Pneumonia is one of the most common and severe diseases associated with Streptococcus pneumoniae infections in children and adults. Etiological diagnosis of pneumococcal pneumonia in children is generally challenging because of limitations of diagnostic tests and interference with nasopharyngeal colonizing strains. Serological assays have recently gained interest to overcome some problems found with current diagnostic tests in pediatric pneumococcal pneumonia. To provide insight into this field, we have developed a protein array to screen the antibody response to many antigens simultaneously. Proteins were selected by experimental identification from a collection of 24 highly prevalent pediatric clinical isolates in Spain, using a proteomics approach consisting of “shaving” the cell surface with proteases and further LC/MS/MS analysis. Ninety-five proteins were recombinantly produced and printed on an array. We probed it with a collection of sera from children with pneumococcal pneumonia. From the set of the most seroprevalent antigens, we obtained a clear discriminant response for a group of three proteins (PblB, PulA, and PrtA) in children under 4 years old. We validated the results by ELISA and an immunostrip assay showed the translation to easy-to-use, affordable tests. Thus, the protein array here developed presents a tool for broad use in serodiagnosis. Molecular & Cellular Proteomics 14: 10.1074/mcp.M115.049544, 2591–2608, 2015.

Streptococcus pneumoniae, also known as the pneumococcus, is a Gram-positive pathogen recognized as a major cause of pneumonia worldwide (1). It resides as a commensal in the nasopharynx of healthy carriers, but in susceptible individuals this bacterium can spread to other body locations and cause disease. The main group risks are the elderly, immunocompromised people and infants. In fact, ~800,000 children die each year because of pneumococcal disease, of which >90% of these deaths occur in developing countries (2, 3). In addition, a high number of pneumococcal infection cases are diagnosed in the developed countries and can be associated with high morbidity in children and are an important factor that influences quality of life and produces significant mortality in adults (4, 5). There are licensed polysaccharide-based vaccines to prevent pneumococcal infections, but their efficacy is limited (3). Therefore, pneumococcal pneumonia remains as an important health problem and once it has occurred, early diagnosis with accurate diagnostic methods is essential in order to provide patients with prompt and appropriate therapy and hence to improve outcome (1).

Although the major burden of pneumococcal infections is caused by pneumonia, the ability to identify S. pneumoniae as a causative agent in lung infections in children is quite limited. Blood cultures are often negative (6, 7). The BinaxNOW test, which measures teichoic acid, is less specific in children than in adults, because healthy carriage in infants can produce false positive results (8). The amplification and quantification of pneumococcal genes (namely spn9802, ply or pcpA) by PCR has been also used, but with lower sensitivity than culture in blood samples in adults and inability to discriminate between carriage and disease in nasopharyngeal and sputum samples (9–11).

The detection of antibody serological markers by any immunoassay is widely used for early diagnosis, epidemiological surveillance, or evaluation of vaccine immunogenicity against many pathogens, including the pneumococcus (6, 12, 13). Serological diagnosis of pneumococcal disease based on a single antigen is often challenging, because of the interference of natural antibodies elicited by previous colonization events. Therefore, to better discriminate between diseased and healthy people, a combination of antigens would be desirable. To this regard, proteomics offers an excellent platform to develop the necessary more sensitive and specific
immunoassays that can be used for the aforementioned purposes.

It is well assumed that surface proteins are those with the highest chance to raise an effective immune response against pathogen infection, as they are sufficiently exposed and accessible to both T and B cells (14, 15). Protein arrays are powerful tools to interrogate the pattern of host humoral responses to infections (16), which allows the study of many antigens simultaneously with a small amount of sample (17) to select a set of antigens with optimal sensitivity and specificity (18, 19). In this work, we have selected a set of 95 pneumococcal surface proteins by experimental identification, using the proteomic approach of “shaving” of live cells with proteases and further liquid chromatography-tandem MS (LC/MS/MS) analysis (20). After producing the selected proteins as recombinant fragments, we have developed the first pneumococcal surface protein chip and probed it with a collection of sera from infected and control children, in order to find proteins that differentiate between pneumococcal or nonpneumococcal infection/health status. Three proteins were proven to discriminate with optimal sensitivity, specificity, and accuracy between nonpneumococcal disease and disease status for <4-year-old children. As a proof-of-concept, we have developed an immunoassay array with such proteins, obtaining the same sensitivity, specificity, and accuracy, thus demonstrating the power of high-throughput technologies for discovering diagnostic biomarkers of infection and its possible translation to an easy-to-use clinical tool.

EXPERIMENTAL PROCEDURES

Ethics Statement for Human Sera Sampling and Use—This research was performed according to the principles expressed in the Declaration of Helsinki. All human sera (DS1) were obtained from patients admitted to Hospital Universitario Infantil Virgen del Rocío (HUVR) in Seville, Spain. All human sera were collected from children <14 years old. Sera were drawn either from patients with a diagnosis of pneumococcal pneumonia (the “patient” group), based on clinical features, radiological imaging, and isolation of the microorganism from a sterile site (blood or pleural fluid), or from healthy children or patients affected by other pathologies different from pneumococcal pneumonia (the “control” group). All sera from patients with pneumococcal pneumonia were obtained within 10 days of hospital admission. Two different sera sets were collected: 71 sera for the protein array test set, and 24 sera for the validation set. Written informed consent was obtained from parents or legal guardians of participating children and the Hospital Universitario Virgen del Rocío Ethic Committee approved the study (code no. 010470, certificate no. 14/2010), for sera to be used within the project in which this work was designed.

1 The abbreviations used are: LC/MS/MS, liquid chromatography-tandem MS; AOM, Acute otitis media; AUC, Area under the curve; CDM, Chemically defined medium; CI, Confidence interval; FDR, False Discovery Rate; GO, Gene ontology; IPD, Invasive pneumococcal disease; MLST, Multilocus sequence typing; PBST, PBS with 0.05% Tween-20; PCV7, Polysaccharide conjugate 7-valent vaccine; PE, pleural empyema; PMEN, Pneumococcal Molecular Epimology Network; SMI, Signal mean intensity; ST, Sequence type; TMD, Transmembrane domain.
belonging to another strain was found, homology with a corresponding protein of strain R6 was given by using protein-BLAST. If homology with R6 was not observed, then the protein accession numbers of other strains were used. Primary predictions of subcellular localization were assigned by using the web-based algorithm LocateP (http://www.cmbi.ru.nl/locatep-db/cgi-bin/locatepdb.py) (25).

Production of Recombinant Proteins—Recombinant proteins were produced as double fusion fragments containing an N-terminal GST fragment and a C-terminal His-tag using the pSpark® I vector (Canvax Biotech, Córdoba, Spain), and expressed in Escherichia coli BL21, as described (21) and according to manufacturers’ instructions. Briefly, recombinant products were purified either by Ni²⁺-agarose affinity chromatography from the E. coli intracellular fraction, dialyzed against PBS and used for protein array printing after measuring the protein concentration by the Bradford assay (26). All the SprXXX proteins were expressed from the R6 strain. The proteins annotated as SP_XXXX were produced from the TIGR4 strain. The pblB gene (annotated as sph_0062 in Hungary 19A-6 strain) was cloned from the isolate #418; the gene SP70585_2286, from the 70585 strain, was also cloned the isolate #418; and the gene SPJ_1652, from the 670-6B strain, was cloned from the isolate #49H. All the primers were designed using the genomes to which the annotated genes belonged. Protein Microarray Fabrication and Probing—Affinity-purified recombinant proteins were printed on glass slides in quintuplicate (6 ng/spot) as detailed in Fig. 1A with split pins (4 x 4 pin tool) using a robotic array spotter (Genomic Solutions, BioRobotics MicroGrid II 610, Huntington, UK). Proteins were distributed into 384-well plates at 2 wells per sample and 30 µl per well. Each component was prepared at 250 µg/ml in printing buffer (150 mM phosphate, pH 8.5, 0.01% sarkosyl) onto Nexterion Slide H 6-D glass slides. As negative controls, we used 12 commercially available irrelevant (i.e. nonrelated to pneumococcus) proteins from different biological sources (supplemental Table S2). The pins were dwelled into the sample wells and blotted 15 times before printing. The humidity level in the arraying chamber was maintained at 55–60% during printing. Each of the components was printed five times in a grid of 140 µm diameter spots with 175-µm pitch. Eight complete arrays were printed on each slide. Printed slides were placed in a slide humidity chamber overnight at 75% relative humidity and stored at -20 °C until use. Probing with human sera was carried out in duplicate for each serum sample. Slides were blocked with 25 mM ethanolamine in 100 mM sodium borate buffer (final pH 8.5) and washed three times for 1 min in PBST and once for 1 min in H2O. Then the slides were allowed to dry by centrifugation (350 x g for 15 min). After that, they were assembled on 16-well slide holders (Nexterion Slide H MPX 16, Schott, Louisville, KY) and 45 µl of a dilution of different sera from the test set (1:200 in PBST) were incubated for 1 h protected from light at room temperature. The different samples were washed twice with 100 µl of PBST for 2 min and then incubated with anti-human IgG-Cy3 (1:1000) or anti-human IgM-Cy5 (1:200), covered tightly with a seal strip, and incubated for 1 h at room temperature. The slides were removed from holders, washed twice for 10 min in PBST, then once in PBS for 10 min and finally centrifuged 350 x g for 15 min. To process the array data, the slides were scanned with a Genepix 4000B microarray scanner (Molecular Devices Corporation, Union City, CA) at photomultiplier voltage settings so that no saturated pixels were obtained. Image analysis was carried out with Genepix Pro 4.1 analysis software (Molecular Devices Corporation). Spots were defined as circular features with maximum diameter of 140 µm. Local background subtraction was performed and corrected median feature intensity was used for initial data processing.

ELISA—To validate the immunoreactivity results obtained by the protein microarray, the significantly discriminant proteins were validated by ELISA using the validation set sera (SDS1). The proteins were coupled to the plate individually or in combination at 1 µg/position. The sera were used at a 1:100 dilution. As secondary antibody, anti-human IgG coupled to peroxidase was used at a 1:1000 dilution. Reaction was developed and stopped according to manufacturer’s instructions and the plate was read at 450 nm.

Immunoblot Printing and Probing—Antibody-based detection of proteins in immunoreactive strips was performed using 1 µg of pneumococcus recombinant proteins Spr0247, Spr0561, Sph_0062, and a 1:1:1 mixture of them. As negative controls, 1 µg of Lys9 (His-tag recombinant protein of S. cerevisiae expressed in E. coli BL21 with pET/100 TOPO cloning system, Invitrogen, Madrid, Spain, according to the manufacturer’s instructions) and 1 µg of commercial available trypsin (Promega) were used. As positive controls, 5 µg of pneumococcal serotype 8 strain total protein extract and 0.5 µg of commercial anti-human IgG produced in goat (Invitrogen) were used. Proteins and extracts were transferred to a nitrocellulose membrane and air dried. Nonspecific sites were blocked by incubation with 5% nonfat milk in T-TBS for 45 min. After two washes with T-TBS, a second 1 h incubation of the membrane with the validation set sera (SDS1), diluted 1:200 in 3% nonfat milk in T-TBS, was carried out. As secondary antibody, rabbit anti-human IgG conjugated to horseradish peroxidase (Sigma, St. Louis, MO), diluted 1:10,000 in T-TBS, was used. After 1 h incubation, membranes were washed three times with TBS and developed with ECL Plus Western blotting Detection System (GE Healthcare, Barcelona, Spain, according to the manufacturer’s instructions). Densitometric analysis was performed using ImageJ v1.48 software.

Data and Statistical Analysis—For analysis of antibody binding to recombinant fragments on the microarray, local background subtraction from 10 surrounding spots was performed and corrected median fluorescence intensity was used for initial data processing. Then, the mean background signal of negative controls was subtracted from each raw spot value after sera hybridization. Negative controls represented hybridizations of nonpneumococcal proteins and buffer spots with sera and secondary antibodies. Both in nonpneumococcal proteins and buffer positions, no reaction with human sera was observed. After background subtraction, negative or zero values were assigned a net value of 0. Then, outlier values for each spot were removed. The two different hybridizations for each serum were averaged to report the signal mean intensity (SMI) values, and the mean and standard deviation (S.D.) were obtained from the five printed spots per protein in each patient and control groups. Finally, data normalization by background was carried out using Microsoft Excel as described (27). Absolute SMI values <500 were not considered for further statistical analysis, to avoid measures close to the detection limit.

The sera were stratified according to children’s age in different groups, as described in detail in the “Results” section: group 1 (G1) comprising sera of children <4 years old; group 2 (G2) comprising children >4 years old; group 3 (G3) comprising controls; and group 4 (G4) consisting of patients sera. All the sera together were named the ALL group.

Normalized data were run in the MeV v4.9.0 software. The Wilcoxon-Mann-Whitney test was applied for experiments involving pairwise comparisons between pneumococcal-infected and control groups. The Benjamini-Hochberg (BH) correction was used to control the false discovery rate. Protein targets were considered as immunogenic candidates if antibody levels were significantly different between pneumococcal-infected and control groups with at least a 1.5-fold difference in their SMI values (BH-adjusted p values < 0.05). Hierarchical clustering was used to group sera samples and antigens into subsets, such that those within each cluster (subset) are more closely related to one another than samples assigned to other clusters. Clustering is based on the degree of similarity between the SMI values
for each individual. MeV v4.9.0 was used to perform the clustering analysis. Receiver operating characteristic (ROC) curve analysis was performed with MedCalc v12.7.8. Sensitivity, specificity and Area Under the Curve (AUC) were determined from the resulting ROC analysis. Extension of ROC curve analysis to combinations of antigens was performed as described (28).

For sample size calculation, a power analysis was carried out using an on-line calculator for microarray experiments (http://bioinformatics.mdanderson.org/MicroarraySampleSize/). For 95 protein antigens, a desired fold difference of 1.5 between controls and patients, the minimum sample size is 34 sera assuming 1 false positive (α = 0.0105), or 23 sera assuming five false positives (α = 0.0526).

RESULTS

Selection of Pneumococcal Proteins for Protein Array Design—The overall aim of this work was to construct a protein array to profile the antibody patterns of sera from children with pneumococcal pneumonia and to assess its utility as a diagnostic tool. We based the design of our pneumococcal protein array on antigens experimentally identified on the surface of a collection of pediatric clinical isolates, as surface proteins are those with the highest probabilities to raise an effective humoral immune response. We applied a successful proteomic approach extensively used by our research group to identify in a fast and reliable way the most surface-exposed proteins, consisting of “shaving” live cells with proteases and further LC/MS/MS analysis.

We analyzed 24 clinical isolates collected from children with invasive pneumococcal disease (IPD), which corresponded to 12 different serotypes. As genetic diversity of pneumococcal surface proteins depends on noncapsular genomic background, we genotyped all the isolates by MLST. We found 20 different clonal sequence types (ST), including several major global clones recognized by the PMEN (http://www.sph.emory.edu/PMEN) as shown in supplemental Table S1. Next, we “shaved” the bacterial cells with trypsin and analyzed the resulting peptides. The number of proteins identified, including cytoplasmic proteins derived from experimental limitations of the strategy (29), ranged between 203 and 687, with yields of predicted surface proteins ranging between ca. 20 and 40%, as already described for adult clinical isolates (supplemental Table S3 and SDS2) (22). To include the best potential antigens in the array, we selected those proteins identified experimentally on the surface of the clinical isolates. The first round of selection was made on proteins being present on a high proportion of isolates (≥50%). Table I shows the list of surface proteins found in ≥50% of clinical isolates. It comprised 17 cell wall proteins with the LPXTG anchoring motif, five membrane proteins with one TMD, eight proteins with more than one TMD, four secreted proteins, and three lipoproteins. The cell wall proteins ZmpB and IgA were identified as many different proteins in a diverse number of sequenced strains included in the search database, as they are highly variable in their N-term. However, they were included in the list as the R6 strain-annotated proteins Spr0581 (ZmpB) and Spr1042 (IgA). In addition, we included in the list the protein PblB, which we annotated as of unknown subcellular localization. For this protein, there is a discrepancy about its localization by different prediction algorithms. LocateP predicts a cytosolic location. However, PSORTB assigns it into the “cell wall” category, although not unambiguously. This protein, found in 20 out of the 24 analyzed isolates, has been demonstrated in Streptococcus mitis to be surface-attached (30, 31).

All the proteins for the array, except two, were identified experimentally in variable numbers of clinical isolates. Table II shows the 95 antigens that were selected for production as recombinant polypeptides to be further printed on the array, classified into subcellular localization compartments, according mainly to LocateP, and to their GO annotation (biological process category). In addition to the 37 proteins previously referred to as surpassing the threshold of being identified in ≥50% of the isolates, the rest of proteins were found in a variable number (between 3 and 11) of the analyzed isolates, except two proteins that were not identified but were selected because of their reported immunogenic and/or protective capacity: the cell wall protein SP_1772, and the membrane protein SP_2093. We also selected nine predicted cytoplasmic proteins, identified experimentally, for three reasons: (1) lack of a signal peptide, but with a clearly recognized extracellular localization and function (LytA, Ply); (2) lack of a signal peptide and main intracellular function, but often reported in pneumococcus and many other microbes to be extracellularly located and even displaying immunogenic/protective activity (possible “moonlighting proteins”: Eno, GAPDH); and (3) controls to demonstrate the serodominancy and discriminatory capacity of extracellular antigens.

The 95 selected proteins were studied for production of recombinant polypeptides in E. coli, to obtain purified fragments to be printed on the array. As a general criterion, we selected the regions in which we found a high concentration of identified peptides by our “shaving” approach, normally coinciding with the most exposed domains. Then, we removed the signal peptides and transmembrane domains. Specifically, for lipoproteins we selected the region nearest to the C-term; for cell wall proteins, the region nearest to the N-term (except for ZmpB and IgA, as they have the LPXTG anchoring motif close to the N-term; for them, we selected the C-term region); for membrane proteins, any predicted extracellular domain in which we identified peptides experimentally; and for secreted and predicted cytoplasmic proteins, any region in which we identified peptides experimentally. The fragments cloned for each protein are shown in SDS3. For the vast majority of proteins obtained by Ni²⁺-agarose affinity purification from the E. coli soluble fraction, purity levels were >95%, as estimated by densitometry analysis of SDS-PAGE gels (supplemental Fig. S1).

Protein Microarray Fabrication and Sera Antibody Profile—After selection of the set of pneumococcal proteins as potential immunoreactive antigens, we built a protein microarray to
| Locus     | Description                                      | GO* | 3H | 6H | 13H | 29H | 46H | 49H | 52H | 56H | 65H | 66H | 67H | 72H | 74H | 80H | 83H | 84H | 85H | 86H | 87H | 98H | 99H | 418165 | 599295 |
|-----------|--------------------------------------------------|-----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Cell Wall | Cell wall surface anchor family protein          | CP,SO,BA,MO | 14/24 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| SP_0368   | SP_0368 - Cell wall surface anchor family protein | CP,SO,BA,MO | 14/24 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| SP70585_1587 | G5 domain family | Unknown | 16/24 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| SP70585_1731 | Neuraminidase | MP | 13/24 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| SPCG_0364 | Cell wall surface anchor family protein          | CP,SO,BA | 13/24 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| SPCG_0599 | prtA Serine protease                             | MP,BR | 18/24 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| SPCG_0603 | bgaA β-galactosidase                             | MP | 14/24 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| spr057    | sTH β-N-acetylmuraminidase                       | MP | 13/24 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| spr047    | pulA Alkaline amylopullulanase                    | MP | 14/24 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| spr0328   | Cell wall surface anchor family protein          | CP,SO,BA,MO | 13/24 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| spr0561   | prtA Cell wall-associated serine protease precursor | MP,BR | 20/24 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| spr0565   | bgaA β-galactosidase precursor                  | MP | 18/24 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| spr0581   | zmpB Zinc metalloprotease                        | MP | 24/24 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| spr1042   | lig Immuno globulin A1 protease                  | MP | 23/24 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| spr1403   | Hypothetical protein                             | Unknown | 13/24 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| spr1431   | lycC 1,4-β-N-acetylmuraminidase                  | MP | 14/24 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| spr1536   | nanA Sialidase A precursor (neuraminidase A)     | MP | 17/24 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| spr1652   | Cell wall surface anchor family protein          | MO | 13/24 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| Membrane proteins (1 TMD) | Surface protein pspA precursor | Unknown | 13/24 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| spr0121   | pspA Penicillin-binding protein 2X               | MP,CP,BR,SO,RS | 14/24 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| spr0304   | bbpX Hypothetical protein                        | Unknown | 12/24 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| spr0334   | Hypothetical protein                             | Unknown | 13/24 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| spr0790   | phfD Pneumococcal histidine triad protein D precursor | Unknown | 16/24 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| spr1823   | bbp2a Penicillin-binding protein 2A              | MP,CP,RS | 12/24 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| Membrane proteins (<1 TMD) | Phage-related minor tail protein | Unknown | 16/24 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| SPAP_0062 | SPAP_0062 - Phage-related minor tail protein | Unknown | 16/24 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| SPAP_0063 | Hypothetical protein                             | Unknown | 15/24 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| SPAP_1059 | Hypothetical protein                             | Unknown | 18/24 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| spr0072   | ftsH Cell division protein                       | MP,CP | 12/24 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| spr0043   | comA competence factor Transporting ATP-binding/permease ComA | Unknown | 12/24 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| spr057    | cellB Competence protein CellB                   | CP,SO,RS | 13/24 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| SPN23F05990 | ZmpB Zinc metalloprotease                          | Unknown | 16/24 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| spr0308   | phtE Pneumococcal histidine triad protein E precursor | Unknown | 13/24 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
Table I—continued

| Locus   | Description                 | GO*         | 3H | 6H | 13H | 29H | 46H | 49H | 52H | 56H | 57H | 65H | 66H | 67H | 72H | 74H | 80H | 83H | 84H | 85H | 86H | 87H | 98H | 99H |
|---------|-----------------------------|-------------|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| spr0931 | Hypothetical protein        | Unknown     | 12/24 | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   |
| spr2021 | General stress protein GSP-781 | Unknown | 13/24 | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   |
| Lipoproteins |                       |             |      |    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| spr0327 | ABC transporter substrate-binding protein | Unknown | 14/24 | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   |     |     |
| spr1975 | Zinc ABC transporter zinc-binding protein | CP,SO,RS,EL,BA | 12/24 | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   |     |     |
| spr1382 | Peptide ABC transporter substrate-binding protein | SO,EL | 13/24 | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   |     |     |
| Unknown |                           |             |      |    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Sph_0082 | PblB                       | PblB        | Unknown | 20/24 | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   | x   |

- aGene Ontology annotations of the biological functions of the identified surface proteins. GO annotated code: metabolic process (MP), cellular process (CP), cellular component organization or biogenesis (CC), biological regulation (BR), single-organism process (SO), response to stimulus (RS), developmental process (DP), establishment of localization (EL), biological adhesion (BA), multi-organism process (MO), reproduction (Re), cell killing (CK), localization (Lo).
- b#: Number of clinical isolates (out of 24) in which each of the listed proteins were identified.
- cProtein categories were established according to LocateP subcellular predictions: lipoproteins were those predicted as lipid-anchored proteins; cell wall proteins, as those having an LPXTG motif; secretory proteins, as those possessing an SPI-type signal peptide; membrane proteins with one transmembrane domain (TMD), as those possessing either a C- or an N-terminally anchored transmembrane region; membrane proteins with >1TMD, those predicted as multi-transmembrane proteins; “surface proteins” means the sum of the previous categories; and cytoplasmic proteins, those without any exporting or sorting signal, and predicted as intracellular proteins.
- dThe PblB protein, although annotated as cytosolic by LocateP, was classified in this study as of unknown subcellular localization, as it has been described to be surface-expressed in Streptococcus mitis and because in the Streptococcus pneumoniae sequence, PsortB predicts it as a cell wall protein.
test its viability for detection of antibody profiles as a way to investigate humoral immune response in children with pneumococcal pneumonia, and to evaluate its potential as a serological tool to diagnose pneumococcal pneumonia in individuals or to discriminate among groups. Proteins were immobilized in quintuplicate on the slides, arranged by categories according to their predicted subcellular localization and nature of surface-attaching motifs (cytoplasmic, transmembrane, lipid-anchored, cell-wall or secreted proteins; Fig. 1A).

Next, we used a collection of sera from children to assess the immunologic responses to the printed proteins. The collection comprised 38 sera from patients diagnosed with pneumococcal pneumonia that was mostly (95%) complicated...

| Location | Number | GO (Biological process) | Locus (Internal ID) |
|----------|--------|-------------------------|---------------------|
| Secretory* | 9 | Carbohydrate metabolic process nucleic acid metabolic process Unknown | spr0867 (63), spr1431 (56), spr1903 (21) SP70585_2286 (59), spr0096 (50), spr0908 (45), spr0931 (36), spr1995 (5), spr2021 (2) |
| Lipoprotein | 21 | Carbohydrate metabolic process Cell adhesion Protein metabolic process Transporter activity | spr1527 (57) spr0906 (33), spr1494 (6), spr1975 (23) spr0679 (17), spr0884 (16) spr0101 (30), spr0146 (29), spr0327 (10), spr1251 (28), spr1257 (80), spr1382 (1), spr1707 (3), spr1712 (78), spr1918 (22) spr0174 (79), spr0747 (37), spr0834 (15), spr1061 (55), spr1645 (52), spr1687 (18) |
| Cytosolic | 9 | Carbohydrate metabolic process Oxidation-reduction process Pathogenesis Protein metabolic process Unknown | spr0642 (13), spr1036 (14), spr1754 (12), spr1825 (68), spr1906 (73) spr1430 (70) spr1739 (11) spr0656 (66) spr0868 (69) |
| Cell Wall | 19 | Carbohydrate metabolic process Pathogenesis Protein metabolic process Unknown | spr0057 (32), spr0247 (19), spr0286 (51), spr0440 (31), spr0565 (43), spr1536 (46) spr0328 (26), spr1652 (25) spr0561 (20), spr1771 (39), spr0581 (86), spr1042 (88) |
| TMD | 37 | Carbohydrate metabolic process Establishment of competence for transformation nucleic acid metabolic process | spr1333 (40), spr1531 (58), spr1823 (53), spr1909 (47) spr0044 (54) spr1584 (83) |
| 1TMD | 25 | Carbohydrate metabolic process Establishment of competence for transformation nucleic acid metabolic process Protein metabolic process Regulation of cell shape Transporter activity Unknown | spr0794 (72), spr2045 (8) spr0304 (41), spr0329 (9), spr0716 (75), spr1577 (67) spr0693 (34), spr1243 (65) SP_2093 (76), spr0121 (4), spr0334 (42), spr0337 (74), spr0907 (44), spr1060 (27), spr1098 (64), spr1370 (38) spr1418 (62) |
| >1TMD | 12 | CPS biosynthesis Establishment of competence for transformation Protein metabolic process Transporter activity Unknown | SP_0348 (97) spr0043 (85), spr0857 (87) spr0012 (84) SPJ_1852 (96), SPP_0072 (94), spr0601 (98), spr1221 (99), spr1496 (92) |

*Protein categories were established according to LocateP subcellular predictions: lipoproteins were those predicted as lipid-anchored proteins; cell wall proteins, as those having an LPXTG motif; secretory proteins, as those possessing an SPI-type signal peptide; membrane proteins with one transmembrane domain (TMD), as those possessing either a C- or an N-terminally anchored transmembrane region; membrane proteins with >1TMD, those predicted as multi-transmembrane proteins; “surface proteins” means the sum of the previous categories; and cytoplasmic proteins, those without any exporting or sorting signal, and predicted as intracellular proteins.
with pleural empyema. We used as controls 33 sera from healthy children \( (n = 21) \) or patients \( (n = 12) \) that were admitted to the hospital because of diverse nonpneumococcal diseases and had similar mean age than that of pneumococcal pneumonia cases (SDS1). The array was reproducible in terms of both protein seroreactivity and sera profiling (supplemental Fig. S2). The sera were stratified according to children’s age in two different groups: group 1 (G1) comprising 43 sera of children <4 years old \( (21 \text{ controls, C1, 25.7 ± 10.8 months; and 22 pneumococcal pneumonia patients, P1, 26.5 ± 11.5 months}) \), and group 2 (G2) comprising 28 sera of children >4 years old \( (12 \text{ control infants, C2, 92.0 ± 38.3 months; and 16 pneumococcal pneumonia patients, P2, 79.9 ± 19.9 months}) \). We considered also other two groups to
study the effect of age stratification within either control or pneumococcus-infected children: group 3 (G3) comprising controls (C1 + C2) and group 4 (G4) consisting of patients sera (P1 + P2). When considering all the sera together, we referred to this as the ALL group (n = 71; 33 controls aged 49.8 ± 40.3 months, and 38 pneumococcal-infected patients aged 48.2 ± 30.6 months). Then, we probed the sera on the microarray and measured both primary (IgM) and secondary (IgG) humoral responses (Fig. 1B and 1C), whose overview can be visualized as a heatmap (Fig. 1D and supplemental Fig. S3).

Identification of Antigens Related to Pneumococcal Disease—We used the normalized serological profiles of children patients and controls to search for serodiagnostic antigens that can discriminate between patients and controls, or to show age/stage-specific evolution. First, we looked at the most seroprevalent protein antigens, according to the SMI values of IgG and IgM levels. SDS4 shows the 20 most immunodominant ones for both IgG and IgM responses, stratified in the above mentioned groups. Considering all the 71 analyzed sera (the ALL group), the SMI values of IgG levels ranged between 13,324, corresponding to seroreactivity against PspC in patient sera, and 816, corresponding to NisP in control sera (Fig. 2A). As expected, for IgM levels the SMI values were lower. In the <4 years-old group (G1, 43 sera), many of the most seroprevalent antigens discriminated between patients and controls (considering patients/controls ratios ≥1) for IgG responses, while most of the responses for the G2 group showed no differences (patients/controls ratios ~1). When comparing age groups, we found a general increase in SMI IgG levels in controls (G3), but not in infected children (G4). The control sera group (G3) was then rearranged in subgroups of 12-month intervals and we studied the kinetics of IgG levels against six proteins showing the highest differences between G1 and G2 groups: as shown in Fig. 2, the IgG levels decreased slightly from birth until 1–2 years old, remained relatively constant until 5–6 years old and increased clearly in the oldest subgroup. The same trend was observed for the rest of the anti-proteins IgGs in the same periods of lifespan (supplemental Fig. S4). Regarding the IgM levels, very little or almost no discrimination was obtained in any of the groups. Therefore, only IgG responses were used in subsequent analyses in the search of serodiagnostic protein biomarkers.

Then, we identified the protein candidates related with pneumococcal pneumonia as those showing significantly different IgG levels against such proteins between controls and pneumococcus-infected children groups, with at least 1.5-fold differences in their SMI values (adjusted p values < 0.05, FDR = 0.05) and absolute SMI values above 500, to avoid measures close to detection limits (Table III). In the ALL group, 10 proteins met these requirements, being PblB that showing the highest patients/controls ratio. Interestingly the number of discriminant proteins increased to 24 in the G1 group, i.e. that of children <4 years old. Of these, one third, i.e. eight proteins had the LPXTG cell wall-anchoring motif. These proteins were among those showing the highest patients/controls ratios (e.g. 5.52 for PrtA, 2.80 for PulA, 2.48 for NanA). Again, PblB showed a high ratio (3.93), being the most significant one. Several predicted secreted proteins were also found to differentiate between patients and controls, being PspC the most discriminant one (3.19). Two cytosolic proteins were also found (Eno and SpxB), but also other two proteins without classical signal/exporting peptides, for which there is extensive literature to be exported: LytA and Ply.

When we considered the IgM response, there was not any single protein that significantly discriminated between patients and controls.

Defining Serodiagnostic Antigen Biomarkers—In order to define a reliable set of proteins as serodiagnostic biomarkers, we tried to improve the sensitivity, specificity, and accuracy of such a serodiagnostic test based on our protein array. To this aim, we carried out a receiver operating characteristic (ROC) curves analysis to study the discriminatory power of different sets of proteins between patient and control sera. ROC curves were generated for each of the protein candidates of the different groups (Table III and SDS4), and the area under the ROC curves (AUC) for each individual antigen is listed in supplemental Table S4 in decreasing order. In the ALL group, the three most discriminatory proteins were PrtA, PblB, and PulA, with AUC values of 0.802, 0.768, and 0.762, respectively. The same three proteins showed also the highest dis-
Then, we extended the analysis to sets of antigens using combinations of ROC curve analysis (Fig. 3). As inputs to the classifier, we used the highest-ranking AUC antigens in combinations of 2, 3, 4, ..., n proteins. In the ALL group, there was not a significant improvement when combining several antigens: considering PrtA and PblB together (the two with the
highest AUC values), there was a slight increase in sensitivity, but a decrease in specificity and accuracy (Table IV). However, a clear improvement was obtained in the G1 group: the best parameters were obtained with the combination of the first three antigens (PblB, PulA and PrtA), as the classifier predicted a 100% sensitivity, 95.8% specificity, and 97.9% accuracy to discriminate pneumococcal pneumonia patients from the control group, being those values equal or higher than considering only PblB or PblB/H11001 PulA.

Validation of Array Data and Immunostrip Probing—The serodiagnostic capacity of the protein array was validated by a colorimetric ELISA assay using an independent set of 24 sera from 4 years-old (12 control infants, C, 31.0 ± 12.4 months; and 12 pneumococcal pneumonia patients, P, 27.6 ± 12.3 months, SDS1) (Fig. 4). With this test, we confirmed the results of the protein arrays, i.e. the combination of more than one antigen resulted in a clear discrimination between infected patients and controls in children <4 years-old according to the ROC curves (Fig. 4A). The AUC analysis showed higher values of sensitivity, specificity and accuracy (100%, 100 and 100% respectively) for the combination than when the antigens were considered individually (Fig. 4B and supplemental Table S5).

Finally, in order to check the application of the discriminatory power of sets of proteins from protein array projects in affordable, ease-of-use serological tests, we developed an immunostrip assay using the three proteins with the highest discriminatory capacity (i.e. PblB, PulA and PrtA) and a mixture of them. The immunostrips were probed with the 24 sera of the independent validation set. As shown in Fig. 5, there was in general a higher IgG response in sera from infected patients compared with controls, both against the individual proteins and against the 1:1:1 mixture, whereas the reaction against positive controls was the same. The spot intensities were analyzed, and the AUC analysis showed the highest values of sensitivity, specificity and accuracy (100% for the three parameters) for the test to discriminate between patients and controls.

DISCUSSION
Protein microarrays are an ideal means to explore in a high-throughput way the humoral responses to many pathological conditions, including infectious diseases (32). In fact, there is extensive literature reporting the use of these platforms to measure the antibody profiles in collections of sera from patients affected by bacterial (19, 33–39), fungal (40),
TABLE IV

Area under the Receiver Operating Characteristics (ROC) curve analysis (AUC) of combinations of protein biomarkers selected from protein arrays. As inputs to the classifier, we used the highest ranking area under the curve (AUC) antigens which were combined up to ten antigens per group.

| Sample | Proteins | Area under the ROC curve (AUC) | Standard error | 95% Confidence interval | z statistic | Significance level P (Area = 0.5) | Sensitivity | Specificity | Accuracy | PPV\(^a\) | NPV\(^a\) |
|--------|----------|-------------------------------|----------------|--------------------------|-------------|-----------------------------------|-------------|-------------|-----------|-----------|-----------|
| G-ALL  | PrtA     | 0.802256                      | 0.0569         | 0.692500 to 0.886310     | 5.312       | <0.0001                           | 92.1        | 71.4        | 81.8      | 76.3      | 90.0      |
|        | PblB     | 0.813534                      | 0.0543         | 0.705219 to 0.895141     | 5.770       | <0.0001                           | 94.7        | 68.6        | 81.6      | 75.1      | 92.9      |
|        | PulA     | 0.806015                      | 0.0563         | 0.696727 to 0.889266     | 5.436       | <0.0001                           | 84.2        | 77.1        | 80.6      | 78.7      | 83.0      |
|        | GaiU     | 0.806015                      | 0.0563         | 0.696727 to 0.889266     | 5.440       | <0.0001                           | 84.2        | 80.0        | 82.1      | 80.8      | 83.5      |
|        | NanA     | 0.806015                      | 0.0552         | 0.696727 to 0.889266     | 5.440       | <0.0001                           | 81.6        | 80.0        | 80.8      | 80.3      | 81.3      |
|        | PulA     | 0.806015                      | 0.0550         | 0.695035 to 0.888085     | 5.377       | <0.0001                           | 81.6        | 77.1        | 79.3      | 78.1      | 80.7      |
|        | Sp1875   | 0.813534                      | 0.0538         | 0.705219 to 0.895141     | 5.829       | <0.0001                           | 86.8        | 74.3        | 80.5      | 77.2      | 85.0      |
|        | 7Prot    | 0.810526                      | 0.0533         | 0.701816 to 0.892797     | 5.829       | <0.0001                           | 84.2        | 74.3        | 79.2      | 76.6      | 82.5      |
|        | 8Prot    | 0.790226                      | 0.0565         | 0.679050 to 0.876765     | 5.140       | <0.0001                           | 92.1        | 67.5        | 78.9      | 72.9      | 89.3      |
|        | 9Prot    | 0.789474                      | 0.0564         | 0.678213 to 0.876164     | 5.134       | <0.0001                           | 92.1        | 65.7        | 79.2      | 72.9      | 89.3      |
|        | 10Prot   | 0.789474                      | 0.0564         | 0.678213 to 0.876164     | 5.134       | <0.0001                           | 92.1        | 65.7        | 79.2      | 72.9      | 89.3      |
| G1     | PblB_a   | 0.981061                      | 0.0195         | 0.889269 to 0.999700     | 24.661      | <0.0001                           | 100.0       | 95.5        | 97.7      | 70.6      | 100.0     |
|        | PblB_a+PulA | 0.990530                | 0.0102         | 0.905324 to 0.999997     | 47.987      | <0.0001                           | 95.5        | 100.0       | 97.7      | 100.0     | 95.7      |
|        | PrtA     | 0.996212                      | 0.0046         | 0.915646 to 1.000000     | 107.82      | <0.0001                           | 100.0       | 95.8        | 97.9      | 96.0      | 100.0     |
|        | LytC     | 0.994318                      | 0.0064         | 0.912133 to 1.000000     | 76.258      | <0.0001                           | 100.0       | 95.8        | 97.9      | 96.0      | 100.0     |
|        | NisP     | 0.992424                      | 0.0074         | 0.908695 to 1.000000     | 66.551      | <0.0001                           | 100.0       | 91.7        | 95.8      | 92.3      | 100.0     |
|        | GalU     | 0.962121                      | 0.0258         | 0.859845 to 0.996338     | 17.941      | <0.0001                           | 100.0       | 87.5        | 93.7      | 88.9      | 100.0     |
|        | NanA     | 0.958333                      | 0.0265         | 0.854258 to 0.995251     | 17.278      | <0.0001                           | 95.5        | 87.5        | 91.5      | 88.4      | 95.1      |
|        | Sp70585_2286 | 0.956439              | 0.0265         | 0.851493 to 0.994664     | 17.198      | <0.0001                           | 95.5        | 87.5        | 91.5      | 88.4      | 95.1      |
|        | PspC     | 0.933712                      | 0.0352         | 0.819577 to 0.985688     | 12.307      | <0.0001                           | 90.9        | 87.5        | 89.2      | 75.7      | 88.6      |
|        | Spr1652  | 0.952652                      | 0.0279         | 0.846020 to 0.993414     | 16.226      | <0.0001                           | 90.9        | 91.7        | 91.3      | 91.6      | 91.0      |

\(^a\) PPV and NPV = Positive and Negative Predictive Values, respectively.
and parasitic infections (41–45). They can be used as a diagnostic tool, for validation and/or discovery of vaccine candidates, and for epidemiological surveillance programs (46). Our work reports for the first time the development of a protein array of pneumococcal antigens, based on their experimental identification on the surface of relevant clinical isolates, and its use as a diagnostic tool to discriminate between pneumococcal-infected children and controls. Surface proteins play

![Graphs and Data](image-url)

**FIG. 4.** Validation of biomarker discovery from protein array results by ELISA using an independent validation set of sera. A, For each protein (PrtA, PulA, PblB) and the 1:1:1 combination of them, it is represented the receiver operating characteristics (ROC) curve and interactive dot diagram to show the diagnostic capacity of the assay (C: controls; P: patients). B, Statistical parameters of the Area Under the ROC Curve (AUC) of the three single proteins and their combination in the ELISA test.

**FIG. 5.** Immunostrip test using the three best protein biomarkers discovered with the protein array. On the upper panel, the dot blot assay using the three individual antigens (PrtA, PulA, PblB) and their 1:1:1 combination is shown. As negative controls, the irrelevant yeast protein Lys9 (NC-1) and commercial trypsin (NC-2) were used. As positive controls, pneumococcal serotype 8 strain total protein extract (PC-1) and commercial anti-human IgG produced in goat (PC-2) were used. The 24 sera of the independent validation set (12 controls and 12 pneumococcal-infected children) were used at a 1:200 dilution to probe the nitrocellulose membranes. Down, the receiver operating characteristics (ROC) curve and interactive dot diagram of the combination of antigens are represented to show the diagnostic capacity of the assay (C: controls; P: patients).
many important roles in the interaction between cells and their environment (47). They have many functions and are targets for drugs and vaccine development, as well as candidates for diagnostic biomarkers, as they have the highest probabilities to be recognized by the immune system (14, 15, 34). Proteomics offers excellent approaches and strategies to identify antigenic surface proteins in microorganisms, especially those most exposed and abundant. A successful approach to identify in a fast and reliable way the “surfome,” i.e. the set of surface proteins, consists of “shaving” live cells with proteases, following LC/MS/MS analysis (20, 48, 49). When applied to a collection of strains or clinical isolates of a given species, it provides a very interesting overview of the presence, frequency, and abundance of the identified proteins throughout the studied population, which allows choosing good candidates for vaccine or diagnostic purposes.

In the present study, we selected the proteins for the protein array design using the “shaving” approach, after analysis of 24 clinical isolates from pneumococcus-infected children. Some of the isolates belonged to the most prevalent and virulent serotypes circulating in Spain in the last years (1, 3, 7F, 14, and 19A) (50). The identifications included numerous cytoplasmic proteins, which may be because of unavoidable lysis (21, 29). However, both the numbers of total identified proteins and percentages of predicted surface ones were very similar to those described for adult clinical isolates (22). Moreover, many of the proteins identified in ≥50% of pediatric clinical isolates in this study were also found in ≥50% of adult clinical isolates. Although we based this protein array platform for being used as a clinical tool for children, it might be also used for adults.

Almost all the proteins selected for the array design were identified experimentally. Although Table I reflects only those found in ≥50% of clinical isolates, many others were identified in a high number (i.e. 6–11) of isolates. Many of these were also identified in a high proportion of adult clinical isolates (22). Our results show the enrichment of LPXTG-cell wall proteins when compared with their predicted figures from the genome, as already demonstrated in previous works for this and other pathogens (20, 21, 48, 49, 51–54). We decided to include also some cytoplasmic proteins, as for some of them their surface localization and immunogenicity/protective activity, are described, like Eno (55, 56) or GAPDH (57). In addition, LytA and Ply are predicted as cytoplasmic because they lack a signal peptide, but their final destination is extracellular (58). We only included three proteins not identified experimentally, either from in silico selection or being described in literature as immunogenic/protective: SP_1772 (PsrP), a cell-wall protein with proven immunogenic/protective capacity (59); SP_2093, a transmembrane protein without any described immunogenic/protective activity that was selected to test its seroreactivity in absence of experimental evidence; and HipO, a cytoplasmic protein that was selected as an internal negative control.

In our view, the most interesting results derived from the proteomics approach is the discovery of the broad surface expression of PblB (20 out of the 24 isolates). This is an unusual surface protein, as it neither has a signal peptide/cell wall sorting motif, nor is it strongly similar to any known bacterial adhesin. Still, it resembles a phage-encoded tail fiber protein (30). In S. mitis, it has been demonstrated that this protein is surface-exposed, and acts as a platelet-binding adhesin (30, 31). In pneumococcus, there is no evidence on the function of pblB-like genes encoded by bacteriophages. It has been described that up to 76% of pneumococcal clinical isolates contain prophages (60, 61). This is the first work that shows experimentally the surface location of this protein in pneumococcus. Moreover, the use of a compilation database containing all the pneumococcal protein sequences available so far has made possible to identify the protein in our study. The use of only reference strains (R6, TIGR4) would have resulted in the missing of this information, as these strains lack PblB.

We selected for recombinant fragment production the regions of the proteins experimentally found in the “shaving” approach. If possible, we selected highly conserved domains across the sequenced proteins available in databases. This was not always possible, as some proteins are highly variable, e.g. ZmpB and IgA. These two proteins were annotated respectively in our study as single entries, but different sequences with <95% similarity are available, with high variability in the C-term. However, we produced the C-term of both proteins from the R6 strain genome, as that region is the most exposed for both cases. About PblB, different forms are also available in the databases. As this is a large protein, we produced two fragments for this: one containing the C-term region, which is more conserved (PblB_a), and the other containing a region close to the M-term, being more variable (PblB_b).

There is a renewed interest in serological diagnosis of pneumococcal infection using surface proteins-based tests. Some epidemiological studies have been conducted, based on reduced sets of proteins showing unclear patterns of antibody response when applied to heterogeneous cohorts of patients (6, 13, 62). Serological diagnosis of pneumococcal disease is challenging because it should discriminate between infected patients and nonsymptomatic carriers (63), and because of difficulties in obtaining acute and convalescent sera from pediatric patients (1). Thus, to investigate the humoral responses to IPD in pediatric empyema cases, we developed an array to test 95 antigens. We used sera from patients with the most severe spectrum of pneumococcal pneumonia: bacteremic pneumonia and/or pleural empyema. This latter severe complication of pneumonia has increased worldwide over the last decade and is generally difficult to diagnose etiologically by blood culture because of high rates of antibiotic treatment prior to sample collection (64, 69).
Among the most IgG-reacting seroprevalent antigens, we found many proteins from all the subcellular compartments that have been already described to be immunogenic: the secreted proteins PspC, PcsB, and LytC, lipoproteins, transmembrane proteins, cell wall proteins like PrtA, classical cytosolic proteins like Eno, and cytosolic proteins described to be extracellular like LytA and Ply. The C-term fragment of PblB (PblB_a) was also found to be highly immunoreactive. The general increase in IgG levels as children get older may be explained by colonization, as the differences are observed in the controls (G1 group) but not in the pneumococcal-infected patients (G4 group). There was also an IgM response, but it was in general terms lower than that of IgG, and did not discriminate between patients and controls. Therefore, we based our study on the IgG responses.

As shown, group management by age stratification was useful to highlight discriminant responses. The clearest ones were those of the G1 group, i.e. children <4 years old. For that group, we obtained the highest number of discriminant antigens and with the highest patient/control ratios of IgG levels. There was a clear enrichment in LPXTG-cell wall proteins: out of the 24 discriminant proteins in the G1 group, eight belonged to this category. This is in agreement with the knowledge that in Gram-positives, the cell wall proteins are generally highly exposed, abundant, highly immunogenic, and even protective. Regarding cell wall proteins, ZmpB and IgA were hardly immunoreactive, which may be because of their variability, therefore hindering their recognition by nonhomologous counterparts. The classical cytosolic proteins Eno and SpxB were also discriminant, as well as the nonclassical extracellular proteins LytA and Ply. Among predicted secreted proteins, the one with the highest discriminant capacity was PspC. We also found that LytC clearly discriminated. We have previously described that this protein is carried by extracellular vesicles released by pneumococcus, and that LytC is highly immunoreactive and even immunogenic (66). However, one of the most seroprevalent antigens, PcsB, did not discriminate at all. Interestingly, we found a highly discriminant capacity for the nonwell-characterized pneumococcal PblB, which was seroprevalent but not among the highest ones.

This study shows that a combination of the highly discriminant proteins discovered in a multiantigen platform for serodiagnostic test can be useful to discriminate between infected and control people. In our case, the combination of PblB, PuA, and PrtA showed such a discriminatory power with very high accuracy, sensitivity, and specificity, as demonstrated by the AUC analysis. The results were also validated by ELISA, and the immunostrip assay also showed that the results may be translated into easy-to-use, affordable tests.

This study has several limitations. It can be argued that in our case the work considered a relatively limited number of sera, but as stated above, legal, ethical, logistic, and technical issues often limit sera sampling in children patients. However, the sample size was above the minimum, according to the power analysis. Although sampling of pneumococcal pneumonia patients was performed generally within the first few days of hospital admission, once microbiological diagnosis of pneumococcal disease was established, it was not performed as fixed time points since disease onset because of variable duration of symptoms prior to hospital admission. Nevertheless, our protein array is an excellent launching platform to be applied in different programs and sera populations.

CONCLUSIONS

We have developed a protein array for its use in the study of humoral responses to pneumococcal infection, based on the selection of experimentally identified proteins. The platform has proven its capacity to measure antibody levels in children's sera and to discriminate between pneumococcal- and nonpneumococcal infected infants. PblB has been shown experimentally for the first time to be expressed on the surface of a large collection of pneumococcal isolates, being immunoreactive and even discriminant. This platform is an excellent means to be used as a diagnostic tool and can be adapted to different population studies. Moreover, it may be also useful in programs of epidemiological surveillance and even for vaccine candidate discovery.
Pneumococcal Serodiagnostic Antigens by Protein Array

PLoS Med. 7, e1000348

3. Pittet, L. F., and Posfay-Barbe, K. M. (2012) Pneumococcal vaccines for children: a global public health priority. Clin. Microbiol. Infect. 18, 25–36

4. Welte, T., Torres, A., and Nathwani, D. (2012) Clinical and economic burden of community-acquired pneumonia among adults in Europe. Thorax 67, 71–79

5. Weycker, D., Strutton, D., Edelsberg, J., Sato, R., and Jackson, L. A. (2010) Failure to elicit seroresponses to pneumococcal surface proteins (pneumococcal histidine triad D) pneumococcal choline-binding protein A, and serine proteinase precursor A) in children with pneumococcal bacteremia. J. Infect. Dis. 192, 906–914

6. Hagerman, A., Posfay-Barbe, K. M., Gram, S., Govindan, V., Dagan, R., and Siegert, C. A. (2012) Identification of seroreactive proteins of Leptospira interrogans serovar copenhagenii using a high-density protein microarray approach. PLoS Negl. Trop. Dis. 7, e2499

7. Madhi, S. A., Kukwanda, L., Cutland, C., and Klugman, K. P. (2005) The impact of a 9-valent pneumococcal conjugate vaccine on the public health burden of pneumonia in HIV-infected and -uninfected children. J. Infect. Dis. 190, 1511–1518

8. Turner, P., Hinds, J., Turner, C., Jankhot, A., Gould, K., Bentley, S. D., Scott, G., Jones, N., and Ananth, V. J. (2010) A multilocus sequence typing scheme for Streptococcus pneumoniae: identification of clones associated with serious invasive disease. J. Clin. Microbiol. 48, 526–532

9. Zhou, M., Boekhorst, J., Francke, C., and Siezen, R. J. (2008) LocateP: genome-scale subcellular-location predictor for bacterial proteins. BMC Bioinformatics 9, 173

10. Bensing, B. A., Cardona, F., Orefici, G., Grandi, G., and Bensi, G. (2009) Protein profiling: shaving live organisms for a fast proteomic identification of surface proteins. J. Proteomics 72, 691–702

11. Brandt, M. G. (2001) Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics 17, S96–S104

12. DeLong, E. R., DeLong, D. M., and Clarke-Pearson, D. L. (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44, 837–845

13. Carvalho Mda, G., Tondella, M. L., McCaustland, K., Weidlich, L., McGee, P. W., Ma, J., Mitchell, T. J., and Sullam, P. M. (2002) Variation and to the quantification of differential expression. Bioinformatics 18, 756–762

14. Davis, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 51, 63–69

15. Durbin, R., and Hardin, J. (2002) Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics 17, S96–S104

16. Enright, M. C., and Spratt, B. G. (1998) A multilocus sequence typing scheme for Streptococcus pneumoniae: identification of clones associated with serious invasive disease. Microbiology 144, 3049–3060

17. Friedrich, J. M., and Edes, E. W. (2007) Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics 17, S96–S104

18. Brazil, M. L., and Weidlich, L., McGee, P. W., Ma, J., Mitchell, T. J., and Sullam, P. M. (2002) Mechanism of cell surface expression of the Streptococcus mitis platelet binding proteins PibA and PiB. Mol. Microbiol. 46, 844–857

19. Bensing, B. A., Rubens, C. E., and Sullam, P. M. (2001) Genetic loci of Streptococcus mitis that mediate binding to human platelets. Infect. Immun. 69, 1379–1386

20. Bensing, B. A., Cardona, F., Orefici, G., Grandi, G., and Bensi, G. (2009) Protein profiling: shaving live organisms for a fast proteomic identification of surface proteins. J. Proteomics 72, 691–702

21. Bensing, B. A., Cardona, F., Orefici, G., Grandi, G., and Bensi, G. (2009) Protein profiling: shaving live organisms for a fast proteomic identification of surface proteins. J. Proteomics 72, 691–702

22. Bensing, B. A., Cardona, F., Orefici, G., Grandi, G., and Bensi, G. (2009) Protein profiling: shaving live organisms for a fast proteomic identification of surface proteins. J. Proteomics 72, 691–702

23. Bombaci, M., Grifantini, R., Moro, M., Reguzzoni, F., Petracca, R., Meoni, E., Balioni, S., Zingaretti, C., Falugi, F., Manetti, A. G., Margariti, I., Musser, J. M., Cardona, F., Orefici, G., Grandi, G., and Bensi, G. (2009) Protein array profiling of tc patient sera reveals a broad range and enhanced immune response against Group A Streptococcus antigens. PLoS ONE 4, e5332

24. Felgner, P. L., Kayala, M. A., Vigil, A., Burke, C., Nakajima-Sasaki, S., Pablo, J., Molina, D. M., Hirst, S., Chew, J. S., Wang, D., Tan, G., Duffield, M., Yang, R., Neel, J., Chantarattana, N., Bancroft, G. J., Pittb, R., and Lertmemongkolchai, G. (2011) Immunity to pneumococcal surface proteins in children with community-acquired pneumonia: a distinct pattern of responses to pneumococcal choline-binding protein A. Clin. Microbiol. Infect. 17, 1237–1238

25. Felgner, P. L., Kayala, M. A., Vigil, A., Burke, C., Nakajima-Sasaki, S., Pablo, J., Molina, D. M., Hirst, S., Chew, J. S., Wang, D., Tan, G., Duffield, M., Yang, R., Neel, J., Chantarattana, N., Bancroft, G. J., Lertmemongkolchai, G., Davies, D. H., Baldi, P., Peacock, S., and Pittb, R. W. (2009) A Burkholderia pseudomallei protein microarray reveals serodiagnostic and cross-reactive antigens. Proc. Natl. Acad. Sci. U.S.A. 106, 13499–13504

26. Lee, S. J., Yang, R., Li, S., Ji, X., Lee, S. I., Park, K., and Jo, S. (2011) Identification of a common immune signature in murine and human systemic Salmonellosis. Proc. Natl. Acad. Sci. U.S.A. 109, 4998–5003
36. Rodríguez-Ortega, M. J., Luque, I., Tarradas, C., and Bárçena, J. A. (2008) Molecular & Cellular Proteomics 14.10

37. Margarit, I., Bonacci, S., Pietrocolo, G., Rindi, S., Ghezzo, C., Bombaci, M., Nardi-Dei, V., Grifantini, R., Speziale, P., and Grandi, G. (2009) Capturing host-pathogen interactions by protein microarrays: identification of novel streptococcal proteins binding to human fibronectin, fibrinogen, and C4BP. *FASEB J.* 23, 3100–3112

38. Vigil, A., Ortega, R., Nakajima-Sasaki, R., Pablo, J., Molina, D. M., Chao, C. C., Chen, W. H., Ching, W. M., and Felgner, P. L. (2010) Genome-wide analysis of humoral immune response to *Coxiella burnetii* infection by protein microarray. *Proteomics* 10, 2259–2269

39. Liang, L., Doskaya, M., Juarez, S., Caner, A., Jasinskas, A., Tan, X., Liang, W. J., Liang, X., Gilman, R. H., Davies, D. H., Tsolis, R. M., Vinetz, J. M., and Felgner, P. L. (2011) Systems biology approach predicts antibody signature associated with Brucella melitensis infection in humans. *J. Proteome Res.* 10, 4813–4824

40. Mochon, A. B., Jin, Y., Ye, J., Kayala, M. A., Wingard, J. R., Clancy, C. J., Nguyen, M. H., Felgner, P., Baldi, P., and Liu, H. (2010) Serological profiling of *Streptococcus* species proteins by protein microarray reveals permanent host-pathogen interplay and stage-specific responses during candidemia. *PLoS Pathog.* 6, e1000827

41. Barry, A. E., Trieu, A., Fowkes, F. J., Pablo, J., Kalantari-Dehaghi, M., Jasinskas, A., Tan, X., Kayala, M. A., Tavul, L., Siba, P. M., Day, K. P., Baldi, P., Felgner, P. L., and Doolan, D. L. (2011) The stability and complexity of antibody responses to the major surface antigen of *Plasmodium falciparum* is associated with age in a malaria endemic area. *Mol. Cell. Proteomics* 10, M111.003262

42. Baum, E., Badu, K., Molina, D. M., Liang, X., Felgner, P. L., and Yan, G. (2012) Protein microarray analysis of antibody responses to Plasmodium falciparum in western Kenyan highland sites with differing transmission levels. *PLoS ONE* 8, e82246

43. Doolan, D. L., Mu, Y., Unal, B., Sundaresh, S., Hirst, S., Valdez, C., Randall, A., Molina, D., Liang, X., Freilich, D. A., Olo, J. A., Blair, P. L., Aguilar, J. J., Baldi, P.,获批, H., and Felgner, P. L. (2008) Profiling humoral immune responses to *P. falciparum* infection with protein microarrays. *Proteomics* 8, 4860–4869

44. Gaze, S., Driguez, P., Pearson, M. S., Mendes, T., Doolan, D. L., Trieu, A., McManus, D. P., Gobert, G. N., Periago, M. V., Correa-Oliveira, R., Cardoso, F. C., Oliveira, G., Nakajima, R., Jasinskas, A., Hung, C., Liang, L., Pablo, J., Bethony, J. M., Felgner, P. L., and Loukas, A. (2014) An immunomics approach to infectious disease discovery: antibody signatures of naturally resistant and chronically infected individuals from endemic areas. *PLoS Pathog.* 10, e1004033

45. Gaudet, E., Pokorny, M., Croucher, N. J., Hiller, N. L., Hu, F. Z., Ehrlich, G. D., Bentley, S. D., Garcia, E., and Mitchell, T. J. (2009) Comparative genomic analysis of ten *Streptococcus pneumoniae* adhesin, block adhesion and protect mice against pneumococcal acute otitis media. *J. Infect.* 16, 375–383

46. Carrillo, M. L., Alzate-Arango, P., Bergmann, S., Rohde, M., Chhatwal, G. S., and Hammerschmidt, S. (2010) Alpha-Enolase of *Streptococcus pneumoniae* is a plasminogen-binding protein displayed on the bacterial cell surface. *Mol. Microbiol.* 40, 1273–1287

47. Bing, E., Feldman, G., Portnoi, M., Daga, R., Overweg, K., Mulhall, F., Chalifa-Caspi, V., Wells, J., and Mizrahi-Nebenzahl, Y. (2004) Glycolytic enzymes associated with the cell surface of *Streptococcus pneumoniae* are antigenic in humans and elicit protective immune responses in the mouse. *Clin. Exp. Immunol.* 138, 290–298

48. Mitchell, A. M., and Mitchell, T. J. (2013) *Streptococcus pneumoniae*: virulence factors and variation. *Clin. Microbiol. Infect.* 16, 411–418

49. Ylostalo, K., Chang, Y., Yamada, T., McManus, D. P., Gobert, G. N., Periago, M. V., Correa Oliveira, R., Cardoso, F. C., Oliveira, G., Nakajima, R., Jasinskas, A., Hung, C., Liang, L., Pablo, J., Bethony, J. M., Felgner, P. L., and Loukas, A. (2014) An immunomics approach to infectious disease discovery: antibody signatures of naturally resistant and chronically infected individuals from endemic areas. *PLoS Pathog.* 10, e1004033

50. Ardanuy, C., Marimón, J. M., Calatayud, L., Giménez, M., Alonso, M., Grau, I., Pallarés, R., Pérez-Trallero, E., and Liñha, J. (2012) Mol. Cell. Proteomics 11, 2365–2369

51. 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

52. Adrian, P. V., Bogaert, D., Omins, M., Rapola, S., Lahdenmäki, M., Klipi, T., de Graaff, K., Káthy, A., and Hemans, P. W. (2004) Development of antibodies against pneumococcal proteins alpha-enolase, immunoglobulin A1 protease, streptococcal lipoprotein rotase A, and putative protease maturation protein A in relation to pneumococcal carriage and *Otis* Media. *Vaccine* 22, 2737–2742

53. Bergmann, S., Rohde, M., Chhatwal, G. S., and Hammerschmidt, S. (2001) *Streptococcus pneumoniae* adhesin, block adhesion and protect mice against pneumococcal acute otitis media. *J. Infect.* 16, 375–383

54. Ramírez, M., Severina, E., and Tomasz, A. (1999) A high incidence of prophage carriage among natural isolates of *Streptococcus pneumoniae* in the United States. *J. Bacteriol.* 181, 3619–3625

55. Romero, P., Croucher, N. J., Hiller, N. L., Hu, F. Z., Ehrlich, G. D., Bentley, S. D., Garcia, E., and Mitchell, T. J. (2009) Comparative genomic analysis of ten *Streptococcus pneumoniae* temperate bacteriophages. *J. Bacteriol.* 191, 4854–4862

56. Simell, B., Melin, M., Lahdenmäki, M., Briles, D. E., Hollingshead, S. K., Klipi, T. M., and Káthy, A. (2007) Antibodies to *Streptococcus pneumoniae* surface protein A families 1 and 2 in serum and saliva of children and the risk of pneumococcal acute otitis media. *J. Infect.* 196, 1528–1536

57. Turner, P., Turner, C., Green, N., Ashton, L., Lwe, E., Jankhot, A., Day, N. P., White, N. J., Nosten, F., and Goldblatt, D. (2013) Serum antibody responses to pneumococcal colonization in the first 2 years of life: results from an SE Asian longitudinal cohort study. *Clin. Microbiol. Infect.* 19, E551–E558

58. Buiging, C. L., Korgenski, K., Daly, J., Ampofo, K., Pavia, A., and Mason, E. O. (2006) Impact of the pneumococcal conjugate vaccine on pneumococcal parapharyngeal empyema. *Pediatr. Infect. Dis. J.* 25, 250–254

59. Obando, I., Muñoz-Almagro, C., Arroyo, L. A., Tarrago, D., Sanchez-Tatay, D., Moreno-Perez, D., Perez-Deza, D., Hernandez-Bou, S., Garcia-Garcia, J. J., Hausdorff, W., Pavia, A., and Mason, E. O. (2006) Impact of the pneumococcal conjugate vaccine on pneumococcal parapharyngeal empyema. *Pediatr. Infect. Dis. J.* 25, 250–254
icles produced by Streptococcus pneumoniae. *J. Proteomics* **106**, 46–60

67. Vizcaíno, J. A., Côté, R. G., Csordas, A., Dianes, J. A., Fabregat, A., Foster, J. M., Griss, J., Alpi, E., Birim, M., Contell, J., O’Kelly, G., Schoenegger, A., Ovelleiro, D., Perez-Riverol, Y., Reisinger, F., Rios, D., Wang, R., and Hermjakob, H. (2013) The PRoteomics IDENTifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* **41**, D1063–D1069

68. Vizcaíno, J. A., Deutsch, E. W., Wang, R., Csordas, A., Reisinger, F., Rios, D., Dianes, J. A., Sun, Z., Farrah, T., Bandeira, N., Binz, P. A., Xenarios, I., Eisenacher, M., Mayer, G., Gatto, L., Campos, A., Chalkley, R. J., Kraus, H. J., Albar, J. P., Martinez-Bartolomé, S., Apweiler, R., Omenn, G. S., Martens, L., Jones, A. R., and Hermjakob, H. (2014) ProteomeX-change provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* **32**, 223–226