Sequence Evolution of the \textit{porB} Gene of \textit{Neisseria gonorrhoeae} and \textit{Neisseria meningitidis}: Evidence of Positive Darwinian Selection

N. H. Smith, J. Maynard Smith, and B. G. Spratt
School of Biological Sciences, University of Sussex

Protein 1 (PI) is a major porin of \textit{Neisseria gonorrhoeae} and \textit{Neisseria meningitidis} and is encoded by a single locus, \textit{porB}. Alleles of the \textit{porB} locus of \textit{N. gonorrhoeae} are assigned to two homology groups, PI(A) and PI(B), on the basis of immunological and structural similarity. In a like manner, alleles of the \textit{porB} locus of the closely related bacterium, \textit{N. meningitidis}, are allocated into class 2 and class 3 homology groups. An individual strain of \textit{N. gonorrhoeae} or \textit{N. meningitidis} expresses either one or other of these porin homology groups but never both, and the antigenic reactions of these highly diverse outer membrane proteins form part of the \textit{N. gonorrhoeae} and \textit{N. meningitidis} serotyping schemes.

A comparison of the number of synonymous and nonsynonymous substitutions per site between the two most divergent alleles of each of these four groups of \textit{porB} alleles shows that PI(A) alleles have accumulated significantly more nonsynonymous substitutions per site than synonymous substitutions. In contrast the distribution of synonymous and nonsynonymous substitutions between alleles of class 2 and class 3 porins are not significantly different from random. We localize the regions of the PI(A) alleles with an excess of amino acid changes to the surface-exposed loops of these outer membrane proteins and suggest that positive Darwinian selection for diversity, driven by the human immune system, can most easily explain the allelic polymorphism and the pattern of synonymous and nonsynonymous substitutions.

Introduction

It has been proposed that the high levels of polymorphism found in foreign-antigen-recognizing proteins of vertebrates are maintained by balancing selection, and this suggestion is supported by a comparison of the pattern of synonymous and nonsynonymous substitutions in the antigen recognition regions of these genes (Hughes and Nei 1988, 1989; Tanaka and Nei 1989). If the human immune system is in an evolutionary "arms race" with pathogens, then it is to be expected that similar selection pressures will be evident among the immunexposed proteins of pathogens. Indeed, an accelerated rate of nonsynonymous nucleotide substitutions, relative to synonymous substitutions, has been demonstrated for the major surface proteins of the malarial parasite, \textit{Plasmodium falciparum} (Hughes 1991, 1992). We here present evidence that the major outer membrane protein, PI, of the bacterial pathogen \textit{Neisseria gonorrhoeae} is also under positive Darwinian selection for diversity in response to pressure from the human immune system.

The true neisseriae are a closely related group of Gram-negative diplococci and are primarily commensals of the mucous membranes of mammals. Two species in particular, \textit{N. meningitidis} (the meningococcus) and \textit{N. gonorrhoeae} (the gonococcus), are important pathogens of humans. The meningococcus is commonly carried in the upper respiratory tract of humans and occasionally invades the host and causes septicaemia and meningitis (meningococcal disease). \textit{Neisseria gonorrhoeae}, on the other hand, is the causative agent of the human venereal disease, gonorrhoea. Humans are the only hosts of the gonococcus which typically causes an infection of the urogenital tract that is eventually limited, presumably by the immune system. Little acquired immunity follows infection, and repeat infection can be common (Ison 1988).

Strains of \textit{N. meningitidis} are virtually identical by DNA-DNA hybridization, and isolates of \textit{N. gonorrhoeae} form a closely related, but distinct, population (Guibourdenche et al. 1986). The application of multilocus enzyme electrophoresis and nucleotide sequencing studies to strains of \textit{N. meningitidis} and \textit{N. gonorrhoeae} has confirmed the division of these bacteria into two
distinct species and has demonstrated high levels of recombination within each species but limited recombination between the species (Maynard Smith et al. 1993; Vázquez et al. 1993).

Protein 1 (PI) encoded by the porB locus, is the major outer membrane protein of the Neisseria and functions as a porin allowing the passage of small molecules through the outer membrane. The extensive antigenic variability of this protein is part of the N. gonorrhoeae and N. meningitidis serotyping schemes involving panels of monoclonal antibodies. Alleles of the porB locus of N. gonorrhoeae have been assigned to two homology groups based on close sequence and immunological relationships and are designated as either PI(A) or PI(B). Alleles within each group are much more similar to each other than they are to members of the other group, and individual strains of N. gonorrhoeae express either a PI(A) or a PI(B) allele. Similarly, alleles of the porB gene of N. meningitidis are allocated into class 2 and class 3 homology groups.

PorB is an excellent candidate for displaying the effects of positive selection; the protein is constitutively expressed at high levels, is required by all meningococci and gonococci, is surface exposed, and elicits a strong immune reaction during infection (Lison 1988). Furthermore, the porB locus does not manifest on/off phase variation or gene conversion mechanisms that change the expression or antigenicity of other surface antigens at high frequency (Seifert 1992). Because individual cells do not have molecular mechanisms to escape immune surveillance of porB, any antigenic variants that arise by mutation during an infection will, presumably, be subject to intense selection.

A model of the structure of neisserial porins predicts that eight regions (loops 1–8) are surface exposed (van der Ley et al. 1991). This model is based on the identification of highly conserved regions capable of forming amphipathic β-sheets that are assumed to be membrane spanning, and highly variable regions of predominantly hydrophilic amino acids that are presumably surface exposed. Sequencing of many different neisserial genes has confirmed that the general structure of these porins consists of nine conserved regions separated by eight regions that are highly variable in both amino acid sequence and length (Carbonetti and Sparling 1987; Gotschlich et al. 1987; Carbonetti et al. 1988; Butt et al. 1990; Elkins et al. 1992; Feavers et al. 1992; Ward et al. 1992; Mee et al. 1993). Epitope mapping of serotype-specific monoclonal antibodies has provided experimental evidence for the surface exposure of several of the predicted loops of the PI(A) proteins (Elkins et al. 1992; Mee et al. 1993).

To examine the selective forces operating on these proteins, we have analyzed the substitutions between the two most divergent alleles of the PI(A), PI(B), class 2 and class 3 homology groups, and we show that the regions of the PI(A) alleles of N. gonorrhoeae that encode surface-exposed loops exhibit a pattern of synonymous and nonsynonymous substitutions that is compatible with positive Darwinian selection for diversity. In contrast, the porB alleles of N. meningitidis do not show clear evidence of diversifying selection, presumably because of differences in the immune response to these two organisms.

DNA Sequences Analyzed

The porin sequences analyzed consist of 11 sequences of the gonococcal PI(A) homology group (Carbonetti and Sparling 1987; Elkins et al. 1992; Mee et al. 1993; a porin from strain P118, GenBank accession number X58073); 3 sequences of the gonococcal PI(B) homology group (Gotschlich et al. 1987; Carbonetti et al. 1988; Butt et al. 1990); 5 partial sequences of the class 2 homology group of Neisseria meningitidis (Feavers et al. 1992); and 4 full sequences of meningococcal class 3 porB alleles (Ward et al. 1992). The 19-codon signal sequence was excluded from further analyses. The position of the eight externally exposed loop regions was taken from the alignment of Ward et al. (1992), and all other segments of the gene, encoding-membrane-spanning and internal loop regions, were designated as nonloop regions.

Results

General Properties of porB Alleles

Alignment of entire porin sequences from different homology groups is problematic because of the extensive differences in size and composition of regions identified as loops. The sequences of the nonloop regions can, however, be simply aligned without insertions or deletions. Within the class 3 and PI(A) homology groups, the alignment of entire sequences is straightforward and does not require insertions or deletions. Alignment of the PI(B) alleles requires a two-codon indel in loop 6, and, because incorrect alignment can seriously interfere with our analysis, the entire sequence of loop 6 was excluded from all further analyses of PI(B) alleles. The two most divergent sequences of the class 2 alleles can be aligned without indels, even though one other allele in the homology group requires insertion of two large gaps in loop regions for optimal alignment.

Aligned sequences from each homology group were compared, and the two most dissimilar alleles in each homology group were chosen for further analyses. The most divergent PI(A) alleles were derived from N. gonorrhoeae strains SU95 and PI-35 (GenBank accession L19963 and L19965), PI(B) alleles from N. gonorrhoeae
strains R10 and MS11 (GenBank accession J03017 and X54024), class 2 alleles from N. meningitidis strains M982 and 2996 (GenBank accession X67938 and X67939), and class 3 alleles from N. meningitidis strains M981 and MC58 (GenBank accession X65531 and X65532).

Synonymous changes in the nonloop regions of the most divergent alleles of the four homology groups were used to demonstrate the relationships shown in figure 1. The high bootstrap value of the branch leading to the PI(A) and class 3 homology groups confirms that alleles of these homology groups, derived from N. gonorrhoeae and N. meningitidis, respectively, are more similar in sequence to each other than they are to the PI(B) alleles of N. gonorrhoeae or the class 2 alleles of N. meningitidis (Ward et al. 1992; Suker et al. 1993). All the porB alleles show evidence of bias in the use of synonymous codons. For example, the PI(A) allele of strain SU95 uses an AAA Lysine codon 19 times, but the AAG codon is not used at all. In general, the codon biases in these highly expressed genes are similar to those in the codon usage table of N. gonorrhoeae based on 11 highly expressed gonococcal genes (West and Clark 1989).

Calculation of Synonymous and Nonsynonymous Substitution Rates

Evidence of the selective forces acting on a gene as two alleles diverge can be gathered by estimating the number of synonymous ($d_s$) and nonsynonymous ($d_N$) changes per site. On the assumption that synonymous and nonsynonymous mutations per site are equally frequent, then the following generalizations can be made. If synonymous and nonsynonymous substitutions are selectively equal, then the ratio $d_s/d_N$ is expected to be close to unity; but if, as is the case in most comparisons between bacterial alleles (Sharp 1991), purifying selection is preferentially eliminating amino acid changes, then the $d_s/d_N$ ratio will be significantly higher. A higher rate of nonsynonymous substitutions compared with synonymous substitutions ($d_s/d_N < 1$) is evidence of positive selection for amino acid change.

The number of synonymous substitutions per synonymous site and the number of nonsynonymous substitutions per site were calculated by the method of Nei and Gojobori (1986) using the MEGA suite of computer programs (Kumar et al. 1993). The correction of Jukes and Cantor (1969) for multiple substitutions, which would be minimal for these comparisons, has not been applied because it tends to inflate the number of substitutions as they are partitioned into smaller regions.

A simple method to calculate the significance of differences in the rates of nonsynonymous and synonymous substitutions is to compare only the two most divergent alleles of each homology group. This method makes less efficient use of the data than the calculation of standard errors (Nei and Jin 1989), but it makes fewer assumptions: since the data are sufficient to demonstrate selection with a high level of significance, the loss of efficiency is not important. The data analyzed are summarized in table 1. All the alleles of porB available had been chosen for sequencing because they represented distinct serotypes and thus were known to have amino acid differences. Preselection of alleles may introduce a bias toward amino acid substitution in the data set, but, by only analyzing the two most divergent sequences in each data set, this small bias has been minimized.

Types of Comparisons

Three types of comparison have been made, as summarized in table 2.

A. Rates of Synonymous and Nonsynonymous Change

The number of synonymous changes per site is compared to the number of nonsynonymous changes per site, in a given gene or region of a gene, for a given pair of strains. It is usual to find higher rates of synonymous substitutions, because of the stronger purifying selection on nonsynonymous mutations (Sharp 1991). In the complete absence of selection, we expect the two rates to be approximately equal. If there is a higher rate of nonsynonymous substitution, this is evidence for positive Darwinian selection on amino acid composition.
Table 1
The Numbers of Substitutions in Synonymous (syn.) and Nonsynonymous (nonsyn.) Sites, the Number of Each Type of Site (sites), and the Ratio of Substitutions per 100 Sites (per 100 sites)

<table>
<thead>
<tr>
<th></th>
<th>TOTAL*</th>
<th>LOOP REGIONS</th>
<th>NONLOOP REGIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>PI(A) alleles:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Substitutions . . .</td>
<td>3.5</td>
<td>42.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Sites . . . . . . . . . .</td>
<td>221.2</td>
<td>702.8</td>
<td>80.7</td>
</tr>
<tr>
<td>Per 100 sites . . . . . . .</td>
<td>1.6</td>
<td>6.0</td>
<td>0.6</td>
</tr>
<tr>
<td>PI(B) alleles:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Substitutions . . .</td>
<td>5.0</td>
<td>10.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Sites . . . . . . . . . .</td>
<td>226.8</td>
<td>706.2</td>
<td>84.8</td>
</tr>
<tr>
<td>Per 100 sites . . . . . . .</td>
<td>2.2</td>
<td>1.4</td>
<td>1.2</td>
</tr>
<tr>
<td>Class 2 alleles:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Substitutions . . .</td>
<td>11.0</td>
<td>42.0</td>
<td>9.0</td>
</tr>
<tr>
<td>Sites . . . . . . . . . .</td>
<td>231.3</td>
<td>737.7</td>
<td>105.0</td>
</tr>
<tr>
<td>Per 100 sites . . . . . . .</td>
<td>4.8</td>
<td>5.7</td>
<td>8.6</td>
</tr>
<tr>
<td>Class 3 alleles:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Substitutions . . .</td>
<td>20.25</td>
<td>39.75</td>
<td>13.25</td>
</tr>
<tr>
<td>Sites . . . . . . . . . .</td>
<td>223.5</td>
<td>712.5</td>
<td>82.25</td>
</tr>
<tr>
<td>Per 100 sites . . . . . . .</td>
<td>9.1</td>
<td>5.6</td>
<td>16.1</td>
</tr>
</tbody>
</table>

* The number of substitutions is given for the entire gene (Total) and partitioned between loop and nonloop regions, of the most divergent alleles of the four homology groups of perB.

B. The Absolute Numbers of Synonymous and Nonsynonymous Changes

Comparisons of kind A are appropriate only for a given pair of strains, since the number of changes per site will increase with the time of separation. However, within limits (Maynard Smith 1994), if the ratio of the absolute numbers of synonymous to nonsynonymous changes is lower in a gene from one pair of strains, than in a gene from another pair of strains, this is evidence for selection even if the ratios are calculated for different pairs of strains.

C. Rates of Nonsynonymous Changes in Different Regions

The number of nonsynonymous changes per site is compared for different regions of a gene, in the same pair of strains. A higher rate of nonsynonymous change in one region can imply either relaxed purifying selection on amino acids or stronger selection for change in that region.

Substitutions in PI(A) Alleles of Neisseria gonorrhoeae

A total of 46 substitutions were recorded between the two most divergent PI(A) alleles and were distributed as 42.5 nonsynonymous substitutions in 702.8 sites and 3.5 synonymous substitutions in 221.2 sites (table 1). There is a significantly higher rate of nonsynonymous than synonymous substitution between these two alleles (comparison 1, table 2). The $d_s/d_N$ ratio of these alleles is 0.27; nonsynonymous substitutions have been more than three times as common as synonymous substitutions during the divergence of these two alleles. This evidence for selection for change in amino acid composition is confirmed by comparing the PI(A) alleles with a housekeeping gene, glnA. The most distantly related of 11 alleles of the glnA gene sequenced from strains of N. gonorrhoeae have 17 synonymous substitutions in 297 synonymous sites and only 9 nonsynonymous substitutions in 960 nonsynonymous sites (N. H. Smith, J. Zhou, and B. G. Spratt, unpublished data): the proportion of synonymous to nonsynonymous substitutions is significantly greater in glnA than in the PI(A) alleles (comparison 4, table 2). In the penA gene Maynard Smith (1994), summing over all comparisons that have not been subject to interspecies recombination, found 21 nonsynonymous and 91 synonymous differences between related strains of Neisseria. Again, the proportion of synonymous to nonsynonymous changes is much greater than in the PI(A) alleles; the $d_s/d_N$ ratio of glnA is 6.1 and for penA is approximately 14.

Substitutions in Loop and Nonloop Regions of PI(A) Alleles

To elucidate the regions of PI(A) that have an excess of amino acid substitution, we allocated the
Table 2
Comparison of Nonsynonymous and Synonymous Substitutions

A. Rates of Nonsynonymous and Synonymous Change

<table>
<thead>
<tr>
<th>Species*</th>
<th>Genes</th>
<th>Region</th>
<th>Substitutions/site</th>
<th>$\chi^2$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 NG</td>
<td>Pl(A)</td>
<td>Total</td>
<td>1.55%:6.05%</td>
<td>7.09</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>2 NM</td>
<td>Class 2</td>
<td>Total</td>
<td>4.8%:5.7%</td>
<td>0.3</td>
<td>NS</td>
</tr>
<tr>
<td>3 NM</td>
<td>Class 3</td>
<td>Total</td>
<td>9.1%:5.6%</td>
<td>3.44</td>
<td>NS</td>
</tr>
</tbody>
</table>

B. Absolute Numbers of Synonymous and Nonsynonymous Changes

<table>
<thead>
<tr>
<th>Species*</th>
<th>Genes</th>
<th>Region(s)</th>
<th>Syn./nonsyn.</th>
<th>$\chi^2$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 NG</td>
<td>Pl(A):glnA</td>
<td>Total</td>
<td>0.08:1.89</td>
<td>27.2</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>5 NM</td>
<td>porB</td>
<td>Loops:nonloops</td>
<td>0.28:2.25</td>
<td>12.7</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>6 NG:NM</td>
<td>porB</td>
<td>Loops</td>
<td>0.035:0.28</td>
<td>7.9</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>7 NG:NM</td>
<td>porB</td>
<td>Nonloops</td>
<td>0.7:2.25</td>
<td>2.3</td>
<td>NS</td>
</tr>
</tbody>
</table>

C. Rates of Nonsynonymous Change in Different Regions

<table>
<thead>
<tr>
<th>Species*</th>
<th>Genes</th>
<th>Region(s)</th>
<th>Substitutions/site</th>
<th>$\chi^2$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 NG</td>
<td>Pl(A)</td>
<td>Loops:nonloops</td>
<td>11.9%:2.1%</td>
<td>28.1</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>9 NM</td>
<td>Class 2</td>
<td>Loops:nonloops</td>
<td>10.9%:0.8%</td>
<td>35.2</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>10 NM</td>
<td>Class 3</td>
<td>loops:nonloops</td>
<td>13.2%:0.2%</td>
<td>55.3</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

* NG, Neisseria gonorrhoeae; NM, N. meningitidis.

Nonsynonymous substitutions between those regions encoding the eight surface-exposed loops and those encoding the nine nonloop regions defined by van der Ley et al. (1991). There are significantly more nonsynonymous substitutions in loop sequences of the PI(A) alleles than in nonloop regions (comparison 8, table 2). This observation is not unexpected because one of the criteria used to define the loop regions was the variability of these regions (van der Ley et al. 1991). Nevertheless, regions that have been identified as surface-exposed loops are evolving much faster by amino acid substitution than the membrane spanning nonloop regions of these porins. The number of synonymous differences in loop and nonloop regions of the PI(A) alleles are too few to carry out a meaningful statistical analysis; however, it is clear from table 1 that the accelerated rate of nonsynonymous change in the loop regions of PI(A) alleles is not accompanied by an accelerated rate of synonymous change.

The significance of nonsynonymous substitutions in nonloop regions of this gene cannot be determined, primarily because the exact extent of the loop regions is not well characterized. For example, eight of the nine amino acid substitutions in nonloop regions of the PI(A) alleles are within five codons of loop sequences and may be partially surface exposed or influenced by the loop regions.

The number of substitutions between the two most divergent PI(B) alleles of Neisseria gonorrhoeae is not sufficient for statistical analysis.

PorB alleles of Neisseria meningitidis

In contrast to PI(A) alleles of N. gonorrhoeae, the rates of synonymous and nonsynonymous substitutions between the two most divergent alleles of the class 2 and class 3 homology groups of N. meningitidis are not significantly different (comparisons 2 and 3, table 2). However, combining data for class 2 and class 3 alleles, the ratio of synonymous to nonsynonymous substitutions is significantly lower in loop than in nonloop regions (comparison 5, table 2), suggesting that there has been selection for amino acid change in the loop regions. This conclusion is confirmed by the higher rates of nonsynonymous change per site in the loop regions (comparisons 9 and 10, table 2). The absolute numbers of nonsynonymous substitutions are 3.5 times more frequent than synonymous substitutions in loop regions, whereas in nonloop regions the distribution is reversed; synonymous substitutions are more numerous than nonsynonymous substitutions.
Comparison between porB Genes of Neisseria gonorrhoeae and Neisseria meningitidis

Differences in the selective pressures operating on the porB genes of N. gonorrhoeae compared with N. meningitidis can be detected by combining the numbers of synonymous and nonsynonymous substitutions from both homology groups of each species. For loop regions of the combined PI(A) and PI(B) data set from N. gonorrhoeae, the distribution of synonymous (1.5) to nonsynonymous (42.5) substitutions is significantly different from the distribution of synonymous (22.25) to nonsynonymous (77.75) substitutions in the loop regions of the class 2 and class 3 alleles of N. meningitidis (comparison 6, table 2). During the divergence of porB alleles of N. gonorrhoeae, nonsynonymous substitutions in loop regions have been significantly more frequent, relative to synonymous substitutions, than in the divergence of the porB alleles of N. meningitidis. In contrast, there is no significant evidence for a difference in the ratio of synonymous to nonsynonymous differences in the nonloop regions of the two species (comparison 7, table 2).

Discussion

We have shown that PI(A) alleles, in contrast to housekeeping genes of Neisseria gonorrhoeae, have a significantly higher rate of nonsynonymous substitution than synonymous substitution and that this effect is most marked in the regions encoding surface-exposed loops of these outer membrane proteins. Elevated substitution rates in one region of a gene relative to another may be observed as a result of relaxed selection, but relaxed selection alone will not preferentially elevate the rate of nonsynonymous substitutions over synonymous substitutions as we have observed in the loop regions of the PI(A) alleles. Positive selection for amino acid changes that increase the functional fitness of a protein may enhance the rate of nonsynonymous substitutions over synonymous changes but will not generate the extreme polymorphism found in these alleles. There is an alternative hypothesis that would explain a higher \( d_s \) than \( d_N \) without assuming positive selection. Imagine a gene in which codon usage is strongly conserved, because of high expression, but containing regions (such as those encoding spacer regions between domains) in which any amino acid substitution was accepted. In the spacer regions, synonymous mutations would be removed by purifying selection, but amino acid changes would be accepted, provided that the new amino acid had the correct codon. For such a protein, \( d_s \) would be greater than \( d_N \), and codon bias would be the same in spacer and nonspacer regions. We have determined for the PI(A) allele of strain PI-35 that the codon bias in the loop regions is different from the codon bias in the nonloop regions. For example, in the loop regions the amino acid Asn is coded by AAT six times and by AAC four times; however, in the nonloop regions the six occurrences of Asn are coded by AAC. We conclude that amino acid changes have occurred in the loop regions even if they involved the use of an unfavorable codon. We therefore reject the alternative hypothesis.

The nucleotide diversity (\( \pi \); Nei 1987, p. 256) of PI(A) alleles is 0.025 and is far in excess of the nucleotide diversity of housekeeping genes of N. gonorrhoeae (Vázquez et al. 1993; E. Feil and B. G. Spratt, unpublished data). The extensive nucleotide diversity of PI(A) alleles and the presence of two major groups of alleles at the porB locus of N. gonorrhoeae implies that some mechanism is positively selecting diversity at this locus against the homogenizing effect of drift or periodic selection. We therefore conclude that the pattern of synonymous and nonsynonymous substitutions in the PI(A) alleles of the porB gene of N. gonorrhoeae is the result of positive Darwinian selection for diversity and that this selection is acting primarily on the surface-exposed loop regions.

Positive selection can maintain allelic polymorphisms in a population for much longer than neutral polymorphisms (Takahata and Nei 1990), and at first sight the distribution of porB homology groups within these neisserial populations may be taken as evidence of polymorphisms that have survived speciation events. The close homology between the class 3 porins of N. meningitidis and the PI(A) porins of N. gonorrhoeae (fig. 1) is most easily explained as descent from a sequence present in the common ancestor of N. gonorrhoeae and N. meningitidis, implying that PI(B) and class 2 alleles, which apparently diverged much earlier, were also present in the common ancestral population. However, before this distribution of homology groups can be taken as evidence of an ancient polymorphism, with the associated conclusion that the polymorphism is maintained by balancing selection, it must be remembered that interspecies barriers to recombination are weak in bacteria, and the distribution of these alleles throughout these species may reflect recent recombination events rather than the prolonged maintenance of polymorphism. Recombination at highly selected loci is likely in Neisseria species for which evidence of interspecies transfer of genes has been well documented (Spratt et al. 1992; Zhou and Spratt 1992).

Although the PI(A) alleles of N. gonorrhoeae show evidence at the nucleotide level of diversifying selection, the selective forces operating on the porB alleles of N. meningitidis are not as clear. Compared with the rest of the gene, the loop regions of the class 2 and class 3 alleles show an accelerated rate of substitution, but, in contrast to the PI(A) alleles, the overall rate of synonymous and
nonsynonymous substitutions per site are equal. These observations can most easily be explained by the action of relaxed selection on the loop regions and purifying selection on the rest of the gene. However, it is more likely that the equal rate of synonymous and nonsynonymous substitutions in these genes is the result of the long-term interaction between weak positive selection for diversity, which reduces the $d_S/d_N$ ratio, and purifying selection, which elevates the relative rate of synonymous change.

The most likely mechanism selecting increased amino acid diversity of PI(A) is the human immune system, although the role of PI in adhesion and invasion may also play a part (Ison 1988). *Neisseria gonorrhoeae* is a sexually transmitted disease that relies on a small cohort of high-frequency transmitters to maintain itself in the human population, and models for the maintenance of this disease in the human population suggest that the basic reproductive rate of the organism ($R_0$) is dependent only on the transmission probability of the pathogen per partner, the length of the infectious state, and the number of partners per unit time (Anderson and May 1991). Epidemiological and serological evidence suggests that for the gonococcus, both the transmission probability and the length of the infectious state can be enhanced by changes in the antigenicity of the PI protein (Ison 1988; Plummer et al. 1989).

Our analyses show that the $d_s/d_N$ ratio of PI(A) alleles is most simply explained by positive Darwinian selection for diversity in response to selection by the human immune system. Diversifying selection, acting in a frequency-dependent manner, can also be invoked to explain the diversity of PI(A) alleles; however, the other three homology groups of porB analyzed here have not, apparently, been subject to the same intense selection for amino acid change. Less intense immune selection on meningococci than on gonococci may explain why the porB alleles of *N. meningitidis* do not show clear evidence at the nucleotide level of positive selection.

The limited data available also indicates that the $d_s/d_N$ ratio of the PI(B) alleles of *N. gonorrhoeae* is higher than the $d_s/d_N$ ratio of PI(A) alleles and suggests that PI(A) alleles have recently been subject to stronger selection. We note that although uncomplicated gonorrhoea is caused by isolates of either PI(A) or PI(B) serotypes, isolates from the blood during disseminated gonococcal infection are almost invariably strains bearing PI(A) alleles (Sandstrom et al. 1984). Similarly, isolates of *N. gonorrhoeae* bearing PI(B) alleles are over-represented in rectal isolates from male homosexuals (Morse et al. 1982). These epidemiological differences may result in differing selection pressures on PI(A) and PI(B) alleles, and as more nucleotide sequence data becomes available, it will be possible to determine whether the two gonococcal homology groups of porB are evolving under different selective pressures.

**Acknowledgment**

This research was supported by The Wellcome Trust.

**Literature Cited**


Paul M. Sharp, reviewing editor

Received September 4, 1994

Accepted December 7, 1994