

# Protectome Analysis: A New Selective Bioinformatics Tool for Bacterial Vaccine Candidate Discovery<sup>§</sup>

Emrah Altindış, Roberta Cozzi, Benedetta Di Palo, Francesca Necchi, Ravi P. Mishra, Maria Rita Fontana, Marco Soriani, Fabio Bagnoli, Domenico Maione, Guido Grandi, and Sabrina Liberatori

New generation vaccines are in demand to include only the key antigens sufficient to confer protective immunity among the plethora of pathogen molecules. In the last decade, large-scale genomics-based technologies have emerged. Among them, the Reverse Vaccinology approach was successfully applied to the development of an innovative vaccine against *Neisseria meningitidis* serogroup B, now available on the market with the commercial name BEXSERO® (Novartis Vaccines). The limiting step of such approaches is the number of antigens to be tested in *in vivo* models. Several laboratories have been trying to refine the original approach in order to get to the identification of the relevant antigens straight from the genome. Here we report a new bioinformatics tool that moves a first step in this direction. The tool has been developed by identifying structural/functional features recurring in known bacterial protective antigens, the so called "Protectome space," and using such "protective signatures" for protective antigen discovery. In particular, we applied this new approach to *Staphylococcus aureus* and Group B *Streptococcus* and we show that not only already known protective antigens were re-discovered, but also two new protective antigens were identified. *Molecular & Cellular Proteomics* 14: 10.1074/mcp.M114.039362, 418–429, 2015.

Although vaccines based on attenuated pathogens as pioneered by Luis Pasteur have been shown to be extremely effective, safety and technical reasons recommend that new generation vaccines include few selected pathogen components which, in combination with immunostimulatory molecules, can induce long lasting protective responses. Such approach implies that the key antigens sufficient to confer protective immunity are singled out among the plethora of

pathogen molecules. As it turns out, the search for such protective antigens can be extremely complicated.

Genomic technologies have opened the way to new strategies in vaccine antigen discovery (1, 2, 3). Among them, Reverse Vaccinology (RV)<sup>1</sup> has proved to be highly effective, as demonstrated by the fact that a new Serogroup B *Neisseria meningitidis* (MenB) vaccine, incorporating antigens selected by RV, is now available to defeat meningococcal meningitis (4, 5). In essence, RV is based on the simple assumption that cloning all annotated proteins/genes and screening them against a robust and reliable surrogate-of-protection assay must lead to the identification of all protective antigens. Because most of the assays available for protective antigen selection involve animal immunization and challenge, the number of antigens to be tested represents a severe bottleneck of the entire process. For this reason, despite the fact that RV is a brute force, inclusive approach ("test-all-to-lose-nothing" type of approach) in their pioneered work of MenB vaccine discovery, Pizza and co-workers did not test the entire collection of MenB proteins but rather restricted their analysis to the ones predicted to be surface-localized. This was based on the evidence that for an anti-MenB vaccine to be protective bactericidal antibodies must be induced, a property that only surface-exposed antigens have. For the selection of surface antigens Pizza and co-workers mainly used PSORT and other available tools like MOTIFS and FINDPATTERNS to find proteins carrying localization-associated features such as transmembrane domains, leader peptides, and lipobox and outer membrane anchoring motifs. At the end, 570 proteins were selected and entered the still very labor intensive screening phase. Over the last few years, our laboratories have been

From the ‡Research Department, Novartis Vaccines and Diagnostics, 53100 Siena, Italy

Received, March 21, 2014 and in revised form, October 31, 2014

Published, MCP Papers in Press, November 3, 2014, DOI 10.1074/mcp.M114.039362

Author contributions: G.G. and S.L. designed research; E.A., R.C., B.D., F.N., R.P.M., M.F., and S.L. performed research; E.A., R.C., B.D., F.N., R.P.M., M.F., M.S., F.B., D.M., G.G., and S.L. analyzed data; E.A., M.S., F.B., D.M., G.G., and S.L. wrote the paper.

<sup>1</sup> The abbreviations used are: RV, Reverse Vaccinology; 5' NT, 5' nucleotidase; AMP, adenosine monophosphate; BCA, Bicinchoninic acid; CFU, Colony Forming Unit; DUF, Domain of Unknown Function; fHbp, Meningococcal Factor H binding protein; GAS, Group A *Streptococcus*; GBS, Group B *Streptococcus*; IPTG, Isopropyl β-D-1-thiogalactopyranoside; LD<sub>50</sub>, 90% lethal dosage; MenB, *Neisseria meningitidis* serogroup B, *Meningococcus* B; ORF, Open Reading Frame; PA, Protective Antigen; PCR, Polymerase Chain Reaction; PS, Protective Signature; PT, Pertussis Toxin; SLO, Streptolysin O from *Streptococcus pyogenes*.

trying to move to more selective strategies. Our ultimate goal, we like to refer to as the “Holy Grail of Vaccinology,” is to identify protective antigens by “simply” scanning the genome sequence of any given pathogen, thus avoiding time consuming “wet science” and “move straight from genome to the clinic” (6).

With this objective in mind, we have developed a series of proteomics-based protocols that, in combination with bioinformatics tools, have substantially reduced the number of antigens to be tested in the surrogate-of-protection assays (7, 8). In particular, we have recently described a three-technology strategy that allows to narrow the number of antigens to be tested in the animal models down to less than ten (9). However, this strategy still requires high throughput experimental activities. Therefore, the availability of *in silico* tools that selectively and accurately single out relevant categories of antigens among the complexity of pathogen components would greatly facilitate the vaccine discovery process.

In the present work, we describe a new bioinformatics approach that brings an additional contribution to our “from genome to clinic” goal. The approach has been developed on the basis of the assumption that protective antigens are protective in that they have specific structural/functional features (“protective signatures”) that distinguish them from immunologically irrelevant pathogen components. These features have been identified by using existing databases and prediction tools, such as PFam and SMART. Our approach focuses on protein biological role rather than its localization: it is completely protein localization unbiased, and lead to the identification of both surface-exposed and secreted antigens (which are the majority in extracellular bacteria) as well as cytoplasmic protective antigens (for instance, antigens that elicit interferon  $\gamma$  producing CD4<sup>+</sup> T cells, thus potentiating the killing activity of phagocytic cells toward intracellular pathogens). Should these assumptions be valid, PS could be identified if: (1) all known protective antigens are compiled to create what we refer to as “the Protectome space,” and (2) Protectome is subjected to computer-assisted scrutiny using selected tools. Once signatures are identified, novel protective antigens of a pathogen of interest should be identifiable by scanning its genome sequence in search for proteins that carry one or more protective signatures. A similar attempt has been reported (10), where the discrimination of protective antigens *versus* nonprotective antigens was tried using statistical methods based on amino acid compositional analysis and auto cross-covariance. This model was implemented in a server for the prediction of vaccine candidates, that is, Vaxijen ([www.darrenflower.info/Vaxijen](http://www.darrenflower.info/Vaxijen)); however, the selection criteria applied are still too general leading to a list of candidates that include ca. 30% of the total genome ORFs very similarly to the number of antigens predicted by classical RV based on the presence of localization signals.

Here we show that Protectome analysis unravels specific signatures embedded in protective antigens, most of them

related to the biological role/function of the proteins. These signatures narrow down the candidate list to ca. 3% of the total ORFs content and can be exploited for protective antigen discovery. Indeed, the strategy was validated by demonstrating that well characterized vaccine components could be identified by scanning the genome sequence of the corresponding pathogens for the presence of the PS. Furthermore, when the approach was applied to *Staphylococcus aureus* and *Streptococcus agalactiae* (Group B *Streptococcus*, GBS) not only already known protective antigens were rediscovered, but also two new protective antigens were identified.

## EXPERIMENTAL PROCEDURES

**Meta-analysis of Bacterial Protective Antigens and Selection of Protectome Candidate Antigens**—Protein sequences corresponding to known bacterial protective antigens and whole proteome set of MenB, GBS, *S. aureus*, *Bordetella pertussis*, and *Streptococcus pyogenes* were downloaded from the Uniprot Knowledgebase (11) and analyzed by BLAST and SMART (12, 13). Each protein antigen was annotated with the relevant features, *i.e.* specific Pfam domains (14), multiple internal repeats, as predicted by PROSPERO (15), and genus- or species-specificity. Taxonomically-restricted genes were identified by BLAST analysis and the cut off values to define specific ORFs were set to an *e*-value  $> 10e^{-5}$  and to a sequence coverage  $< 50\%$ . Multiple sequence alignments were carried out by using ClustalW (16).

**Cloning, Expression, and Purification of Selected GBS Proteins**—GBS 2603 V/R strain was used as source of DNA for amplification of selected Protectome antigens. Gene coding for the protein SAG1333 were cloned as C-terminal His-tag fusion protein and then expressed and purified as already reported (3). PCR primers were designed to amplify the complete gene without predicted signal peptide coding sequences covering the amino acid sequence 28–663. PCR products were cloned by using the Polymerase Incomplete Primer Extension (PIPE) method (17). HK100 competent cells were transformed with PCR products (I-PCR) immediately following amplification with the V-PCR of SpeedET vector (C-term 6xHis tag). Expression was obtained maintaining the cultures at 25 °C for 4h after the induction using arabinose 0.2%. Cells were harvested by centrifugation and lysed in “B-PER buffer” (Pierce) containing lysozyme 1 mg/ml, DNase 0.5 mg/ml, and COMPLETE inhibitors mixture (Roche). The cell-lysate was clarified by centrifugation and applied onto His-Trap HP column (Amersham Biosciences) pre-equilibrated in buffer containing 10 mM imidazole. Protein elution was performed using an imidazole gradient. Protein concentration was estimated using BCA assay (Pierce).

**Cloning, Expression, and Purification of Selected *S. aureus* Proteins**—*S. aureus* NCTC8325 strain was used as source of DNA for amplification. Gene coding for the protein SAOUHSC\_00427, were cloned as N-terminal His-tag fusion protein containing a TEV cleavage sequence. PCR primers were designed to amplify the complete gene without predicted signal peptide coding sequences. PCR products were cloned by using the Polymerase Incomplete Primer Extension (PIPE) method (17). HK100 competent cells were transformed with the gene PCR products (I-PCR) mixed with the V-PCR of pET-TEV vector (N-term 6xHis tag). The final plasmid was transformed in BL21(DE3) strain, suitable for expression. Cultures were grown until OD<sub>600</sub> = 0.4 and induced at 25 °C for 4 h with 1 mM IPTG. Procedures for protein purification were as described above.

**Animal Model for GBS**—A maternal immunization/neonatal pup challenge model of GBS infection was used to verify the protective efficacy of SAG1333, as previously described in Maione *et al.* (3). In brief, CD-1 female mice (6–8 weeks old) received three doses (days

1, 21, and 35) of either 20  $\mu\text{g}$  antigen or PBS combined with Freund's adjuvant before breeding. Mice were bred 2–7 days after the last immunization and within 48 h of birth, pups were injected intraperitoneally with 50  $\mu\text{l}$  of GBS 515 (serotype Ia) strain corresponding to a LD<sub>50</sub>. 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 values were calculated as  $[(\% \text{ dead in control} - \% \text{ dead in vaccine})/\% \text{ dead in control}] \times 100$ . All animal experiments performed for this study were in accordance with the current Italian law and approved by the internal Animal Ethics Committee of Novartis Vaccines and Diagnostics.

**Animal Model for *S. aureus***—The protection efficacy of Protectome selected antigens have been determined *in vivo* using a peritonitis model (18) and a kidney abscess model in mice in which immunized animals were challenged on day 24 by intravenous injection of a sublethal dose of *S. aureus* ( $\sim 2$  to  $6 \times 10^7$  CFU, specific inoculum varied depending on the challenge strain). On day 28, mice were euthanized and kidneys were removed and homogenized in 1% Triton X-100 and plated on agar media in triplicate for determination of colony forming units (CFU). All animal experiments performed for this study were in accordance with the current Italian law and approved by the internal Animal Ethics Committee of Novartis Vaccines and Diagnostics.

**Cellular Impedance Measurement by xCELLigence System**—Alteration of cellular morphology by the combination of AMP and SAG1333 was tested on A459 (adenocarcinomic human alveolar basal epithelial cells) and Raw 264.7 (mouse leukemic monocyte macrophage cell) cell lines by the xCELLigence System (Roche). This system monitors cellular events in real time and measures electrical impedance across interdigitated micro-electrodes integrated on the bottom of tissue culture E-Plates. Impedance measurement with the xCELLigence system background of the E-plates was determined in 50  $\mu\text{l}$  medium and subsequently 50  $\mu\text{l}$  of cell suspension was added ( $5 \times 10^5$  cells/well for A549 cells and  $1 \times 10^6$  cells/well for Raw cells). Cells were incubated for 30 min at room temperature and E-plates were placed into the Real-Time Cell Analyzer (RTCA) station. Cells were grown for 24 h, with impedance measured every 30 min. After 1 day, cells were incubated with different concentrations of SAG1333 or the N-terminal portion of SAG1333 (10–25  $\mu\text{g}/\text{ml}$ ) in either presence or absence of 5 mM AMP. The cellular events were monitored every 15 min for 10 h and every 30 min for the following 14h. A positive control was performed with Streptolysin-O from *Streptococcus pyogenes* (SLO) using a concentration of 20  $\mu\text{g}/\text{ml}$ , while PBS was used as negative control. Impedance was represented by the cell index (CI) values  $[(Z_i - Z_0)/15[\text{Ohm}]]$ ;  $Z_0$ : background resistance,  $Z_i$ : individual time point resistance) and the normalized cell index was calculated as the cell index  $CI_{i,t}$  at a given time point divided by the cell index  $CI_{\text{nml\_time}}$  at the normalization time point (nml\_time). Data were normalized to the point corresponding to the addition of compounds.

**The *In Vitro* Enzymatic Activity of the Full Length SAG1333**—We used a commercially available Protein Tyrosine Phosphatase (PTP) Assay Kit (PTP-101, Sigma) to prove the enzymatic activity of SAG1333 to produce adenosine from AMP by releasing the phosphate. The PTP Assay Kit is based on the *in vitro* colorimetric determination of protein tyrosine phosphatase (PTP) activity and the determination of free phosphate, generated in the dephosphorylation reaction of the substrate, AMP in the present study, using Malachite Green/Ammonium Molybdate reagent. The experiments were carried out according to protocols provided by the kit. Very briefly, 10  $\mu\text{l}$  of 100 mM AMP, 30  $\mu\text{l}$  of SAG1333 (0.95  $\mu\text{g}/\text{ul}$ ), and 10  $\mu\text{l}$  phosphatase reaction buffer were added to microtiter plates. Then the reaction was terminated by the addition of 50  $\mu\text{l}$  Malachite Green/Ammonium Molybdate complex to the reaction mixture. The color was quantitated by spectrophotometry at 650 nm (using an ELISA reader) and

reflected the total amount of free, inorganic phosphate in the sample and thus the relative amount of tyrosine phosphatase activity in the sample. A standard phosphate curve was also constructed using phosphate standard solutions and by plotting the absorbance at 650 nm versus pmoles phosphate. The phosphate content of the all buffers used this experiment were also evaluated by the kit to show their phosphate free state.

## RESULTS

**Strategy to Exploit Protectome Analysis in Vaccine Discovery**—Fig. 1 schematizes the approach here proposed to exploit Protectome analysis to discover vaccine candidates. In essence, the approach consists in two main steps. First, protective antigens (PAs) so far identified in bacterial pathogens are retrieved from the literature (see [supplemental References](#)) and from existing database (19) and subjected to bioinformatics analysis in search of common protective signatures (PS) using Blast, ClustalW, Smart, and Pfam (12, 13, 14, 15). Second, PS are launched against the conserved pan-genome of the pathogen of interest with the aim at identifying proteins that share some of the signatures identified by Protectome analysis. As shown in Fig. 1, for the construction of the “Protectome Database,” data available on PAs of 38 human pathogens ([supplemental Table S1](#)) were collected. For these pathogens a total of 245 antigens have been described to be protective on the basis of the results obtained using one or more biological assays ([supplemental Table S2](#)). The meta-analysis of the “Protectome Space” revealed that known PAs can be grouped in three different categories as follows: (1) most of them (85%) show the occurrence of Pfam domains conserved among different bacterial species; (2) the remaining ones (15%) were identified by Blast analysis as taxonomically-restricted genes, being species/genus-specific proteins with either no domains, species/genus-specific domains or uncharacterized domains described as Domain of Unknown Function (DUF) in the Pfam database; and (3) a third class, overlapping with the above mentioned ones, was identified as proteins with multiple internal repeats that show either internal repeats conserved among different species, for example, the fibronectin binding repeat (PF02986), or genus/species-specific repeats, for example, those found in the Streptococcal M proteins (PF02370) or uncharacterized internal repeats, for example, the surface antigen variable number of repeats domain (PF07244) found in D15 from *H. influenzae*.

In terms of protective signatures, the analysis identified 41 major Pfam domains/clans occurring in multiple PAs and Table I reports the details of the antigens that show the same domain composition and organization together with the sequence similarity within the same subclass. These signatures clearly classified most antigens based on their well-known function or biological role, for example, toxins, iron-uptake systems, adhesins, etc. In some cases, the classification goes together with the taxonomy-associated properties, for example, flagellin domains are clearly found only in antigens from specific Gram-negative bacteria. On the other hand, the do-



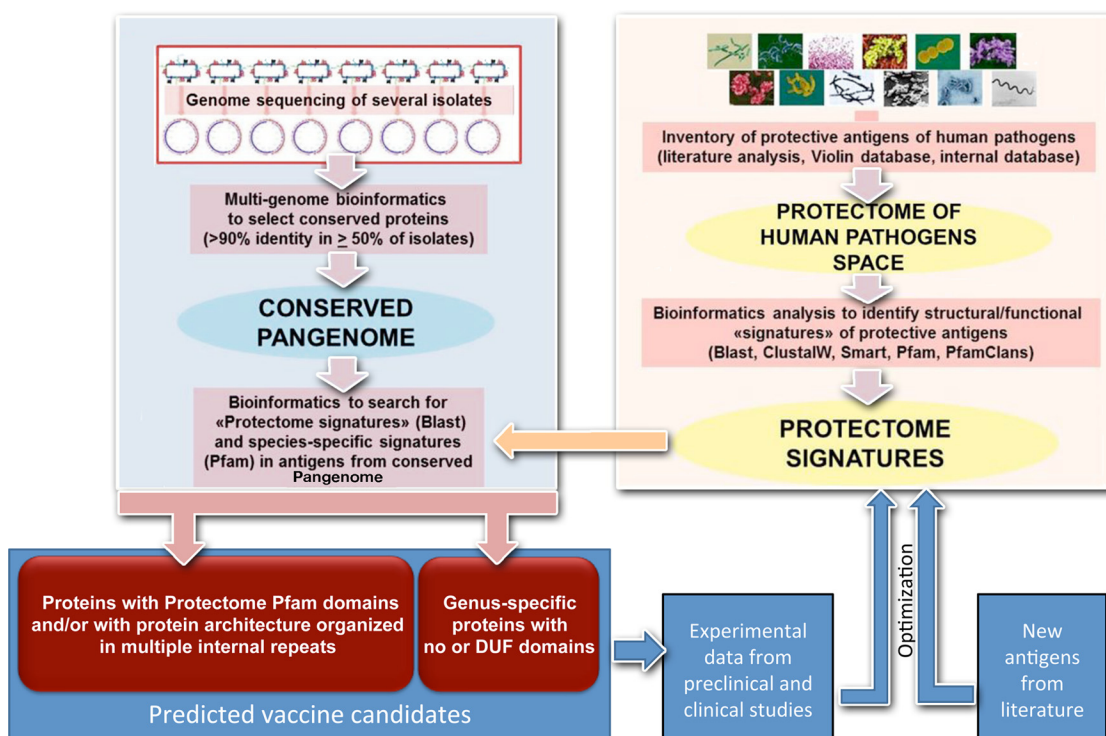


FIG. 1. **The operative steps of Protectome analysis.** The Protectome approach is based on mining the conserved pangenome a bacterial specie by using specific features. Different bioinformatics tools are used to scan for “protective signatures” i.e. (1) the identified Pfam domains, (2) a protein architecture organized in multiple internal repeats, and (3) a species- or genus-specificity of the corresponding genes with either no, species/genus-specific or DUF domains. Based on the currently defined molecular features, the Protectome approach selects less than 5% of the total predicted ORFs.

main-based classification also identified very specific, not obvious, subclasses. For instance, the Peptidase\_S8 domain groups four antigens from four different species, including both Gram-positive and Gram-negative bacteria. In these cases, the presence of antigens with this specific feature could suggest possible common mechanisms of pathogenicity for the grouped bacteria. As shown in the table, in most cases a meaningful relation between different antigens could not be appreciated at the primary sequence level, but become clear when looking at their functional/structural domains and organization. In this regard, also a relatively unspecific feature like the occurrence of multiple internal repeats appear to occur in specific classes of proteins, adhesins in particular, suggesting that this kind of protein architecture may provide advantage in relation to the adhesion process.

**Validation of Protectome Analysis for Vaccine Candidate Selection**—Having demonstrated that PAs belonging to our selected “Protectome space” share one or more PS, we next asked the question whether PS could be used to identify new PAs. However, before doing that, we decided to validate the strategy by demonstrating that it can efficiently single out known PAs included in commercially available vaccines. In particular, we took into consideration the five antigens of *Neisseria meningitidis* B vaccine (BEXSERO®, Novartis Vaccines, 4), and the three antigens of *Bordetella pertussis* vac-

cine (DAPTACEL®, Sanofi Pasteur). To this aim we: (1) defined a “sub-Protectome space” in which the PAs belonging to the two pathogens under investigation were excluded, (2) identified the PS by bioinformatics analysis of the sub-Protectome, and (3) scanned the sequences of the known PAs of the two pathogens for the presence of one or more PS. In parallel, the genome sequences of the isolates belonging to each pathogen and available in public database were compared with identified antigens carrying species-specific signatures. The results of the Protectome-based selection are reported in supplemental Table S3.

Table II summarizes the results of the validation analysis. In essence, all vaccine antigens would have been discovered in that they carry specific PS derived from sub-Protectome analysis. Moreover, when the same workflow was applied to Group A *Streptococcus* (GAS) all the three candidate antigens recently described in an experimental vaccine currently under development (9) were also identified. In particular, each vaccine contains a pathogen-specific protein, that is, the PT for DAPTACEL®, the fHbp for BEXSERO®, and the uncharacterized protein Spy0269 for GAS. The Filamentous Hemagglutinin would have been selected because of its characteristic organization in multiple internal repeats and the other antigens because of their specific functional domains already identified as PS (Table I), that is Pertactin domain is present

TABLE I

Classification of known bacterial protective antigens based on their functional and structural features. References are provided as supplemental material

Functional classes and subclasses		Antigens	%ID	Associated Pfam domains and clans
Flagellins		Flagellins from <i>P. aeruginosa</i> <sup>S1</sup> , <i>E. coli</i> <sup>S2</sup> , <i>C. jejuni</i> <sup>S3</sup> , <i>H. pylori</i> <sup>S4</sup>	9–41%	PF00669, Flagellin_N, PF00700, Flagellin_C
ompA		C3389 from <i>E. coli</i> <sup>S5</sup> , P6 from <i>H. influenzae</i> <sup>S6</sup> , Omp16 from <i>B. abortus</i> <sup>S7</sup> , OmpF from <i>P. aeruginosa</i> <sup>S8</sup>	16–32%	PF00691, OmpA
Toxins	Pore-forming toxins	Pneumolysin from <i>S. pneumoniae</i> <sup>S9</sup> , Streptolysin O from <i>S. pyogenes</i> <sup>S10</sup> , Listeriolysin O from <i>L. monocytogenes</i> <sup>S11</sup>	38–43%	PF01289, Thiol_cytolysin
		Alpha-toxin <sup>S12</sup> and Pantone-Valentine Leukocidin F <sup>S13</sup> from <i>S. aureus</i>	24%	PF07968, Leukocidin
		RTX-I and RTX-II toxins from <i>A. pleuropneumoniae</i> <sup>S14</sup> , Cya from <i>B. pertussis</i> <sup>S15</sup> , Hemolysin from <i>E. coli</i> <sup>S16</sup>	23–55%	PF00353, HemolysinCabind PF02382, RTX.
	Ribosyl transferases	Diphtheriae toxin from <i>C. diphtheriae</i> <sup>S17</sup> , Pertussis toxin from <i>B. pertussis</i> <sup>S17</sup> , Cholera toxin from <i>V. cholerae</i> <sup>S18</sup>	7–11%	CL0084, ADP-ribosyl
	Neurotoxins	Tetanus toxin from <i>C. tetani</i> <sup>S19</sup> , Botulinum toxin from <i>C. botulinum</i> <sup>S20</sup>	33%	PF01742, Peptidase_M27 PF07951, Toxin_R_bind_C, PF07953, Toxin_R_bind_N.
	Glycosyltransferases	Toxin A and Toxin B from <i>C. difficile</i> <sup>S21</sup>	47%	PF11713, Peptidase_C80, PF12919, TcdA_TcdB, PF12920, TcdATcdB_pore
	Binary toxins	Protective antigen from <i>B. anthracis</i> <sup>S22</sup> , C2 toxin from <i>C. botulinum</i> <sup>S20</sup> , Iota toxin component Ib from <i>C. perfringens</i> <sup>S23</sup>	26–37%	PF03495, Binary_toxB
Extracellular enzymes	proteases	C5a peptidases from <i>S. agalactiae</i> and <i>S. pyogenes</i> <sup>S24, S24a</sup> , SpyCEP from <i>S. pyogenes</i> <sup>S25</sup> , PrtA from <i>S. pneumoniae</i> <sup>S26</sup> , NMB1969 from <i>N. meningitidis</i> <sup>S27</sup>	9–29%	PF00082, Peptidase_S8
	Hydrolases with papain-like fold	P60 from <i>L. monocytogenes</i> <sup>S28</sup> , NMB2001 from <i>N. meningitidis</i> <sup>S29</sup> , PcsB from <i>S. pneumoniae</i> <sup>S30</sup> , SpeB from <i>S. pyogenes</i> <sup>S31</sup>	6–21%	CL0125, Peptidase_CA
	Hydrolases with trypsin-like fold	HtrA from <i>H. influenzae</i> <sup>S32</sup> , CT823 from <i>C. trachomatis</i> <sup>S33</sup>	x%	PF00089, Trypsin
	5' nucleotidases	NucA from <i>H. influenzae</i> <sup>S34</sup> , Spy0872 from <i>S. pyogenes</i> <sup>S35</sup>	18%	PF02872, 5_nucleotid_C PF00149, Metallophos.
Iron uptake and other transport systems	heme binding proteins	IsdA and IsdB from <i>S. aureus</i> <sup>S36</sup>	10–32%	PF05031, NEAT
	TonB dependent receptors	IroN, FyuA, IreA and IutA <sup>S37</sup> from <i>E. coli</i> , TbpA from <i>P. haemolytica</i> <sup>S38</sup>	5–11%	PF07715, Plug PF00593, TonB_dep_Rec
	Other transport proteins	OmpP1 from <i>H. influenzae</i> <sup>S39</sup> , OmpP1 from <i>N. meningitidis</i> <sup>S40</sup>	23%	PF03349, Toluene_X
Solute binding	Transferrin binding protein-like proteins	Heparin binding protein GNA2132 from <i>N. meningitidis</i> <sup>S41</sup> , TbpB from <i>P. haemolytica</i> <sup>S38</sup>	8–23%	PF01298, Lipoprotein_5
Proteins	Other solute binding proteins	P39 from <i>B. melitensis</i> <sup>S42</sup> , CjaA from <i>C. jejuni</i> <sup>S43</sup> , GNA1946 from <i>N. meningitidis</i> <sup>S29</sup> , PotD from <i>S. pneumoniae</i> <sup>S44</sup> , ArtJ from <i>C. trachomatis</i> <sup>S45</sup>	6–58%	CL0177, PBP
Secretion and export systems	secretion system effector c-like proteins	SseC from <i>Y. pseudotuberculosis</i> <sup>S46</sup> , and IpaB from <i>Shigella</i> <sup>S47</sup>	19%	PF04888, SseC
	LcrV-like proteins	LcrV from <i>Y. enterocolitica</i> <sup>S48</sup> , LcrV from <i>Y. pestis</i> <sup>S49</sup> , and PcrV from <i>P. aeruginosa</i> <sup>S50</sup>	39–95%	PF04792, LcrV
	Type iii secretory chaperones	YerA from <i>Y. pestis</i> <sup>S51</sup> , CT043 from <i>C. trachomatis</i> <sup>S52</sup>	24%	PF05932, CesT
Putative adhesins and invasins	fibronectin-binding proteins	FnbA from <i>S. aureus</i> <sup>S53</sup> , Fnz from <i>S. equi</i> <sup>S54</sup> , and Sfb from <i>S. pyogenes</i> <sup>S55</sup>	10–25%	PF02986, Fn_bind.
	Proteins with cnab domain	Pili subunits from <i>S. pyogenes</i> <sup>S56</sup> , <i>S. agalactiae</i> <sup>S57</sup> , and <i>S. pneumoniae</i> <sup>S58</sup> , Cna <sup>S59</sup> , SdrD and SdrE <sup>S60</sup> from <i>S. aureus</i>	10–58%	PF05738, Cna_B
	Proteins with uncharacterized multiple internal repeats	SdrD and SdrE <sup>S60</sup> from <i>S. aureus</i> , SAN_1485 from <i>S. agalactiae</i> <sup>S61</sup>	17–55%	PF10425, SdrG_C_C
		Rib <sup>S62</sup> and C protein alpha antigen <sup>S63</sup> from <i>S. agalactiae</i>	48%	PF08829, AlphaC_N PF08428, Rib
	Others	NadA <sup>S41</sup> and NhhA <sup>S64</sup> from <i>N. meningitidis</i> , UspA1 and UspA2 from <i>M. catarrhalis</i> <sup>S65</sup> , YadA and YadC from <i>Y. pestis</i> <sup>S66</sup> , C4424 from <i>E. coli</i> <sup>S67</sup>	8–42%	PF03895, YadA
Proteins with other binding domains		BrkA <sup>S68</sup> and Pertactin <sup>S69</sup> from <i>B. pertussis</i> , Hap from <i>H. influenzae</i> <sup>S70</sup> , NMB1998 from <i>N. meningitidis</i> <sup>S71</sup>	11–31%	PF03212, Pertactin
		PspC from <i>S. pneumoniae</i> <sup>S72</sup> , IgA-binding beta antigen from <i>S. agalactiae</i> <sup>S73</sup>	27%	PF05062, RICH
	lysm domain containing proteins	P60 from <i>L. monocytogenes</i> <sup>S28</sup> , Sip from <i>S. agalactiae</i> <sup>S74</sup> , Intimin from <i>E. coli</i> <sup>S75</sup>	7–11%	PF01476, LysM
	Streptococcal histidine triad proteins	PhtA, PhtB and PhtD <sup>S76</sup> from <i>S. pneumoniae</i>	32–87%	PF04270, Strep_his_triad

TABLE II  
Protectome features identified in antigen components from licensed vaccines

Recombinant vaccine	Pathogen	Antigens	Protectome features
Daptacel®, Sanofi Pasteur (Licenced)	<i>B. pertussis</i>	Pertussis Toxin (PT) Filamentous Haemagglutinin Pertactin	<i>Bordetella</i> -specific Multiple internal repeats PF03212. Pertactin
Bexsero®, Novartis Vaccines (Licensed)	<i>N. meningitidis</i> Serogroup B	Factor H binding protein (fHbp) Neisserial heparin binding protein (NHBA) Neisserial adhesin A (NadA)	<i>Neisseria</i> -specific PF01298. Lipoprotein_5 PF03895. YadA

also in Hap from *H. influenza* and NMB1998 from *N. meningitidis*; the Lipoprotein\_5 is also found in TbpB from *H. influenzae*; the YadA domain is present in multiple antigens including NhhA from *N. meningitidis*, UspA1 and UspA2 from *M. catarrhalis*, YadA and YadC from *Y. pestis*, and C4424 from *E. coli*; the Thiol\_cytolysin domain in Pneumolysin from *S. pneumoniae* and Listeriolysin O from *L. monocytogenes*; the Peptidase\_S8 domain as already mentioned is shared by C5a peptidases from *S. agalactiae*, PrtA from *S. pneumoniae* and NMB1969 from *N. meningitidis*. When looking to single vaccine antigens, we also looked at how many additional proteins encoded in the genomes of the bacteria contain the same PS. We performed this in particular for those antigens that have specific Pfam functional domains (supplemental Table S4 reports the information obtained). As expected, the PS identified class of proteins where other members were proved to be protective, supporting our idea that the biological role/function of a protein can determine its protective properties. As an example, the PS “PF03212 Pertactin” identifies a class of four different proteins in *B. pertussis* genome, two of which are known protective antigens. Also in the case of GAS, the PS “Peptidase\_S8” identified two different proteins in the GAS genome, both protective. In MenB, two YadA-containing proteins were identified, again both protective.

A further validation step would be to establish if proteins with no PS are truly nonprotective. However, addressing this question is extremely complicated. In fact, it would imply testing more than 95% of the total ORFs *in vivo*. To at least partially test the selectivity of our approach we have analyzed the presence of our PS in an independent dataset of nonprotective antigens (10). Interestingly, only two proteins out of 100 in the dataset of known nonprotective antigens contain PS. It is worth noting that proteins with these domains, that is, SBP\_bac\_1 and Bmp, have been found to be protective in other species (see supplemental Table S2). This latter result suggests that these domains can be representative of families that include both protective and nonprotective members and that a further refinement of these features is needed to be able to identify only the protective ones.

**Prediction of PAs for *S. aureus* and GBS by Using Protectome Analysis**—Having validated the proposed approach by “re-discovering” well known PAs we next applied our strategy to two different pathogens, GBS and *S. aureus*, which have

already been the objective of extensive studies aimed at the identification of vaccine candidates (3, 15, 20, 21). Scope of our study was to investigate whether the approach could select the known PAs and could provide indications on novel protective protein candidates. Table III reports the list of the conserved vaccine candidates identified by scanning the available genomes of both pathogens for the presence of protective and species-specific signatures. As shown in the supplemental Table S4 and S5, 59 and 89 proteins were identified for GBS and *S. aureus*, respectively. The list includes all known PAs previously described for both pathogens (Table I and supplemental Table S2), further strengthening the validation analysis carried out on *N. meningitidis*, *B. pertussis*, and GAS. When the candidates were classified according to their cellular localization, it was found that they largely belong to the external compartments (secreted or surface-exposed proteins), in line with the notion that antibody-mediated immunity targeting extracellular proteins constitutes the most represented mechanism of protection. However, not all the predicted surface-exposed proteins are included in the selection list (Fig. 2), indicating that the presence of extracellular localization signals is not a prerequisite. In fact, PA may also be T-cell antigens that do not necessarily required to be surface-exposed or secreted. In addition, PA could carry nonclassical export signals and therefore not predicted by current *in silico* tools. This can be also deduced by analyzing the “surfomes” of GBS and *S. aureus*. Both pathogens have been the object of surface proteome characterization using protease cell surface shaving coupled to mass spectrometry analysis of proteolytic peptides, (22). The study applied to GBS COH1 identified 43 protease-accessible proteins (8). Three independent studies applied to *S. aureus* respectively identified 95, 113, and 96 surface-associated proteins in several strains (23, 24, 25). Also in this case, the Protectome selection does not simply resemble the bacterial surface proteome.

**Protectome Analysis Identified a New Protective Antigen for GBS: The 5' Nucleotidase SAG1333**—We next investigated whether among the list of GBS and *S. aureus* vaccine candidates carrying PS not-yet-discovered PAs exist. To this aim, we first inspected the list of GBS candidates reported in Table III, asking the question which of them had been included in the 589 proteins selected in our previous RV study on the basis of

TABLE III

List of Protectome candidates for *S. aureus* and GBS. Proteins are grouped according to the feature used to select them. \*indicates protective antigens identified in previous studies

<b>CHAP</b>	<b>S. aureus:</b> SAOUHSC_02571, SAOUHSC_02883, SAOUHSC_02979, SAOUHSC_02019, SAOUHSC_00256, SAOUHSC_02576, SAOUHSC_02023
<b>LysM</b>	<b>GBS:</b> SAG1998, SAG1762, SAG1683, SAG1286, SAG0598, SAG0017 <b>S. aureus:</b> SAOUHSC_01501, SAOUHSC_00069*
<b>CHAP, LysM</b>	<b>GBS:</b> SAG0032*, SAG1386, SAG2148
<b>Peptidase_S8</b>	<b>S. aureus:</b> SAOUHSC_00773, SAOUHSC_00671, SAOUHSC_01219, SAOUHSC_00427 <b>S. aureus:</b> SAOUHSC_01949*
<b>5_nucleotid_C</b>	<b>GBS:</b> SAG2053, SAG0676, SAG0416, SAG_1320* <b>S. aureus:</b> SAOUHSC_00025*, SAOUHSC_00107, SAOUHSC_00860
<b>Peripla_BP_2</b>	<b>GBS:</b> SAG1333, SAG1941 <b>S. aureus:</b> SAOUHSC_00074, SAOUHSC_00613, SAOUHSC_00749, SAOUHSC_00976, SAOUHSC_01085, SAOUHSC_02246, SAOUHSC_02430, SAOUHSC_02554*
<b>Lipoprotein_9</b>	<b>GBS:</b> SAG1393, SAG1007 <b>S. aureus:</b> SAOUHSC_00426, SAOUHSC_00844
<b>FbpA</b>	<b>GBS:</b> SAG0776, SAG0971, SAG1641 <b>S. aureus:</b> SAOUHSC_01175
<b>WXG100</b>	<b>GBS:</b> SAG1190 <b>S. aureus:</b> SAOUHSC_00257*, SAOUHSC_00265*
<b>Cna_B</b>	<b>GBS:</b> SAG1039, SAG0230
<b>SdrG_C_C</b>	<b>GBS:</b> SAG0651, SAG1408*, SAG1407*, SAG1404*, SAG0649*, SAG0646*, SAG0645* <b>S. aureus:</b> SAOUHSC_00812*, SAOUHSC_02963*, SAOUHSC_02802*, SAOUHSC_02803*
<b>Cna_B</b>	<b>GBS:</b> SAG1462, SAN_1485*
<b>SdrG_C_C</b>	<b>S. aureus:</b> SAOUHSC_00544*, SAOUHSC_00545*
<b>SBP_bac_3</b>	<b>S. aureus:</b> SAOUHSC_01991, SAOUHSC_02699
<b>SBP_bac_9</b>	<b>GBS:</b> SAG0136, SAG0290, SAG0717, SAG0949, SAG1431, SAG1610, SAG1642, SAG1466 <b>S. aureus:</b> SAOUHSC_00634*, SAOUHSC_02690
<b>Leukocidins</b>	<b>GBS:</b> SAG1533, SAG0535, SAG1234, SAG1938 <b>S. aureus:</b> SAOUHSC_01954, SAOUHSC_01955, SAOUHSC_02241, SAOUHSC_02243, SAOUHSC_02708, SAOUHSC_02709, SAOUHSC_02710, SAOUHSC_01121*
<b>NEAT</b>	<b>S. aureus:</b> SAOUHSC_01081*, SAOUHSC_01082, SAOUHSC_01079*, SAOUHSC_01843
<b>Trypsin</b>	<b>S. aureus:</b> SAOUHSC_01935, SAOUHSC_01936, SAOUHSC_01939, SAOUHSC_01941, SAOUHSC_01942
<b>Ferritin</b>	<b>S. aureus:</b> SAOUHSC_02108
<b>band_7</b>	<b>GBS:</b> SAG0132
<b>Clp Protease</b>	<b>GBS:</b> SAG1585
<b>Strep_his triad</b>	<b>GBS:</b> SAG1233, SAG0907
<b>Genus-specific</b>	<b>S. aureus:</b> SAOUHSC_00030, SAOUHSC_00052*, SAOUHSC_00053*, SAOUHSC_00054*, SAOUHSC_00055*, SAOUHSC_00172, SAOUHSC_00250, SAOUHSC_00251, SAOUHSC_00369, SAOUHSC_00400, SAOUHSC_00402, SAOUHSC_00404, SAOUHSC_00405, SAOUHSC_01025, SAOUHSC_01084, SAOUHSC_01180, SAOUHSC_01317, SAOUHSC_01508, SAOUHSC_01511, SAOUHSC_01512, SAOUHSC_01584, SAOUHSC_01854, SAOUHSC_02008, SAOUHSC_02257, SAOUHSC_02783, SAOUHSC_02788, SAOUHSC_02789, SAOUHSC_02813 <b>GBS:</b> SAG0677, SAG1552, SAG0304, SAG1807, SAG1822, SAG1730, SAG0204, SAG0771, SAG1419, SAG0392, SAG2063

their being secreted or membrane/cell wall-associated proteins (3). Not surprisingly, most of them (50 out of 59) were part of the original selection; the only proteins excluded being nine proteins classified as cytoplasmic and/or unknown. Of the 50 selected proteins six failed to be expressed or were expressed as inclusion bodies and therefore not tested in protection. Among the remaining 44 proteins, five gave low/variable level of protection in the active maternal immunization model and five proteins were highly protective. The latter included the pilus components and SAG0032 (3). When we inspected the six proteins not tested because of failure of expression, one protein, SAG1333, attracted our attention. The protein is annotated as a 5' nucleotidase (5'NT), a class of hydrolytic enzymes that catalyze the hydrolysis of nucleotide into nucleoside and phosphate by cleaving off the phosphate from the 5' end of the sugar moiety. 5'NT are grouped in

different classes on the basis of their substrate specificity, and 5' NTs acting on AMP are of particular interest in bacterial infections, because adenosine is a potent immuno-suppressor exerting its activity by binding to receptors expressed on various innate immune cells, including macrophages. Therefore, we decided to spend additional efforts in the expression of SAG1333 in *E. coli* with the aim of obtaining sufficient material to be tested in the animal model. *sag1333* gene was cloned in a different configuration with respect to what was previously published. In particular, we performed a domain composition analysis using SMART and we identified the regions corresponding to the two predicted functional domains, that is, the Metallophos and the 5'NT domain. Then, we performed a secondary structure prediction and based on that we designed a construct SAG1333<sub>28–663</sub>, lacking the signal peptide and the C-terminal LPXTG motif (Fig. 3A).



FIG. 2. **Signal peptide prediction.** Predicted localization of Protectome antigens compared with the total number of proteins in each sub-cellular fraction in GBS and *S. aureus*, respectively.

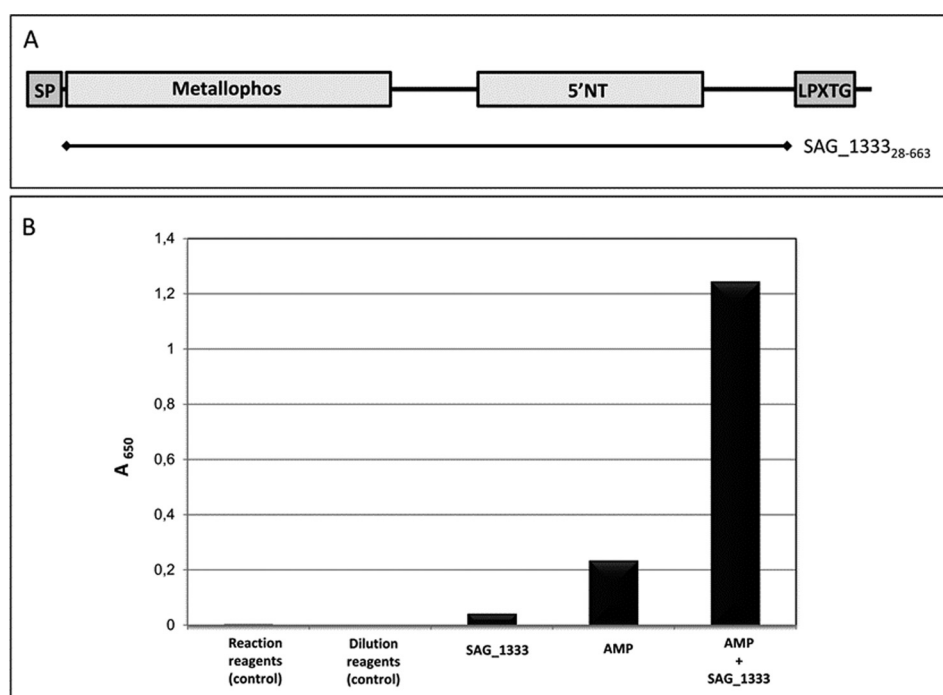
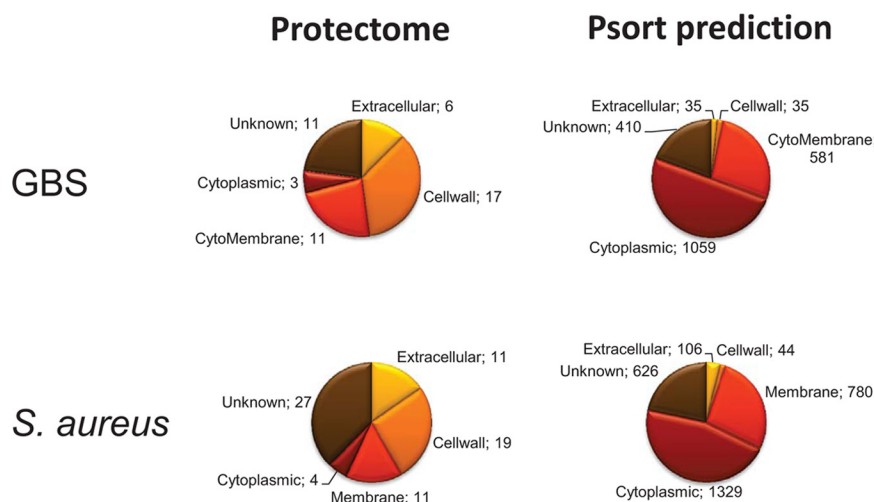


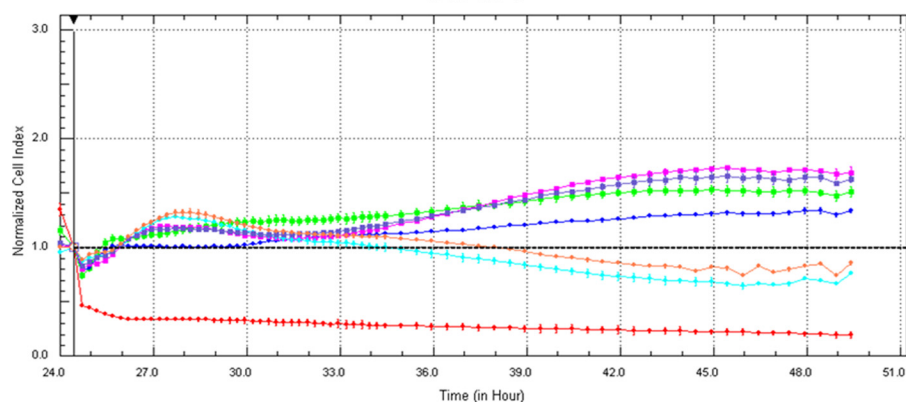
FIG. 3. **In vitro enzymatic activity of SAG1333.** A commercial kit was used to prove the enzymatic activity of SAG1333 to produce adenosine from AMP by releasing the phosphate. The figure shows the measurement of free phosphate (pmol) determined in different compounds. Kit reagents and the enzyme alone were phosphate-free, AMP gave negligible amounts of phosphate. On the other hand, recombinant SAG1333 releases 2354 pmol of free phosphate by releasing the phosphate generated in the dephosphorylation reaction of the AMP.

Using the new cloning strategy the 5'NT was successfully purified from the soluble fraction at a purity estimated to be higher than 85%. Next, we tested its phosphatase activity by following its capacity to release phosphate in the presence of AMP. As shown in Fig. 3B, the addition of AMP to the enzyme resulted in a rapid accumulation of free phosphate. Furthermore, to have indication that the protein can exert an antiphagocytic effect, we analyzed the ability of SAG1333<sub>28-663</sub> to perturb the integrity of macrophages. As shown in Fig. 4, in the presence but not in the absence of AMP, it induced a dramatic decrease of normalized cellular electric resistance.

Interestingly, a similar effect was not observed on A549 lung epithelial cells. Having demonstrated that SAG1333<sub>28-663</sub> has indeed a 5'NT activity on AMP and that the 5'NT-mediated adenosine production can affect the activity of macrophages, we finally asked the question whether 5'NT immunization could elicit protective responses in mice. To this aim, adult female mice were immunized three times at two-week intervals with 20 µg of recombinant 5'NT formulated in aluminum hydroxide and subsequently mated. The 24–48 h old offspring was finally challenged with a lethal dose of GBS 515 serotype Ia strain. As shown in Fig. 5A



## RAW



## A549

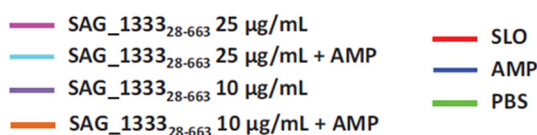
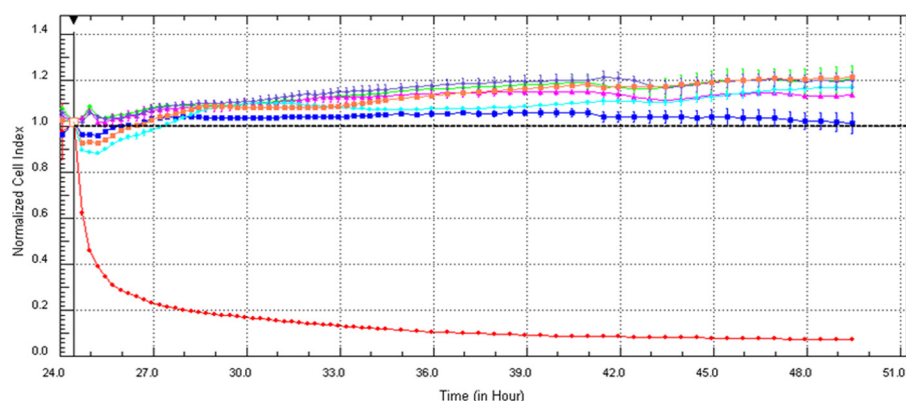


FIG. 4. Real-time analysis of the effect of recombinant SAG1333 on macrophage integrity by xCELLigence system. Graphs report the measurement of cellular Trans-Electric Resistance by xCELLigence. SAG1333 was added at a concentration of 25 and 10 µg/ml; SLO at 20 µg/ml. The following samples were tested: SAG1333 was tested alone or plus 5 mM AMP on Raw cells (high panel) and A549 cells (low panel); Data represent the mean ± S.D. of three independent wells.

most of the pups (67%) from immunized mothers survived the GBS challenge, indicating that functional antibodies elicited by immunization had been transferred from the mothers to the pups at a level sufficiently high to neutralize the infection.

*Protectome Analysis Identified a New Protective Antigen for S. aureus: The LysM Containing Protein SAOUHSC\_00427*—As far as the list of 89 *S. aureus* candidate is concerned, it included all protective antigens reported so far, further validating the selective power of the Protectome approach here proposed. Among the remaining proteins, we focused our attention on SAOUHSC\_00427. This protein appears particularly interesting in that it carries a LysM domain that is shared by a number of virulence factors from different pathogens, including the P60 protein from *Listeria monocytogenes*, the *Escherichia coli* Intimin proteins, and Sip (SAG0032) from GBS. In particular, SAG0032 is a highly protective and conserved antigen which has been the object of intense studies for vaccine development (3). Recombinant SAOUHSC\_00427 was

purified from *E. coli* as His-tag fusion protein and the purified protein was tested for protection using two different mouse models. According to the first model, mice received two intraperitoneal doses of alum-formulated SAOUHSC\_00427 (20 µg/dose) at 2-week intervals. Ten days after immunization, mice were challenged by intraperitoneal injection of a lethal dose ( $2 \times 10^8$  CFUs) of *S. aureus* Newman and animal survival was followed over a period of 15 days. As shown in Fig. 5C, 65% of animals survived the challenge as opposed to the 20% survival of mock immunized mice. In the second model, after immunization, which followed the same protocol as described before, mice were challenged by intravenous injection of a sub-lethal dose of *S. aureus* Newman strain ( $2 \times 10^7$  bacteria) and abscess formation in the kidneys was monitored after 5 days from challenge. As shown in Fig. 5B, immunization resulted in a remarkable reduction (2.5 logs) in the number of CFUs recovered from kidneys. Taken together, the animal studies indicate that SAOUHSC\_00427 is a very promising vaccine candidate that is worth future attention.

### A) Maternal-neonatal model for GBS

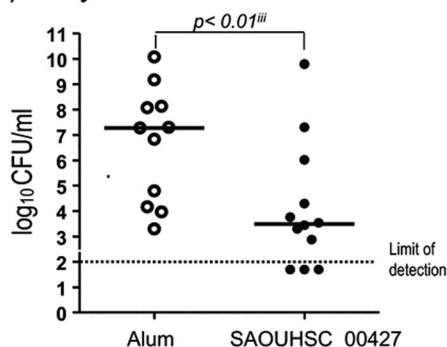
Protein name	No. of mice Alive/Treated (protein)	No. of mice Alive/Treated (alum)	Protection efficacy (%) (i)	P value (ii)
SAG1333	63/89	4/40	67%	<0.001

### B) Peritonitis model for *S.aureus*

Protein name	No. of mice Alive/Treated (protein)	No. of mice Alive/Treated (alum)	Protection efficacy (%) (i)	P value (ii)
SAOUHSC_00427	31/48	12/48	53%	<0.001

- i. Protection efficacy was calculated as  $[(\% \text{ dead}_{\text{control}} - \% \text{ dead}_{\text{vaccine}}) / \% \text{ dead}_{\text{control}}] \times 100$ .  
ii. Significance vs. the control group was evaluated by one-tailed Mann-Whitney U test and Log-rank (Mantel-Cox) test

### C) Kidney abscess model for *S.aureus*.



- iii. Significance vs. the Alum group was evaluated by Mann-Whitney U test

FIG. 5. **Protection studies in relevant *in vivo* models.** Protection conferred by the two antigens were assessed by active maternal immunization/neonatal pup challenge model for GBS (panel A), and the peritonitis (panel B) and kidney abscess model (panel C) for *S. aureus* antigen.

## DISCUSSION

The “Holy Grail” of Vaccinology is the knowledge to select PAs from genome sequence. With this power at their disposal, vaccinologists would dramatically shorten the time to vaccine development, would restrict animal use to toxicology, and would test vaccine efficacy directly in humans.

The Holy Grail is not a reality yet because we are still not capable of recognizing a protective antigen from its primary and/or tertiary structure. At present, the best we can achieve is to define criteria that make a protective antigen protective and then to use high throughput technologies to identify from all pathogen components those fulfilling such criteria. Two different criteria have been extensively applied. The first, the “immunogenicity criterion,” braces the idea that PAs must be immunogenic during natural infection. With this assumption, a number of experimental strategies have been developed to identify pathogen-associated proteins that induce antibody and cell-mediated responses (26, 27, 28). The second, the “compartmentalization criterion,” sustains that if antibodies are mediating the protective response, PAs are either secreted or surface-associated. Therefore, proteomics analyses have been used to experimentally identify this category of antigens (7, 8, 23, 24, 25, 29). More recently, a third criterion has been proposed that essentially combines the immunogenicity and compartmentalization hypotheses: to be PAs they

must be: (1) conserved, (2) well expressed, (3) immunogenic during natural infection, and (4) secreted and/or surface associated. By this way, the number of vaccine candidates to be tested in the animal models of protection, the most demanding and critical part of the whole vaccine discovery process, is remarkably reduced: from a few thousand (the average number of bacterial proteins each of which potentially being a protective antigen) to a few tens and, in the case of the “immunogenic/compartmentalization combined criterion,” to less than ten (9). However, regardless the criterion applied, the identification of PAs still requires a substantial amount of experimental work and it is biased toward the selection of antigens providing antibody-mediated protection. In the present work, we have moved one step forward toward the from-genome-to-vaccine goal. The Protectome approach reaches a substantial filtering of the bacterial proteome without the need of any experimental work and performs an unbiased selection of new vaccine candidates, because all known bacterial PAs were initially included in the Protectome space regardless of their mechanism of protection.

Starting from the assumption that bacterial PAs must have specific structural/functional signatures that make them protective, we created the “Protectome” database including all PAs described so far and we have used different tools to identify common motifs within the Protectome space. As pre-

dicted, the bioinformatics analysis has revealed that PAs can be grouped in families that share “protective signatures” that classify antigens based on their function/biological role (toxins, iron-uptake systems, adhesins, etc.) and/or based on their structural organization (for instance, multiple internal structural motifs of bacterial adhesins). Finally, we have scanned the genomes of different pathogens in search of conserved proteins carrying protective signatures. When applied to different pathogens the approach not only allowed to rediscover already known PAs but also to identify new vaccine candidates that deserve future attention.

Bioinformatics has been the first filtering strategy applied in genome-based vaccine discovery projects to reduce the number of antigens to be tested in animal models. Starting from the assumption that antigens that induce bactericidal antibodies must be surface-exposed, in their pivotal work Pizza and co-workers used PSORT to identify meningococcal genes encoding proteins carrying leader sequence for secretion. In this way ~70% of the genome was excluded from subsequent high throughput analysis but still 600 proteins had to be tested for bactericidal activity to ultimately identify the five PAs currently included in the commercialized vaccine (2, 4). An even higher number of proteins were selected for subsequent testing in the cumbersome active maternal immunization mouse model to select GBS PAs (3). With the bioinformatics approach described here the pool of protective candidates is further reduced down 50–70, thus less than 5% of the total predicted ORFs. Considering that retrospective analyses on a number different pathogens, including *B. pertussis*, *H. pylori*, MenB, GBS, GAS, and *S. aureus*, have shown that the approach would not have missed a single antigen, we believe that the result is remarkable. Furthermore, as we demonstrated for GBS and *S. aureus*, new promising PAs that had not been selected using previous genomic approaches have been identified.

The Protectome method described here has been generated including all bacterial PAs regardless their mechanism of protection (antibody mediated or cell mediated), their compartmentalization (secreted, surface-associated, and cytoplasmic), the type of biological assay used for establishing protective activity (*in vitro* versus *in vivo*), level of protection, formulation (adjuvants) used for inducing protective immune responses, etc. We believe that this unbiased approach is particularly useful when little is known about the type of immune response needed to protect the pathogen of interest. However, in a number of cases, the type of immune response the vaccine should elicit is known. Typical examples are MenB and GBS. In the case of MenB, to be effective vaccination has to induce high bactericidal antibody titers in infants and adolescents, the main vaccine target population. As far as GBS is concerned, the vaccine should protect newborns within their first 90 days of life from delivery. Protection is exclusively antibody-mediated and the vaccine should be administered to women to allow passive transfer of opsonophagocytic antibodies to the fetus. Therefore, for both

pathogens, the only antigens that can induce protective antibodies are those that are surface-associated; secreted toxins, cytoplasmic proteins, and proteins inducing T cell responses cannot elicit the proper immune response. Indeed, the newly identified antigen 5' NT is a surface-associated protein. Although we have not done this analysis, we expect that by using “subprotectomes” tailored on the basis of the immune responses needed for protection, for example, T-cell versus antibody-mediated, the number of selected vaccine candidates would be further reduced.

The animal models used to test protection of the vaccine candidates deserve a comment. It is still not clear how many PAs exist for each pathogen. In the case of pathogens whose pathogenicity is mediated by secreted toxins, single inactivated toxins have been shown to be sufficient to prevent disease. Likewise, one antigen abundantly expressed on the bacterial surface can be sufficient to induce excellent bactericidal/opsonophagocytic antibodies; typical examples are polysaccharides constituting glycoconjugate vaccines, and fHbp of MenB (4, 30, 31). For other pathogens with a much more complex mechanism of pathogenesis, such as GAS and *S. aureus*, effective vaccines are expected to require cocktails of several antigens in order to neutralize different bacterial virulence factors. However, in many cases the animal models used to screen for PAs do not mimic human infection. These models are biased by the fact that animals are infected with large quantities of bacteria and the only antigens that result protective are those that block the bacteremia induced by the challenge. Other antigens that might play important role in inducing protective responses in humans are completely lost because of inadequacy of the models. This aspect has to be constantly kept in mind before excluding or including antigens identified by genomic approaches.

Our bioinformatics approach described here and based on the identification of antigens carrying protective signatures, appears to be the most selective *in silico* strategy for vaccine candidate discovery reported so far. The optimization of the “Protectome space” to be used for protective signature identification and an optimization of the tools used for protective signature selection are expected to bring the candidates down to a number that allow their direct testing in the human model. For instance we have observed that some PS, that is, those associated to specific Pfam domains such as Peptidase\_S8 and Pertactin, are highly selective. On the contrary, PS exist that identify large families of proteins that include both protective and nonprotective antigens. In some cases, these discrepancies can be caused by technical artifacts or use of nonappropriate animal models. However, in other cases, it could be a consequence of the lack of a complete knowledge about protein function and of its associated structural/functional features. A significant step forward would be achieved when a detailed biological characterization of these protective antigens is carried out, leading to a further refinement of PS that would greatly improve the specificity of the Protectome prediction.

[S] This article contains supplemental Tables S1 to S5 and References.

§ To whom correspondence should be addressed: CIBIO, Center of Integrative Biology, University of Trento, Via delle Regole 101, Mat-tarello, Trento, Italy. E-mail: guido.grandi@unitn.it.

## REFERENCES

- Fleischmann, R. D., Adams, M. D., White O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J.-F., Dougherty, B. A., Merrick, J. M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne, J. D., Scott, J., Shirley, R., Liu, L.-I., Glodek, A., Kelley, J. M., Weidman, J. F., Phillips, C. A., Spriggs, T., Hedblom, E., Cotton, M. D., Utterback, T. R., Hanna, M. C., Nguyen, D. T., Saudek, D. M., Brandon, R. C., Fine, L. D., Fritchman, J. L., Fuhrmann, J. L., Geoghagen, N. S. M., Gnehm, C. L., McDonald, L. A., Small, K. V., Fraser, C. M., Smith, H. O., and Venter, J. C. *et al.* (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512
- Pizza, M., Scarlato, V., Masignani, V., Giuliani, M. M., Aricò, B., Coman-ducci, M., Jennings, G. T., Baldi, L., Bartolini, E., Capecci, B., Galeotti, C. L., Luzzi, E., Manetti, R., Marchetti, E., Mora, M., Nuti, S., Ratti, G., Santini, L., Savino, S., Scarselli, M., Storni, E., Zuo, P., Broeker, M., Hundt, E., Knapp, B., Blair, E., Mason, T., Tettelin, H., Hood, D. W., Jeffries, A. C., Saunders, N. J., Granoff, D. M., Venter, J. C., Moxon, E. R., Grandi, G., and Rappuoli, R. *et al.* (2000) Identification of vaccine candidates against serogroup B meningococcus by whole-genome se-quencing. *Science* **287**, 1816–1820
- Maione, D., Margarit, I., Rinaudo, C. D., Masignani, V., Mora, 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., Aricò, B., Savino, S., Santini, L., Brunelli, B., Bambini, S., Biolchi, A., Capecci, B., Cartocci, E., Ciocchi, 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, J. A., Granoff, D., Rappuoli, R., and Pizza, M. (2006) A universal vaccine for serogroup B meningo-coccus. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 10834–10839
- Serruto, D., Bottomley, M. J., Ram, S., Giuliani, M. M., and Rappuoli, R. (2012) The new multicomponent vaccine against meningococcal sero-group B, 4CMenB: immunological, functional, and structural character-ization of the antigens. *Vaccine* **30**, B87–B97
- Grandi, G. (2010) Bacterial vaccine discovery: from “brute force” to high selectivity. *Hum. Vaccin.* **6**, 872–875
- Rodríguez-Ortega, M. J., 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, J. L., and Grandi, G. (2006) Characterization and iden-tification of vaccine candidate proteins through analysis of the group A Streptococcus surface proteome. *Nat. Biotechnol.* **24**, 191–197
- Doro, F., Liberatori, S., Rodríguez-Ortega, M. J., Rinaudo, C. D., Rosini, R., *et al.* (2009) 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* **8**, 1728–1737
- Bensi, G., Mora, M., Tuscano, G., Biagini, M., Chiarot, E., Bombaci, M., Capo, S., Falugi, F., Manetti, A. G., Donato, P., Swennen, E., Gallotta, M., Garibaldi, M., Pinto, V., Chiappini, N., Musser, J. M., Janulczyk, R., Mariani, M., Scarselli, M., Telford, J. L., Grifantini, R., Norais, N., Margari-t, I., and Grandi, G. (2012) Multi high-throughput approach for highly selective identification of vaccine candidates: the Group A Streptococ-cus case. *Mol. Cell. Proteomics* **11**, M111.015693
- Doytchinova, I. A., and Flower, D. R. (2007) VaxiJen: a server for prediction of protective antigens, tumor antigens, and subunit vaccines. *BMC Bioinformatics* **8**, 4
- UniProt Consortium. Activities at the Universal Protein Resource (UniProt) (2014) *Nucleic Acids Res.* **42**, D191–D198
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402
- Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L., Tate, J., and Punta, M. (2014) Pfam: the protein families database. *Nucleic Acids Res.* **42**, D222–D230
- Letunic, I., Doerks, T., and Bork, P. (2012) SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res.* **40**, D302–D305
- Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties, and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680
- Mott, R. (2000) Accurate formula for P-values of gapped local sequence and profile alignments. *J. Mol. Biol.* **300**, 649–659
- Heath, K., and Scott, L. (2009) The Polymerase Incomplete Primer Exten-sion (PIPE) method applied to high-throughput cloning and site-directed mutagenesis. *Methods Mol. Biol.* **498**, 91–103
- Mishra, R. P., Mariotti, P., Fiaschi, L., Nosari, S., Maccari, S., Liberatori, S., Fontana, M. R., Pezzicoli, A., De Falco, M. G., Falugi, F., Altindis, E., Serruto, D., Grandi, G., and Bagnoli, F. (2012) Staphylococcus aureus FruD2 is involved in the early phase of staphylococcal dissemination and generates protective immunity in mice. *J. Infect. Dis.* **206**, 1041–1049
- Xiang, Z., Todd, T., Ku, K. P., Kovacic, B. L., Larson, C. B., *et al.* (2008) VIOLIN: vaccine investigation and online information network. *Nucleic Acids Res.* **36**, D923–D928
- Margarit, I., Bonacci, S., Pietrocchi, G., Rindi, S., Ghezzi, C., *et al.* (2009) Capturing host-pathogen interactions by protein microarrays: identifica-tion of novel streptococcal proteins binding to human fibronectin, fibrino-gen, and C4BP. *FASEB J.* **23**, 3100–3112
- Schluepen, C., Malito, E., Marongiu, A., Schirle, M., McWhinnie, E., Lo Surdo, P., Biancucci, M., Falugi, F., Nardi-Dei, V., Marchi, S., Fontana, M. R., Lombardi, B., De Falco, M. G., Rinaudo, C. D., Spraggon, G., Nissim, M., Bagnoli, F., Grandi, G., Bottomley, M. J., and Liberatori, S. (2013) Mining the bacterial unknown proteome: identification and char-acterization of a novel family of highly conserved protective antigens in Staphylococcus aureus. *Biochem. J.* **455**, 273–284
- Dreisbach, A., van Dijk, J. M., and Buist, G. (2011) The cell surface proteome of Staphylococcus aureus. *Proteomics* **11**, 3154–3168
- Solis, N., Larsen, M. R., and Cordwell, S. J. (2010) Improved accuracy of cell surface shaving proteomics in Staphylococcus aureus using a false-positive control. *Proteomics* **10**, 2037–2049
- Ventura, C. L., Malachowa, N., Hammer, C. H., Nardone, G. A., Robinson, M. A., Kobayashi, S. D., and DeLeo, F. R. (2010) Identification of a novel Staphylococcus aureus two-component leukotoxin using cell surface proteomics. *PLoS One* **5**, e11634
- Dreisbach, A., Hempel, K., Buist, G., Hecker, M., Becher, D., and van Dijk, J. M. (2010) Profiling the surfacome of Staphylococcus aureus. *Proteo-mics* **10**, 3082–3096
- Meinke, A., Henics, T., Hanner, M., Minh, D. B., and Nagy, E. (2005) Antigenome technology: a novel approach for the selection of bacterial vaccine candidate antigens. *Vaccine* **23**, 2035–2041
- Bombaci, M., Grifantini, R., Mora, M., Reguzzi, V., Petracca, R., *et al.* (2009) Protein array profiling of tic patient sera reveals a broad range and enhanced immune response against Group A Streptococcus antigens. *PLoS One* **4**, e6332
- Fenco, O., Frigimelica, E., Buricchi, F., Petracca, R., Galli, G., Faenzi, E., Meoni, E., Bonci, A., Agnusdei, M., Nardelli, F., Bartolini, E., Scarselli, M., Caproni, E., Laera, D., Zedda, L., Skibinski, D., Giovinnazzi, S., Bastone, R., Ianni, E., Cevenini, R., Grandi, G., and Grifantini, R. (2011) Approach to discover T- and B-cell antigens of intracellular pathogens applied to the design of Chlamydia trachomatis vaccines. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 9969–9974
- Berlanda Scorza, F., Doro, F., Rodríguez-Ortega, M. J., Stella, M., Libera-tori, S., Taddei, A. R., Serino, L., Gomes Moriel, D., Nesta, B., Fontana, M. R., Spagnuolo, A., Pizza, M., Norais, N., and Grandi, G. (2008) Proteomics characterization of outer membrane vesicles from the ex-traintestinal pathogenic Escherichia coli DeltatolR IHE3034 mutant. *Mol. Cell. Proteomics* **7**, 473–485
- Dull, P. M., and McIntosh, E. D. (2012) Meningococcal vaccine development—from glycoconjugates against MenACWY to proteins against MenB—potential for broad protection against meningococcal disease. *Vaccine* **30**, B18–B25
- Edwards, M. S., and Gonik, B. (2013) Preventing the broad spectrum of perinatal morbidity and mortality through group B streptococcal vacci-nation. *Vaccine* **31**, D66–D71