doi:10.25143/prom-rsu_2012-11_dt

RĪGAS STRADIŅA UNIVERSITĀTE

Natalja Pronina

THE MOLECULAR BASIS OF PHENYLKETONURIA AND HYPERPHENYLALANINEMIA IN LATVIA

Doctoral Thesis

Speciality – Medical Genetics

Rīga, 2012

RĪGAS STRADIŅA UNIVERSITĀTE

Natalja Pronina

THE MOLECULAR BASIS OF PHENYLKETONURIA AND HYPERPHENYLALANINEMIA IN LATVIA

Doctoral Thesis

Speciality: Medical Genetics

Supervisor:
Dr. med., asoc. Professor Rita Lugovska

ESF
Supported by European Social Fund

RIGA, 2012

CONTENTS

ABBREVIATIONS

5'UTR	- 5' untranslated region
ANOVA	- ANalysis Of VAriance between groups
APS	- Ammonium persulphate
ARS	- Autoregulatory sequence
ATP	- Adenosine-5'-triphosphate
AVs	- Assigned values
BH4	- tetrahydrobiopterin
BLAST	- Basic Local Alignment Search Tool (data base)
BSA	- Bovine serum albumin
cAMP	- Cyclic adenosine monophosphate
CBR1	- Cofactor binding region
CD	- Catalytic domain
cDNA	- Cloned DNA
DGGE	- Denaturation gradient gel electrophoresis
DNA	- Deoxyribonucleic acid
dNTP	- Deoxynucleotide triphosphates
DOPA	- Dihydrophenylalanine
EDTA	- Ethylenediaminetetraacetic acid
EEG	- Electroencephalography
EF	- Executive function
GABA	- γ-Aminobutyric acid
gDNA	- Genomic DNA
HPA	- Hyperphenylalaninemia
IEMs	- Inborn error of metabolisms
kDa	- kilo Dalton
Km	- Michaelis–Menten kinetics (the amount of substrate that produces the half of the maximum velocity of the enzyme)
L-Phe	- L-Phenylalanine
MHP	- Mild hyperphenylalaninemia
MS	- Mass spectrometry
MtDNA	- Mitochondrial DNA
PAH	- Phenylalaninehydroxylase
PAH	- Phenylalaninehydroxylase gene

PCR	- Polymerase chain reaction
Phe	- Phenylalanine
PheOH	- Phenylalaninehydroxylase
PIC	- Polymorphism information content
PKU	- Phenylketonuria
PolyPhen	- Polymorphism Phenotyping
PSIC	- Position-specific-independent-counts
RFLP	- Restriction length fragment polymorphism
SNP	- Single nucleotide polymorphism
STR	- Short tandem repeats
TAE buffer	- Tris-acetate-EDTA buffer
TBE buffer	- Tris-Borate-EDTA buffer
TE buffer	- Tris and EDTA buffer
TEMED	- Tetramethylethylenediamine
Tyr	- Tyrosine
VNTR	- Variable number of tandem repeats

SUMMARY

Phenylketonuria (PKU) is the most common inborn error of amino acid metabolism in Europeans. It is caused by an autosomal recessive deficiency of the hepatic enzyme phenylalanine hydroxylase (*PAH*) that catalyses the irreversible hydroxylation of phenylalanine to tyrosine. The elevated level of phenylalanine effects the energy production, protein synthesis, and neurotransmitter homeostasis in the developing brain and results in the most important manifestation of PKU - mental retardation.

More than 560 different disease-causing mutations in the *PAH* gene have been identified and reported since the gene was discovered in 1986. Mutations differ in residual enzyme activity, and the genotype could be a good predictor of biochemical phenotype in the majority of patients.

The aim of the study was to investigate the molecular basis of phenylketonuria and hyperphenylalaninemia in Latvian patients and evaluate a PAH gene mutation diagnostic strategy in Latvian population.

Analysis of the molecular basis of PKU in Latvia has revealed 20 different mutations in the *PAH* gene. The most common mutation was R408W that accounted for 73% of all PKU chromosomes that gives the high level of homogeneity ("homozygosity") at the *PAH* locus in Latvia (j=0.514). Frequencies of remained 19 mutations ranged from 0.7 to 5.7% of all mutant alleles. The majority of mutations (12/20) were severe and responsible for the classic PKU phenotype that was observed in 91% of PKU patients.

The evaluation of patients' genotypes can provide the additional information for BH_4-responsiveness. According to the study results 13 of 70 (18%) Latvian PKU patients could potentially benefit from chaperon therapy by sapropterin dihydrochloride while remaining 57 (81%) patients with homozygous R408W mutation should keep the low phenylalanine diet as the only effective form of therapy.

Minihaplotype studies have revealed 16 different minihaplotypes associated to *PAH* gene mutations and 20 different minihaplotypes for normal *PAH* alleles. The average probability of heterozygosity for minihaplotypes was about 76% for mutant and 92% for normal chromosomes indicating a greater diversity of normal alleles. Statistical analysis has revealed the significant difference in the distribution of normal and mutant alleles for only two minihaplotypes 3/238 (p=0.0000572) and 8/230 (p=0.0133), and the tendency to statistically significant difference ($p<0.10$) between normal and mutant alleles in the distribution of minihaplotypes 3/242, 7/246 and 8/234.

Analysis of the distribution of the *PAH* gene mutation R408W together with it strong association with a typical East-European VNTR3/STR238 minihaplotype have

confirmed the Balto-Slavic origin of mutation R408W and introduction of this mutation to other European populations by people migrations.

The three-step *PAH* gene mutation detection strategy used in the study is the most effective for routine diagnostics in Latvian population with the sensitivity of the method 99%.

KOPSAVILKUMS

Fenilketonūrija (FKU) ir biežākā iedzimtā aminoskābju vielmaiņas patoloģija Eiropā.

Tās cēlonis ir autosomāli recesīvi pārmantots aknu enzīma fenilalanīna hidroksilāzes (FAH), kas katalizē neatgriezenisku fenilalanīna pārvēršanos tirozīnā, defekts. Paaugstinātais fenilalanīna līmenis ietekmē enerģijas produkciju, olbaltumu sintēzi un neiromediatoru homeostāzi augošās smadzenēs un izraisa svarīgāko FKU izpausmi – garīgu atpalicību.

Kopš gēna atklāšanas 1986. gadā ir atklātas vairāk nekā 560 dažādas slimību izraisošas mutācijas fenilalanīna hidroksilāzes gēnā (*FAH*). Dažādas mutācijas nosaka dažādu enzīma atlieku aktivitāti, kas ļauj genotipu izmantot, lai prognozētu iespējamo bioķīmisko fenotipu vairākumam pacientu.

Pētījuma mērķis bija noskaidrot mutāciju spektru *FAH* gēnā Latvijas FKU un hiperfenilalaninēmijas pacientiem un izveidot tā noteikšanas metodi Latvijā.

Latvijas FKU pacientiem raksturīgas 20 dažādas mutācijas *FAH* gēnā. Visbiežāk sastopamā mutācija ir R408W, kuru atrod 73% FKU pacientu hromosomās, tai raksturīga augstas pakāpes homozigotāte *FAH* lokusā Latvijā (j=0,514). Pārējo 19 mutāciju biežums mutantajās alēlēs svārstās no 0,7% līdz 5,7%. Vairākums no mutācijām (12/20) ir smagas un izraisa klasisku FKU fenotipu, ko novēro 91% pacientu.

Pacientu genotipa izmeklēšana sniedz papildu informāciju par iespēju izmantot šaperonu terapiju. Atbilstoši pētījuma rezultātiem 13 no 70 (18%) Latvijas FKU pacientiem varētu būt efektīva šaperonu terapija ar sapropterīna dihidrohlorīdu, savukārt atlikušajiem 57 (81%) ar homozigotu R408W mutāciju vai kompaunda heterozigotiem ar smagām mutācijām abās alēlēs jāturpina zema fenilalanīna diēta kā vienīgā efektīvā terapijas forma.

Minihaplotipu izpēte atklāja 16 dažādus minihaplotipus saistībā ar *FAH* gēna mutācijām un 20 dažādus minihaplotipus normālajās *FAH* alēlēs. Vidējā heterozigotātes iespēja bija aptuveni 76% mutantajās un 92% normālajās alēlēs, kas parāda normālo alēļu lielāku dažādību. Statistiskā analīze parādīja ticamu atšķirību divu minihaplotipu izplatībā starp mutantām un normālām alēlēm 3/238 (p=0,0000572) un 8/230 (p=0,0133), savukārt trīs 3/242, 7/246 un 8/234 parādīja statistiski ticamu tendenci (p< 0,10) izplatības atšķirībai starp mutantām un normālām alēlēm.

FAH gēna R408W mutācijas cieša saistība ar Austrumeiropai raksturīgu minihaplotipu un tās izplatības analīze apstiprina šīs alēles Baltijas-slāvu izcelsmi un ienākšanu citās populācijās ar salīdzinoši nesenām cilvēku migrācijām visā Eirāzijas kontinentā.

Pētījumā izmantotā trīs posmu FAH gēna mutāciju noteikšanas metode ir efektīvākā ikdienas diagnostikai Latvijas FKU/HFA pacientiem ar analīzes sistēmas jutību 99%.

INTRODUCTION

Metabolic diseases are life threatening inheritable, genetic disorders in which errors of metabolism occur, involving a block where a catalyst or enzyme is absent or malfunctioning. The most part of inherited metabolic disorders are identifies as inborn errors of metabolism or IEMs. IEMs are normally defined as diseases of amino acids, organic acids, the urea cycle, galactosemia, primary lactic acidoses, glycogen storage diseases, lysosomal storage diseases, and diseases involving peroxisomal and mitochondrial respiratory chain dysfunction (Villas-Boas, 2007).

The incidence of inherited metabolic diseases varies from 1 in 10 000 people to diseases that are very rare and affect around 1 in 100 000, and lead to a wide range of special needs in care and education (Hannigan, 2007).

Phenylketonuria (PKU) is the most common inborn error of amino acid metabolism in Europeans with the incidence of about one case per 10 000 live births. It has been frequently described as a paradigm of a Mendelian disorder. PKU was the first metabolic cause of mental retardation to be identified, the first genetic disorder of the central nervous system that could be fully treated by modification of external factors (i.e., the diet), and the first disorder that was successfully diagnosed by universal neonatal screening.

The understanding of the biochemical and molecular basis of PKU was very important for the treatment strategies introduced for these patients and led to significant reduction in morbidity and to an improvement in quality of life.

Untreated PKU is associated with an abnormal phenotype which includes growth failure, poor skin pigmentation, microcephaly, seizures, global developmental delay and severe intellectual impairment. However, since the introduction of newborn screening programs and with early dietary intervention, children born with PKU can now expect to lead relatively normal lives.

The phenylalanine hydroxylase (*PAH)* gene (OMIM 261600) (http://www.pahdb.mcgill.ca) was first cloned in 1983 (Woo, 1983). The most studies of *PAH* gene mutations in different populations were carried out in the 1990s.

Most forms of PKU and hyperphenylalaninaemia (HPA) are caused by mutations in the *PAH* gene on chromosome 12q23.2. Over 560 disease causing mutations have been reported, most of them corresponding to point mutations causing

missense changes. Mutations differ in residual enzyme activity, and the genotype could be a good predictor of biochemical phenotype in the majority of patients.

Despite the fact that dietary treatment remains the mainstay of PKU management, it is multifaceted, challenging, and life long. Key dietary behaviours associated with optimal control of blood phenylalanine (Phe) concentrations include avoidance of high-protein foods and supplementation with special medical foods that often are unpalatable. Dietary compliance is influenced by cognitive, emotional, psychological, and cultural factors. Non-compliance with the dietary prescription is commonplace, particularly during adolescence and adulthood. For all these reasons, the contemporary interest in PKU more often is focused on the development of new therapeutic approaches, but it is still closely related to the biochemical and molecular basis of PKU.

Chaperon therapy by sapropterin dihydrochloride is a novel therapeutic approach that is effective in a subset of individuals with PKU. It is specifically indicated to reduce blood Phe levels in patients with HPA due to tetrahydrobiopterin-(BH4-) responsive PKU and could be used alone or in conjunction with a Phe-restricted diet. The effectiveness of the chaperon therapy depends on mutation residual activity in the *PAH* gene. So, the genotypes can provide the additional information for BH_4 responsiveness and are taken into account in selecting the type of PKU treatment.

Population studies have revealed at least 87 different haplotypes for *PAH* gene mutations but only a few are prevalent, and most are uncommon. Haplotypes 1 through 4 account for more than 80% of PKU-bearing chromosomes and are mainly used to determine the origin of mutations.

The high degree of polymorphism and strong Mendelian segregation of minihaplotypes (combination of VNTR and STR systems) considerably increases the number of cases that are informative and makes it useful for prenatal diagnosis, detection of rare mutations, and carrier screening determination in PKU families.

Among the most common *PAH* mutations is R408W. In eastern European populations, the R408W mutation is strongly associated with RFLP haplotype 2, the three-copy VNTR allele (VNTR 3), and the 238-bp STR allele. Therefore, we expect a high incidence of this mutation in our study.

Large number of Latvian PKU patients has classic (severe) clinical PKU phenotype that could be possible because of the prevalence of severe mutations in Latvian PKU chromosomes.

THE AIM OF THE STUDY:

The aim of the study was to investigate the molecular basis of phenylketonuria and hyperphenylalaninemia in Latvian patients and evaluate a *PAH* gene mutation diagnostic strategy in Latvian population.

THE TASKS OF THE STUDY:

1. To investigate the *PAH* gene mutation spectrum for Latvian PKU patients and their parents;

2. To detect the association between minihaplotypes and mutations at the *PAH* locus in Latvian PKU patients and their parents;

3. To compare distribution of minihaplotypes in mutant and normal *PAH* chromosomes;

4. To estimate the genotype-phenotype correlation in patients with PKU;

5. To compare the frequency of the mutation R408W in Latvian PKU chromosomes with the frequencies of this mutation in other PKU populations in Europe;

6. To evaluate the diagnostic techniques for the introduction of a *PAH* gene mutation detection strategy in the routine diagnostics in Latvia.

Scientific Novelty of the Study

This study is the first study in the Baltic States region investigating minihaplotype associations for full mutation spectrum including rare and novel mutations observed in the *PAH* gene.

Three novel nucleotide changes were identified and two of them assumed to be disease causing while the third one is going to be mild mutation causing MHP.

R408W was found on high relative frequencies and typical Eastern-European minihaplotype 3/238 confirming the Balto-Slavic origin of the mutation.

Practical Novelty of the Study

The evaluation of patients' genotypes provides the additional information for detection of clinical phenotype for PKU patients and selecting of appropriate therapy. Chaperon therapy was introduced for four Latvian PKU patients.

Results of minihaplotypes studies are useful for prenatal diagnosis, carrier screening and detection of rare mutations in Latvian PKU chromosomes.ons

Theses to be defended

- Common phenylketonuria mutation for Latvian patients is R408w tah is one of predominant in the *PAH* gene in European populations and the most predominant in Eastern European populations.
- The three-step *PAH* gene mutation detection strategy used in the study is the most effective for routine diagnostics in Latvian population.

1 LITERATURE REVIEW

1.1 HUMAN METABOLISM AND METABOLIC DISEASES

Metabolism is the sum total of all the chemical reactions constituting the continuing process of breakdown and renewal of the tissues of the human body. Enzymes play an indispensable role in facilitating the process by serving as catalysts in the conversion of one chemical (metabolite) to another, often extracting the energy required for the reaction from a suitable high-energy source, such as ATP. Mutations might affect enzyme activity by affecting the steady-state amount of enzyme protein because of a defect in enzyme production or as a result of abnormally rapid breakdown of the mutant protein. Alternatively, mutations might impair the activity of the enzyme without affecting the amount of enzyme protein by specifically impairing the catalytic properties of the protein (Clarke, 2006).

Metabolic diseases are life threatening inheritable, genetic disorders in which errors of metabolism occur, involving a block where a catalyst or enzyme is absent or malfunctioning. This defect results in the build up of chemicals on one side of the metabolic blockage and a deficiency of vital chemicals on the other. This causes an overdose of one or more, often toxic, chemicals and the shortage of others, which are essential to normal body functioning. The consequences of such chemicals are often fatal, leading to either a slow deterioration with progressive physical and mental issues, or to rapid decline and death (Anonymous, 2011).

Classically, most inherited metabolic disorders are identifies as inborn errors of metabolism or IEMs. IEMs are normally defined as diseases of amino acids, organic acids, the urea cycle, galactosemia, primary lactic acidoses, glycogen storage diseases, lysosomal storage diseases, and diseases involving peroxisomal and mitochondrial respiratory chain dysfunction (Villas-Boas, 2007).

The incidence of inherited metabolic diseases varies from 1 in 10 000 people to diseases that are very rare and affect around 1 in 100 000, and lead to a wide range of special needs in care and education (Hannigan, 2007).

Inborn errors of metabolism are individually rare, but collectively numerous. From a pathophysiological perspective, metabolic disorders can be divided into following three diagnostically useful groups:

- Group 1: disorders which dive rise to intoxication

- Group 2: disorders involving energy metabolism

- Group 3: disorders involving complex molecules

Group 1 includes inborn errors of intermediary metabolism that lead to an acute or progressive intoxication from the accumulation of toxic compounds proximal to the metabolic block. In this group are the inborn errors of amino acid catabolism (phenylketonuria, maple syrup urine disease, homocystinuria, tyrosinemia etc.), most organic acidurias (methylmalonic, propionic, isovaleric etc.), congenital urea cycle defects, sugar intolerances (galactosemia, hereditary fructose intolerance), metal intoxications (Wilson, Menkes, hemochromatosis), and porphyrias. Clinical conditions in this group do not interfere with the embryo-foetal development and present with a symptom-free interval and clinical signs of "intoxication", which may be acute (vomiting, coma, liver failure, thromboembolic complications) or chronic (failure to thrive, developmental delay, ectopia lentis, cardiomyopathy). Circumstances that can provoke acute metabolic attacks include catabolism, fever, intercurrent illness and food intake. Clinical expression is often both late in onset and intermittent (Fernandes, 2006).

Group 2 consists of inborn errors of intermediary metabolism with symptoms due at least partly to a deficiency in energy production or utilisation within liver, myocardium, muscle, brain or other tissues. This group can be divided into mitochondrial and cytoplasmic energy defects. Mitochondrial defects are the most severe and are generally untreatable (congenital lactic acidemias, mitochondrial respiratory chain disorders, fatty acid oxidation and ketone body defects, that are partly treatable). Cytoplasmic energy defects are generally less severe. They include disorders of glycolysis, glycogen metabolism and gluconeogenesis, hyperinsulinism (all treatable disorders), disorders of creatine metabolism (partly treatable), and the new inborn errors of the pentose phosphate pathway (untreatable). Common symptoms are hypoglycaemia, hyperlactatemia, hepatomegaly, severe generalised hypotonia, myopathy, cardiomyopathy, failure to thrive, cardiac failure, circulatory collapse, sudden unexpected death in infancy, and brain involvement. Some of the mitochondrial

disorders and pentose phosphate pathway defects can interfere with the embryo-foetal development and give rise to dysmorphism, dysplasia and malformations (Fernandes, 2006).

Group 3 involves cellular organelles and includes diseases that disturb the synthesis or the catabolism of complex molecules. Symptoms are permanent, progressive, independent of intercurrent events and unrelated to food intake. All lysosomal storage disorders, peroxisomal disorders, disorders of intracellular trafficking and processing such α-1-antitrypsin, congenital disorders of glycosylation, and inborn errors of cholesterol synthesis belong to this group. Almost none are treatable acutely, but enzyme replacement therapy is now available for several lysosomal disorders (Fernandes, 2006).

Considering that part of IEMs is treatable or partly treatable the early diagnosis is very important for the successful treatment outcome.

1.1.1. Inheritance of metabolic diseases

Each of the different metabolic diseases is inherited through one of four routes:

- Autosomal dominant inheritance
- Autosomal recessive inheritance
- MtDNA inheritance
- X-linked inheritance

Autosomal dominant inheritance occurs when a single copy of the diseased gene will dominate the other normal gene. Therefore if a defective gene is inherited from either parent, the child will be affected with the disorder. There is a 50% chance of a child being affected by the disease if either of the parents has a defective gene.

Autosomal recessive inheritance occurs when a child inherits a gene for the disease from both parents. The risk that the offspring of a couple who are both carriers of the disease will be affected is 25%. There is a 50% chance that their child will be a carrier, and a 25% chance that the child will not carry the abnormal gene.

Disorders inherited through the mitochondrial DNA are only passed down from mothers to their children. During fertilisation the mtDNA from father is lost. This means that girls will always pass on a defect in their mtDNA and boys will never pass on a defect in their mtDNA to their children.

X-linked inheritance occurs when diseases are coded on the X chromosomes of genes. For females any disease trait on one of the two X chromosomes is usually masked by the other normal X chromosome. Males inherit the disease because of only one X chromosome (Hannigan, 2007).

1.1.2 Phenotype-genotype correlations for IEMs

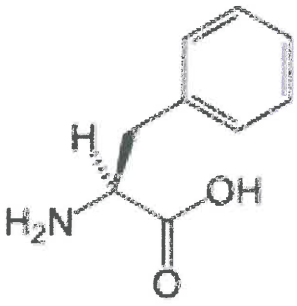
Mutations may affect gene products in many ways; the effects of a single mutation also vary tremendously. Some mutations may totally disrupt the production of any gene product, resulting in severe disease. By contrast, other mutations might have no effect whatsoever apart from a functionally silent change in the nucleotide sequence of the gene. The relationship between genotype and disease phenotype is complex. Severe mutations, such as deletions or insertions, are generally associated with clinically severe disease, and the disease phenotype among different affected individuals tends to be similar. Structurally more subtle mutations, such as those resulting in single amino acid substitutions, are often associated with milder disease phenotypes. Moreover, the disease phenotype often varies markedly between different affected individuals, even within the same family, a reminder that the expression of any genetic information, including disease-causing mutations, is influenced by other genes (gene-gene interactions) and by environmental factors (gene-environment interactions) (Clarke, 2002).

1.2. PHENYLALANINE

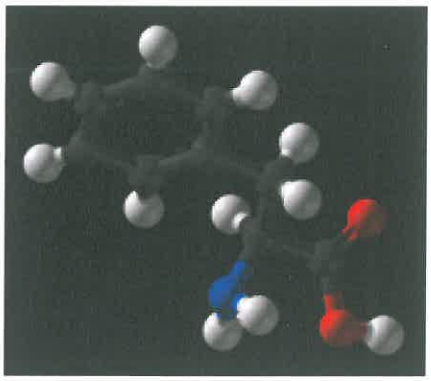
Phenylalanine (Phe) is an essential amino acid that occurs as a constituent of many proteins and is a precursor for tyrosine, the monoamine signalling molecules dopamine, norepinephrine (noradrenaline), and epinephrine (adrenaline), and the skin pigment melanin. Phenylalanine is necessary for growth in infants and for nitrogen equilibrium in adults.

Phenylalanine is a derivative of alanine with a phenyl substituent on the β carbon (Fig. 1.1). This essential amino acid is classified as nonpolar because of the hydrophobic nature of the benzyl side chain. Due to its hydrophobicity, even the free amino acid is not very soluble in water and phenylalanine is nearly always found buried

within a protein. Phenylalanine is an aromatic amino acid and the π electrons of the phenyl ring can stack with other aromatic systems and often do within folded proteins, adding to the stability of the structure (Scott, 1988).



a)



b)

Fig. 1.1. Chemical formula [C6H5CH2CH(NH2)COOH] (a) and 3-D molecular model (b) of phenylalanine. Hydrogen is marked in white sphere, carbon in grey, nitrogen in blue and oxygen in red (source: http://en.wikibooks.org/).

Phenylalanine was first isolated in 1879 and first synthesized in 1882 (McGraw-Hill, 2003). It is found in three forms: L-phenylalanine, the natural form found in proteins; D-phenylalanine (a mirror image of L-phenylalanine that is made in a laboratory), and DL-phenylalanine, a combination of the two forms.

L-Phenylalanine (LPA) is white, powdery solid and an electrically-neutral amino acid, one of the twenty common amino acids used to biochemically form proteins, coded for by DNA. The codons for L-phenylalanine are UUU and UUC.

L-phenylalanine is found in most foods that contain protein such as beef, poultry, pork, fish, milk, yogurt, eggs, cheese, soy products (including soy protein isolate, soybean flour, and tofu), and certain nuts and seeds. A non food source of phenylalanine makes up 50% of artificial sweetener aspartame. This compound is metabolised by the body into several chemical by products including phenylalanine (Nelson & Cox, 2000).

1.2.1 Phenylalanine metabolism

Phenylalanine and tyrosine are discussed together, since tyrosine results from hydroxylation of phenylalanine acid and is the first product in phenylalanine degradation. Because of this, tyrosine is not usually considered to be essential, whereas phenylalanine is. Three-quarters of indigested phenylalanine is metabolised to tyrosine.

This is catalysed by phenylalanine hydroxylase (Fig. 1.2), which is tetrahydrobiopterin dependent. This reaction occurs only in the direction of tyrosine formation, and phenylalanine cannot be synthesized from tyrosine (Devlin, 1997).

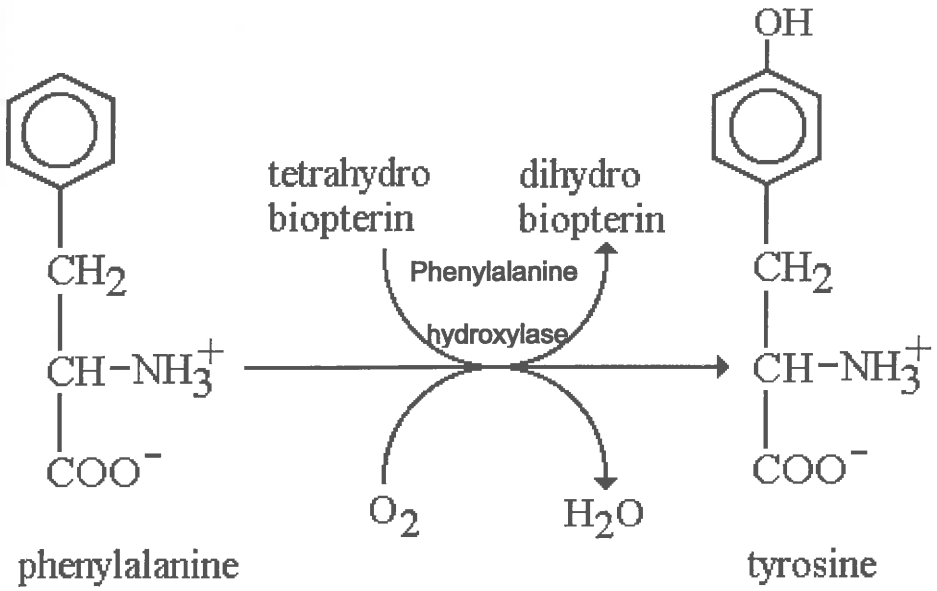


Fig. 1.2. Phenylalanine conversion to tyrosine. The enzyme phenylalanine hydroxylase is responsible for phenylalanine conversion to tyrosine, a mixed function oxygenase with a tetrahydrobiopterin cofactor. Half of the oxygen molecule re-appears in the tyrosine -OH group and the other half is reduced to water. The dihydrobiopterin in the above reaction is an isomer of the folic acid compounds involved in one-carbon metabolism (source: http://www.bmb.leeds.ac.uk/illingworth/bioc1010/index.htm).

Tyrosine (Tyr) is a precursor of the catecholamine hormones: dopamine, noradrenaline and adrenaline; and also the thyroid hormone thyroxine as well as melanine. Catecholamine synthesis (Fig. 1.3) starts with tyrosine hydroxylase that, like phenylalanine, is also dependent on tetrahydrobiopterin. The cofactor initiates the catecholamine synthesis in keratinocytes and melanocytes. Tyrosine hydroxylase produces dihydroxyphenylalanine, also known as DOPA, dioxophenylalanine. DOPA decarboxylase forms dopamine, the active neurotransmitter, from DOPA. In the substantia nigra and some other parts of the brain, this is the last enzyme in this pathway. The adrenal medulla converts dopamine to norepinephrine and epinephrine. Brain plasma tyrosine regulates norepinephrine formation. Estrogens decrease tyrosine concentration and increase tyrosine aminotransferase activity, diverting tyrosine into the catabolic pathway (Devlin, 1997).

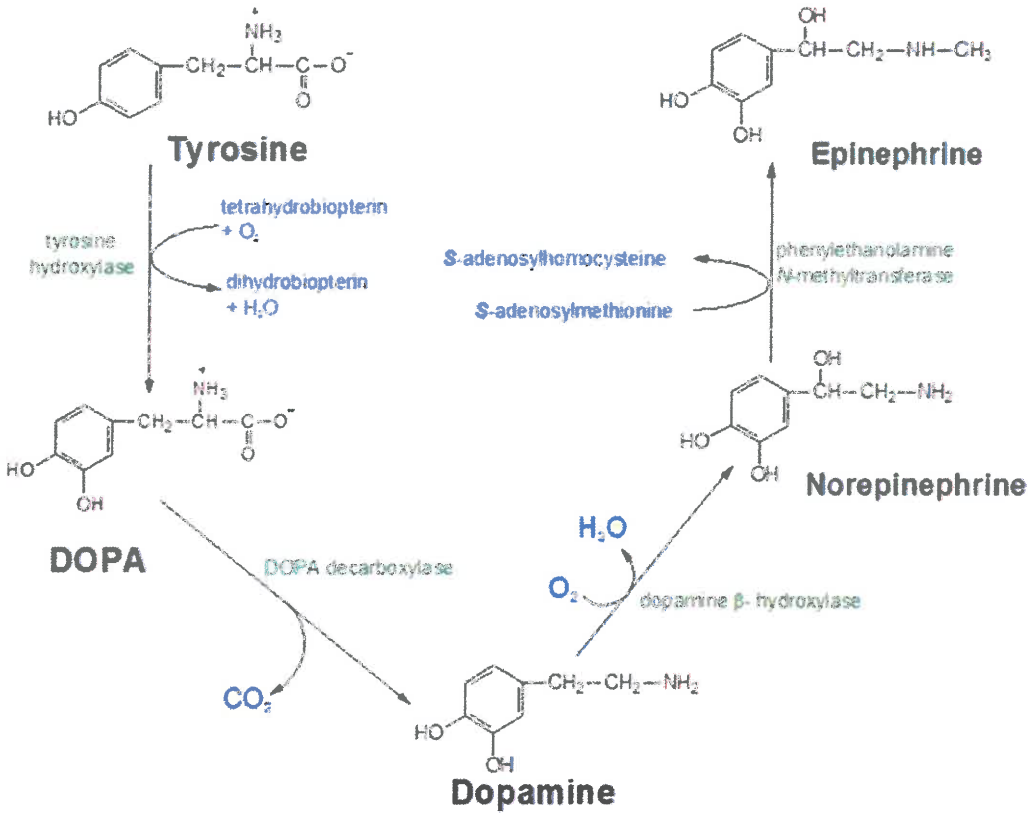


Fig. 1.3. Metabolism of tyrosine (http://themedicalbiochemistrypage.org/nerves.html).

In the presence of a defect in phenylalanine hydroxylase the first compound that accumulates is phenylalanine itself. There are two routes by which the excess phenylalanine can be metabolised: oxidation to tyrosine (the normal and main route for degradation of Phe, and the normal route for biosynthesis of Tyr), and transamination to phenylpyruvate and subsequent further metabolism (a minor route, which comes to the fore when the main route is blocked) (Fig. 1.4). If the enzyme, phenylalanine hydroxylase, which converts phenylalanine into tyrosine, is missing, then phenylalanine undergoes a transamination reaction to make phenylpyruvic acid (Williams, 2008).

Transamination of phenylalanine to form phenylpyruvate normally does not occur unless circulating concentrations exceed 1.2 mmol/L. The elevated level of phenylalanine effects the energy production, protein synthesis, and neurotransmitter homeostasis in the developing brain. Phenylalanine uses the same active transport channel as tryptophan that is the precursor of several substances, including serotonin and niacin, to cross the blood - brain barrier, and, in large quantities, interferes with the production of serotonin. So phenylalanine inhibits the transport of neutral amino acids

across the blood-brain barrier, leading to a selective amino acid deficiency in the cerebrospinal fluid (Scriver, *et al.*, 1995).

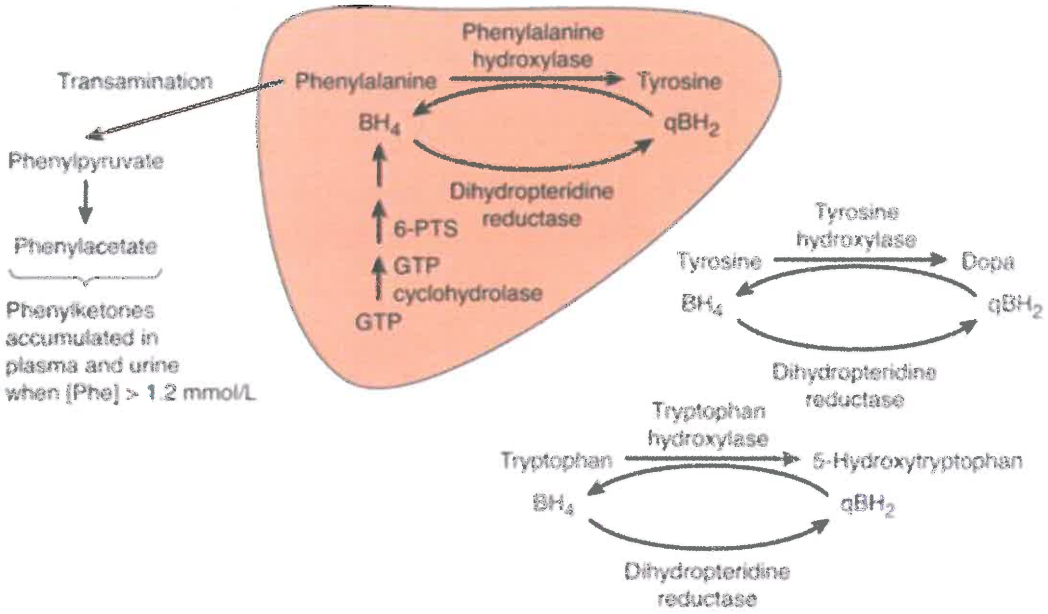


Fig. 1.4. Normal (right) and abnormal (left) phenylalanine metabolism in human (McPhee and Hammer, 2010)

The excessive accumulation of phenylalanine in plasma and tissues and it's metabolism to phenylpyruvate, phenyllactate, and phenylacetate, collectively known as phenylketones, results in the most common inherited metabolic disorder called phenylketonuria (PKU). There are a variety of secondary effects of the accumulation of phenylalanine and its metabolites. Decreased pigmentation has been related to the inhibition of tyrosinase by phenylalanine. Decreased levels of serotonin appear to be due inhibition of 5-hydroxytryptophan decarboxylase by phenylpyruvic, phenyllactic and phenylacetic acids. Decreased amount of epinephrine, norepinephrine and dopamine are presumably caused by inhibition of dopamine decarboxylase. The metabolites that accumulate in PKU also inhibit glutamic acid decarboxylase in brain, and this would decrease levels of 4-aminobutiric acid (GABA) that regulates neuronal excitability throughout the nervous system (Nyhan, *et al.*, 2005).

1.2.2 Human Phenylalanine Hydroxylase enzyme

Phenylalanine hydroxylase (PheOH, 1 phenylalanine 4-monooxygenase, EC 1.14.16.1) is an iron- and tetrahydropterin-dependent enzyme that catalyses the hydroxylation of L-phenylalanine (L-Phe) to L-tyrosine. The reaction is the rate-limiting step in the catabolic pathway of phenylalanine resulting in the complete degradation of the amino acid. PheOH activity is tightly regulated by reversible phosphorylation and substrate activation (Fusetti, 1998).

Human PheOH exists in a pH-dependent equilibrium of homotetramers and homodimers, and is organised into three domains: a 12–19kDa N-terminal domain (residues 1–142) involved in regulation of the enzyme activity followed by a segment of about 38 kDa, which consist of a catalytic domain (residues 143–410) and a 5kDa C-terminal tetramerisation domain (residues 411- 452) (Fig. 1.5) (Erlandsen, 1997; Erlandsen, 2003).

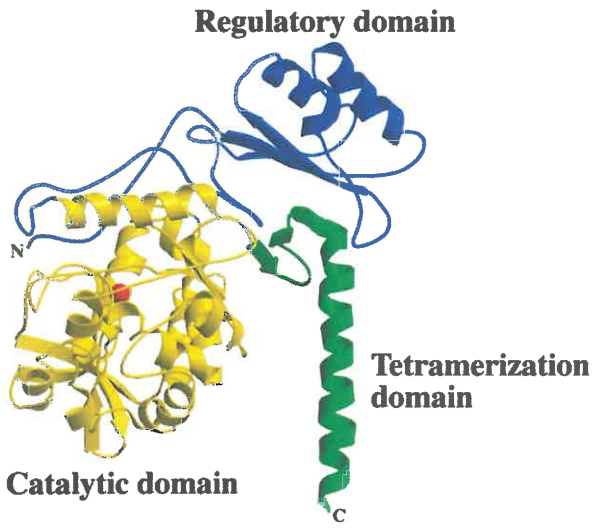


Fig. 1.5. Structure of a monomer of human PAH full-length composite model. The regulatory domain (residues 1–142), the catalytic domain (residues 143–410), and the tetramerisation domain comprise the full-length monomer. The regulatory domain is coloured blue, the catalytic domain is yellow, and the tetramerisation domain is green. The iron is shown as a red sphere (source: PAHdb: http://www.pahdb.mcgill.ca/Information/Molecular/full_dom.html).

PheOH dimeric form contains the regulatory and catalytic domains, a tetrameric form contains the catalytic and tetramerisation domains (Erlandsen, 2003).

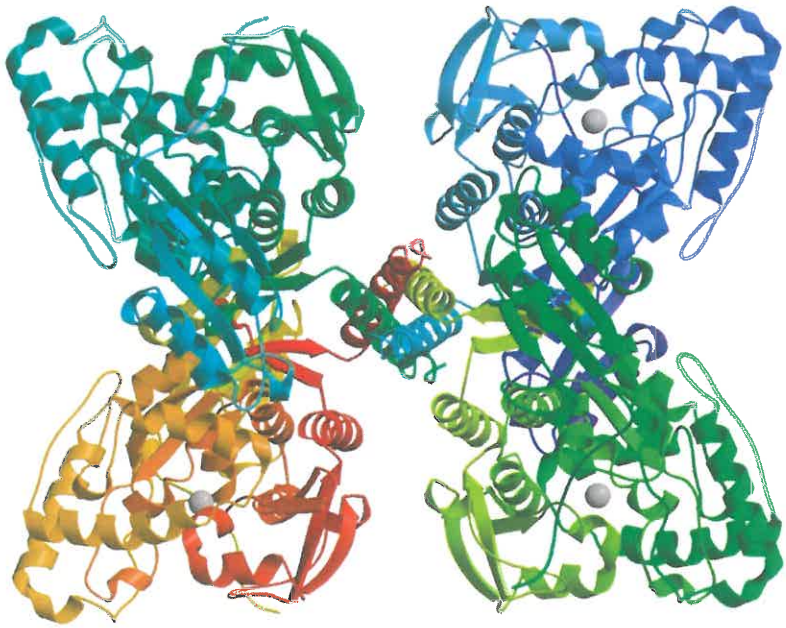


Fig. 1.6. Structure of the phenylalanine hydroxylase full-length model (the structures of catalytic/tetramerization domains and regulatory/catalytic domains were used). The model is coloured from red (N-terminus in monomer A) to blue (C-terminus of monomer D). The iron is shown as a gray sphere in all four monomers making up the tetramer (http://www.pahdb.mcgill.ca/Information/Molecular/full.html).

The regulatory domain of PAH contains an α-β sandwich with an interlocking double $\beta\alpha\beta$ motif ($\beta\alpha\beta\beta\alpha\beta$ topology). The N-terminal autoregulatory sequence (ARS; residues 19–33) extends over the active site in the catalytic domain. An N-terminal truncated form of PAH that includes the catalytic and tetramerisation domains (residues 116–452) crystallised as a tetramer (dimer of dimers). The tetramerisation domain contains two β-strands, forming a β-ribbon, and a 40 Å (4.0 nm) long α-helix (Fig. 1.5). The 4 α-helices (1 from each monomer) pack into a tight antiparallel coiled-coil motif in the center of the tetramer structure. The catalytic domain contains 13 α-helices and 8 β-strands. The active site of human PheOH C-term is located in a deep cleft in the core of each monomer (Fig. 1.6) (Erlandsen, 1997; Erlandsen, 2003).

Phenylalanine hydroxylase is activated by its substrate phenylalanine and through phosphorylation by cAMP-dependent protein kinase at Ser16 in the N-terminal autoregulatory sequence of the enzyme (Horne, 2002).

1.3 PHENYLKETONURIA – PKU

1.3.1 Clinical abnormalities of PKU patients

Phenylketonuria (PKU) is a disorder of aromatic amino acid metabolism in which phenylalanine cannot be converted to tyrosine. The most important and sometimes the only manifestation of PKU is mental retardation. The intelligence of untreated patients is very low, with intelligence quotients usually under 50 but sometimes patients with untreated PKU could have borderline intelligence.

Phenylketonuric infants appear normal at birth. Retardation of development may not be evident of months. Vomiting may be a prominent early symptom. Irritability, an eczematoid rash, and an unusual odour may also be observed very early in life. The odour of the phenylketonuric patient is that of phenylacetic acid. It has variously been described as mousy, barny, wolf-like or musty (Nyhan, 2005).

Patients with PKU are often good looking children. They are fair-haired, fair-skinned and blue-eyed in over 90 percent of the cases (Okano, 1991). The dermatitis is usually mild, and it is absent in three-quarters of the patients. Patients may complain of intractable itching in the absence of visible cutaneous lesions. Sclerodermatous skin also could be infrequently reported in infant with PKU.

Neurological manifestations are not usually prominent, but about a third of the patients may have all of the signs of cerebral palsy. They are spastic, hypertonic and have increased deep tendon reflexes. Only about 5 % have these manifestations to a severe degree. They may have contractures and limitation of mobility. Hyperactivity is quite common, and there may be abnormalities of gait. Another one-third of the patients have very mild neurological signs such as a unilateral Babinski response or hyperactive deep tendon reflexes. Another third of untreated patients have no neurological signs except for mental retardation (Nyhan, 2005).

Seizures occur in about a fourth of the patients. They are usually neither prominent nor difficult to manage. Nevertheless, about 80% have electroencephalograph (EEG) abnormalities. Hyperactivity and behaviour problems are common. Purposeless movements, rhythmic rocking, stereotypy, tremors and athetosis may be seen. Somatic development tends to be normal, but stature may be short. Some patients have minor malformations like widely spaced teeth, pes planus, partial syndactityly, and epicantus. Congenital heart disease appears to be more common in PKU than in the

general population (Verkerk, 1991). Some patients have microcephaly. In the past a majority with untreated PKU required institutional care (Nyhan, 2005).

1.3.2 Diagnosis of PKU

The diagnosis of PKU is made in the neonatal period. This is accomplished by the routine screening of all infants for an elevated concentration of phenylalanine in the blood. It is generally carried out on discharge from hospital after the initiation of protein-containing feedings. A drop of blood collected from the heel on filter paper is analysed for phenylalanine by bacterial inhibition method developed by Guthrie, or by a quantitative determination of the concentration of phenylalanine. A positive screening test is usually repeated from the same blood sample (Nyhan, 2005). The filter paper spots are stable for many years and the PKU screening tests have been reported to have a low error rate (Scriver, 2001).

There are several analytical techniques that can be used for quantitative and semi-quantitative analysis of phenylalanine levels from dry blood spots. With the introduction of expanded newborn screening, most screening laboratories now use tandem mass spectrometry (MS/MS) for the analysis of amino acids including phenylalanine and acylcarnitine species (Bodamer, 2010).

The first test to screen large numbers of newborn infants for the presence of elevated levels of blood phenylalanine was the bacterial inhibition test. It was developed by Dr. Robert Guthrie in 1960 in Buffalo, NY, and becomes known as the Guthrie Test (Schuett, 2009). Bacterial growth on an agar plate is inhibited through the action of a particular chemical ('inhibitor') on a small disc in the middle of the plate. Any structurally related compound, e.g. amino acid (phenylalanine) or metabolite will compete with the inhibitor and initiate bacterial growth. The growth zone around the punch will be proportionate to the amount of phenylalanine, e.g. blood concentration that is brought onto the plate with a dry blood punch. The size of bacterial growth zones from standard blood samples can be compared with those from individual neonatal samples and the blood phenylalanine concentration deduced accordingly in a semiquantitative manner. Antibiotics given to the mothers and/or infants may interfere with the results of any bacterial inhibition test as bacterial growth may have been inhibited (Guthrie, 1996).

Tandem Mass Spectrometry MS/MS was introduced to newborn screening laboratories during the late 1990s (Dhondt, 2007; Bodamer, 2007). The main advantage of this technique is the simultaneous, fully automated analysis of different analytes such as amino acids including phenylalanine and acylcarnitine species. The diagnostic sensitivity of MS/MS for hyperphenylalaninemia (HPA)/PKU is superior compared to other analytical techniques. The measurement of the phenylalanine to tyrosine ratio may help in differentiating between false positives and cases of HPA/PKU (Bodamer, 2010).

Diagnosis of hyperphenylalaninaemia (HPA) is made on the basis of an elevated blood Phe concentration on a repeat blood sample. The upper reference limit for Phe in whole blood or plasma in neonates is <150 μmol/L and slightly lower (<120 μmol/L) in older children (Williams, 2008). Early detection is desirable in order to introduce appropriate treatment and prevent mental retardation (Blau, 2003).

1.3.3 Classification of PKU clinical forms

Phenylketonuria is classified by the severity of hyperphenylalaninaemia. The normal range of blood phenylalanine concentrations is 50–120 μmol/L (Blau, 2010). On the basis of blood Phe concentrations before starting treatment, PAH deficiency can be classified into classic (severe) PKU (Phe >1200 μmol/L), mild (atypical or variant) PKU (Phe = 600–1200 μmol/L) and mild HPA (MHP), where blood Phe is elevated above upper reference limit, but <600 μmol/L.1 (Williams, 2008). Sometimes a moderate classification is included for concentrations of 900–1200 μmol/L (Blau, 2010).

Classification is not always straightforward because phenylalanine concentrations are measured in newborn babies when blood phenylalanine might not have had time to reach its highest value. Classification can also be made on the basis of tolerance for dietary phenylalanine while on diet, which is not always easily and accurately measured. This tolerance is usually not greater than 250 mg per day in classic phenylketonuria, whereas in mild or even moderate phenylketonuria, phenylalanine tolerance can range from 250 to 400 mg per day (Güttler, 1980; Blau, 2010).

1.3.4 PKU treatment

Early detection and treatment of PKU prevent the most obvious and severe consequences of this disorder. The primary treatment of this condition is to reduce the levels of Phe in the blood, thus limiting Phe accumulation in the tissues and minimizing neurotoxic effects seen in patients with elevated blood levels of this amino acid. This is accomplished through lifelong dietary restriction of Phe and supplementation with a formula containing all required amino acids excluding Phe (Gentile, 2010).

Diet therapy. The restriction of dietary phenylalanine remains the mainstay of phenylketonuria management, and usually begins immediately after confirmation of hyperphenylalaninaemia in a neonate. Patients with phenylketonuria have to accept the phenylalanine-free formula and avoid foods rich in protein (egg, meats, fish, eggs, standard bread, most cheeses, nuts, and seeds) and foods and drinks containing aspartame, flour, soya, beer, or cream liqueurs. Low-protein natural foods such as potatoes, some vegetables, and most cereals can be eaten but only in severely restricted amounts. Low-protein variants of some foods exist, such as low-protein bread and low-protein pasta. The required amount of daily protein is largely obtained from manufactured, commercially available phenylalanine-free protein substitutes (Blau, 2010).

During infancy, adherence to the diet is straightforward because the child's parents control nutritional intake. As children get older, adherence to the diet becomes increasingly difficult because meals have to be planned rigorously and children cannot choose the food routinely consumed by their peers (Blau, 2010).

Consequently, compliance with the diet is often poor, especially when the patient reaches adolescence, as evidenced by poor control of blood phenylalanine concentrations in this age group (Walter, 2004; Crone, 2005). Long-term maintenance of the diet is important, because patients find it difficult to return to adequate dietary compliance after a period of eating an unrestricted diet. The difficulty of the dietary regimen, psychosocial and emotional factors, issues relating to family cohesion, commitment of parents to maintaining the diet, knowledge of the disease and attitudes to health-care professionals could be the reasons for suboptimum dietary compliance (Crone, 2005; Bekhof, 2003; Olsson, 2007).

Treated individuals with PKU do not appear sick and often do not feel the effects of poor metabolic control. However, even early- and well-treated patients experience hidden disabilities, such as subtle deficits in executive function (EF), mild

reductions in mental processing speed, social difficulties, and emotional problems that may remain unnoticed for years (Gentile, 2010).

Chaperon therapy by sapropterin dihydrochloride. Sapropterin dihydrochloride is an enzyme cofactor and oral form of tetrahydrobiopterin (BH4). BH4 supplementation is a novel therapeutic approach that is effective in a subset of individuals with PKU. PKU due to PAH deficiency is not associated with BH4 deficiency, but in a subset of individuals with PKU, oral supplementation with further BH4 can lead to reduction in blood Phe concentration. About two-thirds of patients with mild PKU are BH4-responsive (Harding & Blau, 2010).

Three mechanisms involved in BH4-responsiveness have been postulated, including (1) reduced binding affinities for BH4 in PAH Km mutants, (2) the stabilisation of mutant PAH by BH4 through protection of the active tetramers or dimers from either cleavage or degradation by the ubiquitin-dependent proteasome pathway, and finally (3) effects of exogenous BH4 supplementation upon the regulation of BH4 biosynthesis. It is quite likely that the mechanism underlying BH4-responsiveness will be multifactorial (Pey, 2004; Gersting, 2008).

Some mutations are associated with a BH4-sensitive phenotype of phenylketonuria in which giving pharmacological doses of exogenous BH4 results in an increase in the activity of PAH that is sufficient to reduce circulating phenylalanine to a therapeutically relevant extent (Zurflüh, 2008; Kure, 1999). These mutations usually present with substantial residual activity when expressed recombinantly in eukaryotic cell systems and are located in all regions of PAH (Blau, 2010).

1.3.5 Possible PKU treatment strategies in future

Large neutral amino acids. Because phenylalanine competes with other large neutral amino acids for transport across the blood–brain barrier, supplementation with these amino acids other than phenylalanine could provide another potential treatment approach. Giving these amino acids after an oral phenylalanine load abolishes the rise in phenylalanine in the brain of patients with phenylketonuria (Pietz, 1999). A double-blind, placebo-controlled study also showed a reduction in blood phenylalanine from baseline of 39% during short-term treatment with large neutral amino acids. Additionally, large neutral amino acid treatment seems to have a beneficial effect on executive functioning (Matalon, 2007; Schindeler, 2007).

Phenylalanine ammonia lyase. Phenylalanine ammonia lyase is a bacteria-derived enzyme that catalyses the conversion of L-phenylalanine to transcinnamic acid and ammonia without a cofactor requirement (MacDonald, 2007). In the mouse model of phenylketonuria, blood and brain concentrations of phenylalanine were reduced during 90 days of treatment with injections of phenylalanine ammonia lyase (Sarkissian, 2008).

1.3.6 Epidemiology of PKU

The prevalence of phenylketonuria varies widely around the world. In Europe the prevalence is about one case per 10 000 live births, but for some areas of Europe it is higher. Persistent hyperphenylalaninaemia is detected in about one in every 4000 births in Turkey because of high consanguinity within the population, and in Northern Ireland. Finland has the lowest prevalence in Europe with one case per 100 000. In the USA the prevalence is one case per 15 000. In Latin America it varies from about one case per 50 000 to one per 25 000 births; prevalence is generally higher in southern Latin America than elsewhere in that region. Estimates of prevalence rates in Asia vary from about one per 15 000 to one per 100 500 births in regions of China, less than one per 200 000 in Thailand, and about one per 70 000 in Japan. Africa seems to have a very low prevalence of phenylketonuria and Spain has an especially high prevalence of mild hyperphenylalaninaemia (Blau, 2010; Loeber, 2007; Ozalp, 2001; Zschocke, 1997; Guldberg, 1995; Borrajo, 2007; Zhan, 2009; Jiang, 2003; Pangkanon, 2009; Aoki, 2007; NIH Consens Statement, 2000; Desviat, 1999).

Incidence of PKU in Lithuania is 1:9300 (Kasnauskiene, 2003), 1: 6010 in Estonia (Õunap, 1998) and 1: 8000 live-births in Latvia (Purina, 1995).

1.4 PHENYLALANINE HYDROXYLASE (*PAH*) GENE

1.4.1. *PAH* gene location and structure

More than 98 percent of mutations associated with human HPA occur at the *PAH* locus; the remainder are at the loci dedicated to synthesis and regeneration of tetrahydrobiopterin. The *PAH* gene has regulatory elements and an architecture typical of many housekeeping genes. The locus harbours several hundred known alleles, some

of which are polymorphic and neutral in their effect on PAH enzyme activity; most are a cause of HPA (Scriver, 2008).

The gene for PAH is located on chromosome 12 in humans. It is 79, 277 bases long. Its cytogenetic address is 12q22-q24.2, on the long arm of the chromosome (Fig.1.7).

Fig. 1.7. Ideogram of chromosome 12 depicting approximate location of *PAH* on the long portion. *PAH* gene location is denoted by red marker (GeneCards, 2011).

PAH has a number of known orthologs in other organisms, one of the most important of which is found in mice. The mouse is an important model organism for the study of PAH and its role in metabolism, due to the great amount of conservation between the human and mouse *PAH* genes. Figure 1.8 shows the results of a BLAST2 alignment between human and mouse PAH (Anonymous, 2003).

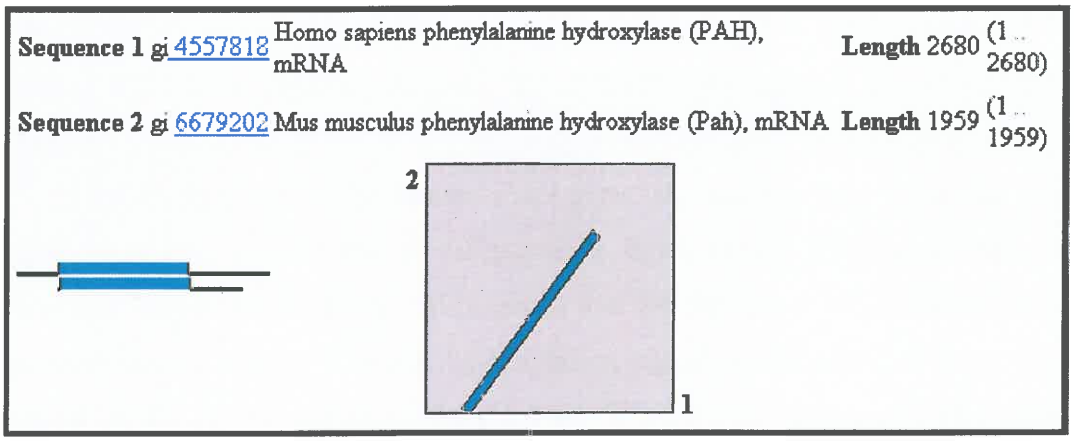


Fig. 1.8. Results of an alignment between human and mouse *PAH* sequences. The line is unbroken and travels fairly evenly from the bottom left corner to the top right corner, meaning that the sequences are highly conserved. A perfect alignment travels directly from bottom left to top right, dividing the grey square into exact halves (source: Anonymous, 2003: http://www.bio.davidson.edu/Courses/Molbio/MolStudents/spring 2003/Kogoy/favorite.html).

PAH encodes the enzyme phenylalanine hydroxylase. This protein's sequence contains 452 amino acids (a 2.4-kb cDNA clone) of 51,862 relative molecular mass (Kwok, 1985). It is primarily expressed in the liver. However, phenylalanine hydroxylase is also expressed in the kidney and spleen. The phenylalanine hydroxylase

enzyme can exist as several different isozymes, which are dimers, tetramers, or longer polymers of the same repeating subunit. Each chain folds to form three domains: an N-terminal regulatory domain, a catalytic domain, and a C-terminal oligomerisation domain (Miranda, 2002).

The cDNA sequence (DiLella, 1986; Konecki, 1992; Kwok, 1985) (GenBank U49897.1) contains 13 exons that constitute approximately 2.9% of the genomic *PAH* sequence. The intronic splice-site nucleotide sequences are all conventional. Exon border types vary; most are type 3 (beginning after the third nucleotide of a codon); codons 118 and 236 introduce type 1 borders spanning introns 3 and 6, respectively; codons 170, 281, and 400 introduce type 2 borders spanning introns 5, 7, and 11, respectively (Scriver, 2008).

A genomic sequence of the *PAH* gene and its flanking regions spanning 171,266 bp has approximately ~27 kbp of 5' untranslated region (5'UTR) upstream from the translation initiation site and approximately 64.5 kbp of 3' sequence downstream from the poly(A) site in the last exon (exon 13). The gDNA nucleotide numbers are in register with the cDNA sequence (which has long served *PAH* mutation nomenclature) because the *PAH* gDNA has been numbered in PAHdb so that the +1 nucleotide is the adenine of the transcription initiation site (ATG) in exon 1. Thus gDNA exons, introns, and the 3'UTR have positive numbers; the 5'UTR has negative numbers (Scriver, 2003).

Exonic sequences in the human *PAH* gene take up less than 3 percent of the genomic sequence between the 5' +1 positions down to the 3' poly(A) tract. The shortest and longest exons are 57 bp (exon 9) and 892 bp (exon 13), respectively; the mean exon size is 170 bp. Three polyadenylation signals [AATAAA] in exon 13 are annotated on the gDNA sequence; the third site is used most frequently. The shortest and longest introns are 556 bp (intron 10) and 17,874 bp (intron 2), respectively, whereas intron 3 is 17,187 bp in length, and the mean intron size is 6390 bp (Table 1.1). These are typical mammalian gene dimensions (Scriver, 2008).

The *PAH* genomic sequence consists of 40.7% GC, slightly above the modal value (37-38%) for human genes. The density of interspersed repeats is 42.2% in the *PAH* gene. Repetitive DNA is often the cause of large genomic deletions and duplications. Intron 2 has a 99% nucleotide identity with the *Alu* repeat element between bp 17,273 and bp 17,546 that might account for a 5' deletion causing PKU (Scriver, 2003). Putative *Alu* repeats in the *PAH* genomic sequence and CpG

dinucleotides (n51198) are potential sites for recurrent mutation in the *PAH* gene (Scriver, 2007).

The 5' untranslated region of the gene has five potential cap sites upstream from the actual methionine translation initiation codon in exon 1 (Konecki, 1992); multiple cap sites are a feature of many housekeeping genes within a 0.5-kb region upstream from the first codon. The *PAH* gene lacks a proximal TATA box (Scriver, 2008).

Table 1.1. cDNA and genomic DNA annotations

exon	cDNA	gDNA	intron	sequence
1	1-60	1-60	1	61-4232
2	61-168	4233-4340	2	4341-22214
3	169-352	22215-22398	3	22399-39585
4	353-441	39586-39674	4	39675-50549
5	442-509	50550-50617	5	50618-61889
6	510-706	61890-62086	6	62087-64271
7	707-842	64272-64407	7	64408-65465
8	843-912	65466-65535	8	65536-70272
9	913-969	70273-70329	9	70330-72792
10	970-1065	72793-72888	10	72889-73444
11	1066-1199	73445-73578	11	73579-76708
12	1200-1315	76709-76824	12	76825-78005
13[a]	1316-1359	78006-78897		

[a] Exon 13 cDNA covers only the translated and termination codons
Exon 13 gDNA extends 3' to include polyadenylation signals

1.4.2 *PAH* gene mutations

Over 560 putative pathogenic *PAH* alleles are documented in the database (www.pahdb.mcgill.ca). The vast majority of alleles are known (or presumed to be) causes of PKU or non-PKU HPA, having been ascertained through patients with a HPA phenotype. The disease-causing mutations fall into five classes: missense, 63%; small deletions, 13%; splice, 11%; putative silent, 7%; stop/nonsense, 5%; small insertions, 1%. Large deletions, once thought to be rare, probably account for approximately 3% of PKU-causing mutations (Fig. 1.9) (Kozak, 2006; Scriver, 2007).

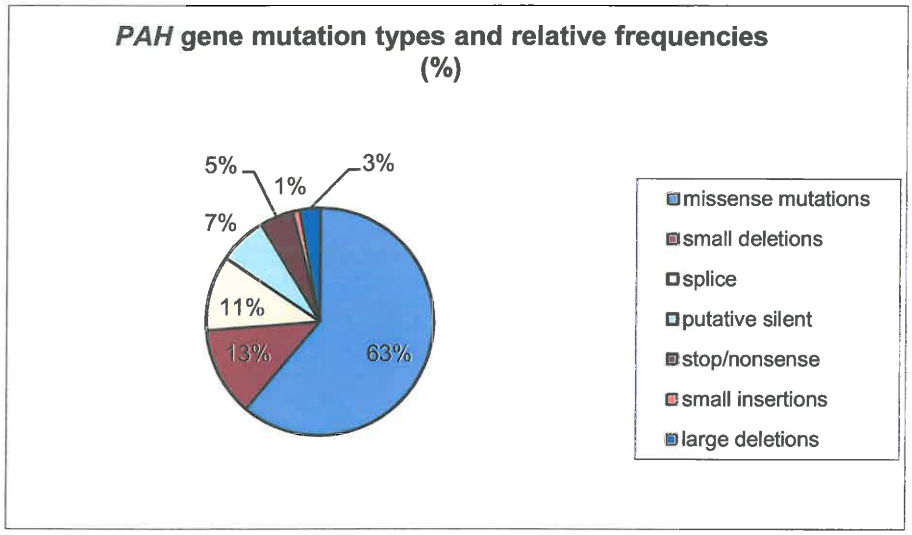


Fig. 1.9. *PAH* gene mutation types and relative frequencies (%).

Most missense mutations map onto the region between and including exon 5 (*PAH* residue 148) and exon 12 (*PAH* residue 438): 57 of these mutations are located in the regulatory domain sequence (residues 1–142); 231 mutations target the catalytic domain (residues 143-410); and 14 are located in the tetramerisation domain sequence (residues 411–452). Small deletions or insertions occur in the *PAH* gene. Large deletions affecting coding regions are rare. Some deletions involve a single exon. More frequently, multiple exons are affected. The relative frequency of large *PAH* gene deletions is apparently low (<1%) but this may reflect poor ascertainment. Most of the large recognized deletions were discovered when a disease-causing point mutation could not be identified by current methods of mutation analysis or when a haplotype configuration lacked one or more of its polymorphic alleles. Large insertions or gene duplications have not yet been reported (Scriver, 2003).

De novo alleles are rare occurrences in the PKU and related phenotypes. M1I and IVS3nt-6 have each been reported once as *de novo* alleles, once from Norway (Eiken, 1996) and E76G has been reported once from Taiwan (Chen, 2002). Another occurrence (unnamed) is reported once from Southern Germany (Aulehla-Scholz and Heilbronner, 2003). *De novo* alleles could not be recognised if the parental alleles have not also been analysed, a practice that is not uniform. Accordingly, one cannot say that these few *de novo* alleles have revealed mutation hotspots and that the *PAH* locus is unusually mutable (Scriver, 2003).

1.4.3 Mutation effect on enzyme function

On the enzyme level there are ''null''-mutations (splicing, termination, and frameshift mutations as well as many missense mutations) that cause severe structural alterations or destroy the catalytic domain and result in the absence of a functioning PAH protein. In contrast, other (mostly missense) mutations interfere primarily with protein folding, regulation, or parameters of enzyme activity and leave variable residual function (Zschocke, 2003).

Mutations can be grouped into five structural categories, based on the distinct expected structural and functional effects of the mutations in each category. Missense mutations and small amino acid deletions are found in three categories: (1) active site mutations, (2) dimer interface mutations, and (3) domain structure mutations. Nonsense mutations and splicing mutations form the category of (4) proteins with truncations and large deletions. The final category, (5) fusion proteins, is caused by frameshift mutations (Jennings, 2000).

Active site mutations are mutations affecting any residue in the active site of enzyme. Most of these residues are strictly conserved in PAH sequences. The iron, bound to the enzyme through the side chains of H285, H290, and E330, is essential for enzymatic activity. Some residues are involved in BH4 binding and mutations in these residues affect the pterin binding to PAH (for example, L249F, L249H, L255V, and L255S). Other mutations in the active site may affect phenylalanine substrate binding, the overall structure of the active site, or have more indirect effects by affecting protein folding and stability.

Domain structure mutations affect either residues forming the hydrophobic cores of the protein domains, or residues that may in other ways be responsible for the structural integrity of the individual domains, the minority of domain structure mutations affect surface residues. These mutations may affect not only the thermodynamic stability of a protein, but also processes that lead to the adoption of the folded structure, such as protein folding and possible non-specific association during folding, and consequently associations with cellular factors involved in this process, such as chaperones, and proteases.

Dimer interface mutations affect residues in the dimer interface. The dimer interface probably participates in the conformational changes occurring upon activation by phenylalanine, and the interface mutations may interfere with the activation process.

Most residues in the dimer interface affected by mutations are strictly conserved in PAH sequences.

Proteins with truncations and large deletions can be caused by nonsense and splicing mutations. While nonsense mutations result in C-terminally truncated proteins, splicing mutations can produce terminal or internal deletions, for example by exon skipping.

Insertion or deletion mutations introducing frameshifts will result in mutant proteins containing a truncated PAH sequence fused to an unrelated sequence at the C-terminus; such proteins are herein referred to as fusion proteins. Frameshift mutations at residues N-terminal to residue Q383 will result in inactive proteins. Indeed it is found that such mutant proteins are all associated with classic PKU. Mutations C-terminal to residue Q383 may indirectly affect the active site, interfere with folding, cause aggregation, or form unstable proteins (Jennings, 2000).

1.4.4 *PAH* gene polymorphic alleles

The *PAH* cDNA sequence contains recognised polymorphic alleles (Fig. 1.10). Three forms of *PAH* polymorphism exist:

1. *Biallelic restriction fragment length polymorphisms* (RFLPs) (DiLella, 1986; Lidsky, 1985) named from the corresponding restriction enzyme (*Bgl*II, *Pvu*IIa, *Pvu*IIb, *Eco*RI, *Alu*I, *Msp*I, *Xmn*I, and *Eco*RV). With the exception of the *Eco*R sites, which still require analysis by Southern blotting, the RFLPs can be analysed by methods based on polymerase chain reaction (PCR) amplification *Bgl*II (Dworniczak, 1991a), *Pvu*IIa (Dworniczak, 1991b), *Pvu*IIb, *Msp*I (Wedemeyer, 1991), and *Xmn*I (Goltsov, 1992a).

2. *Multiallelic polymorphisms*, which include a hypervariable sequence [variable number of tandem repeats (VNTRs)] of 30-bp cassettes harbouring at least 10 alleles (differing by number of repeats) in a *Hind*III fragment 3 kb downstream from the last exon in *PAH* (Goltsov, 1992b; Latorra, 1994), and a series of short tandem [tetranucleotide (TCTA)*n*] repeats (STRs) harbouring at least 9 alleles in the third intron of *PAH* (Giannattasio, 1997; Goltsov, 1993; Zschocke, 1994a).

3. *Single-nucleotide polymorphisms* (SNPs), which are silent (non-RFLP) alleles; for example, a silent c.696A/G polymorphism (q = 0.08–0.63) in codon 232 (Q232Q) (Lichter-Konecki, 1994).

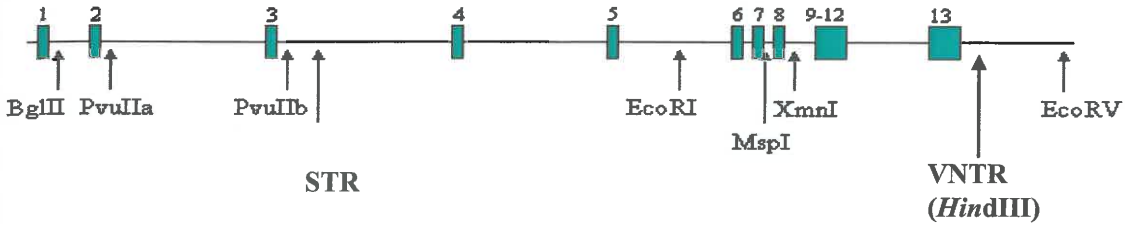


Fig. 1.10. Schematic map of the *PAH* locus indicating exons and sites of the RFLP, STR and VNTR polymorphisms (modified from Kidd and Kidd, 2005).

RFLP, STR, and VNTR alleles can be combined to generate core haplotypes at the extended *PAH* locus. An informative minihaplotype consisting of only the STR and VNTR alleles and accessible to PCR-based analysis has been developed (Eisensmith et al., 1994), as have other approaches (Zschocke et al., 1995). The extended (full) *PAH* haplotypes are named with Arabic numbers (Eisensmith and Woo, 1992), and at least 87 are known (PAHdb: www.pahdb.mcgill.ca).

PAH haplotypes could be generated from combinations of RFLP, STR, and VNTR alleles (Eisensmith and Woo, 1995), but only a few haplotypes are prevalent, and most are uncommon. Despite the large number of haplotypes observed, five of them - haplotypes 1, 2, 4, 5, and 7-account for more than 76% of normal chromosomes throughout Europe, and only four haplotypes - 1 through 4 - account for more than 80% of PKU-bearing chromosomes (Daiger, 1989a).

The apparent shortage of *PAH* haplotypes is explained by linkage disequilibrium across the 100-kb region of the extended haplotype (Chakraborty, 1987; Degioanni and Darlu, 1994; Kidd, 2000). *PAH* haplotype heterogeneity is greater on mutant and normal chromosomes in Europeans (Daiger et al., 1989a) than it is on chromosomes in Asians (Daiger et al., 1989b). *PAH* haplotype diversity is greater in African populations than it is in Europeans, assuming that the latter are descendant of a small founding group emerging out of Africa some 100,000 years ago (Kidd et al., 2000).

1.4.5 Restriction fragment length polymorphism (RFLP) in the *PAH* gene

Southern analysis of many individual genomic DNA samples digested by a battery of restriction enzymes demonstrated the presence of at least eight RFLPs associated with the *PAH* gene. The high degree of heterozygosity of these RFLPs among Caucasian PKU families greatly facilitated the implementation, on the basis of haplotype analysis, of prenatal diagnosis and genetic counselling in PKU families. Seven dimorphic and one trimorphic (*Hin*dIII) RFLP detected by Southern analysis have been used to define distinct haplotypes at the *PAH* locus: *Bgl*II (Dworniczak, 1991a), *Pvu*II(a) (Dworniczak, 1991b) and *Pvu*II(b), *Msp*I (Wedemeyer, 1991), *Xmn*I (Goltsov, 1992a), and *Hin*dIII (Goltsov, 1992b) (Eisensmith and Woo, 1992). Of the 384 (2^7 x 3) possible haplotypes defined by these sites, only 87 are known (PAHdb: www.pahdb.mcgill.ca).

Detailed examination of the amplification fragment containing the polymorphism detected by *Hin*dIII digestion has identified a region containing a VNTR. Comparison of this region in Caucasian PKU families demonstrates the presence of at least six alleles that differ in the number of repeated units (Goltsov et al. 1992a). This latter finding permits an expansion of the existing haplotype system, from 384 possible combinations to at least 768 (2^7 x 6) (Eisensmith and Woo, 1992).

1.4.6 Variable number of tandem repeats in *PAH* gene

The *PAH* gene is linked to a variable number tandem-repeat (VNTR) polymorphism, which is a 30 bp AT-rich (70%) tandem-repeat (Fig. 1.11) system located 3 kb down from the final exon of the gene. This VNTR system is responsible for the previously reported three alleles for the *Hin*dIII polymorphism of the human *PAH* gene which are 4.0, 4.2, and 4.4 kb long (Woo, 1983) and contain 3, 6-9 and 12 copies of VNTR respectively (Goltsov, 1992b). VNTR with 13 repeats was observed in two PKU families from south of Iran and was associated with normal alleles in both families (Kamkar, 2003).

Kindred analysis in phenylketonuria families demonstrates Mendelian segregation of these VNTR alleles, as well as associations between these alleles and certain *PAH* mutations (Goltsov, 1992b).

This polymorphism is diversified enough in European Caucasians to be useful in carrier detection of PKU, with polymorphism information content (PIC) of 70%. In Chinese PKU families, the PIC of this polymorphism is much lower, at 32%. This polymorphism is also important in haplotyping the *PAH* gene and tracing the geographical origin of mutations (Hosseini-Mazinani, 2008; Cali, 1997; Tighe, 2003).

Since the VNTR is a highly polymorphic genetic marker and is inherited in a Mendelian fashion, it is used to give a risk estimation of linked defective allele. In addition, this VNTR may prove useful in studies concerning the origins and distributions of *PAH* mutations in different human populations (Kamkar, 2003).

CACATATATGTAT[A/G]TGCATA[C/T]GTA[C/T]GTA[GG/TG/TA]

Fig. 1.11. Sequence for the 30-bp repeated unit, derived from sequencing of 10 different VNTR alleles. Variable nucleotides are indicated in brackets (Goltsov, 1992b).

Within chromosomes bearing the polymorphic 4.2-kb allele, the VNTR allele containing 8 repeats is most prevalent (about 60%) among both normal and mutant chromosomes. The VNTR allele containing 7 repeats is the second most frequent allele associated with chromosomes bearing the 4.2-kb allele. This VNTR allele is present on about 1/5 of all normal chromosomes and on about 1/3 of all mutant chromosomes of this type. The VNTR allele containing 9 repeats accounts for most of the remaining normal or mutant chromosomes in this class. Only one mutant allele bearing six VNTR units was observed in 295 members of European PKU families (Goltsov, 1992b).

1.4.7 Short tandem repeats (STR) in the *PAH* gene

The STRs are highly polymorphic and inherited stably. Tetrameric STR $(TCTA)_n$ has been described in the human phenylalanine hydroxylase gene harbouring at least 9 alleles (226bp to 258bp) in the third intron of *PAH* (Fig. 1.12). STR within the *PAH* gene has an average level of heterozygosity of about 75% in Orientals and about 80% in European Caucasian populations. This single marker is as informative as haplotype analysis in Europeans and nearly twice as informative as haplotype analysis in Orientals (Goltsov, 1993).

GCCAGAACAATACTGGTTCTGTGGAAAGCAGAAAGACCTATCTGCCTG 49

TCTATCTATCTATCTATCTATCTGTCTATCTATCTATCCGTCTATCTATCC 100

ATCATCTATCATCTACACCTGTCTATCTATCATCTATCTACCTATCTATC 151

TGTCCATCAATCATCTATCATCTATCTATCTATCTATCTATCTATCTATCT 202

ATCTATCTATGTTGTCTGGGAACACTTATGATT 234

Fig. 1.12. The DNA sequence of the region of the *PAH* gene containing the STR system. Tandem repeats (TCTA) are underlined; regions where polymorphism occurs are double underlined. The positions of the PCR primers used to amplify this region are highlighted (Goltsov, 1993).

The most STR alleles are present in both normal and mutant *PAH* chromosomes with a continuous distribution from the smallest (226bp) to the largest (258bp) allele. A higher frequency of the 238bp allele among mutant chromosomes and the higher frequency of the 242bp and 246bp alleles among normal chromosomes were established in Caucasians from different European populations. The two most common alleles, the 238bp and 242bp, together account for about 57% of normal and 68% of mutant chromosomes. The high degree of polymorphism and strong Mendelian segregation of this system makes it useful for prenatal diagnosis and carrier screening determination in PKU families.

STR system has a major advantage over traditional RFLP haplotype analysis. In populations characterised by the presence of a single predominant RFLP haplotype, this system considerably increases the number of cases that are informative (Goltsov, 1993).

1.5 R408W MUTATION IN *PAH* GENE

Among the most common *PAH* mutations is R408W. Mutation R408W is found at relative allele frequencies up to 84% in Europe (Lillevali, 1996; Zschocke, 2003).

The R408W mutation (c.1222C>T; CGG>TGG) (DiLella, 1987), a C to T transition in exon 12 of the *PAH* gene, results in the substitution of tryptophan for arginine ($Arg^{408} \rightarrow Trp^{408}$) at amino-acid residue 408 and is a null mutation associated with <0.3% of normal activity and a severe PKU phenotype (Kayaalp, 1997).

This mutation involves a CpG dinucleotide in a so-called "hypermutable" codon. CpG dinucleotides are under-represented in human genomic DNA, most are located in

CpG islands of promoters, and only a minority (~3%) of the cytosines are methylated. In promoter regions, methylation plays a critical role in gene regulation and expression. Methylation of cytosines occurs also in CpG dinucleotides in exons and is a source of mutability there. Methylated cytosine in its particular dinucleotide context (mCpG) can undergo spontaneous deamination initiating a C>T substitution and it is generally assumed that a CpG to TpG transition reflects the presence of methylated cytosine. This could be the one of explanations for the high relative frequency of the c.1222C>T (p.R408W) allele in the human population (Murphy, 2006).

The R408W mutation is involved in interdomain interactions in a PAH monomer. The most important residues at the interface between the catalytic and tetramerization domains are Leu311, Leu308, and Arg408. Arg408, located in the loop between Cα12 and Tβ1, forms hydrogen bonds to the carbonyl oxygens of Leu308 (at the end of Cα8) and Leu311 (in the loop between Cα8 and Cα9) (Fig. 1.13). Substitution of Arg408 into a larger and bulkier tryptophan would alter the hydrogen bonding network at the interface of the tetramerization and catalytic domains, interfering with the correct positioning of the β-ribbon (Tβ1 and Tβ2) (Erlandsen, 2003).

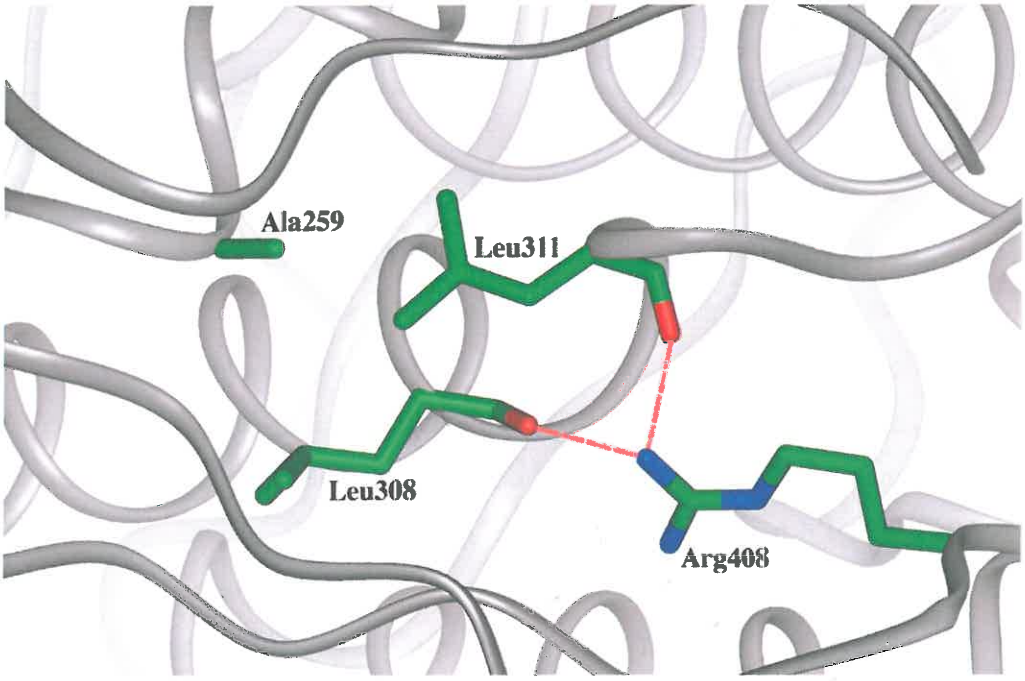


Fig. 1.13. Close of the R408 mutation, including the hydrogen bonding (Source: Erlandsen and Stevens, 1999).

In Europe, the R408W mutation is observed on chromosomes of two major haplotype backgrounds. R408W-2.3 (RFLP haplotype 2/VNTR 3) exhibits a west-to-east cline of relative frequency reaching its maximum in the Balto-Slavic region. R408W-1.8 exhibits an east-to-west cline in north-western Europe peaking in Connacht, the most westerly province of Ireland (Tighe, 2003).

In eastern European populations, the R408W mutation is strongly associated with RFLP haplotype 2, the three-copy VNTR allele (VNTR 3), and the 238-bp STR allele. In north-western European populations, it is strongly associated with RFLP haplotype 1, the VNTR allele containing eight repeats (VNTR 8), and the 242-bp STR allele (Table 1.2). An examination of the linkage between the R408W mutation and highly polymorphic RFLP, VNTR, and STR haplotypes suggests that recurrence is the most likely mechanism to account for the two different major haplotype associations of R408W in Europe (Tighe, 2003).

Table 1.2. Mutation R408W haplotypes at the human *PAH* locus. Plus (+) and minus (-) indicate the presence or absence of a polymorphic restriction site, respectively (Eisensmith and Woo, 1992).

Haplotype	BglII	PvuII(a)	PvuII(b)	EcoRI	MspI	XmnI	VNTR (HindIII)	EcoRV
1.8	-	+	-	-	+	-	8	-
2.3	-	+	-	-	+	-	3	+

The geographic distributions of R408W-1.8 and R408W- 2.3 observed in Europe are the result of human dispersals (Tighe, 2003).

The high relative frequency of the R408W/RFLP haplotype 2/VNTR 3 allele in all three major branches of Slavic peoples (east, west, and south) suggests that it was already present before their expansion and subsequent westward and southern migrations in the 5th and 6th centuries. The lower frequency of this mutant allele in ethnic Tatars and other central Asian natives (<40%) suggests that it was not introduced into Slavs from central Asians, but rather was introduced into central Asians by the more recent eastward expansion of Slavs (Eisensmith, 1995).

The ultimate origin of this allele in Ireland and its present distribution is clearly consistent with the numerous historical and archaeological records of Irish settlements in Scotland, the Isle of Man, Wales, southwest England, and Brittany from the 3rd to the 5th century. The relatively high frequency of the R408W/RFLP haplotype 1/VNTR 8

allele in Norway is probably due to the assimilation of Irish into Norwegian culture, either through intermarriage or slavery, that occurred during the extended presence of Norwegian Vikings in Ireland and northwest Scotland from the late 8th to the middle of the 12th century (Eisensmith, 1995).

1.5.1 Genetic diversity within the R408W mutation lineages in Europe

Variation in the VNTR motif located at the 3' end of the *PAH* locus demonstrates greater genetic diversity within the VNTR-8 lineage as compared with VNTR-3 and may suggest that VNTR-8 is older than VNTR-3. Wild-type VNTR-8 chromosomes were found to exhibit two major cassette organisations: $(a1)_5$-b3-b2-c1 and $(a1)_5$-b5-b2-c1 at relative frequencies of 58% and 40%, respectively, while wild-type VNTR-3 chromosomes are associated with a single invariant cassette organisation, a2-b2-c1 (Table 1.3).

R408W-1.8 chromosomes were predominantly associated with the (a1)5-b5-b2-c1 cassette organisation (98%). R408W-2.3 chromosomes were associated exclusively with the a2-b2-c1 cassette organisation.

Table 1.3. *PAH* VNTR cassette sequence variants (Tighe, 2003).

No.	Cassette	Sequence
1.	a1	CACATATATGTATATGCATATGTACGTATG
2.	a2	CACATATATGTATATGCATACGTACGTATG
3.	a3	CACATATATGTATATGCATATGTACATATG
4.	a4	CACATATATGTATATGCATATGTACATAGG
5.	b1	CACATATATGTATGTGCATATGTACGTATA
6.	b2	CACATATATGTATGTGCATATGTACATAGG
7.	b3	CACATATATGTATGTGCATATGTACGTAGG
8.	b4	CACATATATGTATGTGCATATGTAAGTAGG
9.	b5	CACATATATGTATGTGCATATGTACGTATG
10.	c1	CACATATATGTATGTGCATATGTATGTATA
11.	c2	CACATATATGTATATGCATATGTATGTATA

Analysis of associations between VNTR cassette sequences and STR alleles confirms the evidence of genetic diversity: wild-type VNTR-8 $(a1)_5$-b3-b2-c1 chromosomes exhibit an unimodal distribution (242 bp) whereas (a1)5-b5-b2-c1 chromosomes exhibit an apparently bimodal distribution (234 bp and 242 bp). Wild-type VNTR-3 chromosomes have an unimodal distribution (238 bp). R408W-1.8-[(a1)5-b5-b2-c1] and R408W-2.3-[a2-b2-c1] chromosomes both have unimodal STR allele distributions centered on the 242 bp and 238 bp alleles, respectively. Thereby the VNTR-3 lineage exhibit less diversity than VNTR-8 allele, this relative lack of diversity suggests that the VNTR-3 lineage may have arisen more recently than VNTR-8 in the European population. Conversely R408W-1.8 mutation have accumulated some nucleotide sequence variation within the VNTR and, thus, may potentially be older than R408W-2.3 (Tighe, 2003).

1.6 CORRELATION BETWEEN *PAH* GENOTYPE AND METABOLIC PHENOTYPE

The *PAH* mutation genotype is one of the main determinant of metabolic phenotype in most patients with PAH deficiency. The highly variable metabolic phenotypes of PAH deficiency correlate with *PAH* genotypes.

Some mutations cause complete abolition of PAH function, whereas others are associated with residual in vitro activity in the range of 2%–70%. This continuous spectrum of mutation-related enzyme activities, together with the plethora of possible mutation combinations (genotypes), may explain the observed inter-individual heterogeneity of metabolic phenotypes.

On the basis of the collaborative study (seven centres in France, Italy, Belgium, Germany, and Denmark participated) results, 105 of the *PAH* gene mutations were assigned to one of the four phenotype categories by Guldberg e*t al.* in 1998: "classic or severe PKU" "moderate PKU," "mild PKU," and "mild hyperphenylalaninemia" (MHP) (Appendix I) (Guldberg, 1998).

PAH deficiency has a correlation between genotype and phenotype, implying (1) that the milder of two mutations determines the phenotypic outcome and (2) that gene dosage has a significant effect— that is, a patient homozygous for a mutation with residual enzyme activity may exhibit a milder phenotype than is seen in a patient carrying the mutation in the functionally hemizygous constellation. To convert these

observations into a more formalised system, Guldberg e*t al.* in 1998 assigned each mutation an arbitrary assigned value (AV): AV=1, for classic-PKU mutations; AV=2, for moderate- PKU mutations; AV= 4, for mild-PKU mutations; and AV=8, for MHP mutations. These values were chosen as the lowest positive whole numbers that allow discrimination between the different mutation combinations. By means of this classification, the phenotype resulting from the combination of two mutant *PAH* alleles may be expressed numerically as the sum of the two mutations' AVs (Table 1.4) (Guldberg, 1998).

Table 1.4. Model for phenotypic effect of two mutant *PAH* alleles, expressed as the sum of their AVs (Guldberg, 1998).

	PHENOTYPIC EFFECT WHEN AV OF FIRST MUTATION IS			
AV OF SECOND MUTATION	1 (Classic PKU)	2 (Moderate PKU)	4 (Mild PKU)	8 (MHP)
1 (Classic PKU)	2 (Classic PKU)	3 (Moderate PKU)	5 (Mild PKU)	9 (MHP)
2 (Moderate PKU)		4 (Moderate PKU/mild PKU)	6 (Mild PKU)	10 (MHP)
4 (Mild PKU)			8 (Mild PKU/MHP)	12 (MHP)
8 (MHP)				16 (MHP)

1.6.1 Mutations associated with tetrahydrobiopterin BH_4 – responsiveness

Tetrahydrobiopterin BH4-responsive PAH deficiency is a subgroup of hyper-phenylalaninemia caused by specific mutations in the *PAH* gene.

The spectrum of HPAs caused by PAH deficiency ranges from the mild HPA (MHP), to mild phenylketonuria (mild PKU), and intermediate or classical PKU. Patients with BH4-responsive PAH deficiency belong mostly to the groups of MHP and mild PKU (Blau and Erlandsen, 2004).

Some of potentially BH4-responsive mutations were expressed recombinantly in eukaryotic cell systems or *Escherichia coli* and found to have substantial residual activity (Appendix II). About 64% of all mutations are located in the catalytic domain of PAH, 14% in the regulatory domain, 10% are located in the tetramerisation domain, and 12% are intronic (Appendix III) (Blau and Erlandsen, 2004; Zurfluh, 2008).

Only very few of the described mutations are located within the two cofactor-binding regions CBR1 (V245A, R252W, R261X, and R261Q) and CBR2 (P281A, P281S, and P281L) and only two of them (V245A and R261Q) seem to be associated with BH4-responsiveness. BH4-responsive mutations with their residual activity are listed in Appendix IV.

The allelic *PAH* mutation-combination is the most important indicator of BH4-responsiveness. The BH4-responsiveness in PAH deficiency is characterised by a substantial residual PAH activity of at least one mutant allele. The estimated frequency of BH4-responsiveness is much higher (>75%) in southern regions of Europe with a high frequency of BH4-responsive alleles p.R261Q, p.V388 M, p.I65 T, p.R158Q, or p.L48S, than in central Europe (50–70%), or in some eastern European countries (<40%) with frequent severe alleles p.R408W, p.R252W, or IVS12-1G>A (Zurfluh, 2008).

1.7 DISTRIBUTION OF PHENYLKETONURIA MUTATIONS IN EUROPE

Phenylketonuria is heterogeneous. More than 560 different mutations in the phenylalanine hydroxylase gene have been identified (http:///www.pahdb.mcgill.ca). Twenty-nine (29) different mutations may be regarded as prevalent in European populations (reached relative frequencies of at least 3% in at least two countries). The spectrum of mutations found in individual regions results from a combination of factors including founder effect, range expansion and migration, genetic drift, and probably heterozygote advantage. Mutations vary in their impact on enzyme activity, causing a range of clinical phenotypes from severe PKU to MHP that does not require treatment. PKU is detected by neonatal screening and there is almost complete ascertainment of the disease in most European countries (Zschocke, 2003).

1.7.1 PKU mutations in Northern Europe

Denmark

Denmark was the first country with complete *PAH* mutation data. The mutational spectrum consists of 35 different mutations, including 23 missense mutations, 5 splice mutations, 4 nonsense mutations, and 3 deletions. Seventeen of these

mutations have been reported for the first time. There is one predominant mutation, IVS12+1G>A, which together with mutations R408W-H2 and Y414C accounts for two-thirds of mutant chromosomes (Guldberg, 1993a).

Finland

The incidence of phenylketonuria in Finland is extremely low, probably below 1 in 100 000. Mutations and RFLP haplotypes were published for all four known patients. Mutation R408W was found on four mutant chromosomes (all haplotype 2), and IVS7ntl (haplotype 4), R261Q (haplotype 1), and IVS2ntl (haplotype 11) were found on a single chromosome each. No mutation was found on the remaining chromosome. These findings support a pronounced negative founder effect as the cause of the low incidence of PKU in Finland, and are consistent with existing data regarding the European and Baltic origin of Finnish genes (Guldberg, 1995).

Sweden

More than 50 *PAH* gene mutations have been reported in Swedish PKU patients. Three mutations: R408W, Y414C and IVS12+1G>A, together account for 56% of all PKU chromosomes and 10 relatively infrequent mutations were found on other 17% of mutant chromosomes. The R408W mutation was the most common (22%) with the prevalent RFLP haplotype 2, closely followed by the Y414C (18%) and IVS12+1G>A (16%) mutations (Svensson, 1993). No more recent and complete data are available from Sweden.

Norway

Thirty-three (33) different mutations constituted 99.6% of all mutant alleles (only 1 allele remains unknown) were identified in Norwegian PKU alleles, 23 of these have been identified also in other European countries. Twenty were predicted missense mutations, 6 splice mutations, 4 nonsense mutations and 2 deletion mutations and 1 mutation disrupted the start codon. The 8 most common mutations represented 83.5% of the PKU alleles, with single allele frequencies ranging from 5.9 to 15.7%. Four of these mutations (R261Q, R408W, Y414C, and IVS12+1G>A) are commonly occurring also in PKU patients in other European countries, while the other 4 (G46S, G272X, F299C, and R408Q) have higher frequencies in Norway than in any other country studied. Mutations F299C and R408Q are common along the western and northern coast, G46S

and particularly G272X in the highly populated southeast, while IVS12+1G>A, Y414C, and R408W (with equal proportion of haplotypes 1 and 2) are found throughout the country. Six mutations (I65T, L249F, P281L, Y356X, R158Q, and R252W) have frequencies between 0.8% and 2.1%, and 19 mutations were encountered only once. The majority of PKU mutations were found on the same RFLP/VNTR haplotype backgrounds in Norway as in other European populations, suggesting that only a few of the mutations may represent recurrent mutations (< 3.4%). Among 10 mutations only reported for Norwegian population, 2 *de novo* mutations (0.8%) were detected arisen in Norway. Results are compatible with multiple founder effects and genetic drift for the distribution of PKU mutations within Norway (Eiken, 1996).

Iceland

Iceland shows evidence of a strong local founder effect, with mutation Y377>Tfs (c.1129delT) accounting for 42% of PKU chromosomes. Genealogical examination, extending back more than 5 generations, shows that this mutation has probably arisen in an isolated part of southern Iceland and was enriched by a founder effect. Other common mutations include P281L and F299C, but the total number of Icelandic PKU patients is relatively small (17 patients). Totally 9 different mutations were found in the phenylalanine hydroxylase gene (Guldberg, 1997).

United Kingdom of Great Britain and Northern Ireland

Detailed PKU mutation studies from the British Isles identified marked differences between regions (Zschocke, 1995; Tyfield, 1997; Zschocke, 1997; O'Donnell, 2002).

Examination of phenylalanine hydroxylase gene in the phenylketonuria populations of four geographical areas of the British Isles: the west of Scotland, southern Wales, and south-western and south-eastern England revealed 63 different mutations. In all four regions, two mutations in exon 12 (R408W and IVS12+1G>A) and one mutation in exon 3 (165T) predominated, and together they accounted for 40%-50% of PKU chromosomes. Three other mutations in three exons - F39L, F299C, and L348V - accounted for an additional 11%-19%.

Mutation R408W was the most common in western Scotland, accounting for 31% of mutant chromosomes. The relative frequencies were about half this in Wales and London and one third of this in south-western England. A reverse gradient is seen

for IVS12+1G>A. In south-western England this mutation was the most common accounting for 27% of PKU chromosomes. Two mutations, L48S and IVS10-11G>A, were detected almost exclusively in south-western England and together accounted for 9% of mutant chromosomes (Tyfield, 1997).

The investigation of the mutation spectrum of hyperphenylalaninaemia in the Irish Republic identified 29 distinct mutations in 11 of the 13 *PAH* exons. The three most common mutations were R408W (41.0%), F39L (12.2%) and I65T (10.4%), together accounting for 63.6% of mutant alleles. Mutation F39L showed maximal frequency in Leinster (the east coast), whereas I65T peaked in Ulster (the north/north-west) but R408W occurred with the highest relative frequency (55.8%) in Connacht, the most westerly province. Scandinavian *PAH* mutations (F299C, R408Q, Y414C and G46S) accounted for 6.1% of the mutant alleles were detected (O'Donnell, 2002).

1.7.2 PKU mutations in Western Europe

Germany

The most comprehensive mutation analysis in PKU patients have been carried out in Germany and resulted in four published studies from various regions of the country (Guldberg, 1996a; Zschocke and Hoffmann, 1999; Hennermann, 2000; Aulehla-Scholz and Heilbronner, 2003), all of which show a marked heterogeneity with regard to PKU mutations.

Up to 91 different mutations in *PAH* gene are identified. The most common mutation is R408W (22%) with RFLP haplotype 2, which is more frequent in the East (up to 38%) than in the West; other prevalent mutations include IVS12+1G>A (12%) and Y414C (8%). IVS10–11G>A (10%) is also relatively common, only partly due to a high prevalence of this mutation in recent immigrants from Turkey (Zschocke and Hoffmann, 1999; Zschocke, 2003).

The Netherlands and Belgium

Twenty-one (21) different mutations were found in all thirteen exons of the *PAH* gene in phenylketonuria and hyperphenylalaninemia patients in the Netherlands including 4 novel gene aberrations (Meijer, 1993; van der Sijs-Bos, 1996). Three were the most common: IVS12+1G>A (24%), R261Q (18%) and R158Q (13%) (Zschocke, 2003).

Mutation data from Belgium have only partly been published (Verelst, 1993). Common mutations in The Netherlands and Belgium include IVS12+1G>A, R158Q, R261Q, and P281L, while R408W is remarkably rare (up to 5%) (Zschocke, 2003).

Switzerland, France and Austria

Limited data for Switzerland indicate a high frequency of mutation R261Q (32%) (Eisensmith, 1992).

The full spectrum of *PAH* gene mutations is not available for PKU patients in France. Published data for the whole of France are limited to an old study where 15 mutations were screened in a variable, sometimes not exactly specified number of patients, but mutation data for northern France show allele frequencies similar to those found in Belgium (Zschocke, 2003).

No data about PKU mutations are available for Austria.

1.7.3 PKU mutations in Eastern Europe

The unique feature of eastern European populations with regard to PKU mutations is the predominance of a single mutation, R408W on RFLP haplotype 2. Next relatively common PKU mutation throughout Eastern Europe is R158Q.

Poland

Three the most prevalent mutations R408W (55-68%), IVS10–11G>A (5-10%) and IVS12+1G>A (5.2%) were identified in Polish PKU patients. Mutation R408W frequencies are different in different studies from Poland possibly because of different geographical regions studied and also there is an overlap of authors that may also indicate an overlap of patients (Zschocke, 2003).

Estonia

Molecular analysis of Estonian patients has revealed high genotypic homogeneity in this group, as 84% of the mutant alleles carry the R408W mutation. Five more mutations – IVS12+1G>A, R261Q, R252W, R158Q, S349P – have been detected in low frequencies (Lilleväli, 1996).

Lithuania

Twenty-one (21) different mutations were found in Lithuanian PKU patients, including 10 missense mutations, three splice mutations, four nonsense mutations and four micro-deletions. No mutation was found in exons 4, 8, 11 and 13. The three most common PKU and MHP mutations R408W (73.4 %), R158Q (7.1 %) and A403V (2.0 %) accounted for 83 % of mutant PKU and MHP alleles, with the other mutations being rare (0.5-1.5 %). These results demonstrate that the PKU in Lithuania is relatively homogeneous (Kasnauskiene, 2003).

Czech Republic

A total of 30 different mutations were detected on PKU chromosomes in Czech Republic. 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, R270fsdelSbp, A342P, and IVS11nt-8G>A were identified. These results confirm that PKU is a heterogeneous disorder at the molecular level also in Czech population (Kozak, 1997).

Slovakia

The predominant mutation in unrelated Slovak PKU families of Caucasian origin is R408W, with a frequency of 45.9%. In addition, four other mutations have been identified at relatively high frequencies: IVS12+1G>A, 10.2%; R158Q, 7.1%; R261Q, 7.1%; R252W, 2.0%. Mutations detection was focused to analysis of seven more common mutations in the *PAH* gene (Kadasi, 1995).

Russian Federation

There are incomplete data sets for the European part of Russia and other regions of the Russian Federation.

Thirty-one (31) unrelated phenylketonuria patients from the Moscow region were screened for mutations in the phenylalanine hydroxylase gene at the following codons: 408, 158, 261 and IVS-12. The following mutation frequencies were determined: codon 408 – 56.4%; codon 158 – 8.1%; codon 261 – 3.2%, and IVS-12 – 16% (Charikova, 1993).

Eleven different *PAH* gene mutations were detected in PKU patients in St. Petersburg. Mutations R408W was prevalent with frequency 70,7%, other mutations appeared with frequencies below 5%: R261Q – 4,3%, P281Q – 4,3%, R252W – 2,9%, IVS12+1G>A – 2,1%, R158Q – 1,4%, R261X – 0,7%, R243Q – 0,7%, E280K – 0,7%, IVS10–11G>A – 0,7% and K363fs – 1,4% (Baranovskaya, 1996).

The most prevalent mutation among PKU patients in the Far East of Russia was R408W (63%), with a haplotype background of 2.3. It also showed a very high degree of homozygosity (43%). The other five mutations (R158Q, R261Q, R252W, R261X, and IVS12+1G>A) accounted for 1.7%–6.7% of all PKU alleles. The genetic structure of PKU patients in the Far East of Russia seems to be also relatively homogeneous (Sueoka, 1999).

Ukraine

Recent data from the Ukraine have been published in abstract format. The most frequent mutation found according to their study was R408W (57%). The frequencies of other mutations were: R158Q – 3.5%, R252W – 2.9%, P281L – 2.3%, Y414C – 1.5%, IVS10–11G>A, IVS12+1G>A, R261Q, G272X, S273F, R413P – 0.6-1% (Nechyporenko and Livshits, 2002).

Hungary, Bulgaria and Romania

Data for these countries are restricted, but all three populations show the prevalence of R408W mutation: Hungary – 49%, Bulgaria – 35% and Romania 48%. Next frequent mutation for each country was R158Q (7%), IVS10–11G>A (25%) and c.1089delG (14%) respectively (Zschocke, 2003).

1.7.4 PKU mutations in Southern Europe

Also for the most of south-eastern European countries limited data about *PAH* gene mutations spectrum are available.

Croatia

PKU in Croatia is moderately homogeneous. A total of 21 disease-causing mutations were identified including 13 missense, 2 nonsense and 5 splice site mutations as well as one single nucleotide deletion. Seven of the 21 mutations are C>T or G>A

alterations at known hypermutable CpG sites. Four different mutations (R408W, P281L, R261Q and E390G) account for two thirds (66%) of mutant alleles in the investigated patients. The commonest mutation, R408W on haplotype 2 was found with a relative frequency of 37 %. P281L accounted for 11 %, R261Q and E390G each for 9 % of mutant chromosomes. There were three novel mutations: L249P (c.746T>C) in exon 7, IVS8+2T>C (c.912T>C) in intron 8 and F402L (c.1206T>G) in exon 12 of the *PAH* gene. Twelve mutations were identified on single chromosomes only (Zschocke, 2003b).

Serbia

The Serbian population is characterised by a high number of different mutations in the *PAH* gene: 19 different disease-causing mutations have been identified. The most frequent mutations, L48S (21%), R408W (18%), P281L (9%), E390G (7%) and R261Q (6%), accounted for 60% of all mutant alleles. Less frequent ones were: R158Q (4.4%), I306V (4.4%), IVS12+1G>A (4.4%), Q20X (2.9%), R111X (2.9%), V177L (2.9%), P225T (2.9%), R261X (2.9%), p.S16>XfsX1 (1.5%), S231F (1.5%), R252Q (1.5%), R297H (1.5%), IVS10-11G>A (1.5%), and R413P (1.5%). Homozygosity was observed in three patients only, which gives a frequency of homoallelic PKU genotypes of 8.82%. As expected, each of them was homozygous for one of the three most frequent mutations. This finding suggests that PKU in Serbia is heterogeneous (Stojiljković, 2006; Stojiljković, 2007).

Greece

PKU patients in Greece were examined on the presence of the 9 following mutations: R158Q, R261Q, Y414C, E280K, R252W, P281L, R408Q, IVS10–11G>A and IVS12+1G>A. Preliminary mutation data from Greece showed relative frequencies of over 10% for mutations P281L and IVS10–11G>A, but the total mutation detection rate was only 30% (Traeger-Synodinos, 1994).

Italy

PAH gene mutations data were obtained from five Italian regions, Calabria, Campania, Piemonte, Puglia/Basilicata and Sicilia. The four most frequent mutations were IVS10–11G>A (19.4%), R261Q (13.5%), L48S (9.7%) and R158Q (4.8%) and nine mutations (IVS10–11G>A, R261Q, L48S, R158Q, R252W, R261X, A300S, F55fs,

P281L) represented 65% of total PKU chromosomes. These findings confirm the heterogeneity of the molecular basis of PKU in Italy (Giannattasio, 2001).

Spain

Spanish PKU chromosomes has been found to carry 67 different mutations, among them 17 are unique to this population and 10 of them are described here for the first time. The four most frequent mutations: severe IVS10–11G>A (10%), MHP mutation A403V (8%), and two mild mutations V388M (6%) and I65T (7%), account for only 31% of the total PKU chromosomes. From the remaining mutations 39 are very rare, present only on one or two mutant chromosomes and the others have frequencies between 0.8–4.5%. Mutations IVS12+1G>A and R408W are totally absent in Spain PKU population. The underlying genetic heterogeneity is the basis of the biochemical and clinical diversity of the disorder. The high frequency of mutations with a low degree of severity can explain the relatively higher prevalence of MHP and mild PKU phenotypes in Spain compared with Northern European populations. (Desviat, 1999).

Portugal

Common mutations in Portugal include IVS10–11G>A, R261Q, and V388M. There is no one mutation(s) prevalent in the Portuguese PKU population; rather three groups of mutations exist for which relative frequencies are 9-11% (IVS10–11G>A, R261Q, V388M), 3-6% (I65T, P281L, R252W, R158Q), and below 2% (L348V, Y414C, L311P, Y198fsdel22, R408W, R270K and R261X), respectively (Rivera, 1998).

Thus phenylketonuria in Europe has multiple origins. There have been numerous independent mutation events, and several mutations have independently recurred in different founders. There is not a single PKU mutation that is found in all European populations, nevertheless, there are a limited number of mutations that can be regarded as common.

1.8 *PAH* GENE MUTATIONS' ASSOCIATIONS WITH MINIHAPLOTYPES

Minihaplotypes (combined VNTR and STR) data are used to detect the origin of individual mutation and also for the rapid identification of rare mutations. Compared with conventional haplotypes (combined data from the VNTR site and seven diallelic RFLPs), minihaplotypes are much easier to obtain and are more informative for mutation analysis.

VNTRs are quite stable, and conventional haplotypes generally change through recombination only. In contrast, STR alleles are known to change much more frequently (Weber and Wong, 1993). Minihaplotypes combine the stability of VNTR alleles with the wide spectrum of STR alleles. These qualities make them particularly useful in a diagnostic setting (Zschocke, 1995).

PKU mutations are shared between several populations. Mutations spread with migrating peoples from founder populations to other regions and are frequently distributed over many countries. Analysing minihaplotypes for as many mutations as possible, investigating the origins of mutations, and studying the genetic history of different populations can thus further improve efficiency for diagnostic mutation analysis in PKU (Zschocke, 1995).

Germany

Nevertheless, detailed data about *PAH* gene mutations minihaplotypes are available from few populations. Many countries still have only conventional haplotypes data. The most detailed investigation of minihaplotypes was performed in German population. Minihaplotypes were obtained for all 91 mutations identified in German PKU patients. The most common mutation R408W was associated with minihaplotypes 3/238 (VNTR/STR) in 98% of all alleles, mutation IVS12+1G>A with minihaplotype 8/242 in all cases. Mutations L48S and R158Q both were found exclusively in association with minihaplotype 3/234. Other common mutations R261Q and P281L were associated with several minihaplotypes. Mutation IVS10-11G>A was mostly associated with minihaplotype 7/250 and was the most common in the subgroup of patients of Turkish descent (Appendix V) (Zschocke and Hoffmann, 1999).

Poland

The association of mutations R408W, IVS10-11G>A and A403V common in different European populations with a VNTR and STR in the *PAH* gene was examined in a group of Polish PKU and MHP patients. Additionally, minihaplotypes were established for another 16 mutations, but results were reported only for 10 mutations. The prevalent minihaplotype for mutation R408W was 3/238 but for IVS10-11G>A – 7/250 corresponding to Mediterranean origin. Mutation A403V was associated with 2 minihaplotype variants: 8/242 and 8/246; but mutation P281L with 8/242 and 3/242 (Appendix V) (Zekanowski, 2001).

Italy

PAH gene mutations and minihaplotypes have been determined for 78.5% and 64%, respectively, of the chromosomes studied in five Italian regions, Calabria, Campania, Piemonte, Puglia/Basilicata and Sicilia. Twenty-one (21) different minihaplotypes and twenty-four (24) PKU mutations were found. Among the fifteen minihaplotypes associated to specific *PAH* mutations, most were associated to more than three mutations. Minihaplotypes 3/234 and 3/238 associated with eight and seven mutations, respectively. On the other side, most of PKU mutations, including the most common ones L48S, R261Q, R261X and IVS10-11G>A, are associated to more than one minihaplotype (Appendix V) (Giannattasio, 2001).

Northern Ireland

Minihaplotype data were obtained for all mutant chromosomes in Northern Ireland. Twenty-one (21) different alleles were found. Most common were minihaplotypes 8/242 and 8/246, because of the high frequency of mutations R408W and I65T in the population. Other minihaplotypes for mutation R408W were 3/238, 8/238, 8/230 and 3/242 (Appendix V) (Zschocke, 1995).

The more recent study from Ireland confirmed that the predominant minihaplotype association for R408W was with the 8/242 minihaplotype (49 of 60 alleles, 81.7%). Minihaplotype 3/238 was found in smaller frequencies (9 of 60 alleles, 15%), minihaplotype 8/246 in two chromosomes (Appendix V) (O'Donnell, 2002).

Spain

The most common *PAH* gene mutation in Spain IVS10-11G>A was found in association with 7 minihaplotypes, two of them were predominant – 7/246 and 7/250. Mutation E280K was associated with minihaplotype 9/234 only (Appendix V) (Perez, 1997).

Lithuania and Estonia

Minihaplotypes studies in Lithuanian and Estonian PKU patients were performed only for the most common mutation R408W. In Lithuania R408W mutation was strongly associated with the three-copy VNTR and the 238-bp STR allele. The frequency of this association was 68% (Giannattasio, 1997). R408W mutation, accounting 84% of mutant chromosomes, was found on typical East-European minihaplotype 3/238 in all chromosomes in Estonia (Ounap, 1998).

2 MATERIALS AND METHODS

2.1. SUBJECTS

2.1.1. PKU screening procedure

Latvian PKU patients are initially diagnosed as having PAH deficiency through the National Newborn Screening Program that is being implemented since 1980 in Riga and 1985 in whole Latvia.

The blood specimen for PKU screening must be obtained at 72 hours after birth before mother and child are discharged from the maternity unit. Sampling time mentioned above is important because the phenylalanine level of affected infants rises gradually after birth with little, if any, effect of the amount of protein ingested by the infant. Screening at 72 hours is recommended to maximize detection. The practice of early discharge from the nursery may lead to a falsely negative screening result.

Phe test is performed on blood spotted on filter paper with the fluorometric assay. In case of positive result (Phe > 120 μmol/L or 2mg/ %) the test should be repeated from the same blood sample. If the repeated test is positive, new blood sample request is sent to paediatrician. After confirmation of diagnosis from newly received blood spotted on filter paper a child with parents are called for consultation. Result of Phe analysis taken during the consultation is pre-treatment Phe level that is used to classify the severity of PKU. Accordingly, patients were classified as having classic or severe PKU with plasma Phe >1200 μmol/L, mild PKU with Phe level 360–1200 μmol/L and MHP with Phe <360 μmol/L. PKU dietary treatment was applied to patients with Phe levels >240 μmol/L.

All Latvian PKU patients are being diagnosed and treated in Medical Genetics Clinic in Riga.

2.1.2. DNA sampling

The collection of blood samples for DNA analysis from PKU patients and their first degree relatives was started in 1997. All patients with PKU registered in Medical Genetics Clinic (born before and after 1997) were included in our study with their parents' consent.

Patients with transitory HPA who did not required treatment were not included in this study. However, patients with unclear or borderline parameters were investigated to clarify patients' status.

DNA samples for control group were collected from volunteers from the mixed population and were investigated for the absence of the *PAH* gene sequence changes to analyse the possible population specific polymorphism.

The study was conducted in accordance with the Helsinki convention and approved by the Central Medical Ethics Committee.

2.2. MOLECULAR STUDIES

Molecular studies were based on results of Joint (Lithuania, Italy, Latvia and Germany) research project "Molecular Genetic Testing in Phenylketonuria: a Model to Assess the Quality Control System for Monogenic Disease" (Acronym: MOLGENT) supported by EC INCO-COPERNICUS programme 1998-2002. The purpose of the project was to standardise and optimise molecular genetic testing in PKU. A set of standardised schemes and protocols developed during the project was used in our study.

Equipment used:

- ✓ Genomic DNA Purification Kit („Fermentas", Lithuania)
- ✓ Thermal cycler ("Eppendorf", Germany/"AppliedBiosystems", USA)
- ✓ Microcentrifuge ("Sigma", USA)
- ✓ System for horizontal gel electrophoresis ("Sigma", USA)
- ✓ System for vertical gel electrophoresis ("Sigma", USA)
- ✓ System for DGGE *Ingeny phorU* ("Ingeny International", The Netherlands)
- ✓ UV transilluminator ("Vilber Lourmat", Germany)
- ✓ Image analysis system ("SynGene", UK).
- ✓ ABI 310 Genetic Analyzer ("Applied Biosystems", USA)

Reagents used:

- ✓ Native *Taq* DNA polymerase with BSA („Fermentas", Lithuania)
- ✓ 2mM dNTP Mix („Fermentas", Lithuania)
- ✓ Restriction enzyme *Sty*I („Fermentas", Lithuania)
- ✓ Oligonucleotide primers ("Eurofins MWG-Operon", Germany)

- ✓ BigDye® Terminator v3.1 Cycle Sequencing Kit
- ✓ POP-6™ polymer
- ✓ 10x TBE buffer ("INNO-TRAIN Diagnostik", Germany)
- ✓ Agarose („Fermentas", Lithuania)
- ✓ 6x loading solution („Fermentas", Lithuania)
- ✓ Ethidium bromide stock solution ("Sigma", USA)

2.2.1. Genomic DNA Extraction

Genomic DNA was purified from whole blood leucocytes. Blood samples were collected in EDTA-coated tubes to prevent clotting and DNA degradation.

Genomic DNA Purification Kit („Fermentas", Lithuania) was used for DNA isolation according to the manufacturer instructions. Typically 200 µl of frozen blood was used for DNA isolation with the yield of 2-10 µg. For larger DNA quantity 500 µl of blood were lysed with 1 ml of deionised water, leukocytes spun down (5000 rpm, 2 min) and the pellet resuspended in 50 - 200 µl of TE buffer or sterile deionised water. DNA concentration was measured spectrophotometrically or evaluated visually after electrophoresis in agarose gel.

2.2.2. Identification of *PAH* gene mutation R408W (c.1222C→T)

PAH gene mutation R408W is a transition C→T at the position 1222 of cDNA (Appendix VI) causing amino acid substitution (codon CGG→TGG). Diagnostic identification of mutation R408W is based on the fact that it creates new restriction enzyme *Sty*I site in the exon 12 (Fig. 2-1).

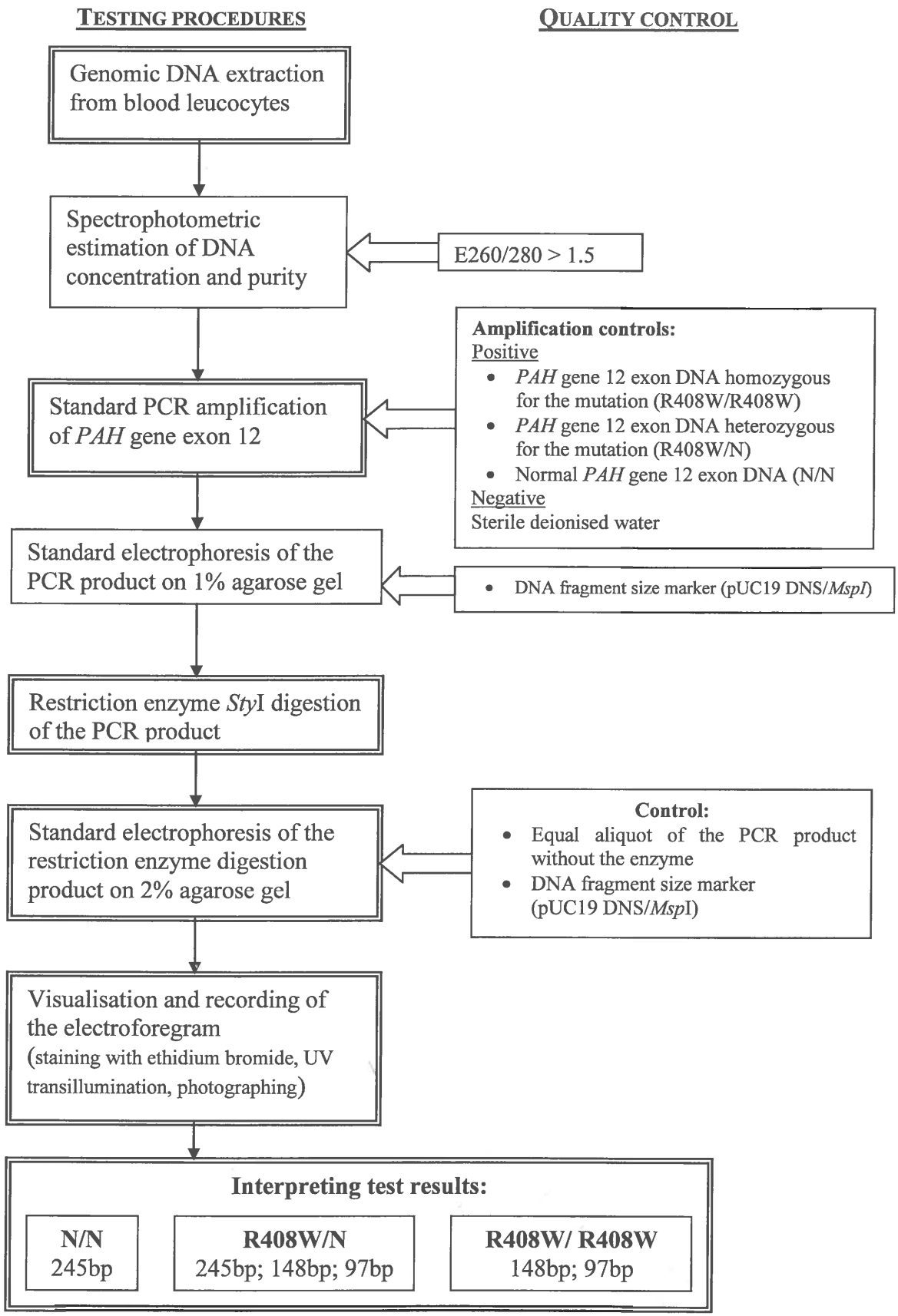


Fig. 2.1. Scheme of the diagnostic testing for the *PAH* gene mutation R408W (MOLGENT 1999-2002).

PAH gene exon 12 was amplified by PCR in a 30µl volume containing 100 to 250 ng of genomic DNA, 0.2 mM of each dNTP, 100mM KCl, 20mM Tris HCl (pH8.0), 2.5mM $MgCl_2$, 1 unit of native *Taq* DNA polymerase ("Fermentas"), and 20pmol of each primer (Forward: 5'- ATGCCACTGAGAACTCTCTT-3' Reverse: 5'-AGTCTTCGATTACTGAGAAA-3'; Appendix VI). As an internal control for reagents contamination, blank sample (template substituted by sterile deionised nuclease-free water) was used.

Also two positive controls and one normal control for the following restriction analysis were amplified:

- *PAH* gene 12 exon DNA homozygous for the mutation (R408W/R408W)
- *PAH* gene 12 exon DNA heterozygous for the mutation (R408W/N)
- Normal *PAH* gene 12 exon DNA (N/N)

PCR was carried out for an initial 5 min at 95°C followed by 30 cycles of 1 min at 95 °C, 1 min at 50°C, 1 min at 72°C, and a final elongation step at 72°C for 10 min. After completion of the thermal cycling protocol, aliquots (5 µl) of PCR product were electrophoresed on 1% agarose gel in 0.5x TBE buffer and visualised by ethidium bromide staining.

Following digestion of *PAH* gene exon 12 PCR product with the restriction enzyme *Sty*I was performed in a total reaction volume of 20µl.

Restriction site for restriction enzyme *Sty*I:

AA
5'...C↓CTTGG...3'
3'...GGAACC↑...5'
TT

Restriction mixture was prepared with the appropriate reaction buffer supplied by the manufacturer („Fermentas", Lithuania") along with the 10U of the enzyme. Reaction mix was incubated at 37°C overnight and restriction products as well as an aliquot of the uncut DNA were visualised in an ethidium bromide stained 2% agarose gel. pUC19 DNS/*Msp*I DNA fragment size marker was used for fragments' length detection.

Undigested *PAH* gene exon 12 PCR product is 245bp DNA fragment. Different *PAH* gene exon 12 alleles digested with restriction enzyme *Sty*I produce different sets of DNA fragments. Interpretation of the restriction enzyme digestion results is shown in Table 2.1.

Table 2.1. Results of *PAH* gene exon 12 PCR-amplified DNA sample digestion with restriction enzyme *Sty*I (MOLGENT 1999-2002).

PAH gene exon 12 locus genotype	*Digestion with StyI*	
	Results	DNA fragments produced
Homozygous (normal) **N/N**	No digestion	245bp
Heterozygous **R408W/N**	N allele undigested, R408W allele digested	245bp, 148bp, 97bp
Homozygous (mutant) **R408W/R408W**	Full digestion	148bp, 97bp

2.2.3. Denaturing gradient gel electrophoresis (DGGE) of *PAH* gene exons

Non-R408W chromosomes were screened for mutations through denaturing-gradient gel electrophoresis (DGGE) of the 13 exons of the *PAH* gene (Guldberg and Güttler, 1994). DGGE signal pointing to a possible nucleotide sequence change can be best visible when several DNA samples of the same gene fragment are in line on the gel; normal band is necessary for comparison. Therefore, strategy "all *PAH* gene exons of a single patient on a single gel" in the case of 16–25 lanes gel is possible only in the case of multiplex DGGE (Figs. 2.2 and 2.3). Multiplex DGGE is applicable only when a good standard DGGE picture is obtained.

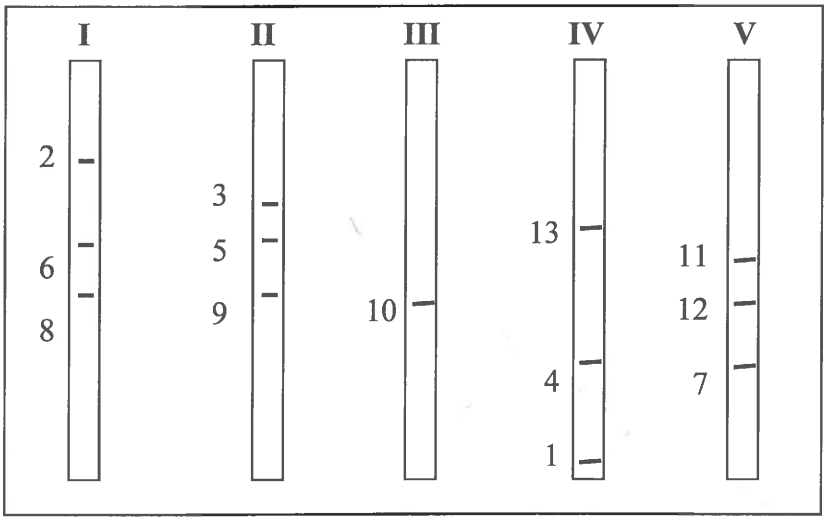


Fig. 2.2. Scheme of multiplex DGGE of 1–13 exons of *PAH* gene. I–V – lanes (groups of lanes) on the gel. Positions of the bands corresponding to each exon are shown on the left of a relevant lane (MOLGENT 1999-2002).

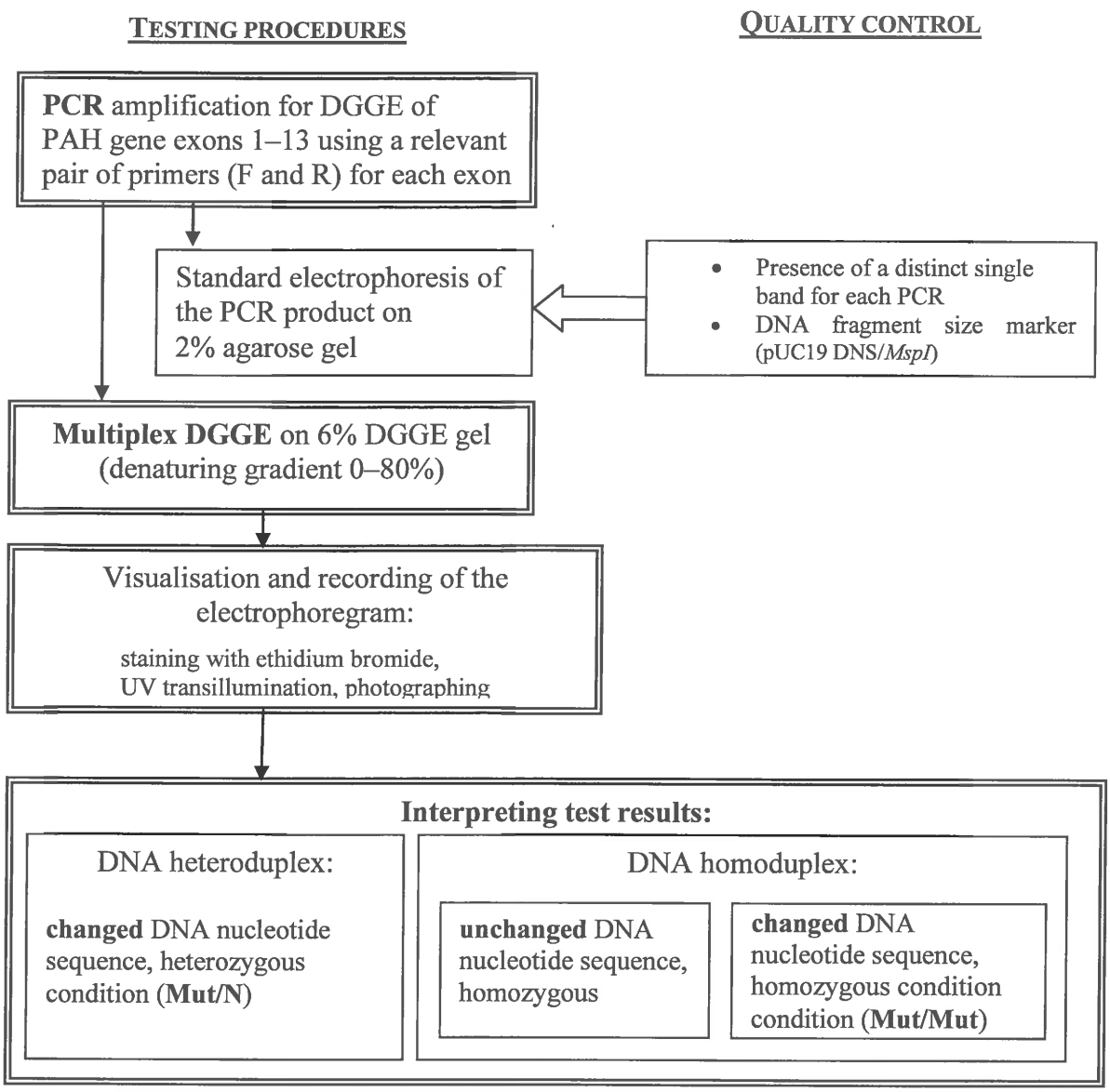


Fig. 2.3. Scheme of the multiplex DGGE of *PAH* gene exons (MOLGENT 1999-2002).

Genomic DNA for the amplification of each of *PAH* gene exon for the DGGE analysis was added to a solution that contained 10x PCR buffer (750 mM Tris-HCl (pH 8.8 at 25°C), 200 mM (NH4)2SO4, 0.1% (v/v) Tween 20), 0.2mM of each dNTP, 2.5mM $MgCl_2$, 20pmol of each primer (Table 2.2) and 1.25U of Native Taq DNA polymerase in a total reaction volume of 50 µl.

PCR for *PAH* gene exons 1-13 was carried out for an initial 5 min at 94°C followed by 30 cycles of 45 sec at 94 °C, 45 sec at 58°C, 45 sec at 72°C. This was followed by a last cycle of 5 min at 72 °C, 5 min at 94 °C, 20 min at 58°C and 20 min at 37 °C for heteroduplex formation.

Table 2.2. GC-clamped oligonucleotide primers for the amplification of *PAH* gene exons (MOLGENT 1999-2002).

Exon	Nucleotide sequence
PAH-1	5'-[50GC]-TTAAAACCTTCAGCCCCACG-3' 5'-TGGAGGCCCAAATTCCCCTAACTG-3'
PAH-2	5'-GAGGTTTAACAGGAATGAATTGCT-3' 5'-[40GC]-TCCTGTGTTCTTTTCATTGC-3'
PAH-3	5'-[40GC]-GCCTGCGTTAGTTCCTGTGA-3' 5'-CTTATGTTGCAAAATTCCTC-3'
PAH-4	5'-ATGTTCTGCCAATCTGTACTCAGGA-3' 5'-[40GC]-CAAGACATAGGCCATGGACT-3'
PAH-5	5'-TCATGGCTTTAGAGCCCCCA-3' 5'-[40GC]-AGGCTAGGGGTGTGTTTTTC-3'
PAH-6	5'-[40GC]-CCGACTCCCTCTGCTAACCT-3' 5'-CAATCCTCCCCCAACTTTCT-3'
PAH-7	5'-[40GC]-GGTGATGAGCTTTTAGTTTTCTTTC-3' 5'-AGCAAATGAACCCAAACCTC-3'
PAH-8	5'-[40GC]-TGGCTTAAACCTCCTCCCCT-3' 5'-CTGGGCTCAACTCATTTGAG-3'
PAH-9	5'-ATGGCCAAGTACTAGGTTGG-3' 5'-[40GC]-GAGGGCCATAGACTATAGCA-3'
PAH-10	5'-[40GC]-TTAACGATCATAGAGTGTGC-3' 5'-ACAAATAGGGTTTCAACAAT-3'
PAH-11	5'-TGAGAGAAGGGGCACAAATG-3' 5'-[40GC]-GCCAACCACCCACAGATGAG-3'
PAH-12	5'-ATGCCACTGAGAACTCTCTT-3' 5'-[40GC]-ACTGAGAAACCGAGTGGCCT-3'
PAH-13	5'-[40GC]-GACACTTGAAGAGTTTTTGC-3' 5'-TTTTCGGACTTTTTCTGATG-3'

GC clamps:

[50GC]=CCCGCCGCCCGCCGCTCGCCCGCCGCGCCCCTGCCCGCCGCCCCCGCCCG

[40GC]=CGCCCGCCGCGCCCCGCGCCCGTCCCGCCGCCCCCGCCCG

After completion of the thermal cycling protocol, aliquots (5 µl) of PCR products were electrophoresed on 2% agarose gel in 1x TBE and analysed after ethidium bromide staining.

DGGE is an electrophoretic separation method based on differences in melting behaviour of double stranded DNA fragments (Fisher and Lermann, 1979). The

electrophoresis takes place in a vertically placed polyacrylamide gel in a gradient of denaturants. In addition the gel should be kept at constant homogeneous temperature ~60°C.

To set up a DGGE gel two acrylamide solutions with denaturant gradient of 0% and 80% are used (Table 2.3).

Table 2.3. Preparing of 0% and 80% Denaturant stock solutions (MOLGENT 1999-2002).

Gel solutions (100ml)	Denaturant stock solutions	
	0%	80%
40% Acrylamide/Bis-acrylamide (37.5 : 1)	15 ml	15 ml
Formamide	—	32 ml
Urea	—	34 g
50×TAE buffer	2 ml	2 ml
ddH_2O	Up to 100 ml	Up to 100 ml

Note: the solutions should be stored in dark glass bottles at 4°C.

Gel contains gradient of formamide and urea linearly increasing from the top the bottom of the gel. Use the gradient maker placed on the top of a magnetic stirrer to make the denaturing gradient.

Add 25µl of TEMED (N,N,N'N'-tetramethylenediamine) and 500µl of 10% APS (ammonium persulfate) to 30ml of each stock solution to prepare working solutions and mix well. Using the gradient maker fill the gel cassette with the gel. Allow the gel to polymerize for about 30 to 60 minutes.

Prepare the samples for the loading into the gel by mixing them with 6x Loading Dye. Five groups of PCR-amplified DNA samples to be loaded on corresponding groups of lanes are prepared immediately before loading according to the scheme on Fig. 3-3:

- ✓ I group of lanes: exon 2 + exon 6 + exon 8;
- ✓ II group of lanes: exon 3 + exon 5 + exon 9;
- ✓ III group of lanes: exon 10;
- ✓ IV group of lanes: exon 13 + exon 4 + exon 1;
- ✓ V group of lanes: exon 11 + exon 12 + exon 7.

Mix 8-10µl of each sample in groups and add 3µl of the 6x loading solution to the samples.

Place the frame with the gel into the bath containing 1× TAE buffer pre-heated to 60°C, flush the slots with syringe fitted with a needle, and load the samples. Run DGGE for about 16 h at 75V. After electrophoresis stain the gel for about 10 min in 1× TAE, containing 0.5 μg/ml of ethidium bromide.

Interpretation of the DGGE results is based on different electrophoretic mobility of double-stranded DNA fragments differing in a single nucleotide pair on the denaturing gradient gel (Fig. 2.4).

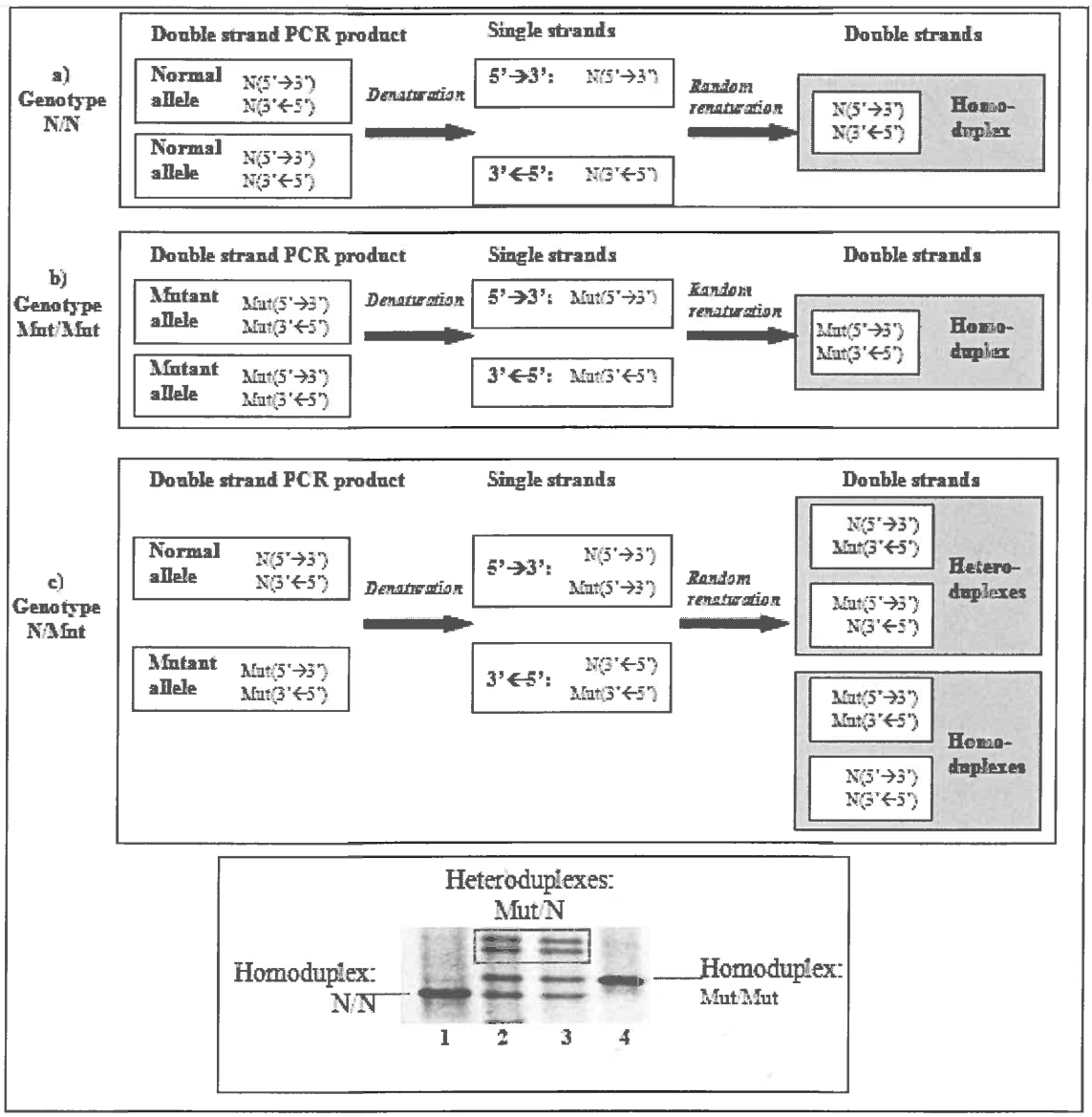


Fig. 2.4. Upper and middle panels: formation of double-stranded DNA fragments with different mobility on denaturing gradient gel. Lower panel: fragment of a DGGE electrophoregram of *PAH* gene exon 12. Lane 1, normal DNA; lanes 2 and 3, mutation in heterozygous condition; lane 4, mutation in homozygous condition (MOLGENT 1999-2002).

DNA homoduplex formation is an indicative of:

a) an **unchanged** DNA nucleotide sequence, homozygous condition (**N/N**) – Fig. 2.4: upper panel (a); lower panel, lane 1;

b) a **change** in the DNA nucleotide sequence, homozygous condition (**Mut/Mut**) – Fig. 2.4: middle panel (b); lower panel, lane 4.

Bands corresponding to Mut/Mut and N/N can be distinguished when several DNA samples of the same gene fragment are in line on the gel (compare lanes 1 and 4 on lower panel of Fig. 2.4). Also, a normal band is necessary for comparison.

DNA heteroduplex formation is an indicative of a **change** in the DNA nucleotide sequence in heterozygous condition: **Mut/N** (Fig. 2.4, lanes 2, 3) or two different mutations (**Mut_1/Mut_2**) in one PCR-amplified DNA fragment. Bands corresponding to Mut_1/Mut_2 and Mut/N form different electrophoretic patterns and can be distinguished when several DNA samples of the same gene fragment are in line on the gel. Also, a normal band is necessary for comparison.

2.2.4. Automated direct sequencing of *PAH* gene exons

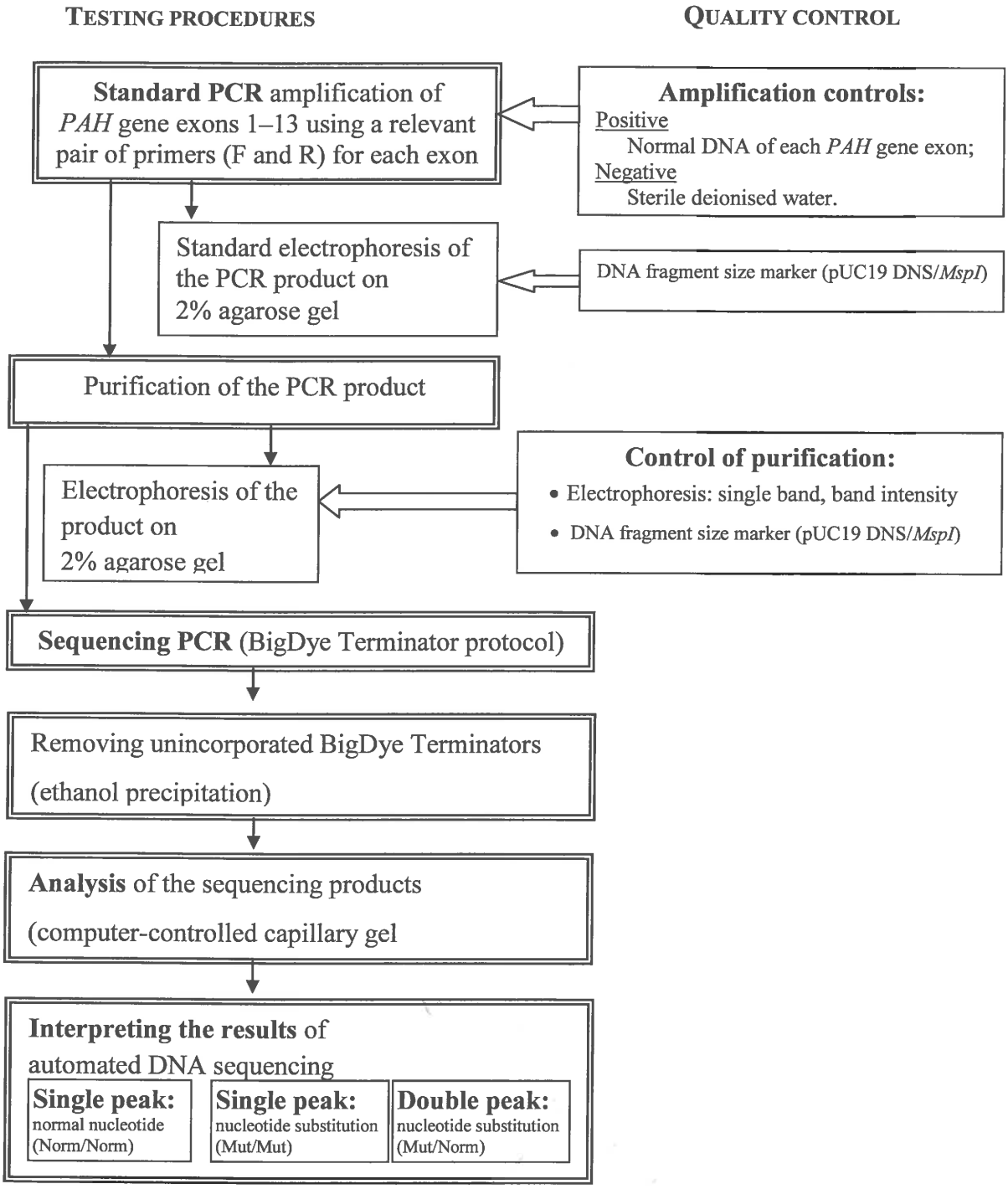


Fig. 2.5. Scheme of the direct automated sequencing of *PAH* gene exons (MOLGENT 1999-2002).

Exons that showed variant electrophoretic patterns on DGGE were next analysed by direct sequencing (Fig. 2.5).

The reaction mixture for amplification of *PAH* gene exons contained 10x PCR buffer (200 mM Tris-HCl (pH 8.3 at 25°C), 200 mM KCl, 50 mM $(NH_4)_2SO_4$), 0.2mM of each dNTP, 2.5mM $MgCl_2$, 25pmol of each primer for an appropriate exon (Table 2.4) and 1U of Hot Start *Taq* DNA Polymerase in a total reaction volume of 50 µl. 200–500 ng of genomic DNA were used for the reaction.

Table 2.4. Primers for the PCR amplification of *PAH* gene exons (MOLGENT 1999-2002).

***PAH* gene exon**	**Oligo-nucleotide primer***	***Nucleotide sequence***
Exon 1	PAH-S-X1-F	TATTATAGGGCGAATTGGGTCTCCTGCGTCCCCCACAC
	PAH-S-X1-R	TATGTAAAACGACGGCCAGTCCCAAATTCCCCTAACTGAG
Exon 2	PAH-S-X2-F	TATTATAGGGCGAATTGGGTGAGTTCATGCTTGCTTTGTC+
	PAH-S-X2R	TATGTAAAACGACGGCCAGTCTGTTCCAGATCCTGTGTTC
Exon 3	PAH-S-X3-F	TATTATAGGGCGAATTGGGTCCTGCGTTAGTTCCAGTGAC
	PAH-S-X3-R	TATGTAAAACGACGGCCAGTTTAATCCCCCAACAGTCTTC
Exon 4	PAH-S-X4-F	TATTATAGGGCGAATTGGGTCCCACTTGCCATCACCATTG
	PAH-S-X4-R	TATGTAAAACGACGGCCAGTATTTTTCCCAGCCCTCGTGT
Exon 5	PAH-S-X5-F	TATAAGGGAACAAAAGCTGGTAACCAAGGGAAGGAGACAT
	PAH-S-X5-R	TATTATAGGGCGAATTGGGTATGAGGGCAAGGGAGAAG
Exon 6	PAH-S-X6-F	TATTATAGGGCGAATTGGGTTGATGGCAGCTCACAGGTTC
	PAH-S-X6-R	TATGTAAAACGACGGCCAGTTCTTCCCCTTCCCTCTCCTC
Exon 7	PAH-S-X7-F	TATTATAGGGCGAATTGGGTACATCTGAAGCCAAGTCTG
	PAH-S-X7-R	TATAAGGGAACAAAAGCTGGAGCAATGAACCCAAACCTC
Exon 8	PAH-S-X8-F	TATTATAGGGCGAATTGGGTGGCTTGGCTTAAACCTCCTC
	PAH-S-X8-R	TATGTAAAACGACGGCCAGTCTCCCTGGGCTCAACTCATT
Exon 9	PAH-S-X9-F	TATTATAGGGCGAATTGGGTTCTATGTGGGCTGTTCTGA
	PAH-S-X9-R	TATGTAAAACGACGGCCAGTCCAGGGGAGTAGGAAAGTT
Exon 10	PAH-S-X10-F	TATGTAAAACGACGGCCAGTCCCCAAAATAATGCTTTACT
	PAH-S-X10-R	TATAAGGGAACAAAAGCTGGACGGATACAAATAGGGTTTC
Exon 11	PAH-S-X11-F	TATAAGGGAACAAAAGCTGGGAATCGGGGTGAGATGAGAG
	PAH-S-X11-R	TATTATAGGGCGAATTGGGTGAGTGGCACCAGTCAGGAG
Exon 12	PAH-S-X12-F	TATGTAAAACGACGGCCAGTAATGGTGCCCTTCACTCAA
	PAH-S-X12-R	TATAAGGGAACAAAAGCTGGTTTTTCCTATGGCGATGGTA
Exon 13	PAH-S-X13-F	TATAAGGGAACAAAAGCTGGTCCAAGAAGCCCACTTATC
	PAH-S-X13-R	TATTATAGGGCGAATTGGGTTGATGAAATGCGACAGATTAC

* F= forward, R= reverse

PCR amplification for *PAH* gene exon proceeded for 30 cycles under the following conditions: denaturing for 1 min at 95°C , annealing for 1 min at 57°C and synthesis for 1 min at 72°C; extended with initial denaturation cycle by 5 min at 95°C and final synthesis cycle by 5 minutes at 72°C.

A 5 µl aliquot of each reaction solution was dissolved in 1% agarose gel and visualised by adding ethidium bromide to assess yield and purity of PCR product before sequencing.

To remove the unincorporated nucleotides and primers PCR products were purified before the direct sequencing procedure by using NucleoSpin® Extract II kit ("Macherey-Nagel", Germany) according to the manufacturer instructions.

After purification PCR products were again checked the on 2% agarose gel. Prepare the loading samples by mixing 3µl of each sample with 0.5µl of 6× loading solution for PCR product analysis. Purified DNA should run as a single band on an agarose gel. Then analysis was continued with the direct sequencing.

BigDye® Terminator v3.1 Cycle Sequencing Kit was used for automated cycle sequencing according to the manufacturer instructions. BigDye is a set of dye terminators labelled with high-sensitivity dyes.

The reaction mix contained the following reagents: 1x BigDye Sequencing Buffer, 1x Terminator Ready Reaction Mix, 3.2 pmol of primer, 3-10ng of template and water to 20µl.

PCR was run per manufacturer's directions. A 5-minute denaturation step at 96°C followed by 25 cycles of 10 seconds denaturation at 96°C, 5 seconds annealing at 50°C and synthesis step at 60°C for 4 minutes. Hold at 4°C until ready to purify.

Purification of extension products by Ethanol/Potassium Acetate Precipitation was the method for preparing extension products for electrophoresis (removing unincluded BigDye Terminators, nucleotides, oligonucleotide primers). Purifying protocol was as follows:

- Add 2 µl of 3 M potassium acetate, pH 5.6 and 50µl of 96% ethanol to the each reaction tube. Mix by pipetting and leave at room temperature for 20 min. Centrifuge at 4°C for 20 min at 13 000 rpm. Aspirate the supernatant.
- Add 200 µl 75% EtOH to the each tube. Vortex thoroughly.
- Centrifuge for 5 min at 13 000 rpm. Aspirate the supernatant.
- Dry the pellets at 60°C for 5 min.

Electrophoresis and data analysis of samples on the ABI PRISM® 310 Genetic Analyzer required the following: POP-6™ polymer, 1-mL syringe, 61-cm, 50-µm i.d. capillary, run module - Seq POP6 (1 mL) E and Dye Set/Primer (Mobility) File - DT310POP6{BDv3}v1.mob.

For preparation and loading of the samples 20-25µl of Deionised formamide to each sample pellet were added and each tube was sealed with a septum. Then samples were mixed thoroughly on a vortex mixer and heated for 2 minutes at 95°C, chilled on ice for 3 min or until ready to load on the instrument.

Capillary electrophoresis was run on an ABI Prism® 310 genetic analyzer and electrophoregram was analysed by ABI DNA™ sequencing software (software for PE Applied Biosystems Genetic Analyzer: ABI PRISM™ 310 Collection and Sequencing Analysis).

The results of the automated capillary gel electrophoresis are presented graphically as curves (i. e. 5'→3' sequences of peaks) in different colours. Each line represents fractions of DNA fragments generated by random incorporation of a definite type the Dye Terminator, which is a marker for a definite 3' terminal nucleotide: red – T, blue – C, green – A, black – G.

DNA sequence calculated from the lines of each colour is placed above the lines matching each nucleotide to a relevant peak.

The results of the direct DNA sequencing are interpreted by the comparison of the identified DNA sequence with the normal DNA sequence of the sequenced DNA fragment. In the case of an automated DNA sequencing the sequence is analysed by relevant computer software. A single peak is interpreted by the software as a relevant nucleotide. A double peak is interpreted by the software as unreadable nucleotide (N nucleotide in the DNA sequence).

<u>All peaks are single:</u>

a) Identified nucleotide sequence is identical to the normal nucleotide sequence: normal nucleotide sequence (Norm/Norm),
b) Identified nucleotide sequence differs from the normal nucleotide sequence: mutation in homozygous condition (Mut/Mut).

<u>Double peak:</u>

Nucleotide substitution (Mut/Norm).

Double peak should be analysed manually to identify which peak corresponds to the normal DNA sequence of the Norm allele and which peak corresponds to the changed DNA sequence of the Mut allele.

Unreadable line (all double peaks) downstream a certain peak:

Insertion or deletion in heterozygous condition (Mut/Norm) is located next to the last single peak. To specify the mutation, double peaks should be analysed manually to identify which peak corresponds to the normal DNA sequence of the Norm allele and which one corresponds to the changed DNA sequence of the Mut allele.

2.2.5. *PAH* gene minihaplotype analysis

Two *PAH* gene satellite systems were analysed to create minihaplotypes: minisatellite system - Variable Number of Tandem Repeats (VNTR) and microsatellite system - Short Tandem Repeats (STR) (Goltsov, 1992; Zschocke, 1994(a)).

To amplify VNTR alleles PCR reaction was performed using 200-500 ng of genomic DNA, 10x PCR buffer (750 mM Tris-HCl (pH 8.8 at 25°C), 200 mM $(NH_4)_2SO_4$, 0.1% (v/v) Tween 20), 0.2mM of each dNTP, 2.5mM $MgCl_2$, 20pmol of each primer: forward 5′-GCCAGAACAACTAC-TGGTTC-3′ and reverse 5′-AATCATAAGTGTTCCCAGAC-3, and 1U of Native *Taq* DNA polymerase in a total reaction volume of 30µl.

Initial denaturation for 4 min at 95°C followed by 30 cycles of 40 sec at 95°C, 40 sec at 55°C and 40 sec at 72°C. The final extension was at 72°C for 10 min.

After completion of the thermal cycling protocol, store the samples at +4°C or -20°C for longer time period. PCR products are checked on 6% PAA gel and visualised using ethidium bromide.

Electrophoretic resolution of the amplified products demonstrates DNA fragments of six discrete sizes -380, 470, 500, 530, 560, and 650 bp that correspond to the 3, 6, 7, 8, 9, or 12 copies of the repeated unit, respectively.

STR alleles were amplified in 20µl reaction volumes containing 50-100 ng DNA in 10x PCR buffer (750 mM Tris-HCl (pH 8.8 at 25°C), 200 mM $(NH_4)_2SO_4$, 0.1% (v/v) Tween 20), 0.2mM of each dNTP, 2.5mM $MgCl_2$, 1U of Native *Taq* DNA polymerase and 10pmol of each primer: forward 5`-6FAM-GCCAGAACAATACTGGTTC-3` and reverse 3`-AATCATAAGTGTTCCCAGAC- 3`

PCR conditions were as follows: denaturation at 95°C for 5 min; 33 cycles of 95°C for 1 min, 57°C for 1 min,72°C for 1 min; final extension at 72°C for 5 min.

After completion of the thermal cycling protocol, store the samples at +4°C or -20°C for longer time period. PCR products are resolved on an Applied Biosystems ABI Prism

310 Genetic Analyser running GeneScan software as follows 1μl of PCR product mix with 20μl of deionised formamide and 0.5μl of GeneScan™ -2500 ROX™ Size Standard (provides labelled fragments of 37, 94, 109, 116, 172, 186, 222, 233, 238, 269, 286, 361, 470, 490, and 536 bases as well as larger fragments.

Heat the loading mix for 3 minutes at 95°C. Immediately chill on ice for a few minutes and load samples.

The corresponding STR fragments' length (226 bp; 230 bp; 234 bp; 238bp; 242 bp; 246 bp; 250 bp; 254bp and 258bp) was calculated according to the calibration curve of the GeneScan™ ROX 2500™ size standard.

2.2.6 Programmes used for the results processing

Polymorphism Phenotyping

Novel amino acid changes found in Latvian PKU chromosomes that were not previously reported were analysed by *PolyPhen* software to gather information about it possible effect on PAH (http://genetics.bwh.harvard.edu/pph/).

PolyPhen (=*Poly*morphism *Phen*otyping) is an automatic tool for prediction of possible impact of an amino acid substitution on the structure and function of a human protein. This prediction is based on straightforward empirical rules which are applied to the sequence, phylogenetic and structural information characterizing the substitution.

The resulting multiple alignment is used by the new version of the PSIC software (**P**osition-**S**pecific **I**ndependent **C**ounts) to calculate the so-called *profile matrix*. Elements of the matrix (profile scores) are logarithmic ratios of the likelihood of given amino acid occurring at a particular position to the likelihood of this amino acid occurring at any position (background frequency).

PolyPhen computes the absolute value of the difference between profile scores of both allelic variants in the polymorphic position. Big values of this difference may indicate that the studied substitution is rarely or never observed in the protein family. *PolyPhen* also shows the number of aligned sequences at the query position. This number may be used to assess the reliability of profile score calculations.

PolyPhen uses empirically derived rules to predict that an nsSNP is

- **probably damaging**, i.e., it is with high confidence supposed to affect protein function or structure

- **possibly damaging**, i.e., it is supposed to affect protein function or structure
- **benign**, most likely lacking any phenotypic effect
- **unknown**, when in some rare cases, the lack of data do not allow *PolyPhen* to make a prediction

Appendix VII contains rules used by *PolyPhen* to predict effect of nsSNPs on protein function and structure. One row corresponds to one rule which may consist of several parts connected by logical "and". For a given substitution, all rules are tried one by one, resulting in prediction of functional effect. If no evidence for damaging effect is seen, substitution is considered benign.

Calculation of Homozygosity

Homozygosity (j) at the PAH locus in the population is determined using the equation $j = \Sigma x_i^2$, where x_i is the frequency of the ith allele. In populations where ascertainment of mutations is not 100%, each of the uncharacterised alleles is defined as having a frequency of 1/N, where N is the total number of mutant chromosomes investigated.

This value is the theoretical frequency of patients carrying two identical mutations. Homozygosity values of different populations reflect their mutational heterogeneity for the particular locus (Guldberg, 1996).

Calculation of Expected Heterozygosity for PKU minihaplotypes

The average level of heterozygosity of the VNTR/STR system was calculated according to the formula provided by Daiger *et al.*: $1 - \sum p_i$ (Daiger, 1989a).

Statistical methods

Statistical calculations were performed with STATISTICA 7 software using sub-module General linear/Non-linear model with one-way ANOVA, distribution binomial and logit function. The result for comparison of R408W mutation frequencies between two geographic populations was expressed as mean frequency ± 95% confidence interval.

Comparisons between minihaplotypes' variants in normal and mutant *PAH* chromosomes with "0" and "1" parameters were done using Chi-square (χ^2) 2x2 contingency table analysis.

Absolute allele frequencies in a population were estimated using incidence data and the Hardy-Weinberg formula: $p^2 + 2pq + q^2 = 1$ (http://www.changbioscience.com/genetics/hardy.html).

Chi-square test was used for comparison of absolute and relative *PAH* allele frequencies, as well as for comparison of two different approaches for classifying the PKU clinical phenotype (http://www.graphpad.com/quickcalcs/index.cfm).

Sensitivity measurements for methods used in our study were performed according to Altman and Bland, 1994. Sensitivity is the proportion of true positives that are correctly identified by the test.

3 RESULTS

A total of 74 Latvian patients with PKU, corresponding to 70 unrelated families and their 110 first degree relatives, were investigated. Most (57/70) of the patients, accounting for 81.5%, were identified in neonatal screening, the remaining (13/70) accounting for 18.5%, when they showed mental retardation in period between 1 and 5 years old. The oldest patient was born in 1967 but the youngest one in 2011. Six patients from five unrelated families were born before the National Newborn Screening Program was launched in the whole country.

Preliminary patients' phenotypes were classified according to the pre-treatment level of Phe: 64.3% (45/70) had severe PKU (Phe > 1200 μmol/L), 22.9% (16/70) had mild PKU (Phe 360–1200 μmol/L), 5.7% (4/70) had MHP (Phe < 360 μmol/L) and in 7.1% (5/70) cases Phe level was unavailable. Using the Phe tolerance (amount of dietary phenylalanine per day to keep plasma concentration of Phe at a safe level) as the additional classification parameter, 91.4% (64/70) patients were classified as having severe PKU, 5.7% (4/70) as having mild PKU and 2.9% (2/70) as having MHP (Table 3.1). Phe measurement units used in Latvia are mg% (1 mg% corresponds to 60 μmol/L).

Table 3.1. Classification of PKU clinical forms in Latvian patients.

PKU clinical form	By Phe pre-treatment level	N	%	By Phe tolerance	N	%
Severe	> 1200 μmol/L (>20 mg%)	45	64.3	250-350 mg	64	91.4
Mild PKU	360–1200 μmol/L (6 - 20 mg%)	16	22.9	350-600 mg	4	5.7
MHP	< 360 μmol/L (< 6 mg%)	4	5.7	< 600 mg	2	2.9
Unclassified	Unknown	5	7.1	—	—	—

3.1 MUTATION SPECTRUM IDENTIFIED FOR LATVIAN PKU PATIENTS

Considering the R408W mutation's high relative frequency in Europe, all new PKU patients primarily were tested for the presence of this mutation. Mutation R408W was analysed by *Sty*I restriction enzyme digestion analysis of *PAH* gene exon 12 (Fig. 3.1).

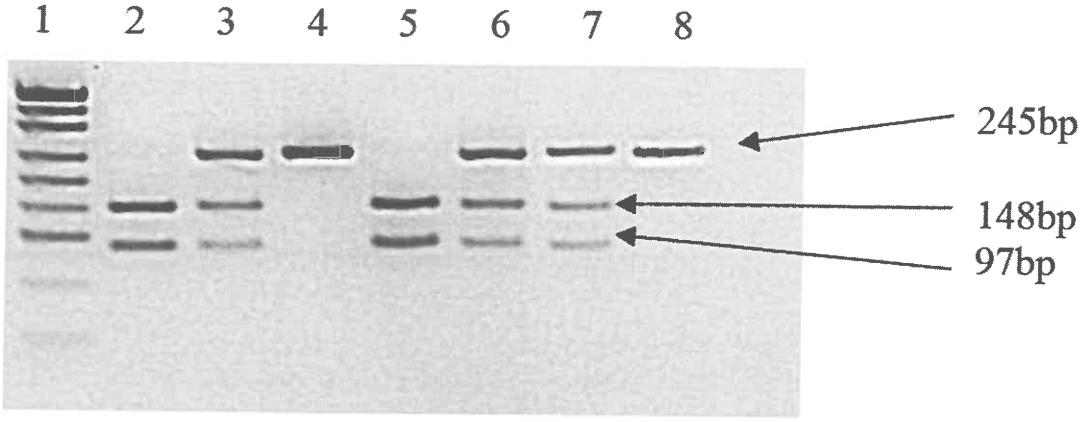


Fig. 3.1. Results of *Sty*I restriction enzyme digestion analysis of *PAH* gene exon 12. Line 1: Marker pUC19DNS/*Msp*I; line 2: control sample for homozygous R408W mutation; line 3: control sample for heterozygous R408W mutation; line 4: normal homozygous control; lines 5, 6 and 7 are PKU patient and his parents, respectively; line 8: undigested control for exon 12.

Thirty-eight (38) non-R408W chromosomes were screened for mutations through denaturing-gradient gel electrophoresis (DGGE) of the 13 exons of the *PAH* gene (Fig. 3.2).

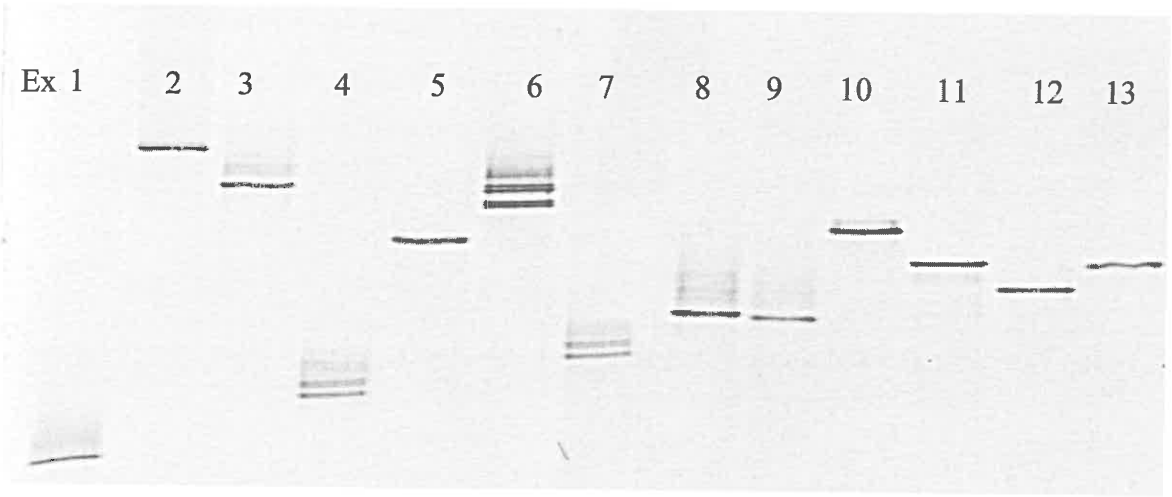


Fig. 3.2. Example for DGGE results: exons 4, 6 and 7 show heteroduplexes that indicate the presence of pathogen or silent mutation in the *PAH* gene.

Exons showing variant electrophoretic patterns were sequenced by fluorescent automated sequencing of *PAH* gene fragments by automated genetic analyzer ABI PRISM™ 310 and BigDye Terminator Sequencing protocol (Perkin Elmer Applied Biosystems).

Direct sequencing results of exon 1

Exon 1 did not show a variant electrophoretic pattern (heteroduplex) in denaturing-gradient gel electrophoresis, but the sequencing analysis of the whole *PAH* gene revealed a DNA sequence polymorphism -71A->C [c.-71A>C] in 5' UTR region on single chromosome (Fig. 3.3).

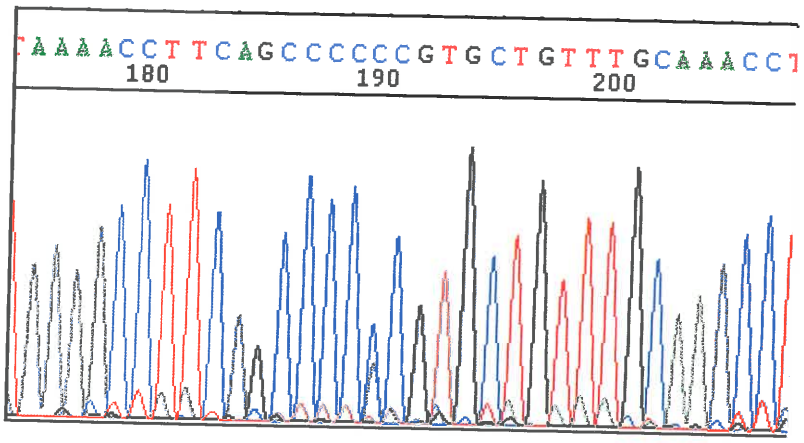


Fig. 3.3. Electropherogram of the *PAH* gene 5'UTR region fragment demonstrates DNA sequence polymorphism -71A->C [c.-71A>C].

Direct sequencing results of exon 2

DGGE analysis of *PAH* gene has shown a heteroduplex of exon 2 for 11 PKU patients. Disease-causing missense mutation L48S [c.143T>C] was found for one PKU patient (0.7%); 11 alleles showed the presence of polymorphism IVS2+19T->C [c.168+19T>C] in intron 2 (Fig. 3.4 (a) and (b)).

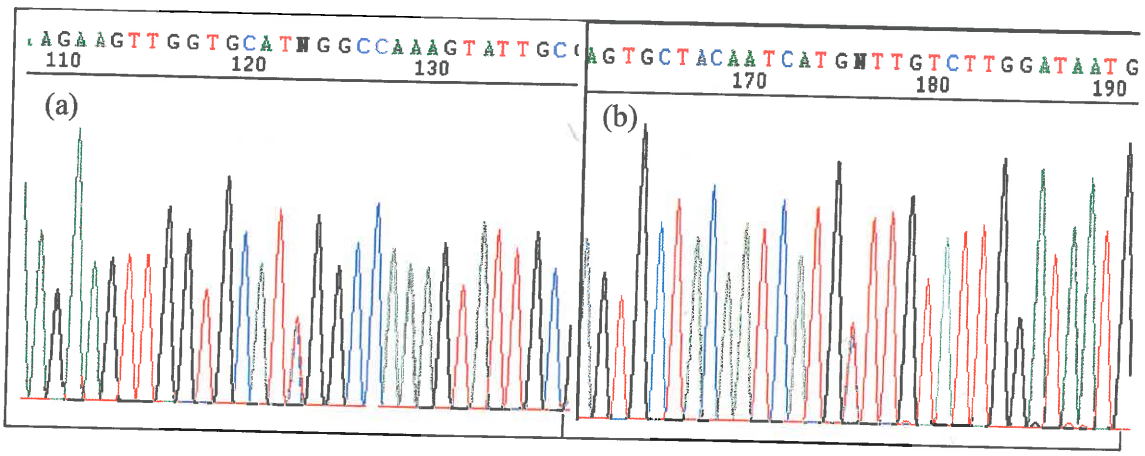


Fig. 3.4. Electropherogram of the *PAH* gene exon 2/intron 2 fragment demonstrates (a) missense mutation L48S [c.143T>C] and (b) DNA sequence polymorphism IVS2+19T->C [c.168+19T>C].

Direct sequencing results of exon 3

Two different mutations were found in exon 3: missense mutation A104D [c.311C>A] and nonsense mutation R111X [c.331C>T] (Fig. 3.5 (a) and (b)). Mutation A104D was identified in 2 PKU chromosomes, corresponding to a frequency of 1.4%; mutation R111X was found only once - 0.7%.

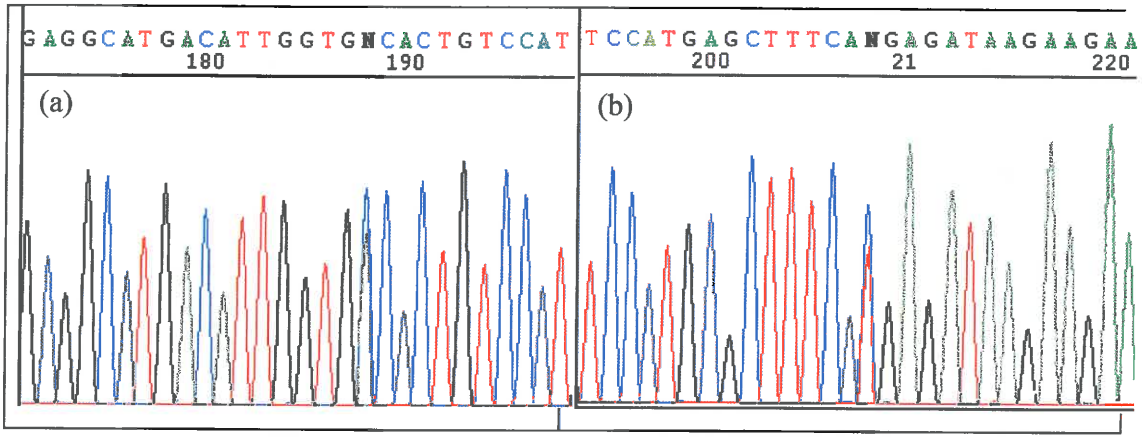


Fig. 3.5. Electropherogram of the *PAH* gene exon 3 fragment demonstrates the presence of two mutations (a) missense mutation A104D [c.311C>A] and (b) nonsense mutation R111X [c.331C>T].

Direct sequencing results of exon 4

No disease-causing mutations were found in exon 4. Two silent mutations for 12 PKU patients IVS3-22C->T [c.353-22C>T] and IVS4+47C->T in intron 3 and intron 4, respectively, were identified by sequencing analysis of exon 4 (Fig. 3.6 (a) and (b)). Silent mutation IVS3-22C->T was found in 6 alleles, IVS4+47C->T – in 8 alleles.

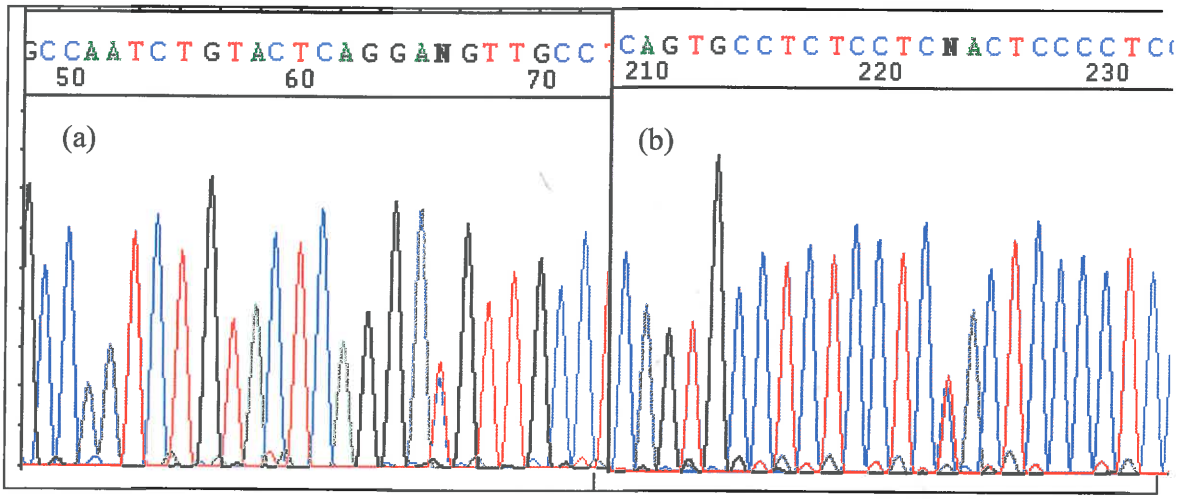


Fig. 3.6. Electropherogram of the *PAH* gene exon 4 fragment demonstrates the presence of two silent mutations (a) IVS3-22C->T [c.353-22C>T] and (b) IVS4+47C->T [c.441+47C>T] in the adjacent introns.

Direct sequencing results of exon 5

Only one missense mutation R158Q [c.473G>A] was found in exon 5 of *PAH* gene, corresponding to a frequency of 3% (Fig. 3.7). No silent mutations were detected in this exon.

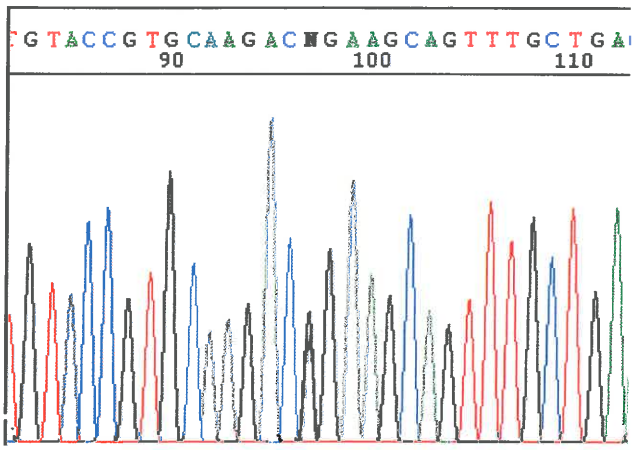
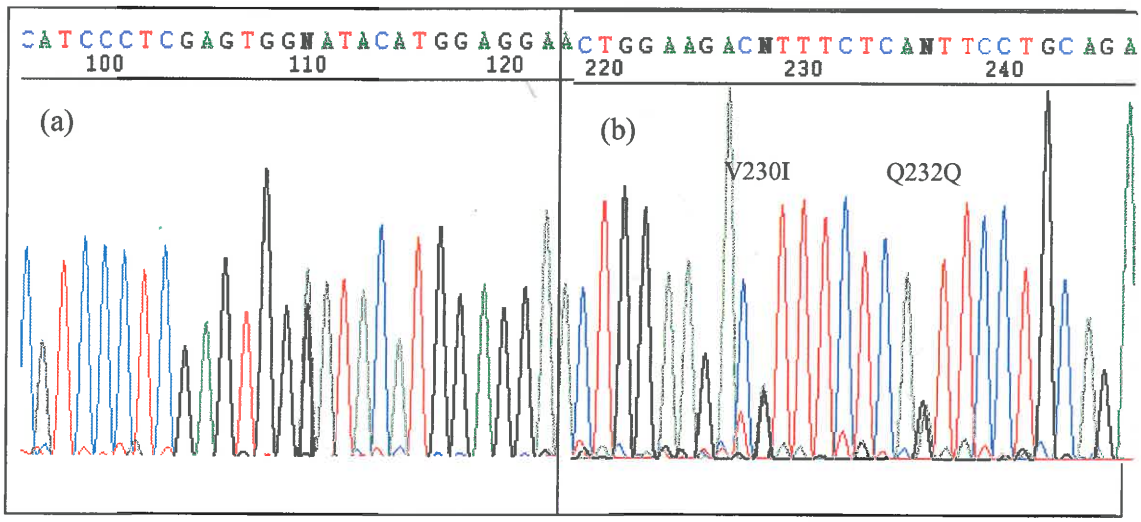


Fig. 3.7. Electropherogram of the *PAH* gene exon 5 fragment demonstrates the presence of mutation R158Q [c.473G>A].

Direct sequencing results of exon 6

Two missense mutations and one small deletion were found in exon 6: E178G [c.533A>G], V230I [c.688G>A] and E221_D222>Efs [c.663_664delAG] (Fig. 3.8 (a), (b) and (c)). Each of these three mutations was found only in one PKU chromosome and corresponded to a frequency of 0.7%.

Sequencing analysis of exon 6 also revealed 2 different silent mutations: Q232Q [c.696A>G] (Fig. 3.8 (b)) in 7 chromosomes and IVS5-54A->G [c.510-54A>G] located in intron 5 in one chromosome (Fig. 3.8 (d)).



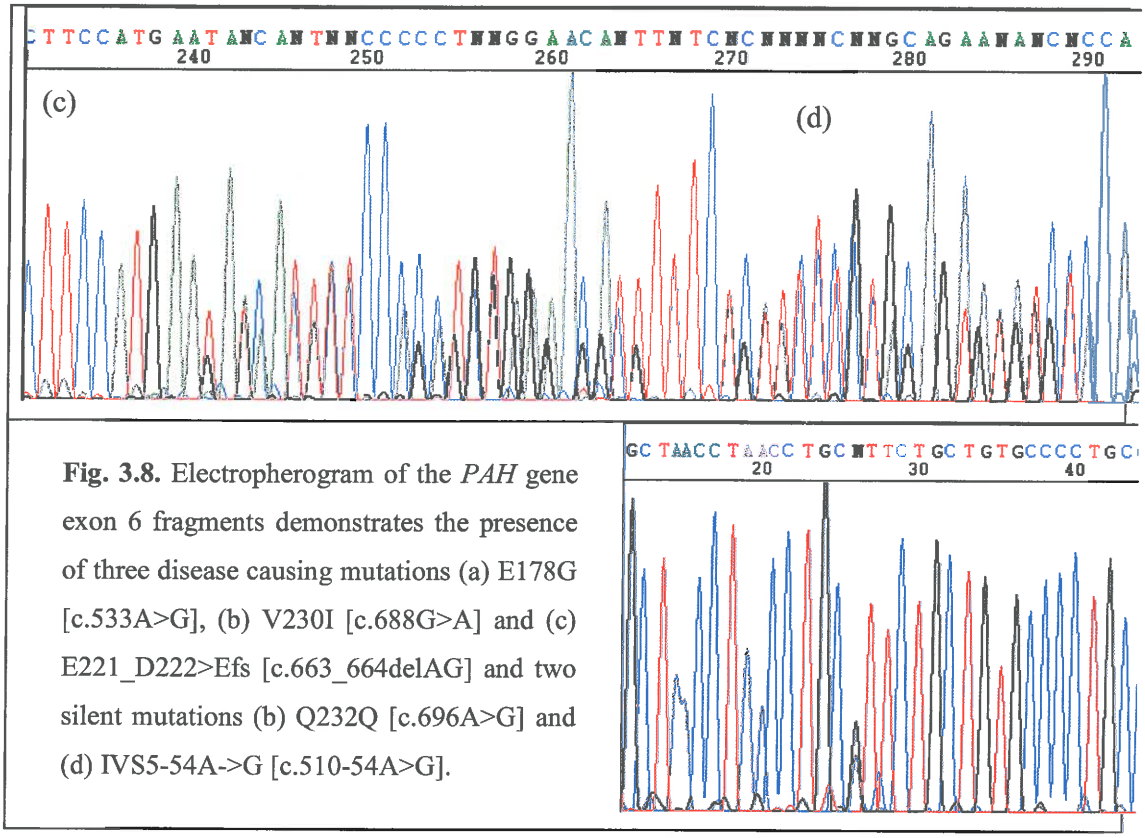


Fig. 3.8. Electropherogram of the *PAH* gene exon 6 fragments demonstrates the presence of three disease causing mutations (a) E178G [c.533A>G], (b) V230I [c.688G>A] and (c) E221_D222>Efs [c.663_664delAG] and two silent mutations (b) Q232Q [c.696A>G] and (d) IVS5-54A->G [c.510-54A>G].

Direct sequencing results of exon 7

The majority of PKU causing mutations (5/20) was found in exon 7: three missense mutations and two nonsense mutations that together corresponded to a frequency of 12.2%. Mutation E280K [c.838G>A] was the next common mutation after the R408W and accounted for 5.7% of mutant alleles. Mutations R261Q [c.782G>A] and P281L [c.842C>T] were identified in relative frequencies 3% and 2.1%, respectively (Fig. 3.9 (a), (b) and (c)).

Both nonsense mutations R261X [c.781C>T] and G272X [c.814G>T] were identified once with relative frequency of 0.7% each (Fig. 3.9 (d) and (e)).

Silent mutation V245V [c.735G>A] occurred in 6 chromosomes (Fig. 3.9 (f)).

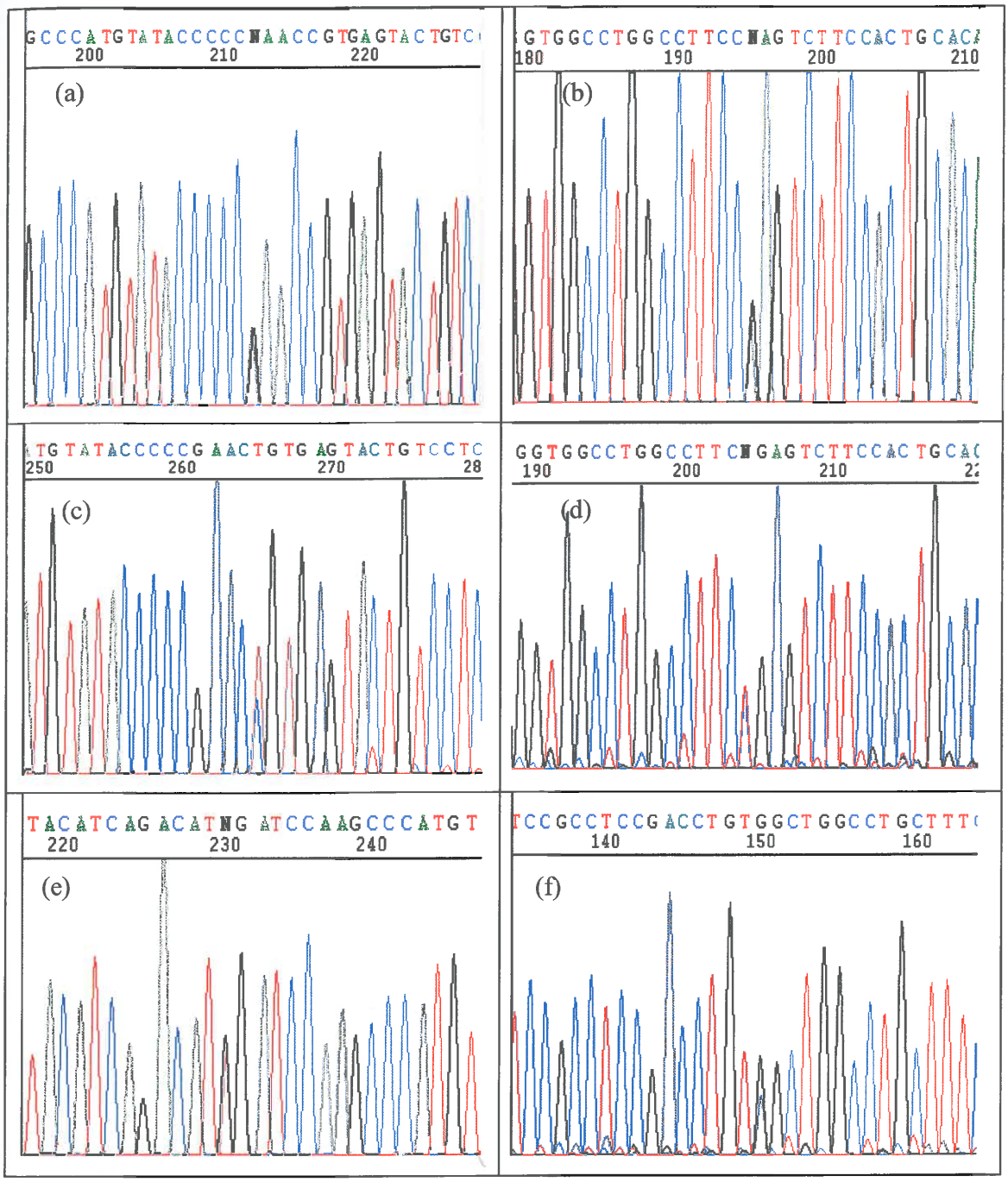


Fig. 3.9. Electropherogram of the *PAH* gene exon 7 fragments demonstrates the presence of 5 different disease causing mutations (a) E280K [c.838G>A], (b) R261Q [c.782G>A], (c) P281L [c.842C>T], (d) R261X [c.781C>T] and (e) G272X [c.814G>T]; (f) one silent mutation V245V [c.735G>A].

Direct sequencing results of exon 8

Exon 8 showed a heteroduplex in denaturing-gradient gel electrophoresis only for one PKU chromosome. Direct sequencing analysis revealed a novel nucleotide change that was not previously reported: P292T [c.874C>A] (Fig. 3.10).

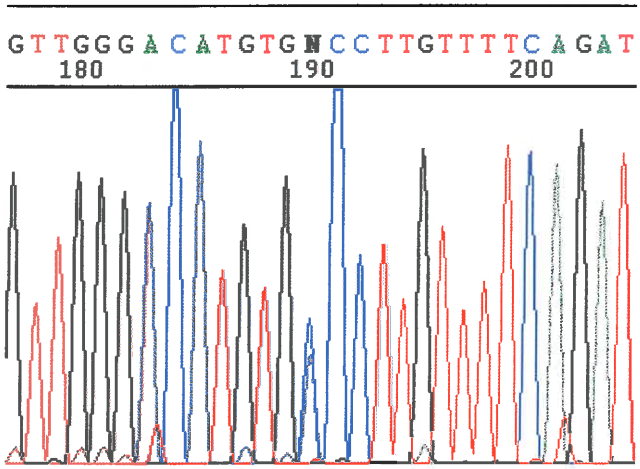


Fig. 3.10. Electropherogram of the *PAH* gene exon 8 fragment demonstrates the presence of a novel nucleotide change c.874C>A that results in proline change to treonine at 292 amino acid position.

This novel mutation was not examined by in vitro expression analysis. Using a *PolyPhen* (=Polymorphism Phenotyping) that is a tool which predicts possible impact of an amino acid substitution on the structure and function of a human protein using straightforward physical and comparative considerations, this variant was predicted to be *probably damaging* with PSIC (**P**osition-**S**pecific **I**ndependent **C**ounts) score difference 2.858. Closest contact with functional site: HIS 336D, distance 3.249 Å (http://genetics.bwh.harvard.edu/ggi/pph/7a5342f2e2a2a0c5e4db58a57d60ee753540c860/2605042.html).

Direct sequencing results of exon 9

One disease-causing missense mutation and one silent mutation were observed in exon 9. Mutation I306V [c.916A>G] was found in a single PKU chromosome, but polymorphism IVS9+ 43G>T [c.969+43G>T] in 4 alleles (Fig. 3.11 (a) and (b)).

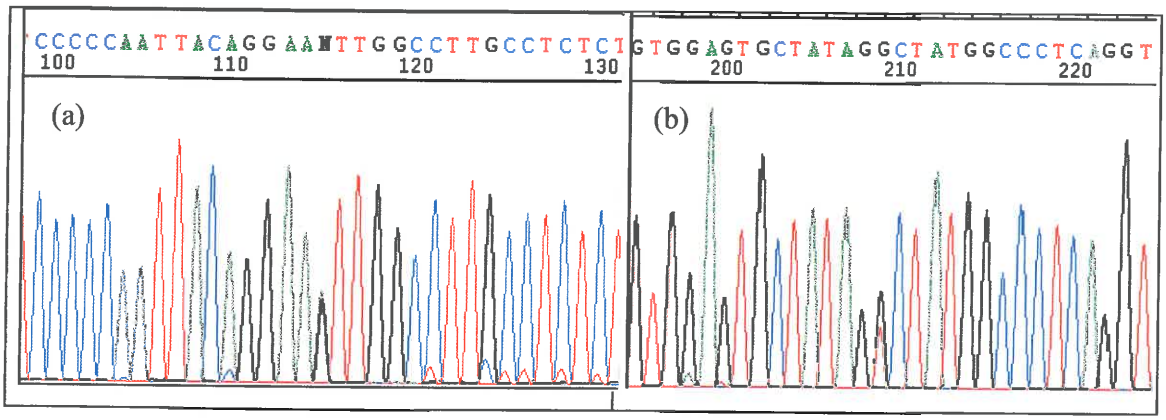


Fig. 3.11. Electropherogram of the *PAH* gene exon 9 fragments demonstrates the presence of (a) mutation I306V [c.916A>G] and (b) polymorphism IVS9+ 43G>T [c.969+43G>T].

Direct sequencing results of exon 10

Exon 10 did not show a variant electrophoretic pattern (heteroduplex) in denaturing-gradient gel electrophoresis and the sequencing analysis of this exon was not performed.

Direct sequencing results of exon 11

Direct sequencing analysis revealed one PKU causing splice site mutation IVS10-11G>A [c.1066-11G>A] in intron 10. Mutation was observed in 2 PKU chromosomes - 1.4%. The novel nucleotide change K371E [c.1111A>G] was found in exon 11 for one HPA patient (Fig. 3.12 (a) and (b)) who had slightly elevated Phe serum level and did not require treatment. In vitro expression analysis for this mutation was not performed, but it was predicted to be *benign* by *PolyPhen* analysis with PSIC score difference: 0.302; closest contact with other chains: PRO 366D, distance 4.896 Å (http://genetics.bwh.harvard.edu/ggi/pph/7a5342f2e2a2a0c5e4db58a57d60ee753540c860/2607880.html).

Silent mutation L385L [c.1155G>C] in exon 11 of *PAH* gene was observed only once (Fig. 3.12 (c)).

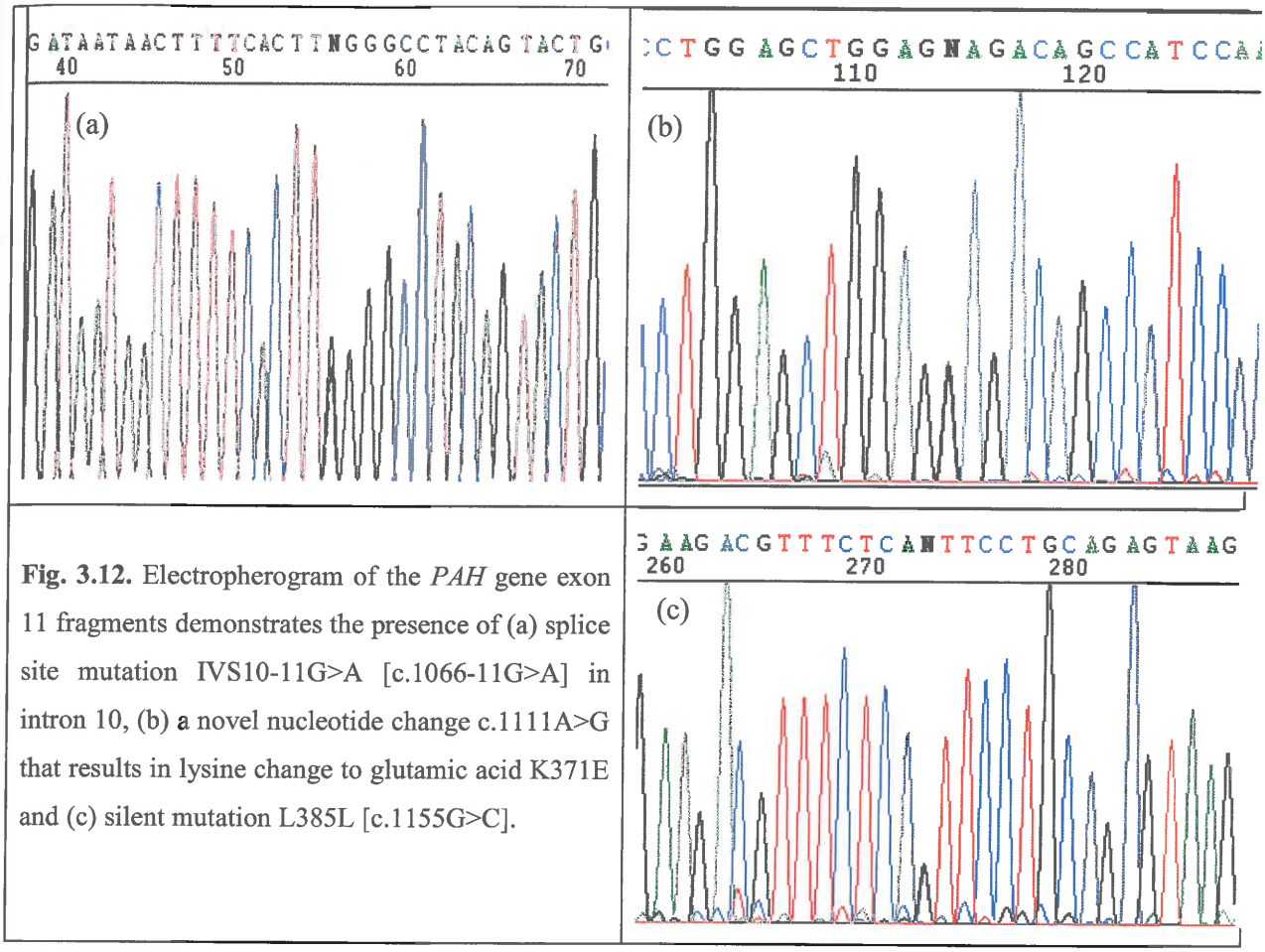
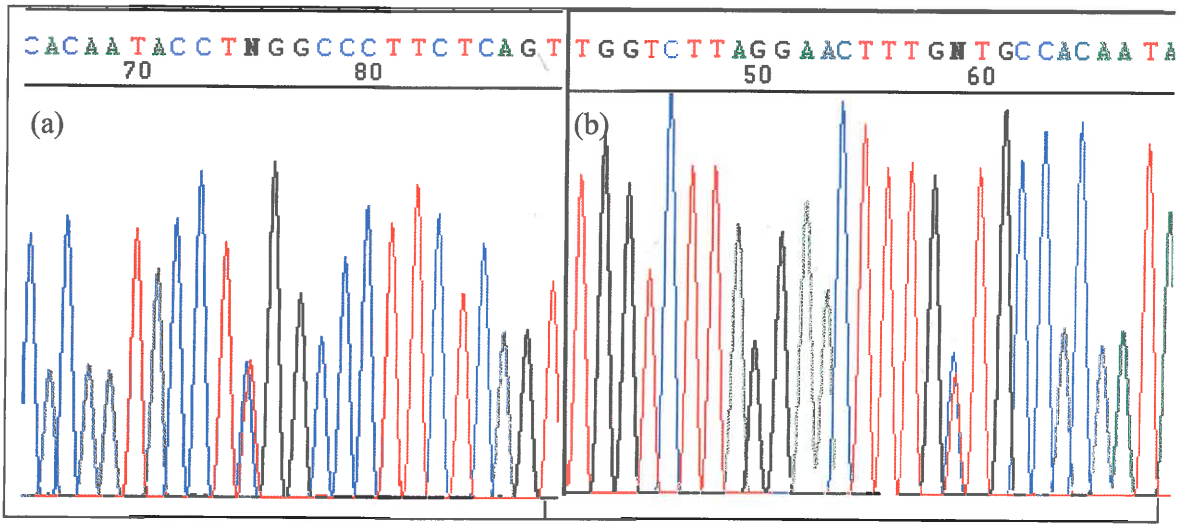


Fig. 3.12. Electropherogram of the *PAH* gene exon 11 fragments demonstrates the presence of (a) splice site mutation IVS10-11G>A [c.1066-11G>A] in intron 10, (b) a novel nucleotide change c.1111A>G that results in lysine change to glutamic acid K371E and (c) silent mutation L385L [c.1155G>C].

Direct sequencing results of exon 12

In addition to the most prevalent mutation R408W [c.1222C>T] in exon 12 mutations A403V [c.1208C>T] and IVS12+1G>A [c.1315+1G>A] were identified by direct sequencing (Fig. 3.13 (a), (b) and (c)). Each of these two mutations was found only once.



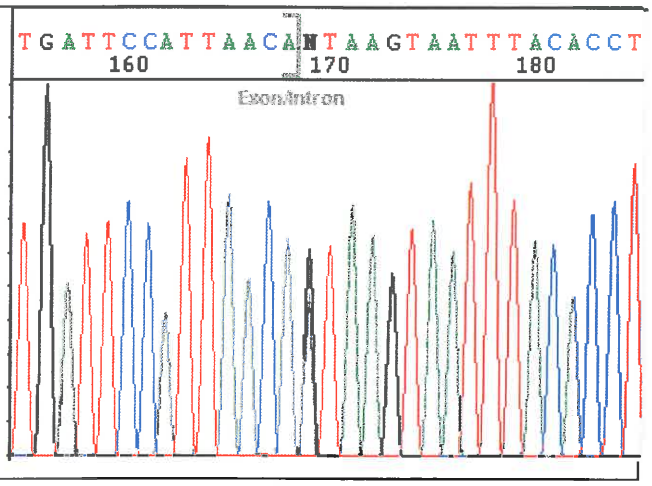


Fig. 3.13. Electropherogram of the *PAH* gene exon 12 fragments demonstrates the presence of (a) the most prevalent mutation R408W [c.1222C>T], (b) missense mutation A403V [c.1208C>T] and (c) splice site mutation IVS12+1G>A [c.1315+1G>A].

Direct sequencing results of exon 13

A novel splice site mutation IVS12-1G>A [c.1316-1G>A] on the border between intron 12 and exon 13 was found in two unrelated PKU chromosomes (1.4%) (Fig. 3.14). In vitro expression analysis to confirm the association of this mutation with PKU was not performed.

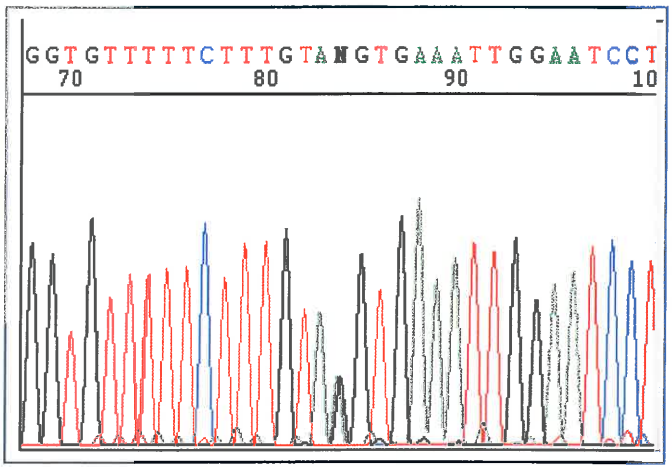


Fig. 3.14. Electropherogram of the *PAH* gene exon 13 fragment demonstrates the presence of a novel nucleotide change c.1316-1G>A that results in splice site mutation IVS12-1G>A.

In general, mutation analysis of 140 independent Latvian PKU chromosomes has revealed 20 different *PAH* gene mutations, representing a mutation detection rate 99%, one allele remained unknown. Thirteen were predicted missense mutations, 3 nonsense mutations, 3 splice site mutations, and one small deletion. PAH molecular lesions were identified in 8 different *PAH* gene exons (2, 3, 5, 6, 7, 8, 9, 11 and 12),

while none was found in exons 1, 4, 10, and 13. Two mutations were identified in intron 12 and one mutation in intron 10.

The most common mutation was R408W, accounted for 73% of mutant alleles (Fig. 3.15). Next most common mutation E280K presented on 5.7% of all PKU chromosomes. The frequency of the other five mutations (R261Q, R158Q, P281L, IVS10-11G>A and A104D) ranged from 1.4% to 3%.

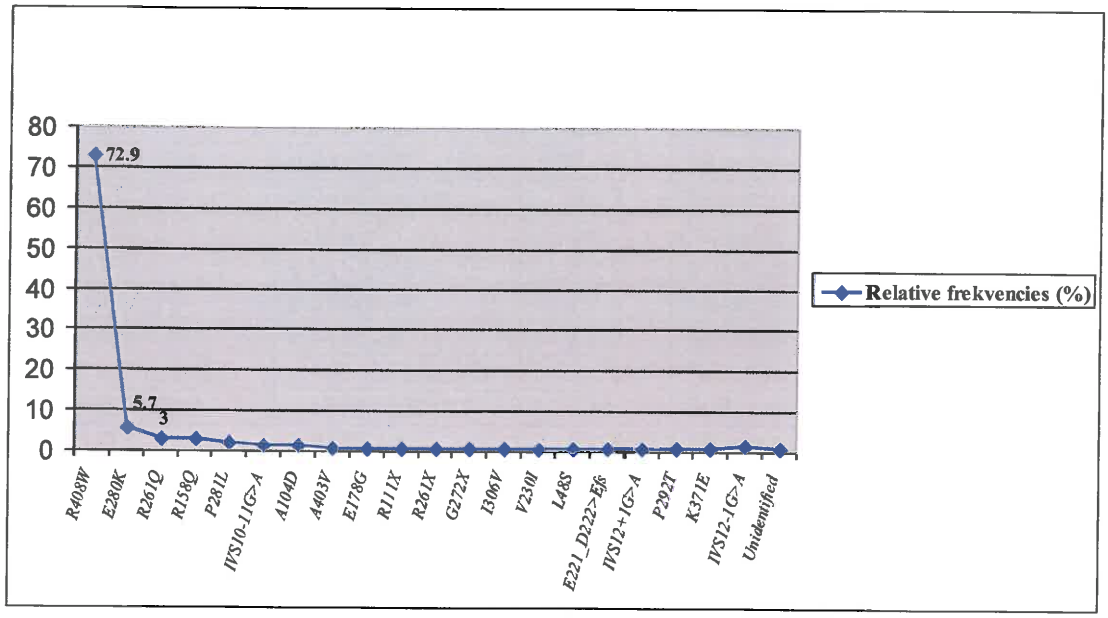


Fig 3.15. *PAH* gene mutation spectrum for Latvian PKU chromosomes.

Twelve mutations were identified on single chromosome only, corresponding to a frequency of 0.7%. Three mutations (P292T, K371E and IVS12-1G>A) had not been previously identified; two of them were found only once, but the third one was identified on two unrelated PKU chromosomes. None of these mutations have been examined by in vitro expression analysis. A novel mutation was assumed to be disease-causing when (1) it was either non-silent or a potential splicing mutation, (2) no other mutation was identified in the coding region of the *PAH* gene, (3) the allele was inherited from the parent who did not carry the other PKU mutation, and (4) the mutation had not been previously identified on normal or mutant chromosomes (Zschocke, 2003). To check the population specific polymorphisms for these novel single nucleotide changes control samples of 100 volunteered individuals without PKU were tested.

PAH gene mutations, their location and frequencies are summarised in table 3.2.

Table 3.2. Frequencies of *PAH* gene mutations in Latvian PKU patients.

No.	*Mutation Name*	*Syst. name*	*Location*	*Characters of mutation*	*No.*	*RF %*
1	R408W	c.1222C>T	Ex 12	Missense	102	72.9
2	E280K	c.838G>A	Ex 7	Missense	8	5.7
3	R261Q	c.782G>A	Ex 7	Missense	4	3.0
4	R158Q	c.473G>A	Ex 5	Missense	4	3.0
5	P281L	c.842C>T	Ex 7	Missense	3	2.1
6	IVS10-11G>A	c.1066-11G>A	I10	Splice site	2	1.4
7	A104D	c.311C>A	Ex 3	Missense	2	1.4
8	A403V	c.1208C>T	Ex 12	Missense	1	0.7
9	E178G	c.533A>G	Ex 6	Missense	1	0.7
10	R111X	c.331C>T	Ex 3	Nonsense	1	0.7
11	R261X	c.781C>T	Ex 7	Nonsense	1	0.7
12	G272X	c.814G>T	Ex 7	Nonsense	1	0.7
13	I306V	c.916A>G	Ex 9	Missense	1	0.7
14	V230I	c.688G>A	Ex 6	Missense	1	0.7
15	L48S	c.143T>C	Ex 2	Missense	1	0.7
16	E221_D222>Efs	c.663_664delAG	Ex 6	Deletion	1	0.7
17	IVS12+1G>A	c.1315+1G>A	I12	Splice site	1	0.7
18	**P292T***	**c.874C>A**	Ex 8	Missense	1	0.7
19	**K371E***	**c.1111A>G**	Ex 11	Missense	1	0.7
20	**IVS12-1G>A***	**c.1316-1G>A**	I12	Splice site	2	1.4
21	Unidentified	-	-	-	1	0.7
Total					**140**	**100**

*Novel mutations are marked in bold

The sensitivity of a used mutation detection tests was calculated as the proportion of alleles with identified *PAH* mutation that test positive for it. This can also be written as:

$$\text{sensitivity} = \frac{\text{number of true positives}}{\text{number of true positives} + \text{number of false negatives}}$$

Using these calculations the sensitivity of approach used for mutation detection in *PAH* gene was detected as 99.28%.

3.2 GENOTYPES IDENTIFIED IN LATVIAN PKU PATIENTS

The most prevalent genotype among Latvian PKU patients was R408W/R408W. Thirty-six (51.4%) of 70 unrelated characterised PKU patients were homozygous for R408W, the remaining 34 patients were compound heterozygous. The homozygosity value (*j*) for the PKU population of Latvia is 0.514.

The majority of compound heterozygote PKU patients had R408W mutation in one allele (42.9%); only four patients (5.7%) had no R408W mutations in their chromosomes (Fig. 3.16).

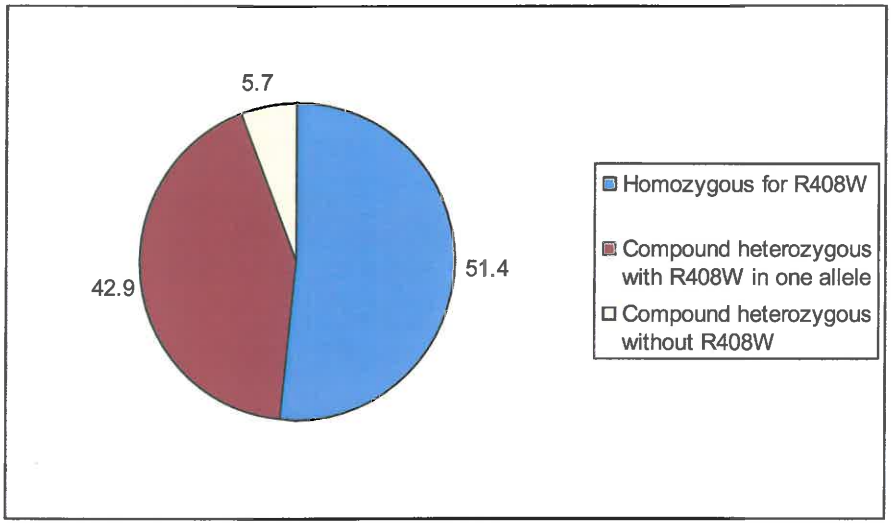


Fig. 3.16. The distribution of genotypes for Latvian PKU patients (%).

The full list of indentified genotypes for Latvian PKU patients and frequency for each genotype are shown on Fig. 3.17.

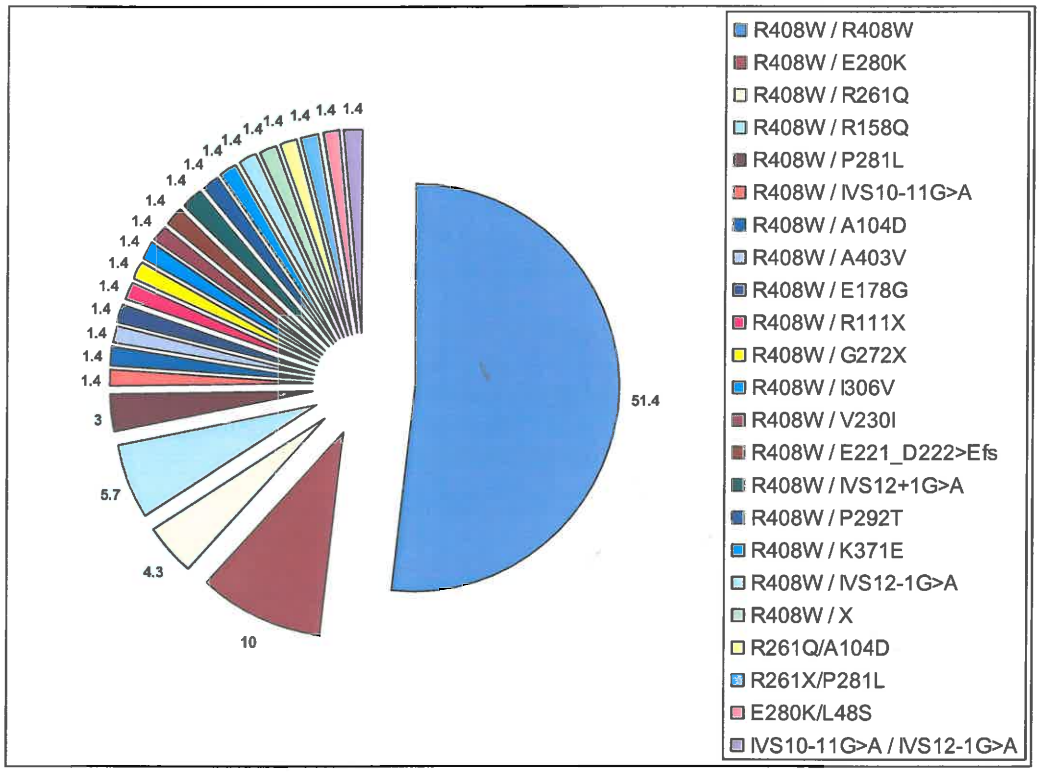


Fig. 3.17. Relative frequencies of genotypes identified in Latvian PKU patients.

Hardy-Weinberg equilibrium was used to determine whether the population is balanced. We used the R408W mutation number (n=102) obtained from our study to calculate the absolute frequencies for PKU genotypes.

Allele frequencies in population:

	A (p)	a (q)	Total
Frequency	0.729	0.271	1
Number	102	38	140

Genotype frequencies in population:

	AA (p^2)	Aa (2pq)	aa (q^2)	Total
Frequency	0.531	0.396	0.074	1
Number	37	28	5	70

According to Hardy-Weinberg equilibrium the absolute frequency for R408W homozygote patients has to be 53.1% and 39.6% for compound heterozygous patients with R408W mutation. Our study did not reveal the presence of other homozygous patients in PKU population that according to Hardy-Weinberg equilibrium has to be up to 1%.

To estimate the difference between the PKU genotypes' absolute frequencies and current relative frequencies in Latvian PKU population we regarded PKU patients without R408W mutation as *aa* homozygous to make calculations (Table 3.3).

Table 3.3. Data used for Chi-square test: AA homozygous - R408W/R408W homozygous patients; Aa heterozygous - R408W/X heterozygous patients; aa homozygous - patients without R408W mutation.

Row #	Category	Observed	Expected #	Expected
1	AA homozygous	36	37	52.857%
2	Aa heterozygous	30	28	40.000%
3	aa homozygous	4	5	7.143%

P value and statistical significance:

Chi squared equals 0.370 with 2 degrees of freedom.

The two-tailed P value equals 0.8312.

By conventional criteria, this difference is considered to be not statistically significant, as we expected. These data means, that the Latvian PKU patients pool is balanced.

In that way our results suggest that for the Latvian population, the allele and genotype frequencies at the *PAH* locus are in Hardy-Weinberg equilibrium. In other words, we can expect these allele frequencies to remain constant over time (barring any specific evolutionary forces acting upon this locus), thus ensuring genetic variation in the population at the *PAH* locus.

3.2.1. Genotype-phenotype correlations in Latvian PKU patients

On the basis of individual data on Phe tolerance and pre-treatment serum Phe levels, the patients were assigned to one of four generally accepted phenotype categories: 64 (91.4%) had severe PKU, four (5.7%) had mild PKU and two (2.9%) had MHP. No one patient was classified as having moderate PKU clinical form. Genotype and phenotype correlation in PKU patients is shown in Table 3.4.

Table 3.4. Observed correlation between *PAH* genotype and phenotype in Latvian PKU patients.

No.	PAH locus genotype	No. of cases	RF%	Clinical phenotype
1	R408W / R408W	36	51,4	Severe PKU
2	R408W / E280K	7	10,0	Severe PKU
3	R408W / R261Q	3	4,3	Severe PKU
4	R408W / R158Q	4	5,7	Severe PKU
5	R408W / P281L	2	3,0	Severe PKU
6	R408W / IVS10-11G>A	1	1,4	Severe PKU
7	R408W / A104D	1	1,4	Severe PKU
8	R408W / R111X	1	1,4	Severe PKU
9	R408W / E221_D222>Efs	1	1,4	Severe PKU
10	R408W / IVS12+1G>A	1	1,4	Severe PKU
11	R408W / E178G	1	1,4	Mild PKU
12	R408W / A403V	1	1,4	Mild PKU
13	R408W / G272X	1	1,4	Mild PKU
14	R408W / I306V	1	1,4	Mild PKU
15	R408W / V230I	1	1,4	MHP
16	R408W / **K371E**	1	1,4	MHP
17	R408W / **P292T**	1	1,4	Severe PKU
18	R408W / **IVS12-1G>A**	1	1,4	Severe PKU
19	R408W / X	1	1,4	Severe PKU
20	R261Q/A104D	1	1,4	Severe PKU
21	R261X/P281L	1	1,4	Severe PKU
22	E280K/L48S	1	1,4	Severe PKU
23	IVS10-11G>A / **IVS12-1G>A**	1	1,4	Severe PKU
	Total	**70**	**100**	

According to multicenter study (Guldberg, 1998) 10 of 17 known *PAH* gene mutations observed in Latvian PKU patients were classified as severe or classic-PKU mutations, one as moderate-PKU mutation, two mild-PKU and 4 as mutations causing MHP. In compliance with Guldberg *et al.,* 1998 proposed model for phenotypic effect of two mutant *PAH* alleles, expressed as the sum of their assigned values (AVs), we calculated the phenotype for Latvian PKU patients using the each mutation assigned value (Table 3.5).

Table 3.5. Genotype and phenotype correlation in Latvian PKU patients according to Guldberg model.

No.	PAH locus genotype	AV_1+AV_2	AVs sum	Clinical phenotype
1	R408W / R408W	1+1	2	Severe PKU
2	R408W / E280K	1+1	2	Severe PKU
3	R408W / R261Q	1+2	3	Moderate PKU
4	R408W / R158Q	1+1	2	Severe PKU
5	R408W / P281L	1+1	2	Severe PKU
6	R408W / IVS10-11G>A	1+1	2	Severe PKU
7	R261X/P281L	1+1	2	Severe PKU
8	R408W / R111X	1+1	2	Severe PKU
9	R408W / E221_D222>Efs	1+1	2	Severe PKU
10	R408W / IVS12+1G>A	1+1	2	Severe PKU
11	R408W / **IVS12-1G>A***	1+1	2	Severe PKU
12	IVS10-11G>A / **IVS12-1G>A**	1+1	2	Severe PKU
13	R408W / G272X	1+1	2	Severe PKU
14	R408W / A104D	1+4	5	Mild PKU
15	R261Q/A104D	2+4	6	Mild PKU
16	E280K/L48S	1+4	5	Mild PKU
17	R408W / I306V	1+8	9	MHP
18	R408W / V230I	1+8	9	MHP
19	R408W / E178G	1+8	9	MHP
20	R408W / A403V	1+8	9	MHP

* we assume a novel splice site mutation as severe (classic)

In comparison with observed correlation between *PAH* genotype and phenotype in our study, some deviations took place using the Guldberg model. Mostly these deviations occurred in classification of severe PKU and MHP (Fig. 3.18). It shows that determining the severity of the disease not only mutation characteristics but also the individual characteristics of the organism have to be taken into account.

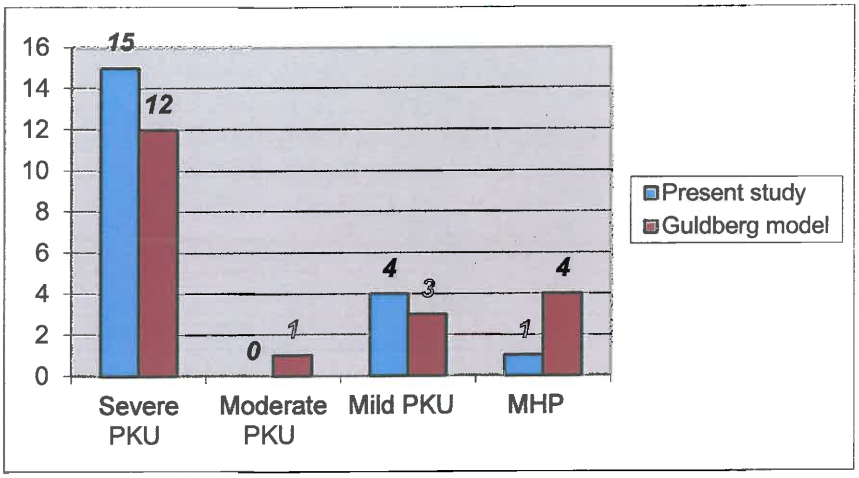


Fig. 3.18. Comparison of genotype and phenotype correlation in Latvian PKU patients using different classification models.

Chi-squared test was used to estimate the difference between two variants of the PKU classification (Table 3.6).

Table 3.6. Data used for Chi-squared test calculations.

Row #	Category	Observed	Expected
1	Severe PKU	15	12
2	Moderate PKU	0	1
3	Mild PKU	4	3
4	MHP	1	4

P value and statistical significance:

Chi squared equals 4.333 with 3 degrees of freedom. The two-tailed P value equals 0.2276. This difference is considered to be not statistically significant. It could be explained by the prevalence of severe mutations in Latvian PKU chromosomes and relatively small number of mutations causing variant PKU phenotypes.

3.2.2 Silent mutations in the *PAH* gene

Nine different silent mutations were found in Latvian PKU patients' *PAH* chromosomes in different combinations (Table 3.7). The most frequent polymorphisms were IVS2+19T->C (found in 11 patients) and IVS4+47C->T (found in 8 patients). Silent mutations were determined in 18 compound heterozygous patients by direct

sequencing analysis of *PAH* gene exons that formed heteroduplexes during DGGE. All silent mutations were found in heterozygote condition.

Table 3.7. Silent mutations found in Latvian PKU patients.

No.	Genotype	SNPs	Location
1	R408W/E221_D222>Efs	IVS3-22C->T IVS4+47C->T V245V	Ex 4 Ex 4 Ex 7
2	R408W/E280K	IVS2+19T->C	I2/Ex 2
3	R408W/E280K	IVS2+19T->C	Ex 2
4	R408W/IVS12+1G>A	Q232Q	Ex 6
5	R111X/R408W	IVS2+19T->C	Ex 2
6	R408W/E280K	IVS2+19T->C IVS4+47C->T	Ex 2 Ex 4
7	R408W/E280K	IVS2+19T->C IVS4+47C->T	Ex 2 Ex 4
8	R408W/IVS10-11G>A	IVS2+19T->C	Ex 2
9	R408W/E280K	IVS2+19T->C IVS4+47C->T	Ex 2 Ex 4
10	R408W/E280K	IVS2+19T->C	Ex 2
11	IVS10-11G>A/IVS12-1G>A	IVS2+19T->C IVS3-22C->T IVS5-54A->G	Ex 2 Ex 4 Ex 6
12	R408W/I306V	IVS3-22C->T IVS4+47C->T Q232Q V245V IVS9+43G->T	Ex 4 Ex 4 Ex 6 Ex 7 Ex 9
13	R408W/V230I	IVS3-22C->T Q232Q V245V	Ex 4 Ex 6 Ex 7
14	R408W/K371E	IVS4+47C->T Q232Q V245V IVS9+43G->T	Ex 4 Ex 6 Ex 7 Ex 9
15	E280K/L48S	IVS2+19T->C IVS3-22C->T Q232Q V245V	Ex 2 Ex 4 Ex 6 Ex 7
16	R408W/P292T	-71A->C IVS4+47C->T Q232Q IVS9+43G->T L385L	5'UTR Ex 4 Ex 6 Ex 9 Ex 11
17	R408W/E280K	IVS2+19T->C IVS4+47C->T	Ex 2 Ex 4
18	X/R408W	IVS3-22C->T Q232Q V245V IVS9+ 43G>T	Ex 4 Ex 6 Ex 7 Ex 9

3.3 ANALYSIS OF *PAH* GENE MUTATIONS' MINIHAPLOTYPES

Results of VNTR and STR systems analysis were used to form *PAH* gene mutations' minihaplotypes (Fig. 3.19 and 3.20). *PAH* minihaplotypes for the mutant chromosomes have been identified in 34 compound heterozygote patients when parents were available and two homozygote patients for mutation R408W.

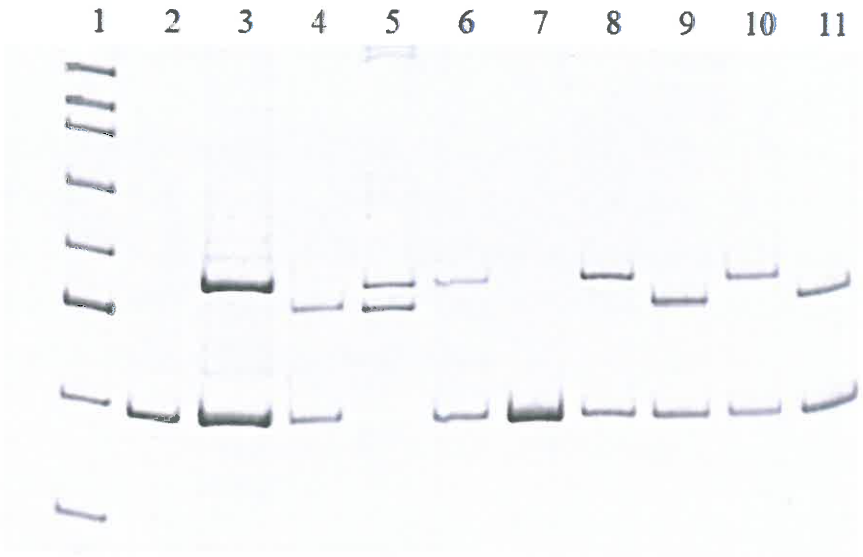
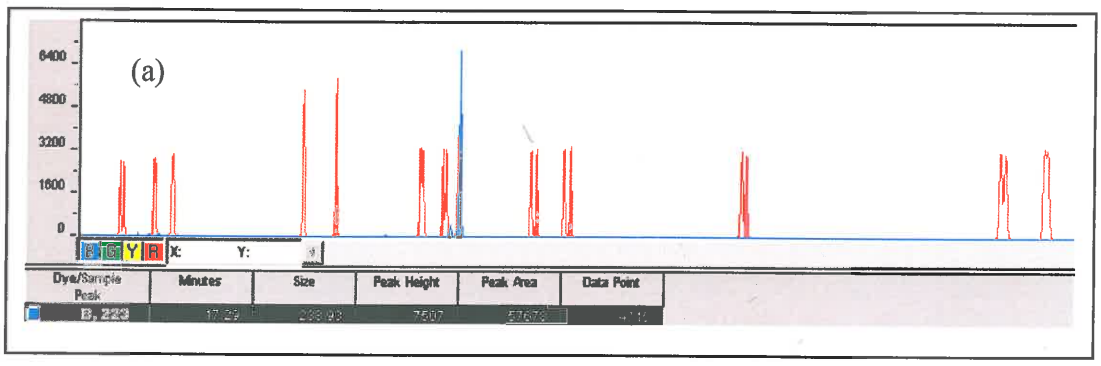


Fig. 3.19. VNTR analysis results for two PKU families. Line 1: Marker pUC19DNS/*Msp*I; lines 3, 5, 9 and 11 are controls for VNTR alleles 3/8, 7/8, 3/7 and 3/7, respectively; lines 2, 4, 6 and 7, 8, 10 are samples of two PKU families with the following VNTR alleles 3/3, 3/7, 3/8 and 3/3, 3/8, 3/8, respectively.



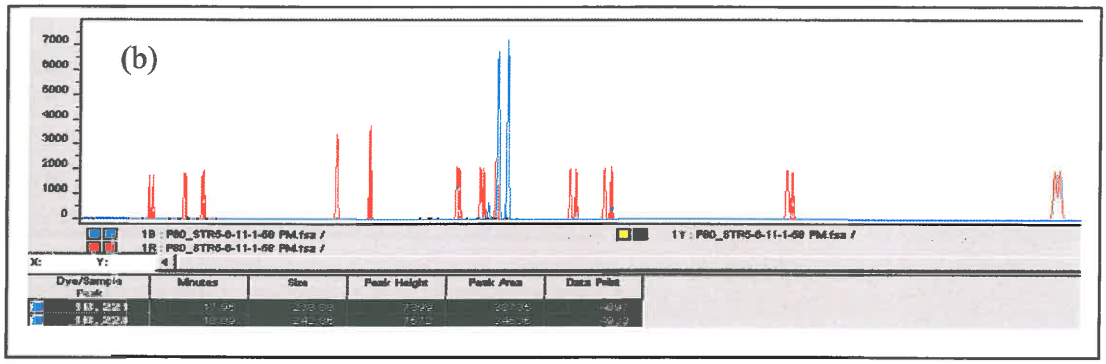


Fig. 3.20. Results of STR system analysis for (a) patient homozygous for 238 STR allele; (b) patient heterozygous for 238 and 242 STR alleles.

Minihaplotypes association with PKU mutations is reported in Table 3.8. Sixteen different *PAH* gene minihaplotypes have been identified associated with Latvian PKU chromosomes. The most frequent minihaplotype was 3/238 that associated with the mutation R408W that is typical for Eastern-European populations and is spread across Europe from the northeast to the southwest.

Table 3.8. Association between VNTR/STR minihaplotypes and PKU causing mutations in the Latvian population.

No.	Mutation	Total alleles investigated	Minihaplotype	Alleles
1	R408W	34	3/238 3/242 3/234 8/238	28 3 2 1
2	E280K	8	9/250 9/246	7 1
3	R261Q	4	3/238 8/238	2 2
4	R158Q	4	3/238 3/234 7/234	2 1 1
5	P281L	3	7/242 8/242	2 1
6	IVS10-11G>A	2	7/250	2
7	A104D	2	8/242	2
8	A403V	1	8/246	1
9	E178G	1	7/242	1
10	R111X	1	8/250	1
11	R261X	1	7/238	1
12	G272X	1	9/234	1
13	I306V	1	3/234	1
14	V230I	1	3/246	1

Table 3.8 (continued)

15	L48S	1	3/234	**1**
16	E221_D222>Efs	1	3/242	**1**
17	IVS12+1G>A	1	8/242	**1**
18	**P292T**	1	8/226	**1**
19	**K371E**	1	3/238	**1**
20	**IVS12-1G>A**	2	7/242	**2**
21	Unidentified	1	3/234	**1**
	Total	**72**		**72**

Among the sixteen minihaplotypes associated to specific *PAH* mutations five were associated to more than one mutation (Table 3.9). On the other side, more common PKU mutations, including the most common mutation R408W and mutations E280K, R261Q, R158Q and P281L, were associated to more than one minihaplotype that could be result of different origin of the mutations (Table 3.8).

Table 3.9. *PAH* minihaplotypes associated to more than one mutation

No.	*Minihaplotype*	*Mutation*	*Alleles*
1	3/238	R408W R261Q R158Q K371E	28 2 2 1
2	3/242	R408W E221_D222>Efs	3 1
3	3/234	R408W R158Q I306V L48S Unidentified	2 1 1 1 1
4	7/242	P281L E178G IVS12-1G>A	2 1 2
5	8/238	R408W R261Q	1 2
6	8/242	A104D P281L IVS12+1G>A	2 1 1

Minihaplotypes for 61 normal *PAH* alleles were obtained from the patients' parents. Most of the minihaplotypes were present on both normal and mutant *PAH* chromosomes. A continuous distribution from the smallest STR allele (226bp) to the largest (250bp) was observed on both allele types. Alleles 254bp and 258bp were absent in our study. The largest VNTR allele 12 was found only on normal *PAH* chromosome.

The most common minihaplotype among mutant alleles was 3/238 due to the high prevalence of R408W mutation. Minihaplotype 3/242 was the most prevalent among normal chromosomes (Table 3.10).

The average probability of heterozygosity for minihaplotypes was about 76% for mutant and 92% for normal chromosomes.

Table 3.10. Frequencies of minihaplotypes on normal and mutant *PAH* chromosomes.

No.	Minihaplotype	Frequency		p_i^2	
		Mutant	Normal	Mutant	Normal
1	3/234	0.0833	0.0492	0.00693	0.00242
2	3/238	0.4583	0.1147	0.21	0.013156
3	3/242	0.0556	0.1475	0.0031	0.02176
4	3/246	0.0139	0.0164	0.00019	0.00027
5	7/230	—	0.0164	—	0.00027
6	7/234	0.0139	0.0327	0.00019	0.00107
7	7/238	0.0139	0.0164	0.00019	0.00027
8	7/242	0.0694	0.0492	0.0048	0.00242
9	7/246	—	0.0492	—	0.00242
10	7/250	0.0278	—	0.00077	—
11	8/226	0.0139	0.0164	0.00019	0.00027
12	8/230	—	0.082	—	0.0067
13	8/234	—	0.0492	—	0.00242
14	8/238	0.0416	0.0984	0.00173	0.00968
15	8/242	0.0556	0.082	0.0031	0.0067
16	8/246	0.0139	0.0655	0.00019	0.00429
17	8/250	0.0139	0.0164	0.00019	0.00027
18	9/234	0.0139	0.0164	0.00019	0.00027
19	9/246	0.0139	0.0164	0.00019	0.00027
20	9/250	0.0972	0.0492	0.00945	0.00242
21	12/230	—	0.0164	—	0.00027
	Total	**1.0000**	**1.0000**	$\Sigma_m = 0.2414$	$\Sigma_n = 0.077616$
	1 - Σ			**0.7586**	**0.9224**

There was a statistically significant difference observed between normal and mutant alleles in the distribution of minihaplotype 3/238 and 8/230 (Fig. 3.21 (a) and (b)).

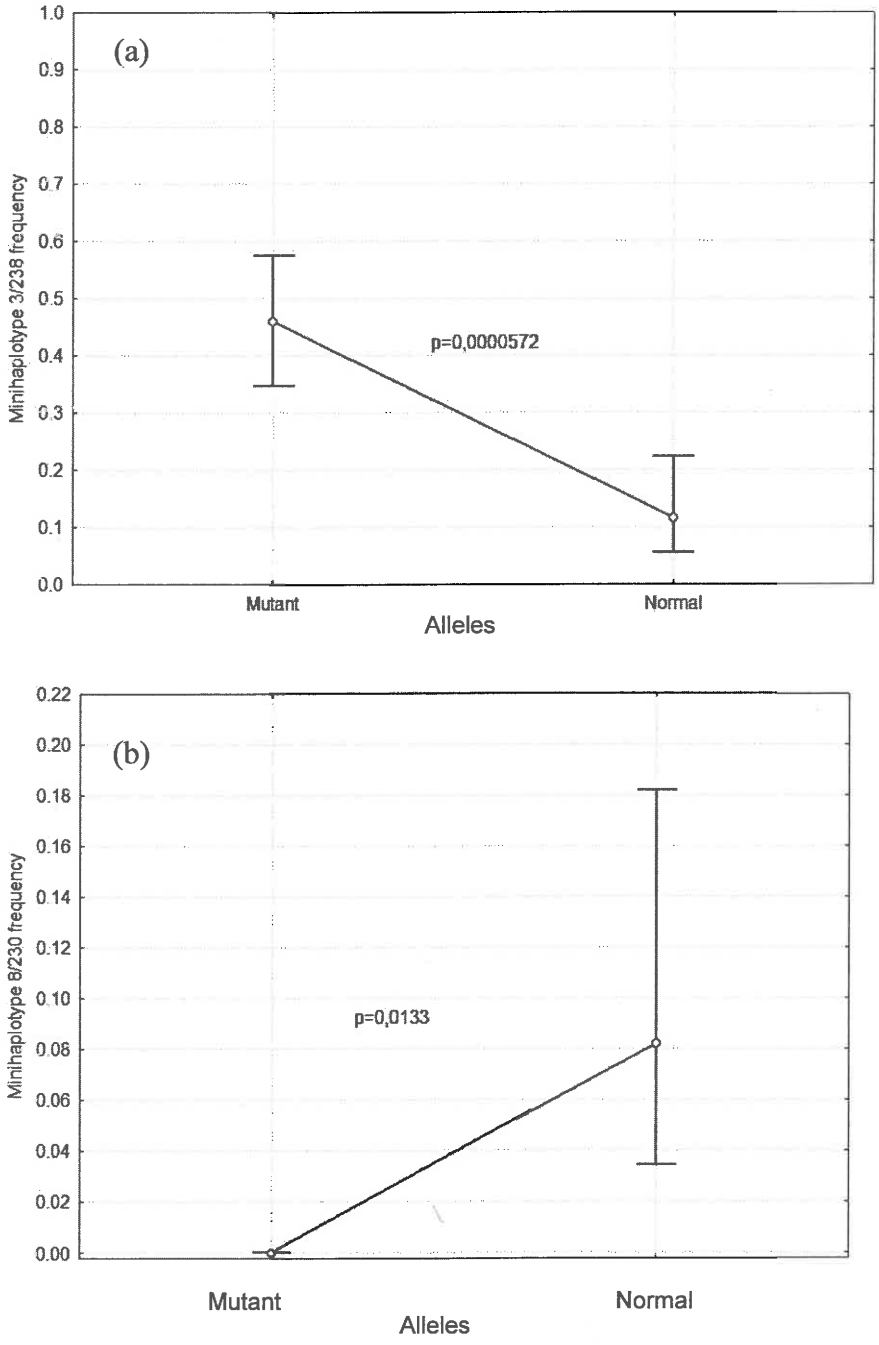


Fig. 3.21. Results of statistical analysis for distribution of minihaplotypes 3/238 (a) and 8/230 (b) between normal and mutant alleles.

Statistical analysis showed the tendency to statistically significant difference ($p<0.10$) between normal and mutant alleles in the distribution of minihaplotypes 3/242, 7/246 and 8/234 (Appendix VIII). There was no statistically significant difference overall in the relative frequencies of other minihaplotypes.

3.4 ANALYSIS OF THE DISTRIBUTION OF *PAH* GENE MUTATION R408W

Distribution of *PAH* gene mutation R408W was similar in Latvia, Lithuania and St. Petersburg region. Estonia keeps the highest number of *PAH* chromosomes harbouring mutation R408W (Fig. 3.22). Statistical analysis comparing the relative frequencies of mutation R408W overall among the different geographical regions showed this to be highly significant for Latvia and other European countries (Table 3.11). Results of statistical analysis confirm the Balto-Slavic origin of mutation R408W and introduction of this mutation to other European populations by people migrations (Fig. 3.23).

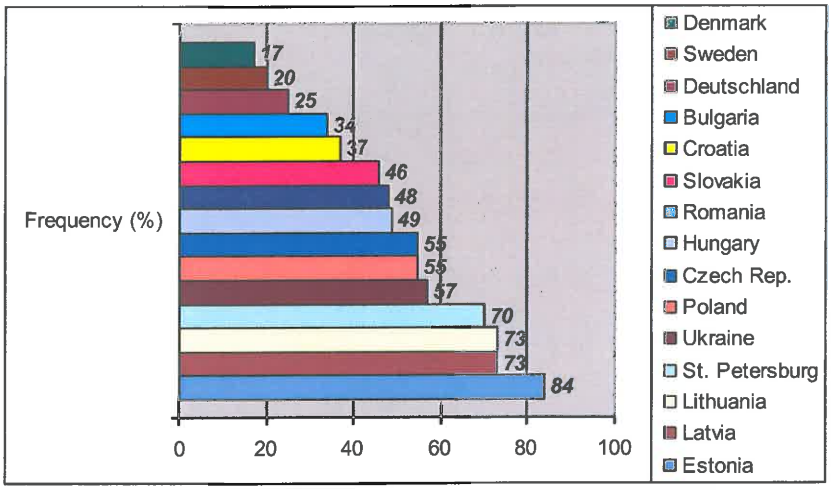


Fig. 3.22. Geographic distribution of the mutation R408W within different European populations.

Table 3.11. Comparison between the frequency of the mutation R408W in Latvian PKU population and other European populations.

Country	Frequency (%)	Alleles investigated	References	$P^{\#}$
Estonia	84	68	Lilleväli, 1996	0.0836
Latvia	73	140	Present study	*
Lithuania	73	184	Kasnauskiene, 2003	—
St. Petersburg	70	140	Baranovskaya, 1996	—
Ukraine	57	202	Nechyporenko and Livshits, 2002	0.00283
Poland	55	182	Zschocke, 2003	0.00109
Czech Rep.	55	266	Kozak, 1997	< 0.001
Hungary	49	70	Zschocke, 2003	< 0.001
Romania	48	44	Zschocke, 2003	< 0.001

Table 3.11 (continued)

Slovakia	46	98	Kadasi, 1995	< 0.001
Croatia	37	79	Zschocke, 2003b	< 0.001
Bulgaria	34	60	Zschocke, 2003	< 0.001
Deutschland	25	438	Zschocke, 1999	< 0.001
Sweden	20	176	Svensson, 1993	< 0.001
Denmark	17	308	Guldberg, 1993a	< 0.001

[#] the *p* value is for the comparison of the relative frequencies between two geographical groups.

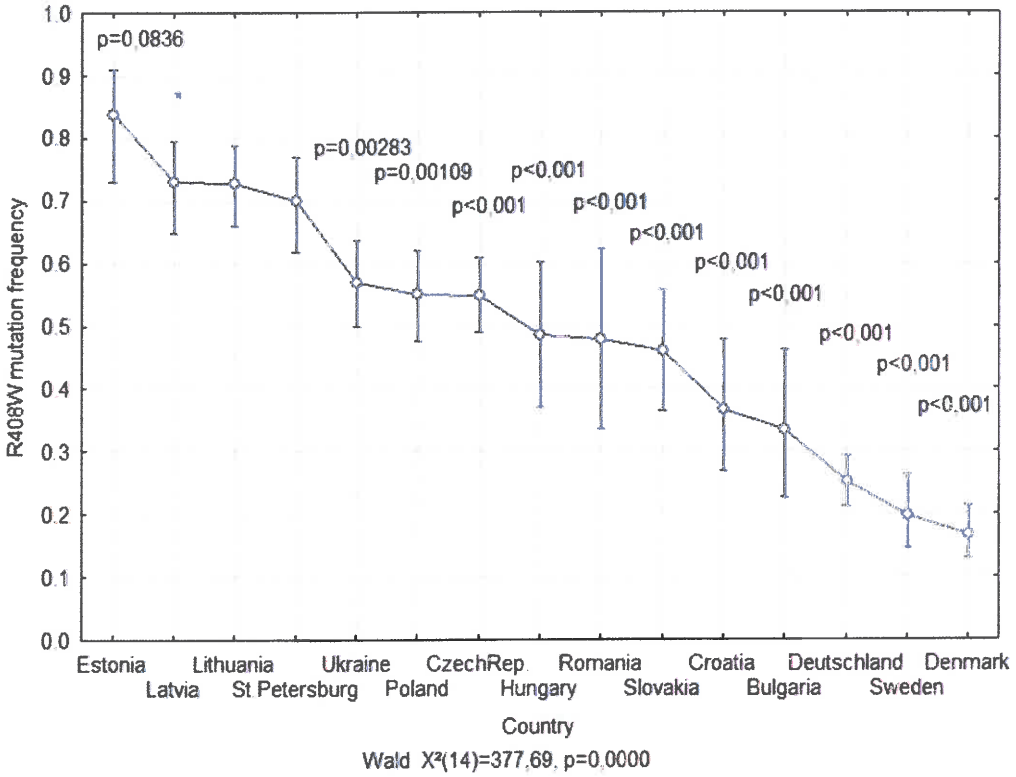


Fig. 3.23. Results of statistical analysis between R408W mutation frequencies for different European countries. Significance level set at 0.05 was used to compare Latvia (*) to the other geographic populations.

4 DISCUSSION

Phenylketonuria is the most common inborn error of amino acid metabolism in Europeans. It has been frequently described as a paradigm of a Mendelian disorder. PKU was the first metabolic cause of mental retardation to be identified, the first genetic disorder of the central nervous system that could be fully treated by modification of external factors (i.e., the diet), and the first disorder that was successfully diagnosed by universal neonatal screening.

The understanding of the biochemical and molecular basis of phenylketonuria and the innovative treatment strategies introduced for these patients during the last 60 years, were transferred to other inborn errors of metabolism and led to significant reduction in morbidity and to an improvement in quality of life.

The *PAH* gene was first cloned in 1983 (Woo, 1983). Almost 30 years have now passed since the first molecular tests were carried out in PKU families. The most studies of *PAH* gene mutations in different populations were carried out in the 1990s and a *PAH* Mutation Analysis Consortium was formed, as a result PAHdb, an online relational database, was created.

Contemporary interest in PKU more often is focused on the development of new therapeutic approaches, but it is still closely related to the biochemical and molecular basis of PKU.

4.1 MUTATIONS IDENTIFIED IN LATVIAN PKU CHROMOSOMES

In our study the most common mutation in Eastern Europe R408W was found in relative frequency of 73% in Latvian PKU patients. Similar frequencies of R408W are found in Lithuania and St. Petersburg region. The highest relative frequency of this mutation is found in Estonia (84%). Statistical analysis did not show a significant difference in the distribution of the R408W allele between Latvia, Lithuania, Estonia and St. Petersburg region. So the high frequency of this mutation is explained by Balto-Slavic origin of R408W mutation that also is confirmed by our results. The relative frequency of R408W is progressively lower the greater the geographic distance from regions with the highest relative frequency ($p<0.001$) suggesting that this mutant allele may have been introduced into these populations by the relatively recent spread of European peoples across the Eurasian landmass.

We did not investigate the haplotype of mutation R408W because the significant differences in comparison with other Eastern European populations were not expected. Thus we assume that mutation R408W in Latvian population is associated with haplotype 2. Instead of this we provided minihaplotype studies for this mutation and found that 97% of R408W alleles are associated with three-copy VNTR allele but 82% with VNTR3/STR238 minihaplotype. This founding also suggests that our assumption about haplotype association could be truthful because for eastern European populations the R408W mutation is strongly associated with RFLP haplotype 2, the three-copy VNTR allele (VNTR 3), and the 238-bp STR allele. Three R408W alleles (9%) were found in association with minihaplotype 3/242, two alleles (6%) in association with minihaplotype 3/234 and only one R408W allele (3%) was found on minihaplotype 8/238. In comparison with Estonian and Lithuanian PKU populations, R408W was found on Eastern-European minihaplotype 3/238 on 98% and 68% of R408W alleles, respectively. Minihaplotype 3/242 for R408W was also found in Polish population on 6 mutant chromosomes (6.5%) and on single chromosome in Estonia. No "Celtic" allele VNTR8/STR242 that is the most common in north-western European populations was found in Latvian PKU chromosomes.

Mutation R408W was also found in association with minihaplotypes 8/230, 8/246, 9/242 and 9/246 in other European populations but on small number of chromosomes. The association of a mutation with different VNTR and STR background is common in different populations and reflects a hypervariability of tandem-repeat DNA sequences. The other mechanisms leading to minihaplotype heterogeneity include a founder effect and a recurrence mechanism of the mutation. R408W allele's codon contains a methylated cytosine, which can experience methylation mediated deamination of the 5mC nucleotide to create thymine (c.1222C>T). In CpG dinucleotides mutation has been estimated to occur much greater than mutations at non-CpG dinucleotides. This process of recurrent mutation provides one explanation for the high relative frequency (<10%) of the R408W allele in European PKU genomes.

The high degree of minihaplotype variability associated with one mutation could also indicate an ancient origin.

PKU in Latvia is homogeneous. Calculation of the homogeneity ("homo-zygosity") at the *PAH* locus in Latvia gives a value of 0.514.

In comparison, the most homogeneous population described so far are Yemenite Jews, in whom a single molecular defect (large deletion covering the exon 3 of the *PAH*

gene) is responsible for all the PKU cases in the population. Accordingly, the homozygosity value for this population is 1 (Avigad, 1990). An analysis of family histories of the Yemenite Jewish community has traced the origin of this defect to a common ancestor from Sana, Yemen, before 18th century (Killeen, 2008).

Other populations of Eastern European countries are also quite homogeneous. Homozygosity value for Estonia is 0.62 (Lilleväli, 1996), for Lithuania is 0.54 (Kasnauskiene, 2003), for Poland is 0.35 (Jaruzelska, 1993), for Southern Poland is 0.44 (Zygulska, 1991). Homozygosity value is rather low in ethnically mixed populations, for example, in Germany – $j = 0.08$ (Zschocke and Hoffmann, 1999). Germany has been at the crossroads of migration throughout the history of Europe. At present, almost 20% of people living in Germany are relatively recent immigrants and the largest ethnic group of non-German origin is the Turkish.

The homozygosity value has to be considered when a *PAH* gene mutation detection strategy for a particular population is being created. The high level of homozygosity may facilitate genotyping and to raise the mutation detection rate.

Relatively high homozygosity level also has an impact on clinical picture of the disease for a significant number of patients. Mutation R408W is a severe mutation that is associated with < 0.3% of normal activity and a severe PKU phenotype. Taking into account that thirty-six (51.4%) of 70 unrelated characterised PKU patients are homozygous for R408W, the prevalent clinical form of the PKU supposed to be severe. Remaining 34 (48.6%) patients are compound heterozygous. The majority of compound heterozygote patients (43%) had R408W mutation in one allele that means that their clinical form of PKU depends on the mutation in other allele of the *PAH* gene. Only four Latvian PKU patients (5.7%) had no R408W mutations in their chromosomes.

Next most common mutation found in Latvian PKU patients was E280K presented on 5.7% of all PKU chromosomes. Mutation E280K is a severe mutation with enzyme residual activity ~0.9% of wild type. Glu280 is located in a stretch of 27 PAH amino acids (His263 to His289), which is highly conserved and is closed to the active site. It forms a salt bridge to Arg158 plus hydrogen-bond to His146 (Appendix IX). The mutation E280K would be expected to affect severely the electrostatic potential in the active site (Jennings, 2000).

Seven of eight (7/8) E280K alleles were found with R408W in homologous allele, these patients had severe form of PKU due to very low enzyme activity determined by both severe *PAH* gene mutations. One patient had genotype E280K/ L48S.

E280K allele, involving a CpG dinucleotide, was found in 3 different haplotypes that suggests the recurrent mechanism for this mutation. In Latvian PKU chromosomes mutation E280K was found in strong association with minihaplotype 9/250, one allele was found on minihaplotype 9/246. In comparison with other populations it was found on different minihaplotypes 7/246 (Germany), 8/238 and 8/246 (Northern Ireland), 9/234 (Spain). Mutation E280K is not very common in European populations and usually was found only in few PKU chromosomes.

Two mutations were detected with frequency 3% each - R261Q and R158Q. The R158Q mutation is a frequent mutation in patients with PKU in European countries but usually represents less than 10% of PKU chromosomes. As mentioned above, Arg158 forms a salt bridge to Glu280, but also forms a hydrogen bond to Tyr268 (Appendix IX). Both of these interactions are important for conserving the shape of the active site, and substitution into a glutamine residue will alter the active site architecture and lower enzymatic activity. Mutation R158Q is classified as severe with the residual activity of 10% for PAH enzyme (Jennings, 2000). However, sometimes it was classified as moderate. So R158Q is one of mutations that are frequently represented in discordant genotype-phenotype associations. A potential cause of these inconsistencies may relate to the biological properties and functions of the mutant protein. It was noticed that some patients with mild PKU have a relatively low phenylalanine tolerance when treated, although plasma phenylalanine levels were only a little above the therapeutic threshold when these patients were untreated. Mutation R158Q is considered to be BH4-responsive. In our study mutation R158Q was found only with R408W mutation in homologous allele and all four patients were characterised as having severe PKU.

Minihaplotypes studies have revealed that mutation R158Q was associated with three different minihaplotypes: 3/238 (two alleles), 3/234 (one allele) and 7/234 (one allele). In other populations this mutations was found on minihaplotype 3/234 (Germany, Italy and Northern Ireland). This mutation also involves a CpG dinucleotide and different minihaplotypes could be the result of the recurrent mechanism of this mutation but available data are not sufficient to make conclusions.

Mutation R261Q is a moderate PKU mutation with residual activity >30% of wild type of *PAH* enzyme. R261Q was found throughout the Europe but the highest frequency was discovered in Switzerland (>30%). Arg261 is located in the cofactor binding domain (CBR1), where it hydrogen bonds to Gln304 and Thr238 (Appendix X). This helps to stabilise the structure of the active site, and substituting the arginine would

destabilise the active site. Mutation R261Q is considered to be BH4-responsive. It was associated with responsiveness both in combination with null mutations and not. However, in another studies it was found not to be responsive.

Genotyping of Latvian PKU patient has revealed that three patients are compound heterozygotes for R261Q and R408W mutations and one patient for mutations R261Q and A104D. As a confirmation of mentioned above all patient were classified as having severe form of PKU in despite of moderate phenotype of R261Q.

Four R261Q alleles were associated with two minihaplotypes – 3/238 and 8/238 equally. In other European populations the most common minihaplotype for R261Q was found 8/238 (Germany and Italy) but it was also found on minihaplotypes 3/246, 7/242, 8/234, 8/242 and 8/246. Association with several minihaplotypes and distribution among different European populations could be result of the presence of hypermutable CpG dinucleotide and also indicate an ancient origin of this mutation.

P281L mutation is not very common in Europe, the highest frequency 19% was found in Iceland, Croatia and Greece showed similar frequencies 11% and 10%, respectively. In other European populations the most common frequency of P281L was 1-3% and did not exceed 6%. Three Latvian PKU chromosomes (2.1%) were found to harbour P281L allele. Pro281 is located in the active site and helps to define the shape of the active site very close to the iron in the PAH structure (Appendix IX). A substitution to a less rigid leucine will change the conformation of the active site by removing the conformational constraints imposed by the proline and resulting in <1% of enzyme activity.

In Latvian PKU patients P281L mutation was found in two different genotypes: R408W/P281L for two patients and R261X/P281L for one patient. All patients had severe PKU that is compatible with metabolic phenotype of mutations.

P281L mutation's the most common minihaplotypes 7/242 and 8/242 both were found in Latvian P281L alleles on two and one chromosome, respectively. Additionally, minihaplotypes 8/238, 8/234 and 3/234 were found in Italian population, minihaplotype 3/242 in one Polish PKU patient. Association with several minihaplotypes and distribution among different European populations could be result of hypermutable CpG dinucleotide in proline codon that indicates the recurrent mechanism of this mutation. Variability of minihaplotypes also could indicate an ancient origin of this mutation.

IVS10-11G>A is the most common mutation in the Mediterranean, particularly in Turkey where it accounts for more than 30% of PKU alleles. Mutation is categorised as severe because completely abolishes PAH enzyme activity. This splicing mutation most likely results in truncated protein lacking C-terminal 97 amino acid residues (residues 356-402). As a result an unstable protein is produced.

We found mutation IVS10-11G>A in two PKU chromosomes corresponding to a frequency of 1.4%. Both patients had severe form of PKU but different genotypes - IVS10-11G>A/IVS12-1G>A and R408W/IVS10-11G>A.

Both Latvian IVS10-11G>A alleles were found in association with the most prevalent for it minihaplotype 7/250, although, it was found associated with minihaplotype 7/230 in Spanish Gypsies, Czech, English and German patients. IVS10-11G>A is probably an ancient mutation that originated long before the end of last ice age and separated into different alleles early in prehistory. The east-west gradient in the Mediterranean basin with the highest focus in Turkey, has suggested a spread from Asia Minor during the Neolithic period (Calì, 1997). Moreover, recent migration has brought the mutation to Northern European countries like Germany.

Mild mutation A104D with residual activity <25% was found on two Latvian chromosomes – 1.4%. A104D is not frequent mutation and presents in small frequencies in Central and Northern Europe. Mutation is located in regulatory domain of PAH enzyme (Appendix XI) and is classified into BH4-responsive alleles. This mutation is associated with variant PKU, and neutral Ala104 is located in a loop between Rα2 and Rβ4 in the regulatory domain. Substitution into a larger and charged residue may destabilise this loop structure.

Mutation A104D was found in two compound heterozygous genotypes with mutations R408W and R261Q in homologous alleles. Both patients had severe PKU despite the fact that mutation R261Q was considered to be BH4-responsive also confirming the lack of the strong correlation between genotype and BH4-responsiveness.

Information about A104D allele's minihaplotypes was available only from German PKU population where it was associated with 8/242 and 8/246 minihaplotypes. Latvian A104D alleles were found in association with 8/242 minihaplotype both. Alanine codon does not involve hypermutable CpG dinucleotide that could be the reason for mutation rarity.

Three nonsense mutations, R111X, R261X and G272X, were found only once corresponding to a frequency of 0.7% each. All three alleles are classified as severe with residual activity <1% of wild type PAH enzyme activity. Mutation R111X is located in regulatory domain (exon 3), mutations R261X and G272X in catalytic domain (exon 7) of the protein. Mutations result in C-terminally truncated protein.

As in the case of mutation R261Q, substitution of the arginine to a premature stop codon (R261X) will necessarily destabilise the active site structure of the enzyme.

Glu272 is located in a loop just before the active site histidines that bind the catalytic iron and the mutation results in a truncated form of PAH that has none of the residues that are responsible for binding iron; thus, no catalytic activity will be observed.

Mutation R111X causes the loss of approximately two-thirds of the PAH polypeptide. It is more common in Orientals, in European populations frequency of R111X was found about 1-2%. It was also found in Turkey, Italy, Sicily, and Australia. Mutation G272X is common in Norway (16%), while R261X frequency varies within 1-2% in Europe.

Mutations G272X and R111X were found in null/null genotypes with R408W mutation in homologous allele. Mutation R261X was found together with P281L allele. All three patients showed severe PKU phenotype.

In Latvian PKU chromosomes mutations R111X, R261X and G272X were found in association with minihaplotypes 8/250, 7/238 and 9/234, respectively.

Minihaplotypes 3/238 and 3/246 was identified for R261X allele in German PKU patients, but minihaplotypes 7/242 (the most common), 3/238 and 3/242 in Italian PKU patients. Since a CpG dinucleotide is involved the mutation could have arisen independently on different populations and minihaplotype backgrounds.

Mutation G272X was found on minihaplotype 8/226 in Germany. No more minihaplotype data are available for G272X allele from other European populations and there is a total lack of data about R111X mutation's minihaplotype association.

Four mild mutations, A403V, E178G, I306V and V230I, were classified as mutations caused MHP with residual activity of PAH enzyme 32%, 39%, 39% and 63%, respectively. Each mutation was found in single chromosome only. Mutations A403V, E178G and V230I were previously reported as BH4-responsive.

Mutation A403V has been detected in southern Europe, and in Spain it is a relatively common mutation (14%).

Ala403 is located at the end of helix Cα12, close to Ala309 in helix Cα8. Alanine or another smaller residue might be necessary for close packing of helices Cα8 and Cα12 (Appendix XI and XII). Substitution into a larger valine might result in a less stable protein because it would require the surrounding protein to adjust and create space for the accommodation of the bulkier side chains. Thus, the BH4 binding site in mutant PAH might be only slightly different as compared with the wild-type PAH structure, explaining the BH4-responsiveness of this genotype.

Mutation V230I has the same effect on the enzyme as mutation A403V due to similar substitution of smaller hydrophobic valine residue to larger isoleucine that in the same would require the adjustment of surrounding protein and creation of space for the accommodation of the bulkier side chains (Appendix XI).

Mutation E178G is not frequent in Europe. E178G is located on the surface of catalytic domain (Appendix XI). Substitution to a small and flexible hydrophobic residue may be very unfavourable, because it can change the fold of the catalytic domain core, which is important for maintaining proper catalytic function. The enzyme is quite susceptible to mutations that destroy the cooperative activation mechanism probably by hindering the transmission of the conformational change.

Mutation I306V is located close to the active site (Appendix XII). It results in change of large buried hydrophobic residue to smaller one and destabilises the protein by creating cavity in the hydrophobic core and, the effect depends on the size of the resulting cavity.

All four mild mutations were found together with R408W mutation in functionally hemizygous genotypes. Three patients were classified as having mild PKU phenotype. Patient with genotype R408W/V230I had MHP and did not require treatment.

Mutations A403V, E178G, I306V and V230I were found in associations with minihaplotypes 8/246, 7/242, 3/234 and 3/246, respectively.

Association with the same minihaplotype for mutation A403V was found in Poland, Italy and Spain. Strong association with minihaplotype 8/242 was identified in German PKU patients; it was also prevalent minihaplotype in Italian and Polish PKU chromosomes.

Minihaplotype data for mutation V230I are available only from German PKU population where it was associated with 3/242 minihaplotype. No data were available for mutations E178G and I306V.

Mutation L48S is quite common in the Southern Europe. It was more common in PKU chromosomes in Serbia (21%) and in South of Italy (11%). L48S mutation was classified as mild and BH4-responsive but was frequently represented in discordant genotype-phenotype associations. Mutation is located in regulatory domain and results in change of buried hydrophobic leucine residue to polar serine (Appendix XI). Burial of polar side chains results in decrease of protein stability. L48S mutation was found in single Latvian PKU chromosome with null mutation E280K in homologous allele. Patient was functionally hemizygous but was classified as having severe PKU form.

In our study mutation L48S was found on minihaplotype 3/234 that was prevalent minihaplotype for it in German and Italian PKU alleles. Minihaplotypes 3/230, 3/238 and 8/238 were also identified for L48S in single Italian PKU chromosome each.

Only one small deletion was identified in *PAH* gene of Latvian PKU patients - mutation E221_D222>Efs. Mutation results in fusion protein. Such kind of mutations introduces frameshifts and result in mutant proteins containing a truncated PAH sequence fused to an unrelated sequence at the C-terminus. No data about the protein residual activity was available but mutation was classified as causing severe PKU.

Mutation E221_D222>Efs was identified in genotype with other severe mutation R408W and PKU form was classified as severe.

We found this mutation in association with minihaplotype 3/242. In two German PKU chromosomes it was found on minihaplotypes 3/238 and 3/242.

The mutation IVS12+1G>A results in a truncated protein lacking the C-terminal 52 residues (residues 401–452). *In vitro* expression of the mutant protein comprising residues 1–400 suggests an unstable protein is produced possibly because residues 409–422 participate in the dimer and domain interfaces. Therefore in genotype with other null mutation as R408W it gives severe PKU phenotype.

IVS12+1G>A is quite common in Scandinavia and it frequency reaches 37% in Denmark. In Latvia it was found in a single PKU chromosome (0.7%) and was associated with minihaplotype 8/242 that is only minihaplotype reported for this mutation in Germany, Italy and Northern Ireland. Latvian patient harbouring this

mutation had R408W in the other allele, so, patient's severe PKU was compatible with metabolic phenotype of both mutations.

Three mutations (P292T, K371E and IVS12-1G>A) identified in present study had not been previously reported; two of them were found only once, but the third one was identified on two unrelated PKU chromosomes. None of these mutations have been examined by *in vitro* expression analysis but also none of them had been observed in the 100 normal subjects tested (200 chromosomes). These results do not fully exclude, but reduce the possibility of population-specific polymorphism. The all 13 exons of *PAH* gene were sequenced for these four patients but other PKU mutations were not found. Of course, *in vitro* expression analysis remains the most effective way to confirm that a "disease-associated" mutation is truly pathogenic.

In one family a PKU patient presented the P292T mutation that was a c.874C>A substitution in the *PAH* gene in exon 8 at amino acid 292, resulting in a missense mutation – hydrophobic proline is substituted by hydrophilic threonine. Mutation was linked to minihaplotype 8/226. The patient had a neonatal diagnosis, with phenylalanine level indicating severe PKU (~25mg%), and received diet therapy soon after. Mutation P292T was found in heterozygosity with severe R408W mutation. Considering that PKU clinical form depends on the combination of mutant alleles inherited, we suggest that mutation P292T has to be associated with low residual activity of PAH enzyme. This variant was also predicted to be probably damaging by using an automatic tool for prediction of possible impact of an amino acid substitution on the structure and function of a human protein (PolyPhen).

In parents: mutation R408W was inherited from patient's mother but the paternal chromosomes did not carry either of the PKU mutations identified in the child. Since the results of paternity testing for this family were compatible with paternity, we concluded that the second mutation has to be arisen *de novo.* Thereby, the risk for another child with PKU in this family will be low.

These findings demonstrate two important points: the necessity of screening the whole coding region of the *PAH* gene for diagnostic purposes on the one hand, and second the usefulness of confirming inheritance of mutations from both parents when possible. Otherwise the prediction of the expected phenotype or the calculation of risk for another child with PKU may be incorrect.

The 1111A>G substitution in exon 11 of the *PAH* gene at amino acid 371 results in a missense mutation – lysine is substituted by glutamic acid. Both amino acids are

classified as polar. This substitution was found in a PKU patient and his father and was associated with minihaplotype 3/238. The patient was diagnosed through the neonatal screening and had slightly elevated phenylalanine level (2.7mg%) that did not require treatment. Mutation K371E also was found in heterozygosity with R408W mutation. This finding suggests that K371E is a mild mutation with enough PAH residual activity for normal clinical phenotype. These findings match with PolyPhen analysis results that predicted this change as benign. Genotype K371E/R408W could be defined as functionally hemizygous. The normal individuals tested for this mutation did not present it.

The c.1316-1G>A substitution in intron 12 of the *PAH* gene results in a mutation. This substitution was found in two unrelated patients with severe PKU. The mutation IVS12-1G>A is located at the boundary of intron 12 and exon 13 and affects the conserved dinucleotide AG at the 3' splice site. According to its location this mutation results in a truncated protein lacking the C-terminal at least 14 amino acid residues (residues 439-452). This mutation also was not found in 200 chromosomes of 100 healthy individuals.

In one patient this mutation was found in association with mutation R408W but in another one in association with another known splice site mutation IVS10-11G>A. Both patients were diagnosed through the neonatal screening. Patient with genotype IVS12-1G>A/R408W received diet therapy soon after and is currently asymptomatic despite the severe clinical form of PKU. Patient with genotype IVS12-1G>A/IVS10-11G>A has severe mental retardation because of parents' refusal to maintain the adequate dietary treatment. In both cases mutation IVS12-1G>A was linked to minihaplotype 7/242.

Remaining unknown allele may harbour large deletion and thus deserve further investigation by other techniques.

During our work we are faced with some difficulties comparing the frequency of mutations in different populations, especially concerning rare mutations in the *PAH* gene. Some studies were conducted in early 1990s and only the most common mutations were identified using that time available techniques.

Minihaplotype analysis is possible in two ways: in case of homozygous mutation or if patient's parents are available for investigation. Unfortunately, only in few populations minihaplotype analysis was performed for full spectrum of *PAH* gene mutations, more often it was made only for more common mutations. All together, these

limitations do not allow tracing the origin of mutations and later distribution to other populations.

Sequencing analysis revealed the presence of 9 different silent mutations in different combinations in Latvian PKU chromosomes. Three of them were located in exons and 6 were located in introns. No specific distribution of *PAH* gene polymorphisms was observed. Polymorphisms detection rate depends on primers used for sequencing analysis. Intron polymorphisms located more distantly from the exon/intron boundaries may remain undetected.

According to Hardy-Weinberg equilibrium Latvian PKU patients' pool is balanced. Statistical analysis confirmed that the observed and expected genotypes' frequencies are not significantly different from one another ($p = 0.8312$). However, our study did not reveal the presence of other homozygous patients in PKU population that according to Hardy-Weinberg equilibrium has to be up to 1%. It could be explained by MHP patients that is part of the PAH deficiency spectrum and is frequently caused by compound heterozygosity for classical PKU mutations and specific mild mutations or, possibly, by two mild mutations. These patients may or may not have been included in the molecular studies. This poses some deviations when absolute allele frequencies in a population are estimated using incidence data and the Hardy-Weinberg formula.

4.2 GENOTYPE-PHENOTYPE CORRELATIONS IN LATVIAN PKU PATIENTS

Compiling data on mutations observed in Latvian PKU chromosomes, we can conclude that the majority of mutations (12/20) are severe and responsible for the severe PKU phenotype. Two mutations are mild and five are MHP causing; only one is responsible for moderate PKU. So the high proportion of patients with severe or classic PKU is explained by the mutation severity. We compared two different approaches for classifying the PKU clinical phenotype and results did not shown the significant difference ($p = 0.2276$). The traditional method for PKU phenotype classification is based on Phe pre-treatment serum level and Phe tolerance. The other model is based on an arbitrary value for phenotypic prediction system. By means of this classification, a phenotype resulting from the combination of two mutant *PAH* alleles may be expressed numerically as the sum of the AVs of two mutations. In our study the majority of PKU patients corresponded to that rule but some deviations were observed. Mostly it relates

to mutations with residual *in vitro* activities. Many factors can influence phenotypic variation in PKU, such as inter-individual variations in intestinal absorption, hepatic uptake of dietary phenylalanine, rate of incorporation of phenylalanine into proteins, rates of influx of phenylalanine across the blood brain barrier, mutations located close to the cofactor binding site and affecting the activity of the enzyme, as well as interactions of the *PAH* gene with other loci.

The efficiency of the method based on AVs estimates will vary depending on the set of mutations in a specific population. For populations in which the most common mutations are null mutations, the system could be highly useful.

Contemporary therapy for PKU is centered upon tight restriction of dietary Phe intake and requires supplementation with special medical foods that supply sufficient essential amino acids and energy from fat and carbohydrate. Institution and maintenance of the PKU diet are difficult, and the required medical foods are often unpalatable. Dietary therapy is recommended for life (Anonymous, 2001), but non-compliance with the dietary prescription is commonplace, particularly during adolescence and adulthood. Hyperphenylalaninaemia in adults is often associated with attention problems, mood instability and poor job performance. Chronically elevated Phe may cause a progressive neurodegenerative disorder affecting white matter that leads to seizures and gait disturbance. Finally, untreated maternal hyperphenyl-alaninaemia during pregnancy is the only teratogen guaranteed to cause birth defects, which include microcephaly, mental retardation and congenital heart disease.

There is no other effective and relatively simple type of treatment that could completely replace the dietary treatment, although, research in this direction is made constantly. However, chaperon therapy by tetrahydrobiopterin supplementation is effective in a subset of individuals with BH4-responsive hyperphenylalaninemia that has been recently described as a variant of PAH deficiency caused by specific mutations in the *PAH* gene.

The evaluation of the data contained in the BIOPKU data base we would expect BH4-responsiveness in about 18% (13/70) of all Latvian patients with PAH deficiency. Seven mutations from 20 observed in Latvian PKU patients have substantial residual activity of PAH (10-39%): mutation R158Q with residual activity 10%, A104D – 26%, A403V – 32%, mutations R261Q, I306V, E178G and L48S with residual activity 39%.

There is still some inconsistency reported of BH4-responsiveness in patients harbouring L48S, R158Q and R261Q mutations. This inconsistency confirms the lack

of the strong correlation between genotype and BH4-responsiveness. In this case, in which this residue is implicated in the interaction of two neighbouring subunits, the second PKU allelic variant in the patient could have a high significance in determining the responsiveness to BH4.

A genotype can be considered as associated with BH4-responsiveness if one of alleles harbours BH4-responsive mutation with a substantial residual PAH activity. Among 70 Latvian PKU patients 13 can be considered as potentially BH4-responsive: three patients with genotype R408W /R261Q, four patients with genotype R408W /R158Q, six patients with genotypes R408W/A104D, R408W/A403V, R408W/E178G, R408W/I306V, R261Q/A104D and E280K/L48S. Twelve patients have null mutation in homologous allele. In comparison, BH4-responsiveness is much higher (<75%) in southern regions of Europe with a high frequency of BH4-responsive alleles. Due to the high prevalence of mutation R408W and other null mutations in Baltic countries, the number of potentially BH4-responsive patients is relatively low.

Despite the fact that there is no strong correlation between genotype and BH4-responsiveness, mutation analysis provides useful information on potential nonresponders in patients harbouring two null alleles and may, to some extent, predict possible BH4-responders.

4.3 MINIHAPLOTYPE STUDIES FOR MUTANT AND NORMAL *PAH* ALLELES

Minihaplotype studies have revealed 16 different minihaplotypes associated to *PAH* gene mutations and 20 different minihaplotypes for normal *PAH* alleles. The most common minihaplotype for mutant alleles was 3/238 due to the high prevalence of mutation R408W among Latvian PKU chromosomes, while for normal alleles more common minihaplotype was 3/242. Distribution of STR alleles is consistent with the previously reported data about higher frequency of the 238bp allele among mutant chromosomes and the higher frequency of the 242bp and 246bp alleles among normal chromosomes in Caucasians from different European populations. Contrary, distribution of VNTR alleles is different from the accepted opinion that the VNTR allele containing 8 repeats is the most prevalent (about 60%) among both normal and mutant chromosomes. The prevalence of VNTR 3 allele among mutant chromosomes is explained by the most common R408W mutation, but distribution of VNTR 8 allele in

normal chromosomes is consistent with this statement. This is partially confirmed by statistical analysis that has revealed the significant difference in the distribution of normal and mutant alleles for only two minihaplotypes 3/238 (p=0.0000572) and 8/230 (p=0.0133).

Statistical analysis showed the tendency to statistically significant difference ($p<0.10$) between normal and mutant alleles in the distribution of minihaplotypes 3/242, 7/246 and 8/234. The approval of these trends requires an investigation of greater number of alleles.

The average probability of heterozygosity for minihaplotypes was about 76% for mutant and 92% for normal chromosomes indicating a greater diversity of normal alleles. The association of minihaplotypes with specific mutations results in it limitation in comparison with normal alleles that makes it useful for prenatal diagnosis and carrier screening determination in PKU families.

One of the limitations in performed minihaplotype studies was insufficient number of analysed chromosomes due to relatively small Latvian PKU population.

4.4 THE MOST EFFECTIVE STRATEGY FOR ROUTINE DIAGNOSTICS OF *PAH* GENE MUTATIONS IN LATVIA

The three-step *PAH* gene mutation detection strategy based on the information of PKU causing mutation spectrum and PKU population homozygosity value can be found to be the best for routine diagnostics for Latvian PKU patients. In the first step, the common mutation R408W detection with restriction enzyme assay that identifies both alleles in >50% of patients and one allele in a further 43% is used. In the second step, denaturant gradient-gel electrophoresis is used to determine the possible location of other *PAH* gene mutations. In the final third step, depending on DGGE results (number of exons that showing variant electrophoretic patterns) direct sequencing analysis or minihaplotypes combining STR and VNTR data could be used to determine rare mutations. In case of prenatal diagnostic or the need to quickly provide information on *PAH* gene mutations minihaplotype analysis could be used after DGGE and prior to direct sequencing analysis. For example, exon 7 harbours the majority of PKU mutations and several silent mutations. Minihaplotype analysis can provide faster and less time-consuming response for determining the nature of sequence changes.

Taking into account that the diagnostic strategy has to be designed to identify a great number of mutations, the detection rate of 99% achieved in our study confirms that the diagnostic approach used had the best possible design.

5 CONCLUSIONS

1. Analysis of the molecular basis of PKU in Latvia has revealed 20 different mutations in the *PAH* gene: the most common mutation R408W accounted for 73% of all PKU chromosomes, the frequencies of remained 19 mutations ranged from 0.7 to 5.7% of all mutant alleles.

2. Minihaplotypes (combinations of *PAH* gene STR and VNTR systems) were determined for all 20 mutations indentified in Latvian PKU chromosomes; a strong association of mutation R408W with VNTR3/STR238 minihaplotype was indicated.

3. The average probability of heterozygosity for minihaplotype system was found lower for mutant chromosomes (0.76) compared to normal *PAH* chromosomes (0.92) indicating a greater diversity of normal alleles.

4. The estimation of genotype-phenotype correlation has revealed that Latvian PKU patients are homogeneous in terms of clinical PKU form due to the high frequency of severe R408W mutation and the high level of homogeneity (j=0.514) at the *PAH* locus.

5. Analysis of the distribution of the *PAH* gene mutation R408W has confirmed the Balto-Slavic origin of mutation R408W and introduction of this mutation to other European populations by people migrations.

6. The three-step *PAH* gene mutation detection strategy used in the study is the most effective for routine diagnostics in Latvian population with the sensitivity of the method 99%.

6 PUBLICATIONS

1. Pronina N, Giannattasio S, Lattanzio P, Lugovska R, Vevere P, Kornejeva A. The molecular basis of phenylketonuria in Latvia. - Hum Mutat., 2003 Apr., V. 21, (4), pp. 398-399.
2. Natalija Pronina, Rita Lugovska, Parsla Vevere. Mutational spectrum in Latvian patients with phenylalanine hydroxylase deficiency. RSU Research articles in medicine and pharmacy, 2008, pp.76-79.
3. Natalija Pronina, Rita Lugovska. Association between minihaplotypes and mutations at the PAH locus in Latvian phenylketonuria patients. Proceedings of the Latvian Academy of Sciences. Section B, Vol. 65 (2011), No. 3/4 (674/675), pp. 73–79.
4. Natalija Pronina, Rita Lugovska. Three novel mutations in the Phenylalanine Hydroxylase Gene (PAH) observed in Latvian patients with Phenylketonuria. Accepted for publication in *RSU Research articles in medicine and pharmacy 2011.*

THESES AND PRESENTATIONS

1. V.Kucinskas, V.Jurgelevicius, D. Stepanoviciute, S. Giannattasio, P.Lattanzio, E.Marra, R.Lugovska, N.Pronina, J.Zschocke. Molecular Genetic Testing in Phenylketonuria; Model to Assess The Quality Control System For Monogenic Diseases. – Abstr.10th International Congress of Human Genetics, 2001, Vienna. - European Journal of Human Genetics, 2001, V. 9, Suppl. 1, p. 299.
2. N. Pronina, R. Lugovska, P. Vevere, A. Kornejeva. Mutation analysis of PAH gene among Latvian patients. Abstr. European Human Genetics Conference, May 25-28, 2002, Strasbourg, France. - European Journal of Human Genetics. 2002, V.10, Supp l, p. 239.
3. N. Pronina, S. Giannattasio, P. Lattanzio, R. Lugovska. The molecular basis of phenylketonuria in Latvia. - Abstr. European Human Genetics

Conference, May 3-6, 2003, Birmingham, England. - European Journal of Human Genetics. 2003, V.11, Supp l, p.171.

4. Pronina N., Lugovska R., Muceniece Z., Olhovaya O., Krumina Z., Lace B., Balode I. Experience of molecular diagnostics for genetic disorders in Latvia.- Abstr. European Human Genetics Conference, May 7-10, 2005, Prague, Czech Republic. - European Journal of Human Genetics. 2005, V.13, Supp l, p.286.

5. Pronina N., Lugovska R., Sterna O., Daneberga Z., Vevere P. Investigation of phenotype-genotype correlations for patients with PKU in Latvia. Journal of Inherited Metabolic Disease. Abstracts. Vol. 30, Suppl.1, 2007, pp. 9.

6. Pronina N., Lugovska R., Sterna O., Daneberga Z. Phenotype-genotype correlations for patients with PKU in Latvia. European Journal of Human Genetics Vol 15 Supp. 1, 2007, pp. 219.

7. Lugovska R., Krumina Z., Lace B., Vevere P., Pronina N., Daneberga Z.,Kornejeva A. Newborn screening and diagnosis of inborn errors of metabolism in Latvian children. Laboratorine Medicina (ISSN 1392-6470), 2008, Vol. 10, Special Suppl, p. 15.

8. N. Pronina, R. Lugovska. Identification of three novel single nucleotide changes in the *PAH* gene observed in Latvian patients with phenylketonuria. European Journal of Human Genetics Vol 19 Supp. 2, 2011, pp. 435.

9. Pronina N., Lugovska R. The role of genotyping to patients with phenylketonuria. Uzstāšanas RSU Zinātniskajā konferencē, 14.-17. aprīlis, 2011.g.

7 ACKNOWLEDGEMENTS

I would like to thank my supervisor Assoc. Professor. *Rita Lugovska* for the opportunity and management of my work.

Special thanks to Dr. *Sergio Giannattasio* for the possibility to acquire the analysis of minihaplotypes in Research Centre of Mitochondrial and Energy metabolism (Centro di Studio sui Mitocondri e Metabolismo Energetico, CSMME) in Bari (Italy), as well as his colleagues in the laboratory: *Paolo Lattanzio, Ersilia Marra, Salvatore Passarella.*

Many thanks to Professor. *Johannes Zschocke* for the cooperation during the European Commission's FP 4 INCO-Copernicus project "Molecular Genetic Testing in phenylketonuria: a model of ASSESS The Quality Control System for monogenic disease" (MOLGENT, 1999-2002) framework.

My sincere thanks to my colleagues from Medical Genetics Clinic, but especially to colleagues from the DNA laboratory: *Kristina Morozova, Olga Sterna, Zanda Daneberga* and *Oksana Osipova.*

I would like to thank colleagues of the RSU Molecular Genetics Research Laboratory for a deep emotional support in the making of the work.

Thanks to *Normunds Jurka* for the help of the statistical data processing and interpretation.

The current work was supported by the European Social Fund (ESF).

8 REFERENCES

1. Altman, D.G., Bland, J.M. (1994). Diagnostic tests. 1: Sensitivity and specificity. *BMJ.* 11;308(6943), 1552.
2. Anonymous. (2001). National institutes of health consensus development panel. National institutes of health consensus development conference statement: phenylketonuria: screening and management, October 16–18, 2000. *Paediatrics,* 108, 972–982.
3. Anonymous. (2003). Davidson College. Biology Department. http://www.bio.davidson.edu/Courses/Molbio/MolStudents/spring2003/Kogoy/favorite.html.
4. Anonymous (2011). The National Information Centre for Metabolic Diseases (NICMD), UK, http://www.climb.org.uk/pro.htm.
5. Aoki, K., Ohwada, M., Kitagawa, T. (2007). Long-term follow-up study of patients with phenylketonuria detected by the newborn screening programme in Japan. *J. Inherit. Metab. Dis.* 30, 608.
6. Aulehla-Scholz, C., Heilbronner, H. (2003). Mutational spectrum in German patients with phenylalanine hydroxylase deficiency. *Hum. Mutat.* 21, 399–400.
7. Avigad, S., Cohen, B. E., Bauer, S., Schwartz, G., Frydman, M., Woo, S.L., Niny, Y., Y. Shiloh. (1990). A single origin of phenylketonuria in Yemenite Jews. *Nature.* 344, 168-70.
8. Baranovskaya, S., Shevtsov, S., Maksimova, S., Kuzmin, A., Schwartz, E. (1996). The mutations and VNTRs in the phenylalanine hydroxylase gene of phenylketonuria in St Petersburg. *J. Inherit. Metab. Dis*. 19, 705-709.
9. Bekhof, J., van Spronsen, F.J., Crone, M.R., van Rijn, M., Oudshoorn, C.G., Verkerk, P.H. (2003). Influence of knowledge of the disease on metabolic control in phenylketonuria. *Eur. J. Pediatr.* 162, 440–442.
10. Blau, N., Bonafé, L., Blaskovics, M.E. (2003). Disorders of phenylalanine and tetrahydrobiopterin metabolism. In: N. Blau, M. Duran, M.E. Blaskovics, K.M. Gibson, Scriver CR. (Eds). Physician's Guide to the Laboratory Diagnosis of Metabolic Diseases. Springer-Verlag Berlin Heidelberg New York; 2nd Edition, p.89-106.
11. Blau, N., Erlandsen, H. (2004). The metabolic and molecular bases of tetrahydrobiopterin-responsive phenylalanine hydroxylase deficiency. *Mol. Genet. Metab.* 82, 101–111.
12. Blau, N., van Spronsen, F.J., Levy, H.L. (2010). Phenylketonuria. *Lancet.* 376, 1417–1427.
13. Bodamer, O.A. (2010). Screening for Phenylketonuria. *Ann Nestlé*. 68, 53–57.
14. Bodamer, O.A., Hoffmann, G.F., Lindner, M. (2007). Expanded newborn screening in Europe 2007. *J. Inherit. Metab. Dis.* 30, 439–444.
15. Borrajo, G.J. (2007). Newborn screening in Latin America at the beginning of the 21st century. *J. Inherit. Metab. Dis.* 30, 466–481.

16. Calì, F., Dianzani, I., Desviat, L.R., Perez, B., Ugarte, M., Ozguc, M., Seyrantepe, V., Shiloh, Y., Giannattasio, S., Carducci, C., Bosco, P., De Leo, G., Piazza, A., Romano, V. (1997). The STR252-IVS10nt546- VNTR7 phenylalanine hydroxylase minihaplotype in five Mediterranean samples. *Hum. Genet.* 100(3-4), 350-355.

17. Chakraborty, R, Lidsky, A.S., Daiger, S.P., Guttler, F., Sullivan, S., DiLella, A.G., Woo, S.L.C. (1987). Polymorphic DNA haplotypes at the human phenylalanine hydroxylase locus and their relationship with phenylketonuria. *Hum. Genet.* 76, 40-46.

18. Charikova, E.V., Khalchitskii, S.E., Antoshechkin, A.G., Schwartz, E.I. (1993). Distribution of Some Point Mutations in the Phenylalanine Hydroxylase Gene of Phenylketonuria Patients from the Moscow Region. *Hum. Hered.* 43, 244–249.

19. Chen, K.J., Chao, H.K., Hsiao, K.J., Su, T.S. (2002). Identification and characterization of a novel liver-specific enhancer of the human phenylalanine hydroxylase gene. *Hum. Genet.* 110, 235–243.

20. Clarke, J.T.R. (2002). A clinical guide to inherited metabolic diseases. Cambridge, U.K. ; New York : Cambridge University Press, 2rd Edition, pp. 306.

21. Clarke, J.T.R. (2006). A clinical guide to inherited metabolic diseases. Cambridge, U.K. ; New York : Cambridge University Press, 3rd Edition, pp. 338.

22. Crone, M.R., van Spronsen, F.J., Oudshoorn, K., Bekhof, J., van Rijn, G., Verkerk, P.H. (2005). Behavioural factors related to metabolic control in patients with phenylketonuria. *J. Inherit. Metab. Dis.* 28, 627–637.

23. Daiger, S.P., Chakraborty, R., Reed, L., Fekete, G., Schuler, D., Berenssi, G., Nasz, I., Brdicka, R., Kamaryt, J., Pijakkova, A., Moore, S., Sullivan S., Woo, S.L.C. (1989a). Polymorphic DNA Haplotypes at the Phenylalanine Hydroxylase (PAH) Locus in European Families with Phenylketonuria (PKU). *Am. J. Hum. Genet.* 45, 310-318.

24. Daiger, S.P., Reed, L., Huang, S.S., Zeng, Y.T., Wang, T., Lo, W.H., Okano, Y., Hase, Y., Fukuda, Y., Oura, T. (1989b). Polymorphic DNA haplotypes at the phenylalanine hydroxylase (PAH) locus in Asian families with phenylketonuria (PKU). *Am. J. Hum. Genet.* 45(2), 319-324.

25. Degioanni, A., Darlu, P. (1994). Analysis of the molecular variance at the phenylalanine hydroxylase (PAH) locus. *Eur. J. Hum. Genet.* 2,166-176.

26. Desviat, L.R., Perez, B., Gamez, A., Sanchez, A., Garcia, M.J., Martinez-Pardo, M., Marchante, C., Boveda, D., Baldellou, A., Arena, J., Sanjurjo, P., Fernandez, A., Cabello, M.L., Ugarte, M. (1999). Genetic and phenotypic aspects of phenylalanine hydroxylase deficiency in Spain: molecular survey by regions. *Eur. J. Hum. Genet.* 7, 386–392.

27. Devlin, T.M. (1997). Textbook of Biochemistry with Clinical Correlations. Wiley-Liss, New York, 4th Edition, pp 1186.

28. Dhondt, J.L. (2007). Neonatal screening: from the 'Guthrie age' to the 'genetic age'. *J. Inherit. Metab. Dis.* 30, 418–422.

29. DiLella, A.G., Kwok, S.C., Ledley, F.D., Marvit, J., Woo, S.L. (1986). Molecular structure and polymorphic map of the human phenylalanine hydroxylase gene. *Biochemistry*. 25(4), 743-749.

30. DiLella, A.G., Woo, S.L. (1987). Molecular basis of phenylketonuria and its clinical applications. *Mol. Biol. Med.* 4(4), 183-192.

31. Dworniczak, B., Wedemeyer, N., Eigel, A., Horst, J. (1991(b)) PCR detection of the PvuII (Ea) RFLP at the human phenylalanine hydroxylase (PAH) locus. *Nucleic Acids Res*. 25; 19(8), 1958.

32. Dworniczak, B., Wedemeyer, N., Horst, J. (1991(a)). PCR detection of the Bg1II RFLP at the human phenylalanine hydroxylase (PAH) locus. *Nucleic Acids Res*. 25;19(8), 1958.

33. Eiken, H.G., Knappskog, P.M., Boman, H., Thune, K.S., Kaada, G., Motzfeldt, K., Apold, J. (1996). Relative frequency, heterogeneity and geographic clustering of PKU mutations in Norway. *Eur. J. Hum. Genet.* 4, 205–213.

34. Eisensmith, R.C., Goltsov, A.A., O'Neill, C., Tyfield, L.A., Schwartz, E.I., Kuzmin, A., Baranovskaya, S.S., Tsukerman, G.L., Treacy, E., Scriver, C.R., Guttler, F., Guldberg, P., Eiken, H.G., Apold, J., Svensson, E., Naughten, E., Cahalane, S.F., Croke, D.T., Cockburn, F., Woo, S.L.C. (1995). Recurrence of the R408W Mutation in the Phenylalanine Hydroxylase Locus in Europeans. *Am. J. Hum. Genet.* 56, 278-286.

35. Eisensmith, R.C., Okano, Y., Dasovich, M., Wang, T., Guttler, F., Lou, H., Guldberg, P., Lichter-Konecki, U., Konecki, D.S., Svensson, E., Hagenfeldt, L., Rey, F., Munnich, A., Lyonnet, S., Cockburn, F., Connor, L.M., Pembrey, M.E., Smith, I., Gitzelmann, R., Steinmann, B., Apold, J., Eiken, H.G., Giovannini, M., Riva, E., Longhi, R., Romano, C., Cerone, R., Naughten, E.R., Mullins, C., Cahalane, S., Ozalp, I., Fekete, G., Schuler, D., Berencsi, G.Y., Nasz, I., Brdicka, R., Kamaryt, J., Pijackova, A., Cabalska, B., Boszkowa, K., Schwartz, E., Kalinin, V.N., Jin, L., Chakraborty, R., Woo, S.L.C. (1992). Multiple origins for phenylketonuria in Europe. *Am. J. Hum. Genet*. 51, 1355–1365.

36. Eisensmith, R.C., Woo, S.L.C. (1992). Updated Listing of Haplotypes at the Human Phenylalanine Hydroxylase (PAH) Locus. *Am. J. Hum. Genet*. 51, 1445-1448.

37. Eisensmith, R.C., Woo, S.L.C. (1995). Molecular genetics of phenylketonuria: From molecular anthropology to gene therapy. *Adv. Genet.* 32, 199–271.

38. Erlandsen, H. and Stevens, R.C. (1999). The Structural Basis of Phenylketonuria. *Mol Genet Metab.* 68(2), 103-125.

39. Erlandsen, H., Fusetti, F., Martinez, A., Hough, E., Flatmark, T., Stevens, R. C. (1997). Crystal structure of the catalytic domain of human phenylalanine hydroxylase reveals the structural basis for phenylketonuria. *Nat. Struct. Biol*. 4, 995–1000.

40. Erlandsen, H., Patch, M.G., Gamez, A., Straub, M., Stevens, R.C. (2003). Structural Studies on Phenylalanine Hydroxylase and Implications Toward Understanding and Treating Phenylketonuria. *Pediatrics.* 112,1557-1565.

41. Fernandes, J., Saudubray, J.-M., van den Berghe, G., Walter, J.H. (2006). Inborn metabolic diseases: diagnosis and treatment. Springer Medizin Verlag Heidelberg, 4th Edition, pp.561.

42. Fisher, S. G., Lerman, L. S. (1979). Length-independent separation of DNA restriction fragments in two-dimensional gel electrophoresis. *Cell* 16, 191-200.

43. Fusetti, F., Erlandsen, H., Flatmarki, T., Stevens, R.C. (1998). Structure of Tetrameric Human Phenylalanine Hydroxylase and Its Implications for Phenylketonuria. *J. Biol. Chem.* 273(27), 16962–16967.

44. GeneCards human gene database. 2011. http://www.genecards.org/cgi-bin/carddisp.pl?gene=PAH

45. Gentile, J.K., Hoedt, A.E.T., Bosch, A.M. (2010). Psychosocial aspects of PKU: Hidden disabilities – A review. *Mol. Gen. and Metab.* 99, S64–S67.

46. Gersting, S.W., Kemter, K.F., Staudigl, M., Messing, D.D., Danecka, M.K., Lagler, F.B., Sommerhoff, C.P., Roscher, A.A., Muntau, A.C. (2008). Loss of function in phenylketonuria is caused by impaired molecular motions and conformational instability. *Am. J. Hum. Genet.* 83(1), 5–17.

47. Giannattasio, S, Lattanzio, P., Bobba A., Marra, E. (1997). Detection of microsatellites by ethidium bromide staining. The analysis of an STR system in the human phenylalanine hydroxylase gene. *Mol Cell Probes*. 11, 81-84.

48. Giannattasio, S., Dianzani, I., Lattanzio, P., Spada, M., Romano, V., Calì, F., Andria, G., Ponzone, A., Marra, E., Piazza, A. (2001). Genetic heterogeneity in five Italian regions: analysis of PAH mutations and minihaplotypes. *Hum. Hered.* 52(3), 154-159.

49. Giannattasio, S., Jurgelevičius, V., Lattanzioa, P., Cimbalistienė, L., Marraa, E., Kučinskas, V. (1997). Phenylketonuria Mutations and Linked Haplotypes in the Lithuanian Population: Origin of the Most Common R408W Mutation. *Hum. Hered.* 47, 155-160.

50. Goltsov, A.A., Eisensmith, R.C., Woo, S.L. (1992a). Detection of the XmnI RFLP at the human PAH locus by PCR. *Nucleic Acids Res*. 25;20(4), 927.

51. Goltsov, A.A., Eisensmith, R.C., Koneckit, D.S., Lichter-Konecki U., Woo, S.L.C. (1992b). Associations between Mutations and a VNTR in the Human Phenylalanine Hydroxylase Gene. *Am. J. Hum. Genet*. 51, 627-636.

52. Goltsov, A.A., Eisensmith, R.C., Naughten, E., Jin, L., Chakraborty, R., Woo, S.L.C. (1993). A single polymorphic STR system in the human phenylalanine hydroxylase gene

permits rapid prenatal diagnosis and carrier screening for phenylketonuria. *Hum. Mol. Genet.* 2(5), 577-581.

53. Güttler, F. (1980). Hyperphenylalaninemia diagnosis and classification of the various types of phenylalanie hydroxylase defi ciency in childhood. *Acta Pediat. Scand.* 280: 1–80.

54. Guldberg, P., Guttler, F. (1994). ''Broad-range'' DGGE for single-step mutation scanning of entire genes: application to human phenylalanine hydroxylase gene. *Nucleic Acids Res*. 22, 880–881.

55. Guldberg, P., Henriksen, K.F., Guttler, F. (1993a). Molecular analysis of phenylketonuria in Denmark: 99% of the mutations detected by denaturing gradient gel electrophoresis. *Genomics.* 17, 141–146.

56. Guldberg, P., Henriksen, K.F., Sipila, I., Guttler, F., de la Chapelle, A. (1995). Phenylketonuria in a low incidence population: molecular characterisation of mutations in Finland. *J. Med. Genet.* 32, 976–978.

57. Guldberg, P., Levy, H. L., Hanley, W. B., Koch, R., Matalon, R., Rouse, B. M., Trefz, F., de la Cruze, F., Henriksen, K. F., and F. Güttler (1996). Phenylalanine hydroxylase gene mutations in the United States: report from the maternal PKU collaborative study. *Am. J. Hum. Genet*. 59, 84-94.

58. Guldberg, P., Mallmann, R., Henriksen, K.F., Guttler, F. (1996a). Phenylalanine hydroxylase deficiency in a population in Germany: mutational profile and nine novel mutations. *Hum. Mutat*. 8, 276–279.

59. Guldberg, P., Zschocke, J., Dagbjartsson, A., Henriksen, K.F., Güttler, F. (1997). A molecular survey of phenylketonuria in Iceland: identification of a founding mutation and evidence of predominant Norse settlement. *Eur. J. Hum. Genet.* 5(6), 376-381.

60. Guldberg, P., Rey, F., Zschocke, J., Romano, V., François, B., Michiels, L., Ullrich, K., Hoffmann, G.F., Burgard, P., Schmidt, H., Meli, C., Riva, E., Dianzani, I., Ponzone, A., Rey, J., Güttler, F. (1998). A European multicenter study of phenylalanine hydroxylase deficiency: classification of 105 mutations and a general system for genotype-based prediction of metabolic phenotype. *Am. J. Hum. Genet.* 63, 71-79.

61. Guthrie, R. (1996). The introduction of newborn screening for phenylketonuria. A personal history. *Eur. J. Pediatr.* 155(1),S4–S5.

62. Hannigan, S. (2007). Inherited metabolic diseases: a guide to 100 conditions. Radcliffe Publishing, pp.167.

63. Harding, C.O., Blau, N. (2010). Advances and challenges in phenylketonuria. *J. Inherit. Metab. Dis.* 33, 645–648.

64. Hennermann, .J.B, Vetter, B., Wolf, C., Windt, E., Buhrdel, P., Seidel, J., Monch, E., Kulozik, A.E. (2000). Phenylketonuria and hyperphenylalaninemia in eastern Germany: a characteristic molecular profile and 15 novel mutations. *Hum. Mutat.* 15, 254–260.

65. Horne, J., Jennings, I.G., Teh, T., Gooley, P.R., Kobe, B. (2002). Structural characterization of the N-terminal autoregulatory sequence of phenylalanine hydroxylase. *Protein Sci.* 11(8): 2041–2047.

66. Hosseini-Mazinani, S.M., Koochmeshgi, J., Khazaee-Koohpar, Z., Hosein-Pur-Nobari, N., Seifati, S.M. (2008). Carrier detection of phenylketonuria in Iranian families by variable number tandem-repeat polymorphism analysis. *East Mediterr. Health J.* 14(6), 1445-1451.

67. Jaruzelska, J., Matuszak, R., Lyonnet, S., Rey, F., Rey, J., Filipowicz, J., Borski, K., Munnich, A. (1993). Genetic background of clinical homogeneity of phenylketonuria in Poland. *J. Med. Genet.* 30, 232-234.

68. Jennings, I.G., Cotton R.G.H., Kobe, B. (2000). Structural interpretation of mutations in phenylalanine hydroxylase protein aids in identifying genotype–phenotype correlations in phenylketonuria. *Eur. J. Hum. Genet.* 8, 683–696.

69. Jiang, J., Ma X, Huang, X., Pei, X., Liu, H., Tan, Z., Zhu, L. (2003). A survey for the incidence of phenylketonuria in Guangdong, China. *Southeast Asian J. Trop. Med. Public Health*, 34 (3): 185.

70. Kadasi, L., Polakova, H., Ferakova, E., Hudecova, S., Bohusova, T., Szomolayova, I., Strnova, J., Hruskovic, I., Moschonas, N.K., Ferak, V. (1995). PKU in Slovakia: mutation screening and haplotype analysis. *Hum. Genet.* 95, 112–114.

71. Kamkar, M., Saadat, M., Saadat, I., Haghighi, G. (2003). Report of VNTR with 13 repeats linked to PAH locus in unaffected members of two PKU families. Iran *Biomed J.* 7(2), 89-90.

72. Kasnauskiene, J., Giannattasio, S., Lattanzio, P., Cimbalistiene, L., Kucinskas, V. (2003). The molecular basis of phenylketonuria in Lithuania. *Hum. Mutat.* 21(4), 398-401.

73. Kayaalp, E., Treacy, E., Waters, P.J., Byck, S., Nowacki, P., Scriver, C.R. (1997). Human phenylalanine hydroxylase mutations and hyperphenylalaninemia phenotypes: a metanalysis of genotype–phenotype correlations. *Am. J. Hum. Genet.* 61, 1309–1317.

74. Kidd, J.R, Pakstis, A.J., Zhao, H., Lu, R.B., Okonofua, F.E., Odunsi, A., Grigorenko, E., Bonne-Tamir, B., Friedlaender, J., Shulz, L.O., Parnas, J., Kidd, K.K. (2000). Haplotypes and linkage disequilibrium at the phenylalanine hydroxylase locus, PAH, in a global representation of populations. *Am. J. Hum. Genet.* 66, 1882-1899.

75. Kidd, J.R., Kidd K.K. (2005). The Population Genetics of PAH. New York: McGraw-Hill. Revised April, 2008, from MMBID Online

http://www.medgen.mcgill.ca/scriver/pah/Update/UpdateChapter77-ThePopulation GeneticsofPAH.html. DOI: http://dx.doi.org/10.1036/ommbid.100

76. Killeen, A.A., Rubin, E., Strayer, D.S. (2008). Developmental and genetic diseases. In McDonald J.M., Michalopoulos G.K., Trojanowski J.Q., Ward P.A. (Ed.), Rubin's pathology (pp.177-229). Lippincott Williams & Wilkins.

77. Konecki, D.S., Wang, Y., Trefz, F.K., Lichter-Konecki, U., Woo, S.L. (1992). Structural characterization of the 5' regions of the human phenylalanine hydroxylase gene. *Biochemistry.* 31(35), 8363-8368.

78. Kozak, L., Blazkova, M., Kuhrova, V., Pijackova, A., Ruzickova, S., Stastna, S. (1997). Mutation and haplotype analysis of phenylalanine hydroxylase alleles in classical PKU patients from the Czech Republic: identification of four novel mutations. *J. Med. Genet.* 34, 893–898.

79. Kozak, L., Hrabincova, E., Kintr, J., Horky, O., Zapletalova, P., Blahakova, I., Mejstrik, P., Prochazkova, D. (2006). Identification and characterization of large deletions in the phenylalanine hydroxylase (PAH) gene by MLPA: evidence for both homologous and non-homologous mechanisms of rearrangement. *Mol. Genet. Metab.* 89, 300–309.

80. Kure, S., Hou, D.C., Ohura, T., Iwamoto, H., Suzukim, S., Sugiyama, N., Sakamoto, O., Fujii, K., Matsubara, Y., Narisawa, K. (1999). Tetrahydrobiopterin-responsive phenylalanine hydroxylase defi ciency. *J. Pediatr.* 135, 375–378.

81. Kwok, S.C., Ledley, F.D., DiLella, A.G., Robson, K.J., Woo, S.L. (1985). Nucleotide sequence of a full-length complementary DNA clone and amino acid sequence of human phenylalanine hydroxylase. *Biochemistry*. 29;24(3),556-561.

82. Latorra, D., Stern, C.M., Schanfield, M.S. (1994). Characterization of human AFLP systems apolipoprotein B, phenylalanine hydroxylase, and D1S80. *PCR Methods Appl.* 3(6), 351-358.

83. Lichter-Konecki, U., Schlotter, M., Konecki, D.S. (1994). DNA sequence polymorphisms in exonic and intronic regions of the human phenylalanine hydroxylase gene aid in the identification of alleles. *Hum Genet*. 94(3), 307-310.

84. Lidsky, A.S., Ledley, F.D., DiLella, A.G., Kwok, S.C., Daiger, S.P., Robson, K.J., Woo, S.L. (1985). Extensive restriction site polymorphism at the human phenylalanine hydroxylase locus and application in prenatal diagnosis of phenylketonuria. *Am. J. Hum. Genet*. 37(4), 619-634.

85. Lillevali, H., Ounap, K., Metspalu, A. (1996). Phenylalanine hydroxylase gene mutation R408W is present on 84% of Estonian phenylketonuria chromosomes. *Eur. J. Hum. Genet*. 4, 296–300.

86. Loeber, J.G. (2007). Neonatal screening in Europe; the situation in 2004. *J. Inherit. Metab. Dis.* 30, 430–438.

87. MacDonald, M.J., D'Cunha, G.B. (2007). A modern view of phenylalanine ammonia lyase. *Biochem. Cell Biol.* 85, 273–282.

88. Matalon, R., Michals-Matalon, K., Bhatia, G., Burlina, A.B., Burlina, A.P., Braga, C., Fiori, L., Giovannini, M., Grechanina, E., Novikov, P., Grady, J., Tyring, S.K., Guttler, F. (2007). Double blind placebo control trial of large neutral amino acids in treatment of PKU: eff ect on blood phenylalanine. *J. Inherit. Metab. Dis.* 30, 153–158.

89. McGraw-Hill. (2003). McGraw-Hill Dictionary of Scientific & Technical Terms. Publisher: McGraw-Hill Professional, 6th Edition, pp. 2380.

90. McPhee, S.J., Hammer, G.D. (2010). Pathophysiology of Disease: An Introduction to Clinical Medicine. McGraw-Hill Medical, 6th Edition, pp. 752.

91. Meijer, H., Jongbloed, R.J., Hekking, M., Spaapen, L.J., Geraedts, J.P. (1993). RFLP haplotyping and mutation analysis of the phenylalanine hydroxylase gene in Dutch phenylketonuria families. *Hum. Genet.* 92, 588–592.

92. Miranda, F.F., Teigen, K., Thórólfsson, M., Svebak, R.M., Knappskog, P.M., Flatmark, T., Martínez, A. (2002). Phosphorylation and mutations of Ser(16) in human phenylalanine hydroxylase. Kinetic and structural effects. *J. Biol. Chem.* 25;277(43),40937-40943.

93. MOLGENT. (1999-2002). FP 4 European Commission Project INCO-Copernicus "Molecular genetic testing in phenylketonuria: a model to assess the quality control system for monogenic disease". European Commission. An overview of Public Health Research projects 1994-2002. (http://ec.europa.eu/research/biosociety/pdf/health.pdf). p.67.

94. Murphy, B.C., Scriver, C.R., Singh, S.M. (2006). CpG Methylation Accounts for a Recurrent Mutation (c.1222C>T) in the Human PAH Gene. *Hum. Mutat.* 27(9), 975-980.

95. Nechyporenko, M.V., Livshits, L.A. (2002). Molecular and genetical study of phenylketonuria in Ukraine (abstract). *Eur. J. Hum. Genet.* 10(1), 236.

96. Nelson, D.L., Cox, M. M. (2000). Lehninger, Principles of Biochemistry. Worth Publishing: New York, 3rd Edition, pp. 1200.

97. Nyhan, W.L., Barshop, B.A., Ozand, P.T. (2005). Atlas of Metabolic Diseases. Hodder Arnold: London, 2nd Edtiton, pp. 788.

98. O'Donnell, K.A., O'Neill, C., Tighe, O., Bertorelle, G., Naughten, E., Mayne, P.D., Croke, D.T. (2002). The mutation spectrum of hyperphenylalaninaemia in the Republic of Ireland: the population history of the Irish revisited. *Eur. J. Hum. Genet.* 10(9), 530-538.

99. Okano, Y., Eisensmith, R.C., Güttler, F., Lichter-Konecki, U., Konecki, D.S., Trefz, F.K., Dasovich, M., Wang, T., Henriksen, K., Lou, H., Woo, S.L.C. (1991). Molecular Basis of Phenotypic Heterogeneity in Phenylketonuria. *N. Engl. J. Med.* 324, 1232-1238.

100. Olsson, G.M., Montgomery, S.M., Alm, J. (2007). Family conditions and dietary control in phenylketonuria. *J. Inherit. Metab. Dis.* 30, 708–715.

101. Õunap, K., Lilleväli, H., Metspalu, A., Lipping-Sitska, M. (1998). Development of the phenylketonuria screening programme in Estonia. *J. Med. Screen.* 5, 22-23.

102. Ozalp, I., Coskun, T., Tokatli, A., Kalkanoğlu, H.S., Dursun, A., Tokol, S., Köksal, G., Ozgüc, M., Köse, R. (2001). Newborn PKU screening inTurkey: at present and organization for future. *Turk. J. Pediatr.* 43, 97–101.

103. PAH Phenylalanine Hydroxylase Locus Specific Database. PAHdb: www.pahdb.mcgill.ca

104. Pangkanon, S., Charoensiriwatana, W., Janejai, N., Boonwanich, W., Chaisomchit, S. (2009).Detection of phenylketonuria by the newborn screening program in Thailand. *Southeast Asian J. Trop. Med. Public Health.* 40, 525–529.

105. Perez, B., Desviat, L.R., Ugarte, M. (1997). Analysis of phenylalanine hydroxylase gene in the Spanish population: Mutation profile and association with intragenic polymorphic markers. *Am. J. Hum. Genet.* 60, 95–102.

106. Pey, A.L., Pérez, B., Desviat, L.R., Martínez, M.A., Aguado, C., Erlandsen, H., Gámez, A., Stevens, R.C., Thórólfsson, M., Ugarte, M., Martínez, A. (2004). Mechanisms underlying responsiveness to tetrahydrobiopterin in mild phenylketonuria mutations. *Hum. Mutat.* 24(5), 388–399.

107. Phenylketonuria (PKU): Screening and Management. NIH Consensus Statement 2000 October 16-18; 17(3): 1–33.

108. Pietz, J., Kreis, R., Rupp, A., Mayatepek, E., Rating, D., Boesch, C., Bremer, H.J. (1999). Large neutral amino acids block phenylalanine transport into brain tissue in patients with phenylketonuria. *J. Clin. Invest.* 103, 1169–1178.

109. Purina, G., Lugovska, R., Sokolova, L. (1995). Medical genetical service in Latvia: developmental trends. *Proceed. Latv. Acad. Sci.* 5/6, 105–108.

110. Rivera, I., Leandro, P., Lichter-Konecki, U., Tavares de Almeida, I., Lechner, M.C. (1998). Population genetics of hyperphenylalaninaemia resulting from phenylalanine hydroxylase deficiency in Portugal. *J. Med. Genet.* 35, 301–304.

111. Sarkissian, C.N., Gamez, A., Wang, L., Charbonneau, M., Fitzpatrick, P., Lemontt, J.F., Zhao, B., Vellard, M., Bell, S.M., Henschell, C., Lambert, A., Tsuruda, L., Stevens, R.C., Scriver, C.R. (2008). Preclinical evaluation of multiple species of PEGylated recombinant phenylalanine ammonia lyase for the treatment of phenylketonuria. *Proc. Natl. Acad. Sci. USA.* 52, 20894–20899.

112. Schindeler, S., Ghosh-Jerath, S., Thompson, S., Rocca, A., Joy, P., Kemp, A., Rae, C., Green, K., Wilcken, B., Christodoulou, J. (2007). The eff ects of large neutral amino

acid supplements in PKU: an MRS and neuropsychological study. *Mol. Genet. Metab.* 91, 48–54.

113. Schuett, V. (2009). Celebrating 75 Years Since the Discovery of PKU. National PKU News Birmingham Children's Hospital in England. 21(2).
114. Scott, T. A. (1988). Chemistry and biochemistry of the amino acids. Barrett, G.C. (Ed.). Chapman and Hall, London, pp. 684.
115. Scriver, C.R. (2007). The PAH Gene, Phenylketonuria, and a Paradigm Shift. *Hum. Mutat* . 28(9), 831-845.
116. Scriver, C.R., Beaudet, A.L., Sly, W.S., Valle, D. (1995). The metabolic basis of inherited disease. Stanbury, J.B., Wyngaarden, J.B., Fredrickson, D.S. (Ed.). McGraw-Hill Information Services Co., New York, 7th Edition, pp. 3006.
117. Scriver, C.R., Hurtubise, M., Konecki, D., Phommarinh, M., Prevost, L., Erlandsen, H., Stevens, R., Waters, P.J., Ryan, S., McDonald, D., Sarkissian, C. (2003). PAHdb 2003: what a locus-specific knowledgebase can do. *Hum. Mutat.* 21(4),333-44.
118. Scriver, C.R., Levy, H., Donlon, J. (2008). Hyperphenylalaninemia: Phenylalanine Hydroxylase Deficiency. In: The Online Metabolic and Molecular Bases of Inherited Diseases. Valle, D., Beaudet, A. L., Vogelstein, B., Kinzler, K.W., Antonarakis, S.E., Ballabio, A., Scriver, C.R., Sly, W.S., Childs, B.(Ed.), chapter 77. http://www.ommbid.com/OMMBID/the_online_metabolic_and_molecular_bases_of_inherited_disease/b/abstract/part8/ch77
119. Scriver, CR, Kaufman S. (2001). Hyperphenylalaninemia: phenylalanine hydroxylase deficiency. In: Scriver CR, Beaudet AL, Sly WS, Valle D, Childs B, Kinzler K, Vogelstein B, editors. The Metabolic and Molecular Bases of Inherited Disease. New York: McGraw-Hill; 8th Edition, pp. 1667-1724.
120. Stojiljković, M., Jovanović, J., Djordjević, M., Grković, S., Cvorkov, Dražić, M., Petručev, B., Tošić, N., Karan Djurašević, T., Stojanov, Lj., and S. Pavlović (2006). Molecular and phenotypic characteristics of patients with phenylketonuria in Serbia and Montenegro. *Clin. Genet.* 70(2), 151-155.
121. Stojiljković, M., Stevanović, A., Đorđević, M., Petručev, B., Tošić, N., Karan-Đurašević, T., Aveić, S., Radmilović, M., Pavlović, S. (2007). Mutations in the PAH gene: A Tool for population genetics study. *Arch. Biol. Sci.* 59(3), 161-167.
122. Sueoka, H., Moshinetsky, A., Nagao, M., Chiba, S. (1999). Mutation screening of phenylketonuria in the Far East of Russia. *J. Hum. Genet.* 44(6), 368-371.
123. Svensson, E., von Dobeln, U., Eisensmith, R.C., Hagenfeldt, L., Woo, S.L. (1993). Relation between genotype and phenotype in Swedish phenylketonuria and hyperphenylalaninemia patients. *Eur. J. Pediatr.* 152, 132–139.

124. The Biology Project. Department of Biochemistry and Molecular Biophysics, University of Arizona, August 25, 2003, The Biology Project – Biochemistry: Phenylalanine F (Phe). http://www.biology.arizona.edu/biochemistry/problem_sets/aa/phenylalanine.html

125. The National Center for Biotechnology Information (NCBI). GenBank: U49897.1: Homo sapiens phenylalanine hydroxylase (PAH) mRNA, complete cds. http://www.ncbi.nlm.nih.gov/nuccore/U49897?report=GenBank.

126. Tighe, O., Dunican, D., O'Neill, C., Bertorelle, G., Beattie, D., Graham, C., Zschocke, J., Cali, F., Romano, V., Hrabincova, E., Kozak, L., Nechyporenko, M., Livshits, L., Guldberg, P., Jurkowska, M., Zekanowski, C., Perez, B., Desviat, R.L., Ugarte, M., Kucinskas, V., Knappskog, P., Treacy, E., Naughten, E., Tyfield, L., Byck, S., Scriver, C.R., Mayne, P.D., Croke, D.T. (2003). Genetic diversity within the R408W Phenylketonuria mutation lineages in Europe. *Hum. Mutat.* 21, 387-393.

127. Traeger-Synodinos, J., Kanavakis, E., Kalogerakou, M., Soulpi, K., Missiou-Tsangaraki, S., Kattamis, C. (1994). Preliminary mutation analysis in the phenylalanine hydroxylase gene in Greek PKU and HPA patients. *Hum. Genet.* 94, 573–575.

128. Tyfield, L.A., Stephenson, A., Cockburn, F., Harvie, A., Bidwell, J.L., Wood, N.A., Pilz, D.T., Harper, P., Smith, I. (1997). Sequence variation at the phenylalanine hydroxylase gene in the British Isles. *Am. J. Hum. Genet.* 60(2), 388-396.

129. van der Sijs-Bos, C.J., Diepstraten, C.M., Juyn, J.A., Plaisier, M., Giltay, J.C., van Spronsen, F.J., Smit, G.P., Berger, R., Smeitink, J.A., Poll-The, B.T., Ploos van Amstel, J.K. (1996). Phenylketonuria in The Netherlands: 93% of the mutations are detected by single-strand conformation analysis. *Hum. Hered.* 46, 185–190.

130. Verelst, P., Francois, B., Cassiman, J.J., Raus, J. (1993). Heterogeneity of Phenylketonuria in Belgium. *Develop. Brain Dysfun.* 6, 97–108.

131. Verkerk, P.H., van Spronsen, F.J., Smit, G.P., Cornel, M.C., Kuipers, J.R., Verloove-Vanhorick, S.P. (1991). Prevalence of congenital heart disease in patients with phenylketonuria. *J. Pediatr.* 119(2):282-283.

132. Villas-Bôas, S.G., Nielsen, J., Smedsgaard, J., Hansen, M.A.E. (2007). Metabolome analysis: an introduction. Wiley-Interscience, pp.311.

133. Walter, J.H., White, F.J. (2004). Blood phenylalanine control in adolescents with phenylketonuria. *Int. J. Adolesc. Med. Health.* 16, 41–45.

134. Weber, J.L., Wong, C. (1993). Mutation of human short tandem repeats. *Hum. Mol. Genet.* 2, 1123-1128.

135. Wedemeyer, N., Dworniczak, B., Horst, J. (1991). PCR detection of the MspI (Aa) RFLP at the human phenylalanine hydroxylase (PAH) locus. *Nucleic Acids Res.* 25; 19(8), 1959.

136. Williams, R.A., Mamotte, J.R. Burnett, C.D.S., (2008). Phenylketonuria: An Inborn Error of Phenylalanine Metabolism. *Clin. Biochem. Rev.* 29(1), 31–41.

137. Woo, S.L.C., Lidsky, A., Guttler, F., Chandra, T., Robson, K. (1983). Cloned human phenylalanine hydroxylase gene allows prenatal diagnosis and carrier detection of classical phenylketonuria. *Nature*. 306, 151-155.

138. Zekanowski, C., Jurkowska, M., Bal, J. (2001). Association between minihaplotypes and mutations at the PAH locus in Polish hyperphenylalaninemic patients. *Hum. Hered.* 51, 117–120.

139. Zhan, J.Y., Qin,Y.F., Zhao, Z.Y. (2009). Neonatal screening for congenital hypothyroidism and phenylketonuria in China. *World J. Pediatr.* 5, 136–39.

140. Zschocke, J. (2003). Phenylketonuria mutations in Europe. *Hum. Mutat.* 21, 345–356.

141. Zschocke, J. (2003)a. Focus on the molecular genetics of phenylketonuria. *Hum. Mutat.* 21(4), 331–332.

142. Zschocke, J., Graham, C.A., Carson, D.J., Nevin, N.C. (1995). Phenylketonuria mutation analysis in Northern Ireland: a rapid stepwise approach. *Am. J. Hum. Genet*. 57(6), 1311-1317.

143. Zschocke, J., Graham, C.A., McKnight, J.J., Nevin, N.C. (1994(a)). The STR system in the human phenylalanine hydroxylase gene: true fragment length obtained with fluorescent labelled PCR primers. *Acta Paediatr. Suppl.* 407, 41-42.

144. Zschocke, J., Hoffmann, G.F. (1999). Phenylketonuria mutations in Germany. *Hum. Genet.* 104(5), 390-398.

145. Zschocke, J., Knauer, A., Sarnavka, V., Fumic, K., Mardesic, D., Hoffmann, G.F., Baric, I. 2003(b). The molecular basis of phenylalanine hydroxylase deficiency in Croatia. *Hum. Mutat.* 21, 399.

146. Zschocke, J., Mallory, J.P., Eiken, H.G., Nevin, N.C. (1997). Phenylketonuria and the peoples of Northern Ireland. *Hum. Genet.* 100, 189–194.

147. Zurflüh, M.R., Zschocke, J., Lindner, M., Feillet, F., Chery, C., Burlina, A., Stevens, R.C., Thöny, B., Blau, N. (2008). Molecular genetics of tetrahydrobiopterin responsive phenylalanine hydroxylase deficiency. *Hum. Mutat.* 29, 167–175.

148. Zurfluh, M.R., Zschocke, J., Lindner, M., Feillet, F., Chery, C., Burlina, A., Stevens, R.C., Thony, B., Blau, N. (2008). Molecular Genetics of Tetrahydrobiopterin-Responsive Phenylalanine Hydroxylase Deficiency. *Hum. Mutat*. 29(1),167-175.

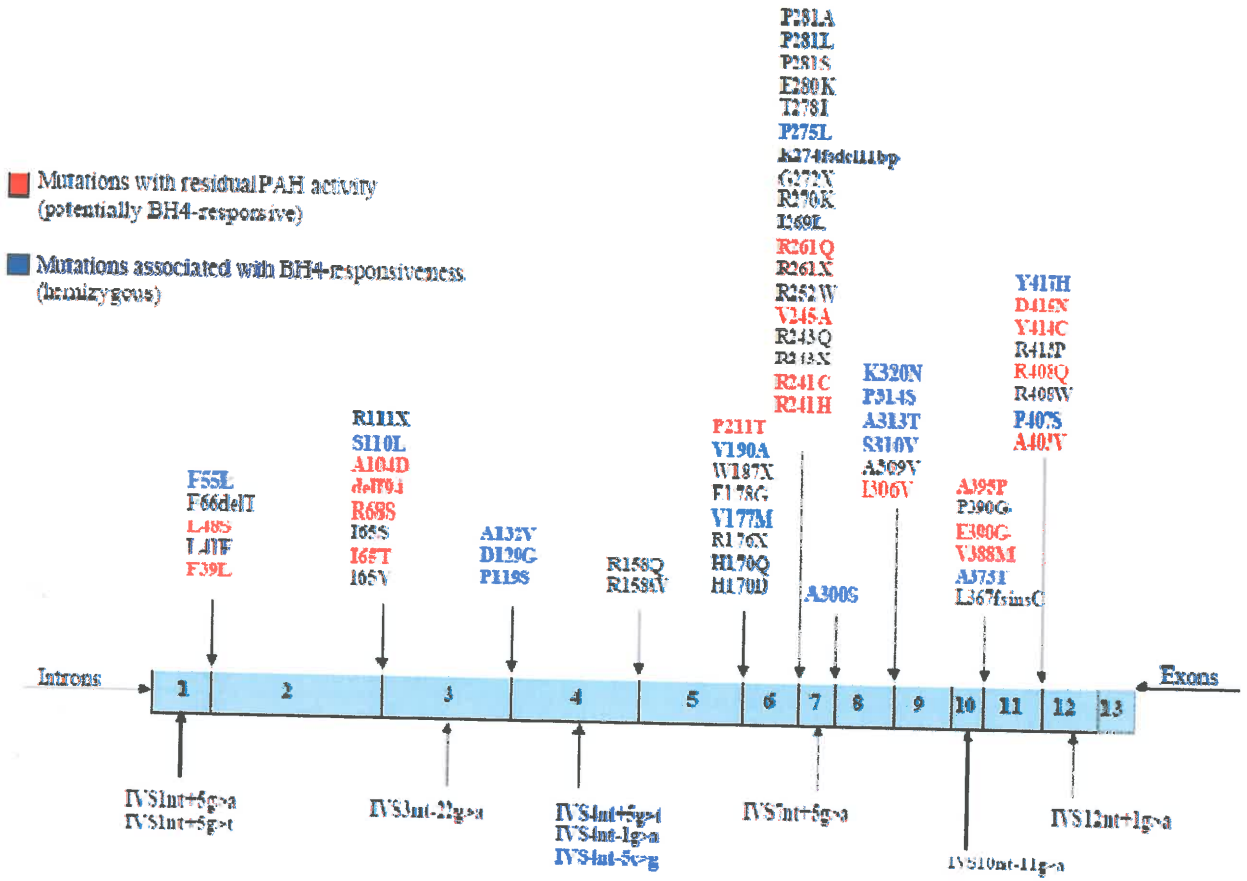
149. Zygulska, M., Eigel, A., Aulehla-Scholz, C., Pietrzyk, J.J., Horst J. (1991). Molecular analysis of PKU haplotypes in the population of southern Poland. Hum Genet. 86(3), 292-294.

APPENDIX

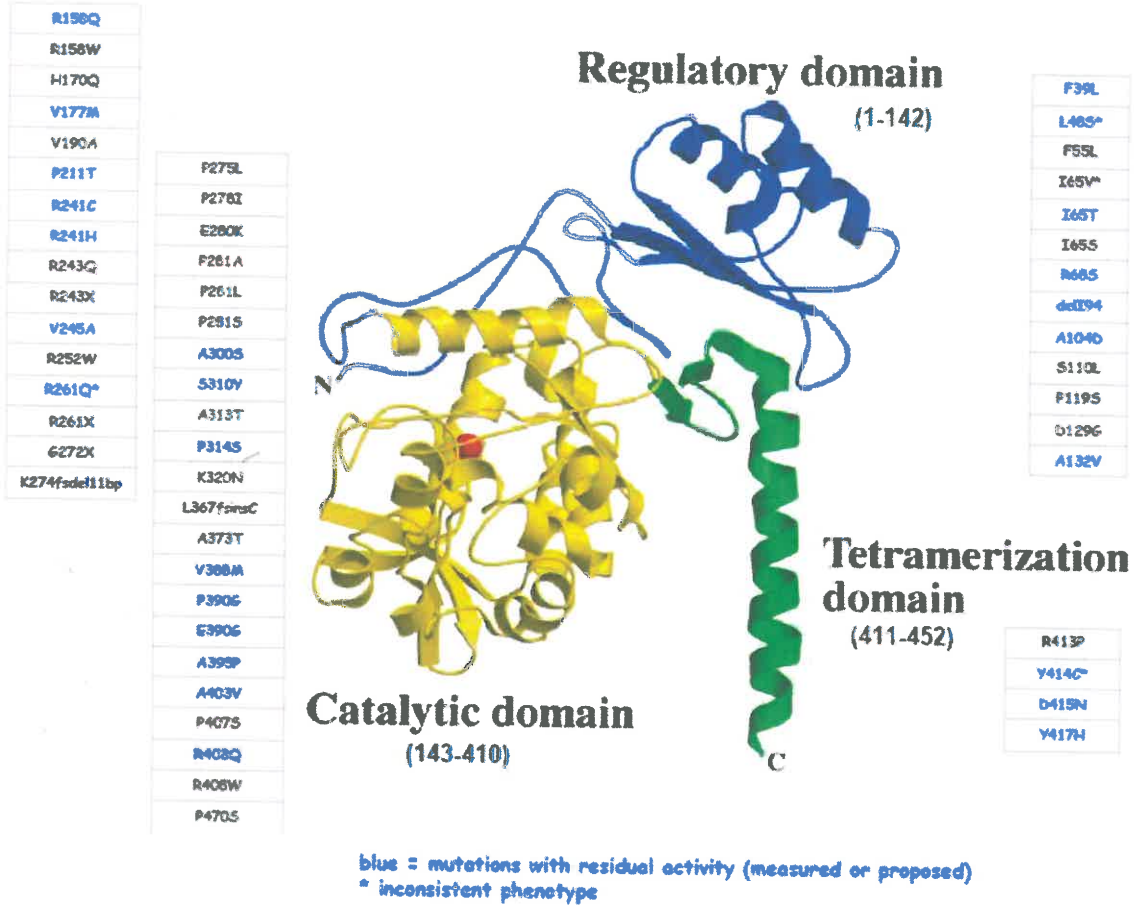
Appendix I. Classification of *PAH* gene mutations according to their caused metabolic phenotype (modified from Guldberg, 1998).

Classic or severe PKU (AV=1)		**Moderate PKU (AV=2)**	**Mild PKU (AV=4)**	**MHP (AV=8)**
M1V	E280K	F39L	(F39L)	A47V
Q20X	P281L	(L48S)	G46S	S87R
IVS1nt5g→t	IVS7nt1g→a	(IVS2nt5g→c)	L48S	T92I
(F39L)	IVS7nt5g→a	I65T	T63P/H64N	R155H
(L48S)	D282N	(R68S)	(I65T)	G171A
F55L	H285Y	(R158Q)	R68S	R176L
F55fsdelT	**S295X**	**E6nt-96A→g**	**A104D**	**E178G**
IVS2nt5g→c	F299C	R261P	IVS4nt-5c→g	V190A
(I65T)*	IVS8nt1g→a	R261Q	I164T	V230I
D84Y	IVS8nt-7a→g	(L311P)	V177A	R241C
P89fsinsC	S310fsdel11bp	L348V	R241H	V245A
I94S	L311P	V388M	A246V	A300S
R111X	F327L	(Y414C)	(R261P)	I306V
R158Q	F331L		Y277D	T380M
I174T	Q336X		G344S	E390G
R176X	A342T		(E390G)	A403V
W187X	A342fsdelG		R408Q	R413S
L194P	G346fsdelG		Y414C	(Y414C)
L197fsdel22bp	G346R			D415N
Y198fsdel22bp	(L348V)			
Y204X	S349P			
Y206X	G352R			
E221_D222fsdelAG	IVS10nt-11g→a			
S231P	IVS10nt-3c→t			
G239S	IVS10nt-1g→a			
R243Q	Y356X			
R243X	S359X			
R252G	K363fsdelG			
R252Q	R367fsinsC			
R252W	A395P			
A259V	IVS11nt-1g→a			
(R261Q)	P407fsdelC			
R261X	R408W			
I269N	(Y414C)			
G272X	IVS12nt1→a			
K274fsdel11bp	K452fsinsA			

* small number of mutations were reported in different phenotype than the majority



Appendix II. Mutations detected in patients with BH4-responsive HPA/PKU (source BIOPKU database; http://www.bh4.org/biopku.html and PAHdb; http://www.pahdb.mcgill.ca).



Appendix III. Mutations detected in patients with (BH4)-responsive PAH deficiency (*PAH* Source: Erlandsen and Stevens, 1999; www.bh4.org/biopku.html; Blau and Erlandsen, 2004).

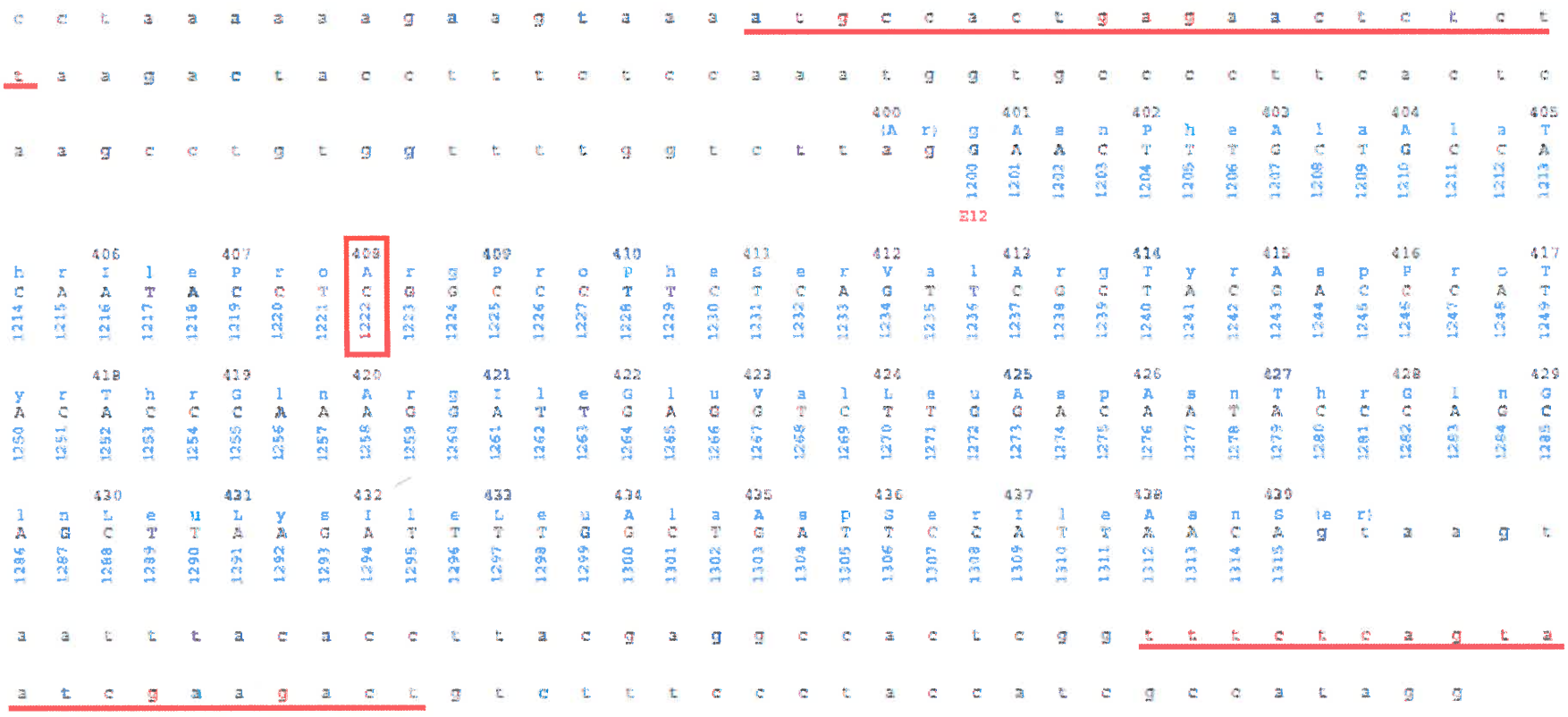
Appendix IV. BH4-responsive mutations with known residual activity (source: Zurfluh, 2008).

No.	Allele	Nucleotide aberration	Residual activity (%)	Domain (CBR) (AR)
1.	p.A403V	c.1208C>T	32	Catalytic
2.	p.R261Q	c.782G>A	38.5	Catalytic (CBR1)
3.	p.Y414C	c.1241A>G	36	Tetramerization
4.	p.A300S	c.898G>T	31	Catalytic
5.	p.V245A	c.734T>C	50	Catalytic (CBR1) (AS)
6.	p.L48S	c.143T>C	39	Regulatory
7.	p.E390G	c.1169A>G	72.5	Catalytic
8.	p.R241C	c.721C>T	25	Catalytic
9.	p.I65T	c.194T>C	25.3	Regulatory
10.	p.R158Q	c.473G>A	10	Catalytic
11.	p.V388M	c.1162G>A	27.5	Catalytic
12.	p.D415N	c.1243G>A	93	Tetramerization
13.	p.R408Q	c.1223G>A	49.7	Catalytic
14.	p.R243Q	c.728G>A	23	Catalytic
15.	p.R413P	c.1238G>C	66	Tetramerization
16.	p.E178G	c.533A>G	39	Catalytic
17.	p.F39del	c.115 117 delTTC	20	Regulatory
18.	p.F39L	c.117C>G	49	Regulatory
19.	p.R68S	c.204A>T	87	Regulatory
20.	p.A395P	c.1183G>C	15.5	Catalytic
21.	p.L348V	c.1042C>G	41	Catalytic
22.	p.P211T	c.631C>A	72	Catalytic
23.	p.P407S	c.1219C>T	94	Catalytic
24.	p.R241H	c.722G>A	23	Catalytic
25.	p.A104D	c.311C>A	26	Regulatory
26.	p.A309V	c.926C>T	44	Catalytic
27.	p.V230I	c.688G>A	63	Catalytic
28.	p.H170D	c.508C>G	43	Catalytic
29.	p.I94del	c.283_285delATC	27	Regulatory
30.	p.T92I	c.275C>T	76	Regulatory
31.	p.V190A	c.569T>C	110	Catalytic
32.	p.A313T	c.937G>A	76	Catalytic
33.	p.A373T	c.1117G>A	56	Catalytic
34.	p.E76G	c.227A>G	47	Regulatory
35.	p.L308F	c.922C>T	49	Catalytic
36.	p.L41F	c.121C>T	10	Regulatory
37.	p.P122Q	c.365C>A	22	Regulatory
38.	p.P244L	c.731C>T	51	Catalytic
39.	p.S87R	c.261C>A	82	Regulatory
40.	p.R176L	c.527G>T	31.5	Catalytic

CBR, cofactor binding region; AS, active site.

Appendix V. *PAH* gene mutations association with minihaplotypes in different populations.

Mutations	Minihaplotypes					
	Germany	**Poland**	**Italy**	**Spain**	**Northern Ireland**	**Ireland**
R408W	3/238 (117) 9/242 (1) 9/246 (1)	3/238 (83) 3/242 (6) 8/238 (1) 8/242 (1)	3/238 (2)	—	8/242 (58) 3/238 (5) 8/238 (2) 8/230 (1) 3/242 (1)	8/242 (49) 8/238 (9) 8/246 (2)
P281L	7,8/242 (23)	8/242 (2) 3/242 (1)	7/242 (3) 8/238 (2) 8/234 (1) 3/234 (1)	7/242 (2) 8/242 (1)	—	—
E280K	7/246 (1)	—	—	9/234 (3)	8/238 (3) 8/246 (1)	—
R261Q	8/238, 242 (30) 3/246 (2) 8/234 (1)	—	8/238 (19) 7/242 (2)	8/238 (2) 8/246 (1)	8/238 (2) 3/246 (1)	—
R261X	3/238, 246 (3)	—	7/242 (5) 3/238 (3) 3/242 (1)	—	—	—
R158Q	3/234 (24)	—	3/234 (5)	—	3/234 (1)	—
A403V	8/242 (9)	8/242 (5) 8/246 (2)	8/242 (4) 8/246 (1)	8/246 (2) 8/242 (1) 8/238 (1)	—	—
L48S	3/234 (15)	—	3/234 (12) 3/238 (1) 3/230 (1) 8/238 (1)	3/234 (1)	—	—
G272X	8/226 (3)	—	—	—	—	—
V230I	3/242 (3)	—	—	—	—	—
A104D	8/242,246 (7)	—	—	—	—	—
IVS10-11G>A	7/250, 246 (45) 7/230 (6) 7/242 (1)	7/250 (13) 7/254 (1)	7/250 (28) 7/246 (6) 9/250 (4) 7/238 (3)	7/246 (6) 7/250 (5) 7/230 (3) 7/254 (2) 7/238 (1) 7/258 (1) 7/262 (1)	7/250 (2)	—
IVS12+1G>A	8/242 (52)	—	8/242 (2)	—	8/242 (8)	—

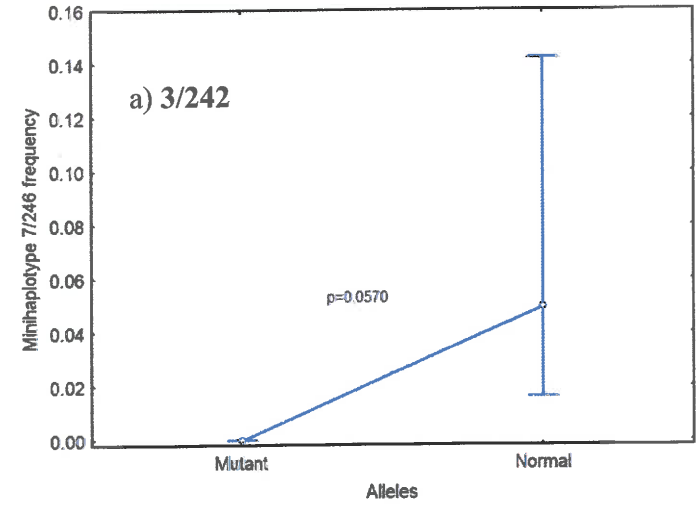


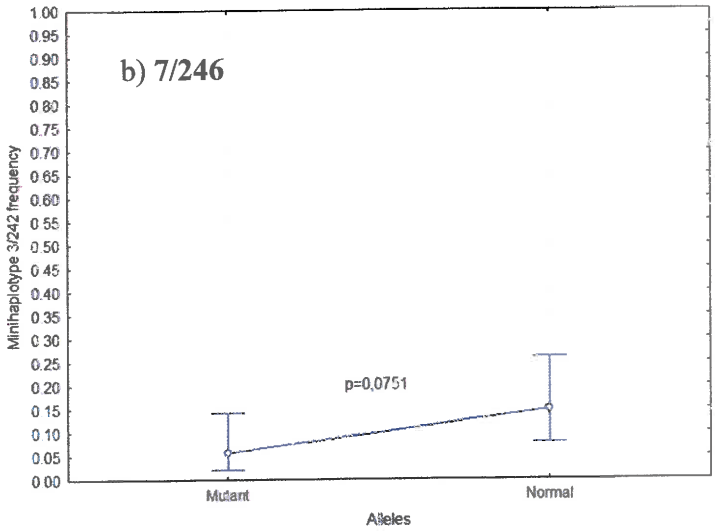
Appendix VI. Genomic DNA sequence of *PAH* gene exon 12 (capital letters) and flanking introns (small letters). Location of the nucleotide and amino acid substitution are typed in red box. PCR oligonucleotide primers are underlined. Recognition sequence for the restriction enzyme *Sty*I is marked by pointer.

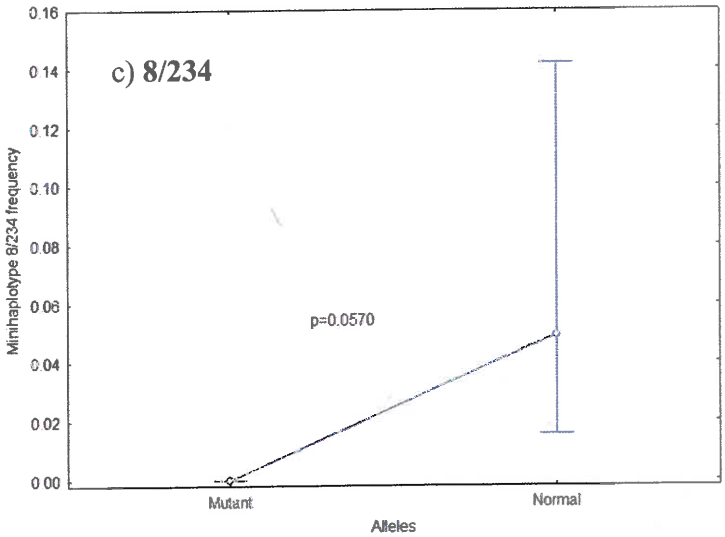
Appendix VII. Rules used by *PolyPhen* to predict effect of nsSNPs on protein function and structure.

	RULES (connected with logical AND)			Prediction	Basis	Effect
	PSIC score difference:	**Substitution site properties:**	**Substitution type properties:**			
1	arbitrary	annotated as a functional site^{+}	arbitrary	**probably damaging**	sequence annotation	functional, functional site
2	arbitrary	annotated as a bond formation site^{++}	arbitrary	**probably damaging**	sequence annotation	structural, bond formation
3	arbitrary	in a region annotated as transmembrane	PHAT* matrix difference resulting from substitution is negative	**possibly damaging**	sequence annotation	functional, functional site, transmembrane
4	arbitrary	in a region predicted as transmembrane		**possibly damaging**	sequence annotation	
5	<=0.5	arbitrary	arbitrary	**benign**	multiple alignment	
6	>1.0	atoms are closer than 3Å to atoms of a ligand	arbitrary	**probably damaging**	structure	functional, functional site, ligand binding
7		atoms are closer than 3Å to atoms of a residue annotated as Binding, Active_site, or Site	arbitrary	**probably damaging**	structure	functional, functional site, indirect
8	in the interval (0.5 - 1.5]	with normed accessibility <=15%	change of accessible surface propensity is >=0.75	**possibly damaging**	structure	structural, buried site, hydrophobicity disruption
9			change of side chain volume is >=60	**possibly damaging**	structure	structural, buried site, overpacking
10			change of side chain volume is <=-60	**possibly damaging**	structure	structural, buried site, cavity creation
11		with normed accessibility <=5%	change of accessible surface propensity is >=1.0	**probably damaging**	structure	structural, buried site, hydrophobicity disruption
12			change of side chain volume is >=80	**probably damaging**	structure	structural, buried site, overpacking
13			change of side chain volume is <=-80	**probably damaging**	structure	structural, buried site, cavity creation
14	in the interval (1.5 - 2.0]		change of accessible surface propensity is >=1.0	**probably damaging**	structure	structural, buried site, hydrophobicity disruption
15			change of side chain volume is >=80	**probably damaging**	structure	structural, buried site, overpacking
16			change of side chain volume is <=-80	**probably damaging**	structure	structural, buried site, cavity creation
17		arbitrary	arbitrary	**possibly damaging**	structure	structural, buried site, cavity creation
18	>2.0	arbitrary	arbitrary	**probably damaging**	multiple alignment	

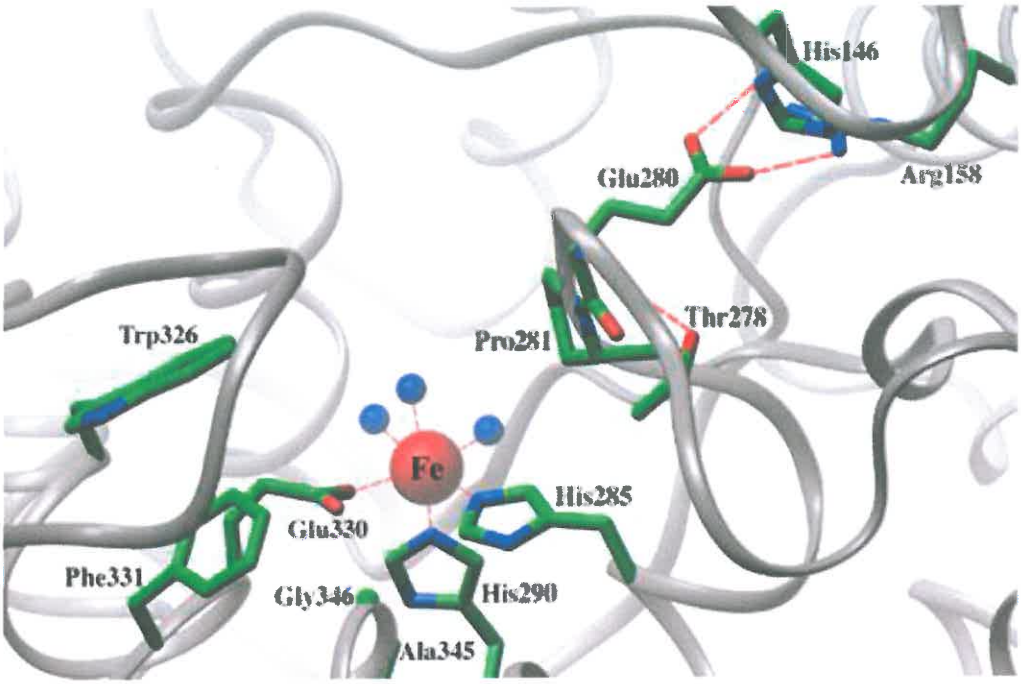
* - predicted hydrophobic and transmembrane matrix



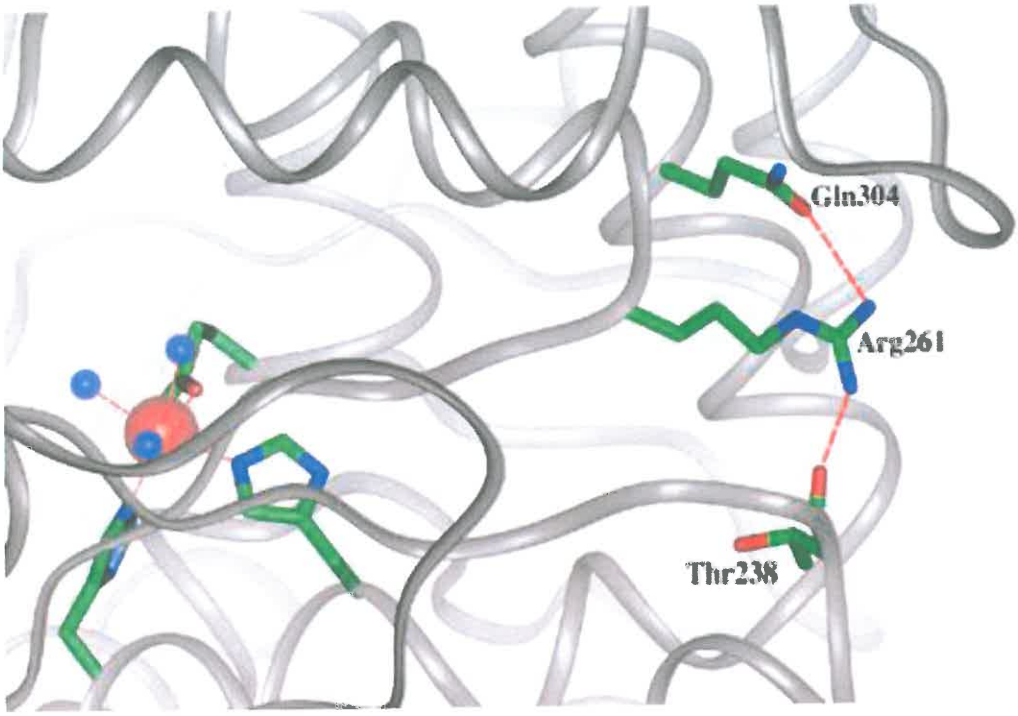




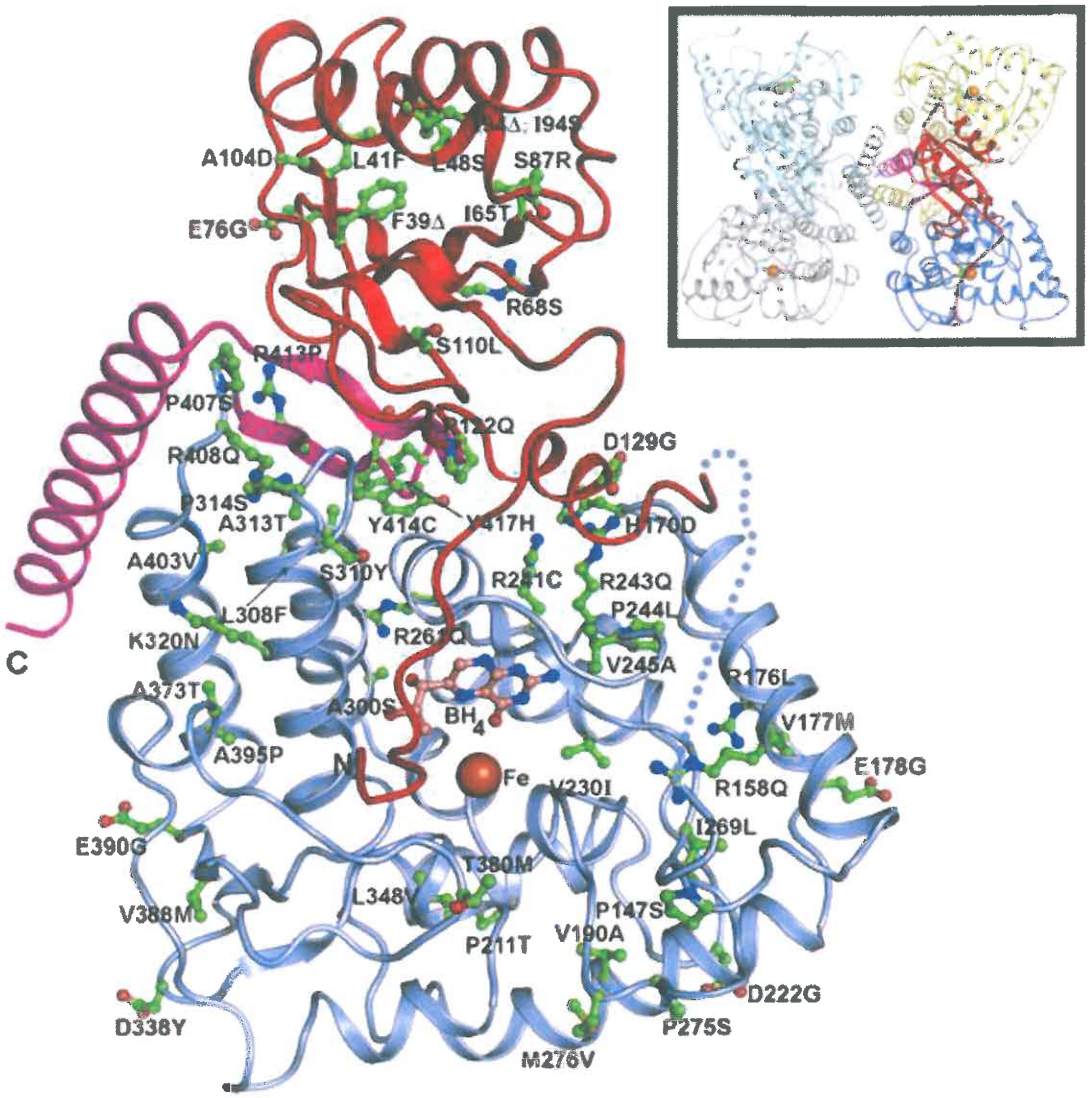
Appendix VIII. Statistical analysis showed the tendency to statistically significant difference between normal and mutant alleles in the distribution of minihaplotypes 3/242 (a), 7/246 (b) and 8/234 (c).



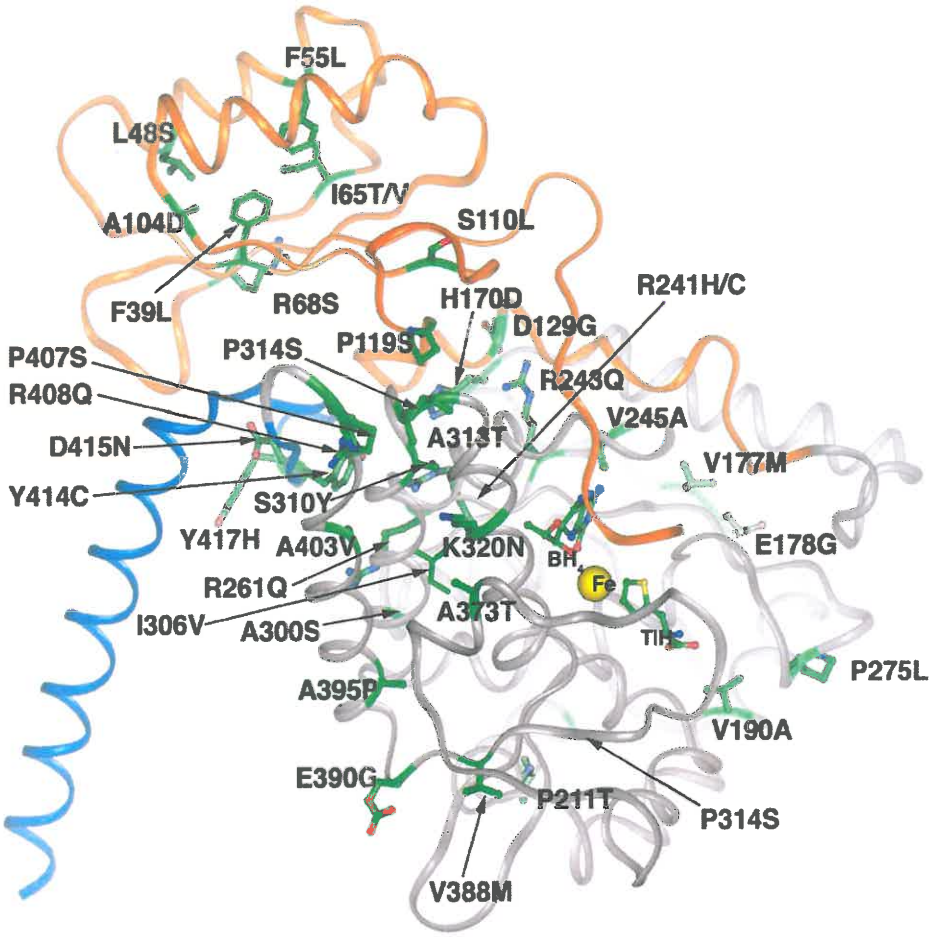
Appendix IX. Close-up of the active site surrounding the catalytic iron. The iron is shown as a red sphere and the three water molecules liganded to the iron are shown as blue spheres. Some residues that have reported PKU mutations (Thr278, Glu280, Pro281, Trp326, Phe331) and are located close to the iron in the active site, together with their interacting residues, are also shown (Erlandsen and Stevens, 1999).



Appendix X. Close-up of the surroundings of residue Arg261, displaying the hydrogen bonding partners Gln304 and Thr238. Arg261 has three associated PKU mutations (R261Q, R261P, and R261X). The active site iron and its ligands are also shown for reference (Erlandsen and Stevens, 1999).



Appendix XI. BH4-responsive mutations are mapped in the 3Dcrystal structure of the PAH monomer. In the active site, the iron atom and BH4 cofactor are shown in red. The N-terminus starting over the active site as well as the rest of the regulatory domain is highlighted in red; catalytic domain in blue; and tetramer domain is in purple. In the top right corner is the native tetramer form of the enzyme (Blau, 2010).



Appendix XII. The backbone of a monomer of the composite model of phenylalanine hydroxylase. The side-chains of the residues involved in BH_4 responsiveness in PKU patients are shown. The carbon atoms are colored green, the nitrogen atoms are blue, and the oxygen atoms are red. The active site iron is shown as a yellow sphere. The cofactor BH_4 is positioned into the structure based on the electron density for the cofactor analogue 7,8-dihydro-L-biopterin crystal structure of human double-truncated phenylalanine hydroxylase (Erlandsen and Stevens, 1999).