Phylogenetic Evidence for Frequent Positive Selection and Recombination in the Meningococcal Surface Antigen PorB

Rachel Urwin,* Edward C. Holmes,† Andrew J. Fox,‡ Jeremy P. Derrick,§ and Martin C. J. Maiden*

*The Peter Medawar Building for Pathogen Research and Department of Zoology, University of Oxford; †Department of Zoology, University of Oxford; ‡Meningococcus Reference Unit, Public Health Laboratory, Withington Hospital, Manchester; and §Department of Biomolecular Sciences, University of Manchester Institute of Science and Technology

Previous estimates of rates of synonymous (dS) and nonsynonymous (dN) substitution among Neisseria meningitidis gene sequences suggested that the surface loops of the variable outer membrane protein PorB were under only weak selection pressure from the host immune response. These findings were consistent with studies indicating that PorB variants were not always protective in immunological and microbiological assays and questioned the suitability of this protein as a vaccine component. PorB, which is expressed at high levels on the surface of the meningococcus, has been implicated in mechanisms of pathogenesis and has also been used as a typing target in epidemiological investigations. In this work, using more precise estimates of selection pressures and recombination rates, we have shown that some residues in the surface loops of PorB are under very strong positive selection, as great as that observed in human immunodeficiency virus-1 surface glycoproteins, whereas amino acids within the loops and the membrane-spanning regions of the protein are under purifying selection, presumably because of structural constraints. Congruence tests showed that recombination occurred at a rate that was not sufficient to erase all phylogenetic similarity and did not greatly bias selection analysis. Homology models of PorB structure indicated that many strongly selected sites encoded residues that were predicted to be exposed to host immune responses, implying that this protein is under strong immune selection and requires further examination as a potential vaccine candidate. These data show that phylogenetic inference can be used to complement immunological and biochemical data in the choice of vaccine candidates.

Introduction

The generation of antigenic diversity has evolved as a strategy for evading immune attack in a wide range of pathogenic and commensal organisms (Deitsch, Moxon, and Wellens 1997). It is effective against both natural and artificially induced immunity and represents a major obstacle to the development of vaccines against pathogens as diverse as Plasmodium falciparum, human immunodeficiency virus (HIV), and Neisseria meningitidis, the meningococcus. In an era when many new vaccine candidates are being identified by genomic techniques, sequence data of the antigen genes obtained from population samples of pathogens can be analyzed by phylogenetic and biochemical modeling techniques to provide a picture of the evolutionary processes acting on these sequences and hence a preliminary evaluation of their vaccine potential.

The meningococcus is an appropriate model system to evaluate this approach because it is a pathogen of global significance which is genetically and antigenically diverse and for which no comprehensive vaccine exists (Pollard and Frasch 2001). Further, large genetically defined isolate collections have been assembled and models of the population biology of this organism are available (Caugant et al. 1987; Maiden et al. 1998). Amongst the candidate vaccine components proposed are the variable outer membrane proteins, the trimeric porins, which act as pores for the passage of solutes into the cell (Tommassen et al. 1990). These molecules are targeted by the host immune response and have been used in meningococcal typing schemes (Bjune et al. 1991a, 1991b; Sierra et al. 1991; van der Ley and Poolman 1992; van der Ley, van der Bielen, and Poolman 1995). Unlike most other Neisseria species, the meningococcus expresses two porins, PorA and PorB. Expression of PorA is regulated at transcription and exhibits three levels depending on the length of the poly-guanidine stretch in the promoter region of the porA gene (van der Ende et al. 1995), whereas there is no evidence to suggest that PorB proteins are subject to phase variation. Mutant meningococcal strains that lack PorB do not grow well (Tommassen et al. 1990), suggesting that PorB has a function essential for growth. PorB is also capable of translocating vectorially into the membranes of mammalian cells (Blake and Gotschlich 1987), and of binding ATP and GTP, which down regulates pore size and alters voltage dependence and ion selectivity (Rudel et al. 1996). These functional and structural characteristics are thought to influence the early stages of neutrophil activation and therefore implicate the PorB protein in meningococcal pathogenesis (Rudel et al. 1996).

A PorB topology model has been constructed on the basis of nucleotide sequence data (Maiden et al. 1991; van der Ley et al. 1991) and, more recently, the structural similarity between the Neisseria porins and the Escherichia coli porins OmpF and PhoE has been exploited to generate a three-dimensional homology model for Neisseria porins (Derrick et al. 1999). These models predicted eight surface exposed “loops” interspersed with highly conserved outer membrane-spanning sequences that formed a “β-barrel” (Kleffel et al.
1985). The antigenically variable epitopes targeted in the host immune response (Saukkonen et al. 1989) were proposed to reside in the surface-exposed loops (McGuinness et al. 1990; Maiden et al. 1991). Serological and molecular characterization of the meningococcal porins have been used for epidemiological analyses of meningococcal carriage and disease (Frasch, Zollinger, and Poolman 1985; Poolman et al. 1986; Maiden et al. 1991), although the variability of these proteins means that they are not always reliable epidemiological markers (Achtman 1995; Urwin et al. 1998a, 1998b).

The meningococcus possesses one of two PorB protein classes, PorB2 or PorB3, which are encoded by alternate allele classes present at the *porB* locus. Phylogenetic analyses show that the *porB3* gene is most closely related to one of the gonococcal porin genes, *porB1a* (Smith, Maynard Smith, and Spratt 1995) and that these gene sequences form a clade together with gonococcal *porB1b*, *N. lactamica* por, and *N. polysacchara* *por* gene sequences (Derrick et al. 1999). The meningococcal *porB2* gene shares sequence similarity both with members of this clade (specifically, the sequence encoding the putative ATP and GTP binding site) and with the porin genes of most human commensal and animal *Neisseria* species, suggesting that *porB2* may have arisen because of interspecies recombination (Derrick et al. 1999). Indeed, the incongruence between phylogenetic trees drawn for individual loop-encoding regions of *porB3* genes (Bash et al. 1995) suggests that inter- and intraspecies recombination is a mechanism that increases genetic variation among *porB* alleles.

Previous comparisons of the rates and distribution of synonymous (*dS*) and nonsynonymous (*dN*) substitutions among meningococcal porin genes have concluded that, unlike the gonococcal PIA and PIB porins which were under positive selection in the surface loop regions, meningococcal *porB* genes were subject to only weak positive selection and purifying selection (Smith, Maynard Smith, and Spratt 1995). This observation supports the hypothesis that PorB is a less important vaccine constituent than the PorA protein. This conclusion was also drawn from studies of the bactericidal activity of monoclonal (Saukkonen et al. 1987) and polyclonal (van der Ley and Poolman 1992) antibodies in mice and immunological results from human vaccine trials (Rosenqvist et al. 1995; Perkins et al. 1998). In addition, some PorB3 proteins expressed on the surfaces of live meningococci have been reported to be poorly accessible for antibody binding (Michaelsen et al. 2001). Consequently, PorB has been deliberately excluded from some vaccine formulations (van der Ley, van der Bijzen, and Poolman 1995). But the previous selection analysis was conducted on very small data sets, and *dN* and *dS* were estimated as mean values across whole or partial gene sequences, making it possible that strongly selected sites were missed in this broadscale comparison (Smith, Maynard Smith, and Spratt 1995). More recently, phylogenetic analyses carried out on a larger set of gonococcal porin gene sequences concluded that there were differences in the evolution of PIA and PIB homology groups, with positive selection driving evolution of the PIA proteins and both positive and purifying selection acting on PIB protein sequences (Posada et al. 2000).

In this work we undertook rigorous maximum likelihood analyses of selection pressures acting on a large set of meningococcal PorB sequences. A likelihood-based approach was also used to determine with more accuracy the extent of recombination in *porB2* and *porB3*. Our study reveals that both genes are subject to exceptionally high rates of positive selection, as well as frequent recombination, which has important implications for the use of PorB as a potential vaccine candidate.

### Materials and Methods

**Meningococcal porB Gene Sequences**

A total of 324 *porB* gene sequences was examined: 121 from GenBank, including previously published sequences (Murakami, Gotschlich, and Seiff 1989; Wolff and Stern 1991; Feavers et al. 1992; Ward, Lambden, and Heckels 1992; Bash et al. 1995; Sacchi et al. 1998; Urwin et al. 1998a, 1998b), and two from complete meningococcal genome sequences (Parkhill et al. 2000; Tettelin et al. 2000); a further 203 *porB* sequences from meningococcal isolates were determined de novo, including 107 isolates from globally representative strain collections, (Maiden et al. 1998), 12 PorB serotyping reference strains (Feavers et al. 1992), 50 isolates from healthy carriers, and 44 disease-causing meningococci from England and Wales.

For de novo sequencing, the propagation of DNA, *porB* gene amplification, and nucleotide sequence determination were as described previously (Feavers et al. 1999; Urwin et al. 1998a). The sequences were assembled with the Staden sequence analysis package (Staden 1996) and all sequences aligned manually in the Seqlab alignment program (Genetics Computer Group, Madison, Wis.) (Devereux, Haereli, and Smithies 1984). The sequence alignment was trimmed at 5' and 3'-ends so that all sequences began at the 5'-end with the thirteenth codon of the sequence encoding the mature protein (GAA in all sequences) and ended seven codons from the 3'-end of this sequence (ATG in *porB2*; GGT in *porB3*) because this corresponded to the length of the shortest sequences in the data set. Calculation of the number of nucleotide differences between pairs of alleles was determined using MEGA version 1.01 (Kumar, Tamura, and Nei 1994). The *porB* allele sequences and alignments can be viewed at http://meisseria.org/typing/porb. There were 125 unique *porB* allele sequences, 46 of which were *porB2* sequences (named *porB2*-1 to *porB2*-46, according to a previously defined nomenclature [Feavers and Maiden 1998]) and 79, which were *porB3* sequences (*porB3*-1 to *porB3*-79). Previously unpublished sequences have been submitted to GenBank, accession numbers AF520356–AF520416.

**Analysis of Selection Pressures**

A maximum likelihood (ML) approach was used to examine selection pressures acting on the meningococ-
cal *porB* genes. Here, $d_S$ and $d_S$ were examined codon-by-codon, using different models of codon substitution that differed in how $d_S/d_S$ ratios (parameter $\omega$) varied along sequences, as well as incorporating information about the phylogenetic relationships of the sequences in question so that comparisons are independent (Yang et al. 2000). Model M0 estimated a single $\omega$ parameter for all sites, whereas the M1 model divided codons into conserved sites ($p_0$), with $\omega_1$ fixed at 0 and neutral sites ($p_1$) with $\omega_2$ estimated from the data. M3 provided a more sensitive test by estimating, from the data, $\omega$ values for three classes of site all of which could be $>1$. The M7 and M8 models both used a discrete beta distribution (with 10 categories and described by parameters $p$ and $q$) to model $\omega$ ratios among sites, although M8, unlike M7, considered an extra class of sites for which $\omega$ could be $>1$. Nested models could be compared using a likelihood ratio test (LRT) in which twice the difference in log likelihood between models was compared with the value obtained under a $\chi^2$ distribution (degrees of freedom equal to the difference in the number of parameters between models). Finally, Bayesian methods were used to determine the probability that a particular codon site fell into the positively selected class. All these analyses used the CODEML program from the PAML package (Yang 1997).

Phylogenetic trees for the two data sets were constructed using the maximum likelihood method available in the PAUP* package (Swofford 1998). The HKY85 model of nucleotide substitution was used, with values for both the transition-transversion ratio and the shape parameter ($\alpha$) of a gamma distribution of rate variation among sites (with eight categories) estimated during tree reconstruction (trees and parameter values available on request).

Recombination Analysis

The extent of recombination in the *porB* sequence data was analyzed by assessing the degree of phylogenetic congruence. Because of the large numbers of sequences available, this analysis was performed on a set of 35 randomly sampled *porB2* and *porB3* alleles. The *porB2* and *porB3* alignments were first split into two equal-sized fragments. ML phylogenetic trees were then estimated for both halves of the alignments using the procedures described above. To establish whether the trees constructed on each half of the alignment were significantly different in topology, as might be expected given frequent recombination, the difference in log likelihood ($\delta$) between the ML tree for the first half of the gene and the ML tree topology for the second half of the gene fitted to the data from the first half, but with branch lengths reoptimized, were compared. The significance of the likelihood differences was assessed using two randomization tests. The first test used Monte Carlo simulation on 100 replicate data sets simulated under the ML model parameters using the program Seq-Gen (Rambaut and Grassly 1997). Maximum likelihood trees were then constructed for each of these simulated data sets, using the procedures described above, and their likelihoods compared with those of ML topology on each data set. If the $\delta$ values for the real data fell within this null distribution of $\delta$ values, then the trees constructed for each half of the gene were not significantly different in topology. In the second test, 200 random trees were created using PAUP*. The likelihoods of these trees were then estimated on the data from the first half of the *porB2* and *porB3* alignments, again with the reoptimization of branch lengths, and the $\delta$ values between these random trees and the two ML trees were then compared. If the $\delta$ values for the two ML trees fell within the 99th percentile of this null distribution then we may say that they are no more similar than two random trees inferred from these data (Holmes, Urwin, and Maiden 1999).

Table 1

<table>
<thead>
<tr>
<th></th>
<th><em>porB2</em></th>
<th><em>porB3</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of alleles</td>
<td>46</td>
<td>79</td>
</tr>
<tr>
<td>Length of sequence (bp)</td>
<td>1,053</td>
<td>894</td>
</tr>
<tr>
<td>No. of variable sites</td>
<td>147</td>
<td>136</td>
</tr>
<tr>
<td>Sequence divergence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean $p$-distance</td>
<td>4.2%</td>
<td>3.9%</td>
</tr>
<tr>
<td>Range</td>
<td>0.1%–7.2%</td>
<td>0.1%–8.2%</td>
</tr>
<tr>
<td>$porB2$ vs. $porB3$</td>
<td>30.9%–32.3%</td>
<td></td>
</tr>
<tr>
<td>No. of amino acid sequences</td>
<td>46</td>
<td>77</td>
</tr>
<tr>
<td>No. of variable sites</td>
<td>70</td>
<td>61</td>
</tr>
<tr>
<td>Mean $p$-distance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loop I</td>
<td>9.8%</td>
<td>27.7%</td>
</tr>
<tr>
<td>Loop II</td>
<td>8.3%</td>
<td>10.0%</td>
</tr>
<tr>
<td>Loop III</td>
<td>6.0%</td>
<td>6.0%</td>
</tr>
<tr>
<td>Loop IV</td>
<td>15.5%</td>
<td>13.9%</td>
</tr>
<tr>
<td>Loop V</td>
<td>25.2%</td>
<td>31.1%</td>
</tr>
<tr>
<td>Loop VI</td>
<td>34.1%</td>
<td>27.4%</td>
</tr>
<tr>
<td>Loop VII</td>
<td>16.7%</td>
<td>35.9%</td>
</tr>
<tr>
<td>Loop VIII</td>
<td>24.1%</td>
<td>13.9%</td>
</tr>
</tbody>
</table>

Structural Models

Structural models for the PorB2 and PorB3 proteins were generated using the software package Modeller (Sali et al. 1995), using the crystal structure of the porin Omp32 from *Comamonas acidovorans* as a template (Zeth et al. 2000; PDB accession 1E54). The sequence alignment was based on that given by Zeth et al. for PorB, with minor modifications.

Results

Nucleotide and Amino Acid Sequence Diversity

A total of 14% (147 of 1,053, table 1) of the nucleotide sites of the 46 unique *porB2* allele sequences were polymorphic, with percentage divergences ($p$ distances) in pairwise comparisons ranging from 0.1% to 7.2% (mean 4.2%). For the 79 unique *porB3* sequences, 15.2% (136 of 894, table 1) of the nucleotide sites were polymorphic, with $p$ distances between 0.1% and 8.2%
(mean 3.9%). Sequence divergence between the porB2 and porB3 allele classes ranged from 30.9% to 32.3%. The amino acid sequences of the putative surface exposed loop regions of the PorB2 and PorB3 proteins, determined according to the structural model for Neisseria porins, were longer in the PorB2 proteins. For the PorB2 proteins, putative loops V and VI were the most variable, with 25 distinct amino acid sequences (25.2% mean divergence) and 19 sequences (34.1% mean divergence), respectively. Putative loop VII was the most variable among PorB3 sequences, with 15 distinct amino acid sequences (35.9% mean divergence), although putative loops I, V, and VI were also variable, with 10 or more amino acid sequences (mean divergence >27%) in each loop region. The structural regions that interspersed the variable loops were largely conserved.

Analysis of Recombination Frequency

The maximum likelihood analysis of tree congruence provided evidence for recombination within both the porB2 and porB3 genes, although not at a rate sufficiently high to completely erase all phylogenetic similarity. Specifically, although the ML tree for the first half of the porB2 gene had a much higher likelihood than that of the ML tree for the second half of the gene fitted to the first half data ($\delta = 254.751$), and this difference was far greater than expected by chance as determined using Monte Carlo simulation (max $\delta = 350.571$), it was less than that seen in random tree topologies (99th percentile of $\delta = 395.806$). Similar results were seen in porB3, where the difference between the two ML trees, $\delta = 350.571$, also fell outside of the null distribution produced by Monte Carlo simulation (max $\delta = 32.876$) but was less than that of random tree topologies (99th percentile of $\delta = 395.806$).

### Table 2

Parameter Estimates for the Maximum Likelihood Analysis of Selection Pressures Acting on the porB2 Gene of Neisseria meningitidis

<table>
<thead>
<tr>
<th>Model</th>
<th>Site Categories ($\omega$)</th>
<th>Likelihood Test</th>
<th>$\chi^2$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>M0</td>
<td>$\omega = 1.085$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M1</td>
<td>$p_0 = 0.908, p_1 = 0.092$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M2</td>
<td>$p_0 = 0.834, p_1 = 0.145, p_2 = 0.21$</td>
<td>M0 vs. M2</td>
<td>1174.738</td>
<td>&lt;0.000</td>
</tr>
<tr>
<td></td>
<td>$\omega_2 = 13.215$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M3</td>
<td>$p_0 = 0.045, p_1 = 0.944, p_2 = 0.011$</td>
<td>M0 vs. M3</td>
<td>1192.657</td>
<td>&lt;0.000</td>
</tr>
<tr>
<td></td>
<td>$\omega_0 = 4.163, \omega_1 = 0.067, \omega_2 = 18.553$</td>
<td>M1 vs. M3</td>
<td>592.996</td>
<td>&lt;0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M2 vs. M3</td>
<td>17.919</td>
<td>&lt;0.000</td>
</tr>
<tr>
<td>M7</td>
<td>$p = 0.008, q = 0.068$</td>
<td>M7 vs. M8</td>
<td>573.963</td>
<td>&lt;0.000</td>
</tr>
<tr>
<td>M8</td>
<td>$p = 0.017, q = 0.107$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p_1 = 0.021, \omega_1 = 12.662$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The maximum likelihood analysis of the selection pressures acting on the porB2 and porB3 alleles provided strong evidence for positive selection (tables 2 and 3). For both porB2 and porB3, the best-supported M3 model estimated $\omega$ ($d_3/d_2$) parameters $\approx 1$, indicative of strong positive selection. Similarly, the M8 model, which could incorporate positive selection, was significantly favored over the M7 model which did not, and estimated $\omega \approx 1$ for both porB2 and porB3. The strength of the inferred selection pressures acting on both porB2 and porB3 was also striking, as was the similarity in selection pressures between these allele classes. For the M3 model, two classes of positively selected sites were apparent. In the case of porB2 ~4.5% of sites fell into a relatively weakly positively selected class, where $\omega_0 = 4.163$, whereas ~1.1% of sites are seemingly subject to very strong positive selection with $\omega_2 = 18.553$. The remaining 94% of sites were highly conserved ($\omega_1 = 0.067$). A similar distribution of sites was apparent in porB3. Here, ~4.6% of sites had an $\omega$ value of 3.229, whereas 0.7% of sites were subject to much stronger positive selection pres-

### Table 3

Parameter Estimates for the Maximum Likelihood Analysis of Selection Pressures Acting on the porB3 Gene of Neisseria meningitidis

<table>
<thead>
<tr>
<th>Model</th>
<th>Site Categories ($\omega$)</th>
<th>Likelihood Test</th>
<th>$\chi^2$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>M0</td>
<td>$\omega = 0.624$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M1</td>
<td>$p_0 = 0.927, p_1 = 0.073$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M2</td>
<td>$p_0 = 0.883, p_1 = 0.103, p_2 = 0.014$</td>
<td>M0 vs. M2</td>
<td>1102.683</td>
<td>&lt;0.000</td>
</tr>
<tr>
<td></td>
<td>$\omega_2 = 10.210$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M3</td>
<td>$p_0 = 0.041, p_1 = 0.952, p_2 = 0.007$</td>
<td>M0 vs. M3</td>
<td>1109.445</td>
<td>&lt;0.000</td>
</tr>
<tr>
<td></td>
<td>$\omega_0 = 3.229, \omega_1 = 0.033, \omega_2 = 13.923$</td>
<td>M1 vs. M3</td>
<td>440.278</td>
<td>&lt;0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M2 vs. M3</td>
<td>6.762</td>
<td>0.034</td>
</tr>
<tr>
<td>M7</td>
<td>$p = 0.008, q = 0.072$</td>
<td>M7 vs. M8</td>
<td>439.640</td>
<td>&lt;0.000</td>
</tr>
<tr>
<td>M8</td>
<td>$p = 0.008, q = 0.061$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p_1 = 0.014, \omega_1 = 10.147$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FIG. 1.—Positive selection among PorB2 and PorB3 sequences. Translated (a) porB2-5 and (b) porB3-2 sequences are used as examples of PorB2 and PorB3 amino acid sequences. The locations of the putative surface-exposed loop sequences (I–VIII) are indicated. Amino acid residues that are subject to positive selection (probability \( \omega = 0.95 \)) are shown in bold type. Sites under strong selection (\( \omega \geq 18.553 \) and 13.923 for PorB2 and PorB3, respectively) are marked below the sequence with black blocks, and sites under weak selection (\( \omega \geq 4.163 \) and 3.229 for PorB2 and PorB3, respectively) are denoted with white blocks.

Bayesian methods were used to identify the sites with the highest probability of falling into the positively selected class under the M3 model (fig. 1). When these selected sites were mapped onto the three-dimensional structural models (fig. 2), it was clear that the majority fell within regions predicted to encode surface-exposed loop regions. Among PorB2 proteins there was evidence of weak or strong positive selection at multiple sites in all surface loops except putative loop II. Four residues in putative loop IV were shown to be under strong selection, six of the 23 amino acids in putative loop V were positively selected and all six selected sites in putative loop VI were under strong selection for amino acid change. For loops IV and V, there were nonselected residues at the center of the loop region, flanked by residues that were much more variable in sequence. Five selected sites were also identified within the structurally constrained membrane-spanning regions of PorB2. At position 39, the replacement of leucine with phenylalanine was always observed with a change at position 42 from methionine to isoleucine or valine. Among PorB3 sequences, most positively selected sites were located in putative surface loops I, V, VI, and VII. Seven sites in loop I were under positive selection compared with only one site in loop I of PorB2. No selected sites were identified in putative loop III, whereas three weakly selected sites were identified within the largely conserved membrane-spanning regions of PorB3.

Finally, because the maximum likelihood approach of Yang et al. (2000) explicitly assumes a phylogenetic tree when estimating selection pressures, we attempted to assess whether high rates of recombination could have produced a false-positive signal for positive selection. We therefore repeated the analysis assuming that sequences were linked by a “star” phylogeny, where the lineages diverge simultaneously from a single root node, thereby removing the affect of phylogenetic history. For both PorB2 and PorB3, this analysis again produced significant evidence for positive selection \( (P < 0.000 \) for M3 vs. M2 and M8 vs. M7) and for extremely high \( \omega \) values at the selected sites—maximum values of 75.052.
Evolution of Meningococcal PorB 1691

Fig. 2.—Ribbon diagrams of models for PorB2 (top) and PorB3 (bottom) with superposition of residues subject to positive selection. Residues under strong selection are shown in red and residues under weak selection are shown in yellow. The diagrams were produced using Molscript (Kraulis 1991).

and 11.305, respectively, under the M3 model, with generally the same sites falling into the positively selected class as in the original analysis (full results available from authors on request).

Discussion

Here phylogenetic and structural modeling techniques have been used to reveal evolutionary processes acting on an established, but controversial, vaccine candidate, the meningococcal PorB protein. Maximum likelihood analysis of selection pressures provided powerful evidence for adaptive evolution of the porB genes of N. meningitidis. In both porB2 and porB3, codons were identified that had been subject to very strong selection pressure ($d_s/d_\epsilon$ of 18.533 and 16.221, respectively). These values were some of the highest seen in any gene studied to date, and similar to the values observed in studies of the intrahost evolution of the HIV-1 env gene, often seen as a paradigm of immune driven positive selection (Nielsen and Yang 1998; Zanotto et al. 1999). The fact that positive selection was still observed at these same sites after removing the effect of phylogenetic history by assuming a star phylogeny indicated that the analysis was unlikely to have been greatly biased by recombination. Mapping of these sites onto refined models of the protein structure of PorB confirmed that the majority of selected sites were located in regions of the protein predicted to be exposed to the host immune system. These results contradict those obtained previously (Smith, Maynard Smith, and Spratt 1995), which provided no evidence for positive selection in the putative loop regions of meningococcal PorB proteins, and in doing so highlight the need to consider each codon separately within amino acid alignments.

In addition, the analysis also revealed that although the PorB2 and PorB3 proteins were subject to very similar selection pressures they exhibited different distributions of positively selected sites. These findings were supported by flow cytometry analysis of live meningococci which demonstrated that some PorB3 variants were not easily accessible for antibody binding (Michaelsen et al. 2001), possibly because of shielding or due to PorB3 extracellular loops being shorter than those of PorB2, so that fewer residues were subject to intense immune selection. In contrast, the gonococcal porins PIA and PIB showed no significant difference in the distribution of selected sites despite reported differences in selection intensities between the two homology groups (Posada et al. 2000). Posada et al. therefore concluded that epidemiological differences between gonococci expressing PIA or PIB proteins were responsible for differences in selection rather than structural differences in the proteins. In meningococci, however, it is possible that the differences in the lengths of the surface loops of PorB2 and PorB3 porins are sufficient to affect the conformation and structure of epitopes presented by these proteins and that this will determine which sites are exposed to selective pressure from the host responses. Furthermore, although there may be some epidemiological differences between meningococci expressing PorB2 and PorB3 proteins, there is no evidence from these data that this has led to concomitant selective differences, although additional studies are required to test this hypothesis further.

The conformational effects of particular amino acid substitutions on PorB structure remain difficult to determine. The homology models used were useful for examining the approximate disposition of residues in space and could provide insights not readily apparent from a sequence alignment. The crystal structure of Omp32 from C. acidovorans was used as the basis for the homology models of PorB2 and PorB3; this is a closer homolog to the Neisseria porins than the E. coli porin crystal structures that were used previously (Derrick et al. 1999). But the size of most of the external loop regions precluded an accurate estimation of their conformations by standard homology modeling techniques, and the location of selected residues shown in external loop regions in figure 2 were approximate.
Figures 1 and 2 illustrate that the strongly and weakly selected sites are not distributed evenly across the loop regions. The presence of conserved residues at the apices of the PorB2 variable loops I, IV, and V suggested that these amino acids were not exposed to the host immune response because of protein folding, or perhaps they fulfill an important role retaining the surface loop structure. For example, the L2 loop in the OmpF protein of E. coli contributes to the stabilization of the porin trimer (Phale et al. 1998) and residues in other loop regions could play analogous roles in stabilizing the protein, using loop-loop interactions. Indeed, although the portions of polypeptide chain joining the ends of the β-barrel strands are frequently referred to as loops, they are likely to contain regions of regular secondary structure, as is seen in other porin structures (Koebnik, Locher, and Van Gelder 2000), and this would place constraints on the sequence variation within these regions in PorB2 and PorB3 proteins. These observations are borne out by the difficulty in mimicking PorB epitopes with linear peptides (Zapata et al. 1992) and by sequence variation identified among serologically similar PorB antigens (Urwin et al. 1998a, 1998b).

Although most residues within the transmembrane β-strands of PorB2 were highly conserved, five positively selected sites were located among these structural regions. According to the structural model, two of these sites (positions 39 and 43 in fig. 1a) were located close to one another on opposite sides of a β-turn and were identified as positively selected with >99% probability (although falling into the weakly rather than the strongly positively selected class). A leucine residue at position 39 was invariably accompanied by a methionine residue at position 43, whereas the phenylalanine residue at position 39 was accompanied by the smaller hydrophobic residues valine or isoleucine at position 43. Within the limitations of the homology model, the side chains of the two residues may be in contact, suggesting that these are compensatory mutations. Elsewhere, the reasons for strong positive selection within the β-barrel region were more difficult to discern. The PorB2 protein contained a number of mutations within the L3 loop, which folds back into the center of the barrel creating a constricted channel for the passage of low molecular weight solutes (Zeth et al. 2000). Residues within the L3 loop were unlikely to be subject to immune selection but are probably involved in ion selectivity: the presence of a number of mutations which resulted in changes in charge within this region of the protein are consistent with this idea (Bauer et al. 1989; Benz et al. 1989; Saint et al. 1996; Schirmer and Phale 1999).

Despite the extensive evidence for recombinational reassortment in PorB proteins, (Vázquez et al. 1995; Cooke et al. 1998; Derrick et al. 1999), this has not been so frequent as to remove all phylogenetic signals. In contrast, congruence tests between different housekeeping genes from N. meningitidis, subject to strong purifying selection, revealed that most gene trees were no more similar to one another than expected by chance (Holmes, Urwin, and Maiden 1999). Overall, the diversity of meningococcal protein antigens explored here exhibits features which phylogenetic and structural models can exploit in the elucidation of the function and vaccine potential of these molecules. The insights obtained can complement data from both genomic studies and experimental studies of immunogenicity in man and animals to provide more complete information on the interactions of pathogen proteins with the hosts and its immune system.

Acknowledgments

This work was supported by The Wellcome Trust (047072), The Royal Society, The Lister Institute, and the United Kingdom Public Health Laboratory Service. We thank Dr. Z. Yang for useful advice regarding CODEML.

LITERATURE CITED


**Sierra, G. V. G., H. C. Campa, N. M. Varcael, I. L. Garcia, P. L. Izquierdo, P. F. Sotolongo, G. V. Casa-


Tommassen, J., P. Vermeij, M. Struyvè, R. Benz, and J. T. Poolman. 1990. Isolation of Neisseria meningitidis mutants deficient in class 1 (PorA) and class 3 (PorB) outer membrane proteins. Infect. Immun. 58:1355–1359.


Peer Bork, reviewing editor

Accepted May 6, 2002