No linkage or association between multiple sclerosis and the myelin basic protein gene in affected sibling pairs

N W Wood, P Holmans, D Clayton, N Robertson, D A S Compston

Abstract
Myelin basic protein was examined as a candidate gene for susceptibility to multiple sclerosis using two adjacent amplification fragment length polymorphisms (AmpFLPs), containing seven and six highly informative alleles respectively. No allelic association was found with multiple sclerosis, comparing 77 cases and 88 controls, and there was no evidence for linkage in 73 affected sibling pairs, using the methods of identity by descent and identity by state.

(J Neurol Neurosurg Psychiatry 1994;57:1191–1194)

Multiple sclerosis affects about one in 800 people in the United Kingdom and is a major cause of neurological disability in young adults. Family and twin studies implicate genetic and environmental factors in the aetiology; the number of genes conferring susceptibility is not known but recurrence risks in families indicate the involvement of more than one locus. Laboratory studies have implicated genes that regulate the immune response including those encoding class II histocompatibility products and variable regions of T cell receptor and immunoglobulin chains. The contribution that these make to susceptibility remains unclear.

Attention has therefore turned to structural genes of myelin as candidates for susceptibility in multiple sclerosis. Investigation of T cell specificity in samples from patients with multiple sclerosis implicates myelin basic protein as a potential autoantigen and the search for a target autoantigen has also been guided by animal studies.

Myelin basic protein comprises about 35% of the structural proteins of CNS myelin. It is localised to the major dense line and seems necessary for compaction of CNS myelin. The gene encoding myelin basic protein has been cloned and mapped to chromosome 18qter. It consists of seven exons and alternative splicing accounts for four isoforms which have been described. Specific mutations disrupt myelination in rodents.

Five studies of myelin basic protein polymorphisms in multiple sclerosis have been reported based either on population associations or linkage analysis in families with several affected members. Initial reports providing evidence both for a population association and linkage to polymorphisms of the myelin basic protein gene have not been confirmed. We studied linkage in 73 affected sibling pairs recruited from throughout the United Kingdom using a tetranucleotide repeat to the myelin basic protein gene. The AmpFLPs produced were used to assign haplotypes and analysis of identity by descent and state failed to show linkage. A comparison of allele frequencies between 77 patients with multiple sclerosis and 88 controls showed no significant differences.

Methods

PATIENTS
We identified 416 families with more than one member affected by multiple sclerosis, of whom about half contained affected sibling pairs. This report concerns the first 77 families in whom the diagnosis of multiple sclerosis was verified in the affected siblings as probable or definite disease by the Poser criteria. The age of affected subjects ranged from 20 to 73 (mean 46) years. We obtained samples of peripheral blood from affected members and as many as other family members as possible, so as to provide maximally informative pedigrees. This study forms part of a systematic analysis for linkage to several candidate gene loci in these families. The index from each family was used to provide a cohort of unrelated cases for comparison of allele frequencies with 88 white controls matched for age and sex and recruited from volunteer blood donors. The families were recruited from all health regions and neurology centres in the British Isles.

MICROSATELLITE MARKERS
A polymorphic (ATGG) repeat has been described, beginning at base pair 518, 5' of the human myelin basic protein gene; it contains a (TGGA) repeat shown to be polymorphic by Boylan and coworkers. We used a pair of primers to amplify two adjacent tetranucleotide repeats, (ATGG) and (TGGA), respectively. The forward primer is located upstream of the first repeat and the reverse primer binds to two sites; the first lies between the two repeats and the second is downstream from (TGGA).

We obtained two highly polymorphic AmpFLPs using these primers in a polymerase chain reaction. The largest encompasses both repeating sequences and varies in length from 208–232 bp. The second examines just the (ATGG) repeat and ranges in length from 122–142 bp.

To obtain the AmpFLPs, 5–10 ng of...
genomic DNA was amplified in the presence of 1.5 mM MgCl₂ and P₃₂P 5' end labelled primer. The cycling parameters were 94°C for five minutes followed by 27 cycles of 94°C for one minute, 55°C for one minute, and 72°C for one minute. An aliquot of the denatured product was then separated on a vertical 0.4 mm thick 6% polyacrylamide denaturing gel (Sequagel, National Diagnostics). Electrophoresis was performed for two to three hours at 120 W in 1 x Tris-borate EDTA. Autoradiography was performed for one to seven days with Kodak XAR-5 high speed film.

Population allele frequencies were obtained from 176 chromosomes for locus A and 172 chromosomes for locus B, using the panel of 88 normal controls.

Seven alleles were defined for locus A and six for locus B, providing highly informative markers; the polymorphic information contents were 0.80 and 0.69, respectively. Genotypes were established for the affected pairs and, where available, their parents and/or other siblings. Haplotypes were constructed from the two genotypes.

STATISTICAL METHODS
The data were analysed in an identity by descent comparison. Where unambiguous haplotype assignment was not possible, a correction was made for the probability of finding a particular genotype or haplotype given allele frequencies present in the normal population and the multiplex families. Identity by descent has advantages over other methods for linkage analysis. Penetrance and age of onset are not important as only affected people are studied, and no assumptions have to be made concerning the mode of inheritance. The expected rate of sharing under the null hypothesis is for random assortment between the genes of interest and the disease state, resulting in 1/4, 1/2, and 1/4 of pairs sharing zero, one, or two haplotypes, respectively. Evidence for linkage is provided by a significant bias towards sharing one or both genotypes.36-37

Identity by state analysis, in which the presence or absence of a particular allele is determined and no attempt is made to identify its parental origin was also carried out. This method takes account of parental homozygosity and the presence of a common allele, which may be present in both parents. These factors influence expected rates of sharing which must be calculated for each allele and are no longer 1/4, 1/2, 1/4 as in identity by descent.38 Identity by state is simple to perform, but lacks power and ignores data available from parents and other siblings.

The figure illustrates the inheritance of these polymorphisms, demonstrating identity by descent and identity by state. The father is homozygous at locus A but locus B increases the available information. Paternal heterozygosity can be defined at this locus, allowing genotypes to be assigned and showing that the affected siblings inherited one genotype identity by descent.

Results
Table 1 summarises the results of the identity by descent analysis; no alleles could be assigned in one affected person from four families. The observed frequencies of sharing zero, one, or two haplotypes at locus A in the remaining 73 pairs were 0.29, 0.48, and 0.23; corresponding frequencies at the B locus were 0.25, 0.50, and 0.25. Neither deviate from the expected frequencies of 0.25, 0.50, and 0.25.

The number of members amongst the 73 families expected to share zero, one, or two haplotypes identical by state was 8.6, 38.6, and 25.9 for locus A; these did not differ from the observed numbers of 10, 36, and 27 (table 2). For locus B, the observed numbers sharing zero, one, or two haplotypes were eight, 40, and 25, compared with the expected numbers of 5.9, 40 and 30.1.

The frequency of alleles identified at locus A or B in the index case from 77 sibling pairs did not differ significantly from the frequencies found in 88 normal controls after correction for multiple comparisons (table 3).

Discussion
Genetic factors are involved in the aetiology of multiple sclerosis. The HLA region has been most studied, mainly in association studies,39 and to date no firm evidence for linkage has been reported.39 It is likely that the contribution made by this region is small and insufficient to account for genetic susceptibility in

Table 1 Identity by descent in 73 affected sibling pairs

<table>
<thead>
<tr>
<th>Haplotypes shared</th>
<th>Locus A</th>
<th>Locus B</th>
<th>Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.29</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>1</td>
<td>0.48</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>0</td>
<td>0.23</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>lod score</td>
<td>0.109</td>
<td>0.00</td>
<td></td>
</tr>
</tbody>
</table>

Table 2 Identity by state in 73 affected sibling pairs

<table>
<thead>
<tr>
<th>Haplotypes shared</th>
<th>Locus A</th>
<th>Locus B</th>
<th>Observed</th>
<th>Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>27</td>
<td>26</td>
<td>25</td>
<td>30</td>
</tr>
<tr>
<td>1</td>
<td>36</td>
<td>38</td>
<td>40</td>
<td>37</td>
</tr>
<tr>
<td>0</td>
<td>10</td>
<td>9</td>
<td>8</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 3 Allele frequencies in 77 patients with multiple sclerosis and 88 controls

<table>
<thead>
<tr>
<th>Allele</th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Locus A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0.054</td>
<td>0.068</td>
</tr>
<tr>
<td>B</td>
<td>0.160</td>
<td>0.253</td>
</tr>
<tr>
<td>C</td>
<td>0.253</td>
<td>0.264</td>
</tr>
<tr>
<td>D</td>
<td>0.123</td>
<td>0.157</td>
</tr>
<tr>
<td>E</td>
<td>0.247</td>
<td>0.140</td>
</tr>
<tr>
<td>F</td>
<td>0.199</td>
<td>0.097</td>
</tr>
<tr>
<td>G</td>
<td>0.089</td>
<td>0.039</td>
</tr>
<tr>
<td>Locus B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.039</td>
<td>0.029</td>
</tr>
<tr>
<td>2</td>
<td>0.175</td>
<td>0.195</td>
</tr>
<tr>
<td>3</td>
<td>0.249</td>
<td>0.333</td>
</tr>
<tr>
<td>4</td>
<td>0.078</td>
<td>0.052</td>
</tr>
<tr>
<td>5</td>
<td>0.045</td>
<td>0.045</td>
</tr>
<tr>
<td>6</td>
<td>0.416</td>
<td>0.333</td>
</tr>
</tbody>
</table>

*χ² = 4.9 (1df); 0.025 < P < 0.05; p mend = 0.1, NS. All other comparisons are non-significant.
multiple sclerosis. The search for other susceptibility genes has largely concentrated on those encoding proteins central to immune function, including T cell receptor $\alpha$ and $\beta$ chains\textsuperscript{10-14} and immunoglobulin heavy chain genes.\textsuperscript{15} The presence of autoreactive T cells in samples obtained from patients with multiple sclerosis\textsuperscript{23} and the antigenicity of myelin basic protein in experimental autoimmune encephalomyelitis has also stimulated interest in the susceptibility role of the myelin basic protein gene.

Boylan et al.\textsuperscript{12} reported that the 5' end of the human myelin basic protein gene contains the repetitive sequence (TGGA), which is highly polymorphic, and has a heterozygosity of at least 45%. Analysis of 65 patients compared with 63 controls showed a significant increase in the 2-14-2-15 alleles in patients with multiple sclerosis. Tienari et al.\textsuperscript{20} showed linkage to the 5' region of the myelin basic protein gene in 17 Finnish families and a population association with one allele which conferred a relative risk of 9.1. Others have failed to confirm linkage in studies from the United States\textsuperscript{31} and Canada\textsuperscript{32} and the population association has also not been confirmed.\textsuperscript{22} We used a highly polymorphic AmpFLP system from the same 5' region of the myelin basic protein gene and failed to find a linkage in over 70 families. No association was seen comparing one member from each sibling pair with unrelated controls.

Linkage analysis is difficult in a polygenic and multifactorial disease such as multiple sclerosis in which most multiplex families contain only two affected members. Some, but not all, of the methodological problems are resolved by the affected sibling pair method. These include variable age of onset, incomplete penetrance, and mode of inheritance but not clinical heterogeneity. We only included patients who fulfilled the probable or definite categories of the Poser criteria but the possibility of disease heterogeneity remains. A second important consideration is the number of families required to prove or refute linkage. This depends on several variables including the contribution to disease susceptibility made by the gene in question, its polymorphism information content, and the number of genes involved in susceptibility. We have studied a large population with a highly informative polymorphism and find no evidence to support linkage or a population association with multiple sclerosis. It is important in association and linkage studies that the alleles are clearly distinguishable; the polymorphism described by Polymeropoulos et al.\textsuperscript{18} has the advantage of producing a relatively small product; differences in allele size are readily discernible (see figure) and it is unlikely that technical factors affected our assignment of alleles. The AmpFLP used by Tienari et al.\textsuperscript{20} has larger products with relatively small differences between alleles making it more difficult to determine their exact sizes.

Another possible explanation for differences between the study from Finland and elsewhere relates to genetic heterogeneity. The high lod score and relative risk for multiple sclerosis in a small number of families and cases suggests a prominent role for the myelin basic protein gene in that community, but our population, and those studied elsewhere, are more likely to be heterogeneous than may be the case for a relatively isolated part of Finland. The method of analysis can also significantly influence the demonstration of linkage. Tienari and colleagues\textsuperscript{20} used a technique that requires estimates to be made both for mode of inheritance and degree of penetrance; and it may be significant that several of their families contained a higher than expected number of affected sibpairs suggesting that the disease may not be identical to that being studied elsewhere. These factors may have far reaching effects on the lod score determined from the analysis of pedigrees. The identity by descent method avoids these two major difficulties; it is robust and informative when looking at complex genetic traits.

Despite the intriguing results of Tienari et al.\textsuperscript{20} the balance of opinion must now shift towards the view that a polymorphism in the myelin basic protein gene does not contribute substantially to the risk in the large number of patients with multiple sclerosis. This should not deter detailed population and family studies of other candidate genes that encode structural components of the myelin.

We are grateful to Professor G C Ebers and Ms Carole Anderson, University Hospital, London, Ontario, for technical assistance and helpful discussion. This work was supported by the Multiple Sclerosis Society of Great Britain and Northern Ireland.