Phenylketonuria in Britain: genetic analysis gives a historical perspective of the disorder but will it predict the future for affected individuals?

L A Tyfield

Most hyperphenylalaninaemia is caused by a loss of the activity of the enzyme phenylalanine hydroxylase (PAH). It has been known for some time that there are varying degrees of hyperphenylalaninaemia (HPA) and at its most severe it results in phenylketonuria (PKU), a condition that requires strict dietary management if the severe neurological sequelae naturally associated with the untreated condition are to be prevented. At its mildest, it results in a non-PKU hyperphenylalaninaemia (non-PKU HPA), which does not require dietary restriction to maintain concentrations within limits that will ensure normal physical, neurological, and cognitive development. Between these two extremes there is a whole range of phenotypic variability, most of which is related to allelic variation at the PAH locus. ( Disorders of hydroxylation resulting from defective enzymes in the synthesis and regeneration of the cofactor pathway will not be considered here.)

PKU is one of the inborn errors of metabolism and the genetics of the disorder are straightfoward—they follow a typical Mendelian autosomal recessive pattern of inheritance. Until recently, this was relevant usually in relation to the risk of a couple, known to be carriers, of having another affected child (1 in 4) or to the chance of a sibling of an affected child being a carrier (2 in 3) and then, to the population risk of a partner with no known family history carrying a mutant gene. With the arrival of the age of molecular genetics and the cloning of the PAH gene, the genetics of PKU have taken on a different significance and extensive examination of the gene over the past decade has unearthed a plethora of information that is relevant both to the severity of the disease in individual families and to the history of the gene in various populations.

The gene

The gene is on the long arm of chromosome 12 in the band region q22–q24.1 It has 13 coding regions (exons) within 90 kb of DNA and the transcribed mRNA is ~2.4 kb long. The PAH region of chromosome 12 is rich in polymorphic markers and these can be used to describe both mutant and normal chromos (fig 1). Originally, a diallelic haplotype system incorporating seven restriction fragment length polymorphic markers (RFLPs) was used to describe the genetic background of this locus. This has been expanded to include a multi-allelic short tandem repeat system (STR), a multi-allelic HindIII system of variable number of tandem repeats (VNTR), and several silent polymorphisms within the coding region of the gene. The latter arise from a substitution, usually in the last base of a codon, with no resulting change in the amino acid sequence of the protein. In theory, the polymorphic systems could manifest more than 6000 different haplotypes, but in reality, only a small number of these have been described on contemporary chromosomes. In recent years, there has been a decline in describing the diallelic haplotypes, probably because of the large amount of work involved. Frequently, an individual is heterozygous at more than one polymorphic site and, therefore, DNA is often needed from two or three family members if complete haplotypes are to be constructed. In addition, efforts have tended to concentrate on defining the genetic background of mutant rather than normal chromosomes. Nevertheless, the high degree of heterozygosity at the PAH locus is extremely useful for prenatal diagnosis and carrier testing whenever these are required.

PAH mutations

For any genetic disorder (except sickle cell disease) genetic heterogeneity is the rule rather than the exception. To the end of February 1997, more than 310 mutations were reported to the PAH Mutation Analysis Consortium database. (The curators of the PKU Mutations Analysis Consortium are at the Montreal Children's Hospital, Montreal, Quebec, Canada. The database is accessible on the Internet (http://www.mcgill.ca/pahdb); in February 1997 there were 91 members of the consortium in 29 countries. More than 5000 PKU chromosomes have been studied worldwide. Most mutations (almost 60%) are missense mutations, which cause an amino acid change in the PAH protein. In addition, there are nonsense mutations that introduce a premature
The table includes only those codons that carry multiple mutations and/or multiple haplotype associations of individual mutations.

The Cpg dinucleotide and the mutated methylated cytosine is on the template strand; hence the observed base change on the template strand is C to A transition.

The Cpg dinucleotide spans two codons (in the case of E280K, the sequence is CCC GAA at codons 279 and 280, respectively). The mutation is on the template strand and thus the observed change is a G → A transition on the template strand.

The lower case indicates that this is not the first nucleotide of the subsequent intron.

stop codon, deletions and insertions (some of which alter the reading frame), and splice site mutations which alter mRNA processing. The distribution of the numbers of mutations along the gene are also shown in fig 1.

Several mutations occur at Cpg dinucleotides. These are hypermutable sites in the genome and the hypermutability usually is related to the methylation status of the cytosine residue. Spontaneous deamination of cytosine, methylated at the 5' ring position, will change cytosine to thymine (with a guanine to adenine transition on the corresponding strand). In the PAH gene, 24 codons involve Cpg sites. Although mutations have not been reported at seven of them, 24 mutations involving the Cpg dinucleotides have been reported at the remaining 17. Often, their hypermutability is observed either by the appearance of different mutations within the same codon or by the appearance of the same mutation on multiple haplotype backgrounds (table 1). Information about the relative frequencies of mutations on different haplotypes within different populations can be useful in defining the relative age and geographical origin of that mutation.

**Population distributions of mutations**

The PAH gene has been studied extensively in many populations throughout the world. Globally, R408W (a substitution of tryptophan for arginine at amino acid position 408 of the PAH protein) is the most common mutation, accounting for 30.4% of mutant chromosomes, and it has been reported on eight different diallelic haplotype backgrounds (table 1). In addition, in many populations of northern Europe R408W is the most common mutation and here it is found principally on haplotype 1 and 2. The haplotype 2 association is more prevalent in north-east Europe and there appears to be a centre of diffusion for this association in Belarus. In north-west Europe the mutation is associated predominantly with haplotype 1 and the centre of diffusion for this association is Ireland. The two haplotype patterns are similar in that the six polymorphic sites from the 5' end of the gene are identical but they differ at the HindIII and EcoRV sites centred at the 3' end. It is possible that the mutation arose only once and was transferred by the recombination event from one haplotype to the other. However, the mutation occurs at one of the hypermutable sites in the gene and detailed examination of the STR and VNTR sequences of haplotypes 1 and 2, on both mutant and normal chromosomes, indicates that almost certainly it arose as two independent events.

In the British Isles, three mutations are most prevalent: R408W in exon 12; IVS12+1G→A, which is a splice site mutation at the exon 12 boundary; and I657T, which is in exon 3. Together, these account for about 50% of

---

**Table 1**

<table>
<thead>
<tr>
<th>Codons</th>
<th>Sequence change</th>
<th>Mutation</th>
<th>Haplotypes associations</th>
</tr>
</thead>
<tbody>
<tr>
<td>111</td>
<td>CGA → TGA</td>
<td>R111X</td>
<td>1, 4, 5, 28</td>
</tr>
<tr>
<td>158</td>
<td>CGG → CAG*</td>
<td>R158Q</td>
<td>1, 2, 4, 7, 16, 28</td>
</tr>
<tr>
<td>241</td>
<td>CGG → TGG</td>
<td>R158W</td>
<td></td>
</tr>
<tr>
<td>243</td>
<td>CGC → CAC*</td>
<td>R241H</td>
<td>1</td>
</tr>
<tr>
<td>252</td>
<td>CGA → CAA*</td>
<td>R241C</td>
<td>1</td>
</tr>
<tr>
<td>281</td>
<td>CGG → CCG</td>
<td>R243Q</td>
<td>4, 7, 9</td>
</tr>
<tr>
<td>297</td>
<td>CGG → TGA</td>
<td>R243X</td>
<td>4</td>
</tr>
<tr>
<td>328</td>
<td>CGG → CAG*</td>
<td>R252Q</td>
<td>1</td>
</tr>
<tr>
<td>351</td>
<td>CGA → CAA*</td>
<td>R252W</td>
<td>1, 6, 7, 8, 42, 69</td>
</tr>
<tr>
<td>381</td>
<td>CGA → CAA*</td>
<td>R261Q</td>
<td>1, 2, 4, 22, 24, 28</td>
</tr>
<tr>
<td>397</td>
<td>CGA → CAA*</td>
<td>R261X</td>
<td>2, 3, 4</td>
</tr>
<tr>
<td>428</td>
<td>GAA → AAT†</td>
<td>E280K</td>
<td>1, 2, 14, 16, 28</td>
</tr>
<tr>
<td>428</td>
<td>CCG → CTG†</td>
<td>P281L</td>
<td>1, 4</td>
</tr>
<tr>
<td>408</td>
<td>CGG → TGG</td>
<td>R408W</td>
<td>1, 2, 4, 5, 13, 34, 41, 44</td>
</tr>
<tr>
<td>408</td>
<td>CGG → CAG*</td>
<td>R408Q</td>
<td>4, 12</td>
</tr>
</tbody>
</table>

---

**Figure 1**

(A) Diagram of the PAH gene. Exons are represented as thin vertical lines; introns are the spaces between them. Relative positions of the diallelic polymorphic sites and the multi-allelic STR and VNTR polymorphic systems are shown. (B) Numbers of mutations that have been reported within the exons of the gene and at splice site mutations or mutations/polymorphisms within introns and 5' and 3' untranslated regions of the gene. Data taken from reference 7.
PKU chromosomes but there are some significant differences in the relative distribution of these mutations in different geographical areas (fig 2). In western Scotland, for example, R408W accounts for about 31% of PKU chromosomes, whereas in south-west England it is found on only 10% of PKU alleles. With IVS12nt1 the relative frequencies are reversed: it is found on 27% of mutant chromosomes in south-west England and on only 5% in western Scotland. These differences in relative frequencies are highly significant (p < 0.001) and almost certainly reflect different historical events in the founding populations of the two geographical areas. R408W is also the most common mutation in Ireland. There has been considerable cross-migration between western Scotland and Northern Ireland since prehistoric times, therefore, it is not surprising that the most common mutation is the same in both geographical areas. However, one in five alleles carrying R408W in western Scotland are associated with a haplotype 2 background, compared with only one in 10 in Northern Ireland, and an almost exclusive haplotype 1 association in the Republic of Ireland. Thus, despite similarities in the relative frequencies of the mutation, differences in the frequencies of its haplotype association show the modern population of western Scotland to be more heterogeneous than that of Northern Ireland.

The splice site mutation, IVS12nt+1, which is the most common in south-west England, also is most common in Denmark, where it accounts for 37% of mutant genes. It has been proposed to be Danish in origin. Its appearance in the British Isles could date from the 6th to the 10th centuries AD, when repeated coastal invasions and inland migrations established the Danish kingdom over a large part of central Britain. Alternatively, the mutation could have been introduced first by the Angles, Saxons, or Jutes who settled in much of east and south Britain during the 5th century AD. In either case, one would anticipate a high relative frequency of IVS12nt+1 in the eastern coastal regions of England. The fact that the mutation is associated almost exclusively with a diallelic haplotype 3 background suggests that it is relatively new.

The relative frequencies of these mutations in London and south Wales, where samples have come principally from in and around Cardiff, are almost identical (fig 2). The diverse grandparental origin of many PKU individuals in south Wales suggests that the similarities between the Welsh and London samples are more likely to reflect the heterogeneity that is seen in the populations of two large capital cities, rather than any similarity in the indigenous populations of the two areas.

The distribution of the 165T mutation is interesting because it is quite common in the populations of the United Kingdom and Spain, as well as in areas of the world that have been inhabited over the last three centuries by immigrants of Anglo-Celtic and Spanish origin. It is relatively rare in other European populations. Owing to its high frequency in Ireland, it has been proposed to be “Celtic” in origin, but its association with the greatest numbers of haplotypes (encompassing diallelic, STR, and VNTR polymorphisms) suggests that it is one of the oldest at the PAH locus and probably predates the arrival of the Celts in Ireland. A hypermutable sequence is not involved.

Another splice site mutation, IVS10nt-11, is most common in populations of southern Europe, including those of Mediterranean and Arabic origin. It occurs with the highest frequency in Turkey, and an east–west gradient in its relative frequency in south European populations has led to the proposal that it originated in the Middle East and spread west and north by the migration of Neolithic farmers. In the British Isles, this mutation and another, which is common in south European populations (L48S), appear almost exclusively in south-west England, where together they account for 10% of mutant chromosomes. This is highly significant (p < 0.001) and although they may have been brought to south-west England through the megalithic builders from Hibernia, the association of each with multiple genetic backgrounds suggests that there were multiple origins. The association of IVS10nt-11 with many haplotypes suggests also that it is very old.

Rare mutations
In most population studies where mutation ascertainment is virtually complete, five or six mutations account for between 30% and 50% of mutant alleles and the remainder are usually multiple and varied. In the PKU population of the British Isles, for example, 72 different mutations have been identified on 500 independent chromosomes from western Scotland,
south Wales, south-west and south-east England. Fifty five mutations (76%) have been found on fewer than 1% of chromosomes and 35 (almost half) were found only once. Of these, 15 mutations had not been reported previously to the Mutation Analysis Consortium database. Eleven mutations were found on only two alleles in apparently unrelated individuals (some in the same geographical area, others in a different area). Where genealogical evidence is lost, detailed examination of the haplotype background on which the mutations are found may provide evidence of an identity by descent. In this regard, sometimes one of the more common mutations has been found in a particularly unusual haplotype association. I65T, for example, which occurs almost entirely on a haplotype 9 in the British Isles, was found on a very rare haplotype 49 background. This mutation could have arisen through a recombination between a mutant haplotype 9 and a normal haplotype 4 which is common in western England. Its appearance in the same unusual association on a single allele in an Australian study39 almost certainly represents an identity by descent, particularly in view of the Anglo-Celtic origin of many Australian settlers over the last century.

It is worth mentioning that PKU is very rare among people of African origin. It has been proposed39 that mutations at the PAH locus arose after the original migration and diaspora out of Africa over 100 000 years ago. Furthermore, it is suggested that the mutations now found in black individuals will have been introduced into the families through ancestral matings with non-blacks. In our own study only four of almost 600 mutant alleles that we have examined in the British Isles are from black children. The parents of one child are both Afro-Caribbean and two children are of mixed race (Afro-Caribbean/Welsh and Afro-Caribbean/Jewish). Although the two children in whom both mutations have been identified both carry one that is common in northern Europe (I65T and R408W), the mutation on the other allele is very rare (P122Q and R252Q). Although R252Q has been reported in other populations within Europe, China, and the United States,30 P122Q, associated with extremely rare but related haplotypes, has been reported on only two additional alleles in Spanish individuals.39

Carrier testing

In most countries of the developed world, population screening for PKU is undertaken in the neonatal period. Almost complete ascertainment of affected individuals enables a true incidence of PKU to be determined in any population. The disorder is rare in Japan and virtually unheard of in Finland. In other European countries and those of North America the incidence is generally about one in 10 000 individuals. A finer delineation of populations shows that within relatively small geographical distances there can be relatively large differences in incidence. Within the British Isles alone, for example, hyperphenylalaninaemia is found in approximately one in 12 000 newborn infants in south-west England; in south Wales, it is approximately one in 10 000; in western Scotland the figure is nearer one in 7500; in Northern Ireland it is approximately one in 4500. Assuming Hardy-Weinberg equilibrium (where there is the possibility of only two alleles in the population “on”, “normal” and “off”, the frequency of the normal allele) in all of these populations (p² + 2pq + q² = 1) the carrier frequency (2pq) will vary from one in 35 in (in Northern Ireland) to one in 55 (in south-west England).

In this regard, carrier testing for a mutation at the PAH gene can be undertaken only for individuals, who are direct relatives of an affected family member. If the mutation has not been identified, this is a relatively straightforward procedure but if not, gene tracking using any of the polymorphic markers can be undertaken to establish whether the consultant has an allele in common with the affected individual. When known carriers, or known PKU individuals, ask for the carrier status of their partners to be determined, biochemical analysis is still the first test to offer. A phenylalanine: tyrosine ratio measured on a sample of blood taken at midday after an overnight fast will determine carrier status. If this test shows that the individual is likely to be a carrier, molecular genetic analysis can be undertaken. However, as there are over 310 mutations characterised, a definitive answer is not always guaranteed.

Genotype/phenotype correlations

In vitro expression analysis shows that each mutation has a particular quantitative effect on enzyme activity. Some mutations are found to be associated with a complete loss of enzyme activity, whereas others are associated with residual activity, ranging from 5% to 70%. It can be anticipated that various combinations...
of mutations would result in a full spectrum of biochemical phenotypes, ranging from classic PKU (requiring strict dietary management) to mild non-PKU HPA (in which dietary restriction of phenylalanine is not necessary).

Some mutations (R297P, A322G, T380M, D425N, as a few examples) appear to be associated only with a non-PKU HPA; when they occur in combination with a null mutation, the diagnostic phenylalanine concentrations in these individuals and phenylalanine concentrations on an unrestricted intake are never above ~450 μmol/l. At the other end of the spectrum, it is probably safe to say that the highest diagnostic phenylalanine concentrations are usually found in infants, whether breast or formula fed, who have the lowest associated PAH activity. However, diagnostic concentrations in infants carrying two null mutations can range from ~1000 to >3000 μmol/l and there is a large overlap with those carrying at least one mutation which is believed to be associated with residual enzyme activity (fig 3).

This also applies to the severity of the disorder as determined by the degree of dietary phenylalanine intake, the relative ease of dietary management or, ultimately, the clinical outcome. Generally, those children with a complete PAH activity tend to have the greatest dietary phenylalanine restriction at all stages of development but, once again, there are large overlaps between those with two null mutations and individuals carrying a less deleterious one (unpublished results). Thus, for any individual phenylketonuric infant, a lower diagnostic phenylalanine concentration in the neonatal period will not necessarily imply residual enzyme activity and, therefore, predict a less severe biochemical phenotype.

The fact that additional genetic, constitutional, and environmental factors contribute to the overall phenotype is also manifest by the large variation in the intellectual outcome of untreated subjects with the same genotype, both within and between families. The precise biological mechanisms by which phenotypic variation occurs are not known, but it is certain that the contributing factors are multiple and varied. Treacey et al. have shown that although there is an identical hydroxylation of phenylalanine to tyrosine in affected siblings with the same genotype, there can be variation in phenylalanine use through additional pathways involved in phenylalanine homeostasis.

Therefore, it is clear that although in vitro expression systems are extremely important and useful laboratory tools for establishing the more proximate effects of a particular mutation such as the effects on mRNA synthesis, protein synthesis, or enzyme activity, they allow only a relative order of severity of various mutations to be established. The more ultimate effects in affected subjects (ease of dietary biochemical control or clinical outcome in treated or untreated patients) must take into account additional complementary influences that are involved both in phenylalanine metabolism in particular and neurophysiological development in general. Ultimately, a simple correlation may not be always apparent between genotype at the PAH gene and clinical phenotype.

Results of the PAH gene analysis presented here were part of a multicentre study in the British Isles. We are deeply indebted to many clinicians throughout England, Scotland, and Wales who sent us samples from their PKU patients and to the patients themselves and their families who participated in the project. We gratefully acknowledge the generosity of Professor M. G. E. Taylor, Dr. Susan Byck who gave us primers for denaturing gradient gel electrophoresis analysis. Mrs Anne Stephenson carried out the genetic analyses in the laboratory at Southmead Hospital and she was supported by a grant from the Wellcome Trust.