BACTERIAL CHARACTERISATION AND ULTRASTRUCTURE

Secondary structure and molecular analysis of interstrain variability in the P5 outer-membrane protein of non-typable *Haemophilus influenzae* isolated from diverse anatomical sites

DIANNE C. WEBB*† and A. W. CRIPPS*

*Research Centre for Human and Biomedical Science, Faculty of Applied Science, University of Canberra, ACT 2601, Australia and †Membrane Biochemistry Group, Division of Biochemistry and Molecular Biology, John Curtin School of Medical Research, Australian National University, PO Box 334, Canberra City, ACT 2601, Australia

The sequence of the non-typable *Haemophilus influenzae* (NTHi) P5 outer-membrane protein from a range of clinical isolates is presented and represents the first analysis of the heterogeneity in P5 from NTHi isolates from diverse anatomical sites. The basis of the previously observed inter-strain variation in the electrophoretic mobility is attributed to heterogeneity in three hypervariable regions. Alignment of the P5 sequences identified regions which are highly conserved and align with the transmembrane region predicted for the homologous *Escherichia coli* protein, OmpA. Variable regions correspond to surface-exposed loops, of which the first loop falls into subclasses. However, these subclasses fail to correlate with anatomical predisposition. Although P5 has been proposed as a fimbrial protein composed of coiled coils, both structural analysis by circular dichroism of purified P5 and computer analysis of the multiply aligned sequences predict a high proportion of β strand with no evidence of coiled coil structure. A detailed model of P5 is presented.

Introduction

While non-typable *Haemophilus influenzae* (NTHi) forms a minor component of the oropharyngeal microbiota [1], it is also an opportunist pathogen commonly associated with chronic bronchitis, chronic obstructive pulmonary disease (COPD), pneumonia and otitis media [2], as well as with sinusitis, conjunctivitis, infections of the female genital tract and neonatal sepsis [3]. Chromosome fingerprint analysis of NTHi isolates from patients with COPD demonstrated genetic diversity [4]. However, analysis of genetic types by multi-locus enzyme electrophoresis suggested that isolates from persistent carriage or chronic infections are more diverse than isolates from acute infections [5]. The major outer-membrane protein (OMP) profile of chronic bronchitic isolates varies in the electrophoretic mobility of proteins P2 (protein b, c) and P5 (protein d) [4]. Furthermore, sequence analysis of the heat-modifiable protein, P5, from five isolates and their

variants from chronic bronchitics, has identified four hypervariable regions that are proposed to be surface-exposed and subject to the immune pressure of the host [6]. How these regions vary amongst isolates from acute infections and amongst isolates from different anatomical sites is not known. Immunisation with P5 confers partial protection against transtubular challenge with the homologous strain but not with heterologous strains in the chinchilla otitis model [7], a result that also indicates that the immunodominant, surface-exposed regions are heterogeneous.

Despite this heterogeneity, P5 may play a role in NTHi pathogenesis as an adhesin by binding to respiratory mucin [8, 9]. Presumably, this binding is mediated by the surface-exposed domains of P5, which may need to be selectively conserved to enhance pathogenicity, compared with non-essential surface domains. There are conflicting reports regarding the structure of P5. Sirakova *et al.* [7] report a structure composed of coiled coils which assemble into fimbriae. However, this fimbrial structure is at odds with the high identity (50%) of P5 to *Escherichia coli* OmpA [10], an eight-stranded β -barrel-forming protein, consistent with the four

Received 31 Oct. 1997; revised version accepted 9 March 1998.

Corresponding author: D. C. Webb.

hypervariable surface-exposed regions of P5 identified in the chronic bronchitis isolates [6].

The purposes of this study were two-fold. Firstly, in order to compare P5 heterogeneity from both acute infections and infections from diverse anatomical sites, the complete *p5* gene from 13 clinical isolates was sequenced, as well as the DNA encoding the first variable region from a further 10 isolates. Secondly, to resolve the controversy regarding the P5 structure, circular dichroic (CD) analysis was performed on purified P5, as well as computer analyses for secondary structure, including the presence of coiled coil domains. Detailed analysis of the basis of the homology of the N-terminal P5 transmembrane region to OmpA, and the highly conserved nature of the C-terminus amongst these isolates supports a β -barrel structure with a periplasmic tail with no evidence for the presence of coiled coils. A model of the P5 N-terminus is proposed.

Materials and methods

Bacterial strains

Most NTHi isolates were obtained from clinical samples by Barratt and Smith Pathology, ACT, Australia. Two isolates (UC19 and 103) were from our laboratory collection and were originally derived from sputum samples of patients with chronic bronchitis. Four isolates (UC5, 9, 14, 16) were obtained from throat swabs of healthy child-care workers and are regarded as commensals. This panel of strains (Table 1), which was chosen to represent a diversity of both biotypes and anatomical predisposition, was screened for absence of agglutination with anti-b capsule antiserum (Murex Diagnostics, Temple Hill, Dartford) and the requirement for growth factors X (haemin) and V (nicotinamide adenine dinucleotide, NAD) with Microring XV impregnated disks (Microdiagnostics, Brisbane, Qld, Australia) on Brain Heart Infusion (BHI) Agar (Oxoid, West Heidelberg, Victoria, Australia). Isolates were biotyped by the APINH identification system for *Neisseria* and *Haemophilus* (bioMérieux, Marcy-l'Etoile, France).

Preparation of chromosomal DNA and sequence analysis of the *p5* gene

NTHi strains were cultured overnight at 37°C in 3 ml of BHI broth supplemented (sBHI) with NAD and

haemin, each at 10 μ g/ml (both from Sigma). Cells were pelleted in a microfuge, washed in phosphate-buffered saline (PBS: 140 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, pH 7.3), resuspended in 400 μ l of buffer containing 50 mM glucose, 10 mM EDTA and 25 mM Tris, pH 8.0, then lysed in 0.1 M NaOH, SDS 0.5%. The lysate was digested at 37°C with heat-treated RNAase 1.25 mg/ml for 15 min and then proteinase K (Sigma) 2.5 mg/ml for 30 min. The digest was phenol extracted, ethanol precipitated, and the dried pellet was resuspended in 500 μ l of H₂O. Oligonucleotides (Gibco BRL, Vic, Australia) used for PCR amplification of the *p5* genes were based on the fimbrin P5-homologue [7], except for the inclusion of the restriction endonuclease sites (underlined) for *Eco*R1 in the forward primer, 96–16 (5'-CTATTCGAT-GACGAATTCGAGGACATCAAAATG-3') and *Pst*I in the reverse primer, 96–17 (5'-CCTTTTCCTGTACTG-CAGCTTTCGTAAAC-3'). The last triplet in 96–16 encodes the methionine P5 start site and 96–17 starts 8 bp after the stop codon. The PCR reaction contained chromosomal DNA 50 ng, 0.1 mM dNTPs (Boehringer Mannheim Pty Ltd, Castle Hill, NSW, Australia), 20 pmol of each primer, 2 mM MgCl₂, and reaction buffer (10 mM KCl, 6 mM (NH₄)₂SO₄, 20 mM Tris-HCl, pH 8.0, 2 mM MgCl₂, Triton X-100 0.1%, BSA 10 mg/ml) in a total volume of 50 μ l. Native *Pfu* DNA polymerase (Stratagene, Willoughby, NSW, Australia) was added after 5 min at 94°C. This was followed by 30 cycles of 94°C for 45 s, 56°C for 1 min and 72°C for 3 min in a thermal cycler (Perkin-Elmer, Mulgrave, Victoria, Australia). Unincorporated nucleotides were removed from the reaction mix with Bresaclean resin (Bresatec, Adelaide, SA, Australia). The vector M13mp19 [11], and the PCR products were digested with *Pst*I and *Eco*R1. After removing the digested ends with Bresaclean resin, the *p5* PCR products were ligated into M13mp19. *E. coli* strain TG1 (K12 Δ (*lac-pro*), *supE*, *thi*, *hsdD5*, /*F'* *traD36*, *proAB*, *lacIq*, *lacZ*, Δ M15) was used as a recipient for blue/white screening of recombinants and the presence of an insert in M13 was confirmed by restriction analysis. The *p5* genes were sequenced from single-stranded DNA prepared from phage recovered from the supernate of recombinant cultures with the Wizard purification kit (Promega, Annandale, NSW, Australia).

All basic recombinant DNA methods used were as described by Sambrook *et al.* [12]. Nucleotide sequence analysis was performed by the dideoxy chain termination method of Sanger *et al.* [13] with the PRISM ready reaction Dye-Deoxy terminator cycle sequencing and an Applied Biosystems automated sequencing system, Model 373A (Applied Biosystems, Perkin Elmer, Mulgrave, Vic, Australia). Three primers were used for sequencing the *p5* gene: the M13 –40 primer (5'-GTTTTCCAGTCACGAC-3') and two *p5* internal primers (Gibco BRL, Vic, Australia): 96–21 (5'-GCTATGAAGTGTTAGACG-GTTTAG-3') and 96–19 (5'-GATGTAACTTTCG-

Table 1. NTHi isolates used for P5 characterisation

NTHi isolation site	UC strain number
Ear	28, 32, 63, 77, 78, 84
Sinus	36
Sputum	1, 19, 27, 33, 35, 47, 103
Conjunctiva	2, 22, 29
Vagina	3, 26
Throat	5, 9, 14, 16

CATTGG-3'). Only the M13 primer was used for sequencing the first variable region. Sequencing was performed on at least two independent clones per strain, and assembled with the MacVector program (Apple Macintosh). The predicted protein sequences were aligned with the Genetics Computer Group PileUp program [14] from the University of Wisconsin accessed through the Australian National Genomic Information Service. Secondary structure predictions were performed with the SSPRED [15] and the COILS computer programs [16].

Preparation of NTHi outer-membrane component

The procedure for isolating a component enriched in OMPs was essentially as described by Carlone *et al.* with modifications [17]. Bacteria were grown on chocolate agar plates (BHI with chocolate horse blood 5%), resuspended in 10 ml of cold 10 mM HEPES (*N*-2 hydroxyethyl piperazine-*N'*-2 ethansulphonic acid) buffer, pH 7.4, per plate and pelleted by centrifugation at 5000 *g* for 10 min. After resuspension in 1 ml of HEPES buffer, bacteria were sonicated on ice with six bursts of 10 s each at power setting 8 (Ystrom Systems Sonicator, USA). Cellular debris was removed by centrifugation at 15 000 *g* for 2 min at 4°C. The total cell membrane was pelleted by centrifugation at 15 000 *g* for 30 min and the OMPs were collected by centrifugation, as above, after solubilisation of the cytoplasmic membrane with 800 µl of sodium-*N*-lauroyl sarcosinate (Fluka Chemicals, Castle Hill, NSW, Australia) 1% in HEPES buffer for 30 min at room temperature with intermittent mixing. The pellet was resuspended in 200 µl of HEPES buffer.

SDS-PAGE

Samples for protein analysis were either boiled or heated at 50°C for 30 min (to determine the heat modifiability) in sample buffer (62.5 mM Tris-HCl, pH 6.8, glycerol 10%, β-mercaptoethanol 5%, SDS 2% and bromophenol blue 0.001%). Proteins were separated electrophoretically by denaturing Tris-glycine 4–20% SDS-PAGE of pre-poured gels (Novex Australia, Terrey Hills, NSW, Australia), with TBE buffer (89 mM Tris base, pH 8.3, 89 mM boric acid and 2 mM EDTA) and a constant voltage of 125 V. Proteins were detected with colloidal Coomassie G-250 (Novex) and Mark 12 (Novex) wide range mol. wt standards were used.

Purification of P5

P5 was purified essentially as described previously [18]. NTHi strain UC19 was harvested from 100 chocolate agar plates, washed and resuspended in 10 mM HEPES buffer, pH 7.4. A component enriched for OMPs was prepared as above from which a component enriched in P5 was prepared by differential detergent solubilisation. P5 was purified by hydroxy-

apatite chromatography with a 5-ml Econo-Pac CHT-II column (BioRad Laboratories, Regents Park, NSW, Australia) and a 40–600 mM NaPi gradient, pH 7.7, containing SDS 0.05%.

Circular dichroism analysis

Before CD analysis P5 was dialysed against PBS containing SDS 0.05%. The CD spectrum from 195 to 240 nm at 1-nm intervals with an integration time of 4 s/nm was recorded for 0.5 ml of 10⁻⁶ M solution of UC19 P5 at room temperature with a Jobin Yvon CD6 Dichrograph (Division d'Instruments SA, Cedex, France) and a 0.1-cm cell. Twenty spectra were averaged and corrected for buffer background. Analysis of the secondary structure was performed with the CONTIN program [19].

Results

Genetic variation in P5 among non-typable H. influenzae clinical isolates

The molecular basis of the inter-strain polymorphism of P5 amongst 13 NTHi clinical isolates was determined by comparison of the amino-acid sequence of the proteins encoded by the corresponding *p5* genes, which were amplified by PCR and sequenced. These sequences were also compared with the previously published sequences of the fimbrin P5-homologue, which was derived from an NTHi strain isolated from a middle ear effusion [7], an *H. influenzae* type b (Hib) strain [10], an *H. influenzae* Rd strain [20] and *E. coli* OmpA [21] (Fig. 1). Sequences were aligned by the PileUp program with strain UC19 selected as the reference strain and amino acids numbered from the first residue of the mature protein. Most variation amongst isolates occurred in three regions – residues 22–39, 70–84 and 125–137 – while the fourth region, previously identified in chronic bronchitic isolates as variable, was conserved in all isolates except the more diverse conjunctival isolate, strain UC2.

Interestingly, among these 16 isolates, the first variable region fell into three types: a longer region encompassing the motif GINNNGAIK was found in P5 from four of the six respiratory isolates (UC1, UC19, UC35 and UC36), both ear isolates (UC28 and UC32) and the 'fimbrin' protein [7]. The other two types, which are both shorter, less polymorphic regions, encompass either the motif GLRALARE or GVRAMGKQ, and were found in the conjunctival isolates (strains UC2, UC22 and UC29), the respiratory isolates UC27 and UC103, and both the Hib and Rd strains.

To determine whether ear and respiratory isolates, in general, had a higher probability of the GINNNGAIK motif, the region of the *p5* genes encoding this region was sequenced from a further 10 isolates, which

Fig. 1. Amino-acid alignment of the P5 proteins. The P5 amino-acid sequences from the NTHi clinical isolates which were analysed as part of this study, as well as the NTHi fimbrin P5-homologue (fim), the b-capsulate strain (Hib), the Rd strain and *E. coli* OmpA were aligned by the PileUp program (GCG). The proposed transmembrane regions of the β -barrel are shaded and the proposed external loops are marked above the sequences by double-ended arrows. The secondary structure predictions (SSPRED) are below the bold line under the alignment where E represents regions of β -strand and H represents helical regions which are boxed and numbered.

included four ear, two sputum, as well as four commensal isolates. The corresponding P5 variable region one motifs from all the isolates sequenced in this study and their corresponding sites of isolation are shown in Table 2. Analysis of this expanded

number of isolates failed to demonstrate a clear correlation between anatomical predisposition and the motif in the first variable region. Two of the commensal isolates also contained the previously unseen motifs, GIRSNAQ and GIREHAP.

Table 2. Comparison of NTHi isolation site with the motif found in P5 surface loop one

GINNNGAIK		GLRALARE		GVRAMGKQ	
Strain	Source	Strain	Source	Strain	Source
UC1	Sputum	UC22	Conjunctival	UC2	Conjunctival
UC3	Vagina	UC26	Vagina	UC77	Ear
UC5	Throat	UC27	Sputum	UC78	Ear
UC16	Throat	UC29	Conjunctival		
UC19	Sputum	UC33	Sputum		
UC28	Ear	UC47	Sputum		
UC32	Ear	UC63	Ear		
UC35	Sputum	UC84	Ear		
UC36	Sinus	UC103	Sputum		

Two throat commensal isolates, UC9 and UC14, contain the unique motifs GIRSNAQ and GIREHAP and are not included in the table.

Secondary structure of P5

Circular dichroism analysis. As the structure of P5 is controversial, circular dichroism was used to determine the secondary structure of P5 purified from strain UC19. P5 is a heat-modifiable protein which will only migrate in SDS-PAGE with an electrophoretic mobility comparable with its true mol. wt (37 kDa) if it has been extensively boiled in sample buffer containing β -mercaptoethanol 10%. If P5 is treated in sample buffer at lower temperatures, it retains a folded structure that migrates faster in SDS-PAGE than the denatured form [18]. SDS-PAGE of the purified P5 showed the presence of a protein with the characteristic electrophoretic mobility also seen in P5 in the sarkosyl-insoluble OMP component, indicating that the purified P5 was correctly folded (Fig. 2). The CD spectrum of the purified P5 exhibited a minimum negative peak at

213 nm (Fig. 3) and corresponds to a β -strand content of 49%, similar to the recent CD analysis of OmpA which demonstrated 40% β -strand content [22].

Coiled coil analysis. Previous analysis of the fimbrin-P5 homologue with the COILS program [16] has led to the suggestion that this protein is composed of coiled coil domains that assemble into a fimbrial structure analogous to streptococcal M protein [7]. Analysis of strain UC19 P5 with the COILS program (Fig. 4) failed to detect any areas of coiled coil formation. A similar result was seen for the fimbrin P5-homologue (data not shown), which is 94% identical to strain UC19 P5. The predictive capacity of the program for the coiled coil structures in a tetravalent hybrid molecule composed of the N-terminal domain of streptococcal M protein from serotypes 24, 5, 6 and 19 M [23] was used as a comparison.

Fig. 2. SDS-PAGE of purified strain UC19 P5. Samples were either boiled (lanes 2 and 4), or heated to 50°C (lanes 3 and 5) for 30 min to visualise the characteristic heat modifiability of P5. Lane 1, mol. wt markers; 2 and 3, P5 purified from NTHi strain UC19; 4 and 5, sarkosyl-insoluble OMPs from UC19. Sizes of markers (kDa) are indicated on the left.

Secondary structure prediction. The SSPRED computer prediction method uses information in multiple sequence alignments of substituted but structurally related proteins to predict α -helix, β -strand and coil states of secondary structure with a mean accuracy of 72.2% [15]. This program was applied to the multiple alignment of 16 P5 sequences (the 13 fully sequenced P5 proteins from this study and the ‘fimbrin’, Hib and Rd P5 proteins). As seen in Fig. 1, seven β -strand regions were predicted in the 200 amino acid N-terminus. Three α -helical regions were predicted for the C-terminus, of which one (helical region two) corresponds to a peptidoglycan binding domain [24]. No coiled domains were predicted by this program.

Proposed structure of P5 based on the OmpA structure

A model of the N-terminus of P5 is proposed in Fig. 5 and is based on the alignment of the P5 amino-acid sequences with OmpA and the OmpA model as proposed by Vogel and Jähnig [25]. In this model, the variable regions form the surface-exposed loops of the eight-stranded β -barrel structure. The amino acids in the remaining C-terminus of the protein are highly

Fig. 3. CD spectrum of strain UC19 P5. This is the average of 20 spectra which were measured in PBS containing SDS 0.05% and corrected for buffer background. The units of mean residue ellipticity are degrees cm² dmol⁻¹.

Fig. 4. Comparison of the coiled coil analysis of strain UC19 P5 and a tetravalent hybrid molecule composed of the N-terminal domain of streptococcal M protein from four serotypes. Analysis was performed with the COILS program and values above zero indicate the probability of a coiled coil structure.

Fig. 5. The proposed model of the 200 residue N-terminal of P5 from strain UC19. The transmembrane charged residues are either circled or in diamonds (indicating positive and negative charge) and are identical to the charged residues found in the transmembrane domain in OmpA except for P5 Lys-111 which in OmpA is equivalent to Arg-97. The glycine residues equivalent to OmpA residues are underlined (mostly found on the hydrophilic face), and the residues that form the aromatic rings (see text) are numbered. The transmembrane region, composed of β -strands, retains an amphipathic nature and the hydrophobic face is in bold. The hypervariable surface exposed loops are numbered and the motif in loop one is shaded.

conserved, with 94% identity amongst the 16 NTHi isolates.

Discussion

Previous analysis of P5 from chronic bronchitic NTHi isolates identified four hypervariable regions which were likely to be surface exposed and thus subject to the immune pressure of the host [6]. As it is suggested that NTHi isolates from acute infections are more diverse than either carrier or chronic respiratory infection isolates [5], it was of interest to determine if the heterogeneity in P5 from various anatomical sites represented the same diversity. The hypervariable regions in this analysis occur predominantly in three regions, which correspond to the first three surface-exposed loops, similar to the hypervariable regions identified by Duim *et al.* [6]. In contrast, the fourth hypervariable region identified in the Duim study is

conserved in all except one of the isolates in the present study. Hypervariability in this region may be caused by the more prolonged immune pressure experienced during chronic infections more typically found in chronic bronchitis. However, as a longitudinal study of the patients from whom the NTHi isolates were derived in this study was not performed, it is difficult to draw any conclusions regarding the persistence of infection and hypervariability in loop four.

The existence of subclasses of loop one is intriguing. The shorter loop containing either the GLRALARE motif or the variant GVRAMGKQ is less heterologous than the longer loop containing the GINNNGAIK motif and it is tempting to speculate that, unlike the longer loop, the shorter loop is less exposed and less subject to the immune pressure of the host. Both variants of the shorter loop contain two positively charged residues (lysine and arginine) interceded with three non-polar residues, suggesting some functional conservation.

Transversal of *E. coli* across the blood-brain barrier is attributed to OmpA-mediated attachment to brain microvascular endothelial cells (BMEC) by interaction with the GlcNAc β 1-4GlcNAc moiety of BMEC glycoproteins [26]. The importance of the first two surface-exposed OmpA loops is demonstrated by the inhibition of this attachment by synthetic peptides which correspond to these loops [27]. While the pathogenic mechanisms of *E. coli* in neonatal meningitis are likely to be different from those proposed for the usually non-invasive NTHi, a role for P5-mediated attachment to host structures is suggested by the observation that the sialic acid-containing oligosaccharides of respiratory mucin bind to P5 [8, 9]. Presumably, this attachment must be mediated by surface-exposed domains. However, the contribution that the P5 loop-one motifs or other surface loops make to the pathogenicity of NTHi can only be speculative at this stage.

Previous molecular analysis of P5 has suggested that it is composed of coiled coil domains which assemble into a fimbrial structure [7]. However, the homology of P5 to *E. coli* OmpA, a β -barrel-forming protein, and the identification of hypervariable surface-exposed loops [6] seems to be at odds with this proposal. The CD spectrum of purified P5 demonstrates a high content of β -strand, which is consistent with CD analysis of OmpA [22]. The SSPRED prediction of seven β -strands in the 200 N-terminal amino acids is also consistent with a β -barrel structure.

Although computer prediction methods have their limitations, no coiled coil structures were predicted for P5 by the two programs, COILS and SSPRED. Coiled coils, composed of two or three α -helices, usually at least four heptads long [28] wound around each other in a supercoil [29], are commonly found in fibrous proteins and are also predicted in the fimbrial streptococcal M protein [30]. There are only three helical domains of any length predicted in P5, all found in the C-terminus, and clearly one of these domains is the peptidoglycan-binding motif. The 94% identity of the 138 amino acids in the C-terminus of the P5 from these isolates is also consistent with a periplasmic location, shielded from the immune pressure of the host. Previous analysis of the fimbrin P5-homologue with the COILS program [7] (data were not shown) is difficult to reconcile with this current analysis, especially as similar analysis of streptococcal M protein unambiguously predicted a high probability for the presence of coiled coils.

Genes that encode fimbrin proteins are usually parts of operons that encode, as well as the structural protein, proteins involved in transport and assembly. This includes chaperones and outer-membrane ushers that form anchor proteins which connect the fimbriae to the cell [31]. The genomic organisation of the *p5* gene does not suggest that same organisation seen

with other fimbrial operons, at least in the Rd strain [20]. However, the possibility that P5 plays some role in the biogenesis of fimbriae cannot be excluded and may help explain some of the anomalies associated with the lack of fimbriae in the P5 mutant strain [7].

The model proposed for the 200 N-terminal residues of P5 is consistent with a number of features associated with the OmpA model. The amphipathic nature of the transmembrane β strands is well conserved in the P5 model, in which most residues, if not identical to OmpA, are conservatively substituted and consistent with the barrel structure of a hydrophobic face which interacts with the lipid environment of the membrane and a hydrophilic face to the interior of the barrel. Nine of the 10 charged residues distributed throughout the middle of the transmembrane region in OmpA are identical to residues in P5, while the conservative substitution of Lys-111 in P5 for the equivalent Arg-97 in OmpA in transmembrane strand five accounts for the tenth. Ten of the 12 glycine residues in the transmembrane domain of OmpA are also present in P5, where they are mostly found on the hydrophilic face of the barrel. A ring of aromatic residues (P5: Phe-7, Tyr-53, Tyr-100, Tyr-152 and Tyr-198; OmpA: Trp-7, Tyr-43, Tyr-129, Tyr-168, Phe-170) is found at equivalent places in both the OmpA and P5 structures on the hydrophobic face of the barrel at the border between the polar and non-polar regions at the periplasmic side of the membrane. Aromatic rings, which are a feature characteristic of porin structures and identified in the X-ray structure of the *Rhodobacter capsulatus* porin [32], may play a role in stabilising the structure within the membrane. At the outer face of the membrane, the ring is less defined: there are four aromatic residues which retain identity amongst the P5 isolates (P5: Tyr-47, Tyr-69, Phe-146 and Trp-166); however, only Tyr-69, Phe-146 and Trp-166 have identity with OmpA (OmpA: Trp-57, Phe-123 and Trp-143). The proposed periplasmic loops are all short in both models, which is similar to the *R. capsulatus* porin funnel-type structure of long polar loops external to the outer membrane, but with a flatter surface at the periplasmic face.

In conclusion, this study represents the first analysis of P5 heterogeneity in NTHi isolates from diverse anatomical sites. The characterisation of the surface-exposed regions of P5 will enhance future understanding of the role these domains play in host interaction and NTHi pathogenicity. Clearly, P5 forms a β -barrel structure analogous to OmpA, with no evidence of coiled coil structure.

We are extremely grateful to G. Ewart, G. Cox, F. Gibson and D. Bastin for helpful discussion, and to G. Cox for the generous provision of experimental resources. DNA sequencing was performed by C. McCrae of the Biomolecular Resource Facility, Centre for Molecular Structure and Function, ANU. The assistance of P. Guggen of the Research School of Chemistry, ANU, in the CD analysis is

gratefully acknowledged. D. W. was supported by a University of Canberra Postgraduate Research Award and a Collaborative Research Scholarship provided by The John Curtin School of Medical Research, ANU.

References

- Kilian M, Heine-Jensen J, Bülow P. *Haemophilus* in the upper respiratory tract of children. A bacteriological, serological and clinical investigation. *Acta Pathol Microbiol Scand B. Microbiol Immunol* 1972; **80**: 571–578.
- Murphy TF, Apicella MA. Nontypable *Haemophilus influenzae*: a review of clinical aspects, surface antigens, and the human immune response to infection. *Rev Infect Dis* 1987; **9**: 1–15.
- Turk DC. The pathogenicity of *Haemophilus influenzae*. *J Med Microbiol* 1984; **18**: 1–16.
- Groeneveld K, van Alphen L, Eijk PP, Jansen HM, Zanen HC. Changes in outer membrane proteins of nontypable *Haemophilus influenzae* in patients with chronic obstructive pulmonary disease. *J Infect Dis* 1988; **158**: 360–365.
- van Alphen L, Cautant DA, Duim B, O'Rourke M, Bowler LD. Differences in genetic diversity of nonencapsulated *Haemophilus influenzae* from various diseases. *Microbiology* 1997; **143**: 1423–1431.
- Duim B, Bowler LD, Eijk PP, Jansen HM, Dankert J, van Alphen L. Molecular variation in the major outer membrane protein P5 gene of nonencapsulated *Haemophilus influenzae* during chronic infections. *Infect Immun* 1997; **65**: 1351–1356.
- Sirakova T, Kolattukudy PE, Murwin D *et al.* Role of fimbriae expressed by nontypeable *Haemophilus influenzae* in pathogenesis of and protection against otitis media and relatedness of the fimbrial subunit to outer membrane protein A. *Infect Immun* 1994; **62**: 2002–2020.
- Reddy MS, Bernstein JM, Murphy TF, Faden HS. Binding between outer membrane proteins of nontypeable *Haemophilus influenzae* and human nasopharyngeal mucin. *Infect Immun* 1996; **64**: 1477–1479.
- Miyamoto N, Bakaletz LO. Selective adherence of nontypeable *Haemophilus influenzae* (NTHi) to mucus or epithelial cells in the chinchilla Eustachian tube and middle ear. *Microb Pathog* 1996; **21**: 343–356.
- Munson RS, Grass S, West R. Molecular cloning and sequence of the gene for outer membrane protein P5 of *Haemophilus influenzae*. *Infect Immun* 1993; **61**: 4017–4020.
- Hu N, Messing J. The making of strand-specific M13 probes. *Gene* 1982; **17**: 271–277.
- Sambrook J, Fritsch EF, Maniatis T. Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press. 1989.
- Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 1977; **74**: 5463–5467.
- Devereux J, Haerberli P, Smithies O. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res* 1984; **12**: 387–395.
- Mehta PK, Heringa J, Argos P. A simple and fast approach to prediction of protein secondary structure from multiply aligned sequences with accuracy above 70%. *Protein Sci* 1995; **4**: 2517–2525.
- Lupas A, van Dyke M, Stock J. Predicting coiled coils from protein sequences. *Science* 1991; **252**: 1162–1164.
- Carlone GM, Thomas ML, Rumschlag HS, Sottnek FO. Rapid microprocedure for isolating detergent-insoluble outer membrane proteins from *Haemophilus* species. *J Clin Microbiol* 1986; **24**: 330–332.
- Munson RS, Granoff DM. Purification and partial characterization of outer membrane proteins P5 and P6 from *Haemophilus influenzae* type b. *Infect Immun* 1985; **49**: 544–549.
- Provencher SW, Glöckner J. Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* 1981; **20**: 33–37.
- Fleischmann RD, Adams MD, White O *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 1995; **269**: 496–512.
- Mowa NR, Nakamura K, Inouye M. Gene structure of the OmpA protein, a major surface protein of *Escherichia coli* required for cell-cell interaction. *J Mol Biol* 1980; **143**: 317–328.
- Sugawara E, Steiert M, Rouhani S, Nikaido H. Secondary structure of the outer membrane proteins OmpA of *Escherichia coli* and OprF of *Pseudomonas aeruginosa*. *J Bacteriol* 1996; **178**: 6067–6069.
- Dale JB, Chiang EY, Lederer JW. Recombinant tetravalent group A Streptococcal M protein vaccine. *J Immunol* 1993; **151**: 2188–2194.
- Koebnik R. Proposal for a peptidoglycan-associating alpha-helical motif in the C-terminal regions of some bacterial cell-surface proteins. *Mol Microbiol* 1995; **16**: 1269–1270.
- Vogel H, Jähnig F. Models for the structure of outer-membrane proteins of *Escherichia coli* derived from raman spectroscopy and prediction methods. *J Mol Biol* 1986; **190**: 191–199.
- Prasadarao NV, Wass CA, Kim KS. Endothelial cell GlcNAc β 1-4GlcNAc epitopes for outer membrane protein A enhance traversal of *Escherichia coli* across the blood-brain barrier. *Infect Immun* 1996; **64**: 154–160.
- Prasadarao NV, Wass CA, Weiser JN, Stins MF, Huang S-H, Kim KS. Outer membrane protein A of *Escherichia coli* contributes to invasion of brain microvascular endothelial cells. *Infect Immun* 1996; **64**: 146–153.
- O'Shea EK, Rutkowski R, Kim PS. Evidence that the leucine zipper is a coiled coil. *Science* 1989; **243**: 538–542.
- Cohen C, Parry DAD. α -helical coiled coils and bundles: how to design an α -helical protein. *Proteins Struct Funct Genet* 1990; **7**: 1–15.
- Phillips GN, Flicker PF, Cohen C, Manjula BN, Fischetti VA. Streptococcal M protein: α -helical coiled-coil structure and arrangement on the cell surface. *Proc Natl Acad Sci USA* 1981; **78**: 4689–4693.
- Hultgren SJ, Abraham S, Caparon M, Falk P, St Geme JW, Normark S. Pilus and nonpilus bacterial adhesins: assembly and function in cell recognition. *Cell* 1993; **73**: 887–901.
- Weiss MS, Abele U, Weckesser J, Welte W, Schiltz E, Schulz GE. Molecular architecture and electrostatic properties of a bacterial porin. *Science* 1991; **254**: 1627–1630.