Mutation and haplotype analysis of phenylalanine hydroxylase alleles in classical PKU patients from the Czech Republic: identification of four novel mutations

Libor Kozak, Michaela Blázková, Viera Kuhrová, Anna Pijáčková, Šárka Růžičková, Sylvie Šťastná

Abstract
Mutations, haplotypes, and other polymorphic markers in the phenylalanine hydroxylase (PAH) gene were analysed in 133 unrelated Czech families with classical phenylketonuria (PKU). Almost 95% of all mutant alleles were identified, using a combination of PCR and restriction analysis, denaturing gradient gel electrophoresis (DGGE), and sequencing. A total of 30 different mutations, 16 various RFLP/VNTR haplotypes, and four polymorphisms were detected in 266 independent mutant chromosomes. The most common molecular defect observed in the Czech population was R408W (54.9%). Each of the other 29 mutations was present in no more than 5% of alleles and 13 mutations were found in only one PKU allele each (0.4%). Four novel mutations G239A, R270fsdel5bp, A342P, and IVS11nt-8g→a were identified. In 14 (5.1%) alleles, linked to four different RFLP/VNTR haplotypes, the sequence alterations still remain unknown. Our results confirm that PKU is a heterogeneous disorder at the molecular level. Since there is evidence for the gene flow coming from northern, western, and southern parts of Europe into our Slavic population, it is clear that human migration has been the most important factor in the spread of PKU alleles in Europe. (J Med Genet 1997;34:893–898)

Keywords: phenylketonuria; mutation detection; polymorphism

Phenylketonuria (PKU, MIM 261600) is an autosomal recessive disorder caused by mutations in the phenylalanine hydroxylase (PAH) gene. More than 290 mutations have been detected in this gene and cited in the PAH Gene Mutation Analysis Consortium Database. Different mutations cause different degrees of PAH enzyme impairment. Total absence or severe reduction of PAH activity or stability results in classical PKU with a serum phenylalanine level above 1200 μmol/l (>20 mg Phe/100 ml).

The PAH gene, spanning 90 kb, contains a whole array of polymorphic sites. After introducing diallelic RFLP haplotypes, other polymorphic markers, such as multiallelic systems of VNTR and STR, silent mutations, and polymorphisms in flanking intronic sequences, were found. All these polymorphisms put together enabled us to differentiate the many variant PAH alleles much more precisely than would have been possible by conventional haplotyping.

Numerous studies have shown the effectiveness of DNA polymorphisms for identification and tracking of both normal and mutant chromosomes during various waves of human migration. In many cases, strong associations were found between PKU mutations and specific RFLP haplotypes in different populations. There are multiple geographically and ethnically distinct origins for PKU within the European population. For more accurate determination of the specific founding events and gene flow it would be necessary to obtain the missing information from less frequently studied populations.

During a previous analysis of some of the Czech PKU alleles, done using restriction enzyme based assays or allele specific oligonucleotide hybridisation (ASO), 11 PKU mutations were detected. Advance in new PAH mutation identification was the result of the introduction of denaturing gradient gel electrophoresis (DGGE) technology. The present study exploits this technique for identification of mutations in a large sample of mutant PKU alleles. The linkage disequilibrium among these mutations and RFLP/VNTR haplotypes, together with additional polymorphisms in the Czech PKU population, is analysed here.

Materials and methods

PATIENTS
The total incidence of classical and mild forms of PAH deficiency in the Czech population is 1:9000. The proportion of mild and classical forms is approximately 1:20. In order to obtain a phenotypically homogeneous sample, we selected 133 (32.5%) unrelated families from the Czech Republic with only classical PKU, treated in two specialised metabolic centres cooperating with our laboratory. Among them, 130 families were white and three were of Gypsy ethnic background. All PKU patients were classified as having classical PKU on the basis of plasma phenylalanine concentrations over 20 mg/100 ml during the initial
MUTATION AND HAPLOTYPE ANALYSIS

Genomic DNA was extracted from 10 ml of EDTA anticoagulated blood by standard methods. Mutations L48S, I65T, R158Q, R252W, R261Q, G272X, IVS10nt546, R408W, Y414C, and IVS12nt1g→a were detected by enzymatic digestion of PCR products containing natural or amplification created restriction enzyme sites under the conditions summarised previously. The silent mutations Q232Q, V245V, and L385L were specified by digestion of amplified exons 6, 7, and 11 with Ddel, AlaI, and Mbol, respectively.7 Haplotypic analysis was performed by PCR at five sites, BglII, PvuIIa, PvuIIb, MspI, and XmnI.28–31 The VNTR system at the HindIII site was analysed following the procedure of Goltsov et al.1 The Southern blotting technique was used for detection of two remaining polymorphic sites, EcoRI and EcoRV.23 RFLP/VNTR haplotypes were numbered in accordance with the nomenclature of Eisensmith and Woo.32

DGGE ANALYSIS

DGGE experiments were performed under the conditions described by Guldberg et al.,32–37 but four different primers were used (exon 2 sense primer: 5′-CTT-GCF-CTG-TCC-ATG-GAG-G-3′, exon 7 sense primer: 5′-GCG-CCT-AGT-GCC-TCT-GAC-TCA-3′, exon 7 antisense primer: 5′-CTT-GTG-GAC-CAG-CCA-GCA-A-3′, exon 12 antisense primer: 5′-GCG-AC-TGA-GAA-ACC-GAG-TGG-CCT-3′. [GC] is the GC clamp sequence identical to that reported by Guldberg et al.33)

To generate heteroduplexes, PCR products were denatured at 100°C for five minutes, then held at 65°C for one hour and at 37°C for one hour. Electrophoresis was run for 4.5 hours at 170 V in a bath of 1 X TAE buffer with a constant temperature of 57°C. The gels were stained with ethidium bromide and photographed.

SEQUENCE ANALYSIS

DNA fragments which displayed an abnormal DGGE pattern were analysed using solid phase sequencing.3 The downstream primer of each pair used for the amplification of the corresponding exon regions of the PAH gene was biotinylated. Strand separation of the PCR product was obtained using streptavidin coated magnetic beads M280 (Dynal). The resulting single stranded PCR product was used as the sequencing template. Sequencing reaction was carried out by the dideoxy chain termination method using Sequenase 2.0 DNA Sequencing Kit (USB), alpha-[35S]dATP (Amersham), and non-biotinylated amplification primer as sequencing primer. The electrophoresis was run on BIO-RAD sequencing apparatus in 6% polyacrylamide, 8.0 mol/l urea at a constant 22 mA for two and four hours.

Results

MUTATION ANALYSIS

A three step approach was used for the detection of PKU mutations. In the first step, prevalent mutations were screened by PCR and restriction analysis as described under Materials and methods. As the second step, “broad range” DGGE analysis was performed in order to localise the remaining unknown PKU mutations in the entire coding region and flanking intronic sequences of the PAH gene. An example of specific heteroduplex/homoduplex band pattern for different genotypes in exon 11 is shown in fig 1. During the last step, sequencing analysis was run to identify exactly nucleotide alterations shown by DGGE as abnormal band pattern. Fig 2 shows novel point mutations found in our study.

Results are summarised in table 1. With the strategy used, we were able to identify disease causing mutations on 252 (94.9%) of 266 mutant alleles. A total of 30 different mutations were found, including 18 missense, five splice site, four nonsense, and three deletion mutations. Eight different PAH molecular lesions were identified in exon 7; while none was found in exons 4, 9, and 13. The most frequent mutation, R408W, accounts for 54.9% of all PKU alleles; the frequency of the four other most prevalent mutations (R158Q, IVS12nt1g→a, R243X, and IVS10nt546) ranges from 5% to 3%. Thirteen mutations
were detected on only one each of 266 Czech PKU alleles, corresponding to a frequency of 0.4%.

Interestingly, 5.1% (14) of alleles still remain unknown (table 2). All 14 PKU patients with unidentified PKU alleles were compound heterozygotes carrying one previously determined mutation.

Four mutations have not previously been described, and thus are described in detail in table 3. Three of them, G239A on haplotype 1.7 (fig 2A), A342P on haplotype 4.3 (fig 2B), and R270fsdel5bp on haplotype 72.12 (results not shown), were private mutations each found on one allele only. Haplotype 4.3 linked mutation IVS11nt-8g→a (fig 2C) was detected on four alleles.

In addition, four known polymorphisms (three silent mutations, Q323Q, V245V, and L385L, and one intron variation IVS3nt-22t→c) were also identified by DGGE and confirmed by restriction or sequence analysis or both.

**Genotype-Polymorphic Marker Association**

Table 1 shows the association between 30 observed PKU mutations, RFLP/VNTR haplotypes, and other polymorphisms found in 252 mutant PAH genes. Sixteen different RFLP/VNTR mutant haplotypes were identified. The prevalent mutation, R408W, was almost exclusively associated with haplotype 2.3 (145 of 146 alleles). Four mutations were associated with more than one haplotype. Two of the mutations were found on haplotypes with identical VNTR (IVS10nt546 on haplotypes 6.7 and 34.7, R243Q on haplotypes 1.8 and 32.8). The other two were on the background of different VNTR haplotypes (R408W on haplotypes 2.3 and 5.9, R158Q on haplotypes 4.3, 2.3, and 7.8). To confirm the identity, all of the four mutations found on unusual haplotypes underwent sequencing analysis. The remaining mutations were tightly linked with one RFLP/VNTR haplotype. Haplotype 4.3 carried the most variant mutations (eight, 26.7%). Seven different mutations occurred with haplotype 1.8 and three with haplotype 1.7. All the other haplotypes carried one or two various mutations.

As shown in tables 1 and 2, four other polymorphic markers were found to be associated with specific mutant alleles, and thus could be helpful as relative markers for identification of these specific mutations. Silent mutations Q232Q and L385L were always present on the haplotype 3.8 and haplotype 7.8 background alleles bearing mutations IVS12nt1g→a and G272X, respectively (however, in one case there was also mutation R158Q, and in three cases there were unknown mutations). A single L385L was identified on the “Gypsy PKU allele”, with the 69.3 haplotype carrying mutation R252W.

Among the 14 unidentified PKU alleles, four different haplotypes were identified (table 2), and therefore we assume that there might be at least four more mutations present in our population. The majority of unknown PAH lesions are linked to haplotype 4.3 alleles.
Table 1  Frequencies of 30 PKU mutations and their association with RFLP/VNTR haplotypes and other polymorphisms in 252 mutant PAH genes

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Haplotype</th>
<th>RLFP/VNTR</th>
<th>Polymorphism</th>
<th>No of alleles</th>
<th>Frequency (%)</th>
<th>Total frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>R408W</td>
<td>2.3</td>
<td>—</td>
<td>—</td>
<td>145</td>
<td>54.5</td>
<td>54.9</td>
</tr>
<tr>
<td>R158Q</td>
<td>5.9</td>
<td>—</td>
<td>—</td>
<td>6</td>
<td>2.3</td>
<td>4.9</td>
</tr>
<tr>
<td>G272X</td>
<td>—</td>
<td>1.4</td>
<td>Q323Q+V245V+IVS3nt-22t-c</td>
<td>10</td>
<td>3.8</td>
<td>4.1</td>
</tr>
<tr>
<td>R243X</td>
<td>4.3</td>
<td>—</td>
<td>Q323Q+L385L</td>
<td>11</td>
<td>4.1</td>
<td>4.1</td>
</tr>
<tr>
<td>R261Q</td>
<td>1.8</td>
<td>—</td>
<td>Q323Q+V245V+IVS3nt-22t-c</td>
<td>9</td>
<td>3.4</td>
<td>3.4</td>
</tr>
<tr>
<td>R270fsdel5bp</td>
<td>3.8</td>
<td>—</td>
<td>—</td>
<td>5</td>
<td>1.9</td>
<td>3.0</td>
</tr>
<tr>
<td>L48S</td>
<td>1.9</td>
<td>—</td>
<td>—</td>
<td>3</td>
<td>1.1</td>
<td>1.1</td>
</tr>
<tr>
<td>P228T</td>
<td>1.9</td>
<td>—</td>
<td>—</td>
<td>4</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>R243Q</td>
<td>1.8</td>
<td>—</td>
<td>—</td>
<td>3</td>
<td>1.1</td>
<td>1.1</td>
</tr>
<tr>
<td>Unknown</td>
<td>32.8</td>
<td>—</td>
<td>—</td>
<td>21</td>
<td>0.4</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Table 2  Association of 14 PKU alleles carrying unknown mutations with RFLP/VNTR haplotype and other polymorphisms and their frequencies

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Haplotype</th>
<th>RLFP/VNTR</th>
<th>Polymorphism</th>
<th>No of alleles</th>
<th>Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unknown</td>
<td>1.8</td>
<td>—</td>
<td>—</td>
<td>2</td>
<td>0.7</td>
</tr>
<tr>
<td>Unknown</td>
<td>4.3</td>
<td>—</td>
<td>Q323Q+V245V</td>
<td>7</td>
<td>2.6</td>
</tr>
<tr>
<td>Unknown</td>
<td>7.8</td>
<td>—</td>
<td>Q323Q+L385L</td>
<td>3</td>
<td>1.1</td>
</tr>
<tr>
<td>Unknown</td>
<td>21.12</td>
<td>—</td>
<td>—</td>
<td>14</td>
<td>5.1</td>
</tr>
</tbody>
</table>

Table 3  Novel PAH mutations identified in the Czech PKU population

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Localisation</th>
<th>Nucleotide change</th>
<th>Change in protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>G239A</td>
<td>Exon 7</td>
<td>GGT-&gt;GCT</td>
<td>Gly-&gt;Ala</td>
</tr>
<tr>
<td>A342P</td>
<td>Exon 10</td>
<td>CCA-&gt;CCA</td>
<td>Ala-&gt;Pro</td>
</tr>
<tr>
<td>R270fsdel5bp</td>
<td>Exon 7</td>
<td>delACATG</td>
<td>Frameshift</td>
</tr>
<tr>
<td>IVS11nt-8g&lt;-&gt;a</td>
<td>Intron 11</td>
<td>ATGG&lt;--TAGG</td>
<td>Splice defect</td>
</tr>
</tbody>
</table>

Discussion

A highly efficient method, based on the combination of "broad range" DGE and sequencing analysis, was used for identification of PAH mutations in the Czech classical PKU population. By detailed scanning of the PAH gene, a mutation detection rate of 94.9% was achieved. Four novel mutations (G239A, A342P, R270fsdel5bp, and IVS11nt-8g<->a) were identified in our samples. This confirms that the final number of PKU mutations has not yet been reached, yet other mutations could be identified in less frequently studied populations. The role of these new mutations in enzyme deficiency can be established only by expression analysis. Without this analysis we can, however, assume that all four novel mutations severely alter PAH activity; compound heterozygosity for these mutations results in classical PKU.

Interestingly, in our study, sequence alterations in 5.1% of alleles (14 of 266) still remain unspecified. This finding does not correlate with the results obtained in north-west European \(^{14,26}\) and Sicilian populations \(^{27}\) where the percentage of mutations identified by the same methods was between 98.2 and 99.6%. Preliminary results from a study done in Poland,\(^ {31}\) as well as from the study presented here, suggest that more unusual PKU alleles may be present in the Slavic population. According to our results there are at least four such alleles (table 2) in the Czech PKU population. There are several possible reasons for this. The unknown lesions may possibly lie deep in the intrinsic regions and may exert some effect on splicing. This hypothesis will be tested by ectopic transcripts. Impairment of the 5' promoter and 3' polyadenylation regions of the PAH gene can also be the cause of PKU. Recently, at least two protein binding sites which can play a role in transcriptional activity of the PAH gene in the 5' TATA-less promoter region were found.\(^ {31}\) Sequencing analysis of these regions is now in progress. It is also possible that some other genes are involved in the pathogenesis which leads to the classical phenotype of PKU.

Strong linkage disequilibrium among PKU mutations and haplotypes, together with additional polymorphisms in the PAH gene, provides a useful tool in the population genetics of PKU. As can be seen in table 3, 14 different mutant alleles carrying 30 non-identical mutations were found. As expected from previous studies,\(^ {14,28,31}\) R408W is the major mutation in our Slavic population with frequencies of 54.5% on allele H2.3/- and 0.4% on allele H5.9/-.. The frequencies of the other 29 mutations, however, did not exceed 5%. Almost half of all the mutations (13 of 30) were unique mutations, each detected on only one of 266 Czech PKU alleles. The observed heterogeneity of mutant alleles, similar to various western European populations, results particularly from human migration in Europe. Numerous waves of migrants flowed through the territory of the present Czech Republic, which was permanently settled by Slavic tribes in the 5th to 6th centuries AD. When the migrating populations integrated, the Slavic population gene pool of that time must have been enriched by many various mutant alleles.

The mutations IVS12nt1g<->a, G272X, and F299C of Scandinavian origin, tightly linked to corresponding haplotypes and polymorphisms (table 1), represent 7.1% of all our mutant alleles. Mutation F39 with haplotype 1.8 and 165T with haplotype 9.8, both predominant in
Ireland, the south-west of England, and in Irish and Scottish immigrants in Australia, were present in our samples uniquely. Mutation IVS10nt456 (H.6.7), prevalent in Turkey and Greece, mutation V388M (H.1.8), frequent in the Iberian Peninsula (UgarTE, personal communication), and finally L485 (H.4.3) and IVS7nt14–a (H.4.3), common in Italy and Sicily, are good examples of gene flow from southern Europe to our population.

Recurrence mutation, among many other processes, also contributed to our variant mutant gene pool. The two mutations R158Q identified on alleles H2.3- and H7.8- Q232Q+L385L, are most likely of such an origin, since they differ considerably from the predominant allele R158Q-H4.3/ Q232Q+V245Y+IVSnt22t-c. The recurrence of mutation R243Q in two separate populations and its consecutive spreading can also not be excluded. This mutation has been previously described in China in association with haplotype 4, and in Japan and Korea with an unknown haplotype. The linkage of R234Q to haplotype 1.8 and 32.8, observed in our population, was novel.

The mutations R408W, R158Q, IVS10nt546, and R234Q were observed on more diverse alleles (table 1). Alleles H2.3- and H5.9- carrying mutation R408W, differ from each other in several polymorphic sites. As mentioned above for the case of mutation R158Q, it could be the recurrent mutation in the CpG hotspot codon. On the other hand, the alleles IVS10nt546-H6.7- and IVS10nt546-H34.7- differ only in one RFLP site and thus the observed heterogeneity is most probably the result of mutation abolishing one of the polymorphic sites. The unusually combined alleles R234Q-H1.8- and R234Q-H32.8- differ in the first three RFLP sites on the 3’ end of the gene. The most probable cause of formation of haplotype 32.8 with R234Q is recombination between haplotype 1.8 carrying R243Q and haplotype 7.

There appears to be an interesting group of six Gypsy PKU alleles included in our study. As previously published, five R252W mutated alleles (H69.3/L385L) were detected exclusively in Gypsy patients. Two of these patients were homozygous for R252W and the third was a compound heterozygote R252WR408W-H2.3. The families are unrelated; however, because of common ancestors from Eastern Slovakia, a common origin of the mutation can be assumed. The mutation R252W was found in Europe mainly on haplotype 7 (many contributors to the PAH Gene Mutation Analysis Consortium, 1996). Since the Gypsies form a special part of the European population, the mutation R252W on the Gypsy PKU allele might be a recurrent mutation in the CpG hotspot codon 252. It is also probable that the allele R252W-H69.3/L385L was formed from the original allele R252W-H7.8/ Q232Q+L385L.

Linkage disequilibrium enables us not only to elucidate the mechanisms of origin and spreading of PKU mutations in various human populations, but also to single out the specific mutations from the wide spectrum of these mutations. Based on the results presented, the most efficient procedure for postnatal and prenatal DNA analysis of PKU in the Czech population was established. In the first step, the mutation R408W is examined, together with the number of VNTR. Then, the "broad range" DGGE of exons 1-12 is run. The suspected mutation is determined from the association of allele R408W and the number of VNTR and polymorphisms detected. The mutation can then be directly identified by restriction enzyme based assays or solid phase sequencing. The procedure described proved to be highly efficient and reliable for prompt determination of PKU genotypes in our population.

We thank Eva Šašková for DNA preparation, Dr P Guldberg for providing DNA samples and the DGGE analysis, and Dr J Apold for providing some biotinylated PCR primers. This study was supported by grant GACR-302/95/2535 from the Grantová agentura České republiky.

6 Eiken HG, Knappskog PM, Guldberg P, Apold J. DGGE analysis as supplement to SSCP analysis of the phenylalanine hydroxylase gene: detection of eight (one de novo, seven inherited) of nine remaining Norwegian PKU mutations. Hum Mutat 1996;8:19-22.


