

Nonclassically Secreted Proteins as Possible Antigens for Vaccine Development: A Reverse Vaccinology Approach

Mauricio de Alvarenga Mudadu · Viviane Carvalho ·
Sophie Yvette Leclercq

Received: 14 August 2014 / Accepted: 21 January 2015 /
Published online: 12 February 2015
© Springer Science+Business Media New York 2015

Abstract Reverse vaccinology strategies have already been applied to a variety of microorganisms and have contributed significantly to vaccine development. However, most of the studies focused on an individual organism or on proteins with signature sequence motifs commonly found in known secreted proteins from bacteria. In this work, we applied a reverse vaccinology strategy based on conservation, virulence, and nonclassically surface exposure criterions to identify potential antigens in two microorganisms with significant degree of genomic plasticity among isolates (*Streptococcus pneumoniae* and *Leptospira* spp.), which imposes a major limitation to the production of a multistrain component vaccine. PSORTb 3.0.2 was run to predict the subcellular localization of the proteins. OrthoMCL was run to identify groups of the most conserved proteins between strains. Virulence prediction was done for the most conserved proteins, and SecretomeP was run to predict the nonclassically secreted proteins among the potential virulence factors. Based on the above criteria, we identified 37 proteins conserved between 16 genomes of *S. pneumoniae* and 12 proteins conserved between 5 leptospiral genomes as potential vaccine candidates.

Keywords Streptococcus · *Leptospira* · Nonclassically secreted proteins · Antigens · Vaccine

Introduction

Together to the sequencing of the first genome and the development of efficient bioinformatics tools, the vaccine development concept has greatly evolved and, in the last two decades, entered the so called reverse vaccinology (RV) era. RV is an *in silico* approach that takes advantages of the many genome sequences available in the public domain. It involves the mining of these sequences using bioinformatics tools where it is possible to predict which proteins among thousands are worthy of further laboratory investigation based on several

M. de Alvarenga Mudadu
Embrapa Southeast Livestock, São Carlos, São Paulo, Brazil

V. Carvalho · S. Y. Leclercq (✉)
Research and Development Center, Ezequiel Dias Foundation (Funed), Belo Horizonte, MG, Brazil
e-mail: sophie.leclercq@funed.mg.gov.br

characteristics such as prevalence, cellular localization, role in the pathogenesis, or immunogenicity. The first studies using this approach led to the identification of several vaccine candidates for several microorganisms such as serogroup B *Neisseria meningitidis* [1], *Streptococcus pneumoniae* [2], and *Chlamydia pneumoniae* [3]. Despite its great success, RV has shown some limitations especially for microorganisms that have a high genetic heterogeneity. For example, the analysis of group B streptococcus (GBS) has shown that the genome is divided in a “core genome” that contains the genes shared by all the strains and the “dispensable genome” that probably represents the adaptation of the strain in a specific environment, absent in at least one strain [4]. This fact led to the conclusion that RV applied to the genome of only one strain would probably not be enough to identify candidates suitable for the development of a universal vaccine. It opens the door for the second phase of RV: the comparative RV, where the objective is to screen multistrain genomes to identify conserved proteins [5, 6]. Proteins with signature sequence motifs commonly found in known secreted or surface-exposed proteins from bacteria are ideal, as they have the highest chance of raising an effective immune response. As a consequence, proteins with known cytoplasmic functions are discarded from most RV studies. However, there is a body of evidence that several typical cytoplasmic proteins could appear on the surface of bacteria and have a role in pathogenesis such as adhesion, plasminogen binding, and modulation of host immune response [7, 8]. The term “moonlighting protein” is now widely used to describe these proteins which sequences containing neither known sequence motifs for surface anchoring nor identified secretion signals but appear on the bacterial surface to take on additional activities. Interestingly, some of them have already been shown as highly immunogenic and even protective in mice in several disease models [9–11].

In this work, we applied a RV strategy based on conservation, virulence, and nonclassically surface exposure criterions to two microorganisms with significant degree of genomic plasticity among isolates which imposes a major limitation to the production of a multistrain component vaccine.

S. pneumoniae is a multiserotype, Gram-positive bacterium and causes invasive infection such as meningitis and bacteremia, as well as many mucosal diseases such as pneumonia, sinusitis, and otitis, and the burden of diseases that it causes is amongst the greatest in the world [12]. The actual vaccine is a lipopolysaccharide-conjugated vaccine that protects only against the serotypes incorporated in its formulation among more than 90 existing serotypes. Its effectiveness in preventing invasive pneumococcal disease (IPD) in infants and community-acquired pneumonia (CAP) has been clearly observed. However, it has been shown that the vaccine exerts a selective pressure on the serotypes not included in the vaccine [13]. This phenomenon of serotype replacement created a necessity for the development of new-generation recombinant protein vaccines, especially based on virulent factors conserved between strains [14].

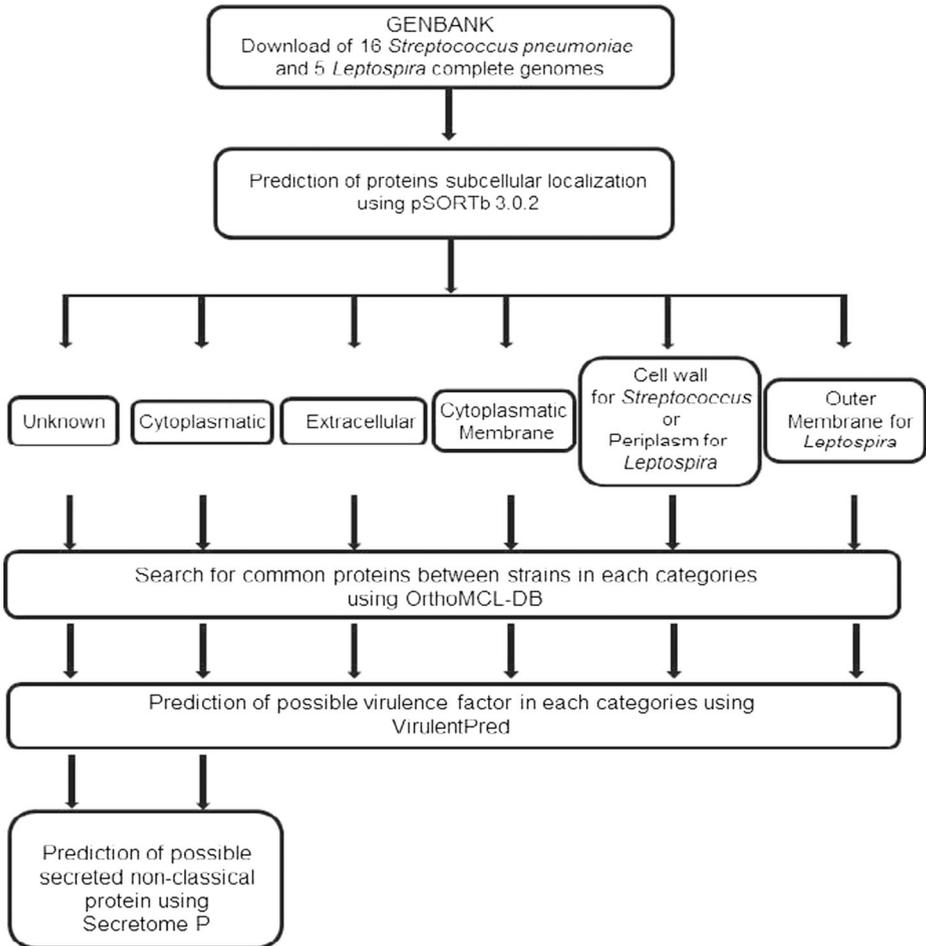
Leptospira spp. comprise a heterogeneous group of pathogenic (and saprophytic) species belonging to the order Spirochaetes that cause fever, chills, headache, and severe myalgia in the early phase of the disease. Progression to multiorgan system complications occurs in 5 to 15 % of cases with mortality rates of 5 to 40 % [15]. It has a global distribution with a higher incidence in the tropics and subtropics, ranging from 10 to 100 human cases per 100,000 individuals. Leptospirosis has also a great economic impact in the agricultural industry since the disease affects the livestock inducing abortions, infertility, reduced milk production, and death [16]. Currently available veterinarian vaccines, based on inactivated whole bacteria, are of low efficacy and do not confer cross-protective immunity against the large number of pathogenic serovars (>200). A vaccine licensed for human is still required [17].

In our current study, we identified 37 proteins conserved between 16 genomes of *S. pneumoniae* and 12 proteins conserved between 5 leptospiral genomes using an in silico

analysis. Among them, 19 and 7 proteins, respectively, are potential nonclassically surface-exposed proteins and could represent new antigens for vaccine development.

Methodology

The sequences of steps applied in this work are represented in the following workflow.



Sequence Retrieval

The complete genomes of 16 *S. pneumoniae* strains representative of 11 different serotypes and 5 *Leptospira* strains were downloaded from NCBI. The proteomes of all strains used in the analyses were obtained by retrieving the coding sequence (CDS) protein sequences from the genome GenBank files. Tables 1 and 2 show the accession number, the strain, the serotype/serovar, and the number of protein sequences retrieved for each one.

Subcellular Localization Prediction

PSORTb v.3.0 is the most commonly used software to predict the localization of a protein in prokaryotes. It uses a combination of six analytical modules, each of which analyzes one biological feature known to influence the subcellular localization of a protein, such as the signal peptide for exportation, transmembrane alpha helices, motifs that show characteristics of specific functions, and others [18]. Each module returns a likelihood score (between 1 and 10) of a protein being at a specific localization, and a cutoff of 7.5 is considered reliable to assign a single localization. If the score is below this threshold, the protein can be considered with unknown or multiple localizations. In this study, the software was run choosing “bacteria” as the “organism” and “Gram positive” was chosen in the “Gram stain” option for *S. pneumoniae* and “Gram negative” for *Leptospira*. The prediction of the localization was performed for all CDS proteins from the complete genomes described above. The output for each genome was separated in five files according to the localization.

Search for the Most Conserved Proteins Between Strains

OrthoMCL is an algorithm that creates groups of homologous proteins based on their sequence similarity [19]. OrthoMCL is based on the reciprocal BLAST best hits identified by pairwise alignments of proteomes. We decided to use OrthoMCL since it provides a good method for the identification of true orthologs among all genomes of the same genera (groups of 16 proteins for *Streptococcus* and 5 proteins for *Leptospira*). To ascertain OrthoMCL to select more conserved sequences, we added an additional BLAST filter to select only alignments with ≥ 98 % similarity. The protein groups identified are suggested to be most conserved among strains and then to be wider effect vaccine candidates—supposedly more effective against all strains simultaneously.

Table 1 Accession numbers and characteristics of the 16 *S. pneumoniae* genomes

Accession numbers	Strain	Serotype	Number of proteins
NC_003028	TIGR4	4	2.105
NC_008533	D39	2	1.914
NC_010380	Hungary19A	19A	2.155
NC_010582	CGSP14	14	2.206
NC_011072	G54	19 F	2.037
NC_011900	ATCC 700669	23 F	1.990
NC_012467	P1031	1	2.073
NC_012468	70585	5	2.202
NC_012469	Taiwan19 F-14	19 F	2.044
NC_014494	AP200	11A	2.216
NC_014498	670-6b	6B	2.352
NC_017591	inv104	1	1.820
NC_017592	OXC141	3	1.823
NC_017593	inv200	14	1.929
NC_017769	st556	19 F	2.148
NC_018594	spna45	3	1.926

Table 2 Accession numbers and characteristics of the five *Leptospira* genomes used (splitted by chromosomes)

Accession numbers	Specie	Serovar	Strain	Number of proteins
NC_005823.1 (Chr 1)	<i>L. interrogans</i>	Copenhageni	Fiocruz L1-130	3.399
NC_005824.1 (Chr 2)	<i>L. interrogans</i>	Copenhageni	Fiocruz L1-130	268
NC_004342.2 (Chr 1)	<i>L. interrogans</i>	Lai	str 56601	3.390
NC_004343.2 (Chr 2)	<i>L. interrogans</i>	Lai	str 56601	293
NC_017551.1 (Chr 1)	<i>L. interrogans</i>	Lai	str IPAV	3.418
NC_017552.1 (Chr 2)	<i>L. interrogans</i>	Lai	str IPAV	293
NC_008508.1 (Chr 1)	<i>L. borgpetersenii</i>	Hardjo-bovis	L550	2.703
NC_008509.1 (Chr 2)	<i>L. borgpetersenii</i>	Hardjo-bovis	L550	242
NC_008510.1 (Chr 1)	<i>L. borgpetersenii</i>	Hardjo-bovis	JB197	2.645
NC_008511.1 (Chr 2)	<i>L. borgpetersenii</i>	Hardjo-bovis	JB197	23

Virulence Factor Prediction

VirulentPred is software that predicts the virulence of bacterial proteins based on a machine learning classification method known as bilayer cascade Support Vector Machine (SVM) [20]. The first-layer SVM classifiers were trained and optimized with different individual protein sequence features. In addition, a similarity search-based module was also developed using a dataset of virulent and nonvirulent proteins as BLAST database. The results from the first layer (SVM scores and PSI-BLAST result) were cascaded to the second-layer SVM classifier to train and generate the final classifier. In this work, we used the output of the OrthoMCL groups to search for possible virulent factors between the more conserved proteins using a threshold ≥ 1 to minimize the occurrence of false positive.

Nonclassically Secreted Protein Prediction

SecretomeP is software that runs an algorithm based on artificial neural networks that predicts nonclassically secreted protein in bacteria or mammalian cells [21]. The prediction method assigns a score to each protein between 0 and 1 where a score above 0.5 is considered indicative of secretion. Nonclassically secreted proteins should obtain a score exceeding the normal threshold of 0.5 and should not simultaneously get a prediction of containing a signal peptide. Here, we used the list of virulent proteins predicted to be “cytoplasmic” or “unknown” as an entry for the SecretomeP.

Results

Prediction of Subcellular Localization and Selection of the Most Conserved Proteins

After running PSORTb 3.0 and comparing the degree of conservation between proteins with OrthoMCL according to the localization, 970 and 63 sequences were selected with ≥ 98 % of similarity between strains of *S. pneumoniae* and *Leptospira*, respectively (Tables 3 and 4).

Table 3 Number of conserved proteins among *S. pneumoniae* genomes regarding the localization

Number of genomes	C	CW	CM	E	U
16	617	8	264	6	75
≥15	728	11	326	9	104
≥14	770	12	352	10	125

C cytoplasm, CW cell wall, CM cytoplasmic membrane, E extracellular, U unknown

Prediction of Virulence and Nonclassically Secreted Proteins

For further investigation, we selected the protein sequences that were conserved among all the genomes analyzed, to narrow the panel of proteins to be screened. All proteins identified in this step were screened for potential role in virulence using the VirulentPred software. One hundred proteins conserved among the 16 *S. pneumoniae* strains were found to be putative virulent factors (data not shown). Among them, 75 were found as cytoplasmic localized by PSORT 3.0. These proteins are usually discarded from further studies, as they are not expected to be presented to the host immune system. However, some proteins have already been detected on the surface of microorganisms even when they do not possess classical peptide signal or surface exposure motifs. For this reason, we searched for nonclassically secreted proteins in the categories unknown and cytoplasmic using SecretomeP. We found 19 putative nonclassically secreted proteins. In total, 37 proteins were considered possible virulence factors, surface-exposed and conserved between the 16 *Streptococcus* genomes (Table 5).

Twenty-six proteins were found to be potential virulent factors conserved between the five *Leptospira* strains, especially flagellar proteins and ribosomal proteins (data not shown). After screening for nonclassically secreted proteins in the categories unknown and cytoplasmic using SecretomeP, we found seven putative nonclassically secreted proteins, all but one are ribosomal proteins. In total, 12 proteins were considered possible virulence factors, surface-exposed and conserved between the five *Leptospira* genomes (Table 6).

Discussion

RV has already been used for studying *Streptococcus* and *Leptospira* spp. However, these studies focused on an individual organism or in proteins with signature sequence motifs commonly found in known secreted proteins from bacteria [2, 22]. The novelty of the present study lies in the comparison of entire genomes and the incorporation of nonclassically secreted proteins in the analysis. Using this approach, we identified 37 proteins conserved between 16 genomes of *S. pneumoniae* and 12 proteins conserved between five leptospiral genomes potentially exposed on the surface of the bacteria and with a possible role in virulence.

Table 4 Number of conserved proteins among *Leptospira* genomes regarding the localization

Number of genomes	C	OM	CM	E	U	P
5	43	0	5	1	10	4
≥4	43	0	5	1	13	4
≥3	1545	57	598	46	793	31

C cytoplasm, OM outer membrane, CM cytoplasmic membrane, E extracellular, U unknown, P periplasm

Table 5 Possible virulent factors, surface-exposed and conserved between the 16 genomes of *Streptococcus pneumoniae* regarding the initial localization by PSORT 3.0

Accession number	Protein	Localization
NP_345553.1	Hypothetical protein	E
YP_873926.1	Cell wall surface	CW
NP_345111.1	Transmembrane Vexp1	CM
NP_345112.1	ABC transporter ATP-binding	CM
NP_345113.1	Transmembrane Vexp3	CM
NP_345146.1	ABC transporter ATP-binding	CM
NP_345252.1	ABC transporter ATP-binding	CM
NP_345455.1	Translocase subunit SecG	CM
NP_345479.1	Adhesion lipoprotein	CM
NP_345510.1	Iron ABC transporter	CM
NP_345815.1	ABC transporter/permease	CM
NP_345823.1	Glycosyl transferase	CM
NP_345826.1	psr protein	CM
NP_345999.1	Cation efflux family	CM
NP_346050.1	Glycosyl transferase	CM
NP_346072.1	Sensor histidine kinase	CM
NP_346344.1	PAP2 family protein	CM
NP_346435.1	Translocase subunit SecE	CM
NP_346504.1	Phosphate ABC transporter	U
NP_346626.1	MreC	U
NP_345126.1	Tributyryn esterase	U
NP_344809.1	Ribosomal protein S12	C
NP_344870.1	Cell division protein FtsL	C
NP_344909.1	Valonate decarboxylase	C
NP_344962.1	Ribosomal protein L28	C
NP_345019.1	Transcriptional regulator MerR	C
NP_345329.1	Ribosomal protein S20	C
NP_345441.1	Ribosomal protein L35	C
NP_345578.1	Ribosomal protein L27	C
NP_345752.1	DNA binding protein	C
NP_345872.1	Ribosomal protein S21	C
NP_345887.1	Methyltransferase	C
NP_346552.1	Ribosomal protein L32	C
NP_346553.1	Ribosomal protein L33	C
NP_345279.1	Hypothetical protein	C
NP_345723.1	Hypothetical protein	C
NP_346048.1	Hypothetical protein	C

E extracellular, *U* unknown, *CM* cytoplasmic membrane, *C* cytoplasmic

Among these vaccine candidates, none matched with the ones described in the pioneer RV studies [2, 22] due to our different focus. In fact, as shown in Tables 3 and 4, the number of conserved proteins increased when the number of compared genomes decreased. When we

Table 6 Possible virulent factors, surface-exposed and conserved between the five genomes of *Leptospira* regarding the initial localization by PSORT 3.0

Accession number	Protein	Localization
NP_712848.2	Flagellar basal rod protein (FlgG)	E
NP_712198.1	Endoflagellar filament core protein	P
NP_712200.1	Flagellin protein	P
NP_712598.2	Endoflagellar hook-filament protein (FlgL)	P
NP_712262.1	Flagellar motor switch protein (FliM)	CM
NP_710628.1	Hypothetical protein	U
NP_710356.1	Ribosomal protein L34	C
NP_711032.1	Ribosomal protein L27	C
NP_711201.1	Ribosomal protein L31	C
NP_711424.1	Ribosomal protein L35	C
NP_711444.1	Ribosomal protein L32	C
NP_713597.1	Ribosomal protein S12	C

E extracellular, *OM* outer membrane, *P* periplasm, *U* unknown, *CM* cytoplasmic membrane, *C* cytoplasmic

looked at the potential candidate antigens present in less than 16 or 5 genomes, respectively, or with less than 98 % similarity, proteins such as choline-binding protein and signal peptidase were identified for *Streptococcus* and LipL22 and LipL23 for *Leptospira* matching the previous studies (result not shown). In the same way, we also identified pullulanase which was indicated as a promising vaccine candidate in a recent RV approach based on conservation, immunogenicity, and human proteome similarity of some streptococcal selected proteins [23]. All together, these observations support the reliability of our results.

Interaction between cell and their environment is critically mediated by surface proteins. In the case of pathogenic bacteria, these molecules are often virulence factors. In both microorganisms, we identified some expected class of virulent factors especially localized in the cytoplasmic membrane or periplasm. For example, ABC transporter ATP-binding protein and psr protein transcription regulator have been recognized as *Streptococcus pneumoniae* virulent factors in lung infection [24, 25]. Inactivation of some flagellar proteins has proved their importance in the bacteria mobility and virulence of *Leptospira interrogans* [26, 27]. Furthermore, some flagellins have been shown to be immunogenic and even protective [28, 29]. Mutation in the gene of the flagellar protein FliM has been related to *Helicobacter pylori* pathogenicity by reducing either the ability of the bacteria to attach to gastric epithelial cells or the intensity of bacteria–host cell interactions [30]. Thus, these studies indicated that these classes of proteins could be evaluated to incorporate a vaccine formulation.

Nevertheless, proteins with more than one transmembrane domain are generally difficult to obtain as recombinant one and are theoretically more embedded under the surface, and then their use for vaccine development could be more tedious. Notably, we identified few or none conserved virulence factors localized in the cell wall, outer membrane, or secreted. This kind of protein is more likely to be variable, part of the dispensable genome that confers a selective advantage in specific environmental conditions, whereas the cytoplasmic proteins are more likely to be part of the core genome that mainly encodes factors for functions that contribute to the major metabolic pathways.

However, as it has been long shown that some cytoplasmic proteins can be displayed at the surface of cell, we focused on searching for nonclassically secreted protein. The so-called

moonlighting proteins are proteins with an additional biological activity when in a different location that they normally occupy. Most of the moonlighting proteins in bacteria have been primarily identified as glycolytic enzymes, other metabolic enzymes, or molecular chaperones, normally classified as cytoplasmic. The most classical examples were GAPDH and enolase that have been shown to have a role in adhesion to host epithelial components and plasminogen or induce a modulation of the host immune response. Thus, it has been suggested that moonlighting proteins contribute to bacterial virulence [7, 8].

In *Streptococcus*, we identified proteins such as phosphatase ABC transporter, tributyrin esterase, and methyltransferase already known to be involved in bacterial pathogenesis [24, 31]. But, curiously, our pipeline returned, as nonclassically secreted protein, a high number of ribosomal proteins that were not, at first glance, the best candidate for vaccine development. However, ribosomal proteins have been found in the exoproteome of many bacteria [32–34] and mounting evidence points to their alternative extracellular location where they might perform nonribosomal function. Interestingly, Ogunniyi and colleagues (2012) [25] demonstrated that the ribosomal proteins L29 and L33 were upregulated in *S. pneumoniae* after infection in the lung, and the ribosomal protein L33 was identified by our pipeline. Furthermore, ribosomal proteins have already shown to be immunogenic and sometimes protective. For example, in a previous study, we showed that the pneumococcal ribosomal protein S9 decreased the bacteremia in mice 24 h after challenge in a model of sepsis [35]. In another work, Moffitt et al. (2012) [36] identified several pneumococcal antigens from the soluble fraction of a killed whole-cell vaccine. Among them, the ribosomal protein S1 was a potent inducer of Th17 and was able to statistically reduce the colonization of nasal mucosal tissue of immunized mice compared with the control group in a model of nasopharyngeal carriage. Then, it is reasonable to think that these proteins might have a protective role in other bacteria, too.

A new prediction tool, named Jenner-predict server, has recently been developed to predict vaccine candidate in bacteria based on host–pathogen interactions [37]. The performance of web server was evaluated against known protective antigens from diverse classes of bacteria reported in Protegen database and datasets used for VaxiJen server development and, then, was applied to *S. pneumoniae* and *Escherichia coli* proteomes. Although the authors state that the prediction accuracy was better than other previously developed tools, it is based on the first premise that the candidate should be classified as “noncytosolic” by the software PSORT 3.0. Thus, none of the cytoplasmic candidates predicted by our method, in *S. pneumoniae*, were pointed out by the Jenner server. Several known candidates such as pneumolysin, surface protein PspA, choline-binding family protein known to be extracellular, or cell wall component were identified by Jaiswal and collaborators but not by our method. The fact that we did not identify these classical candidates could be explained by our more stringent criterions with respect to the conservation between strains. Among the proteins classified as “cytoplasmic membrane,” only two, an iron ABC transporter and an adhesion lipoprotein, were simultaneously identified by the Jenner server and our method probably due to the number of transmembrane helices. Jaiswal and collaborators used less than three membrane domains as a cutoff. Interestingly, the tributyrin esterase classified as unknown was identified in common by both our method and Jenner-predict server. Furthermore, this protein has been identified as essential for lung infection [24], indicating that this protein could be a strong new candidate for vaccine development against *S. pneumoniae*.

Among the possible limitations of our method is the election of false vaccine candidates, due to the accuracies of the software used, which are not optimal. Thus, biological confirmation is necessary. Furthermore, our very stringent criterion could explain why we found principally ribosomal proteins. And, even if these proteins are really exposed and immunogenic, one can

concern about their possible homology with human proteins. We performed sequence alignment of the unknown and cytoplasmic protein categories identified here with the human genome using Basic Local Alignment Search Tool. In average, the sequence identity varies between 22 and 53 %. Notably, the ribosomal proteins L32, L33, and L35; one hypothetical protein of *S. pneumoniae*; the ribosomal protein L35; and the hypothetical protein of *Leptospira* ssp. have no homology with the human genome (results not shown).

The results obtained with *Leptospira* genomes were more restricted and less interesting than with *Streptococcus*. This could be due to the fact that we compared two different species *L. interrogans* and *Leptospira borgpetersenii* instead of several serotypes of the same species and indicated that finding common antigens to develop a broad vaccine is more unlikely to occur. In conclusion, our pipeline gave us the possibility to identify, besides the canonical virulence factors, cytoplasmic proteins with putative extracellular location. They might be new moonlighting proteins, and we think that this kind of proteins should not be neglected as possible subunit vaccine candidates. Some of them are already under investigation in our laboratory to further confirm their surface localization and immunogenicity.

References

- Pizza, M., Scarlato, V., Masignani, V., Giuliani, M. M., Aricò, B., Comanducci, M., et al. (2000). Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science*, *287*, 1816–1820.
- Wizemann, T. M., Heinrichs, J. H., Adamou, J. E., Erwin, A. L., Kunsch, C., Choi, G. H., et al. (2001). Use of a whole genome approach to identify vaccine molecules affording protection against Streptococcus pneumoniae infection. *Infection and Immunity*, *69*, 1593–1598.
- Montigiani, S., Falugi, F., Scarselli, M., Finco, O., Petracca, R., Galli, G., et al. (2002). Genomic approach for analysis of surface proteins in Chlamydia pneumoniae. *Infection and Immunity*, *70*, 368–379.
- Tettelin, H., Masignani, V., Cieslewicz, M. J., Eisen, J. A., Peterson, S., Wessels, M. R., et al. (2002). Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V Streptococcus agalactiae. *PNAS*, *99*, 12391–12396.
- Maione, D., Margarit, I., Rinaudo, C. D., Masignani, V., Mora, M., Scarselli, M., et al. (2005). Identification of a universal group B streptococcus vaccine by multiple genome screen. *Science*, *309*, 148–150.
- Giuliani, M. M., Adu-Bobie, J., Comanducci, M., Arico, B., Savino, S., Santini, L., et al. (2006). A universal vaccine for serogroup B meningococcus. *Proceedings of the National Academy of Sciences*, *103*, 10834–10839.
- Henderson, B., & Martin, A. (2011). Bacterial virulence in the moonlight: multitasking bacterial moonlighting proteins are virulence determinants in infectious disease. *Infection and Immunity*, *79*, 3476–3491.
- Wang, G., Xia, Y., Cui, J., Gu, Z., Song, Y., Chen, Y. Q., et al. (2013). The roles of moonlighting proteins in bacteria. *Current Issues in Molecular Biology*, *16*, 15–22.
- Ling, E., Feldman, G., Portnoi, M., Dagan, R., Overweg, K., Mulholland, F., & Chalifa-caspi, V. (2004). Glycolytic enzymes associated with the cell surface of *Streptococcus pneumoniae* are antigenic in humans and elicit protective immune responses in the mouse. *Clinical and Experimental Immunology*, *138*, 290–298.
- Kositanont, U., Saetun, P., Krittanai, C., Doungchawee, G., Tribuddharat, C., & Thongboonkerd, V. (2007). Application of immunoproteomics to leptospirosis: towards clinical diagnostics and vaccine discovery. *Proteomics Clinical Applications*, *1*, 400–409.
- Li, W., Hu, X., Zhang, X., Ge, Y., Zhao, S., Hu, Y., & Ashman, R. B. (2011). Immunization with the glycolytic enzyme enolase confers effective protection against *Candida albicans* infection in mice. *Vaccine*, *29*, 5526–5533.
- Henriques-Normark, B., & Tuomanen, E. I. (2013). The pneumococcus: epidemiology, microbiology, and pathogenesis. *Cold Spring Harbor Perspectives in Medicine*, *3*, a010215.
- Falup-Pecurariu, O. (2012). Lessons learnt after the introduction of the seven valent pneumococcal conjugate vaccine toward broader spectrum conjugate vaccines. *Biomedical Journal*, *35*, 450–456.
- Feldman, C., & Anderson, R. (2014). Review: current and new generation pneumococcal vaccines. *Journal of Infection*. In press.
- Lim, V. K. E. (2011). Leptospirosis: a re-emerging infection. *Malaysian Journal of Pathology*, *33*, 1–5.

16. Grooms, D. L. (2006). Reproductive losses caused by bovine viral diarrhoea virus and leptospirosis. *Theriogenology*, *66*, 624–628.
17. Verma, R., Khanna, P., & Chawla, S. (2013). Whole-cell inactivated vaccines: future prospects. *Human Vaccines & Immunotherapeutics*, *9*, 763–765.
18. Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., et al. (2010). PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics*, *26*, 1608–1615.
19. Fischer, S., Brunk, B. P., Chen, F., Gao, X., Harb, O. S., Iodice, J. B., et al. (2011). Using OrthoMCL to assign proteins to OrthoMCL-DB or to cluster proteomes into new ortholog groups. *Current Protocols in Bioinformatics*, *35*, 6.12.1–6.12.19.
20. Garg, A., & Gupta, D. (2008). VirulentPred: a SVM based prediction method for virulent proteins in bacterial pathogens. *BMC Bioinformatics*, *9*, 62–74.
21. Bendtsen, J. D., Kiemer, L., Fausbøll, A., & Brunak, S. (2005). Non-classical protein secretion in bacteria. *BMC Microbiology*, *5*, 58–71.
22. Gamberini, M., Gómez, R. M., Atzingen, M. V., Martins, E. A., Vasconcellos, S. A., Romero, E. C., et al. (2005). Whole-genome analysis of *Leptospira interrogans* to identify potential vaccine candidates against leptospirosis. *FEMS Microbiology Letters*, *244*, 305–313.
23. Talukdar, S., Zutchi, S., Prashanth, K. S., Saikia, K. K., & Kumar, P. (2014). Identification of potential vaccine candidates against *Streptococcus pneumoniae* by reverse vaccinology approach. *Applied Biochemistry and Biotechnology*, *172*, 3026–3041.
24. Hava, D. L., & Camilli, A. (2002). Large-scale identification of serotype 4 *Streptococcus pneumoniae* virulence factors. *Molecular Microbiology*, *45*, 1389–1406.
25. Ogunniyi, A. D., Mahdi, L. K., Trappetti, C., Verhoeven, N., Mermans, D., Van der Hoek, M. B., et al. (2012). Identification of genes that contribute to the pathogenesis of invasive pneumococcal disease by in vivo transcriptomic analysis. *Infection and Immunity*, *80*, 3268–3278.
26. Liao, S., Sun, A., Ojcius, D. M., Wu, S., Jinfang Zhao, J., & Yan, J. (2009). Inactivation of the *fliY* gene encoding a flagellar motor switch protein attenuates mobility and virulence of *Leptospira interrogans* strain Lai. *BMC Microbiology*, *9*, 253–263.
27. Lambert, A., Picardeau, M., Haake, D. A., Sermswan, R. W., Srikram, A., Ben Adler, A., & Murray, G. A. (2012). FlaA proteins in *Leptospira interrogans* are essential for motility and virulence but are not required for formation of the flagellum sheath. *Infection and Immunity*, *80*, 2019–2025.
28. Dai, B., You, Z., Chen, Z., Yan, H., & Fang, Z. (2000). Protection against leptospirosis by immunization with plasmid DNA encoding 33 kDa endoflagellin of *L. interrogans* serovar lai. *Chinese Medical Sciences Journal*, *15*, 14–19.
29. Sakolvaree, Y., Maneewatch, S., Jiemsup, S., Klaysing, B., Tongtawe, P., Srimanote, P., et al. (2007). Proteome and immunome of pathogenic *Leptospira* spp. revealed by 2DE and 2DE-immunoblotting with immune serum. *Asian Pacific Journal of Allergy and Immunology*, *25*, 53–73.
30. Zhang, Z. W., Dorrell, N., Wren, B. W., & Farthing, M. J. (2002). *Helicobacter pylori* adherence to gastric epithelial cells: a role for non-adhesin virulence genes. *Journal of Medical Microbiology*, *51*, 495–502.
31. Lamarche, M. G., Wanner, B. L., Crépin, S., & Harel, J. (2008). The phosphate regulon and bacterial virulence: a regulatory network connecting phosphate homeostasis and pathogenesis. *FEMS Microbiology Review*, *32*, 461–473.
32. Gieffing, C., Meinke, A. L., Hanner, M., Henics, T., Bui, M. D., Gelbmann, D., et al. (2008). Discovery of a novel class of highly conserved vaccine antigens using genomic scale antigenic fingerprinting of pneumococcus with human antibodies. *Journal of Experimental Medicine*, *205*, 117–131.
33. Mendum, T. A., Newcombe, J., McNeilly, C. M., & McFadden, J. (2009). Towards the immunoproteome of *Neisseria meningitidis*. *PLOS One*, *4*, e5940.
34. Bergh, P. V., Heller, M., Braga-Lagache, S., & Frey, J. (2013). The *Aeromonas salmonicida* subsp. *Salmonicida* exoproteome: global analysis, moonlighting proteins and putative antigens for vaccination against furunculosis. *Proteome Science*, *11*, 44–56.
35. Araújo, H. F., Campos, P. C., Camargo, D. R. A., Pereira, F. N. R., Samuel, M. L., Oliveira, M. A., et al. (2012). Immune response and protective efficacy of S9 ribosomal protein of *Streptococcus pneumoniae* in a model of sepsis. *Canadian Journal of Microbiology*, *58*, 1055–1062.
36. Moffitt, K. L., Malley, R., & Lu, Y. J. (2012). Identification of protective pneumococcal T(H)17 antigens from the soluble fraction of a killed whole cell vaccine. *PLoS One*, *7*, e43445.
37. Jaiswal, V., Chanumolu, S. K., Gupta, A., Chauhan, R. S., & Rout, C. (2013). Jenner-predict server: prediction of protein vaccine candidates (PVCs) in bacteria based on host-pathogen interactions. *BMC Bioinformatics*, *14*, 211–222.