Structural and Evolutionary Inference from Molecular Variation in Neisseria Porins

JEREMY P. DERRICK,1 RACHEL URWIN,2 JANET SUKER,3 IAN M. FEAVERS,3 AND MARTIN C. J. MAIDEN2,*

Department of Biomolecular Sciences, UMIST, Manchester M60 1QD,1 Division of Bacteriology, National Institute for Biological Standards and Control, South Mimms, Potters Bar, Hertfordshire EN6 3QG,2 and Wellcome Trust Centre for the Epidemiology of Infectious Disease, Department of Zoology, University of Oxford, Oxford OX1 3PS,2 United Kingdom

Received 20 November 1998/Returned for modification 14 December 1998/Accepted 18 February 1999

The porin proteins of the pathogenic Neisseria species, Neisseria gonorrhoeae and Neisseria meningitidis, are important as serotyping antigens, putative vaccine components, and for their proposed role in the intracellular colonization of humans. A three-dimensional structural homology model for Neisseria porins was generated from Escherichia coli porin structures and N. meningitidis PorA and PorB sequences. The Neisseria sequences were readily assembled into the 16-strand β-barrel fold characteristic of porins, despite relatively low sequence identity with the Escherichia proteins. The model provided information on the spatial relationships of variable regions of peptide sequences in the PorA and PorB trimers and insights relevant to the use of these proteins in vaccines. The nucleotide sequences of the porin genes from a number of other Neisseria species were obtained by PCR direct sequencing and from GenBank. Alignment and analysis of all available Neisseria porin sequences by use of the structurally conserved regions derived from the PorA and PorB structural models resulted in the recovery of an improved phylogenetic signal. Phylogenetic analyses were consistent with an important role for horizontal genetic exchange in the emergence of different porin classes and confirmed the close evolutionary relationships of the porins from N. meningitidis, N. gonorrhoeae, Neisseria lactamica, and Neisseria polysaccharea. Only members of this group contained three conserved lysine residues which form a potential GTP binding site implicated in pathogenesis. The model placed these residues on the inside of the pore, in close proximity, consistent with their role in regulating pore function when inserted into host cells.

The genus Neisseria comprises a number of species associated with mucosal carriage and disease in animals and humans (3, 26); these species include Neisseria meningitidis, a major cause of meningitis and septicaemia worldwide (6), and Neisse- ria gonorrhoeae, the etiological agent of gonorrhoea (31). Comparative studies of these two pathogenic species with others not frequently associated with disease, some of which are closely related (16), are helpful in understanding both the pathogenesis of meningococcal and gonococcal infections and the evolutionary events that led to their emergence. Such studies are particularly valuable when they target cell components potentially involved in immunity or other host-pathogen interactions, such as outer membrane proteins (OMPs) (35–37).

The Neisseria porins, a distinct class within the porin superfamily (20), are major components of the outer membranes of all members of the genus Neisseria (11, 35, 37). Most Neisseria species express only one, referred to as Por, the meningococcus exceptionally expressing two, PorA and PorB (17). The gonococcus is the only other Neisseria species known to have a porA gene, which is not expressed due to frameshift and promoter mutations (11). The meningococcal and gonococcal porins are targets for serological typing schemes (14, 15), candidates for inclusion in vaccines (13), and have been implicated in pathogenesis (29).

Interest in the antigenic variability of the Neisseria porins, from both typing and vaccination perspectives, has resulted in the generation of many sequences of their genes. Detailed interpretation of this database of sequences has been hindered by the lack of a molecular structure for the gene products, although a two-dimensional model has been proposed on the basis of sequence similarity (23, 36). This model comprises a β-barrel porin structure, similar to that of the Escherichia coli porins (9), with regions that are relatively conserved among Neisseria sequences forming the β barrel. Other regions, which are less well conserved among and within species, form surface-exposed loops which protrude from the bacterial surface into the surrounding environment (23, 36). While this model was useful, it was not three-dimensional, being proposed before the molecular structures of the E. coli porins were established; the assignment of residues to membrane-embedded or surface-exposed environments was somewhat uncertain.

In this work, we have used molecular modelling techniques to generate a three-dimensional homology model of the Neisseria porins using E. coli porin structures, enabling us to identify more precisely the structural roles of individual residues. Phylogenetic analyses of the porins from different Neisseria species were enhanced when the model was used to assist sequence alignment and to identify structurally conserved parts of the proteins, identifying features likely to be important in the emergence of the pathogenic Neisseria species.

MATERIALS AND METHODS

Neisseria strains and sequences from databases. The Neisseria porin sequences were obtained by translation of nucleotide sequences obtained by direct nucleotide sequence determination in this study or from GenBank, as detailed in Table 1. The porin of Neisseria polysaccharea was excluded from the analysis, as it was essentially identical to the Neisseria lactamica porin. Multiple examples of the porins were not included, with the exception of five additional N. gonorrhoeae...
porins classified elsewhere as “intermediate” and “hybrid” (8) and both Neisseria sicca porins, which were not closely similar.

For the purposes of this work, a nomenclature system suggested previously for Neisseria porins (11) has been adopted and extended. As there is a Por1 monomer in the gonococcus, its expressed porin, previously identified as Por, has been renamed PorB for consistency with the meningococcal nomenclature. Within meningococcal and gonococcal PorB porins, distinct allele classes are distinguished with numbers: the meningococcal class 2 OMP becomes PorB2, and the class 3 OMP becomes PorB3. The gonococcal PIA porin becomes PorB1a, and the PIB protein becomes PorB1b. These designations were intended to maintain as much consistency with previous nomenclature as possible while indicating more accurately the relationships among the proteins and the genes that encode them. The porins of other species have been labelled Por1 and Por2 indicated the two alleles identified in Por, as there remains no evidence for a Por1 gene in these species. The two distinct porins present in Neisseria flavescens were identified as Por1 and Por2, as neither bore a close relationship to the meningococcal and gonococcal PorA and PorB proteins. For N. sicca, the designations Por-1 and Por-2 indicated the two alleles identified in this species, which were insufficiently diverse to be regarded as representative of distinct allele classes. Final designation of the N. flavescens and N. sicca porins requires further study.

**Nucleotide sequence determination of Neisseria porin genes.** Chromosomal DNA was extracted from each of the Neisseria strains by the Isoquick DNA extraction procedure (Orca Research). The porin genes were amplified with PCR primers 27 and 28 (12), purified as described previously (10), and their sequences were determined on both strands with Big Dye terminators (PE Applied Biosystems). The reaction products were sequenced using the Applied Biosystems Prism 377 automated sequencer, and the sequences were assembled with the Staden sequence analysis package (33).

**Phylogenetic analyses.** Nucleotide sequence alignments derived from the peptide sequence alignments shown in Fig. 1 were analyzed with MEGA (22), which was used to generate nonsynonymous distances, estimated by the method of Nei and Gojobori (27), with deletions excluded in pairwise comparisons. In some analyses, only sites within regions predicted to form the β barrel of the protein were included to minimize the effect of the large diversity present in the loop regions. The distance matrices obtained were used to reconstruct phylogenetic trees by the neighbor-joining method; these phylogenies were represented as unrooted radial trees with the program TREEVIEW (data not shown). The
distance data were also visualized as split graphs generated by use of the split decomposition analysis technique with the program SPLITSTREE version 2.4 (18).

**Nucleotide sequence accession numbers.** The novel sequences determined in this study have been deposited in GenBank under accession no. AF121870 to AF121876 (Table 1).

**RESULTS**

**Novel porin sequences and alignments.** The novel porin sequences obtained were consistent with those previously described for various *Neisseria* porins and are shown in Fig. 1, aligned with the sequences for the two *E. coli* porins with known three-dimensional structures and one example of each previously sequenced porin gene. The alignment was based on that of Jeanteur et al. (20), modified to take account of the additional structural information available (9). The initial alignment made was that of the *E. coli* porins and the meningococcal PorA and PorB proteins; this alignment was refined and the other *Neisseria* porins were added after construction of the structural homology models. Correct alignment of meningococcal PorA and PorB with the *E. coli* porins was critical for the generation of accurate models, and each putatively conserved residue in the *E. coli* crystal structures was examined to verify that there was a sound structural basis for its conservation between *Neisseria* and *E. coli* porins. The most striking feature of the alignment was that the *Neisseria* sequences began at the second β strand, within the β-barrel framework, effectively truncating the first β strand, unlike the situation for *E. coli*. This alignment provided more convincing homology models and was also justified from sequence variations within the *Neisseria* porins that would otherwise have occurred within the β-barrel framework.

**Homology models.** Given the high degree of sequence divergence between the *Neisseria* and *E. coli* porins, the β-barrel framework of the modelled porins was constructed remarkably well. Both the PorA and the PorB models formed a well-defined main-chain hydrogen-bonding pattern characteristic of the β-sheet motif, although no hydrogen-bonding or backbone dihedral-angle restraints were used in the calculation. Larger loop regions were less well modelled by this procedure and showed a variety of conformations, depending on the starting alignment and other restraints. Loop regions within proteins are notoriously difficult to model, and a more reliable calculation of their conformations requires further experimental data. The PorA monomer model is illustrated in Fig. 2.

The limitations of the MODELLER package required that the porin models be calculated as monomers, although they exist in a trimeric state within the outer membrane (25). A more realistic model, which constructed the trimer from its constituent monomers, was produced with X-plor (4) by placing the monomers within the original trigonal crystallographic unit cell that was used for the *E. coli* OmpF porin. Several cycles of conjugate gradient energy refinement provided the final model, which was readily accommodated into a trimer without substantial modifications. The PorA trimer model is shown in Fig. 3, viewed from the top. A similar trimer was assembled within the OmpF crystalgraphic unit cell with the PorB monomer model. The fact that the modelled PorA and PorB monomers fit well into the trigonal space group of the OmpF trimer with no major steric clashes indicated that these models were likely to be correct and, further, that the other *Neisseria* porins were likely to conform to these models.

The models had several structural features that are common to porins with known three-dimensional structures. Each model had a well-defined pore that was lined with hydrophilic and hydrophobic side chains of appropriate dimensions for the passage of low-molecular-weight solutes. The distribution of aromatic amino acid side chains within the PorA and PorB models was similar to that established for other porins: the majority of the Phe, Tyr, and Trp side chains point outward in interactions with phospholipids. A subset of these residues interact around the threefold symmetry axis of the PorA trimer and may play a role in stabilizing the quaternary structure of the porin.

**Phylogenetic relationships of Neisseria porin genes.** A number of phylogenetic reconstructions were made with nucleotide sequence alignments based on the peptide sequence alignments shown in Fig. 1. The final phylogeny, shown in Fig. 4, was reconstructed with the following considerations. Due to the strategy used in sequencing, some of the porin genes were truncated; regions corresponding to the first 12 and last 8 amino acids were missing (12, 23). These regions were there-
fore excluded from all sequences for the phylogenetic analyses. As these regions were relatively small and located in parts of the protein that are relatively conserved, the consequent loss of phylogenetic resolution was minimal. Given the large distances among sequences, nonsynonymous distances, estimated as described previously (27), were used. Further, only those regions of the gene encoding the \( \beta \) barrel were included to remove any distortions of the phylogeny introduced by signals from the rapidly evolving antigenically variable loops (32). Finally, split decomposition, which does not assume a tree-like phylogenetic structure (1), was used to ensure that any signal implying that recombination had occurred in the evolution of the \textit{Neisseria} porin genes was identified.

**DISCUSSION**

The antigenic variation of human pathogens remains incompletely understood, impeding vaccine development and complicating molecular typing, which relies on the characterization of such variation to distinguish isolates. An improved understanding of the complex interactions of the host immune system with variable antigens requires a multidisciplinary approach that integrates the inferences available from structural, immunological, and evolutionary data. At present, this integration is most readily achieved for protein antigens, where the relationships among genetic change in the pathogen, antigen structure, and interaction with the host immune system are relatively easily investigated. The porins of the genus \textit{Neisseria} are an instructive example for a number of reasons, including their intrinsic interest, the large number of sequences available, and the fact that the genus includes both pathogenic and commensal species, whose lifestyle imposes constant selection pressure from the host immune response.

\textit{Neisseria} porin antigens comprise regions of relatively conserved sequence, which are predicted to form the \( \beta \)-barrel structure of the proteins, interspersed with more variable regions, which form the putative surface-exposed loops. With two exceptions, these loops are highly variable within and among the \textit{Neisseria} species. These exceptions, putative loops II and III, have structural roles in the three-dimensional model. Putative loop II is important in monomer-monomer interactions within the porin trimer, while putative loop III is sequestered in the pore of each monomer, potentially influencing pore function (Fig. 3). The other loops, which exhibit diversity in both length and sequence, are the determinants of antigenic variability in species that have been extensively studied. Loops I, IV, V, and VI (Fig. 1) show the greatest length variation. The hypervariable regions of meningococcal PorA, VR1 and VR2 (24), correspond to loop regions I and IV of the model (36). In the monomer PorA model, these loops appear to be widely spaced (Fig. 2), making it unlikely that the hypervariable regions would come into direct contact across a single monomer. However, the view of the PorA trimer from above (Fig. 3) indicates that the proposed locations of VR1 and VR2 (loops I and IV) are in close proximity between adjacent monomers, raising the possibility that these regions may interact to some extent across the monomer-monomer interface. In principle, therefore, a PorA epitope could extend across more than one polypeptide chain in a PorA trimer.

Comparison of the pairing of \( \beta \) strands and the positioning...
of the loop regions present in the three-dimensional PorA model with those predicted in the two-dimensional model of van der Ley et al. (36) showed good agreement for the locations of loops I, IV, V, VIII and for the β strands, with the exception of strands 1, 2, 4, 5, and 14. The deviations observed were due at least in part to the fact that the lengths of the β strands vary in the OmpF crystal structure, illustrating the additional information that a three-dimensional structure can add to a model based on sequence alignments.

The variability of the loop sequences, which is likely to be the result of strong positive selection (32), may distort the phylogenetic signal present in the sequences. One result would be to make closely related porins appear more diverse. For example, a recent study proposed hybrid and intermediate porin classes for the gonococcal porins (8). Including porins from other species, limiting the analysis only to those parts of the protein that form the β barrel, and applying more sensitive phylogenetic methods to the alignments generated here show that these proteins cluster closely with other members of their respective porin classes, the apparent differences being introduced mainly by the highly variable loop sequences (Fig. 4).

Whatever the cause of the star phylogeny, the porins of N. canis, N. elongata, and N. mucosa appear to share a common ancestor, occupying one branch of the star, as do those of N. denitrificans and N. animalis. The two N. sicca variants are diverse, perhaps indicating difficulties in the definition of this species. The two separate porin genes present in one N. flavescens isolate are sufficiently diverse that they may have been assembled in the same species relatively recently by horizontal genetic exchange.

Meningococcal PorA and gonococcal PorA are the most diverse members of the family of Neisseria porins and might constitute an outgroup relative to the other porins. The PorB porins of N. meningitidis and N. gonorrhoeae and the PorB porin of N. lactamica (and the closely related N. polysaccharea), with the exception of meningoococal PorB2, form a well-resolved and supported bifurcating tree. The interrelationships of these porins indicate that interspecies genetic exchange of porin genes has occurred during the emergence of the present-day species. It is possible that the common ancestor of the meningococcus and the gonococcus possessed a gene ancestral to the porB1b and porB3 genes. The porB1b gonococcal porin gene...
appears to have arisen by the replacement of a porB1a allele with a gene sharing a common ancestor with the N. lactamica porin gene.

The meningococcal porB2 gene is a special case which appears to have arisen as a result of intragenic recombination during the evolution of the pathogenic Neisseria species. The split graph shows that PorB2 occupies an intermediate position between N. gonorrhoeae and N. meningitidis PorB and the N. lactamica protein on the one hand and the commensal-animal Neisseria porins on the other, with a network of relationships suggesting a phylogenetic signal that is inconsistent with a bifurcating tree. The data contributing to these relationships are illustrated in Fig. 5, which shows the amino acid identities within the β-barrel-forming regions among PorB2, PorB3, and the N. flavescens Por2 protein (similar results were obtained in comparisons of other representative porins from the commensal-animal and pathogenic groups) (data not shown). PorB2 was most similar to the N. flavescens porin (19 amino acid identities in structural regions) but was also related to PorB3 (17 amino acid identities). However, while the identities between PorB2 and the N. flavescens porin were distributed throughout the sequence, those between PorB2 and PorB3 were mainly localized in β sheets on either side of loop II. This region includes the three lysine residues (positions 66, 82, and 113 in Fig. 1) putatively identified as being involved in the binding of ATP and GTP. This binding provides a gating mechanism that down-regulates pore size and changes voltage dependence and ion selectivity when PorB proteins are inserted into mammalian cells (29). In the PorB model, the side chains of these three residues protrude into the center of the pore, providing a plausible steric explanation of the role of GTP in obstructing the passage of solutes and changing the selectivity of the porin. As this gating mechanism, which is only observed in PorB2, PorB3, PorB1a, PorB1b, and the N. lactamica and N. polysaccharea proteins, is conserved, it is tempting to suppose that PorB proteins with this characteristic were important in the evolution of the pathogenic Neisseria species, perhaps by promoting intracellular invasion and thereby improving carriage in humans. The fact that there is a network of interrelationships among PorB2, the porins of the pathogenic Neisseria species, and the porins of the commensal-animal group suggests that PorB2 may have evolved by recombination of two proteins ancestral to the pathogenic and commensal porins, resulting in an antigenically different porin that retained the GTP binding site.

While advances in automated sequence technology have enabled large numbers of nucleotide sequences of variable antigens to be sequenced rapidly, interpretation of these data continues to be limited by the lack of structural data, which remain comparatively difficult to obtain. The insights available from structural biology can enhance understanding of the evolutionary, immunological, and pathological significance of the gene sequence data. The approach taken in the present work, of combining experimental observations, structural modelling, and phylogenetic analyses, is valid, notwithstanding the wide sequence variation between the Neisseria and E. coli porins, as the β-barrel porin structure is well conserved in a wide variety of species. The porin model presented here not only extends the two-dimensional models proposed earlier, enabling the design of further structural and immunological analyses, but also permits more accurate phylogenetic analyses, providing insights into the biochemical mechanisms that may have contributed to the evolution of the pathogenic Neisseria species.

ACKNOWLEDGMENTS

J.P.D. is a fellow of the Lister Institute of Preventive Medicine. M.C.J.M. is a Wellcome Trust senior fellow in biodiversity. We thank Eddie Holmes for comments on the manuscript and advice on the phylogenetic analyses.

REFERENCES