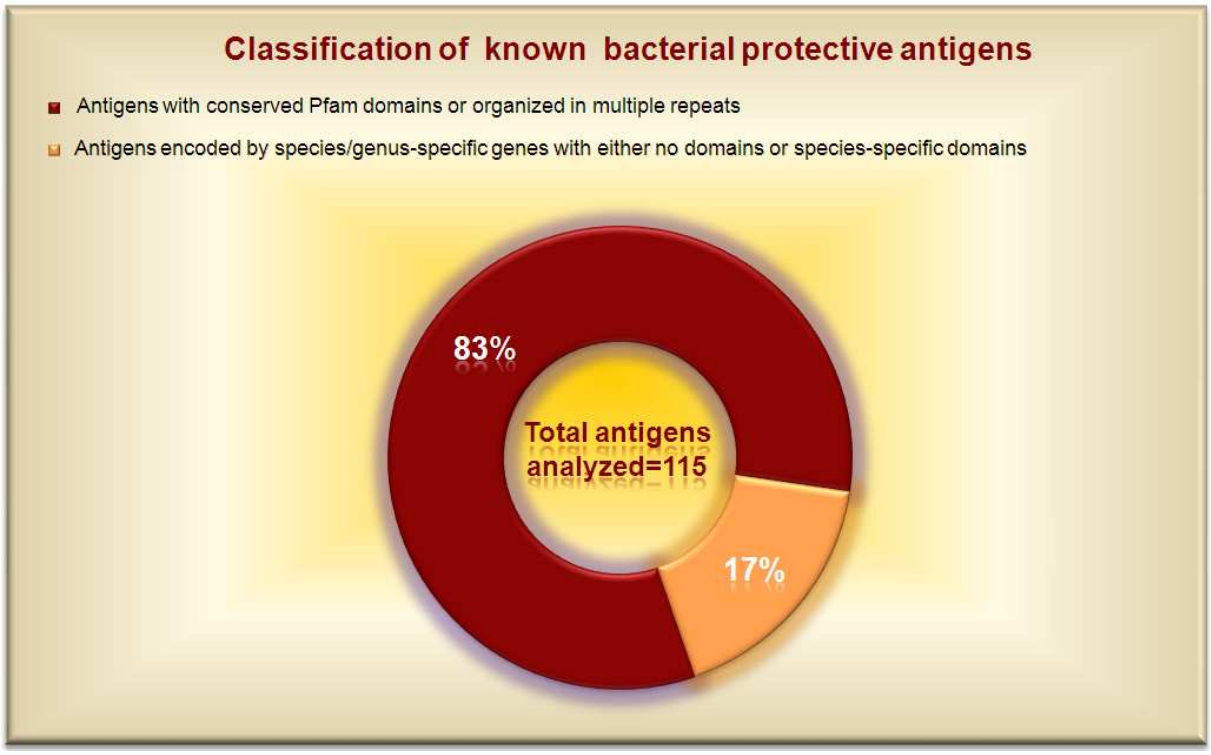


Figure 11: Classification of known bacterial protective antigens. 83 % of the proteins have conserved Pfam domains or organized multiple repeats while 13 % are encoded by species/genus specific genes with either no domains or species-specific domains.

Then we classified the Pfam domains based on their frequency of occurrence in the dataset of bacterial protective antigens and this analysis revealed that 35 Pfam domains are found in more than one protective antigen either from the same species or from different species and groups (Table 6).

Table 6. The recurring Pfam Domains within different protective antigens.

No	Recurring Pfam domains	Number of Species	Observed in
1	PF00691. OmpA.	6	Gram -
2	PF00082. Peptidase_S8	5	Both Gram - and +
3	PF01476. LysM.	5	Both Gram - and +
4	PF05738. Cna_B	4	Gram +
5	PF02872. 5_nucleotid_C., PF00149. Metallophos	3	Both Gram - and +
6	PF10425. SdrG_C_C	2	Gram +
7	PF04270. Strep_his_triad	3	Gram +
8	PF00669, Flagellin_N, PF00670, Flagellin_C	3	Gram +
9	PF01289, Thiol_cytolysin	3	Gram +
10	PF03895. YadA. 1 hit.	3	Gram -
11	PF08428. Rib	2	Gram +
12	PF05257. CHAP domain	2	Gram +
13	PF00877. NLPC_P60.	2	Both Gram - and +

14	PF02608. Bmp	2	Gram +
15	PF01547. SBP_bac_1.	2	Both Gram - and +
16	PF00756. Esterase.	2	Both Gram - and +
17	PF06013 WXG100	2	Gram +
18	PF00353. HemolysinCabind.	2	Gram -
19	PF01742. Peptidase_M27	2	Gram +
20	PF01473. CW_binding_1.	2	Gram +
21	PF03212. Pertactin	2	Gram -
22	PF02395. Peptidase_S6.	2	Gram -
23	PF04792. LcrV. 1 hit.	2	Gram -
24	PF01497. Peripla_BP_2.	2	Gram +
25	PF00593. TonB_dep_Rec., PF07715. Plug.	2	Gram -
26	PF01298. Lipoprotein_5.	2	Gram -
27	PF02876: Stap_Strp_tox_C	2	Gram +
28	PF06013: WXG100	2	Gram +
29	PF00089. Trypsin. 1 hit.	2	Both Gram - and +
30	PF00118. Cpn60_TCP1. 1 hit.	2	Gram -
32	PF00497. SBP_bac_3.	1	Gram -

33	PF11713. Peptidase_C80.	1	Gram +
34	PF05031. NEAT	1	Gram +
35	PF04488. Gly_transf_sug. 1 hit. PF11713. Peptidase_C80. 1 hit.	1	Gram +

On the other hand, primary sequence (BLAST) analysis revealed that only in very few cases protective antigens from different species show a significant sequence similarity (>50% ID). Table 6 reports representative examples of known bacterial protective antigens, their functional classification and the associated molecular features identified by the meta-analysis carried out in the present study.

Table 7 – **Functional classification of known protective antigens and their associated molecular features, including Pfam domain/clans and occurrence of multiple internal repeats.** Examples are provided for each functional class, with the corresponding % of identity obtained by primary sequence alignment. Relation between different protective antigens become obvious only looking at their domain composition and protein architecture organization

Classification	**Sub-classification**	**Examples**	**% ID**	**Pfam Families and Clans**
Flagellins		Flagellins from P.aeruginosa[63], E. coli [64], C. jejuni [65], H. pylori [66]	9- 41 %	PF00669, Flagellin_N PF00700, Flagellin_C
Major outer membrane proteins		C3389 from E. coli [67] , P6 from H. influenzae[68] Omp16 from B. abortus [69] outer membrane porin F from P. aeruginosa [70]	16-32 %	PF00691. OmpA

Toxins	Pore-forming toxins	Ply from S. pneumoniae [71], streptolysin O from S. pyogenes[72], Listeriolysin O from L. monocytogenes [73]	38-43 %	PF01289, Thiol_cytolysin
		Alpha-toxin[74] and Panton-Valentine Leucocidin F [75] from S. aureus	24 %	PF07968. Leukocidin
		Hemolysin from E. coli [26]		PF00353. HemolysinCabind.
	Ribosyl transferases	Difteriae toxin from *C. diphtheriae* [76], Pertussis toxin from *B. pertussis* [77], Cholera toxin from *V. cholerae* [78]	7-11%	CL0084. ADP-ribosyl

	Neurotoxins	Tetanus toxin from *C. tetani*[79] , Botulinum toxin from *C. botulinum*[80]	33 %	PF01742. Peptidase_M27
	Second messenger pathway activators	toxin A and toxin B from *C. difficile*[81]	47 %	PF01473. CW_binding_1. PF04488. Gly_transf_sug PF11713. Peptidase_C80.
Extracellular enzymes	Proteases	C5a peptidases from *S. agalactiae*[82], Spy0416 from *S. pyogenes*[38] ,prtA *from S. pneumonia*[83] , NMB1969 from *N. meningitidis*[84]	9-29 %	PF00082. Peptidase_S8
		SpeB from *S. pyogenes*[85]		PF01640. Peptidase_C10

		Penicillin-binding protein (D-alanyl-D-alanine carboxypeptidase) from *S.pyogenes* [86]		PF07943. PBP5_C PF00768. Peptidase_S11
	5'nucleotidases	NucA from *H. influenza* [87], Spy0872 from *S. pyogenes* [88], SAG_1333 from *S. agalactiae* [this study]	19-64 %	PF02872. 5_nucleotid_C PF00149. Metallophos.
Iron uptake and transport systems	Heme binding proteins	IsdA and IsdB from *S. aureus* [89, 90]	18 %	PF05031. NEAT
	TonB dependent receptors	IroN[91] ,FyuA [91], ChuA [92], Iha [92], IreA [92]and IutA [92]from *E.coli* , TbpA and TbpB from *P. haemolytica* [93]	10- 32 %	PF07715. Plug PF00593. TonB_dep_Rec

	Iron transport	PiaA from *S. pneumoniae* [94]		PF01497. Peripla_BP_2
Other solute binding proteins		P39 from *B. melitensis* [95], potD from *S. pneumoniae* [96], CjaA from *C. jejuni* [97], GNA1946 from *N. meningitidis* [21]	5-11%	CL0177. PBP
Adhesive structures with binding properties to ECM and/or serum proteins	Cna_B domain containing proteins	*Pili subunits from S. pyogenes* [98], *S. agalactiae* [99], *S. pneumoniae* [100], Cna [101], *SdrD*[102] *and SdrE*[102] from *S. aureus*	6- 58 %	PF05738. Cna_B
	Proteins containing variable numbers of internal repeats	FnbpA [101], clumping factor A[102], SdrC [102]and SdrD[102] from *S. aureus,* SAN_1485 from *S. agalactiae* [40]	10- 58 %	PF10425. SdrG_C_C

		M protein from *S. pyogenes* [103]		PF02370. M.
		Leucine rich protein LrrG from *S. agalactiae* [104]		Leucine rich repeats
		Rib protein [105] and C protein alpha-antigen from *S. agalactiae* [106]		PF08829. AlphaC_N. PF08428. Rib
		Factor H binding protein from *N. meningitidis* [107]		PF08794.Lipoprot_C
	Others	Heparin binding protein GNA2132 from *N. meningitidis*[108], TbpB from *H. influenzae* [109]	11 %	PF01298. Lipoprotein_5

		NadA from *N. meningitidis* [110] , UspA1 and UspA2 from *M. catharralis* [111], *YadA* and YadC from *Y. pestis* [112], C4424 *from E. coli* [92]	8- 23 %	PF03895. YadA
		PhtA [113], PhtB [113], SP_1174 [29] from *S. pneumoniae* , SAG_0907 from *S. agalactiae* this study	12- 65 %	PF04270. Strep_his_triad
		BmpA from *B. Burgdorferi* [114] , SAG_0954 from *S. agalactiae* this study	12 %	PF02608. Bmp
		Pertactin from *B. pertussis* [115], Hap from *H. influenza* [116], NMB1998 from *N. meningitidis* [117]	3-28 %	PF03212. Pertactin PF02395. Peptidase_S6.

Proteins with generic cell wall binding domains	LysM domain-containing proteins	P60 from *L. monocytogenes* [118] Intimin from *E. coli* [119], Ebps from *S. aureus* [120] , TspA from *Neisseria meningtidis* serogroup B [158], SAG_1386 from *S. agalactiae* [this study]	30 %	PF01476. LysM
	CHAP domain-containing proteins	PcsB from *S. pneumoniae*[29]		PF05257. CHAP domain
	Choline-binding motifs	PspA from *S. pneumoniae* [121]		PF01473. CW_binding_1
		PspC from *S. pneumoniae* [121]		PF05062. RICH.

3.3 The MetaVaccinology approach: a new knowledge-driven predictive discovery tool based on conserved molecular features in bacterial protective antigens

Based on the results described above, we formulated the hypothesis that the occurrence of the identified molecular features can be predictive of possible protective properties in other proteins from other species of interest. In order to test this hypothesis, predictive rules based on the occurrence of (i) selected MetaV Pfam domains, (ii) a protein architecture organized in multiple internal repeats and (iii) species- or genus-specificity of the corresponding genes were defined, and an *in silico* workflow for the genome wide selection of possible new protective antigens in bacterial genomes arranged as reported in Figure 12**.** The design of such approach is intended as a knowledge-driven predictive tool for the selection of possible new vaccine candidates in both gram positive and gram negative bacteria.

Etymologically, Meta**-** (from Greek: μετά = "after", "beyond", "with", "adjacent", "self"), is a prefix used in English (and other Greek-owing languages) to indicate a concept which is an abstraction from another concept, used to complete or add to the latter. We named our approach as MetaVaccinology, since the knowledge of the predictive rules were abstracted by a species-independent meta-analysis of a comprehensive from dataset of protective antigens and applied to predict new antigens.

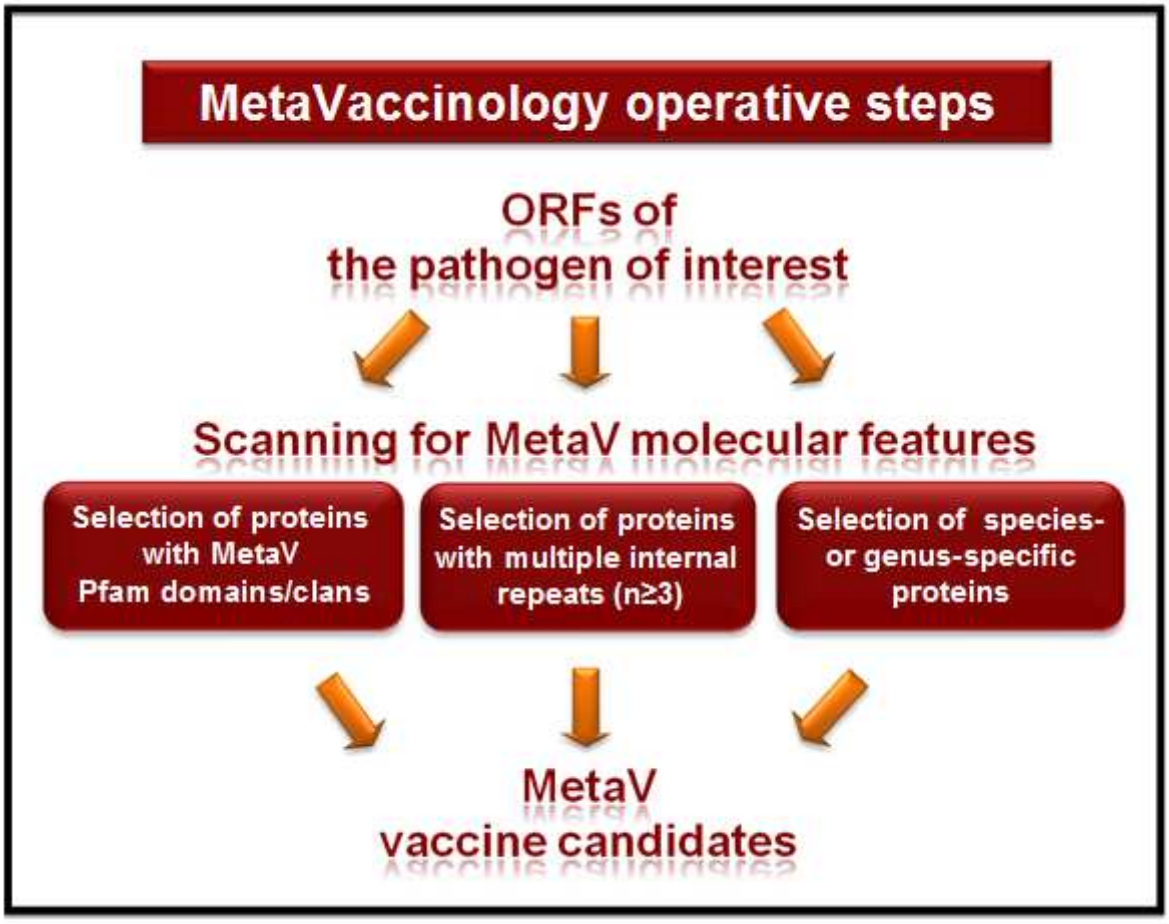


Figure 12: The operative steps of MetaVaccinology

3.4. MetaVaccinology (MetaV) applied to GBS

In order to provide an initial proof-of-concept of the MetaV selection approach, we applied our strategy to GBS. We used the available GBS complete genome of 2603V/R (serotype V), the corresponding protein sequence information from the UniProt knowledgebase was scanned for the occurrence of the MetaV molecular features. Occurrence of the all MetaV core Pfam domains were scanned, and 14 of those that are present in more than one protective antigen were used for selection. As a second step, we used Prospero (see M&M) to scan the

GBS genome for ORFs with more than two internal repeats. GBS specific ORFs were obtained from the genomic analysis carried out by Tettelin *et al.* (61). A total of 61 MetaV candidates were selected by this way (Table 8).

The list of sixty one selected proteins is shown. Proteins are grouped according to the core domains used to select them. For each protein the following information is reported: NCBI gene ID, protein annotation and predicted subcellular localization.

Table 8. The list of GBS MetaVaccinology candidates

No	CHAP Domain	Annotation	Localization
1	SAG1998	hypothetical protein	membrane
2	SAG1762	conserved hypothetical protein	outside
3	SAG1683	immunogenic secreted protein, putative	outside
4	SAG1286	conserved hypothetical protein	membrane
5	SAG0598	N-acetylmuramoyl-L-alanine amidase	unknown

6	SAG0017	PcsB protein	outside
	LysM Domain		
7	SAG2148	LysM domain protein	outside
8	SAG1386	conserved hypothetical protein	membrane
9	SAG0032	surface immunogenic protein	outside
	Cna_B		
10	SAG1408	cell wall surface anchor family protein	LPXTG
11	SAG1407	cell wall surface anchor family protein	LPXTG
12	SAG1404	cell wall surface anchor family protein	LPXTG

13	SAG0651	hypothetical protein	unknown
14	SAG0649	cell wall surface anchor family protein	LPXTG
15	SAG0646	cell wall surface anchor family protein	LPXTG
16	SAG0645	cell wall surface anchor family protein	LPXTG
	Peptidase_S8		
17	SAG2053	serine protease	LPXTG
18	SAG0676	serine protease, subtilase family	LPXTG
19	SAG0416	Protease	LPXTG
20	SAG1236	C5a peptidase	LPXTG
	5_nucleotid_C		

21	SAG1333	5'-nucleotidase family protein	LPXTG
22	SAG1941	2`,3`-cyclic-nucleotide 2`-phosphodiesterase	LPXTG
	Internal Repeats		
23	SAG1350	Surface antigen-related protein	Outside
24	SAG0433	surface protein Rib	LPXTG
25	SAK0517	c protein alpha-antigen precursor	Outside
26	SAG1283	cell wall surface anchor family protein	LPXTG
27	SAG1996	cell wall anchor protein-related protein	LPXTG
28	SAG0421	conserved hypothetical protein	LPXTG
29	SAK0186	iga fc receptor precursor	LPXTG

30	SAG2063	pathogenicity protein, putative	LPXTG
31	SAG1331	R5 protein	LPXTG
	Strep_his_triad		
	SAG1233	streptococcal histidine triad family protein	Outside
32	SAG0907	streptococcal histidine triad family protein	Lipoprotein
	FbpA		
33	SAG1190	adherence and virulence protein A	unknown
	Bmp Domain		
34	SAG0954	lipoprotein	Lipoprotein
35	SAG0405	lipoprotein	Lipoprotein

	WXG100		
36	SAG1039	conserved hypothetical protein	Cytoplasm
37	SAG0230	conserved hypothetical protein	Cytoplasm
	SdrG_C_C		
38	SAG1462	cell wall surface anchor family protein	LPXTG
	NLPC_P60.		
39	SAG0926	Tn916, NLP/P60 family protein	Outside
	Band_7		
40	SAG0132	SPFH domain/Band 7 family protein	Outside
	Lipoprotein_9		

41	SAG0776	YaeC family protein	Lipoprotein
42	SAG0971	Putative uncharacterized protein	Lipoprotein
43	SAG1641	YaeC family protein	Outside
	Gly_transf_sug		
44	SAG1167	Polysaccharide biosynthesis protein CpsM(V)	cytoplasm
	***Streptococci*-specific**		
45	SAG0265	conserved hypothetical protein	outside
46	SAG0371	hypothetical protein	outside
47	SAG0771	cell wall surface anchor family protein	LPXTG
48	SAG0833	hypothetical protein	outside

49	SAG0973	nisin-resistance protein, putative	membrane
50	SAG1037	hypothetical protein	membrane
51	SAG1127	conserved hypothetical protein	outside
52	SAG1419	hypothetical protein	lipoprotein
53	SAG1491	hypothetical protein	outside
54	SAG1473	hypothetical protein	outside
55	SAG1745	hypothetical protein	outside
56	SAG2056	chromosome assembly-related protein	outside
57	SAG2121	hypothetical protein	outside
58	SAG2021	Surface antigen protein	LPXTG

59	SAG1197	Hyaluronate lyase	outside
60	SAG0392	Similar to fibrinogen-binding protein	LPXTG
61	SAG1589	Amino acid permease	membrane

We focused our attention on a representative set of 9 proteins out of 61 MetaV candidates, in particular, (i) two extracellular enzymes, *i.e.* one peptidase S8 and one 5' nucleotidase, (ii) a basic membrane protein, (iii) two proteins with multiple internal repeats, (iv) a histidine triad protein, and (v) three proteins with generic cell-wall binding domains, *i.e.* two LysM and a CHAP domain-containing proteins (Table 9).

Table 9: The MetaV antigens of GBS selected to be tested in animal model.

No	2603 V/R	Annotation	Pfam Domains
1	SAG_0416	protease, putative	PF00082. Peptidase_S8
2	SAG_1350	surface antigen-related protein	Internal Repeats
3	SAG_1333	5'-nucleotidase family protein	PF02872. 5_nucleotid_C
4	SAG_0771	cell wall surface anchor family protein	Internal Repeats
5	SAG_0954	Lipoprotein	PF02608.Bmp
6	SAG_0907	streptococcal histidine triad family protein	PF04270. Strep_his_triad
7	SAG_0017	PcsB protein	PF05257. CHAP domain
8	SAG_1386	conserved hypothetical protein	PF01476. LysM
9	SAG_2148	LysM domain protein	PF01476. LysM

3.5 New Protective Antigens are identified by MetaVaccinology against GBS

The corresponding gene sequences from the 2603V/R strain were cloned in *E. coli* and nine proteins were successfully expressed and purified either as full-length or single domains (see Materials and Methods). In order to assess their efficacy in term of elicited protective immunity against natural infection, we performed *in vivo* experiments in a maternal immunization/neonatal pup challenge mouse model using GBS COH1 and 515 as challenge strains. Female mice received three doses (days 1, 21, 35) of either 20 µg antigen or PBS combined with Freund's adjuvant. Mice were then mated, and the resulting offspring challenged with a dose of GBS calculated to kill 80 to 90% of the pups. Survival of pups was monitored at day 2 after the challenge and protection values were calculated as [(% dead in control – % dead in vaccine)/% dead in control] x 100. Among the nine antigens tested in the animal model, four were able to induce a significant level of protection as compared to the control group, *i.e.* the 5' nucleotidase SAG1333 (67%), the histidine-triad protein SAG0907 (50%), the Bmp protein SAG0954 (41%) and the LysM domain-containing protein SAG1386 (33%) (Table 10).

Table 10. Protective activity of the selected GBS proteins compared with PBS as negative control. Protection was calculated as [(% $dead_{Control}$ - % $dead_{Vaccine}$) / % $dead_{Control}$] x100.

(GBS-01) GBS 515 as challenge strain	Survival	Survival %	Protective Efficacy
Name of the protein	n° animals		
SAG_1333	63 of 89	70	67
SAG_0907	29 of 66	44	38
SAG_0954	33 of 50	66	41
SAG_1386	9 of 20	45	28
PBS	7 of 77	9	*

(GBS-02) GBS COH1 as challenge strain	Survival	Survival %	Protective Efficacy
Name of the protein	n° animals		
SAG_0907	23 of 39	59	50
SAG_1386	33 of 50	66	41
PBS	7 of 40	18	*

3.5.1 SAG_1333

Please look at chapter about functional characterization of SAG_1333

3.5.2 SAG_0907

SAG_0907 is a histidine triad motif containing protein like PhtA, PhtB, or PhtD that are protective cell surface-exposed pneumococcal proteins (113). Humans produce antibodies to Pht proteins upon exposure to pneumococcus, and immunization of mice has provided protective immunity against sepsis and pneumonia and reduced nasopharyngeal colonization. The function of these proteins is not identified yet but there is an on-going discussion on their immune evasion and Factor H binding activity (122, 123). *S. pyogenes* have 2 different virulence factors that contain this motif; HtpA and Spy1361. HtpA-immunized mice

survived after challenge with GAS strains (isolated from patients) for significantly longer periods than sham-immunized mice and Spy 1361 is recently suggested as a vaccine candidate and hypothesized that it might play a role in GAS intracellular invasion (124). All these proteins are characterized by having an histidine (HxxHxH) motif that usually occurs multiple times.

3.5.3 SAG_0954

SAG0954 has a Bmp domain as Bmp proteins of *Borrelia burgdorferi*. BmpA of *B. burgdorferi* plays a significant role in mammalian infection by the Lyme disease spirochete and is an important antigen for the serodiagnosis of human infection. It is reported that BmpA-directed antibodies significantly inhibited the adherence of live *B. burgdorferi* to laminin (125) and *B. burgdorferi* lacking bmpA/B were unable to persist in the joints and failed to induce severe arthritis (126). Another Bmp domain containing protein P48 of *M. agalactiae* is described as an invariable, constantly expressed, immunodominant, surface lipoprotein. A recent report showed that genetic immunization with the immunodominant antigen P48 of *M. agalactiae* stimulates a mixed adaptive immune response in BALBc mice (127,128).

3.5.4 SAG_1386

SAG_1386 is a small LysM domain containing protein. Our MetaV results revealed that LysM (Lysin Motif) domain is one of the most recurring Pfam domain within the protective antigens: P60 from *L. monocytogenes*, intimin from *E. coli* , Ebps from *S. aureus* etc.

Bacteria retain certain proteins at their cell envelopes by attaching them in a non-covalent manner to peptidoglycan, using specific protein domains, such as the prominent LysM domain. LysM containing proteins are composed of not only truly secreted proteins, but also (outer-) membrane proteins, lipoproteins or proteins bound to the cell wall in a (non-)covalent manner. The motif typically ranges in length from 44 to 65 amino acid residues and binds to various types of peptidoglycan and chitin, most likely recognizing the N-acetylglucosamine moiety (129).

3.6 The MetaV approach is able to catch distantly related protective antigens

Figure 13 reports the domain organization of the 4 newly identified protective antigens, together with the corresponding %ID calculated from the primary sequence alignment to other known protective antigens belonging to the same MetaV class derived from other bacterial species, as reported in Table 7. This comparison clearly shows that sequence alignment would not be enough to catch the relationship between these distantly related proteins that share, on the other hand, the same domain composition and, in some cases, also the same domain organization at the whole protein architecture level. In the present study, we show that they also shared conserved protective properties in different bacterial species, suggesting that these proteins can be involved in basic mechanisms of bacterial pathogenesis common to distant human pathogens (look at SAG_1333 functional characterization).

(A) **(B)**

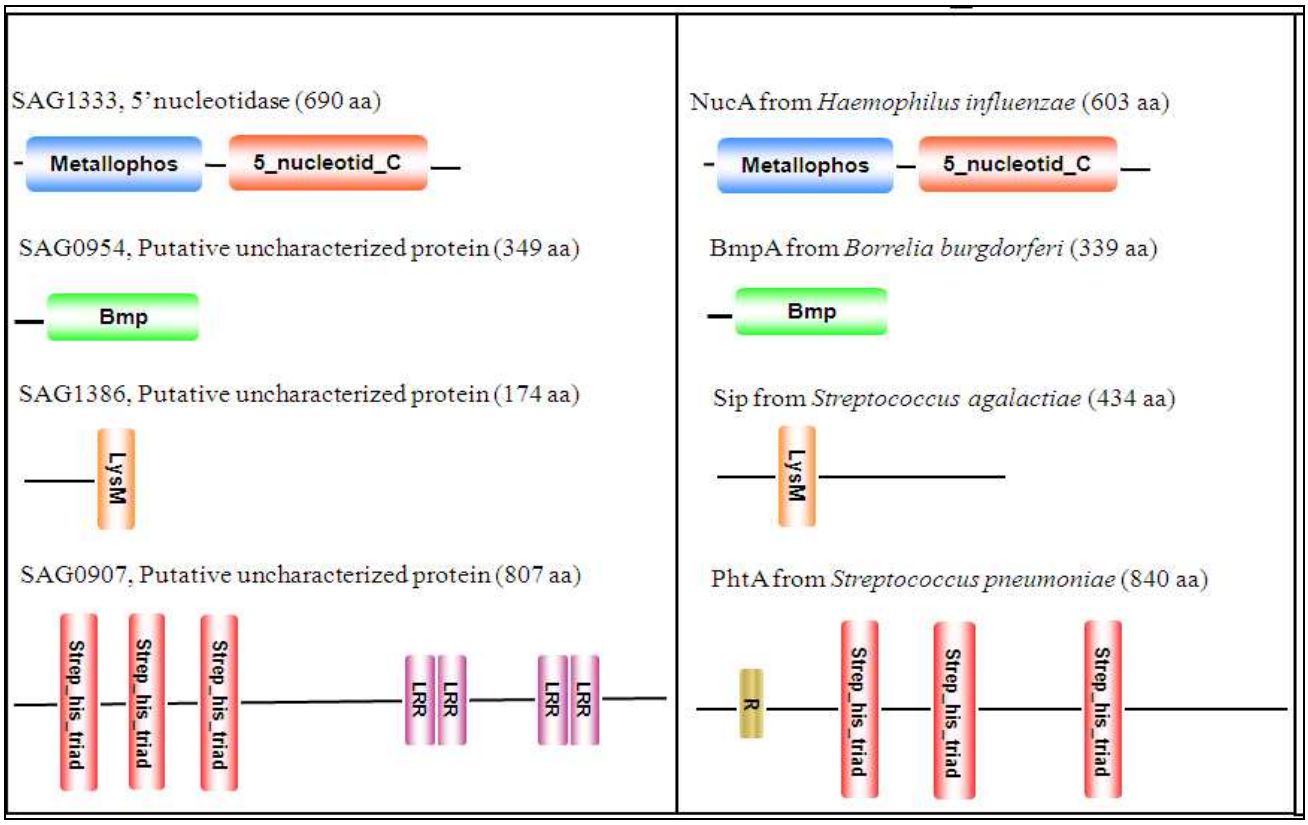


Figure 13. Schematic representation of domain architectures of the four new protective antigens identified in this study (A) and of other known protective antigens belonging to the same MetaV class derived from other bacterial species (B). Locations of domains in panels A and B are according to Pfam and SMART databases. Domains are drawn to approximate scale.

3.7 MetaVaccinology as a Basic Science Tool for GBS

3.7.1 Expression Analysis of New Protective Antigens on GBS Surfome

In order to prove expression of these four new protective antigens on the surface compartment of GBS, we used the surface digestion method as described previously (38, 40). Surface digestion is a new procedure that allows the rapid and selective identification of bacterial surface-exposed proteins, the pool of proteins which are entirely or partially exposed on the surface of the bacteria. The method uses proteolytic enzymes (trypsin and proteinase K) to 'shave' the bacterial surface and the peptides generated are separated from the whole cells and identified by mass spectrometry.

8 different GBS strains with a complete genome sequence, which represent the most important disease-causing serotypes: Ia (515 and A909), Ib (H36B), II (18RS21), III (COH1 and NEM316), and V (2603 and CJB111) were used for these analyses. Very briefly, the exponentially growing live bacterial cells were collected and treated with trypsin and/or PK, in order to shave the bacterial surface and generate peptides from protein domains exposed on the extracellular space. Peptides released into the supernatant were analyzed by nanoLC/MS/MS followed by database search against the whole GBS proteomes.

The surface proteome of the analyzed GBS strains consisted of 47 proteins, most of them (>90%) *in silico* predicted as extracellular or surface-associated. Table 11 shows the identified number of peptides corresponding to the protective antigens. This result indicates that SAG_1333, SAG_0954, SAG_0907 and SAG_1386 are expressed on the surface and

readily accessible to extracellular protease action and thus exposed on *S.aureus* surface, where it is expected to be accessible to circulating antibodies as well.

Table 11: The list of four new protective proteins identified on the surface of different GBS strains. For each protein the following information is reported: NCBI gene ID, protein annotation, predicted localization and identified number of peptides for corresponding protein.

GENE ID		Annotation	Localization	GBS Strain and number of peptides identified on the surfome							
				2603 V/R	18RS21	515	A909	CJB111	COH1	H36B	NEM316
1	SAG_1333	5'-nucleotidase family protein	LPXTG	**8**		**1**			**6**		
2	SAG_0954	lipoprotein	Lipoprotein	**1**							
3	SAG_0907	streptococcal histidine triad family protein	Lipoprotein			**2**					
4	SAG_1386	conserved hypothetical protein	Membrane						**1**		

3.7.2 SAG_1333 Functional Characterization: An Immune Evasion Protein?

The protectivity of 5'nucleotidases were previously shown against *H. influenzae* and Group A *Streptococcus*. In the case of GBS, the 5'-nucleotidase family protein was identified as an immunoreactive extracellular protein (131) and very interestingly, its transcription was 8.1 fold upregulated in human amniotic fluid (132). This is very significant since the neonatal GBS infection can result from fetal aspiration or ingestion of the infected amniotic fluid. Very recently, Schneewind *et al.* (2010) reported that 5'nucleotidase of *S. aureus* (adsA) is a critical virulence factor which the synthesis of adenosine in blood, escape from phagocytic clearance, and subsequent formation of organ abscesses are all dependent on. Moreover, the AdsA homologue of *Bacillus anthracis* enabled escape from phagocytic clearance thus they suggest the bacterial pathogens may exploit the immunomodulatory characteristic of adenosine to escape host immune responses (133). All these 5'nucleotidases from different pathogens have the same domain organization but low sequence identity that is very well-matched with our MetaV hypothesis (Figure 14).

The functional characterization studies on the secreted ATP utilizing enzymes of *Vibrio cholerae and Burkholderia cepacia* showed that 5'-nucleotidases play an important role in allowing pathogens to evade host defence (134, 135) It is also reported that the growth medium of *V. cholerae* fractions, harboring 5' nucleotidase, Ndk, and presumably other ATP-utilizing enzymes are causing enhanced macrophage and mast cell death by activating P2Z receptors. **(**134**)** Additionally, the level of secretion of the 5-nucleotidase *B. cepacia* was reported as lower in the environmental strains than in the clinical strains that show the

importance of the protein for virulence (135).

Vasu Punj *et. al* stated that 5'-nucleotidase of *V. cholerae* dephosphorylates the 5'-terminal phosphates from nucleoside phosphates such as AMP, ADP, or ATP and different ionic forms of ATP and adenine nucleotides have differential agonistic activities towards P2Z receptor activation of macrophages (136, 137). Secretion of 5'-nucleotidase by *V. cholerae* VB1 cells that can generate adenosine, AMP, and ADP from ATP can modulate macrophage cell death through multiple mechanisms. Indeed, it has been reported that a continuous generation of adenosine within the human epidermoid carcinoma cells can lead to an intracellular nucleotide imbalance with pyrimidine starvation, triggering suicidal processes ending up in apoptosis of the cells (138). In a recent study on *T. spiralis* larvae, it has shown that 5-nucleotidase enzyme converts substrate specifically AMP to adenosine. (139)

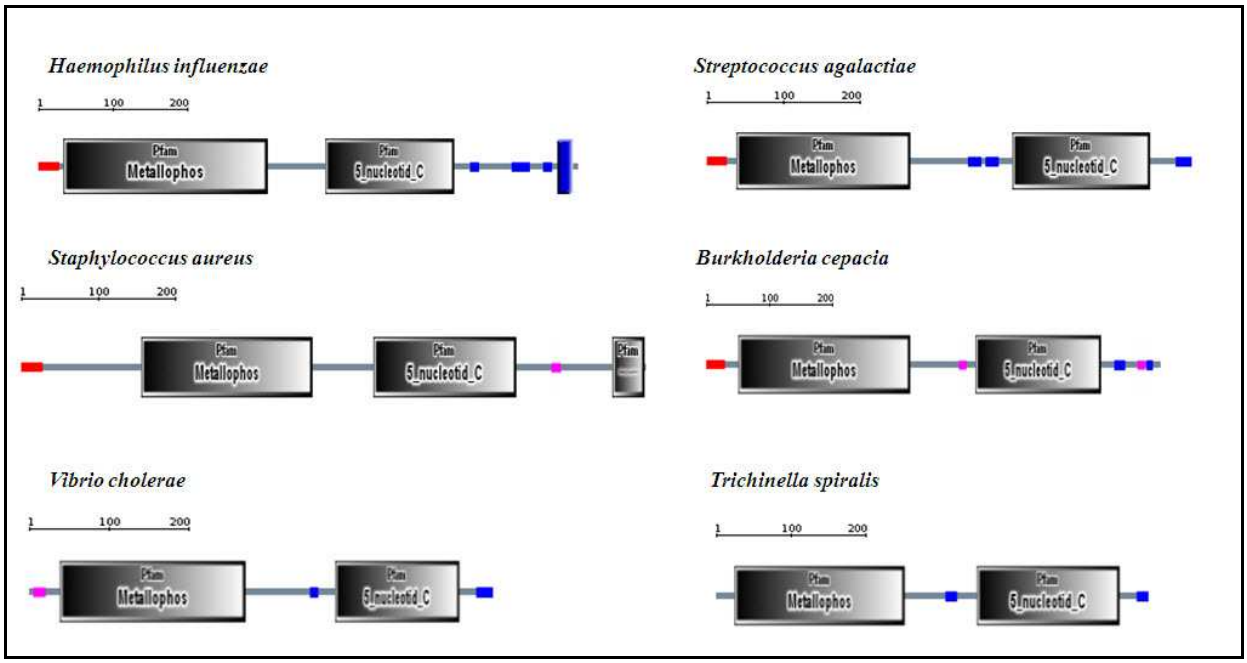


Figure 14. The protein architectures of different 5' nucleotidase proteins from different pathogens.

Adenosine is a purine nucleoside, which is produced also inside the human body under metabolic stress like hypoxic conditions, acute or chronic inflammatory tissue insults. The synthesis of adenosine involves the catabolism of adenine nucleotides (ATP, ADP and AMP) by the action of extracellular ectonucleotidases i.e. CD39 or nucleoside triphosphate dephosphorylase (NTPD) and CD73 or 5′-ectonucleotidase. Once adenosine is released in the extracellular environment, it binds to different types of adenosine (i.e. adenosine A1, A2A, A2B and A3 receptors) receptors expressed on various innate immune cells [Neutrophils, macrophages, mast cells, dendritic cells and natural killer cells]. Thus, depending on the type of adenosine receptor to which it binds, adenosine modulates innate immune response

during various inflammatory conditions [i.e. chronic (cancer, asthma) as well as acute (sepsis, acute lung injury) inflammatory diseases]. Besides expressing various other receptors, macrophages also express all the four types of adenosine receptors (i.e. adenosine A1, A2A, A2B and A3 receptors) (140).

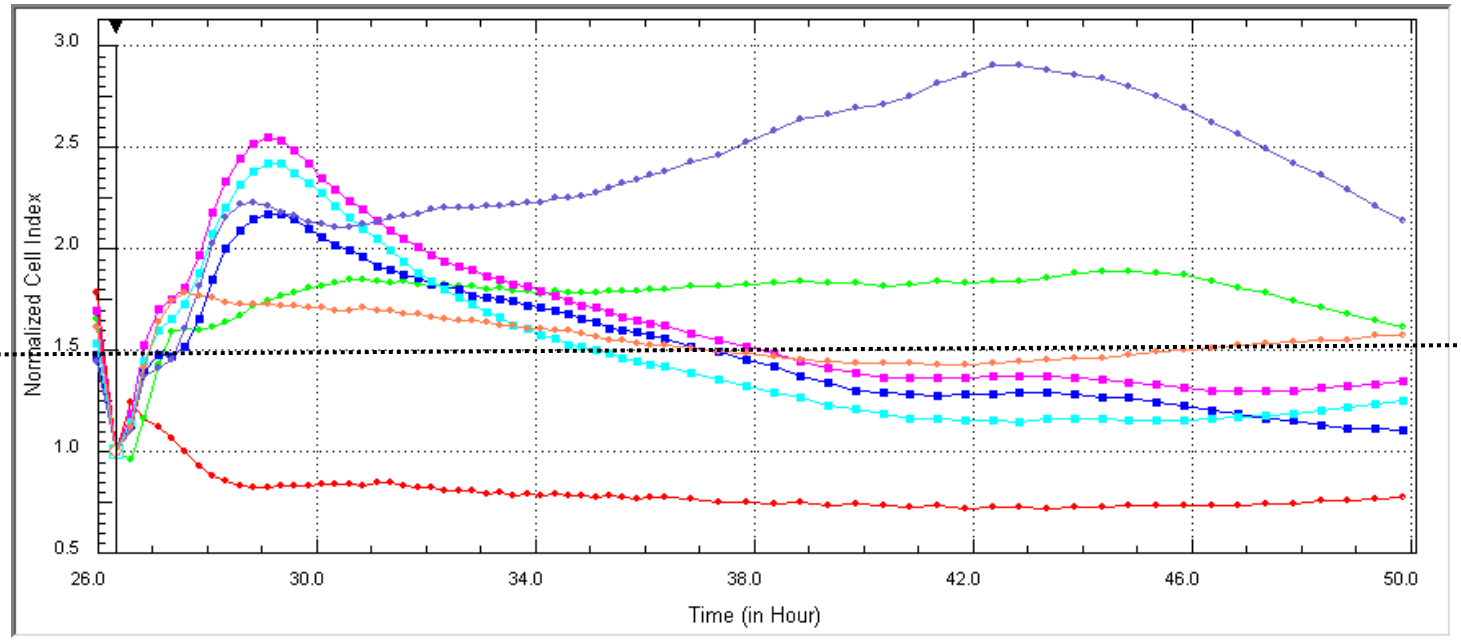
Exogenous adenosine prevents differentiation of monocyte into macrophages and blocks monocyte development at a stage, which resembles phenotypically to dendritic cells. The binding of adenosine to adenosine A1 receptors on monocytes promotes transformation of these cells into multinucleated giant cells but binding of adenosine to adenosine A2 receptors prevents generation of these giant cells. Besides affecting the maturation of monocytes into macrophages, adenosine also suppresses the phagocytic function of macrophages by binding to the adenosine A2 receptors expressed on monocytes or macrophages (140).

In the view of all these knowledge about 5'nucleotidases and adenosine, we constructed a hypothesis about the function of SAG_1333 and decided to test it. The hypothesis was assuming the transformation of AMP to adenosine by SAG_1333 enzymatic activity and the direct or indirect effects of adenosine on macrophages.

In order to test this, we used xCELLigence System from Roche that monitors cellular events in real time without the incorporation. The impedance measurement provides quantitative information about the biological status of the cells, including cell number, viability, and morphology.

The results of our experiments show that SAG_1333 protein or AMP alone has no toxic effect on macrophages. On the other hand, in the presence of AMP (5mM), SAG_1333 is toxic for macrophages and causing significant cell death (Figure 15). The killing effect of SAG_1333 is slower than the pore forming toxin (positive control of the experiment); this data is overlapping with our data since the effect is driven by a product of an enzymatic reaction. Very interestingly, further experiments on epithelial cells showed that AMP+SAG_1333 have no toxic effect on these cells. This observation is quite remarkable that the effect could be immune cells or macrophages specific. in addition, the N-terminal of the SAG_1333 (Metallophos domain) has also no effect on the macrophages either presence or absence of the AMP that shows the function of the SAG_1333 is dependent to C terminal (PF02872, 5_nucleotid_C). The function of 5_nucleotid_C domain is predicted to catalyze the hydrolysis of phosphate and this is matching with the hypothesized function of SAG_1333 that is converting AMP to adenosine and the phenotype observed for macrophages is a consequence of this adenosine production.

Figure 15I shows the effect of SAG_1333 on raw cells. It is observed that in the 5th hour after the addition of SAG_1333+AMP, the number of the macrophages starts to reduce while AMP and PBS alone has no effect. Figure 15II shows the results of the experiment based on carcinomic human alveolar basal epithelial cells (A549) and no effect observed. In the third experiment we used N-terminal of SAG_1333 and we did not observe any toxic effect on raw cells (Figure 15III).



PBS

Pore forming toxin 20 µg/mL

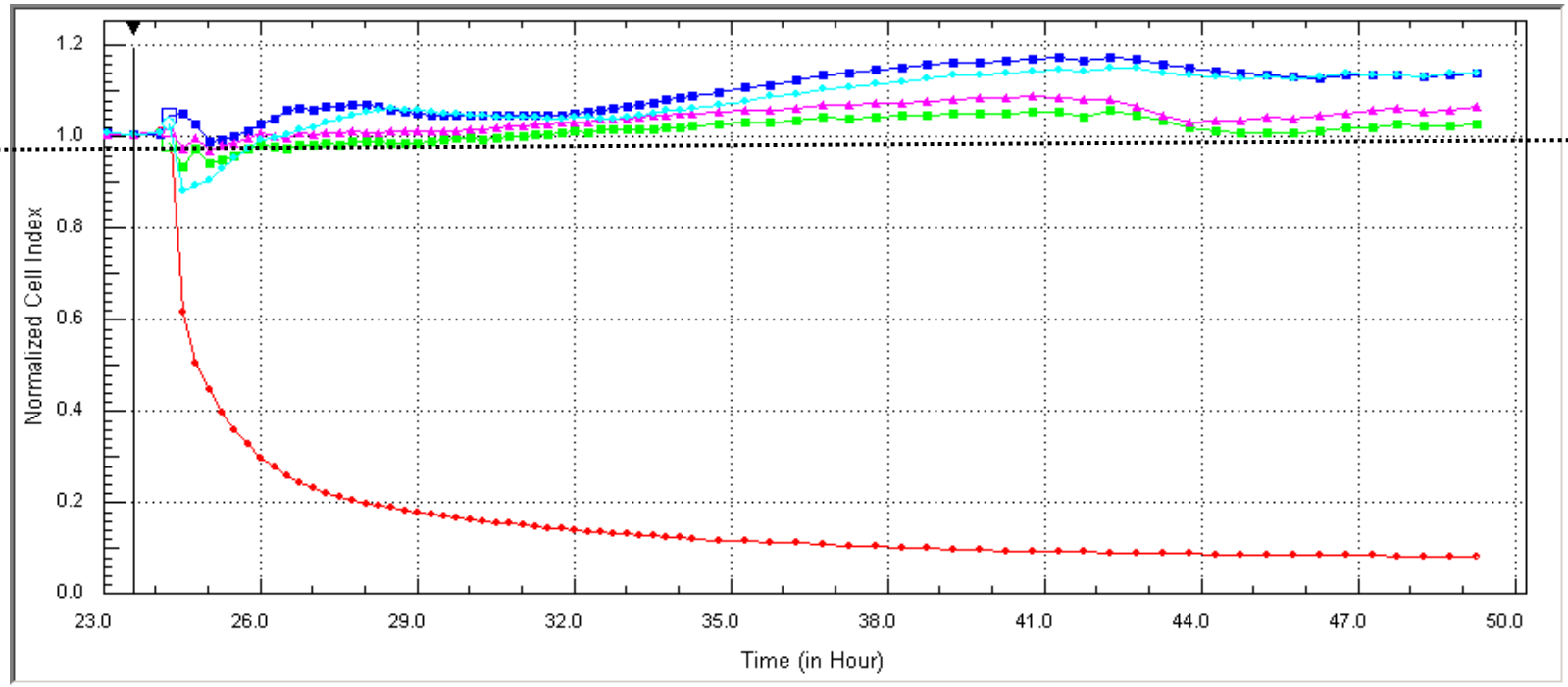
AMP 5 mM

SAG_1333 100 µg/mL

SAG_1333 100 µg/mL + 5mM AMP

SAG_1333 50 µg/mL + 5mM AMP

SAG 1333 25 µg/mL + 5 mM AMP



PBS

Pore forming toxin 20 µg/mL

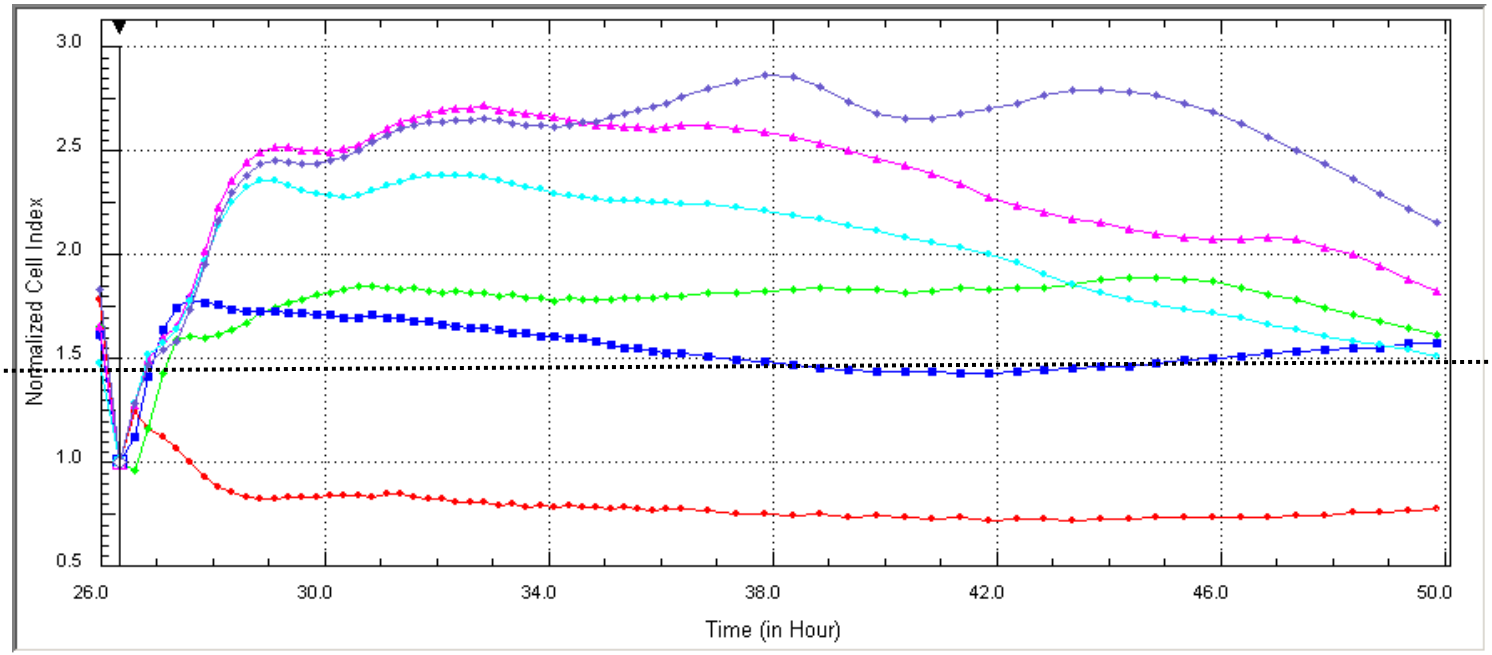
AMP 5 mM

SAG_1333 100 µg/mL

SAG_1333 100 µg/mL + 5mM AMP

SAG_1333 50 µg/mL + 5mM AMP

SAG 1333 25 µg/mL + 5 mM AMP



PBS

Pore forming toxin 20 µg/ml

AMP 5 mM

N-term of SAG_1333 100 µg/ml

N-term of SAG_1333 100 µg/ml + 5 mM AMP

N-term of SAG_1333 50 µg/ml + 5 mM AMP

Figure 15: xCELLigence System is used to monitor the effect of SAG_1333, SAG_1333+AMP, AMP alone and the controls on raw cells (I-first,III-third) and epithelial cells (II-second). Every color corresponds to a component added to cells and the figures show the effect on the cell numbers by time. The dashed line shows the boundary of the "negative controls".

3.7.3 Discussion on the SAG_1333 function

3.7.3.1 Why Adenosine is killing the Macrophages? A speculation: effect of Nitrix oxide?

The question about the mechanism of adenosine action on macrophages remains unanswered but there are some data in the literature that we could speculate on. It is shown that adenosine and its receptor agonists enhanced the production of nitric oxide (NO) in lipopolysaccharide (LPS)-treated RAW 264.7 cells. The results of this study indicate that the enhancement effects of adenosine on NO production in macrophages could be mediated by the extracellular adenosine receptors as well as the downstream metabolites of adenosine. The ubiquitous free radical, NO plays an important role in many biological processes including the regulation of the inflammatory response. High concentrations of either exogenous or endogenous inductible **nitric oxide** synthase (iNOS) derived NO have been shown to induce apoptosis in murine macrophage cell lines. However, despite the apparent reduced capacity of human macrophages in comparison to murine macrophages, to generate iNOS derived NO, human macrophages do undergo apoptosis in response to exogenous NO (141). The causes of macrophage apoptosis in response to nitric oxide are known to be DNA damage and mitochondrial stress which up-regulate p53, release cytochrome c, and activate caspases (142).

A study confirms the importance of NO for GBS mediated macrophage apoptosis. GBS infection was studied in murine macrophage-like J774A.1 cells and gene expression was analyzed before apoptosis. Inhibition of iNOS gene expression by use of N(G)-monomethyl-L-arginine (NMMA) inhibited apoptosis, whereas inhibition of TNF-alpha and IL-1 biological activity did not. Macrophages from congenic iNOS-deficient mice were less

susceptible to apoptosis than were macrophages from C57BL/6 mice. These results show that NO is an important mediator of GBS-induced murine macrophage apoptosis but does not contribute to antimicrobial activity or cytotoxicity in human monocyte-derived macrophages (143).

Moreover, Marriott HM *et al.* also reported similar results on pneumococcal-associated macrophage apoptosis. They showed that NO-mediated macrophage apoptosis during pneumococcal infection involves mitochondrial membrane permeabilization and in the presence of inducible nitric oxide synthase (iNOS) inhibition pneumococcal infection results in macrophage necrosis. The study provides the evidence of the at lower concentrations, NO contributes to pneumococcal killing; at higher concentrations it facilitates MMP (mitochondrial membrane permeabilization)-mediated apoptosis (144). These findings implicate NO as an important factor in macrophage apoptosis during GBS and pneumococcal infection.

3.7.3.2 GBS and Macrophage Apoptosis

The ability of pathogens to promote apoptosis may be important for the initiation of infection, bacterial survival, and escape from the host immune response. In fact, because apoptosis occurs without the release of cellular components, it does not usually lead to inflammation. Therefore, apoptosis may be advantageous for the pathogen because it might avoid the triggering and recruitment of non specific host defense mechanisms. Furthermore, macrophage death could also contribute to delaying or hindering the development of a

specific immune response.

Fettucciari *et al.* demonstrated that serotype III GBS induces apoptosis in infected macrophages. Their investigation showed that down-regulation of β-hemolysin expression, by growth of GBS in glucose-supplemented media, inhibited macrophage apoptosis, which suggests that the surface-bound β-hemolysin of GBS is the bacterial factor responsible for stimulating apoptosis in infected macrophages. However, there has been no direct demonstration that GBS β-hemolysin plays a role in apoptosis of macrophages (145)

In a following study, Ulett *et al.* investigated the role that β-hemolysin plays in nonopsonic phagocytosis, intracellular survival, and apoptosis in infected macrophages. They compared the effect of phagocytosis of a β-hemolysin–deficient isogenic mutant of serotype III GBS with that of the wild-type β-hemolytic parental strain and the ability of these strains to survive in macrophages and induce host-cell apoptosis. Growth conditions in which glucose levels were high were used to inhibit β-hemolysin expression in GBS strains, and the effect on the induction of macrophage cell death was assessed. The results of this study demonstrated that apoptosis in serotype III GBS–infected macrophages does not depend on β-hemolysin per se but on a factor coregulated with β-hemolysin by glucose (146).

3.7.3.3 SAG_1333 may be an important factor for GBS mediated macrophage apoptosis

All these references from literature combined to our experimental data shows that SAG_1333, the 5'nucleotidase protein of GBS, may be the unknown factor that is responsible from

GBS mediated macrophage apoptosis. One more encouraging data is coming from a recently published article that shows the PH regulation of SAG_1333 (147). The inverse ratio between PH and glucose is well known; the high glucose corresponds to lower PH in the medium. This result indirectly shows that SAG_1333 is regulated by glucose as the factor pointed out by Ulett GC *et al.*

We need further microscopic experiments in order to understand the effect of the SAG_1333 on the macrophages. Additional experiments with different cell types could help to show the specificity of the effect to macrophages and or immune cells.

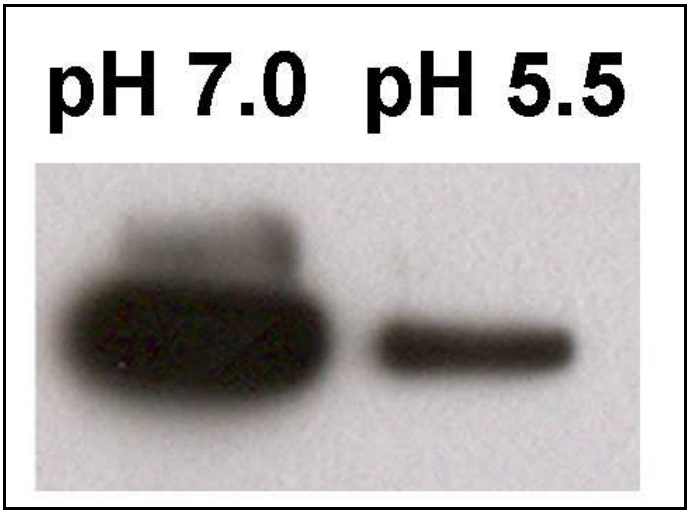


Figure 16. Immunoblot analysis on cell extracts of GBS grown at pH 7.0 or pH 5.5. Bacterial extracts were fractionated by SDS-PAGE, transferred to nitrocellulose, and the blots developed with mouse antisera specific for SAG1333 At pH 5.5 a reduction of SAG_1333 was observed (147)

3.8 LysM Domain and Protectivity

LysM domain is one of the most recurring MetaV domains within protective antigens. Figure 18 shows the proteins architectures of LysM domain containing protective antigens; intimin (eae) from *E. coli*, surface immunogenic protein from GBS, elastin binding protein from *S. aureus*, P60 from *L. monocytogenes* and lastly spr0096 from *S. pneumoniae*. The sequence identities within these protective antigens are 3-30 % and it is impossible to detect any homology by looking primary sequence. On the other hand, all these protective antigens carry a LysM domain. The 2 protective antigens that are identified during present study, SAG_1386 and SAOUHSC_00427 are also LysM domain containing proteins.

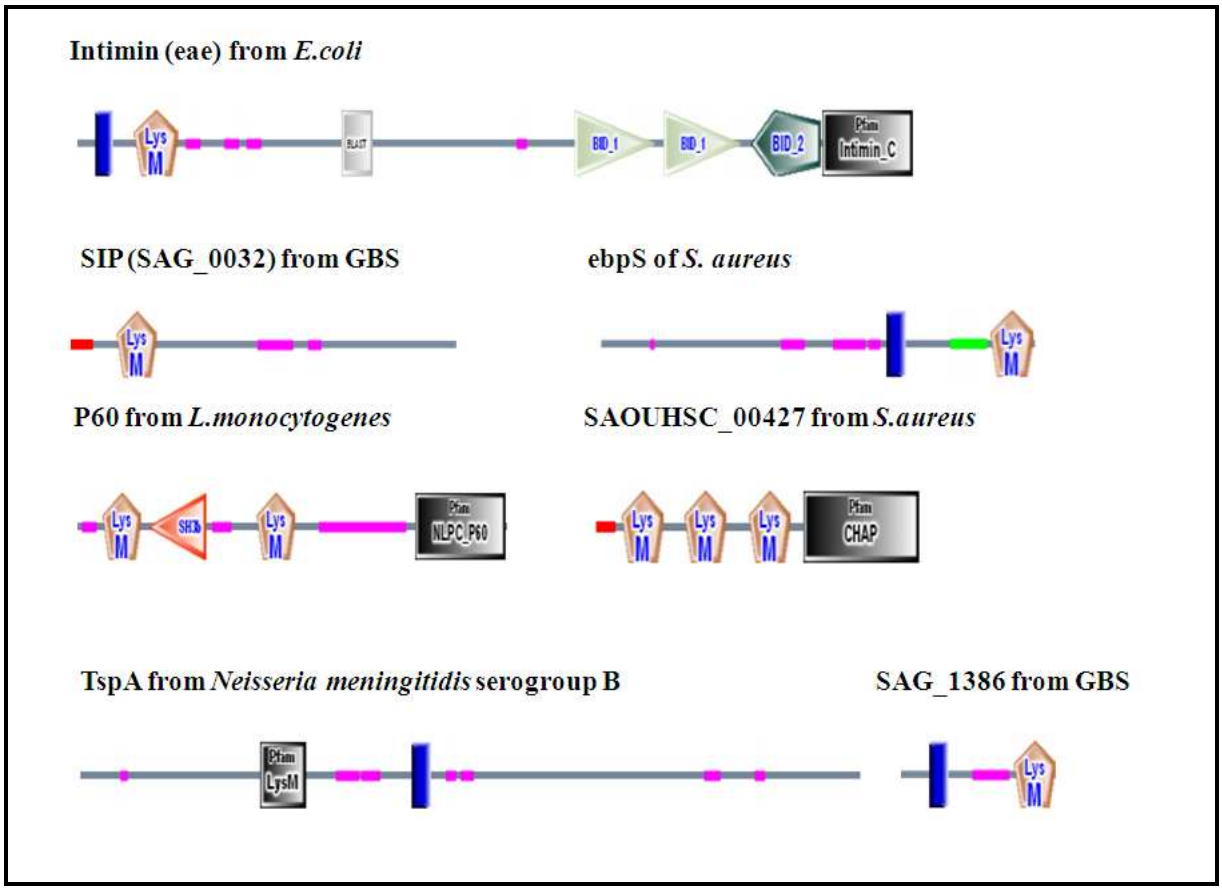


Figure 17: Protein architecture of different LysM domain containing protective antigens from different bacterial pathogens.

3.9 MetaVaccinology (MetaV) applied to *Staphylococcus aureus*

The results reported above for GBS are the first proof of the concept for MetaVaccinology as a new in-silico vaccine discovery tool. In order to show the efficiency of MetaVaccinology as a prediction instrument, we decided to repeat the experiments for *S. aureus*. The available complete genome of *S. aureus* strain NCTC 8325 is used, the corresponding protein sequence information from the UniProt knowledgebase was scanned for the occurrence of the MetaV

molecular features. PROSPERO were used to detect internal repeat containing proteins. 9 of all MetaV candidates are selected to be tested in animal model that are shown below (Table 12).

Table 12. The MetaV antigens of *S.aureus* selected to be tested in animal model.

No	ORF	Annotation	Pfam Domain
1	SAOUHSC_00400	conserved hypothetical protein	No Pfam domain
2	SAOUHSC_00256	Putative uncharacterized protein	PF05257. CHAP
3	SAOUHSC_00994	bifunctional autolysin precursor (Atl)	PF01510. Amidase_2 PF01832. Glucosaminidase
4	SAOUHSC_00671	Secretory antigen SsaA-like protein	PF05257. CHAP PF01476. LysM
5	SAOUHSC_00427	N-acetylmuramoyl-L-alanine amidase sle1	PF05257. CHAP PF01476. LysM
6	SAOUHSC_01949	Intracellular serine protease	PF00082. Peptidase_S8
7	SAOUHSC_00994	Bifunctional autolysin	PF01510. Amidase_2. PF01832. Glucosaminidase
8	SAOUHSC_02576	Secretory antigen SsaA	PF05257. CHAP
9	SAOUHSC_02979	Putative uncharacterized protein	PF05257. CHAP PF01832. Glucosaminidase.

3.9.1 Identification of 5 new protective antigens against *Staphylococcus aureus*

Among the MetaV candidates, we decided to test 9 of them in animal model. In order to assess their efficacy in term of elicited protective immunity against natural infection, we performed *in vivo* experiments in mouse model using Newman as challenge strain. Survival of mice were monitored for fifteen days and protection values were calculated as as [(% $dead_{Control}$ - % $dead_{Vaccine}$) / % $dead_{Control}$] x100. Among the nine antigens tested in the animal model, five were able to induce a significant level of protection as compared to the control group *i.e.* putative uncharacterized protein SAOUHSC_00256 (% 57), conserved hypothetical protein SAOUHSC_00400 ((% 49) for SA-01 schema and N-acetylmuramoyl-L-alanine amidase sle1, SAOUHSC_00427 (% 90), intracellular serine protease SAOUHSC_01949 (% 60), putative uncharacterized protein SAOUHSC_02979 (% 81) for SA-02 schema (Table 13).

3 of 5 newly identified protective antigens; SAOUHSC_00256 and SAOUHSC_02979 are carrying a CHAP domain as PcsB from *S. pneumoniae* while SAOUHSC_00427 is carrying both LysM and CHAP domains. SAOUHSC_01949 has Peptidase_S8 domain as C5a peptidases from *S. agalactiae*, Spy0416 from *S. pyogenes* ,prtA *from S. pneumonia* NMB1969 from *N. meningitidis* . SAOUHSC_00400 is a *S. aureus* specific protein that does not carry any defined Pfam domains.

Table 13. Protective activity of the selected *S. aureus* proteins compared with Alum as negative control. Protection values were calculated as as [(% dead$_{Control}$ - % dead$_{Vaccine}$) / % dead$_{Control}$] x100.

(SA-01) Name of the protein	Survival n° animals	Survival %	Protective Efficacy
SAOUHSC_00256	10 of 16	63	57
SAOUHSC_00671	1 of 16	6	0
SAOUHSC_00994	6 of 16	38	29
SAOUHSC_00400	9 of 16	56	49
Alum	2 of 16	13	*

(SA-02) Name of the protein	Survival n° animals	Survival %	Protective Efficacy
SAOUHSC_00427	15 of 16	94	90
SAOUHSC_01949	12 of 16	75	60
SAOUHSC_02979	14 of 16	88	81
SAOUHSC_02576	9 of 16	56	29
Alum	6 of 16	38	*

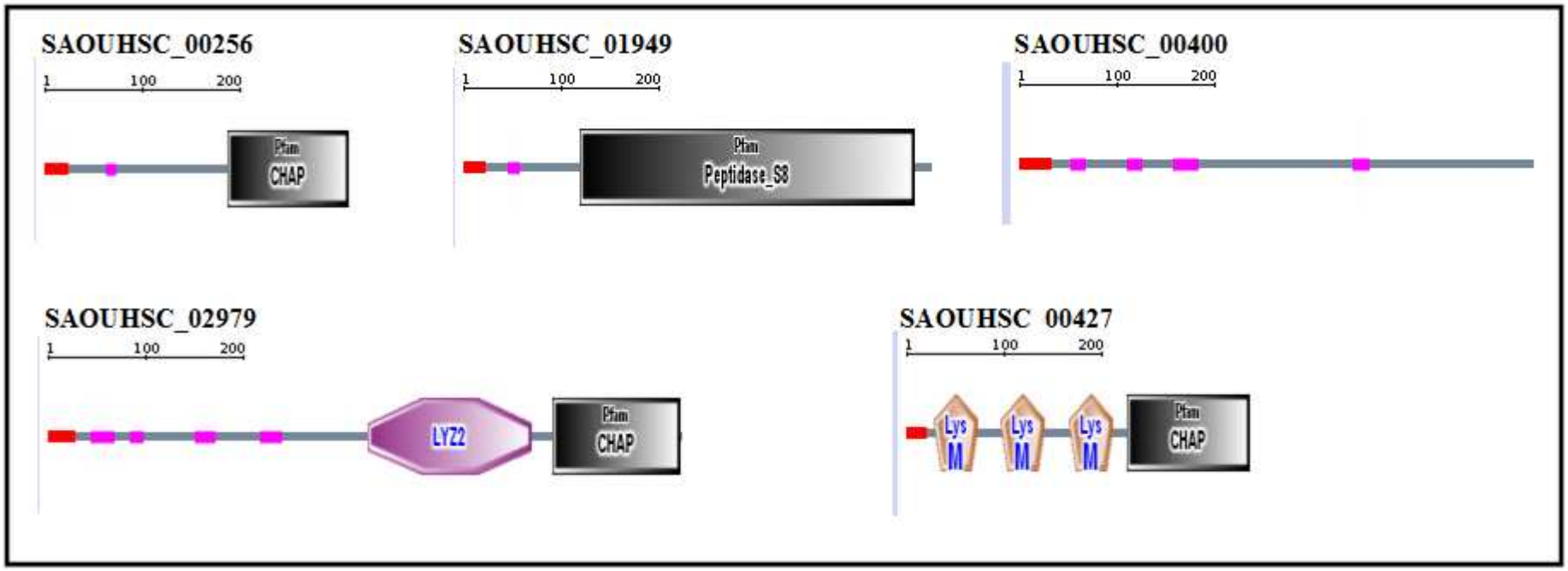


Figure 18. Protein architecture of new protective antigens against *S. aureus*

Very recently, Holtfreter *et. al* published a review of the current knowledge about antibody responses against *S. aureus* which challenges the adaptive immune system with abroad and highly variable antigen repertoire (148). They collected all the results of the different techniques to map the core and the variable *S. aureus* immuneproteomes, which aims to constitute the knowledge base for the design of effective *anti-S.aureus* vaccine compositions. All vaccination strategies against *S. aureus* rely on a central feature of adaptive immunity: immunological memory. The high susceptibility to *S. aureus* infection in patients with immunoglobulin deficiency and the wealth of escape mechanisms developed by *S. aureus* that allow it to interfere with antibody function are strong arguments in favor of an important role for adaptive immunity for this particular pathogen.

2 of 5 protective antigens that are identified by MetaV approach (i) SAOUHSC_00256 (staphyloxanthine biosynthesis protein, putative) and (ii) SAOUHSC_02979 (putative uncharacterized protein) are reported as immunogenic proteins only against sera collected from healthy individuals but not from carriers or patients. Moreover, SAOUHSC_00427 is reacted with sera both collected from healthy individuals and the patients. SAOUHSC_00400 and SAOUHSC_01949 are not identified any of these studies (148)

This knowledge is particularly interesting for vaccine design if the presence of immune response is really specific to healthy individuals. We could speculate that these antibodies are so protective that *S. aureus* can not colonize and/or infect these individuals who have antibodies against these antigens.

There is a need for functional characterization of the new protective antigens. Remarkably, 3 protective antigens carry (cysteine, histidine-dependent amidohydrolases/peptidases) a CHAP domain that is often found in association with other domains that cleave peptidoglycan. Cleavage of peptidoglycan plays an important role in bacterial cell division, cell growth and cell lysis. A review on CHAP domains published by Sanger Institute revealed that several known peptidoglycan amidases fall into CHAP family which includes two different peptidoglycan cleavage activities: L-muramoyl-L-alanine amidase and D-alanyl-glycyl endopeptidase activity. The family includes the amidase portion of the bifunctional glutathionylspermidine synthase/amidase enzyme from bacteria and pathogenic trypanosomes. The large number of multifunctional hydrolases suggests that they might act in a cooperative manner to cleave specialized substrates (149). A very recent systematic analysis of CHAP domain in 12 *S. aureus* genomes and 44 staphylococcal phage genomes revealed that there are 234 putative CHAP-containing proteins for *S. aureus* (150).

SAOUHSC_00427 (N-acetylmuramoyl-L-alanine amidase sle1) is the only functionally characterized protein within our new candidates. Kajimura *et al.* reported that this peptidoglycan hydrolase preferentially cleaved *N*-acetylmuramyl-L-Ala bonds in dimeric cross-bridges that interlink the two murein strands in the peptidoglycan. An insertion mutation of *sle1* impaired cell separation and induced *S. aureus* to form clusters suggesting Sle1 is involved in cell separation of *S. aureus*. The Sle1 mutant revealed a significant decrease in pathogenesis using an acute infection mouse model. Atl (SAOUHSC_00994) is the major autolysin of *S. aureus*, which has been implicated in cell separation of *S. aureus*. Generation of an *atl*/*sle1* double mutant revealed that the mutant cell separation was

heavily impaired suggesting that *S. aureus* uses two peptidoglycan hydrolases, Atl and Sle1, for cell separation. Unlike Atl, Sle1 was not directly involved in autolysis of *S. Aureus* (151). On the other hand, our results show that although both of these proteins have a similar and essential function for cell seperation, Atl is not a protective antigen against *S. aureus* (SA-01)

3.10 Comparison of MetaV with Other Vaccine Discovery Tools

3.10.1 Reverse Vaccinology and MetaV

Reverse Vaccinology and other genomics-based vaccine discovery tools commonly rely on the prediction of potential cell surface/secreted proteins for the identification of new vaccine candidates and predicted subcellular localization is available for almost all the known bacterial proteins through dedicated or general databases, *e.g.* the PSORTdb (http://db.psort.org/) and UniProtKB (http://www.uniprot.org/). Specific algorithms suitable for the *in silico* identification of novel surface-exposed and, thus, antibody accessible proteins mediating a protective response are used, mostly the signal peptide.

In order to understand the efficacy of signal peptide (SP) prediction, we analyzed the presence/absence of SP within all protective antigens that are shown in Table 6. A specific SP could not be detected for 39 out of 91 protective antigens and moreover final localization of 23 proteins are predicted as, cytoplasmic membrane for 5, unknown for 16 and periplasmic for 2 by PsortB. These 39 antigens without a signal peptide could be classified according to their function/structure; the flagellin proteins that are within the best vaccine candidates for many pathogens were totally out of the detection. As a second group, the main Clostridium toxins; tetanus toxin from *C. tetani*, botulinum toxin of *C. botulinum*, toxin A/B from *C.*

difficile, the pore forming toxins, hemolysin of *E.coli*, Listeriolysin O of *L. monocytogenes* and Ply of *S. pneumoniae* do not contain a signal peptide. In addition to those, SdrC, SdrD, SdrE proteins of *S. aureus* and and LysM domain containing proteins intimin from *E. coli*, Ebps from *S. aureus* and SAG_1386 from GBS do not have a predictable SP (Table 14)

This is particularly meaningful, since there is an increasing awareness that secretion pathways in bacteria are far more complex than expected, there are seven different secretion systems identified in Gram-negative and six others for Gram-positive bacteria, yet. The identification of WXG100 domain containing ESAT-6 (esxA) and CFP-10 (esxB) proteins that are the main candidates for an effective subunit vaccine against *M. tuberculosis* caused the recent discovery of a new secretory pathway for Gram + bacteria. Both of these proteins lack a distinguishable Sec-signal sequence, which suggests the existence of a specialized secretion pathway. Several independent studies have demonstrated that the genes that surround the ESAT‐6- and CFP‐10-encoding genes are involved in the production of such a specialized secretion system; type VII secretion (152).

The MetaV approach has the advantage of being independent from the prediction of potential cell surface/secreted proteins for the identification of possible new vaccine candidates, because it relies on the occurrence of specific molecular features, either functional or structural, over the protein whole length, regardless to the presence of a classical signal peptide for secretion.

The Reverse Vaccinology approach successfully applied to GBS. In total, 589 proteins are selected as surface exposed proteins and among thoses 312 were expressed as

recombinant proteins, purified and tested for protection against GBS. Four proteins were found to elicit protective immunity in an animal model. This is a very comprehensive screening of the possible candidates but labor intensive and expensive to be performed. On the other hand, MetaV, by using the extraction of the knowledge of RV approach, reduces the number of proteins to be tested and theoretically enriches the number of protective antigens within this list.

Table 14. The predicted localization of main protective antigens and the presence/absence of signal peptide.

	Organism	Annotation	Signal peptide	PsortB prediction
Flagellins	*Pseudomonas aeruginosa*	A-type flagellin	-	Extracellular
	Pseudomonas aeruginosa	B-type flagellin	-	Extracellular
	Escherichia coli (strain K12)	flagellin	-	Extracellular
	Campylobacter jejuni	Flagellin A	-	Extracellular
	Campylobacter jejuni	Flagellin B	-	Extracellular

	Helicobacter pylori	Flagellin A	-	Extracellular
	Helicobacter pylori	Flagellin B	-	Extracellular
Major outer membrane proteins	*Escherichia coli*	C3389	-	CytoplasmicMembrane
	H. Influenzae	P6	+	Outer membrane
	B. abortus	Omp16	-	Outer membrane
	P. aeruginosa	outer membrane porin F	+	Outer membrane
Pore-forming toxins	*S. pneumoniae*	Ply	-	Extracellular

	S. Pyogenes	streptolysin O	+	Extracellular
	L. monocytogenes	Listeriolysin O	-	Extracellular
	S. aureus	Alpha-toxin	+	Extracellular
	S. aureus	Panton-Valentine Leucocidin F	+	Extracellular
	E. coli	Hemolysin	-	Extracellular
Ribosyl transferases	*C. diphtheriae*	Difteriae toxin	+	Unknown
	B. pertussis	Pertussis toxin (subunit 1)	+	Extracellular

		Pertussis toxin (subunit 2)	+	Unknown
		Pertussis toxin (subunit 3)	+	Unknown
	V. cholerae	Cholera enterotoxin subunit A	+	Unknown
		Cholera enterotoxin subunit A	-	Unknown
Neurotoxins	*C. tetani*	Tetanus toxin	-	Extracellular
	C. botulinum	Botulinum toxin type A	-	Extracellular

		Botulinum toxin type B	-	Extracellular
Second messenger pathway activators	*C. difficile*	toxin A	-	Extracellular
		toxin B	-	CytoplasmicMembrane
Proteases	*S. agalactiae*	C5a peptidase	+	Cell wall
	S. pyogenes	Spy0416	+	Cell wall
	S. pneumonia	prtA	+	Cell wall
	N. meningitidis	NMB1969	+	Outer Membrane

	S. pyogenes	SpeB	+	Extracellular
	S. pyogenes	D-alanyl-D-alanine carboxypeptidase	+	Cytoplasmic Membrane
5'nucleotidases	*H. influenza*	NucA	+	Cell wall
	S. pyogenes	Spy0872	+	Cell wall
	S. agalactiae	SAG_1333	+	Cell wall
Heme binding proteins	*S.aureus*	IsdA	+	Cell wall
	S.aureus	IsdB	+	Cell wall

TonB dependent receptors	*E.coli*	IroN	-	OuterMembrane
	E.coli	FyuA	+	OuterMembrane
	E.coli	ChuA	+	OuterMembrane
	E.coli	Iha	+	OuterMembrane
	E.coli	IreA	-	OuterMembrane
	E.coli	IutA	+	OuterMembrane
	P. haemolytica	TbpA	+	OuterMembrane

	P. haemolytica	TbpB	-	OuterMembrane
Iron transport	*S. pneumonia*	PiaA	+	Unknown
	B. melitensis	P39	+	Periplasmic
	S. pneumoniae	potD	+	Unknown
	C. jejuni	CjaA	+	Periplasmic
	N. meningitidis	GNA1946	+	Cytoplasmic Membrane
Cna_B domain containing	*S. pyogenes*	Ancillary protein 1	-	Cell wall

proteins	*S. agalactiae*	SAG_0645	+	Cell wall
	S. pneumoniae	RrgA	+	Unknown
	S. aureus	Cna	+	Cell wall
	S. aureus	SdrD	-	Cell wall
	S. aureus	SdrE	-	Cell wall
Proteins containing variable numbers of internal repeats	*S. aureus*	FnbpA	+	Cell wall
	S. aureus	clumping factor A	+	Cell wall

	S. aureus	SdrC	-	Cell wall
	S. agalactiae	SAN_1485	-	Cell wall
	S. pyogenes	M protein	+	Cell wall
	S. agalactiae	LrrG	-	Cell wall
	S. agalactiae	Rib protein	+	Cell wall
	S. agalactiae	C protein alpha-antigen	-	Cell wall
	N. meningitidis (serogroup B)	Factor H binding protein	+	Unknown

Others	*N.meningitidis* (serogroup B)	Heparin binding protein GNA2132	+	Unknown
	H. influenzae	TbpB	-	OuterMembrane
	N. meningitidis	NadA	+	OuterMembrane
	M. catharralis	UspA1	+	OuterMembrane
	M. catharralis	UspA2	+	OuterMembrane
	Y. pestis	YadA	-	OuterMembrane
	Y. pestis	YadC	-	OuterMembrane

	E. coli	C4424	+	Unknown
	S. pneumoniae	PhtA	-	Unknown
	S. pneumoniae	PhtB	+	Unknown
	S. pneumoniae	SP_1174	+	Unknown
	S. agalactiae	SAG_0907	-	Cell wall
	B. burgdorferi	BmpA	-	Unknown
	S. agalactiae	SAG_0954	+	Unknown

	B. pertussis	Pertactin	+	OuterMembrane
	H. influenzae	Hap	-	Unknown
	N. meningitidis	NMB1998	-	OuterMembrane
LysM domain-containing proteins	*L. monocytogenes*	P60	+	Extracellular
	E.coli	Intimin	-	OuterMembrane
	S. aureus	Epbs	-	Cell wall
	S. agalactiae	SAG_1386	-	Cytoplasmic Membrane

CHAP domain-containing proteins	*S. pneumoniae*	PcsB	+	Extracellular
Choline-binding motifs	*S. pneumoniae*	PspA	+	Cell wall
	S. pneumoniae	PspC	+	Cell wall

3.10.2 MetaVaccinology and ANTIGENome technology

The expression of vaccine candidates during natural infection in humans is compulsory for subunit vaccines. In order to assess the capability of MetaVaccinology to predict immunogenic proteins that are recognized by human humoral immune system, we used the results of recently published article of Meinke *et al* (2010). Very briefly, ANTIGENome technology was used to identify new vaccine candidates by using a broad range of sera and cervical secretions obtained from either healthy or GBS colonized women against genomic surface display libraries. They identified 35 most frequently selected immunogenic proteins within 167 others (153).

We decided to analyze these 35 most immunogenic proteins; regarding to their coincidence with our MetaV selection. The protein architecture of all proteins are analyzed and the results show that among all immunogenic proteins; 27 (% 77) are covered by MetaV approach: (i) 9 proteins carry MetaV Pfam domains and (ii) 18 proteins are Streptococci-specific. The 8 proteins that are out of MetaV list are mostly cytoplasmic enzymes like ThiI/ Probable tRNA sulfurtransferase or NH3-dependent NAD+ synthetase. Table 15 shows the Meta analysis, annotation and domain composition of the immunogenic proteins identified by ANTIGENome. Two of four protective antigens that we identified by MetaV approach; gbs1403 (SAG_1333) and gbs0918 (SAG_0907) are present in the immunogenic proteins described in this paper, on the other hand SAG_1386 and SAG_0954 are not.

The same group used ANTIGENome technology in order to have a comprehensive *in vivo* antigenic profile of *Staphylococcus aureus* N315 strain. A total of 23 antigenic proteins

were reported as immunogenic proteins frequently identified by bacterial surface display, according to their reactivity with individual sera from patients and healthy individual (27). Among the nine MetaVaccinology antigens tested in the animal model, five were able to induce a significant level of protection as compared to the control group; *i.e.* SAOUHSC_00256 (SA0270), SAOUHSC_00400 (SA0394) SAOUHSC_00427 (SA0423), SAOUHSC_01949 (no homolog for *S. aureus N315* strain) and SAOUHSC_02979 (SA2437). None of these protective antigens are identified by the ANTIGENome technology as immunogenic, thus could be never identified as vaccine candidates by this approach.

As already mentioned before Silva Holtfreter *et. al* published a review concerning the current knowledge about antibody responses against *S. aureus* (148). 2 of 5 protective antigens that are identified by MetaV approach (i) SAOUHSC_00256 (staphyloxanthine biosynthesis protein, putative) and (ii) SAOUHSC_02979 (putative uncharacterized protein) are reported as immunogenic proteins only against sera collected from healthy individuals but not from carriers or patients. Moreover, SAOUHSC_00427 is reacted with sera both collected from healthy individuals and the patients. SAOUHSC_00400 and SAOUHSC_01949 are not identified any of these studies.

All these results prove the fact that MetaV could also predict the *in vivo* immunogenic antigens that give a significant advantage to any vaccine discovery tool. Moreover, the protective antigens identified by MetaV can not be identified by ANTIGENome.

Table 15. Meta-Analysis of immunogenic proteins identified by ANTIGENome technology against GBS

	Immunogenic proteins	**Annotation**	**Domains or Specificity**
No	**Proteins with MetaV domains**		
1	gbs2008	Similar to C5A peptidase	PF00082. Peptidase_S8.
2	gbs1306	Laminin-binding protein	PF04270. Strep_his_triad
3	gbs1805	Similar to secreted unknown protein	PF05257. CHAP domain
4	gbs1308	C5a peptidase, authentic frameshift	Peptidase_S8

5	gbs0031	Surface immunogenic protein	LysM domain
6	gbs1925	Similar to pneumococcal histidine triad protein B precursor	PF04270. Strep_his_triad
7	gbs0918	Weakly similar to histidine triad protein, putative lipoprotein	PF04270. Strep_his_triad
8	gbs0016	Glucan-binding protein B	PF05257. CHAP domain
9	gbs1403	Similar to 5'-nucleotidase, putative PG bound protein	PF02872. 5_nucleotid_C.
	Streptococci-specific proteins		
10	gbs2018	BibA	PF00746. Gram_pos_anchor

11	gbs1356	Putative PG linked protein	PF08363. GbpC PF00746. Gram_pos_anchor
12	gbs0983	Similar to plasmid protein	No Pfam domain
13	gbs2106	Protein of unknown function/lipoprotein, putative	PF06737. Transglycosylas
14	gbs0997	Hypothetical protein	PF01223. Endonuclease_NS
15	gbs0986	Surface antigen protein	PF00746. Gram_pos_anchor
16	gbs0995	Hypothetical protein	No Pfam domain
17	gbs0428	Similar to fibrinogen-binding protein	PF11966. SSURE PF04650. YSIRK_signal

18	gbs1565	Hypothetical protein	No Pfam domain
19	gbs0489	Acetyltransferase, GNAT family	PF00583. Acetyltransf_1
20	gbs1144	Unknown protein	PF00746. Gram_pos_anchor
21	gbs1790	Hypothetical protein	No Pfam domain
22	gbs1638	Amino acid permease	PF08363. GbpC PF00746. Gram_pos_anchor
23	gbs1879	Endopeptidase O (pepO)	PF01431. Peptidase_M13 PF05649. Peptidase_M13_N
24	gbs1242	CpsG, beta-1,4-galactosyltransferase	PF04101. Glyco_tran_28_C

25	gbs1143	Putative PG linked protein	No Pfam domain
26	gbs0456	Cell wall surface anchor family protein	No Pfam domain
27	gbs1270	Hyaluronate lyase	PF02278. Lyase_8. PF02884. Lyase_8_C. PF08124. Lyase_8_N.
	Proteins Not Selected with MetaV		
28	gbs1087	FbsA	PF08017. Fibrinogen_BP
29	gbs0931	Pyruvate kinase	PF00224. PK. PF02887. PK_C.
30	gbs1116	Xanthine permease (pbuX)	PF00860. Xan_ur_permease.

31	gbs0947	Similar to l-lactate dehydrogenase	PF02866. Ldh_1_C. PF00056. Ldh_1_N.
32	gbs1442	ThiI/ Probable tRNA sulfurtransferase	PF02568. ThiI PF02926. THUMP
33	gbs0286	NH3-dependent NAD+ synthetase	PF02540. NAD_synthase
34	gbs0437	Glucose-6-phosphate isomerase (pgi)	PF00342. PGI
35	gbs0235	Glycine betaine/carnitine/choline ABC transporter	PF00005. ABC_tran. PF00571. CBS.

3.10.3 MetaVaccinology and Surfome analysis

The first use and description of surfome the analysis as a vaccine discovery tool was by Rodriguez-Ortega *et al*. A relevant result of this work was the demonstration that comprehensive characterization of surface-exposed proteins can lead to new vaccine candidate discovery. Among the 14 identified surface proteins tested, one protein, Spy0416, conferred high protection levels (38). Secondly, this approach was used in order to analyze the surfome of Group B Streptococcus, COH1 strain and to identify new vaccine candidates. When applied to the Group B *Streptococcus* COH1 strain, 43 surface-associated proteins were identified, including all the protective antigens described in the literature as well as a new protective antigen, the cell wall-anchored protein SAN_1485 belonging to the serine-rich repeat protein family (40).

Of those two new protective antigens identified by surface digestion, Spy0416 is composed of DUF1034, PA (PF00082), Peptidase_S8 (PF00082) YSIRK_signal (PF04650) domains and SAN_1485 is composed of a SdrG_C_C (PF10425) domain. Peptidase_S8 and SdrG_C_C are two MetaV core Pfam domains that are identified in 5 (both Gram -/+) and 2 different pathogens respectively (table 6). Both of these proteins are also the potential MetaV candidates because of their domain organization. In contrary, the protective antigens identified by our study are not the potential surfome analysis candidates. Surface digestion methods uses the obtained peptide numbers as a selection criteria of potential vaccine candidates since it is assumed that surface exposition correlates with the number of peptides obtained. If we had used this criteria to select our proteins, by using our surfome analysis results, SAG_0954 (1 peptide), SAG_0907 (2 peptides) and SAG_1386 (1 peptide) would

never been selected (table 16). Moreover, SAG_1333 (the most protective of four antigens against GBS) could not be identified on the surfome of COH1 in our previous study (40); although the protein was reported to be highly immunogenic by using human sera that shows its *in vivo expression* (153).

We could also compare MetaV and surfome analysis for *S. aureus*. There is a very recently published article on *S. aureus* surfome thatVentura C. L. *et al.* analyzed the cell surface proteome of USA300 strain LAC. A total of 113 identified proteins were associated with the surface of USA300 during the late-exponential phase of growth *in vitro* (154). Even though, the work is called surfome analysis, a high percentage of the proteins are cytoplasmic that causes question marks about the quality of the surfome analysis. We compared the protective antigens that were identified by using MetaV in order to understand if they were detectable by this surfome analysis. Of those 5 protective antigens only N-acetylmuramoyl-L-alanine amidase sle1 SAOUHSC_00427 (SAUSA300_0438) was identified on the surface of USA_300 strain while SAOUHSC_00256 (SAUSA300_0651), SAOUHSC_00400 (SAUSA300_0408) SAOUHSC_00427 (SAUSA300_0438), SAOUHSC_01949 (SAUSA300_1763), SAOUHSC_02979 (SAUSA300_2579) could not.

The protective antigens must either well expressed on the surface or secreted ones. So why could not we show this well expression by using surfome analysis for our protective antigens? This could be due to diversity between *in vivo* and *in vitro* conditions that causes a significant difference on the expression of the proteins. Unlike the genome, proteome is a dynamic composition and the expression profile is directly effected by the external conditions.

Likewise, the *in vivo* proteome of a pathogen is also dynamic that could adapt immediately to new external conditions.

GBS is a major neonatal pathogen that is able to adapt to a variety of host environments, including both rectal and vaginal maternal carriage, growth in amniotic fluid and at various neonatal body sites. Transcriptomics studies of GBS could let us to explain its ability to adapt different environmental conditions. Moreover, these studies could help us to explain and understand, why we could not identify all these protective antigens as well exposed proteins on the surface of GBS by using *in vitro* conditions.

3.10.4 Transcriptomics: Another Genomics Tool to Understand Pathogenicity of GBS

James Musser and his colleagues published three different articles about the GBS transcriptome. In the first study, to enhance understanding of how GBS adapts during invasive infection, they performed a whole-genome transcriptome analysis of GBS after incubation with whole human blood. Global changes occurred in the GBS transcriptome rapidly in response to blood contact following shift from growth in a rich laboratory medium. The transcripts of relatively few proven virulence genes were up-regulated during the first 90 minutes. However, a key discovery was that genes encoding proteins involved in interaction with the host coagulation/fibrinolysis system and bacterial-host interactions were rapidly up-regulated. Extensive transcript changes also occurred for genes involved in carbohydrate metabolism, including multi-functional proteins and regulators putatively involved in pathogenesis. Additionally, they discovered that an incubation temperature closer to that occurring in patients with severe infection and high fever (40 degrees C) induced

additional differences in the GBS transcriptome relative to normal body temperature (37 degrees C) (155).

In a second study, to understand the response to temperatures encountered in the various hosts, they conducted a whole genome transcriptome analysis for organisms grown at 30 degrees C and 40 degrees C. They identified extensive transcriptome remodeling at various stages of growth, especially in the stationary phase (significant transcript changes occurred for 25% of the genes). A large proportion of genes involved in metabolism were up-regulated at 30 degrees ^{o}C in stationary phase. Conversely, genes up-regulated at 40 degrees ^{o}C relative to 30 degrees C include those encoding virulence factors such as hemolysins and extracellular secreted proteins with LPXTG motifs. Over-expression of hemolysins was linked to larger zones of hemolysis and enhanced hemolytic activity at 40 degrees ^{o}C. A key theme identified by this study was that genes involved in purine metabolism and iron acquisition were significantly up-regulated at 40 degrees C (156).

Thirdly, they used amniotic fluid to grow bacteria and characterized the transcriptome of GBS grown in human amniotic fluid (AF) comparing it with the transcriptome in rich laboratory medium. They discovered that GBS significantly remodels its transcriptome in response to exposure to human amniotic fluid. GBS grew rapidly in human AF and did not exhibit a global stress response. The majority of changes in GBS transcripts in AF compared to THY medium were related to genes mediating metabolism of amino acids, carbohydrates, and nucleotides. The majority of the observed changes in transcripts affects genes involved in basic bacterial metabolism and is connected to AF composition and nutritional requirements of the bacterium. Importantly, the response to growth in human AF included significant

changes in transcripts of multiple virulence genes such as adhesins, capsule, and hemolysin and IL-8 proteinase what might have consequences for the outcome of host-pathogen interactions (157).

Lastly, Soriani *et al.* performed a comparative global gene expression analysis of GBS at acidic and neutral pHs. They found that the transcription of 317 genes was increased at pH 5.5 relative to that at pH 7.0, while 61 genes were downregulated. The global response to acid stress included the differential expression of genes involved in transport, metabolism, stress response, and virulence. Known vaccine candidates, such as BibA and pilus components, were also regulated by pH. These results imply that the translocation of GBS from the acidic milieu of the vagina to the neutral pH of the neonatal lung signals the up-regulation of GBS virulence factors and conversion from a colonizing to an invasive phenotype (147).

All these articles show the fascinating ability and harmony of a pathogen to adapt different conditions by immediately shifting its expression profile. It is difficult to observe all these changes on the surfome by using limited *in vitro* conditions. Moreover one of the most significant outcomes of these studies concerning our results is about the expression of 5' nucleotidase SAG1333 (gbs1403). It was 8.1 times up-regulated in amniotic fluid that shows the importance of this antigen in amniotic fluid, just before transfering to lungs of the neonate (159). Additionally, at pH 5.5 a significant reduction was observed that could be interpreted as the down-regulation of the protein while colonizing on the vagina (137).

Table 16: GBS Surfome analysis Results

The list of fourty seven proteins identified on the surface of GBS. Proteins are grouped according to predicted localization. For each protein the following information is reported: NCBI gene ID, protein annotation and identified number of peptides.

GENE ID and Localization		Annotation	GBS Strain and number of peptides identified on the surfome							
			2603 V/R	18RS21	515	A909	CJB111	COH1	H36B	NEM316
	LPXTG									
1	SAG_0416	protease, putative (reticulocyte binding protein)	**45**	**40**		**1**	**1**		**6**	**2**

2	SAG_0421	conserved hypothetical protein	**1**		**1**			**1**		**1**
3	SAG_0771	cell wall surface anchor family protein	**5**	**1**	**1**				**1**	
4	SAG_1473	cell wall surface anchor family protein	**5**	**3**	**1**			**1**		
5	SAK_0186	iga fc receptor precursor				**57**	**2**	**1**	**1**	
6	SAG_1333	5'-nucleotidase family protein	**8**		**1**			**6**		
7	SAG_1941	2`,3`-cyclic-nucleotide 2`-phosphodiesterase						**1**		
8	SAG_0433	surface protein Rib	**7**							

9	SAG_2063	pathogenicity protein, putative		**1**				**3**		
10	SAN_1485	cell wall surface anchor family protein						**11**		
11	SAG_1331	R5 protein	**1**							
12	SAG_1474	amidase family protein					**1**	**1**		
13	SAG_1462	cell wall surface anchor family protein					**1**			**2**
14	SAL_1486	Pili backbone protein			**7**					
	Outside									

15	SAG_0017	PcsB protein	**1**		**2**		**1**		**1**	**2**
16	SAG_0290	ABC transporter, substrate-binding protein	**2**	**1**	**3**	**1**		**6**	**1**	**2**
17	SAG_1350	surface antigen-related protein	**6**		**1**		**2**	**4**		**1**
18	SAG_1361	conserved hypothetical protein	**1**					**2**		**1**
19	SAG_1441	maltose/maltodextrin ABC transporter	**1**			**3**		**1**		
20	SAG_1683	immunogenic secreted protein, putative	**3**		**1**	**3**	**1**	**7**		**2**
21	SAG_2043	cAMP factor		**1**						

22	SAG_0032	group B streptococcal surface immunogenic protein	**12**	**9**	**9**	**1**	**9**	**8**	**5**	**2**
23	SAG_0714	conserved hypothetical protein			**1**			**6**		
24	SAG_1371	conserved hypothetical protein						**2**	**1**	
25	SAI_2340	c protein alpha-antigen precursor			**2**	**2**			**8**	
26	SAG_0635	acid phosphatase precursor, class B						**1**		
27	SAG_1126	surface protein Spb1	**2**					**2**		
	Lipoprotein									

28	SAG_0776	lipoprotein, putative		**1**	**1**			**1**		**1**
29	SAG_1007	iron-compound ABC transporter		**1**					**1**	**1**
30	SAG_0757	peptidyl-prolyl cis-trans isomerase, cyclophilin-type			**1**			**1**		**1**
31	SAG_0383	conserved hypothetical protein				**2**	**3**			
32	SAG_1393	iron compound ABC transporter						**1**		
33	SAG_0954	lipoprotein	**1**							
34	SAG1530	peptidyl-prolyl cis-trans isomerase, cyclophilin-type						**1**	**1**	

35	SAG_1419	lipoprotein, putative							**1**	
36	SAG_0907	streptococcal histidine triad family protein			**2**					
37	SAG_1610	amino acid ABC transporter, substrate-binding protein								**2**
	Membrane									
38	SAG_0479	cell division protein FtsZ						**1**		**1**
39	SAJ_1257	C5a peptidase			**3**			**4**		**1**
40	SAG_1386	conserved hypothetical protein						**1**		

41	SAG_0615	transmembrane protein Vexp3								**1**
42	SAG_1466	amino acid ABC transporter			**2**			**1**		
43	SAG_0455	conserved hypothetical protein							**1**	
44	gbs1286	putative protein of unknown function/lipoprotein								**1**
45	SAG_0146	penicillin-binding protein 4	**1**		**1**			**1**		**1**
	Cytoplasmic									
46	SAG_0959	L-lactate dehydrogenase	**1**							

47	SAG_0606	conserved hypothetical protein	**3**	**1**	**1**			**1**		**2**

4. CONCLUSIONS

A comprehensive list of bacterial antigens described as potential vaccine candidates, based on *in vivo* animal models and/or on *in vitro* assays, was created from literature data and from on line available databases, *e.g.* the VIOLIN web site. 115 different protective antigens from 23 bacterial pathogens, 13 gram negative and 10 are gram positive, were considered for this process.

A systematic analysis of these antigens was carried out at the molecular level, using different bioinformatics tools and looking for conserved molecular features. The results of this analysis revealed that protective antigens from different species rarely show a conserved protein architecture over the whole sequence and a significant sequence similarity (>50% ID).

On the other hand, the results of this meta-analysis showed that protective antigens have recurring functional and/or structural units, which were in most cases associated to either specific domain from the Pfam databases or to a conserved protein architecture organized in a variable number of multiple internal repeats. The rest of the protective antigens not showing these properties were found to be either species-specific or genus-specific ORFs by genomic analysis

We used these features as “vaccine signatures” in a predictive selection process of new protective antigens in other species, and in the present study, we show the results obtained

from the application to GBS and *S. aureus*.

Among the nine antigens tested in the animal model against GBS, four were able to induce a significant level of protection as compared to the control group, *i.e.* the 5' nucleotidase SAG1333, the histidine-triad protein SAG0907, the Bmp protein SAG0954 and the LysM domain-containing protein SAG1386.\

Among the nine antigens tested in the animal model against *S. aureus*, five were able to induce a significant level of protection as compared to the control group *i.e.* putative uncharacterized protein SAOUHSC_00256, conserved hypothetical protein SAOUHSC_00400, N-acetylmuramoyl-L-alanine amidase sle1 SAOUHSC_00427, intracellular serine protease SAOUHSC_0194 and putative uncharacterized protein SAOUHSC_02979.

In order to check the expression of the MetaVaccinology identified protective antigens on the surface compartment of GBS, we used surfome analysis for 8 different strains of GBS. The expression of protective antigens on the surface compartment of GBS was confirmed by this way.

Our results indicate that the MetaVaccinology selection process is a powerful discovery tool in vaccine research and we expect this approach can be of particular interest for projects in the pre-discovery phase, allowing the fast identification of protective antigens with conserved "vaccine signatures".

The new identified proteins share common functional/structural domains with antigens proved to be protective and relevant to the pathogenesis of other important pathogens, *i.e. Haemophilus influenzae*, *Staphylococcus aureus* and *Streptococcus pneumoniae*. For this reason, the MetaVaccinology approach can also be exploited as a basic science tool, as it provides insights about possible common mechanisms of pathogenesis in different bacterial species.

Related to this last point, the role of 5'nucleotidase in immune evasion for GBS was investigated and our study shows that SAG_1333 has a toxic effect on the macrophages in the presence of AMP.

The identification of common mechanisms of pathogenesis for such a different range of pathogens, including Gram positive and Gram negatives, can also open the way to the design of new antimicrobials and other therapeutic treatments.

As a final comment, MetaVaccinology could be exploited to make vaccines against any bacterial pathogen, including the most dangerous ones *i.e Mycobacterium tuberculosis*. On the other, we think that MetaV could be also applied for the parasites *i.e* the *Plasmodium falciparum* which is not a bacteria but a protozoan parasite causing malaria.

REFERENCES

1) Hilleman MR. Vaccines in historic evolution and perspective: a narrative of vaccine discoveries. Vaccine. 2000 Feb 14;18(15):1436-47.

2) Riedel S. Edward Jenner and the history of smallpox and vaccination. Proc (Bayl Univ Med Cent). 2005 Jan;18(1):21-5.

3) Datta SK, Bhatla N, Burgess MA, Lehtinen M, Bock HL. Women and vaccinations: From smallpox to the future, a tribute to a partnership benefiting humanity for over 200 years. Hum Vaccin. 2009 Jul;5(7):450-4. Epub 2009 Jul 20.

4) Borysiewicz LK. Prevention is better than cure. Lancet. 2010 Feb 6;375(9713):513-23. Epub 2009 Dec 23.

5) Stanley A Plotkin Vaccines: past, present and future, Nature Medicine 11, S5 - S11 (2005)

6) Galambos L. A century of innovation in vaccines. Vaccine. 1999 Oct 29;17 Suppl 3:S7-S10.

7) Immunization against diseases of public health importance

http://www.who.int/mediacentre/factsheets/fs288/en/index.html

8) André FE. Vaccinology: past achievements, present roadblocks and future promises. Vaccine. 2003 Jan 30;21(7-8):593-5.

9) Mäkelä PH. Vaccines, coming of age after 200 years. FEMS Microbiol Rev. 2000 Jan;24(1):9-20.

10) McCullers JA. Evolution, benefits, and shortcomings of vaccine management. J Manag Care Pharm. 2007 Sep;13(7 Suppl B):S2-6.

11) Bijal Trivedi , Freelance Science Writer Profile of Rino Rappuoli PNAS July 18, 2006 vol. 103 no. 29 10831-10833

12) Pizza M, Covacci A, Bartoloni A, Perugini M, Nencioni L, De Magistris MT, Villa L, Nucci D, Manetti R, Bugnoli M, *et al.* Mutants of pertussis toxin suitable for vaccine development. Science. 1989 Oct 27;246(4929):497-500.

13) WHO: Hepatitis B

http://www.who.int/immunization/topics/hepatitis_b/en/index.html

14) Davide Serruto, Rino Rappuoli. Post-genomic vaccine development FEBS Lett. 2006 May 22;580(12):2985-92.

15) Barbara Capecchi, Davide Serruto, Jeannette Adu Bobie, Rino Rappuoli and Mariagrazia Pizza. The Genome Revolution in Vaccine Research . Curr Issues Mol Biol. 2004 Jan;6(1):17-27.

16) Rinaudo C.D, John L. Telford, Rino Rappuoli, and Kate L. Seib. Vaccinology in the genome era. J Clin Invest. 2009 Sep;119(9):2515-25

17) Rappuoli Rino. Reverse vaccinology, a genome-based approach to vaccine development. Vaccine 2001 Mar 21;19(17-19):2688-91.

18) Stefania Bambini and Rino Rappuoli. The use of genomics in microbial vaccine development. Drug Discov Today. 2009 Mar;14(5-6):252-60.

19) Tettelin H, Saunders NJ, Heidelberg J, *et al*. Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. Science 2000;287(5459):1809–15.

20) Jeannette Adu-Bobie, Barbara Capecchi, Davide Serruto, Rino Rappuoli, Mariagrazia Pizza. Two years into reverse vaccinology. Vaccine. 2003 Jan 30;21(7-8):605-10.

21) Giuliani MM, Adu-Bobie J, Comanducci M, Aricò 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, Masignani 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. A universal

vaccine for serogroup B meningococcus. Proc Natl Acad Sci U S A. 2006 Jul 18;103(29):10834-9.

22) 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. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome".Proc Natl Acad Sci U S A. 2005 Sep 27;102(39):13950-5. Epub 2005 Sep 19.

23) Muzzi A, Masignani V, Rappuoli R. The pan-genome: towards a knowledge-based discovery of novel targets for vaccines and antibacterials. Drug Discov Today. 2007 Jun;12(11-12):429-39. Epub 2007 May 7.

24) Muzzi A, Masignani V, Rappuoli R. The pan-genome: towards a knowledge-based discovery of novel targets for vaccines and antibacterials. Drug Discov Today. 2007 Jun;12(11-12):429-39.

25) 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, Nardi Dei V, Vegni F, Fraser C, Mancuso G, Teti G, Madoff LC, Paoletti LC, Rappuoli R, Kasper DL, Telford JL, Grandi G. Identification of a universal Group B streptococcus vaccine by multiple genome screen. Science. 2005 Jul 1;309(5731):148-50.

26) 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. Identification of protective and broadly conserved vaccine antigens from the genome of extraintestinal pathogenic Escherichia coli. Proc Natl Acad Sci U S A. 2010 May 18;107(20):9072-7.

27) Etz H, Minh DB, Henics T, Dryla A, Winkler B, Triska C, Boyd AP, Söllner J, Schmidt W, von Ahsen U, Buschle M, Gill SR, Kolonay J, Khalak H, Fraser CM, von Gabain A, Nagy E, Meinke A. Identification of in vivo expressed vaccine candidate antigens from Staphylococcus aureus. Proc Natl Acad Sci U S A. 2002 May 14;99(10):6573-8. Epub 2002

May 7.

28) Meinke A, Henics T, Hanner M, Minh DB, Nagy E. Antigenome technology: a novel approach for the selection of bacterial vaccine candidate antigens. Vaccine. 2005 Mar 18;23(17-18):2035-41.

29) Giefing C, Meinke AL, Hanner M, Henics T, Bui MD, Gelbmann D, Lundberg U, Senn BM, Schunn M, Habel A, Henriques-Normark B, Ortqvist A, Kalin M, von Gabain A, Nagy E. Discovery of a novel class of highly conserved vaccine antigens using genomic scale antigenic fingerprinting of pneumococcus with human antibodies. J Exp Med. 2008 Jan 21;205(1):117-31.

30) Fritzer A, Senn BM, Minh DB, Hanner M, Gelbmann D, Noiges B, Henics T, Schulze K, Guzman CA, Goodacre J, von Gabain A, Nagy E, Meinke AL. Novel conserved group A streptococcal proteins identified by the antigenome technology as vaccine candidates for a non-M protein-based vaccine. Infect Immun. 2010 Sep;78(9):4051-67. Epub 2010 Jul 12.

31) Wasinger, V. C., Cordwell, S. J., Cerpa-Poljak, A., Yan, J. X.,Gooley, A. A., Wilkins, M. R., Duncan, M. W., Harris, R., Williams, K. Humphrey-Smith, I. Progress with gene product mapping of the Mullicutes Mycoplasma genitalium. Electrophoresis (1995), 16: 1090–1094.

32) O'Farrell PH. High resolution two-dimensional electrophoresis of proteins. J Biol Chem, 1975, 250: 4007–4021.

33) Hanash, S. Disease proteomics. Nature, 2003, 422: 226-232.

34) Walters MS, Mobley HL. Bacterial proteomics and identification of potential vaccine targets. Expert Rev Proteomics. 2010 Apr;7(2):181-4.

35) Steen H, Mann M. The ABC's (and XYZ's) of peptide sequencing. Nat Rev Mol Cell Biol. 2004 Sep;5(9):699-711.

36) Altindiş E, Tefon BE, Yildirim V, Ozcengiz E, Becher D, Hecker M, Ozcengiz G.Immunoproteomic analysis of Bordetella pertussis and identification of new immunogenic proteins. Vaccine. 2009 Jan 22;27(4):542-8. Epub 2008 Nov 24.

37) Jagusztyn-Krynicka EK, Dadlez M, Grabowska A, Roszczenko P. Proteomic technology

in the design of new effective antibacterial vaccines. Expert Rev Proteomics. 2009 Jun;6(3):315-30.

38) Rodríguez-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. Characterization and identification of vaccine candidate proteins through analysis of the group A Streptococcus surface proteome. Nat Biotechnol. 2006 Feb;24(2):191-7. Epub 2006 Jan 15.

39) Vivona S, Gardy JL, Ramachandran S, Brinkman FS, Raghava GP, Flower DR, Filippini F. Computer-aided biotechnology: from immuno-informatics to reverse vaccinology. Trends Biotechnol. 2008 Apr;26(4):190-200.

40) Doro F, Liberatori S, Rodríguez-Ortega MJ, Rinaudo CD, Rosini R, Mora M, Scarselli M, Altindis E, D'Aurizio R, Stella M, Margarit I, Maione D, Telford JL, Norais N, Grandi G. Surfome analysis as a fast track to vaccine discovery: identification of a novel protective antigen for Group B Streptococcus hypervirulent strain COH1. Mol Cell Proteomics. 2009 Jul;8(7):1728-37.

41) Mulder NJ, Kersey P, Pruess M, Apweiler R. In silico characterization of proteins: UniProt, InterPro and Integr8. Mol Biotechnol. 2008 Feb;38(2):165-77. Epub 2007 Oct 4.

42) Buljan M, Bateman A. The evolution of protein domain families. Biochem Soc Trans. 2009 Aug;37(Pt 4):751-5.

43) Sammut SJ, Finn RD, Bateman A. Pfam 10 years on: 10,000 families and still growing. Brief Bioinform. 2008 May;9(3):210-9. Epub 2008 Mar 15.

44) Schultz J, Copley RR, Doerks T, Ponting CP, Bork P. SMART: a web-based tool for the study of genetically mobile domains. Nucleic Acids Res. 2000 Jan 1;28(1):231-4.

45) Rajagopal L. Understanding the regulation of Group B Streptococcal virulence factors. Future Microbiol. 2009 Mar;4(2):201-21.

46) Heath PT, Feldman RG. Vaccination against group B streptococcus. Expert Rev Vaccines. 2005 Apr;4(2):207-18.

47) Johri AK, Paoletti LC, Glaser P, Dua M, Sharma PK, Grandi G, Rappuoli R. Group B

Streptococcus: global incidence and vaccine development. Nat Rev Microbiol. 2006 Dec;4(12):932-42.

48) Schaffer AC, Lee JC. Vaccination and passive immunisation against Staphylococcus aureus. Int J Antimicrob Agents. 2008 Nov;32 Suppl 1:S71-8. Epub 2008 Aug 30.

49) David MZ, Daum RS. Community-associated methicillin-resistant Staphylococcus aureus: epidemiology and clinical consequences of an emerging epidemic. Clin Microbiol Rev. 2010 Jul;23(3):616-87.

50) Francois P, Scherl A, Hochstrasser D, Schrenzel J. Proteomic approach to investigate MRSA. Methods Mol Biol. 2007;391:179-99.

51) García-Lara J, Foster SJ. Anti-Staphylococcus aureus immunotherapy: current status and prospects. Curr Opin Pharmacol. 2009 Oct;9(5):552-7. Epub 2009 Sep 3.

52) Meoni E, Faenzi E, Frigimelica E, Zedda L, Skibinski D, Giovinazzi S, Bonci A, Petracca R, Bartolini E, Galli G, Agnusdei M, Nardelli F, Buricchi F, Norais N, Ferlenghi I, Donati M, Cevenini R, Finco O, Grandi G, Grifantini R. CT043, a protective antigen that induces a CD4+ Th1 response during Chlamydia trachomatis infection in mice and humans. Infect Immun. 2009 Sep; 77(9):4168-76. Epub 2009 Jul 13.

53) Montigiani, S., *et al*. 2002. Genomic approach for analysis of surface proteins in Chlamydia pneumoniae. Infect. Immun. 70:368–379.

54) Ariel N, Zvi A, Grosfeld H, Gat O, Inbar Y, Velan B, Cohen S, Shafferman A. Search for potential vaccine candidate open reading frames in the Bacillus anthracis virulence plasmid pXO1: in silico and in vitro screening. Infect. Immun. 2002; 70:6817–6827.

55) Ariel N, Zvi A, Makarova KS, Chitlaru T, Elhanany E, Velan B, Cohen S, Friedlander AM, Shafferman A. Genome-based bioinformatic selection of chromosomal Bacillus anthracis putative vaccine candidates coupled with proteomic identification of surface-associated antigens. Infect. Immun. 2003; 71:4563–4579.

56). Yang, X., Hudson, M., Walters, N., Bargatze, R.F., nd Pascual, D.W. 2005. Selection of protective epitopes for Brucella melitensis by DNA vaccination. Infect. Immun. 73:7297–7303.

57) Lauer P, Rinaudo CD, Soriani M, Margarit I, Maione D, Rosini R, Taddei AR, Mora M, Rappuoli R, Grandi G, Telford JL. Genome analysis reveals pili in Group B Streptococcus. Science. 2005 Jul 1;309(5731):105.

58) Serruto D, Rappuoli R, Scarselli M, Gros P, van Strijp JA. Molecular mechanisms of complement evasion: learning from staphylococci and meningococci. Nat Rev Microbiol. 2010 Jun;8(6):393-9.

59) Xiang Z, Todd T, Ku KP, Kovacic BL, Larson CB, Chen F, Hodges AP, Tian Y, Olenzek EA, Zhao B, Colby LA, Rush HG, Gilsdorf JR, Jourdian GW, He Y. VIOLIN: vaccine investigation and online information network. Nucleic Acids Res. 2008 Jan;36(Database issue):D923-8. Epub 2007 Nov 19.

60) Zhai Y, Saier MH Jr. A simple sensitive program for detecting internal repeats in sets of multiply aligned homologous proteins. J Mol Microbiol Biotechnol. 2002 Jul;4(4):375-7.

61) Tettelin H, Masignani 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. Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V Streptococcus agalactiae. Proc Natl Acad Sci U S A. 2002 Sep 17;99(19):12391-6. Epub 2002 Aug 28.

62) Heath E. Klock and Scott A. Lesley The Polymerase Incomplete Primer Extension (PIPE) Method Applied to High-Throughput Cloning and Site-Directed Mutagenesis *Methods in Molecular Biology: High Throughput Protein Expression and Purification, vol. 498*

63) Campodónico VL, Llosa NJ, Grout M, Döring G, Maira-Litrán T, Pier GB. Evaluation of flagella and flagellin of *Pseudomonas aeruginosa* as vaccines. Infect Immun. 2010 Feb;78(2):746-55. Epub 2009 Dec 7.

64) Roy K, Hamilton D, Ostmann MM, Fleckenstein JM. Vaccination with EtpA glycoprotein or flagellin protects against colonization with enterotoxigenic *Escherichia coli* in a murine model. Vaccine. 2009 Jul 23;27(34):4601-8. Epub 2009 Jun 11.

65) Baqar S, Applebee LA, Gilliland TC Jr, Lee LH, Porter CK, Guerry P. Immunogenicity and protective efficacy of recombinant *Campylobacter jejuni* flagellum-secreted proteins in mice. Infect Immun. 2008 Jul;76(7):3170-5. Epub 2008 Apr 21.

66) Skene C, Young A, Every A, Sutton P. *Helicobacter pylori* flagella: antigenic profile and protective immunity. FEMS Immunol Med Microbiol. 2007 Jul;50(2):249-56. Epub 2007 May 23.

67) Durant L, Metais A, Soulama-Mouze C, Genevard JM, Nassif X, Escaich S. Identification of Candidates for a Subunit Vaccine against Extraintestinal Pathogenic Escherichia coli Infect Immun. 2007 Apr;75(4):1916-25. Epub 2006 Dec 4.

68) Munson RS Jr, Granoff DM. Purification and partial characterization of outer membrane proteins P5 and P6 from *Haemophilus influenzae* type b. Infect Immun. 1985 Sep;49(3):544-9.

69) Pasquevich KA, Estein SM, García Samartino C, Zwerdling A, Coria LM, Barrionuevo P, Fossati CA, Giambartolomei GH, Cassataro J. Immunization with recombinant Brucella species outer membrane protein Omp16 or Omp19 in adjuvant induces specific CD4+ and CD8+ T cells as well as systemic and oral protection against *Brucella abortus* infection. Infect Immun. 2009 Jan;77(1):436-45. Epub 2008 Nov 3.

70) von Specht B, Knapp B, Hungerer K, Lücking C, Schmitt A, Domdey H. Outer membrane proteins of *Pseudomonas aeruginosa* as vaccine candidates. J Biotechnol. 1996 Jan 26;44(1-3):145-53.

71) Alexander JE, Lock RA, Peeters CC, Poolman JT, Andrew PW, Mitchell TJ, Hansman D, Paton JC. Immunization of mice with pneumolysin toxoid confers a significant degree of protection against at least nine serotypes of *Streptococcus pneumoniae*. Infect Immun. 1994 Dec;62(12):5683-8.

72) US Application Publication (Source: USPTO), Publication No. US 2010/0158935 A1 published on 24-Jun-2010, Application No. US 12/711869 filed on 24-Feb-2010

73) Bahey-El-Din M, Casey PG, Griffin BT, Gahan CG. Lactococcus lactis-expressing listeriolysin O (LLO) provides protection and specific CD8(+) T cells against Listeria monocytogenes in the murine infection model. Vaccine. 2008 Sep 26;26(41):5304-14. Epub 2008 Aug 6.

74) Park HM, Yoo HS, Oh TH, Kim D, Han HR. Immunogenicity of Alpha-Toxin, Capsular Polysaccharide (CPS) and Recombinant Fibronectin-Binding Protein (r-FnBP) of Staphylococcus aureus in Rabbit J Vet Med Sci. 1999 Sep;61(9):995-1000.

75) Brown EL, Dumitrescu O, Thomas D, Badiou C, Koers EM, Choudhury P, Vazquez V, Etienne J, Lina G, Vandenesch F, Bowden MG. The Panton-Valentine leukocidin vaccine protects mice against lung and skin infections caused by Staphylococcus aureus USA300. Clin Microbiol Infect. 2009 Feb;15(2):156-64. Epub 2008 Dec 22.

76) Clare K. Schmitt, Karen C. Meysick, and Alison D. O.Brien Bacterial Toxins: Friends or Foes? Emerging Infectious Diseases Vol. 5, No. 2, April.June 1999

77) Poolman JT, Hallander HO. Acellular pertussis vaccines and the role of pertactin and fimbriae. Expert Rev Vaccines. 2007 Feb;6(1):47-56.

78) Sánchez J, Holmgren J. Cholera toxin structure, gene regulation and pathophysiological and immunological aspects Cell Mol Life Sci. 2008 May;65(9):1347-60.

79) Blencowe H, Lawn J, Vandelaer J, Roper M, Cousens S. Tetanus toxoid immunization to reduce mortality from neonatal tetanus. Int J Epidemiol. 2010 Apr;39 Suppl 1:i102-9.

80) Smith LA, Rusnak JM. Botulinum neurotoxin vaccines: past, present, and future. Crit Rev Immunol. 2007;27(4):303-18.

81) Giannasca PJ, Warny M. Active and passive immunization against Clostridium difficile diarrhea and colitis. Vaccine. 2004 Feb 17;22(7):848-56.

82) Cleary PP, Matsuka YV, Huynh T, Lam H, Olmsted SB. Immunization with C5a peptidase from either group A or B streptococci enhances clearance of group A streptococci from intranasally infected mice Vaccine. 2004 Oct 22;22(31-32):4332-41.

83) AZEH I, NAU R, BETHE G, NEUHAUS A, WELLMER A, HEINZ HP, ZYSK G; Immunization of Mice with the Cell Wall-Associated Serine Protease PrtA of Streptococcus pneumoniae Protects against Invasive Pneumococcal Infection. *Abstr Intersci Conf Antimicrob Agents Chemother Intersci Conf Antimicrob Agents Chemother.* 2001 Dec 16-19; 41: abstract no. G-2038.

84) Turner DP, Wooldridge KG, Ala'Aldeen DA. Autotransported serine protease A of Neisseria meningitidis: an immunogenic, surface-exposed outer membrane, and secreted protein. Infect Immun. 2002 Aug;70(8):4447-61.

85) Kuo CF, Wu JJ, Lin KY, Tsai PJ, Lee SC, Jin YT, Lei HY, Lin YS. Role of streptococcal pyrogenic exotoxin B in the mouse model of group A streptococcal infection. INFECTION AND IMMUNITY Aug. 1998, p. 3931–3935

86) Fritzer A, Senn BM, Minh DB, Hanner M, Gelbmann D, Noiges B, Henics T, Schulze K, Guzman CA, Goodacre J, von Gabain A, Nagy E, Meinke AL. Novel conserved group a streptococcal proteins identified by the antigenome technology as vaccine candidates for a non-m protein-based vaccine. Infect Immun. 2010 Sep;78(9):4051-67. Epub 2010 Jul 12.

87) Zagursky RJ, Ooi P, Jones KF, Fiske MJ, Smith RP, Green BA. Identification of a Haemophilus influenzae 5'-Nucleotidase Protein: Cloning of the nucA Gene and Immunogenicity and Characterization of the NucA Protein INFECTION AND IMMUNITY, May 2000, p. 2525–2534

88) Fritzer A, Senn BM, Minh DB, Hanner M, Gelbmann D, Noiges B, Henics T, Schulze K, Guzman CA, Goodacre J, von Gabain A, Nagy E, Meinke AL. Novel conserved group a streptococcal proteins identified by the antigenome technology as vaccine candidates for a non-m protein-based vaccine. Infect Immun. 2010 Sep;78(9):4051-67. Epub 2010 Jul 12.

89) Kuklin, NAA *et al.* novel Staphylococcus aureus vaccine: iron surface determinant B induces rapid antibody responses in rhesus macaques and specific increased survival in a murine S. aureus sepsis model. Infect Immun. 2006 Apr;74(4):2215-23.

90) Kim HK, Dedent A, Cheng AG, McAdow M, Bagnoli F, Missiakas DM, Schneewind O. IsdA and IsdB antibodies protect mice against Staphylococcus aureus abscess formation and lethal challenge. Vaccine. 2010 Aug 31;28(38):6382-6392. Epub 2010 Mar 10.

91) Durant L, Metais A, Soulama-Mouze C, Genevard JM, Nassif X, Escaich S. Identification of Candidates for a Subunit Vaccine against Extraintestinal Pathogenic Escherichia coli Infect Immun. 2007 Apr;75(4):1916-25. Epub 2006 Dec 4.

92) Alteri CJ, Hagan EC, Sivick KE, Smith SN, Mobley HL. Mucosal immunization with iron receptor antigens protects against urinary tract infection. PLoS Pathog. 2009 Sep;5(9):e1000586. Epub 2009 Sep 18.

93) Potter AA, Schryvers AB, Ogunnariwo JA, Hutchins WA, Lo RY, Watts T. Protective capacity of the Pasteurella haemolytica transferrin-binding proteins TbpA and TbpB in cattle. Microb Pathog. 1999 Oct;27(4):197-206.

94) Brown JS, Ogunniyi AD, Woodrow MC, Holden DW, Paton JC. Immunization with Components of Two Iron Uptake ABC Transporters Protects Mice against Systemic Streptococcus pneumoniae Infection. INFECTION AND IMMUNITY, 6702–6706.2001 Nov. 2001, p. 6702–6706

95) Letesson JJ, Tibor A, van Eynde G, Wansard V, Weynants V, Denoel P, Saman E. Humoral immune responses of Brucella-infected cattle, sheep, and goats to eight purified recombinant Brucella proteins in an indirect enzyme-linked immunosorbent assay. Clin Diagn Lab Immunol. 1997 Sep;4(5):556-64.

96) Shah P, Swiatlo E. Immunization with polyamine transport protein PotD protects mice against systemic infection with Streptococcus pneumoniae. Infect Immun. 2006 Oct;74(10):5888-92.

97) Buckley AM, Wang J, Hudson DL, Grant AJ, Jones MA, Maskell DJ, Stevens MP. Evaluation of live-attenuated Salmonella vaccines expressing Campylobacter antigens for control of C. jejuni in poultry. Vaccine. 2010 Jan 22;28(4):1094-105. Epub 2009 Oct 22.

98) Mora M, Bensi G, Capo S, Falugi F, Zingaretti C, Manetti AG, Maggi T, Taddei AR, Grandi G, Telford JL. Group A Streptococcus produce pilus-like structures containing protective antigens and Lancefield T antigens, PNAS October 25, 2005 vol. 102 no. 43 15641–15646

99) Rosini R, Rinaudo CD, Soriani M, Lauer P, Mora M, Maione D, Taddei A, Santi I, Ghezzo C, Brettoni C, Buccato S, Margarit I, Grandi G, Telford JL. Identification of novel genomic islands coding for antigenic pilus-like structures in Streptococcus agalactiae. Mol Microbiol. 2006 Jul;61(1):126-41.

100) Gianfaldoni C, Censini S, Hilleringmann M, Moschioni M, Facciotti C, Pansegrau W, Masignani V, Covacci A, Rappuoli R, Barocchi MA, Ruggiero P. Streptococcus pneumoniae Pilus Subunits Protect Mice against Lethal Challenge. Infect Immun. 2007 Feb;75(2):1059-62. Epub 2006 Dec 4.

101) Zhou H, Xiong ZY, Li HP, Zheng YL, Jiang YQ. An immunogenicity study of a newly fusion protein Cna-FnBP vaccinated against Staphylococcus aureus infections in a mice model. Vaccine 24 (2006) 4830–4837

102) Stranger-Jones YK, Bae T, Schneewind O. Vaccine assembly from surface proteins of Staphylococcus aureus. Proc Natl Acad Sci U S A. 2006 Nov 7;103(45):16942-7. Epub 2006 Oct 30.

103) Olive C, Clair T, Yarwood P, Good MF. Protection of mice from group A streptococcal infection by intranasal immunisation with a peptide vaccine that contains a conserved M protein B cell epitope and lacks a T cell autoepitope. Vaccine. 2002 Jun 21;20(21-22):2816-25.

104) Seepersaud R, Hanniffy SB, Mayne P, Sizer P, Le Page R, Wells JM. Characterization of a novel leucine-rich repeat protein antigen from group B streptococci that elicits protective immunity.. Infect Immun. 2005 Mar;73(3):1671-83.

105) Stålhammar-Carlemalm M, Stenberg L, Lindahl G. Protein rib: a novel group B streptococcal cell surface protein that confers protective immunity and is expressed by most strains causing invasive infections. J Exp Med. 1993 Jun 1;177(6):1593-603.

106) Gravekamp C, Kasper DL, Paoletti LC, Madoff LC. Alpha C protein as a carrier for type III capsular polysaccharide and as a protective protein in group B streptococcal vaccines. Infect Immun. 1999 May;67(5):2491-6.

107) Pizza M, Donnelly J, Rappuoli R. Factor H-binding protein, a unique meningococcal vaccine antigen. Vaccine. 2008 Dec 30;26 Suppl 8:I46-8.

108) Serruto D, Spadafina T, Ciucchi L, Lewis LA, Ram S, Tontini M, Santini L, Biolchi A, Seib KL, Giuliani MM, Donnelly JJ, Berti F, Savino S, Scarselli M, Costantino P, Kroll JS, O'Dwyer C, Qiu J, Plaut AG, Moxon R, Rappuoli R, Pizza M, Aricò B. Neisseria meningitidis GNA2132, a heparin-binding protein that induces protective immunity in humans. Proc Natl Acad Sci U S A. 2010 Feb 23;107(8):3770-5. Epub 2010 Feb 3.

109) Webb DC, Cripps AW. Immunization with recombinant transferrin binding protein B enhances clearance of nontypeable Haemophilus influenzae from the rat lung. Infect Immun. 1999 May;67(5):2138-44.

110) Seib KL, Oriente F, Adu-Bobie J, Montanari P, Ferlicca F, Giuliani MM, Rappuoli R, Pizza M, Delany I. Influence of serogroup B meningococcal vaccine antigens on growth and survival of the meningococcus in vitro and in ex vivo and in vivo models of infection. Vaccine. 2010 Mar 11;28(12):2416-27. Epub 2010 Jan 10.

111) Meier PS, Troller R, Grivea IN, Syrogiannopoulos GA, Aebi C. The outer membrane proteins UspA1 and UspA2 of Moraxella catarrhalis are highly conserved in nasopharyngeal isolates from young children. Vaccine. 2002 Mar 15;20(13-14):1754-60.

112) Murphy BS, Wulff CR, Garvy BA, Straley SC. Yersinia pestis YadC: a novel vaccine candidate against plague. Adv Exp Med Biol. 2007;603:400-14.

113) Adamou JE, Heinrichs JH, Erwin AL, Walsh W, Gayle T, Dormitzer M, Dagan R, Brewah YA, Barren P, Lathigra R, Langermann S, Koenig S, Johnson S. Identification and characterization of a novel family of pneumococcal proteins that are protective against sepsis. Infect Immun. 2001 Feb;69(2):949-58.

114) Yang X, Izadi H, Coleman AS, Wang P, Ma Y, Fikrig E, Anguita J, Pal U. Borrelia burgdorferi lipoprotein BmpA activates pro-inflammatory responses in human synovial cells through a protein moiety. Microbes Infect. 2008 Oct;10(12-13):1300-8. Epub 2008 Aug 5.

115) Poolman JT, Hallander HO. Acellular pertussis vaccines and the role of pertactin and fimbriae. Expert Rev Vaccines. 2007 Feb;6(1):47-56.

116) Cutter D, Mason KW, Howell AP, Fink DL, Green BA, St Geme JW 3rd. Immunization with Haemophilus influenzae Hap adhesin protects against nasopharyngeal colonization in experimental mice. J Infect Dis. 2002 Oct 15;186(8):1115-21. Epub 2002 Sep 16.

117) Serruto D, Adu-Bobie J, Scarselli M, Veggi D, Pizza M, Rappuoli R, Aricò B. Neisseria meningitidis App, a new adhesin with autocatalytic serine protease activity. Mol Microbiol. 2003 Apr;48(2):323-34.

118) Huleatt JW, Jacobs AR, Tang J, Desai P, Kopp EB, Huang Y, Song L, Nakaar V, Powell TJ. Vaccination with recombinant fusion proteins incorporating Toll-like receptor ligands induces rapid cellular and humoral immunity. Vaccine. 2007 Jan 8;25(4):763-75. Epub 2006 Aug 22.

119) Gu J, Liu Y, Yu S, Wang H, Wang Q, Yi Y, Zhu F, Yu XJ, Zou Q, Mao X. Enterohemorrhagic Escherichia coli trivalent recombinant vaccine containing EspA,

intimin and Stx2 induces strong humoral immune response and confers protection in mice. Microbes Infect. 2009 Sep;11(10-11):835-41. Epub 2009 May 7.

120) Zhou H, Xiong ZY, Li HP, Zheng YL, Jiang YQ. An immunogenicity study of a newly fusion protein Cna-FnBP vaccinated against Staphylococcus aureus infections in a mice model. Vaccine 24 (2006) 4830–4837

121) Ferreira DM, Darrieux M, Silva DA, Leite LC, Ferreira JM Jr, Ho PL, Miyaji EN, Oliveira ML. Characterization of Protective Mucosal and Systemic Immune Responses Elicited by Pneumococcal Surface Protein PspA and PspC Nasal Vaccines against a Respiratory Pneumococcal Challenge in Mice. CLINICAL AND VACCINE IMMUNOLOGY, May 2009, p. 636–645

122) Ogunniyi AD, Grabowicz M, Mahdi LK, Cook J, Gordon DL, Sadlon TA, Paton JC. Pneumococcal histidine triad proteins are regulated by the Zn2+-dependent repressor AdcR and inhibit complement deposition through the recruitment of complement factor H. FASEB J. 2009 Mar;23(3):731-8. Epub 2008 Oct 29.

123) Melin M, Di Paolo E, Tikkanen L, Jarva H, Neyt C, Käyhty H, Meri S, Poolman J, Väkeväinen M. Interaction of pneumococcal histidine triad proteins with human complement. Infect Immun. 2010 May;78(5):2089-98. Epub 2010 Mar 1.

124) Reid SD, Montgomery AG, Voyich JM, DeLeo FR, Lei B, Ireland RM, Green NM, Liu M, Lukomski S, Musser JM. Characterization of an extracellular virulence factor made by group A Streptococcus with homology to the Listeria monocytogenes internalin family of proteins. Infect Immun. 2003 Dec;71(12):7043-52.

125) Verma A, Brissette CA, Bowman A, Stevenson B. Borrelia burgdorferi BmpA is a laminin-binding protein. Infect Immun. 2009 Nov;77(11):4940-6. Epub 2009 Aug 24.

126) Pal U, Wang P, Bao F, Yang X, Samanta S, Schoen R, Wormser GP, Schwartz I, Fikrig E. Borrelia burgdorferi basic membrane proteins A and B participate in the genesis of Lyme arthritis. J Exp Med. 2008 Jan 21;205(1):133-41. Epub 2007 Dec 31.

127) Rosati, S., Robino, P., Fadda, M., Pozzi, S., Mannelli, A., Pittau, M., 2000. Expression and antigenic characterization of recombinant Mycoplasma agalactiae P48 major surface protein. Vet. Microbiol. 71, 201–210.

128) Chessa B, Pittau M, Puricelli M, Zobba R, Coradduzza E, Dall'ara P, Rosati S, Poli G, Alberti A. Genetic immunization with the immunodominant antigen P48 of Mycoplasma agalactiae stimulates a mixed adaptive immune response in BALBc mice. Res Vet Sci. 2009 Jun;86 (3):414-20. Epub 2008 Nov 13.

129) Buist G, Steen A, Kok J, Kuipers OP. LysM, a widely distributed protein motif for binding to (peptido)glycans. Mol Microbiol. 2008 May;68(4):838-47.

130) Holtfreter S, Kolata J, Bröker BM. Towards the immune proteome of Staphylococcus aureus - The anti-S. aureus antibody response. Int J Med Microbiol. 2010 Feb;300(2-3):176-92. Epub 2009 Nov 3.

131) Trigo G, Ferreira P, Ribeiro N, Dinis M, Andrade EB, Melo-Cristino J, Ramirez M, Tavares D. Identification of immunoreactive extracellular proteins of Streptococcus agalactiae in bovine mastitis. Can J Microbiol. 2008 Nov;54(11):899-905.

132) Sitkiewicz I, Green NM, Guo N, Bongiovanni AM, Witkin SS, Musser JM. Transcriptome adaptation of group B Streptococcus to growth in human amniotic fluid. PLoS One. 2009 Jul 1;4(7):e6114.

133) Thammavongsa V, Kern JW, Missiakas DM, Schneewind O. Staphylococcus aureus synthesizes adenosine to escape host immune responses. J Exp Med. 2009 Oct 26;206(11):2417-27. Epub 2009 Sep 28.

134) Punj V, Zaborina O, Dhiman N, Falzari K, Bagdasarian M, Chakrabarty AM. Phagocytic Cell Killing Mediated by Secreted Cytotoxic Factors of Vibrio cholerae. Infect Immun. 2000 Sep;68(9):4930-7.

135) Anatoli Melnikov, Olga Zaborina, Neelam Dhiman, B. S. Prabhakar, A. M. Chakrabarty and William Hendrickson. Clinical and environmental isolates of Burkholderia cepacia exhibit differential cytotoxicity towards macrophages and mast cells. Molecular Microbiology (2000) 36(6), 1481±1493

136) Di Virgilio, F.. The P2Z purinoreceptor—an intriguing role in immunity, inflammation and cell death. Immunol. Today 199516:524–528.

137) P2-purinergic receptors: subtype-associated signaling and structure. Harden, T. K., J. L. Boyer, and R. A. Nicholas. 1995. Annu. Rev. Pharmacol. Toxicol. 35:541–579.

138) Wiendl, H. S., C. Schneider, and A. Ogilvie. Nucleotide metabolizing ectoenzymes are upregulated in A431 cells periodically treated with cytostatic ATP leading to partial resistance without preventing apoptosis. Biochim.Biophys. Acta 1998. 1404:282–298.

139) Gounaris K. Nucleotidase cascades are catalyzed by secreted proteins of the parasitic nematode Trichinella spiralis. Infect Immun. 2002 Sep;70(9):4917-24.

140) Kumar V, Sharma A. Adenosine: an endogenous modulator of innate immune system with therapeutic potential. Eur J Pharmacol. 2009 Aug 15;616(1-3):7-15

141) Min HW, Moochhala S, Eng KH. Adenosine and its receptor agonists regulate nitric oxide production and RAW 264.7 macrophages via both receptor binding and its downstream metabolites-inosine. Life Sci. 2000 Mar 31;66(19):1781-93

142) Yook YH, Kang KH, Maeng O, Kim TR, Lee JO, Kang KI, Kim YS, Paik SG, Lee H. Nitric oxide induces BNIP3 expression that causes cell death in macrophages. Biochem Biophys Res Commun. 2004 Aug 20;321(2):298-305.

143) Ulett GC, Adderson EE. Nitric oxide is a key determinant of group B streptococcus-induced murine macrophage apoptosis. J Infect Dis. 2005 May 15;191(10):1761-70. Epub 2005 Apr 13.

144) Marriott HM, Ali F, Read RC, Mitchell TJ, Whyte MK, Dockrell DH. Nitric oxide levels regulate macrophage commitment to apoptosis or necrosis during pneumococcal infection. FASEB J. 2004 Jul;18(10):1126-8.

145) Fettucciari K, Rosati E, Scaringi L, Cornacchione P, Migliorati G, Sabatini R, Fetriconi I, Rossi R, Marconi P. Group B Streptococcus induces apoptosis in macrophages.J Immunol. 2000 Oct 1;165(7):3923-33.)

146) Ulett GC, Bohnsack JF, Armstrong J, Adderson EE. Beta-hemolysin-independent induction of apoptosis of macrophages infected with serotype III group B streptococcus. J Infect Dis. 2003 Oct 1;188(7):1049-53. Epub 2003 Sep 15.

147) Santi I, Grifantini R, Jiang SM, Brettoni C, Grandi G, Wessels MR, Soriani M. CsrRS regulates group B Streptococcus virulence gene expression in response to environmental pH: a new perspective on vaccine development. J Bacteriol. 2009 Sep;191(17):5387-97. Epub 2009 Jun 19.

148) Holtfreter S, Kolata J, Bröker BM. Towards the immune proteome of Staphylococcus aureus - The anti-S. aureus antibody response. Int J Med Microbiol. 2010 Feb;300(2-3):176-92. Epub 2009 Nov 3.

149) Bateman A, Rawlings ND. The CHAP domain: a large family of amidases including GSP amidase and peptidoglycan hydrolases. Trends Biochem Sci. 2003 May;28(5):234-7.

150) Zou Y, Hou C. Systematic analysis of an amidase domain CHAP in 12 Staphylococcus aureus genomes and 44 staphylococcal phage genomes. Comput Biol Chem. 2010 Aug;34(4):251-7. Epub 2010 Jul 17.

151) Kajimura J, Fujiwara T, Yamada S, Suzawa Y, Nishida T, Oyamada Y, Hayashi I, Yamagishi J, Komatsuzawa H, Sugai M. Identification and molecular characterization of an N-acetylmuramyl-L-alanine amidase Sle1 involved in cell separation of Staphylococcus aureus. Mol Microbiol. 2005 Nov;58(4):1087-101.

152) Abdallah AM, Gey van Pittius NC, Champion PA, Cox J, Luirink J, Vandenbroucke-Grauls CM, Appelmelk BJ, Bitter W. Type VII secretion--mycobacteria show the way. Nat Rev Microbiol. 2007 Nov;5(11):883-91.

153) Meinke AL, Senn BM, Visram Z, Henics TZ, Minh DB, Schüler W, Neubauer C, Gelbmann D, Noiges B, Sinzinger J, Hanner M, Dewasthaly S, Lundberg U, Hordnes K, Masoud H, Sevelda P, von Gabain A, Nagy E. Immunological fingerprinting of group B streptococci: from circulating human antibodies to protective antigens. Vaccine. 2010 Oct 8;28(43):6997-7008. Epub 2010 Aug 21.

154) Ventura CL, Malachowa N, Hammer CH, Nardone GA, Robinson MA, Kobayashi SD, DeLeo FR. Identification of a novel Staphylococcus aureus two-component leukotoxin using cell surface proteomics. PLoS One. 2010 Jul 16;5(7):e11634.

155) Mereghetti L, Sitkiewicz I, Green NM, Musser JM.Extensive adaptive changes occur in the transcriptome of Streptococcus agalactiae (group B streptococcus) in response to incubation with human blood. PLoS One. 2008 Sep 4;3(9):e3143.

156) Mereghetti L, Sitkiewicz I, Green NM, Musser JM. Remodeling of the Streptococcus agalactiae transcriptome in response to growth temperature. PLoS One. 2008 Jul 30;3(7):e2785.

157) Sitkiewicz I, Green NM, Guo N, Bongiovanni AM, Witkin SS, Musser JM.Transcriptome adaptation of group B Streptococcus to growth in human amniotic fluid. PLoS One. 2009 Jul 1;4(7):e6114.

158) Kizil G, Todd I, Atta M, Borriello SP, Ait-Tahar K, Ala'Aldeen DA. Identification and characterization of TspA, a major CD4(+) T-cell- and B-cell-stimulating Neisseria-specific antigen. Infect Immun. 1999 Jul;67(7):3533-41.