ARTICLES

nature biotechnology

Characterization and identification of vaccine candidate proteins through analysis of the group A *Streptococcus* surface proteome

Manuel J Rodríguez-Ortega, Nathalie Norais, Giuliano Bensi, Sabrina Liberatori, Sabrina Capo, Marirosa Mora, Maria Scarselli, Francesco Doro, Germano Ferrari, Ignazio Garaguso, Tiziana Maggi, Anita Neumann, Alessia Covre, John L Telford & Guido Grandi

We describe a proteomic approach for identifying bacterial surface-exposed proteins quickly and reliably for their use as vaccine candidates. Whole cells are treated with proteases to selectively digest protruding proteins that are subsequently identified by mass spectrometry analysis of the released peptides. When applied to the sequenced M1_SF370 group A *Streptococcus* strain, 68 PSORT-predicted surface-associated proteins were identified, including most of the protective antigens described in the literature. The number of surface-exposed proteins varied from strain to strain, most likely as a consequence of different capsule content. The surface-exposed proteins of the highly virulent M23_DSM2071 strain included 17 proteins, 15 in common with M1_SF370. When 14 of the 17 proteins were expressed in *E. coli* and tested in the mouse for their capacity to confer protection against a lethal dose of M23_DSM2071, one new protective antigen (Spy0416) was identified. This strategy overcomes the difficulties so far encountered in surface protein characterization and has great potential in vaccine discovery.

© 2006 Nature Publishing Group http://www.nature.com/naturebiotechnology

Bacterial surface proteins play a fundamental role in the interaction between the bacterial cell and its environment[1–6]. They are involved in adhesion to and invasion of host cells, in sensing the chemical and physical conditions of the external milieu and sending appropriate signals to the cytoplasmic compartment, in mounting defenses against host responses and in toxicity. Hence, surface proteins are potential targets of drugs aimed at preventing bacterial infections and diseases[7]. Moreover, because surface proteins are likely to interact with the host immune system, they may become components of effective vaccines. Vaccines based on surface-exposed and secreted proteins are already commercially available and others are in development[8,9].

Despite the biological relevance of bacterial surface proteins, their characterization is still incomplete. This is mostly owing to difficulties in defining the protein composition and topology on the bacterial surface. There are three main methods currently in practice to identify surface proteins. The first method is based on surface protein prediction by genome analysis using algorithms such as PSORT[10,11]. The method is rapid but is not fully reliable and is not quantitative. The second approach employs separation of membrane and cell wall fractions from the cytoplasmic fraction and then identification of proteins by two-dimensional (2D)-electrophoresis or 2D-chromatography coupled to mass spectrometry. This approach, which has been used in several bacteria[12–18], is reasonably quantitative. However, it is not sufficiently selective and several cytoplasmic proteins contaminate the identified membrane proteins. Furthermore, neither of these first two approaches specifically identifies those proteins that actually extend beyond the cell wall and polysaccharide capsule into the extracellular milieu. Finally, in the third approach, membrane proteins are first defined using one of the two methods described above and then surface localization is confirmed by producing polyclonal antibodies against the recombinant forms of each predicted protein and by assaying antibody binding to whole bacterial cells. Although this method, recently used to identify vaccine candidates for group B *Streptococcus*[9], is quantitative and does identify truly surface accessible proteins, it is extremely labor intensive. Here we describe a new procedure that allows the rapid and selective identification of bacterial surface-exposed proteins, the pool of proteins which are entirely or partially exposed on the outside of bacterial cells. The method uses proteolytic enzymes to 'shave' the bacterial surface and the peptides generated are separated from the whole cells and identified by mass spectrometry. To demonstrate the power of the method, we present the characterization of the complete set of M1_SF370 group A *Streptococcus* (GAS) surface-exposed proteins. We show that 95% of the identified proteins belong to four protein families of predicted surface-associated proteins, most of which are accessible to polyclonal antibodies raised against the corresponding recombinant proteins. We also show that the proteins identified include most of the protective antigens described in the literature and at least one new antigen capable of conferring consistent protection in the mouse against challenge with a virulent GAS strain. Therefore, the approach represents a valuable tool to study surface protein organization and to identify new vaccine candidates.

Chiron Vaccines, Via Fiorentina, 1 53100 Siena, Italy. Correspondence should be addressed to G.G. (guido_grandi@chiron.com).

Received 10 August 2005; accepted 18 November 2005; published online 15 January 2006; doi:10.1038/nbt1179

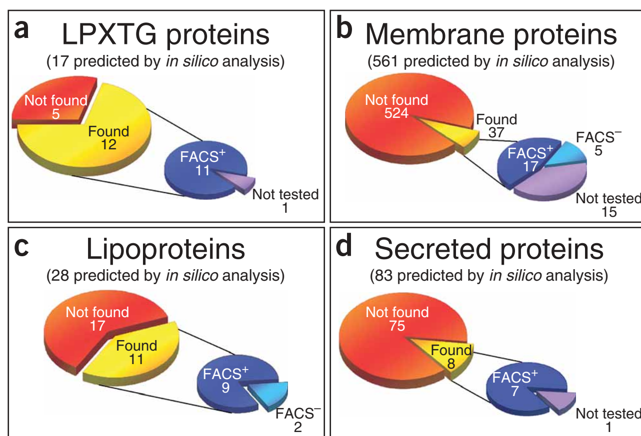

Figure 1 The M1_SF370 surface proteome. The 72 proteins belonging to the M1_SF370 surface proteome are grouped into families based on their predicted cellular location. Red areas of each pie indicate the number of PSORT-predicted proteins that have not been found in the surface proteome, whereas the yellow areas represent the number of identified proteins belonging to each protein family. The blue pies illustrate how many of the identified proteins have been confirmed to be surface-exposed by following antigen-specific antibody binding to whole bacterial cells using FACS analysis.

RESULTS

Analysis of M1_SF370 GAS strain

To identify the surface-exposed proteins of the completely sequenced M1_SF370 strain[19], exponentially growing bacteria were collected and treated with either trypsin or proteinase K to shave the bacterial surface of exposed protein domains. This treatment did not impair cell integrity as assessed by plating the bacteria before and after protease digestion (data not shown). Peptides released into the supernatant were concentrated and analyzed by tandem mass spectrometry (MS/MS). A total of 72 proteins were identified (**Fig. 1** and **Supplementary Table 1** online): 37 from the matching of more than one peptide per protein, and 35 from single-peptide matching. Forty-three proteins were deduced from the trypsin-derived peptides, 18 proteins from the proteinase K–derived peptides and 11 proteins from both trypsin and proteinase K–derived peptides. From sequence analysis, the 72 proteins could be grouped into four major families: the cell wall–anchored family containing LPXTG-like motifs (12 proteins), the lipoprotein family (11 proteins), the transmembrane protein family (37 proteins) and the secreted-protein family (8 proteins). Only four proteins predicted by PSORT to reside in the cytoplasmic compartment were found in the protease-sensitive fraction (**Supplementary Table 1** online). They included the elongation factor Tu, reported to be membrane-associated in other bacteria[20,21], two ribosomal proteins, for which there is also evidence of extracellular functions[22,23], and a hypothetical protein possibly involved in cell wall localization and side-chain formation.

Confirmation of protein surface exposure

The almost complete absence of peptides from predicted cytoplasmic proteins suggested that the procedure was selective for the identification of surface-exposed proteins. To confirm this, we carried out two

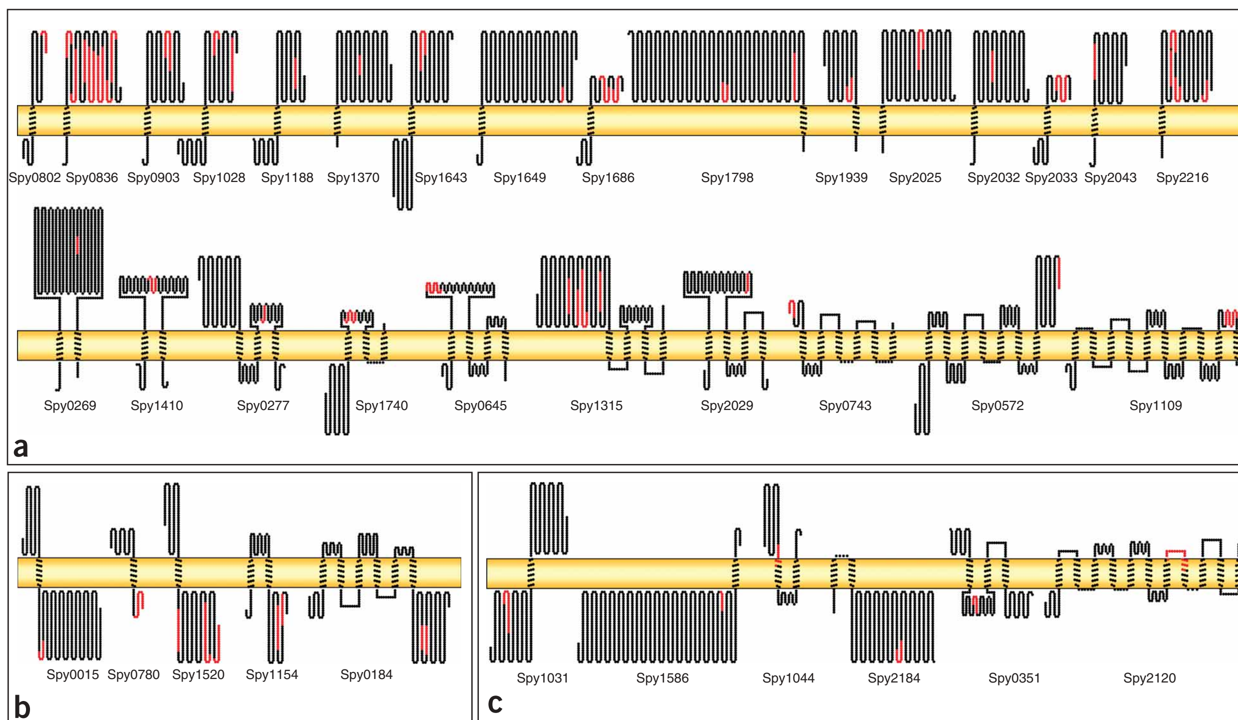

Figure 2 Topological analysis of the 37 surface-exposed proteins belonging to the membrane protein family. Each of the 37 membrane proteins belonging to the M1_SF370 surface proteome was subjected to PSORT topological prediction. (**a**) All proteins whose predicted topology is consistent with the experimental data are shown; the proteolytic peptides (shown in red in the figure) derived by protease treatment of whole cells are in fact located on domains protruding out of the membrane. (**b**,**c**) Shown are the 11 proteins showing discrepancies between *in silico*-predicted topology and experimental data (see text for details) and whose identification was deduced from the characterization of more than one (**b**) or one (**c**) peptide.

© 2006 Nature Publishing Group http://www.nature.com/naturebiotechnology

Table 1 List of reported group A *Streptococcus* protective antigens and their identification in the M1_SF370 surface proteome

Gene locus[a], protein name/function	Protein localization[b]	Readout[c]	Presence of the gene in M1_SF370	Presence of the protein in M1_SF370 surface proteome
NT01SP0102[d] + *spy0128* + *spy0130*, pilin proteins[30]	Cell wall	Survival	Yes	Yes
AAD39085, R28 (ref. 31)	Cell wall	Survival	No	No
spy0843, hypothetical protein[32]	Cell wall	Survival	Yes	Yes
spy1357, GRAB (protein G-related alpha 2M-binding protein)[33]	Cell wall	Bactericidal assay	Yes	Yes
spy2010, C5a peptidase[34]	Cell wall	Colonization	Yes	Yes
spy2018, M protein[35]	Cell wall	Survival	Yes	Yes
spyM18_2046, SPA (streptococcal protective antigen)[36]	Cell wall	Survival	No	No
sps1200, FBP54 (ref. 37)	Cell wall	Survival	No	No
M6_spy0157, Sfb I[38]	Cell wall	Survival	No	No
spy2039, SpeB[39]	Extracellular	Survival	Yes	No
sps0560, SpeA[40]	Extracellular	Survival	No	No
spy0385, Ferrichrome ABC transporter[41]	Lipoprotein	Bactericidal assay	Yes	No
spy1245, Putative phosphate ABC transporter[41]	Lipoprotein	Bactericidal assay	Yes	No
spy1274, Putative amino acid ABC transporter, periplasmic amino acid-binding protein[41]	Lipoprotein	Bactericidal assay	Yes	Yes
spy1390, Putative protease maturation protein[41]	Lipoprotein	Bactericidal assay	Yes	Yes
spy1558, hypothetical protein[41]	Lipoprotein	Bactericidal assay	Yes	No

[a]Gene locus names according to NCBI (http://www.ncbi.nlm.nih.gov/). [b]Protein localization is as predicted by PSORT (http://psort.nibb.ac.jp). The cell wall–anchoring LPXTG motifs were identified using the program Findpatterns, included in the GCG Wisconsin Package version 10.0. [c]Protection readouts are reported as described in the cited references. [d]Gene locus not annotated in NCBI; coded protein was identified from genome downloaded from TIGR (http://www.tigr.org/).

© 2006 Nature Publishing Group http://www.nature.com/naturebiotechnology

types of analysis. First, we analyzed the 37 transmembrane proteins (**Supplementary Table 1** online) with PSORT topological prediction analysis and asked whether the MS/MS-identified peptides were located within domains predicted to be on the external side of the membrane. For 26 out of 37 proteins experimental MS/MS data were consistent with PSORT predictions (**Fig. 2**). In contrast, for the remaining 11 proteins, the identified peptides mapped either in domains predicted to be intracellular or, in two cases, embedded in the membrane. However, from manual inspection of the annotations, we came to the conclusion that the PSORT-predicted transmembrane organization of at least 6 out of 11 proteins (**Supplementary Table 1** online) should be revisited. In particular, the two peptides derived from the putative cell division protein Spy0015, homologous to the FtsH protein family, are located within a conserved protein domain known to be extracellular in FtsH proteins[24–26]. The four peptides of Spy1520, a second putative cell division protein homologous to FtsZ, are located within the C-terminal part of the molecule, a region that in *Bartonella bacilliformis* is immunogenic and surface-exposed[27]. The two hypothetical proteins Spy0351 and Spy2184 are likely to be lipoproteins rather than integral membrane proteins. In fact, their predicted transmembrane regions have a very poor PSORT score and therefore we expect them to completely extend out of the membrane. This is also supported by the presence, adjacent to their predicted leader-peptide cleavage sites, of the cysteine, which in lipoproteins is

Table 2 Surface proteome of group A *Streptococcus* M23_DSM2071 strain and comparison with the M1_SF370 surface proteome

Gene locus[a], protein name/function	Protein localization[b]	Presence of the gene in SF370	Cloned in *E. coli*
spy0416, putative cell envelope proteinase	Cell wall	Yes	Yes
spy0843, hypothetical protein	Cell wall	Yes	Yes
spy1357, GRAB precursor	Cell wall	Yes	Yes
spy2010, C5A peptidase precursor	Cell wall	Yes	Yes
spy2018, M protein	Cell wall	Yes	Yes
NT01SP0246, putative oligopeptidepermease[c]	Lipoprotein	Yes	Yes
spy0604, hypothetical protein	Lipoprotein	Yes	Yes
spy1274, putative amino acid ABC transporter, periplasmic amino acid–binding protein	Lipoprotein	Yes	Yes
spy1882, putative acid phosphatase	Lipoprotein	Yes	Yes
spy0802, hypothetical protein	Membrane	Yes	Yes
spy0836, conserved hypothetical protein	Membrane	Yes	Yes
spy0903, putative ABC transporter (binding protein)	Membrane	Yes	Yes
spy1111, putative zinc-containing alcohol dehydrogenase	Membrane	No	No
spy1798, hypothetical protein sharing similarity with several eukaryotic proteins	Membrane	Yes	Yes
spy0216, putative regulatory protein, RofA related	Extracellular	No	No
spy0469, putative 42 kDa protein	Extracellular	Yes	Yes
spy0611, elongation factor Tu	Cytoplasm	Yes	No

[a]Gene loci are named according to NCBI (http://www.ncbi.nlm.nih.gov/). [b]Protein localization is as predicted by PSORT (http://psort.nibb.ac.jp). The cell wall-anchoring LPXTG motifs were identified using the program Findpatterns, included in the GCG Wisconsin Package version 10.0. [c]Gene locus not annotated in NCBI; coded protein was identified from genome downloaded from TIGR (http://www.tigr.org/).

Table 3 Protective activity of Spy0416

Antigen	Alive/challenged mice	Survival (%)	Statistical significance[a]
Spy0416	7/10	70	$P < 0.01$
Purified *E. coli* GST	0/10	0	–
M23 protein	7/8	88	$P < 0.01$

Mice were immunized three times with affinity-purified proteins (20 μg/dose) and then challenged intranasally with 10^6 CFUs of group A *Streptococcus* M23_DSM2071 strain. Survival rates were calculated 10 d after challenge. Purified Spy0416 was an N-terminal GST fusion constituted by the first 842 amino acids of the protein. Affinity-purified recombinant *E. coli* GST and purified His-tagged M23 protein were used as negative and positive controls, respectively. [a]Statistical significance was calculated by applying the Student's *t*-test.

© 2006 Nature Publishing Group http://www.nature.com/naturebiotechnology

used as the anchor residue to the bacterial surface. Spy1154 has two predicted transmembrane regions, the second of which has a very poor PSORT score (−0.32 as opposed to −8.12 of the first transmembrane region). If, as is likely, the prediction of the second transmembrane region were wrong, the C-terminal part of the molecule, which carries a typical sortase domain, would be exposed on the surface, which would be consistent with the mechanism of action of sortases[28]. Finally, Spy0184, a putative glycine betaine binding–permease protein, is predicted to have six transmembrane regions, one of which has a poor score. Again, if the weak transmembrane region were ignored, the topological organization would change and the C-terminal region, where the two MS/MS-identified peptides fall, would become surface-exposed. Indeed, polyclonal antibodies against the C-terminal domain of the protein efficiently bound GAS M1_SF370 whole cells when tested by fluorescence-activated cell sorting (FACS; data not shown).

The second type of study was based on FACS analysis using protein-specific antibodies. Mouse polyclonal antibodies were produced against 51 recombinant proteins selected from among the M1_SF370 surface-exposed proteins, and the antibodies were tested for their capacity to bind whole bacteria. All but seven sera were positive in the assay, confirming surface exposure of the proteins (**Fig. 1** and **Supplementary Table 1** online). For the remaining 21 proteins for which antisera were not available, we expect that a similar proportion would also be FACS positive. Furthermore, several of the proteins appeared well expressed/exposed on the surface, as judged by fluorescent intensity, which, in some cases, was in the same range as that observed with the major surface antigen, the M protein[29].

In conclusion, considering that (i) all proteins identified have a highly significant MASCOT score (**Supplementary Table 1** online), (ii) a large proportion of the 72 proteins is compatible with the corresponding PSORT-predicted localization and topology and (iii) a good match exists between the mass spectrometry data and the FACS data, we believe that the method leads to a reliable list of surface-exposed proteins. Obviously, we cannot rule out that some of the identified proteins may represent experimental artifacts. In this respect, we have listed the proteins on the basis of the likelihood of being truly exposed on the surface (**Supplementary Table 2** online). The highest probability score can be assigned to 52 proteins in that they were either confirmed by FACS analysis (44 proteins) or, despite the fact that FACS data were not available (5 proteins) or were negative (3 proteins), were identified with more than one peptide.

Application to vaccine discovery

From our previous experience with *Meningococcus B* and group B *Streptococcus* we know that protective antigens must be well expressed and exposed on the surface of the bacterial cell[8,9]. Hence, the comprehensive analysis of surface-exposed proteins may be an ideal approach to the identification of vaccine candidates. In support of this, 7 of the 11 reported GAS protective antigens[30–41] whose genes are present in M1_SF370 are part of the M1_SF370 surface proteome (**Table 1**). We therefore decided to investigate whether proteins identified here could elicit protective responses in a mouse model of infection and disease. To this end, we analyzed the complete set of surface-exposed proteins of M23_DSM2071, a GAS strain that, unlike M1_SF370, is highly virulent in mice. It should be noted that, because the genome sequence of M23_DSM2071 is not available, only the proteins whose genes are shared with M1_SF370 or any of the other six GAS strains whose sequences are available in the public databases (http://www.tigr.org/ and http://www.ncbi.nlm.nih.gov/) could be identified. Furthermore, in contrast to M1_SF370, M23_DSM2071 produces a robust amount of capsule (data not shown), which is expected to limit the number of proteins readily accessible to protease action. A total of 17 proteins were unambiguously identified: 5 cell wall–anchored proteins, 4 lipoproteins, 5 membrane proteins, 2 secreted proteins and 1 cytoplasmic protein (**Table 2**). All of the identified proteins have a homolog in M1_SF370 and all but two (the putative zinc-containing alcohol dehydrogenase and the putative, RofA-related, regulatory protein) were also included in the M1_SF370 surface proteome (**Supplementary Table 1** online). Of the 17 proteins specific to M23_DSM2071, 14 proteins were successfully expressed in *Escherichia coli* as either soluble histidine-fusions or glutathione *S*-transferase (GST)-fusions (**Table 2**). Each recombinant protein was used to immunize mice intraperitoneally and the mice were subsequently challenged intranasally with an LD_{90} dose (10^6 colony forming units (CFUs) in 50 μl) of M23_DSM2071 (LD_{90} is defined as the dose that kills 90% of the challenged mice). Two proteins were protective in this model: the M protein (90% survival rate) and Spy0416 (70% survival rate) (**Table 3**). Spy0416 is a putative cell envelope proteinase carrying a typical cell wall–anchoring LPXTG motif, whose protective activity has never been reported before.

DISCUSSION

In silico analysis of the complete sequence of >100 bacterial genomes predicts that surface-associated proteins constitute between 30% and 40% of all bacterial proteins. A considerable proportion of these proteins are poorly characterized in terms of function, level and kinetics of expression, and topology. Surface proteins of several bacteria have been analyzed using a variety of proteomic approaches. However, reliable methods capable of providing detailed pictures of surface protein organization in bacteria are still unavailable. Particularly challenging has been the identification of the pools of proteins that extend beyond the cell wall and polysaccharide capsule and constitute a subgroup of membrane proteins that are expected to play key roles in communicating and interacting with the environment. Moreover, in pathogenic bacteria these proteins are the most promising vaccine candidates, as we also have recently demonstrated[8,9].

GAS is responsible for a variety of clinical syndromes, ranging from uncomplicated pharyngitis and impetigo to pneumonia, sepsis, necrotizing fasciitis and streptococcal toxic shock syndrome. Furthermore, untreated streptococcal pharyngitis and skin infections may lead to acute rheumatic fever, chronic rheumatic heart disease and acute glomerulonephritis[42]. Despite the importance of this pathogen, to the best of our knowledge, the systematic analysis of its surface antigens is limited to one publication[43] in which the authors analyzed mutanolysin cell wall extracts of three different GAS strains by a combination of 2D gel/mass spectrometry analysis and immunoblots with human sera. The drawback of this approach is that digestion of the

© 2006 Nature Publishing Group http://www.nature.com/naturebiotechnology

peptidoglycan cell wall increases susceptibility to lysis and release of cytoplasmic proteins. In fact, of the 74 proteins identified in these experiments, 75% were predicted to be cytoplasmic and less than 25% were experimentally demonstrated to be on the surface.

The approach described here provides the most extensive and detailed map of the surface-exposed antigens of a GAS isolate to date. Of the 72 surface-exposed proteins identified in the M1_SF370 strain, 95% were predicted to belong to the categories of cell wall–anchored proteins, lipoproteins, transmembrane-spanning proteins and secreted proteins. Genome analysis predicts that the numbers of GAS M1_SF370 proteins constituting these protein families are 17, 28, 561 and 83, respectively (**Fig. 1**). Therefore, although a large proportion of all predicted cell wall–anchored proteins and lipoproteins were identified (70% and 40%, respectively), only a small fraction (6.5%) of these were transmembrane proteins. This indicates that the majority of these proteins are deeply embedded in the membrane and/or hidden by cell wall components, are poorly expressed or both. Although we cannot rule out that some membrane-associated proteins were not identified because they are either too resistant or too sensitive to protease digestion, we believe that the number of false-negative proteins is limited. First, most of the cell wall proteins and lipoproteins, which are designed to extend out of the membrane, were identified, whereas very few secreted proteins were identified; this is in keeping with the notion that only a minor fraction of secreted proteins remain associated with the bacteria. Second, no biochemical or structural reasons are known that make membrane proteins either more sensitive or more resistant to protease treatment than the cell wall proteins and lipoproteins are. Moreover, we have recently described the presence of pili in GAS and we have shown that pili correspond to the trypsin-resistant T antigens described by Lancefield and used for serotyping of GAS isolates[30]. Despite their trypsin-resistant nature, thanks to the use of more than one protease in the shaving step, the pilin proteins (NT01SP0102 and Spy0128) were part of the surface proteome (**Supplementary Table 1** online). Therefore, even the most recalcitrant proteins could be identified.

Several membrane proteins are known to be shielded by other surface-exposed protein complexes and, in the case of capsulated bacteria, by the polysaccharide capsule that surrounds the bacterial cells. In line with this, M23_DSM2071, a strain which produces more capsule than M1_SF370, was found to contain a relatively small number of surface-exposed proteins (17) compared to the 72 proteins of M1_SF370. Furthermore, we have recently found that M3_CDCSS-90, a highly virulent strain that, when grown in liquid culture, has three times as much capsular polysaccharide as M1_SF370 (data not shown), contained only ten major surface-exposed proteins (**Supplementary Table 3** online). Although the genome sequences of both M3_CDCSS-90 and M23_DSM2071 are not available and therefore a few proteolytic peptides could not have been identified because of the sequence variation and gene variability within strains, overall our data indicate that the capsule plays a role in determining the extent of protein accessibility on the bacterial surface. This is consistent with what we have recently demonstrated for a group B *Streptococcus* protein, which was only accessible to antibody recognition in a capsule-negative mutant, and not in the parental wild-type encapsulated strain[9].

An important aspect of the approach proposed here is its utility in refining the protein topology derived from computer-assisted predictors such as PSORT. Because the procedure selectively identifies peptides proteolytically generated from surface-exposed domains, by matching the topology prediction with peptide data, one can either confirm or refine the membrane organization of each identified protein. In general, we found a good correlation between theoretical and experimental data. However, the topology of 30% of the transmembrane proteins identified did not fit with peptide data. Manual inspection of each transmembrane protein highlighted the difficulty that algorithms such as PSORT encounter in assigning topological organization of proteins carrying transmembrane regions with low probability scores. In this respect, the procedure could be successfully used in future refinements of current algorithms.

The search for effective vaccines to prevent group A streptococcal infections has been ongoing for many decades. Based on the original observation by Lancefield[44] that protection against GAS correlates in humans with the levels of bactericidal antibodies, the search for vaccine candidates has been focused on surface antigens. Indeed, the most promising vaccine candidate so far described is the surface M protein, which is a primary virulence determinant and which is used for GAS serotyping[45]. However, the M protein is a highly variable antigen such that GAS is currently classified in more than 100 different serotypes. Therefore, although vaccine studies based on M protein are well underway[2,46], the identification of other, more conserved protective surface antigens would be highly desirable. A relevant result from our work is the demonstration that comprehensive characterization of surface-exposed proteins can lead to new vaccine candidate discovery. Among the 14 identified surface proteins tested, one protein, Spy0416, conferred high protection levels. This is a remarkable result, considering the paucity of protective antigens that have been identified to date. Spy0416 is a 1,647 amino acid cell wall protein, which shares 48% similarity with the C5a peptidase precursor and has a calcium-dependent serine protease activity[47]. Spy0416 has a homolog in group B *Streptococcus* (cspA) that was proposed to be involved in virulence by potentially protecting the bacterium from opsonophagocytic killing[48]. Interestingly, Lei and coworkers recently found that a 31-kDa, N-terminal fragment of Spy0416 is released in the supernatant of GAS cultures and that the protein is well recognized by sera from GAS-infected patients[49]. Based on the seven available genome sequences, the protein appears to be highly conserved (over 98% identity) and preliminary data on surface expression on a panel of 20 different GAS strains indicate that Spy0416 is a major component of over 70% of circulating strains (S.C. and G.B., unpublished data).

METHODS

Bacterial surface digestion. *Streptococcus pyogenes* M1_SF370, M3_CDCSS-90 and M23_DSM2071 (DSMZ) were grown in Todd-Hewitt broth (THB) at 37 °C and 5% CO_2, until an OD_{600} of 0.4 (exponential phase) was reached. Bacteria were harvested by centrifugation at 3,500*g* for 10 min at 4 °C, and washed three times with PBS. Cells were resuspended in one-hundredth volume of PBS containing 40% sucrose (pH 7.4 for trypsin digestion and pH 6.0 for proteinase K digestion). Digestions were carried out with 20 μg trypsin (Promega) or 5 μg proteinase K (Sigma) in the presence of 5 mM DTT, for 30 min at 37 °C. The digestion mixtures were centrifuged at 3,500*g* for 10 min at 4 °C, and the supernatants (containing the peptides) were filtered using 0.22-μm pore-size filters (Millipore). Protease reactions were stopped with formic acid at 0.1% final concentration. Before analysis, salts were removed by off-line high-performance liquid chromatography (HPLC), with a 7-min gradient of 2–80% acetonitrile in 0.1% formic acid. Peptide fractions were concentrated with a Speed-vac centrifuge (Savant), and kept at –20 °C until further analysis.

Protein identification technology by nano-LC/MS/MS. Two different platforms were used for the chromatographic separation of peptides and further identification by tandem mass spectrometry (MS/MS). In the first, peptides were separated by 2-D-nano-liquid chromatography (LC; Dionex). In the first dimension, peptides were loaded on a strong cation exchange (SCX) column

(10 cm × 320 μm inner diameter (i.d.), packed with cross-linked particles functionalized with sulfoethyl groups) and eluted isocratically by applying five NaCl solutions of increasing concentrations (0.01, 0.05, 0.1, 0.5 and 1 M). In the second dimension, peptides were separated by a reversed-phase C18 analytical column (15 cm × 75 μm i.d., C18 PepMap100, 3 μm, 100 Å) through a C18 trap column (PepMap C18 μ-precolumn, 300 μm i.d. × 1 mm). Peptides were eluted with a 45-min gradient of 5–50% of 80% acetonitrile in 0.1% formic acid at a flow rate of 300 nl/min. Eluates were continuously spotted onto an Anchor-Chip matrix-assisted laser desorption ionization (MALDI) target (Bruker Daltoniks), prepared with a thin layer of a saturated solution of α-cyano-4-hydroxycynnamic acid in acetone, every 60 s using a Proteineer FC robot (Bruker Daltoniks). After fraction collection, every spot was manually recrystallized with 0.6 μl of ethanol/acetone/0.1% trifluoroacetic acid (6:3:1). Mass spectrometry analysis was performed automatically with an Ultraflex MALDI-time-of-flight (TOF) instrument, under the control of the WARP LC software, version 1.0 (Bruker Daltoniks). First, MS spectra of all the spotted fractions were acquired for peak selection over the *m/z* range of 900–4,000, after which MS/MS spectra of selected peaks were obtained. After spectra acquisition, the individual MS/MS spectra acquired for each of the precursor ions were combined, smoothed and centroided using FlexAnalysis software, version 2.4 (Bruker Daltoniks), and then transferred to BioTools software, version 3.0 (Bruker Daltoniks) to create a batch .xml file containing the peak lists of any of the individual MS/MS spectra. Search and identification of peptides were performed in batch mode with a licensed version of MASCOT, in a local database containing the 1,819 proteins derived from the complete genome sequences of *S. pyogenes* (downloaded from http://www.tigr.org/ and http://www.ncbi.nlm.nih.gov/). The MASCOT search parameters were (i) species, *S. pyogenes* strain SF370; (ii) allowed number of missed cleavages (only for trypsin digestion), 6; (iii) variable post-translational modifications, methionine oxidation; (iv) peptide tolerance, ±100 p.p.m.; (v) MS/MS tolerance, ±0.75 Da and (vi) peptide charge, +1. Only significant hits as defined by MASCOT probability analysis were considered. The score thresholds for acceptance of protein identifications from at least one peptide were set by MASCOT as 18 for trypsin digestion and 36 for proteinase K digestion.

In the second platform, peptides were separated by nano-LC on a CapLC HPLC system (Waters) connected to a Q-ToF Micro Electron Spray Ionization (ESI) mass spectrometer equipped with a nanospray source (Waters). Samples were loaded onto an Atlantis C18 NanoEase column (100 μm i.d. × 100 mm, Waters), through a C18 trap column (300 μm i.d. × 5 mm, Dionex). Peptides were eluted with a 50-min gradient of 2–60% of 95% acetonitrile, in a solution of 0.1% formic acid at a flow rate of 400 nl/min. The eluted peptides were subjected to an automated data-dependent acquisition program, using the MassLynx software, version 4.0 (Waters), where a MS survey scan was used to automatically select multicharged peptides over the *m/z* range of 400–2,000 for further MS/MS fragmentation. Up to three different components were subjected to MS/MS fragmentation at the same time. For all the samples, a second nano-LC-MS/MS analysis was carried out for the selective fragmentation of mono-charged peptide species. After data acquisition, the individual MS/MS spectra were combined, smoothed and centroided by MassLynx. Search and identification of peptides were performed in batch mode with a licensed version of MASCOT, in a local database (see above), after converting the acquired MS/MS spectra in .pkl file. The MASCOT search parameters were (i) species, *S. pyogenes* genome sequences; (ii) allowed number of missed cleavages (only for trypsin digestion), 6; (iii) variable post-translational modifications, methionine oxidation; (iv) peptide tolerance, ±500 p.p.m.; (v) MS/MS tolerance, ±0.3 Da and (vi) peptide charge, from +1 to +4. As for the previous platform, only significant hits as defined by MASCOT probability analysis were considered. The score thresholds for acceptance of protein identifications from at least one peptide were set by MASCOT as 18 for trypsin digestion and 36 for proteinase K digestion.

Expression and purification of GAS proteins and preparation of immune sera. Cloning, expression and purification of recombinant proteins, and preparation of immune sera, were performed as described[9]. Briefly, the chromosomal DNA of M1_SF370 was used for gene amplification. PCR primers were designed so as to amplify genes without predicted signal-peptide coding sequences. PCR products were introduced into plasmid expression vectors so as to generate recombinant proteins fused either with 6-His or glutathione GST. His-tagged proteins were obtained by cloning in pET21b+ vectors (Novagen). *E. coli* BL21DE3 cells (Novagen) were used as the recipient. GST fusions were obtained by using pGEX-NN plasmids and *E. coli* BL21SI (Novagen) as the recipient. Affinity-purified proteins (three doses of 20 μg/dose) were used to immunize groups of adult CD1 mice (six mice per group). Mice were bled 2 weeks after the third immunization. The sera of each group were pooled and stored at −80 °C.

Capsule quantification by colorimetric assay. GAS strains were grown in 10 ml of THB and when the cultures reached an OD_{600} of 0.4, cells were collected by centrifugation, washed twice with PBS and resuspended in 0.5 ml H_2O. Hyaluronic acid polysaccharide was extracted twice by adding an equal volume of chloroform, and the organic phase was collected by centrifugation. For hyaluronic acid determination, 100 μl of properly diluted organic phase were added to 1.9 ml of a solution containing 0.2% (wt/vol) Stains All Reagent (Sigma), 0.6% acetic acid and 50% formamide, and absorbance was measured at 640 nm and compared with the standard curve of hyaluronic acid experimentally defined using known amounts of polysaccharide.

FACS analysis of surface proteins. Bacteria were grown in THB to an OD_{600} of 0.4, washed twice with PBS, suspended in newborn calf serum (NCS, Sigma), incubated for 20 min at 23 °C and dispensed into a 96-well (20 μl/well). Eighty μl of preimmune or immune mouse sera, diluted in PBS containing 0.1% BSA, was added to the bacterial suspension to a final dilution of 1:200 and incubated on ice for 30 min. After washing twice with 0.1% BSA in PBS, bacteria were incubated on ice for 30 min in 10 μl of goat anti-mouse IgG, $F(ab')_2$ fragment-specific-R-phycoerythrin-conjugated (Jackson Immunoresearch Laboratories) diluted 100-fold in PBS containing 0.1% BSA, 20% NCS. After incubation, bacteria were washed with PBS/0.1% BSA, suspended in 200 μl PBS and analyzed using a FACS Calibur cytometer (Becton Dickinson) and Cell Quest Software (Becton Dickinson).

Screening of protective antigens in the mouse model. Five-week-old female CD1 mice (ten mice/group) were immunized with 20 μg of recombinant proteins administered intraperitoneally at 0, 21 and 35 d with complete Freund adjuvant (priming dose) and incomplete Freund adjuvant (booster doses). Blood samples were collected before the first and after the third immunizations. Immunized mice were challenged intranasally with 50 μl of THB containing 10^6 CFUs of M23_DSM 2071 strain grown in THB broth at an OD_{600} of 0.4. CFU number of the infecting dose was verified by plating on 5 TH plates supplemented with 0.5% yeast extract and 5% defibrinated sheep blood. After challenge, mice were controlled daily for survival and the final survival rate was calculated 10 d thereafter.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS
We thank J.M. Musser and G. Orefici for providing some of the strains used in this work and Marco Tortoli for his expert assistance in animal studies. We are also grateful to Giorgio Corsi for the artwork and Antonietta Maiorino for her expert secretarial assistance. Manuel J. Rodríguez-Ortega had an Intra-European Marie Curie postdoctoral fellowship from the EU. This work was in part supported by National Institutes of Health/ National Institute of Allergy and Infectious Diseases grant number 5U1060595-02, and in part by the Italian Ministry of Education, University and Research grant no. 10811.

COMPETING INTERESTS STATEMENT
The authors declare competing financial interests (see the *Nature Biotechnology* website for details).

Published online at http://www.nature.com/naturebiotechnology/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Lindahl, G., Stalhammar-Carlemalm, M. & Areschoug, T. Surface proteins of *Streptococcus agalactiae* and related proteins in other bacterial pathogens. *Clin. Microbiol. Rev.* **18**, 102–127 (2005).
2. Courtney, H.S., Dale, J.B. & Hasty, D.L. in *Handbook of Bacterial Adhesion: Principles, Methods and Applications* (eds. An, Y.N. & Friedman, R.J.) 553–579 (Human Press, Totowa, NJ, 2002).

© 2006 Nature Publishing Group http://www.nature.com/naturebiotechnology

3. Lin, J., Huang, S. & Zhang, Q. Outer membrane proteins: key players for bacterial adaptation in host niches. *Microbes Infect.* **4**, 325–331 (2002).
4. Niemann, H.H., Schubert, W.D. & Heinz, D.W. Adhesins and invasins of pathogenic bacteria: a structural view. *Microbes Infect.* **6**, 101–112 (2004).
5. Ton-That, H., Marraffini, L.A. & Schneewind, O. Protein sorting to the cell wall envelope of Gram-positive bacteria. *Biochim. Biophys. Acta* **1694**, 269–278 (2004).
6. Janulczyk, R. & Rasmussen, M. Improved pattern for genome-based screening identifies novel cell wall–attached proteins in gram-positive bacteria. *Infect. Immun.* **69**, 4019–4026 (2001).
7. Galperin, M.Y. & Koonin, E.V. Searching for drug targets in microbial genomes. *Curr. Opin. Biotechnol.* **10**, 571–578 (1999).
8. Pizza, M. *et al.* Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* **287**, 1816–1820 (2000).
9. Maione, D. *et al.* Identification of a universal group B *Streptococcus* vaccine by multiple genome screen. *Science* **309**, 148–150 (2005).
10. Nakai, K. Protein sorting signals and prediction of subcellular localization. *Adv. Protein Chem.* **54**, 277–344 (2000).
11. Doytchinova, I.A., Taylor, P. & Flower, D.R. Proteomics in vaccinology and immunobiology: an informatics perspective of the immunone. *J. Biomed. Biotechnol.* **2003**, 267–290 (2003).
12. Molloy, M.P. *et al.* Proteomic analysis of the *Escherichia coli* outer membrane. *Eur. J. Biochem.* **267**, 2871–2881 (2000).
13. Phadke, N.D. *et al.* Analysis of the outer membrane proteome of *Caulobacter crescentus* by two-dimensional electrophoresis and mass spectrometry. *Proteomics* **1**, 705–720 (2001).
14. Molloy, M.P. *et al.* Profiling the alkaline membrane proteome of *Caulobacter crescentus* with two-dimensional electrophoresis and mass spectrometry. *Proteomics* **2**, 899–910 (2002).
15. Nouwens, A.S. *et al.* Complementing genomics with proteomics: the membrane subproteome of *Pseudomonas aeruginosa* PAO1. *Electrophoresis* **21**, 3797–3809 (2000).
16. Rhomberg, T.A. *et al.* Proteomic analysis of the sarcosine-insoluble outer membrane fraction of the bacterial pathogen *Bartonella henselae*. *Proteomics* **4**, 3021–3033 (2004).
17. Sabarth, N. *et al.* Identification of surface proteins of *Helicobacter pylori* by selective biotinylation, affinity purification, and two-dimensional gel electrophoresis. *J. Biol. Chem.* **277**, 27896–27902 (2002).
18. Guina, T. *et al.* Proteomic analysis of *Pseudomonas aeruginosa* grown under magnesium limitation. *J. Am. Soc. Mass Spectrom.* **14**, 742–751 (2003).
19. Ferretti, J.J. *et al.* Complete genome sequence of an M1 strain of *Streptococcus pyogenes*. *Proc. Natl Acad. Sci. USA* **98**, 4658–4663 (2001).
20. Marques, M.A., Chitale, S., Brennan, P.J. & Pessolani, M.C. Mapping and identification of the major cell wall-associated components of *Mycobacterium leprae*. *Infect. Immun.* **66**, 2625–2631 (1998).
21. Dallo, S.F., Kannan, T.R., Blaylock, M.W. & Baseman, J.B. Elongation factor Tu and E1 beta subunit of pyruvate dehydrogenase complex act as fibronectin binding proteins in *Mycoplasma pneumoniae*. *Mol. Microbiol.* **46**, 1041–1051 (2002).
22. Spence, J.M. & Clark, V.L. Role of ribosomal protein L12 in gonococcal invasion of Hec1B cells. *Infect. Immun.* **68**, 5002–5010 (2000).
23. Kurar, E. & Splitter, G.A. Nucleic acid vaccination of *Brucella abortus* ribosomal L7/L12 gene elicits immune response. *Vaccine* **15**, 1851–1857 (1997).
24. Tomoyasu, T. *et al.* Topology and subcellular localization of FtsH protein in *Escherichia coli*. *J. Bacteriol.* **175**, 1352–1357 (1993).
25. Akiyama, Y., Kihara, A., Mori, H., Ogura, T. & Ito, K. Roles of the periplasmic domain of *Escherichia coli* FtsH (HflB) in protein interactions and activity modulation. *J. Biol. Chem.* **273**, 22326–22333 (1998).
26. Amara, R.R., Shanti, S. & Satchidanandam, V. Characterization of novel immunodominant antigens of *Mycobacterium tuberculosis*. *Microbiology* **144**, 1197–1203 (1998).
27. Padmalayam, I., Anderson, B., Kron, M., Kelly, T. & Baumstark, B. The 75-kilodalton antigen of *Bartonella bacilliformis* is a structural homolog of the cell division protein FtsZ. *J. Bacteriol.* **179**, 4545–4552 (1997).
28. Paterson, G.K. & Mitchell, T.J. The biology of Gram-positive sortase enzymes. *Trends Microbiol.* **12**, 89–95 (2004).
29. Lancefield, R.C. Studies on the antigenic composition of group A hemolytic *Streptococci*. *J. Exp. Med.* **78**, 465–476 (1943).
30. Mora, M. *et al.* Group A Streptococcus produce pilus-like structures containing protective antigens and Lancefield T antigens. *Proc. Natl Acad. Sci. USA* **102**, 15641–15646 (2005).
31. Stalhammar-Carlemalm, M., Areschoug, T., Larsson, C. & Lindahl, G. The R28 protein of *Streptococcus pyogenes* is related to several group B streptococcal surface proteins, confers protective immunity and promotes binding to human epithelial cells. *Mol. Microbiol.* **33**, 208–219 (1999).
32. Reid, S.D. *et al.* Postgenomic analysis of four novel antigens of group *A Streptococcus*: growth phase-dependent gene transcription and human serologic response. *J. Bacteriol.* **184**, 6316–6324 (2002).
33. McMillan, D.J. *et al.* Identification and assessment of new vaccine candidates for group A streptococcal infections. *Vaccine* **22**, 2783–2790 (2004).
34. Ji, Y., Carlson, B., Kondagunta, A. & Cleary, P.P. Intranasal immunization with C5a peptidase prevents nasopharyngeal colonization of mice by the group A *Streptococcus*. *Infect. Immun.* **65**, 2080–2087 (1997).
35. Massell, B.F., Michael, J.G., Amezcua, J. & Siner, M. Secondary and apparent primary antibody responses after group A streptococcal vaccination of 21 children. *Appl. Microbiol.* **16**, 509–518 (1968).
36. Dale, J.B., Chiang, E.Y., Liu, S., Courtney, H.S. & Hasty, D.L. New protective antigen of group A *Streptococci*. *J. Clin. Invest.* **103**, 1261–1268 (1999).
37. Kawabata, S. *et al.* Systemic and mucosal immunizations with fibronectin-binding protein FBP54 induce protective immune responses against *Streptococcus pyogenes* challenge in mice. *Infect. Immun.* **69**, 924–930 (2001).
38. Guzman, C.A., Talay, S.R., Molinari, G., Medina, E. & Chhatwal, G.S. Protective immune response against *Streptococcus pyogenes* in mice after intranasal vaccination with the fibronectin-binding protein SfbI. *J. Infect. Dis.* **179**, 901–906 (1999).
39. Kuo, C.F. *et al.* Role of streptococcal pyrogenic exotoxin B in the mouse model of group A streptococcal infection. *Infect. Immun.* **66**, 3931–3935 (1998).
40. Roggiani, M. *et al.* Toxoids of streptococcal pyrogenic exotoxin A are protective in rabbit models of streptococcal toxic shock syndrome. *Infect. Immun.* **68**, 5011–5017 (2000).
41. Lei, B., Liu, M., Chesney, G.L. & Musser, J.M. Identification of new candidate vaccine antigens made by *Streptococcus pyogenes*: purification and characterization of 16 putative extracellular lipoproteins. *J. Infect. Dis.* **189**, 79–89 (2004).
42. Cunningham, M.W. Pathogenesis of group A streptococcal infections. *Clin. Microbiol. Rev.* **13**, 470–511 (2000).
43. Cole, J.N. *et al.* Surface analyses and immune reactivities of major cell wall–associated proteins of group A *Streptococcus*. *Infect. Immun.* **73**, 3137–3146 (2005).
44. Lancefield, R.C. Persistence of type-specific antibodies in man following infection with group A streptococci. *J. Exp. Med.* **110**, 271–292 (1959).
45. Lancefield, R.C. Current knowledge of type-specific M antigens of group A *Streptococci*. *J. Immunol.* **89**, 307–313 (1962).
46. Hu, M.C. *et al.* Immunogenicity of a 26-valent group A streptococcal vaccine. *Infect. Immun.* **70**, 2171–2177 (2002).
47. Fernandez-Espla, M.D., Garault, P., Monnet, V. & Rul, F. *Streptococcus thermophilus* cell wall–anchored proteinase: release, purification, and biochemical and genetic characterization. *Appl. Environ. Microbiol.* **66**, 4772–4778 (2000).
48. Harris, T.O., Shelver, D.W., Bohnsack, J.F. & Rubens, C.E. A novel streptococcal surface protease promotes virulence, resistance to opsonophagocytosis, and cleavage of human fibrinogen. *J. Clin. Invest.* **111**, 61–70 (2003).
49. Lei, B., Mackie, S., Lukomski, S. & Musser, J.M. Identification and immunogenicity of group A *Streptococcus* culture supernatant proteins. *Infect. Immun.* **68**, 6807–6818 (2000).

© 2006 Nature Publishing Group http://www.nature.com/naturebiotechnology