LUND UNIVERSITY

The unique structure of Haemophilus influenzae protein E reveals multiple binding sites for host factors.

Singh, Birendra; Tamim, Al-Jubair; Thunnissen, Marjolein; Mörgelin, Matthias; Riesbeck, Kristian

Published in:
Infection and Immunity

DOI:
10.1128/IAI.01111-12

2013

Link to publication

Citation for published version (APA):
Singh, B., Tamim, A.-J., Thunnissen, M., Mörgelin, M., & Riesbeck, K. (2013). The unique structure of Haemophilus influenzae protein E reveals multiple binding sites for host factors. *Infection and Immunity*, *81*(3), 801-814. https://doi.org/10.1128/IAI.01111-12

Total number of authors:
5

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

The unique structure of *Haemophilus influenzae* protein E reveals multiple binding sites for host factors

Birendra Singh,[a] Tamim Al-Jubair,[a] Marjolein M. Thunnissen,[b] Matthias Mörgelin,[c] and Kristian Riesbeck [a*]

Medical Microbiology, Department of Laboratory Medicine Malmö, Lund University, Skåne University Hospital, SE-205 02 Malmö, Sweden[a]; Centre of Molecular Protein Science, Lund University, Getingevägen 60, SE-221 00 Lund, Sweden[b]; Section of Clinical and Experimental Infectious Medicine, Department of Clinical Sciences, Lund University, SE-221 84 Lund, Sweden[c].

*Address correspondence and reprint requests to Dr. Kristian Riesbeck, Medical Microbiology, Department of Laboratory Medicine Malmö, Lund University, Skåne University Hospital, SE-205 02 Malmö, Sweden. E-mail: kristian.riesbeck@med.lu.se

Running title: Structure of the adhesin protein E

Keywords: Crystal structure, *Haemophilus influenzae*, laminin, plasminogen, Protein E, vitronectin

Abbreviations used in this article: COPD, chronic obstructive pulmonary disease; ECM, extracellular matrix; Ln, laminin; NTHi, non-typable *Haemophilus influenzae*; PE, protein E; Vn, vitronectin

Abstract

Haemophilus influenzae protein E (PE) is a multifunctional adhesin, involved in direct interactions with lung epithelial cells and host proteins, including plasminogen and the extracellular matrix proteins vitronectin and laminin. We recently crystallized PE and successfully collected X-ray diffraction data to 1.8 Å. Here we solved the structure of a recombinant version of PE and analyzed different functional regions. It is a dimer in solution and in the asymmetric unit of the crystals. The dimer has a structure that resembles a flattened β-barrel. It is however not a true β-barrel as there are differences in both the hydrogen bonding pattern and the shape. Each monomer consisted of a 6-stranded antiparallel β-sheet with a rigid α-helix at the C-terminal tethered to the concave side of the sheet by a disulfide bridge. The laminin/plasminogen binding region (residues 41-68) is exposed, while the vitronectin binding region (residues 84-108) is partially accessible in the dimer. The dimerized PE explains the simultaneous interaction with laminin and vitronectin. In addition, we found this unique adhesin being present in many bacterial genera of the family *Pasteurallaceae* and also orthologues in other unrelated species (*Enterobacter cloacae* and *Listeria monocytogenes*). Peptides corresponding to the surface-exposed regions PE24-37, PE74-89, and PE134-156 were immunogenic in the mouse. Importantly, these peptide-based antibodies also recognised PE at the bacterial surface. Taken together, our detailed structure of PE explains how this important virulence factor of *H. influenzae* simultaneously interacts with host vitronectin, laminin or plasminogen promoting bacterial pathogenesis.

Introduction

Haemophilus influenzae is an important Gram-negative respiratory pathogen that causes, for example, acute otitis media in children, and exacerbations in patients with chronic obstructive pulmonary disease (COPD), but also invasive disease such as meningitis and sepsis (1). Encapsulated *H. influenzae* is categorized into six different serotypes comprising a to f, whereas the remaining non-capsulated *H. influenzae* is designated non-typable (NTHi) (2). *H. influenzae* resides in the mucosa, and NTHi is mainly associated with infections in the respiratory tract, whereas encapsulated *H. influenzae* including Hib causes invasive disease. Until the 1990s type b (Hib) was the most common serotype, but a dramatic reduction of Hib cases was observed after introduction of a conjugate vaccine against Hib. However, an increasing incidence of invasive disease caused by non-type b *H. influenzae* has recently been reported from several countries (2-4).

In contrast to the very efficient vaccine against Hib, any suitable vaccine has not yet been implemented for NTHi. Lipopolysaccharide (LPS) and surface-exposed antigenic proteins of Gram negative pathogens are generally predicted as putative starting points for screening of suitable vaccine candidates (5). However, not all surface proteins or LPS are suitable for eliciting protection in the host against a particular pathogen. In recent years, several surface adhesin proteins including HMW-1 and -2 (6), pilA (7), P6 (8), and Protein D (9) have been analyzed for their vaccine potential. Some of these NTHi surface proteins showed initial protection in experimental models, and Protein D now is included in the vaccine Synflorix® giving partial protection against *H. influenzae* in humans (10, 11). Structural data are, however, only available for a few of these vaccine candidates.

We have described the role of a hitherto unknown *H. influenzae* protein E (PE) in interactions with host epithelial cells and in subversion of the host innate immune response (12). Protein E is a 16 kDa surface lipoprotein of *H. influenzae* that functions as an adhesin and induces a pro-inflammatory response during infection leading to IL-8 secretion and up-regulation of ICAM-1 (CD54) in both cell lines and primary epithelial cells originating from patients with COPD. An isogenic *pe* mutant showed defective adhesion and internalization of host epithelial cells. Furthermore, by using a peptide mapping approach we suggested that the amino acid region 84-108 is involved in binding to epithelial cells. Importantly, immunization with the PE 84-108 peptide showed significantly better pulmonary clearance in a mouse model in comparison to immunization with a non-related control peptide (12). When the *pe* gene was sequenced in all *Haemophilus* spp., including Hib and NTHi clinical isolates, we found that PE is a ubiquitous *Haemophilus* outer membrane protein (13). The active vitronectin-binding region PE 84-106 was found to be 100% conserved. PE homologues were also present in other members of the *Pasteurallaceae* family including *Aggregatibacter* spp., *Actinobacillus* spp., *Mannheimia succiniciproducens*, and *Pasteurella multocida* (13).

Vitronectin (Vn) and laminin (Ln) are amongst other proteins found in the extracellular matrix (ECM) (14, 15). In addition, Vn plays a crucial role in maintaining the homeostasis in the regulation of the complement system, *i.e.*, the innate immunity. We recently observed that NTHi binds Vn via surface-exposed PE, and that this interaction leads to an increased serum resistance. The peptide region covered by amino acids 84-108 is the Vn interaction domain (14-16), and further analysis revealed that in particular the residues K85 and R86 are involved Vn binding (17). In addition to the PE-dependent Vn-binding, we observed that the region comprising amino acids PE 41-68 interacts with Ln, an abundant ECM protein in the basement membrane, and that

this interaction leads to better adhesion of NTHi to host tissues (18). Interestingly, the Ln and Vn binding sites on the PE molecule are completely separate and do not interfere with each other during binding. The outer membrane PE thus both functions as an adhesin and concurrently protects bacteria from complement-mediated killing (18). More recently, we observed that PE also binds plasminogen that is ultimately converted into plasmin, and this consequently leads to degradation of complement 3 (C3) and dampening of the host innate response (19).

We recently published a technical report on crystallization data of a recombinant variant PE and a Se-methionine labeled PE (20). In the present study, the X-ray diffraction data at 1.8 Å resolution was analyzed in detail to obtain the structure of PE. Based upon the crystal, we show that the PE molecule is a dimer in the asymmetric unit. The structure of the monomer contained six antiparallel β-sheets connected with loops. At the C-terminal end, a rigid α-helix was found that was fixed at its position by a disulphide bond to the top of the β-sheet. The laminin and plasminogen binding regions of PE were exposed at the surface of the molecule, while the Vn binding region, previously defined by a peptide mapping approach was partially exposed. Furthermore, we used a structure based selection of the exposed regions to verify their localizations and respective immunogenicity by raising antibodies in mice. These structural findings provide insight in the regions that are involved in interactions with host proteins and a possible mode of PE-mediated host interactions by *H. influenzae*.

Material and Methods

Protein expression, purification and crystallization

PE and SeMet-PE were recombinantly expressed in *E. coli* as inclusion bodies and purification was performed as described previously (20). In brief, *E. coli* BL21 (DE3) harbouring plasmids

encoding for PE and SeMet-PE were grown in 1 litre LB with 50 µg/ml kanamycin, and protein expression was induced by addition of 1mM IPTG. Cells were lysed by sonication and inclusion bodies were collected by centrifugation. The inclusion bodies were washed with 5 M Urea, and dissolved in 10 ml of 8 M Urea. Refolding of PE was performed by the dilution method in refolding buffer containing 50 mM Tris-HCl pH 7.8, 500 mM NaCl, 5 mM DTT, 0.005% Tween-20, and 2M Urea. The refolded proteins were dialysed against 50 mM Tris-HCl, pH 7.8 buffer containing 135 mM NaCl and loaded into a Q-Sepharose^{TM} Fast Flow anion exchange (GE healthcare Biosciences, Uppsala, Sweden) column equilibrated with 50 mM Tris-HCl, pH 7.8 buffer containing 135 mM NaCl and 2 mM DTT. The flow-through of the column which contained PE were collected and further purified by Gel filtration column Superdex 200^{TM} (GE healthcare Biosciences) equilibrated with 50 mM Tris-HCl, pH 7.5 buffer containing 500 mM NaCl and 2 mM DTT (20). The protein purity was estimated by SDS-PAGE and concentrations were measured by a Nano-drop spectrophotometer.

The recombinant variant of PE and SeMet PE were concentrated up to 5 mg/ml. Crystallization conditions were screened by using commercial kits as described elsewhere (20). Crystals of recombinant variant PE were obtained in 100 mM SPG buffer pH 6.0, 25% w/v polyethylene glycol 1500. The SeMet PE produced crystals in 100 mM MES, pH 6, 200 mM NaCl and 20% PEG 6000. The size of the SeMet PE crystals was futher improved by using the microseeding technique. Both PE proteins produced crystals within 4-5 days of incubation at 25 ^{o}C. Details of the crystallization strategies can be obtained from (20).

Data collection and refinement

Crystals were quickly cooled in the presence of a cryo protectant solution in a MiTeGen loop using a stream of nitrogen gas at 100K before exposing them to the X-ray beam. The universal cryo solution containing 32% w/v glycerol, 32% w/v ethylene, 36% w/v sucrose and 2% glucose, was used for the recombinant variant PE crystals. Suitable crystals were incubated in drops of mother liquor to which an equal volume of the cryo solution was added. After about 30 seconds of incubation, they were mounted in loops and placed in a cryogenic N_2 gas stream at 100K. In contrast to the cryo conditions for the recombinant variant PE crystals, 15% w/v PEG 400 in reservoir solution was used for SeMet PE crystals. All data collections were performed at station I911-3, at the MAX IV Laboratory (Lund University Sweden). The station was equipped with MD2 goniostat (Maatel, Cedex, France), and a MAR225 CCD detector (Marresearch, Norderstedt, Germany).

To obtain reliable phase information, three data sets, peak (PK), point of inflection (PI) and remote (RM) were collected at the K edge of selenium (around 0.9795 Å) from the same SeMet containing crystal (see table 1 for data collection details and statistics). Data for PK and PI sets were collected first, with reduced exposure times such that the influence of radiation damage on the phasing could be minimized. The data for RM were, however, collected with longer exposure times in order to obtain data to the diffraction limit of the crystals (see Table 1). All data were integrated and scaled using the program XDS (21). The phasing power was monitored by using HKL2MAP (22), a graphical interface to a set of programs from the SHELX-suite. Six of the 8 possible Se positions could be identified using Patterson methods in SHELXC (23). Subsequent refinement and phasing of the structure was achieved by using auto-SHARP (24) using data from all 3 datasets.

Autosharp built 246 residues out of 282 possible using arp/warp (25). The model was inspected using COOT (26), and subsequent rounds of refinement and model building were performed by using the Phenix package (27) and Coot. In total 130 residues for each chain could be traced with 10 residues missing at the N-terminus and 1 residue at the C-terminal end of each of the protein chains. During refinement an additional 127 water molecules were added to the model. The final model had a good stereochemistry (r. m. s deviation from ideal geometry in bonds = 0.017 Å, in angles 1.65^{o}) and 97.3% of the residues lied in the most favoured part of the Ramachandran plot.

Data for the two different crystals forms of the recombinant variant PE were also collected. XDS was used for data processing and scaling (Table 1). The solvent content of the 2nd type of crystals was relatively low (around 30%) which probably was the reason for the diffracting power of these crystals. Both crystal forms were monoclinic and details of the different datasets are described in Table 1. The structure of the SeMet mutant form of PE was used as starting model in Molecular Replacement using Phaser (28) for the first crystal form. For the 2nd crystal form of the recombinant variant PE protein, the first form was used as search model and Phaser was used for the Molecular Replacement. Both structures were refined using TLS refinement within the Phenix suite and model inspection was performed using COOT. During refinement difference density indicated the presence of ethylene and glycerol molecules from the cryo-solutions, and these were placed accordingly. All coordinates and structure factors for the different models were submitted to the PDB database with codes xxxx for SeMet mutant for, xxxx for PE the recombinant variant PE 1 and xxxx for PE the recombinant variant PE 2.

Immunization of mice

Based upon the crystal structure of the PE, peptides were designed for immunization and conjugated with KLH (Innovagen AB, Lund, Sweden). Mice (BALB/c; 6 in each group) were immunized with KLH-conjugated peptides according to a standard immunization protocol (12). In brief, 50 µg peptide with complete Freunds adjuvant were subcutaneously administered. After 4 weeks, booster doses of 50 µg peptide with alumiminum hydroxide were injected the following three weeks. Blood was drawn one week after the last booster dose, and specific Abs were immunopurified using the same peptide, followed by testing in sandwich ELISAs against the various peptides or recombinant PE. The animal experiments were done according to the general rules established by the Swedish Government (Svensk författningssamling 1988:534, Ändringsförfattning 2012:256; Riksdagen, Riksgatan 1, SE-111 28, Stockholm, Sweden). An ethical permit (No. M193-11) was obtained from Malmö/ Lund District Court (Djurförsöksetiska nämnden, Tingsrätten, Byggmästaregatan 2, SE-222 37 Lund, Sweden). This body approved the protocol used in this study.

ELISA

PE peptides or recombinant PE22-160 (100 ng) were coated on 96-well PolySorb microtiter plates (Nunc-Immuno, Roskilde, Denmark) in 100 µl of 100 mM Tris-HCl, pH 9.0 for 15 h at 4°C. Coated plates were washed 3 times with PBS and blocked with 2.5% BSA in PBS containing 0.05% Tween-20 (PBST). Serum or purified antibodies in 2.5% BSA+PBST were added to the wells and allowed to bind for 1 h at room temperature. After washes with PBST, HRP-conjugated rabbit anti-mouse pAb (Dako, Denmark) were added in 2.5% BSA+PBST.

Finally, plates were washed 4 times with PBST and developed by horseradish peroxidase (HRP) substrate containing 20 mM tetramethylbenzidine and 0.1 M potassium citrate. After colour development, reactions were terminated by 1 M H_2SO_4, and finally plates were read at 450 nm in a microplate reader.

Flow cytometry

NTHi 3655 (12) from overnight cultures were grown in broth until OD_{600} 0.8. Thereafter, bacteria were washed twice with PBS containing 1 % BSA and incubated with purified mouse anti-PE peptide antiserum according to a standard protocol. After washing, bacteria were incubated with FITC-conjugated secondary rabbit-anti mouse pAbs (Dako Sweden, Stockholm, Sweden) followed by two washes and thereafter flow cytometry analysis (EPICS®XL-MCL, Coulter, Hialeah, FL).

Dynamic light scattering and Electron microscopy

Dynamic Light scattering (DLS) experiments were performed by using a Zetasizer Nano ZS (Malvern Instruments, Worcestershire, UK). Different concentrations of PE (1 mg/ml, 3 mg/ml and 5mg/ml) were used to collect data in triplicates and mean values were plotted. We used negative staining and transmission electron microscopy (TEM) to visualize the PE molecules as described elsewhere (29).

Results

Protein E is a dimer

The gel filtration profile showed that the majority (>85%) of recombinant PE molecules existed as dimers in solution, whereas only a minor fraction consisted of an oligomeric form during these conditions (Fig. 1A-B). The pooled fractions 13-18 in figure 1A were also analyzed by dynamic light scattering. These experiments confirmed that approx. 85% of PE that eluted as a single peak represented dimers, whereas the minor peak containing the remaining 10-15%, were tetramers (Fig. 1C). In addition, Transmision Electron Microscopy (TEM; negative staining) of recombinant PE also verified that > 85% of PE existed in dimers and minor fraction as tetramer or few monomers (Fig. 1D). Taken together, these results clearly indicated that PE is a dimer.

When the crystal structure was determined, we found that PE is also present as a dimer in the asymmetric unit. Each PE monomer consists of a β-sheet formed by 6 antiparallel β-strands (β1 39-43, β2 46-59, β3 64-78, β4 88-99, β5 102-114, β6a 118-122 and β6b 130-134) (Fig. 2A). In addition, a longer α-helix (138-151) packs on the concave face of the sheet (Fig. 2B) and strands β1, β2 and β3 are curved around it. Strand β6 consists out of two shorter strands β6a and β6b, interrupted by a short loop from residues 123 to 129. After the loop, the β2b lines up to strand β3, and proper hydrogen bond interactions for a sheet are formed. The edge of the sheet is formed by strand β1 and part of strand β2 at one side and strands β6a and β6b at the other side. The α helix is tethered to the β-sheet through a conserved disulphide bond between cysteine residues 99 and 148. This disulfide bond connects the top of the central strand of the β-sheet and one end of the α-helix together. The α–helix shields the upper half of the concave β-sheet and in

particular the part formed by strands β1, β2, β3 and partly β4, is protected from solvent by the helix while the remaining part of the face is exposed.

The convex side of the sheet is not protected by other structural elements in the PE monomer. However, the two monomers in the asymmetric unit pack together through the convex site of each of their β-sheets to form a nearly continuous antiparallel β-sheet that resembles a flattened barrel-like dimer structure. The dimer formation is nearly identical in the different crystal forms presented. The resulting dimer is 55Å x 44Å x 25Å in size (Fig. 3A-B). The barrel has a wedge like shape with the top more narrow than the bottom side. Loops connecting the β-strands at the top of the barrel (Fig. 2; loops 3, 5 and 7) are smaller than the loops found at the bottom of the barrel (loops 1, 2, 4 and 6). No electron density for amino acids 17-27 (before loop 1) of PE could be observed in the crystal structure, and they are hence not modeled. At each side of the barrel, strand β6b from one monomer and the end of strand β2 and loop 3 from another monomer lined up to close the barrel (Fig. 4A) with similar interactions between the two monomers at each of the sides of the barrel. The interactions between these two segments are not through main chain atom hydrogen bonding interactions as seen in true β-barrels but rather through interactions from atoms from side chains from residues in loop 52 to 61 from one monomer (named PE_A) to main chain atoms in the strand β6b in the other monomer (named PE_B). The central two residues are Asn60 and TrpA57. The Nδ2 of Asn60 forms a hydrogen bond with the backbone oxygen of Thr131, whereas the Oδ1 makes a hydrogen bond with the backbone nitrogen from the same Thr131. The side chain of Trp57 occupies a hydrophobic pocket formed by the aliphatic part of the side chain of Lys129, the methyl moiety of Thr105 and the side chain of His67. The Nε1 of Trp57 make a hydrogen bond with the backbone oxygen of Lys129 (Fig. 4A). Other interactions are formed between Asp59 and Thr131, Gln61 and Ser133.

Tyr94 from strand β4 completes the main interactions through a hydrogen bond to the Nz of Lys129 (Fig. 4A). The dimer interactions are not extensive and only 625 $Å^2$ are buried in the interface according to an analysis using Pisa (30).

Both the top and bottom sides of the barrel-like feature are hollow and deep pockets are formed (Fig. 4B). The pocket at the wider bottom is larger, not well defined or well protected from solvent. The volume of the pocket is 857 $Å^3$ as calculated by Pocketfinder (http://www.modelling.leeds.ac.uk/pocketfinder/). This pocket is lined with a number of charged residues, most notable 4 arginines (Arg108 and Arg121 from each subunit) that come together in the center of the pocket to form a ring like arrangement (Fig. 4B). Two Asp residues (Asp110 from each monomer) are positioned between two of the Arg residues and provide some charge compensation (Fig. S1). Further charged residues that line the inner pocket are Lys125 and Lys126 (Fig. S1).

The second pocket at the top of the barrel is much more confined in space and better defined. It has a flattened shape with dimensions of 14Å x 8Å and is about 10Å deep with a volume of 363 $Å^3$ (Fig. 4B). The central part of the pocket is lined with four Ile residues (Ile65 and Ile96 from each monomer), whereas at the inside walls of the pocket more polar residues (His67, Asp59 and Thr105 from each monomer) are found (Fig. 4B and 4C). The rim of the pocket is also polar of character and residues His103, Asn98 and Asn101 are located here. Since this pocket is very well defined in shape and protected from the solvent, it seems likely that it has a ligand binding feature. The residues that are lining the walls of both pockets are, however, not particularly conserved within a family of homologous proteins from different pathogens as described in next sections.

Functional host factor-binding regions of PE are exposed at the surface of the molecule

We previously used a peptide mapping approach to identify the PE amino acid regions that were involved in binding to epithelial cells (12). A peptide consisting of the PE84-108 amino acid region directly interacted with lung epithelial cells. More recently, a similar region was verified that bound Vn (16, 17). In the present crystal structure, the 84-108 amino acid region is located partly in loops 4 and 5, and in beta strands 4 and 5 (Fig. 5A). The structure shows that loop 4 plays a role at the base of protein, where amino acids Asn73-Arg89 are well exposed (Fig. 5B). In contrast, the Ser90-Arg108 part is located in beta strands 4 and 5 that are positioned in the core of the PE dimer. Thus, the loop 4 region upward to amino acid 84 seems more likely to be the real epithelial cell interacting domain (Fig. 5B). Recently we showed that amino acids Lys85 and Arg86 are important for the PE-dependent interaction with vitronectin (17). Importantly, in loop 4, the side chains for Lys85 and Arg86 are well exposed (Fig. 5A and B). The linear sequence of loop 4 in all *Haemophilus* spp. is highly conserved, which suggests similar ligand specificity, while in other pathogens this loop is only partially conserved (Fig. 6A).

Since the multifunctional PE also interacts with laminin, we previously applied a peptide mapping approach to determine the specific laminin binding amino acid region in PE. Interestingly, the region consisting of PE aa 41-68 was directly interacting with Ln (18). The same peptide also has affinity to plasminogen (19). This particular region is located in the loop 1 between strands β1 and β2 and beta strand β2 and loop 2. This whole region is partly exposed at the surface and presents diagonally along with the whole length of the dimer (Fig. 5C and D). In parallel with loop 4, the laminin/plasminogen binding region (loop 3) is also conserved among *Haemophilus* spp., but it is diverse in other pathogens. However, loop 3 has acidic amino acids

that are highly conserved in most of the PE homologues for which sequence information is available (Fig. 6A).

PE is a unique adhesin of the Pasteurallaceae *family*

All pathogenic bacterial species are equipped with specific adhesins that very precisely recognize host tissues in particular anatomical niches. This group of proteins has numerous structurally diverse members. In general, adhesins may be fimbriae, pili or other non-pilus adhesins (31). There are only a limited numbers of adhesins for which structures have been determined and that limits the categorization of this group of proteins (http://supfam.cs.bris.ac.uk). The structural classification of adhesins performed on the basis of domains involved in recognition of host proteins has been performed in the database available at http://supfam.cs.bris.ac.uk/SUPERFAMILY/cgi-bin/scop.cgi?sunid=49401. This includes a limited number of proteins, such as, collagen-binding, fibrinogen-binding, pilus subunits, PapG adhesion receptor-binding, F17-C type adhesin and Dr-family adhesins. However, a large number of other known bacterial adhesins are not yet categorized due to lack of structural information on the domain recognitions.

Pasteurellaceae comprises a large and diverse family of Gram-negative proteobacteria (32). Recently this family was classified into 17 different genera (www.pasteurellaceae.life.ku.dk/). The latest BLAST update in the expasy (NCBI BLAST2) database showed that there are 11 different bacterial pathogens that have a *pe* gene homologue (Fig. 6A). The detailed similarity matrix revealed a sequence identity between 34.1-98.8% and similarity 55.3-99.4% (Fig. 6B), and primary amino acid sequence based phylogenetic tree showed 5 separate clusters (Fig. 6C). *Haemophilus* spp. grouped into a single cluster with

65.3%-98.8% identity and 77.8-99.4% similarity. *Actinobacillus actinomycetemcomitans* showed 70.9-78.4% sequence identity and 80.8-88.6% similarity with *A. aphrophilus* and *A. segnis*, respectively. In the third cluster, *A. succinogenes* showed 43.6% identity and 65.4% similarity with *M. succiniciproducens* (Fig. 6B). *Pasteurella* spp. also have PE analogues that are approximately 34% similar to *H. influenzae* PE, whereas PE derived from *P. multocida* is only 45.9% identical and 64.8% similar to *P. dagmatis*. In addition to these *Pasteurellaceae* family members, *Enterobacter cloacae*, and *Listeria monocytogenes* have PE orthologues (E3G4R8 and E1UDL7, respectively), albeit with lower homology (Fig. 6B and C). We suggest that PE belongs to a unique group of bacterial outer membrane proteins in the family *Pasteurallaceae*.

PE belongs to the cystatin family of proteins

To verify the folding pattern, the PE structure was matched with a fold database (33). The arrangement of a 6-stranded antiparallel β-sheet and α–helix packing at one face as seen within the monomer, was observed in a number of different proteins. In these proteins, however, the other side of the sheet is normally shielded by other secondary structure elements and thus the sheet and helix are part of larger fold arrangements, which are not related to each other. The closest structures, which have the same minimal arrangement, can be seen in the monellin/cystatin family fold (34), which is also composed of a single antiparallel β-sheet with an α-helix protecting the concave face of the sheet. Proteins in this family are either cysteine protease inhibitors (cystatin-like) or sweet proteins from plants (monellin-like). The basic folds of, *e.g.*, monellin and PE superimpose well (Fig. 7). However, the position of the helix within the secondary structure topology of the monellin/cystatin family fold is different from what it is within the PE structure. In the monellin/cystatin family it is placed after the first strand of the

sheet, whereas it is positioned at the C-terminal end of the PE molecule. The helix for the monellin/cystatin family is positioned in the centre of the beta-sheet face and protects the whole face of the protein. The helix in PE is, however, tethered to the top of the sheet by the disulphide bridge and therefore part of the sheet is more exposed. Thus, it seems that the PE and the monellin/cystatin most probably are analogues of each other and probably are products of convergent evolution.

The most homologous regions for all the species are located in the secondary structure elements especially within the central strand β3. Interestingly, one region with a very high homology is not associated with a secondary structure element but instead it is found in loop 6 and consists of the fingerprint [FY]111-[YH]-X-[DE]-F-W-G-X-G119. This loop between β5 and β6 folds back over the convex inner side of the sheet and forms a small hydrophobic core on the inside bottom of the β-barrel like dimer. The aromatic residues in the fingerprint interact with residues from the different strands as well as with the N-terminal loop 1. Especially the residues Tyr/His112 and Trp116 are involved in multiple interactions with residues from different secondary structure elements (Fig. 6A). Using the Prosite webserver (35) to scan the database for sequences with similar fingerprints, it seems that this fingerprint is unique and it might thus be an indicator of this particular fold family.

Structural approach for mapping immunogenic and surface-exposed regions of PE by raising peptide antibodies

Previously we reported that immunization with PE 84-108 mediated protection in a mouse pulmonary clearence model (12). The ubiquitous presence of PE among NTHi clinical isolates

further signified the importance of this protein as a vaccine candidate (13). To verify our crystal model as well as to analyze the immunogenic capacity of various surface-exposed regions of PE, we designed a series of peptides as outlined in figure 8A and B (PE24-37, PE74-89, PE84-108, and PE134-156). In addition, a peptide based upon the cryptic region PE104-128 was used as a negative control. KLH-conjugated peptides were used for subcutaneous immunization of BALB/c mice. Sera from immunized mice (*n*=6 in each group) were pooled, peptide Abs were purified using affinity chromatography, and finally tested in ELISA. All peptides produced specific Abs directed against their corresponding peptides (Fig. 8C). In contrast, the resulting antibodies showed different recognition patterns against recombinant full length PE22-160 in ELISA. Antibodies directed against PE24-37, PE74-89, and PE134-156 recognised the PE22-160 molecule, whereas the anti-PE84-108 Ab recognised PE22-160 with less efficiency (Fig. 8D). However, purified peptide antibodies directed against PE24-37 and PE84-108 displayed cross reactivity against both peptides in ELISA. As expected, the negative control Ab against PE104-128 did not detect full length PE.

We also wanted to test whether our peptide Abs recognized PE when the molecule was surface-exposed in its native PE form on the clinical isolate NTHi 3655. Bacteria were incubated with Abs followed by flow cytometry analysis. In parallel with the results obtained by ELISA, anti-PE24-37, anti-PE74-89, and anti-PE134-156 Abs all detected PE at the bacterial surface, whereas the anti-84-108 Ab showed a minor shift when compared to the negative control consisting of the secondary antibody only (Fig. 8E). Taken together, Abs directed against surface-exposed epitopes of the PE molecule that were defined by designed peptides fit very well with the crystal structure.

Discussion

Protein E is an outer membrane protein found in both encapsulated *H. influenzae* and NTHi. We have shown that NTHi equipped with the multifunctional adhesin PE binds to several different host proteins including Vn, Ln and plasminogen (13, 16, 17, 19). The majority of NTHi infections are in general non-invasive albeit an increasing trend of invasive NTHi disease has been observed the last 5 years (3). Active degradation of host tissues and the ECM is, however, a property of typeable *H. influenzae* such as Hib. Adherence mediated by the PE-Ln interaction thus could provide a firm anchorage to host basal lamina and tissues (Fig. 9). In contrast to Hib, NTHi may also use Ln as a ligand, particularly when the epithelial mucosa is destroyed by a prior viral infection or mixed bacterial infections. In parallel, plasminogen bound to PE degrades C3b and inhibits the complement-mediated innate immunity. The protease activity of plasmin also degrades the ECM (Fig. 9), which eventually will increase the host damage. Experimental evidence suggests that Ln and plasminogen both share a similar binding domain in PE (16, 19). This might of course be conditional according to the abundance and affinity of ligands. An interesting finding is that recombinant PE simultaneously binds Ln and Vn (18). We have also observed that laminin LG1-5 domains of the alfa chain interact with PE and that this interaction can be inhibited by heparin (18). The PE loop 3 region has a few acidic residues (residues D59 and E62) that might be involved in the binding to the LG4-5 domains of laminin. The present data of a dimerized PE molecule (Fig. 3A) explains these *in vitro* observations. However, it is at present unclear whether the dual ligand-binding mode of the PE molecule also exists *in vivo*.

Protein E is a lipoprotein that has a signal peptide at the N-terminal followed by Cys16 (12). According to the general described mechanisms of lipoprotein transport and lipidation mechanisms, the protein is transported to the outer membrane followed by addition of lipid

chains and removal of the signal peptide (36, 37). The Cys16 residue is thus predicted to be involved in lipidation and functions as an anchor of PE on the outer membrane of bacteria. The N-terminal of the dimer is supposed to face towards the membrane side of the bacteria, whereas the C-terminal is facing the outside.

The presence of a well-formed pocket on the top side of the dimer (Fig. 3B) is intriguing. The shape and accessibility of the pocket is very reminiscent of binding pockets in smaller ligand binding proteins. However, a clear function for this pocket cannot be deduced and due the variability of the residues lining the pocket it does not seem to be associated with a conserved function within the family of PE-like proteins. It has been suggested, however, that a high level of diversity within interacting surfaces of adhesins when in contact with their host proteins, stems from the need of the pathogen to change these recognition sites in order to avoid an immune response, while retaining a common biological function (38). This possibility needs to be further investigated. Identification of a putative ligand to the smaller and well defined pocket of PE would be an asset in these investigations.

Protein E is known as one of the *H. influenzae* Vn-binding proteins with the highest affinity (K_d=4.0x10^{-7} M), and most importantly, Vn bound to PE protects NTHi from the membrane attack complex (MAC). The vitronectin-binding region PE84-108 that is embedded in the structure is mostly located in the dimer interface. Recently, we further suggested that Lys85 and Arg86 of PE are involved in binding to Vn (17). Lys85 and Arg86 are located in the end of the exposed loop 4. Thus, a probable region for Vn interactions may be mostly located between amino acids Leu78-Ala88 of loop 4. Vn is also known to function as a bridge molecule that interacts with bacterial surface proteins and the host integrins via an RGD motif (14, 39).

Similarly, the PE-mediated recruitment of Vn by *H. influenzae* might also be involved in adhesion and internalization into epithelial cells.

Protein E meets the criteria of being suitable as a vaccine candidate, *i.e.*, its ubiquitous presence, highly conserved nature, favoured immunogenicity and protective role in model animals (12). In our preliminary experiments, the structure based approach for designing antibodies against the surface-exposed part of the PE molecule recognised recombinant PE in addition to PE at the surface of bacteria (Fig. 8E). This indicated that these peptides may be appropriate for immunization against NTHi.

PE has a unique three dimensional structure that is involved in adhesion of *H. influenzae* to epithelial cells and binds and/ or recruits multiple host proteins. The PE structure is formed by β-sheets with an α-helix covering one side. Within the crystal structure the protein is present as a dimer with the two monomers forming a β-barrel like structure. The dimer of PE resembles a flattened β-barrel like structure, as it seems to form a continuous up and down antiparallel β sheet resulting in formation of large pockets inside the molecule. However, in a number of aspects it differs from true β-barrel structure. The first one is that at the sides of the monomers interacting with each other, the contact surfaces are not made through β-strands, but instead loops are used in an extended conformation. The hydrogen bonding system is not according to that observed in a β-sheet; here instead hydrogen bonds are built up through side chain interactions. Although the interaction area is limited, the presence of dimers is also seen in gel filtration, DLS and TEM experiments, concluding that this is not likely to be an artefact during crystallisation. The other unusual aspect of this "barrel-like" structure is the way the barrel fans out at the bottom side. This is mainly achieved by the high twist in the separate sheets of each

monomer. Only the top part of the barrel is intact and continuous while the bottom part is sheared and open at the sides. This is an unusual arrangement and not observed in other proteins as judged by fold recognition searches using Dali (33).

The protein folding pattern has a partial match with the monellin/cystatin family. However, cystatins are in general monomeric, whereas monellin is a heterodimer formed by an A chain of 42 aa and a B-chain of 50 aa. A higher oligomeric variant of monellin has been observed in crystal structures (40). The dimer arrangement of PE is, however, different from monellin, and for PE a more side to side barrel like dimer is found. Proteins in the monellin/cystatin family have been studied extensively for their capacity to aggregate (41). However, the aggregation of PE has not yet been fully investigated. Besides this similarity with monellin/cystatin, there are no other bacterial adhesins that have this kind of structural appearance. However, orthologues belonging to the PE adhesin sub-family are also present in *E. cloacae* and *L. monocytogenes*. We conclude from sequence comparisons within this set of proteins that the residues, which are important for the specific features of the fold, are very well conserved. These conserved patches at the surface of PE that might be important for the structure of the common fold, are shown in figure S2. The specific role of these proteins in the various bacterial species is not yet elucidated. In fact, the best characterized member of this family is PE, which is described in the current paper. We would therefore propose that PE and its homologs belongs to a new group of adhesins that is particularly present in the family of *Pasteurellaceae*.

Acknowledgements

This work was supported by Alfred Österlund, the Anna and Edwin Berger, Greta and Johan Kock, Lars Hierta, FLÄK, the Gyllenstiernska Krapperup, the Marianne and Marcus Wallenberg

Foundation, the Physiographical Society, the Swedish Medical Research Council (grant number 521-2010-4221, www.vr.se), and the Cancer Foundation at the University Hospital in Malmö, and Skåne County Council´s research and development foundation. We would like to thank the staff at the MAX-IV laboratories and the crystallization facility for their valuable help.

References

1. **Murphy T F, Faden H, Bakaletz L O, Kyd J M, Forsgren A, Campos J, Virji M, and Pelton S I**. 2009. Nontypeable *Haemophilus influenzae* as a pathogen in children. Pediatr. Infect. Dis. J. **28**:43-48.
2. **van Wessel K, Rodenburg G D, Veenhoven R H, Spanjaard L, van der Ende A, and Sanders E A**. 2011. Nontypeable *Haemophilus influenzae* invasive disease in The Netherlands: a retrospective surveillance study 2001-2008. Clin. Infect. Dis. **53**:e1-7.
3. **Resman F, Ristovski M, Ahl J, Forsgren A, Gilsdorf J R, Jasir A, Kaijser B, Kronvall G, and Riesbeck K**. 2011. Invasive disease caused by *Haemophilus influenzae* in Sweden 1997-2009; evidence of increasing incidence and clinical burden of non-type b strains. Clin. Microbiol. Infect. **17**:1638-1645.
4. **Rubach M P, Bender J M, Mottice S, Hanson K, Weng H Y, Korgenski K, Daly J A, and Pavia A T**. 2011. Increasing incidence of invasive *Haemophilus influenzae* disease in adults, Utah, USA. Emerg. Infect. Dis. **17**:1645-1650.
5. **Grandi G**. 2001. Antibacterial vaccine design using genomics and proteomics. Trends Biotechnol. **19**:181-188.

6. **Giufre M, Carattoli A, Cardines R, Mastrantonio P, and Cerquetti M**. 2008. Variation in expression of HMW1 and HMW2 adhesins in invasive nontypeable *Haemophilus influenzae* isolates. BMC Microbiol. **8**:83.

7. **Jurcisek J A, Bookwalter J E, Baker B D, Fernandez S, Novotny L A, Munson R S, Jr., and Bakaletz L O**. 2007. The PilA protein of non-typeable *Haemophilus influenzae* plays a role in biofilm formation, adherence to epithelial cells and colonization of the mammalian upper respiratory tract. Mol. Microbiol. **65**:1288-1299.

8. **Chang A, Kaur R, Michel L V, Casey J R, and Pichichero M**. 2011. *Haemophilus influenzae* vaccine candidate outer membrane protein P6 is not conserved in all strains. Hum. Vaccin. **7**:102-105.

9. **Dicko A, Odusanya O O, Diallo A I, Santara G, Barry A, Dolo A, Diallo A, Kuyinu Y A, Kehinde O A, Francois N, Borys D, Yarzabal J P, Moreira M, and Schuerman L**. 2011. Primary vaccination with the 10-valent pneumococcal non-typeable *Haemophilus influenzae* protein D conjugate vaccine (PHiD-CV) in infants in Mali and Nigeria: a randomized controlled trial. BMC Public Health. **11**:882.

10. **Prymula R, Kriz P, Kaliskova E, Pascal T, Poolman J, and Schuerman L**. 2009. Effect of vaccination with pneumococcal capsular polysaccharides conjugated to *Haemophilus influenzae*-derived protein D on nasopharyngeal carriage of *Streptococcus pneumoniae* and *H. influenzae* in children under 2 years of age. Vaccine. **28**:71-78.

11. **Murphy T F**. 2009. Current and Future Prospects for a Vaccine for Nontypeable *Haemophilus influenzae*. Curr. Infect. Dis. Rep. **11**:177-182.

12. **Ronander E, Brant M, Eriksson E, Morgelin M, Hallgren O, Westergren-Thorsson G, Forsgren A, and Riesbeck K**. 2009. Nontypeable *Haemophilus influenzae* adhesin protein E: characterization and biological activity. J. Infect. Dis. **199**:522-531.

13. **Singh B, Brant M, Kilian M, Hallstrom B, and Riesbeck K**. 2010. Protein E of *Haemophilus influenzae* is a ubiquitous highly conserved adhesin. J. Infect. Dis. **201**:414-419.

14. **Singh B, Su Y C, and Riesbeck K**. 2010. Vitronectin in bacterial pathogenesis: a host protein used in complement escape and cellular invasion. Mol. Microbiol. **78**:545-560.

15. **Singh B, Fleury C, Jalalvand F, and Riesbeck K**. 2012. Human pathogens utilize host extracellular matrix proteins laminin and collagen for adhesion and invasion of the host. FEMS Microbiol. Rev. **36**:1122-1180

16. **Hallstrom T, Blom A M, Zipfel P F, and Riesbeck K**. 2009. Nontypeable *Haemophilus influenzae* protein E binds vitronectin and is important for serum resistance. J. Immunol. **183**:2593-2601.

17. **Singh B, Jalalvand F, Morgelin M, Zipfel P, Blom A M, and Riesbeck K**. 2011. *Haemophilus influenzae* protein E recognizes the C-terminal domain of vitronectin and modulates the membrane attack complex. Mol. Microbiol. **81**:80-98.

18. **Hallstrom T, Singh B, Resman F, Blom A M, Morgelin M, and Riesbeck K**. 2011. *Haemophilus influenzae* protein E binds to the extracellular matrix by concurrently interacting with laminin and vitronectin. J. Infect. Dis. **204**:1065-1074.

19. **Barthel D, Singh B, Riesbeck K, and Zipfel P F**. 2012. *Haemophilus influenzae* Uses the Surface Protein E To Acquire Human Plasminogen and To Evade Innate Immunity. J. Immunol. **188**:379-385.

20. **Singh B, Al Jubair T, Fornvik K, Thunnissen M M, and Riesbeck K**. 2012. Crystallization and X-ray diffraction analysis of a novel surface-adhesin protein: protein E from *Haemophilus influenzae*. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. **68**:222-226.

21. **Kabsch W**. 2010. Xds. Acta Crystallogr. D Biol. Crystallogr. **66**:125-132.

22. **Pape. T a S, R.** 2004. HKL2MAP: a graphical user interface for phasing with SHELX programs. J. Appl. Cryst. **37**:843-844.

23. **Sheldrick G M**. 2008. A short history of SHELX. Acta. Crystallogr. A. **64**:112-122.

24. **Vonrhein C, Blanc E, Roversi P, and Bricogne G**. 2007. Automated structure solution with autoSHARP. Methods Mol. Biol. **364**:215-230.

25. **Langer G, Cohen S X, Lamzin V S, and Perrakis A**. 2008. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nat. Protoc. **3**:1171-1179.

26. **Emsley P, Lohkamp B, Scott W G, and Cowtan K**. 2010. Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. **66**:486-501.

27. **Adams P D, Afonine P V, Bunkoczi G, Chen V B, Davis I W, Echols N, Headd J J, Hung L W, Kapral G J, Grosse-Kunstleve R W, McCoy A J, Moriarty N W, Oeffner R, Read R J, Richardson D C, Richardson J S, Terwilliger T C, and Zwart P H**. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. **66**:213-221.

28. **McCoy A J, Grosse-Kunstleve R W, Adams P D, Winn M D, Storoni L C, and Read R J**. 2007. Phaser crystallographic software. J. Appl. Crystallogr. **40**:658-674.

29. **Engel J, and Furthmayr H**. 1987. Electron microscopy and other physical methods for the characterization of extracellular matrix components: laminin, fibronectin, collagen IV, collagen VI, and proteoglycans. Methods Enzymol. **145**:3-78.

30. **Krissinel E, and Henrick K**. 2007. Inference of macromolecular assemblies from crystalline state. J Mol. Biol. **372**:774-797.

31. **Soto G E, and Hultgren S J**. 1999. Bacterial adhesins: common themes and variations in architecture and assembly. J. Bacteriol. **181**:1059-1071.

32. Kuhnert P, and Christensen, H., ed. 2008. *Pasteurellaceae: Biology, Genomics and Molecular aspects*. Caister Academic Press, Norfolk, UK.

33. **Holm L, and Rosenstrom P**. 2010. Dali server: conservation mapping in 3D. Nucleic Acids Res. **38**:W545-549.

34. **Esposito V, and Temussi, P. A.** 2011. Cystatins: a versatile family. BioMolecular Concepts. **2**:95-102.

35. **de Castro E, Sigrist C J, Gattiker A, Bulliard V, Langendijk-Genevaux P S, Gasteiger E, Bairoch A, and Hulo N**. 2006. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Res. **34**:W362-365.

36. **Hayashi S, and Wu H C**. 1990. Lipoproteins in bacteria. J. Bioenerg. Biomembr. **22**:451-471.

37. **Nakayama H, Kurokawa K, and Lee B L**. 2012. Lipoproteins in bacteria: structures and biosynthetic pathways. FEBS. J. **279**:4247-4268.

38. **Johnson S, Tan L, van der Veen S, Caesar J, Goicoechea De Jorge E, Harding R J, Bai X, Exley R M, Ward P N, Ruivo N, Trivedi K, Cumber E, Jones R, Newham L,**

Staunton D, Ufret-Vincenty R, Borrow R, Pickering M C, Lea S M, and Tang C M. 2012. Design and Evaluation of Meningococcal Vaccines through Structure-Based Modification of Host and Pathogen Molecules. PLoS Pathog. **8**:e1002981.

39. **Bergmann S, Lang A, Rohde M, Agarwal V, Rennemeier C, Grashoff C, Preissner K T, and Hammerschmidt S**. 2009. Integrin-linked kinase is required for vitronectin-mediated internalization of *Streptococcus pneumoniae* by host cells. J. Cell. Sci. **122**:256-267.

40. **Bujacz G, Miller M, Harrison R, Thanki N, Gilliland G L, Ogata C M, Kim S H, and Wlodawer A**. 1997. Structure of monellin refined to 2.3 a resolution in the orthorhombic crystal form. Acta Crystallogr. D Biol. Crystallogr. **53**:713-719.

41. **Esposito V, Guglielmi F, Martin S R, Pauwels K, Pastore A, Piccoli R, and Temussi P A**. 2010. Aggregation mechanisms of cystatins: a comparative study of monellin and oryzacystatin. Biochemistry. **49**:2805-2810.

42. **Corpet F**. 1988. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res. **16**:10881-10890.

43. **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S**. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28**:2731-2739.

Figure legends

FIG 1 Purification of recombinant PE. (A) The gel filtration profile of the recombinant variant PE (approx. 2 mg protein purified by anion exchange as described in the Material and Methods) was injected in an equilibrated Superdex 200^{TM} column. The separation chromatogram of PE is shown here along with a standard molecular weight filtration marker. (B) Fractions collected from (A) were separated using 12% SDS-PAGE followed by staining with Commassie blue R250. A similar gel filtration profile and purity pattern was also observed with SeMet PE (data not shown). (C) Dynamic light scattering (DLS) pattern of the recombinant variant PE purified by gel filtration. (D) Transmission electron microscopy showing PE molecules. Bar represents 100 nM.

FIG 2 Secondary structure elements and the PE monomer. (A) PE$^{28\text{-}159}$ amino acids showing the secondary structure. In total, 6 β-strands, 8 loops and one C-terminal helix exist. (B) In the monomer, 6 antiparallel β-strands form the β sheet. A longer α-helix packs on the concave face of the sheet, where strands β1, β2 and β3 are curved around it. It is tethered to the β-sheet through a conserved disulphide bond between cysteines 99 and 148.

FIG 3 The PE dimer. (A) A cartoon representation of the PE dimer. (B) The surface of the PE dimer shown from the top and bottom cavities in addition to charge distribution of the molecule. The top surface of the molecule is neutral in charge, whereas the bottom side of the dimer is basic. The positive and negative charges are shown in blue and red, respetively.

FIG 4 Dimer interface and PE internal cavities. (A) Stereoview showing side chains in the loop from residues 52 to 61 from one monomer (green A) to main chain atoms in the strand β6b in the other monomer (red B). Two central residues are AsnA60 and TrpA57. The Nδ2 of AsnA60 forms a hydrogen bond with the backbone oxygen of ThrB131 and Oδ1 makes a hydrogen bond with the backbone nitrogen from the same ThrB131. The side chain of TrpA57 occupies a hydrophobic pocket formed by the aliphatic part of the side chain of Lys B129, the methyl moiety of Thr B105 and the side chain of His A67. The Nε1 of TrpA57 makes a hydrogen bond with the backbone oxygen of LysB129. Other interactions are formed between Asp A59 and Thr B131, Gln A61 and Ser133. Tyr A94 from strand β4 completes the main interactions through a hydrogen bond to the Nz of Lys B129. (B) Top and bottom pockets of the dimer in red and blue, respectively. The side chains of Arg 108 and 121 in each monomer are shown. (C) Schematic drawing of the residues lining the walls of the top pocket of the dimer. The coloured halves represent the two monomers and the coloured residues inside the pocket are located at the bottom wall of the pocket.

FIG 5 Binding sites of the various host factors on the PE dimer. (A) Ribbon diagram of a PE dimer that shows the Vn-binding region. The two separate monomers are shown in green and light blue, whereas the amino acid region 84-108 (based on peptide mapping) is shown in dark magenta and blue colour in the two monomers. (B) Surface of the PE dimer with exposed K85 and R86 side chains along with other residues. (C) The Ln/plasminogen binding region of PE. Ribbon model that shows the Ln/plasminogen binding region in red colour. Amino acids of the binding region are shown as stick models. For ease of following the chain, several residues along

this fragment are numbered. (D) The surface of PE with the Ln and plasminogen binding region indicated as red.

FIG 6 PE homologues also exist in other pathogens. (A) BLAST results revealed that several other pathogens have PE homologues. Retrieved sequences were aligned by using MultAlin online tool (42). Highly conserved amino acids are shown in capital red colour, and partially conserved amino acids are in blue colour. Secondary structures are shown at the top with beta strands as arrows and helices as cylinders. (B) Identity and similarity matrix of protein sequences analyzed by using the pairwise alignment tool Needle (www.ebi.ac.uk). (C) Rooted phylogenetic tree showing 5 different clusters. *Haemophilus* spp. devised in a single cluster and distant related sequences *Pasteurella* spp., and *E. cloacae*, and *L. monocytogenes* were devised in two separate clusters. Sequences were analyzed for phylogenetic relations by using MEGA5 software (43).

FIG 7 PE has a monellin-like fold. Superimposed structure of Monellin (PDB code: 2O9U) chain A (blue) and PE monomer (green). The beta sheets arrangements in the protein backbone are very similar. However, helices are present at the C-terminus in PE and at the N-terminus in monellin.

FIG 8 Mapping of surface-exposed immunogenic regions using mouse anti-PE peptide Abs. (A) A ribbon diagram shows what PE amino acid regions (indicated with different colours) that were

selected for immunization of a series of mice. (B) Surface structure of selected regions as shown in (*A*). (C) Results from ELISA that demonstrate Ab recognition of peptides that were coated on microtiter plates. (D) Recognition of full length recombinant PE22-160 by peptide Abs shown in ELISA. (E) Flow cytometry profiles of NTHi showing surface recognition of PE at the surface of bacteria using anti-PE peptide Abs.

FIG 9 Schematic representation of the multiple functions of PE. (A) *H. influenzae* PE interacts with the host epithelium. The PE molecule mediates binding to the epithelial surface utilizing a hitherto unknown receptor. This interaction contributes to bacterial adhesion and induction of a proinflammatory response by the epithelial cells. PE binds to Ln, which contributes in adhesion of *H. influenzae* to the basement membrane and the host ECM. In addition to Ln, *H. influenzae* binds plasminogen by using PE. When plasminogen is bound to PE it is converted into active plasmin by host urokinase plasminogen activator (uPA) or tissue plasminogen activator (tPA). Active plasmin may help in bacterial invasion and degradation of the ECM. (B) PE inhibits the membrane attack complex (MAC) at the surface of *H. influenze* that is accomplished by the binding of Vn. Vitronectin is a well known complement regulator that inhibits the terminal complement pathway by interacting with the C5b-7 complex assembly, and also inhibits C9 polymerization during formation of a lytic pore. PE is able to bind both Ln and Vn at the same time and these both ligands do not interfere with each other, thus adhesion and MAC inhibition can be carried out simultaneously.

Table 1. Data processing, phasing and refinement statistics.

	PE 1	**PE 2**	**SeMet peak**	**SeMetinf**	**SeMet rem**
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$
a (Å)	44.2	54.7	44.1	44.1	44.1
b (Å)	57.3	42.5	56.9	56.9	56.9
c (Å)	61.4	56.8	61.4	61.4	61.4
β (°)	96.05	116.3°			
Wavelength (Å)	1.000	1.000	0.97918	0.97942	0.9700
Maximum resolution (Å)	1.8	2.1	2.6	2.7	2.3
$V_{Matthews}$(Da/Å^3) Dimer in asymmetric unit	2.28	1.75	2.26	2.26	2.26
Solvent contend (%)	46.0	29.7	45.7	45.7	45.7
Total observations	115628	32032	62016	60725	101090

Unique reflections	28219	12867	17365	16137	26407
Rmerge (%)	6.2 (45.7)	5.6 (45.9)	5.5 (20.2)	4.7 (15.7))	9.4 (50.4)
Completeness (%)	99.5 (98.6)	91.4 (79.8)	94.4 (68.2)	98.8 (94.3)	99.0 (95.1)
Multiplicity					
Average (I)/(ΣI)	12.34 (2.64)	11.94 (2.03)	16.37 (4.19)	20.45 (6.76)	11.40 (2.85)
Phasing					
FOM	-	-			0.32
Phasing Power (isomorpous)	-	-	0	0.713 (0.538)[#]	0.331 (0.253)
Phasing Power (anomalous)	-	-	1.079	1.050	0.416
Cullis R-factor (isomorphous)	-	-	0	0.617 (0.642)	0.899 (0.976)
Cullis R-factor (anomalous)	-	-	0.843	0.826	0.958
Refinement					

Refinement range (Å)	29-1.8	28-2.1	-	-	29-2.3
Rcrys (%)	20.07	18.17	-	-	18.68
Rfree (%)	23.10	24.46	-	-	25.49
No of water molecules	127	50	-	-	119
Other	3 GOL, 2EDO	1 GOL, 2 EDO	-	-	
RMSD bond length (Å)	0.013	0.07	-	-	0.009
RMSD angles (°)	1.498	1.055	-	-	1.208
Favorable region	99.2	98.4	-	-	98.1
Additional allowed region	0.8	1.6	-	-	1.9
Disallowed	0	0	-	-	0

Figure 1

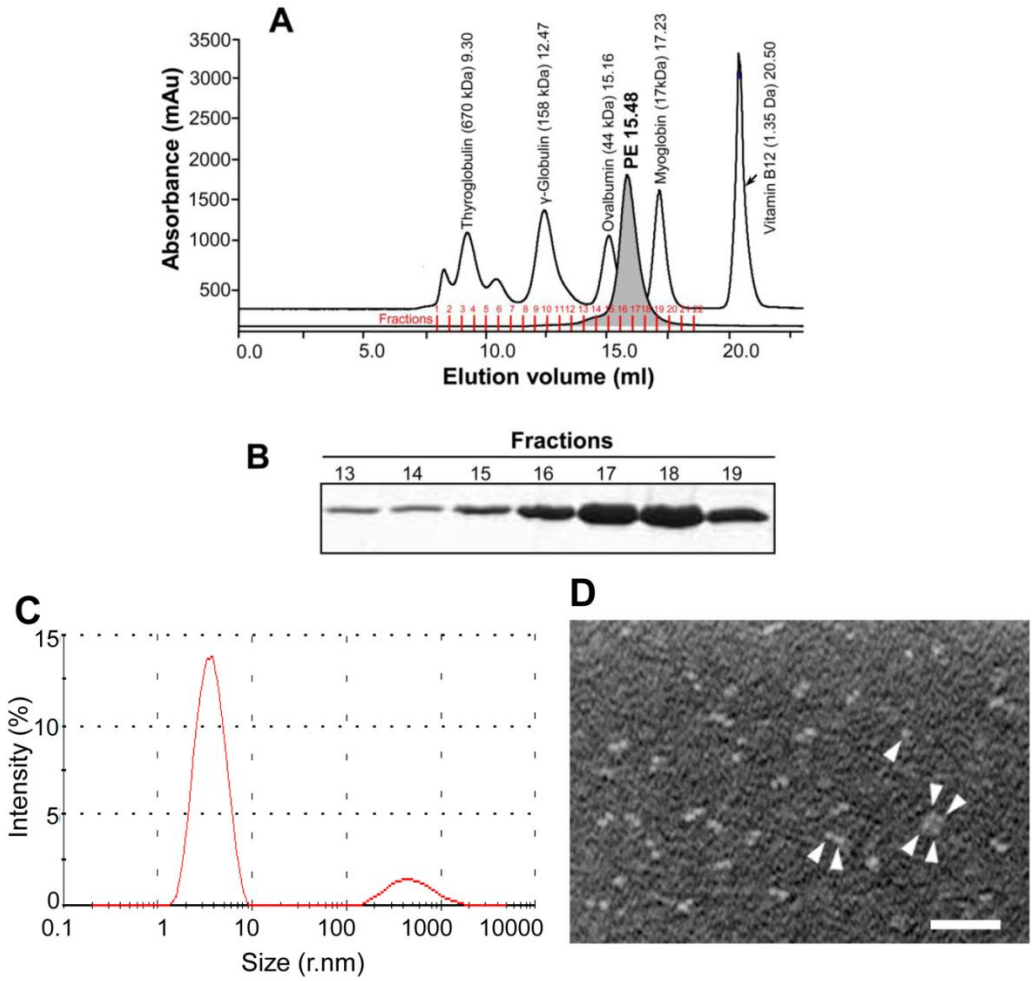

A
Absorbance (mAu)
3500
3000
2500
2000
1500
1000
500
Thyroglobulin (670 kDa) 9.30
γ-Globulin (158 kDa) 12.47
Ovalbumin (44 kDa) 15.16
PE 15.48
Myoglobin (17kDa) 17.23
Vitamin B12 (1.35 Da) 20.50
Fractions
0.0
5.0
10.0
15.0
20.0
Elution volume (ml)
B
Fractions
13
14
15
16
17
18
19
C
Intensity (%)
15
10
5
0
0.1
1
10
100
1000
10000
Size (r.nm)
D

Figure 2

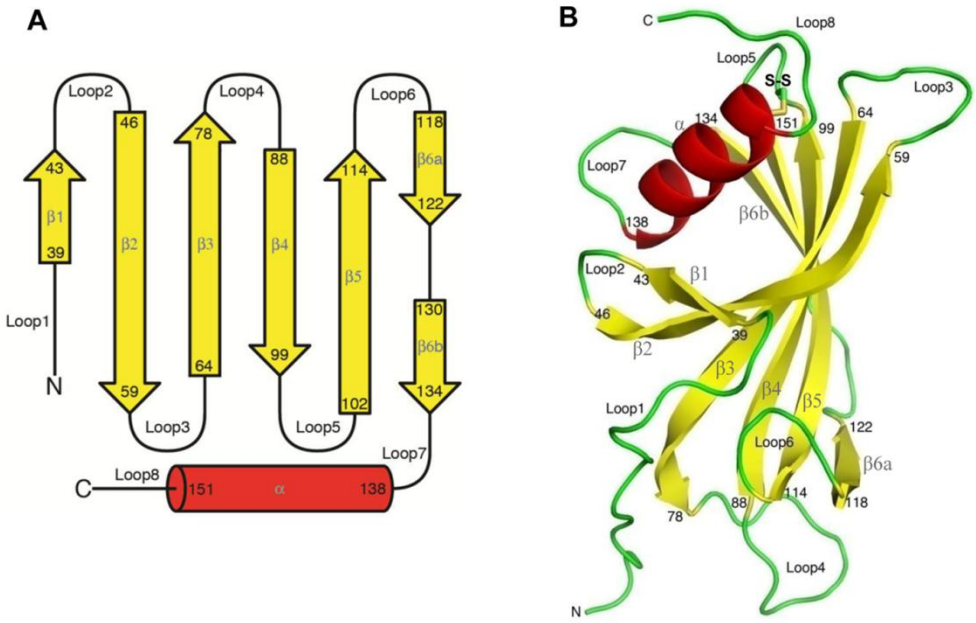


Figure 3

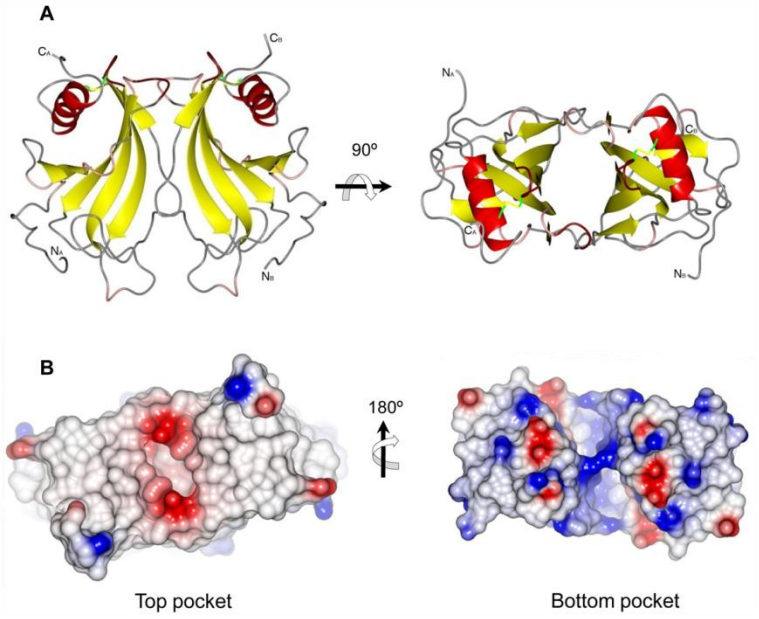


Figure 4

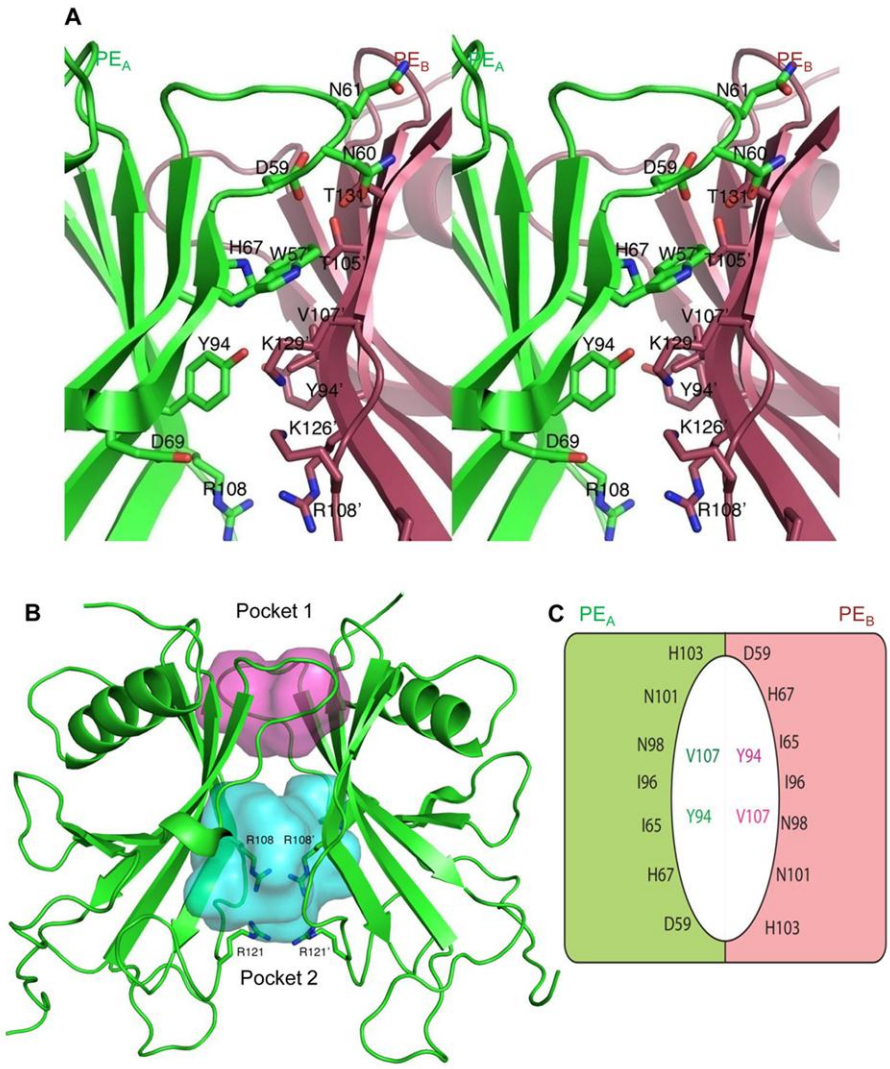

A
PE_A
PE_B
N61
D59
N60
T131'
H67
W57
P105'
V107'
Y94
K129'
Y94'
K126'
D69
R108
R108'
B
Pocket 1
R108
R108'
R121
R121'
Pocket 2
C
PE_A
PE_B
H103
D59
N101
H67
N98
V107
Y94
I65
I96
I96
Y94
V107
I65
N98
H67
N101
D59
H103

Figure 5

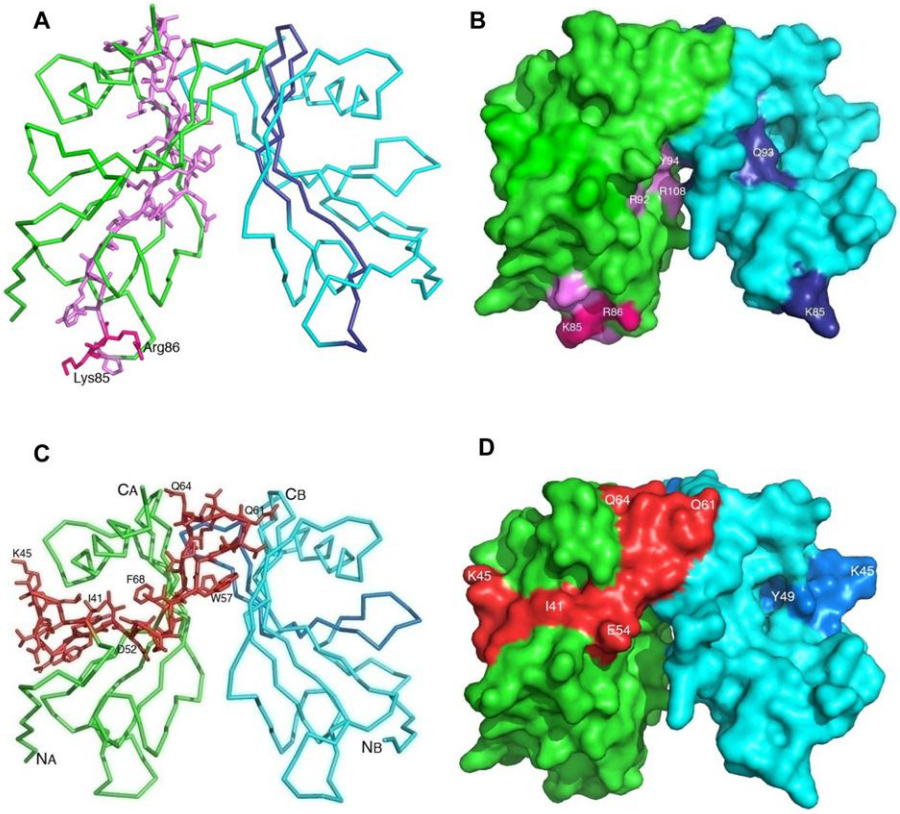

A
Arg86
Lys85
B
Q93
Y94
R108
R92
R86
K85
K85
C
CA
Q64
Q61
CB
K45
F68
I41
W57
D52
NA
NB
D
Q64
Q61
K45
K45
I41
Y49
E54

Figure 6

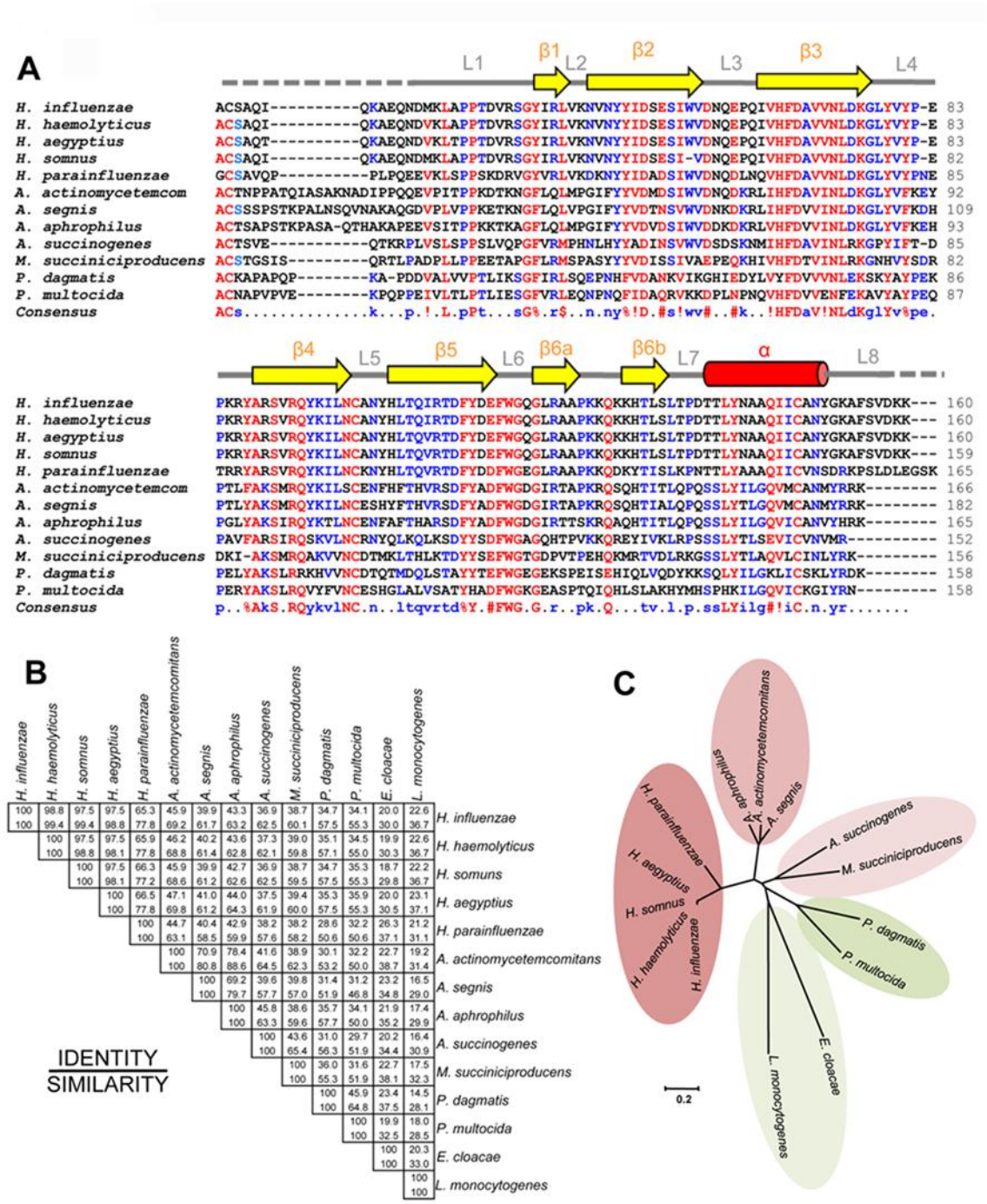

A
L1
β1
L2
β2
L3
β3
L4
H. influenzae ACSAQI----------QKAEQNDMKLAPPTDVRSGYIRLVKNVNYYIDSESIWVDNQEPQIVHFDAVVNLDKGLYVYP-E 83
H. haemolyticus ACSAQI----------QKAEQNDVKLAPPTDVRSGYIRLVKNVNYYIDSESIWVDNQEPQIVHFDAVVNLDKGLYVYP-E 83
H. aegyptius ACSAQT----------QKAEQNDVKLTPPTDVRSGYIRLVKNVNYYIDSESIWVDNQEPQIVHFDAVVNLDKGLYVYP-E 83
H. somnus ACSAQI----------QKAEQNDMKLAPPTDVRSGYIRLVKNVNYYIDSESI-VDNQEPQIVHFDAVVNLDKGLYVYP-E 82
H. parainfluenzae GCSAVQP----------PLPQEEVKLSPPSKDRVGYVRLVKDKNYYIDADSIWVDNQDLNQVHFDAVVNLDKGLYVYPNE 85
A. actinomycetemcom ACTNPPATQIASAKNADIPPQQEVPITPPKDTKNGFLQLMPGIFYYVDMDSIWVDNQDKRLIHFDAVINLDKGLYVFKEY 92
A. segnis ACSSSPSTKPALNSQVNAKAQGDVPLVPPKETKNGFLQLVPGIFYYVDTNSVWVDNKDKRLIHFDVVINLDKGLYVFKDH 109
A. aphrophilus ACTSAPSTKPASA-QTHAKAPEEVSITPPKKTKAGFLQLMPGIFYYVDVDSIWVDDKDKRLVHFDVVINLDKGLYVFKEH 93
A. succinogenes ACTSVE----------QTKRPLVSLSPPSLVQPGFVRMPHNLHYYADINSVWVDSDSKNMIHFDAVINLRKGPYIFT-D 85
M. succiniciproducens ACSTGSIS---------QRTLPADPLLPPEETAPGFLRMSPASYYYVDISSIVAEPEQKHIVHFDTVINLRKGNHVYSDR 82
P. dagmatis ACKAPAPQP--------KA-PDDVALVVPTLIKSGFIRLSQEPNHFVDANKVIKGHIEDYLVYFDVVVNLEKSKYAYPEK 86
P. multocida ACNAPVPVE--------KPQPPEIVLTLPTLIESGFVRLEQNPNQFIDAQRVKKDPLNPNQVHFDVVENFEKAVYAYPEQ 87
Consensus ACs.....k...p.!.L.pPt...sG%.r$..n.ny%!D.#s!wv#..#k..!HFDaV!NLdKglYv%pe.
β4
L5
β5
L6
β6a
β6b
L7
α
L8
H. influenzae PKRYARSVRQYKILNCANYHLTQIRTDFYDEFWGQGLRAAPKKQKKHTLSLTPDTTLYNAAQIICANYGKAFSVDKK--- 160
H. haemolyticus PKRYARSVRQYKILNCANYHLTQIRTDFYDEFWGQGLRAAPKKQKKHTLSLTPDTTLYNAAQIICANYGKAFSVDKK--- 160
H. aegyptius PKRYARSVRQYKILNCANYHLTQVRTDFYDEFWGQGLRAAPKKQKKHTLSLTPDTTLYNAAQIICANYGKAFSVDKK--- 160
H. somnus PKRYARSVRQYKILNCANYHLTQVRTDFYDEFWGQGLRAAPKKQKKHTLSLTPDTTLYNAAQIICANYGKAFSVDKK--- 159
H. parainfluenzae TRRYARSVRQYKILNCKNYHLTQVRTDFYDDFWGEGLRAAPKKQDKYTISLKPNTTLYAAAQIICVNSDRKPSLDLEGSK 165
A. actinomycetemcom PTLFAKSMRQYKILSCENFHFTHVRSDFYADFWGDGIRTAPKRQSQHTITLQPQSSLYILGQVMCANMYRRK-------- 166
A. segnis PTLYAKSMRQYKILNCESHYFTHVRSDFYADFWGDGIRTAPKRQSQHTIALQPQSSLYTLGQVMCANMYRRK-------- 182
A. aphrophilus PGLYAKSIRQYKTLNCENFAFTHARSDFYADFWGDGIRTTSKRQAQHTITLQPQSSLYILGQVICANVYHRK-------- 165
A. succinogenes PAVFARSIRQSKVLNCRNYQLKQLKSDYYSDFWGAGQHTPVKKQREYIVKLRPSSSLYTLSEVICVNVMR---------- 152
M. succiniciproducens DKI-AKSMRQAKVVNCDTMKLTHLKTDYYSEFWGTGDPVTPEHQKMRTVDLRKGSSLYTLAQVLCINLYRK--------- 156
P. dagmatis PELYAKSLRRKHVVNCDTQTMDQLSTAYYTEFWGEGEKSPEISEHIQLVQDYKKSQLYILGKLICSKLYRDK-------- 158
P. multocida PERYAKSLRQVYFVNCESHGLALVSATYHADFWGKGEASPTQIQHLSLAKHYMHSPHKILGQVICKGIYRN--------- 158
Consensus p..%AkS.RQykvlNC.n..ltqvrtd%Y.#FWG.G.r..pk.Q...tv.l.p.ssLYilg#!iC.n.yr.....
B
H. influenzae
H. haemolyticus
H. somnus
H. aegyptius
H. parainfluenzae
A. actinomycetemcomitans
A. segnis
A. aphrophilus
A. succinogenes
M. succiniciproducens
P. dagmatis
P. multocida
E. cloacae
L. monocytogenes
100 98.8 97.5 97.5 65.3 45.9 39.9 43.3 36.9 38.7 34.7 34.1 20.0 22.6 H. influenzae
100 99.4 99.4 98.8 77.8 69.2 61.7 63.2 62.5 60.1 57.5 55.3 30.0 36.7
100 97.5 97.5 65.9 46.2 40.2 43.6 37.3 39.0 35.1 34.5 19.9 22.6 H. haemolyticus
100 98.8 98.1 77.8 68.8 61.4 62.8 62.1 59.8 57.1 55.0 30.3 36.7
100 97.5 66.3 45.9 39.9 42.7 36.9 38.7 34.7 35.3 18.7 22.2 H. somuns
100 98.1 77.2 68.6 61.2 62.6 62.5 59.5 57.5 55.3 29.8 36.7
100 66.5 47.1 41.0 44.0 37.5 39.4 35.3 35.9 20.0 23.1 H. aegyptius
100 77.8 69.8 61.2 64.3 61.9 60.0 57.5 55.3 30.5 37.1
100 44.7 40.4 42.9 38.2 38.2 28.6 32.2 26.3 21.2 H. parainfluenzae
100 63.1 58.5 59.9 57.6 58.2 50.6 50.6 37.1 31.1
100 70.9 78.4 41.6 38.9 30.1 32.2 22.7 19.2 A. actinomycetemcomitans
100 80.8 88.6 64.5 62.3 53.2 50.0 38.7 31.4
100 69.2 39.6 39.8 31.4 31.2 23.2 16.5 A. segnis
100 79.7 57.7 57.0 51.9 46.8 34.8 29.0
100 45.8 38.6 35.7 34.1 21.9 17.4 A. aphrophilus
100 63.3 59.6 57.7 50.0 35.2 29.9
100 43.6 31.0 29.7 20.2 16.4 A. succinogenes
100 65.4 56.3 51.9 34.4 30.9
100 36.0 31.6 22.7 17.5 M. succiniciproducens
100 55.3 51.9 38.1 32.3
100 45.9 23.4 14.5 P. dagmatis
100 64.8 37.5 28.1
100 19.9 18.0 P. multocida
100 32.5 28.5
100 20.3 E. cloacae
100 33.0
100 L. monocytogenes
IDENTITY
SIMILARITY
C
H. parainfluenzae
H. aegyptius
H. somnus
H. haemolyticus
H. influenzae
aphrophilus
A. actinomycetemcomitans
A. segnis
A. succinogenes
M. succiniciproducens
P. dagmatis
P. multocida
E. cloacae
L. monocytogenes
0.2

Figure 7

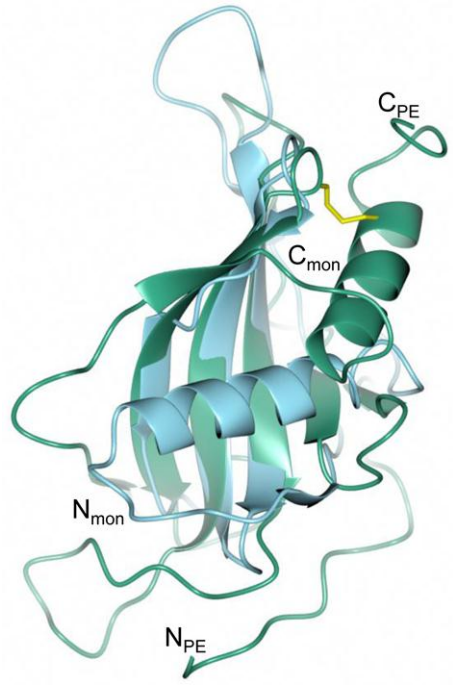


Figure 8

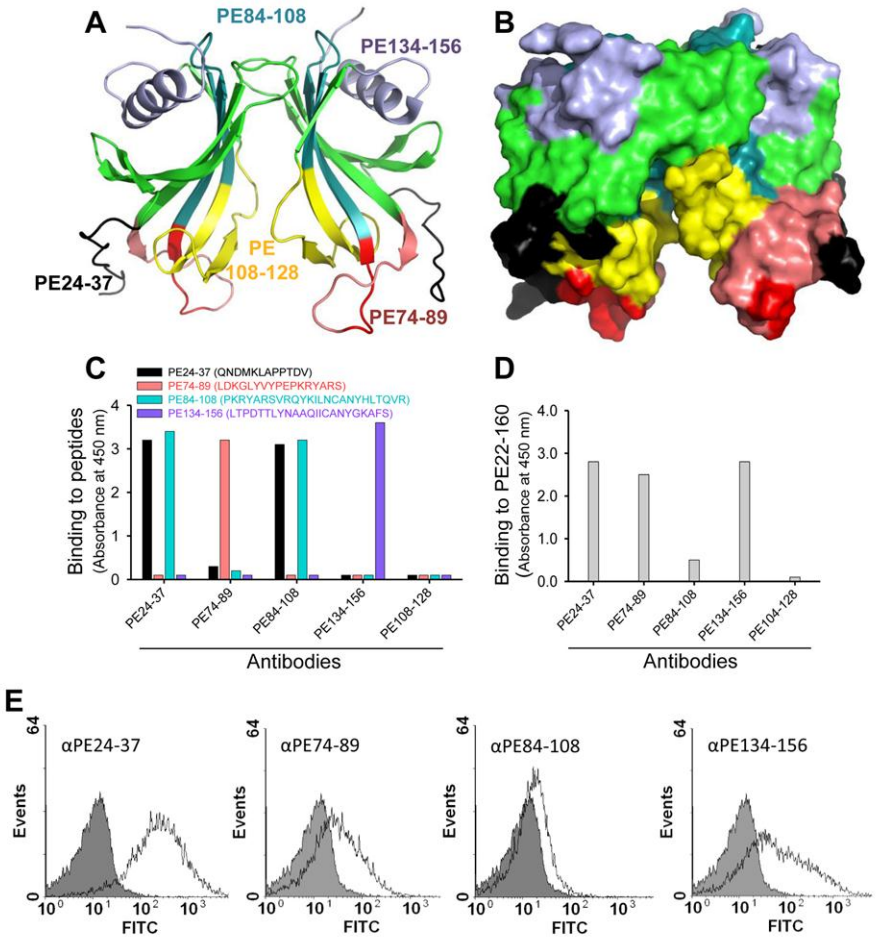
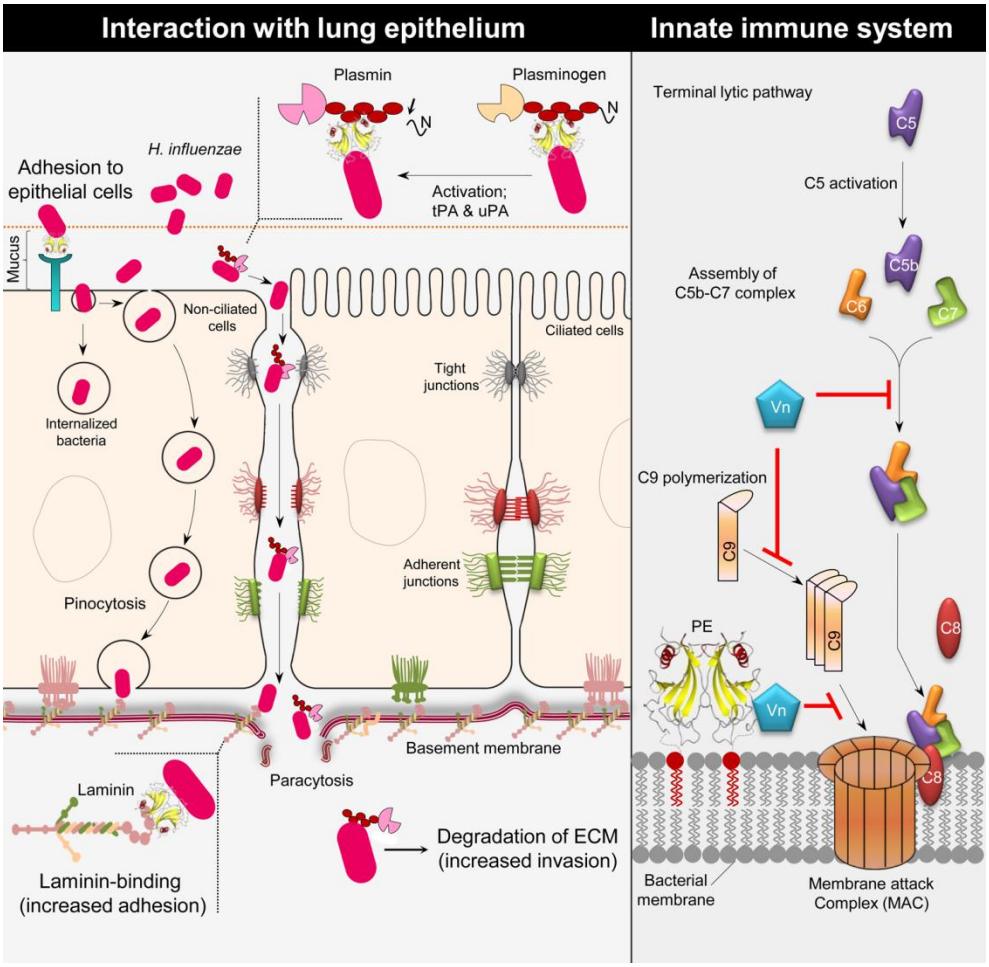

A
PE84-108
PE134-156
PE
108-128
PE24-37
PE74-89
B
C
PE24-37 (QNDMKLAPPTDV)
PE74-89 (LDKGLYVYPEPKRYARS)
PE84-108 (PKRYARSVRQYKILNCANYHLTQVR)
PE134-156 (LTPDTTLYNAAQIICANYGKAFS)
Binding to peptides
(Absorbance at 450 nm)
PE24-37
PE74-89
PE84-108
PE134-156
PE108-128
Antibodies
D
Binding to PE22-160
(Absorbance at 450 nm)
PE24-37
PE74-89
PE84-108
PE134-156
PE104-128
Antibodies
E
αPE24-37
αPE74-89
αPE84-108
αPE134-156
Events
FITC

Figure 9


Interaction with lung epithelium
Innate immune system
Plasmin
Plasminogen
N
N
Activation;
tPA & uPA
H. influenzae
Adhesion to
epithelial cells
Mucus
Non-ciliated
cells
Ciliated cells
Tight
junctions
Internalized
bacteria
Adherent
junctions
Pinocytosis
Basement membrane
Paracytosis
Laminin
Laminin-binding
(increased adhesion)
Degradation of ECM
(increased invasion)
Terminal lytic pathway
C5
C5 activation
C5b
Assembly of
C5b-C7 complex
C6
C7
Vn
C9 polymerization
C9
C9
C8
PE
Vn
C8
Bacterial
membrane
Membrane attack
Complex (MAC)