Population Genetic Variation at the Human Phenylalanine Hydroxylase locus

Kevin C. Carter

Department of Biology

McGill University

Montreal, Canada

July 1996

A thesis submitted to the Faculty
of Graduate Studies and Research
in partial fulfilment of the requirements
for the degree of Master of Science.

ISBN 0-612-19800-6

Canada

This thesis is dedicated to my family whose support
and encouragement allow me to continue towards
my lifelong goals.

ABSTRACT

Denaturing gradient gel electrophoresis (DGGE) and sequencing of the *PAH* locus has found 38 different mutations on 141 chromosomes in the PKU patients resident in Quebec; mutation analysis is now 92.5% complete. Two novel disease producing alleles (K42I, R157N) and one silent allele (IVS6 nt-55) were discovered in this project; these mutations remain unique to the Quebec population. Three novel mutation-[haplotype] combinations were also found (S67P [H1], G218V [H2], V245A [H7]); they are not at hypermutable sites and are therefore compatible with a single homologous recombination event between two different haplotypes. Whereas mutation types (missense 64%, nonsense 6%, spiice 9%, frameshifts 6%, silent 15%), resemble those in world populations, the Quebec allele profile differs from that of any European population, reflecting range expansion, founder effects, genetic drift and assimilation. Furthermore, when analyzed by geographic region a stratification of *PAH* alleles is apparent, reflecting the different demographic histories of Western and Eastern Quebec and Montreal.

RÉSUMÉ

Les techniques de séquençage et d'électrophorèse sur gradient de gel dénaturant ont identifiées 38 mutations différentes; l'analyse des mutations est maintenant complète à 92.5% pour les 141 chromosomes de patients québécois atteints de la PCU. Ce projet a permis de découvrir des mutations spécifiques à la population québécoise, soit une allèle silencieuse (IVS6 nt-55) et deux nouvelles allèles (K42I et R157N) causant la maladie. De plus, trois nouvelles associations de mutations et de haplotypes ont été découvertes (S67P [H1], G218V [H2], V245A [H7]); ces mutations ne sont pas à des sites hypermutables et sont donc compatibles avec une recombinaison homologue simple entre deux haplotypes différents. Même si les types de mutations retrouvés (faux-sens (64%), non-sens (6%), épissage (9%), décalage du cadre de lecture (6%), silencieuse (15%)) ressemblent à ceux de la population mondiale, le profil allélique québécois diffère de celui de toutes populations européennes et suggère des causes de migration, d'effets fondateur, de dérive génétique et d'assimilation. D'ailleurs, lorsque le Québec est séparé on régions distinctes, une stratification des allèles *PAH* est apparente, ce qui correspond aux différents passés démographiques de Montréal, de l'ouest et de l'est de la province du Québec.

ACKNOWLEDGMENTS

This thesis could not have been possible without the help of many colleagues and friends. Many people contributed to this work in both a professional and personal manner, the following list only touches on these contributions.

First and foremost I would like to thank my supervisor and friend, Charles Scriver. He is always willing to give people a chance, and by taking a chance with me has given me one of the best experiences and greatest opportunities of my life. I hope I can continue at the level of excellence that he brings out in every member of his team. He is often imitated, never duplicated.

Thanks to IREP for providing the studentship for this project.

Special thanks goes to my supervisory committee, Rima Rozen and Ken Morgan, whose helpful suggestions and constant input helped in the maturation of this project.

Very special thanks to Susan Byck, who constantly put up with a barrage of questions and always found the time to teach, demonstrate, proof-read and brain storm with me. Without her assistance this project would not have happened.

Thanks to Per Guldberg for taking the time to show me the secrets of DGGE and the night life of Copenhagen. His advice and hospitality is greatly appreciated.

Thanks to Gilles Doucet and Rachel Laframboise whose collaboration provided the samples from the Eastern and Western Quebec regions.

Lynne Prevost and Huguette Rizziero deserve special thanks for putting up with my constant requests and helping me deal with McGill and hospital bureaucracy.

Thanks to Dr. Kurt Sittman for allowing me to TA his genetics course and for being a sounding board for all my crazy population genetics ideas.

The following deserve thanks for their valuable assistance and collaborations. Eileen Treacy for allowing me to be a part of the *in vivo* study. Emre Kayaalp for providing genotpye-phenotype data. Paula Waters and Stacy Hewson for expressing the novel mutations of Quebec. Paula again, for starting the mutation detction as a sort of "warm up" to her project. Liem Hoang for helping to set up a patient database. Gay Elkas for providing patient information. Brent Richards for providing valuable haplotype information and being an entertaining lab partener . Karan Manhas for keeping me busy and providing computer assistance and database information. Christineh Sarkissian and Margaret Fuller for introducing me to the world of enu mice and providing an entertaining laboratory atmosphere.

The following members of the deBelle lab have reinforced my belief that school cannot be accomplished (and life is nothing) without good friends.

Special thanks to Pierre J. Ledoux for being the greatest friend anyone could ask for. From showing me the way home from the Q.E. to walking the streets of Montreal at 5 a.m. he has always been there for me. This thesis would have been possible without him ... but it wouldn't have been half as fun.

Thanks to Emre Kayaalp for opening up my eyes to the world of funk (I'm a funkier person because of you) and for doing your dishes for me, you are a great friend.

Thanks to Dan Wilson for always being there, opening your home for every party (bottle cap fights) and for reminding everybody that Tuesday night was cheap movie night.

Thanks to Stephane Roy for educating me on the finer points of wine consumption. I will always think of you when I look for the indent in the bottom of the bottle and know that this is a good wine.

Thanks to Omar J. Chahal for leaving completely unmentionable messages on my answering machine and for pool and a movie.

Thanks to Stacy, Fran, Claudia and Susan for keeping myself and the above mentioned males a little bit civilized. It's refreshing to know that not everybody can be dragged down to our level.

And thanks to everybody on the seventh floor (and surrounding areas) of the Montreal Children's. Everybody made these two years a pleasure and I couldn't imagine a better place to work or better people to work with.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF FIGURES continued

LIST OF TABLES

LIST OF ABBREVIATIONS

ACRS	Artificially created restriction site
ASO	Allele specific oligonucleotide
DGGE	Denaturing gradient gel electrophoresis
[H]	Haplotype
HPA	Hyperphenylalaninemia
PAH	Phenylalanine hydroxylase (gene)
PAH	Phenylalanine hydroxylase (enzyme)
PCR	Polymerase chain reaction
phe	Phenylalanine
PKU	Phenylketonuria
RFLP	Restriction fragment length polymorphism
SSCP	Single strand conformation polymorphism
STR	Short tandem repeat
TRH	Tryptophan hydroxylase
TYH	Tyrosine hydroxylase
UTR	Untranslated region
VNTR	Variable number tandem repeat

Chapter 1
Introduction

1.1.0 History of the hyperphenylalaninemias (HPA) 1934-1996:

In 1934 Fölling described an autosomal recessive disorder of phenylalanine metabolism which was found in persons with profound mental retardation (Fölling 1934). Penrose confirmed this finding by publishing a pedigree and named this disease phenylketonuria (PKU) due to the phenylpyruvic acid found in the urine of affected individuals (Penrose 1935). In 1946 Penrose suggested that this disease was potentially treatable and questioned its non-random distribution and 2 percent carrier frequency in Europeans (Penrose 1946). In 1953 the first attempt at treating PKU was undertaken by placing a mentally retarded patient on a diet low in phenylalanine (phe) (Bickel et al. 1953). Based on the moderate yet measurable success of the low phenylalanine diet, this method of treatment for PKU was further investigated. It was discovered that diagnosis at birth and precise control of phenylalanine intake are required to avoid all neuropsychological consequences of hyperphenylalninemia (HPA) (Smith et al. 1974). Today, a newborn screening program and a complex diet therapy have become available for HPA (Scriver et al. 1980a; 1980b). This screening test and diet therapy have become two of the most widely practiced genetic tests and therapies with near total ascertainment of affected individuals in many populations. The questions posed by Penrose in 1946 (non-random distribution of PKU and elevated carrier frequencies) are still studied and pondered today.

1.2.0 The biochemistry of L-phenylalanine hydroxylation:

With five genes that are responsible for HPA (McKusick OMIM 261600) (Scriver et al. 1995), genetic heterogeneity exists (Figure 1-1). In order for catalytic hydroxylation of L-phenylalanine to occur, both phenylalanine hydroxylase (PAH; EC 1.14.14.1) and a tetrahydrobiopterin cofactor ((6R)-L-erythro-5, 6,7, 8-tetrahydrobiopterin, or BH_4) must be present and the pterin cofactor must then be recycled (reduced) to act as a catalyst in the hydroxylating reaction. Thus other loci are involved. To reduce BH_4, pterin-4α-carbinolamine dehydratase (PHS/PHD) and dihydropteridine reductase (DHPR; EC 1.6.99.7) are required. Additional loci are required to synthesize BH_4; these are guanosine triphosphate cyclohydrolase I (GTPCH; EC 3.5.3.16), 6-pyruvoyltetrahydropterin synthase (6-PTPS; EC 4.6.1.10), and sepiapterin reductase (SR; EC 1.1.1.153), each acting sequentially (Kaufman 1993). While mutations at five of these genes can cause HPA (*PAH*, *PCD/DCOH*, *QDHR*, *GTPCH* and *PTS*) (Scriver 1994), mutations at the PAH locus account for 99% of hyperphenylalaninemias and contribute to both classical PKU and non-PKU hyperphenylalaninemia (non-PKU HPA) (Danks et al. 1979).

1.3.0 The screening test for HPA:

A newborn screening program for HPA is now in effect in many populations. Testing for HPA was first accomplished through an assay of urinary metabolites of phenylalanine (Moncrief et al. 1968). As this method proved to be insensitive it was succeeded by a more accurate method which screens blood phenylalanine and misses only one in seventy infants with PKU (Holtzman et al. 1986). Measuring blood phenylalanine

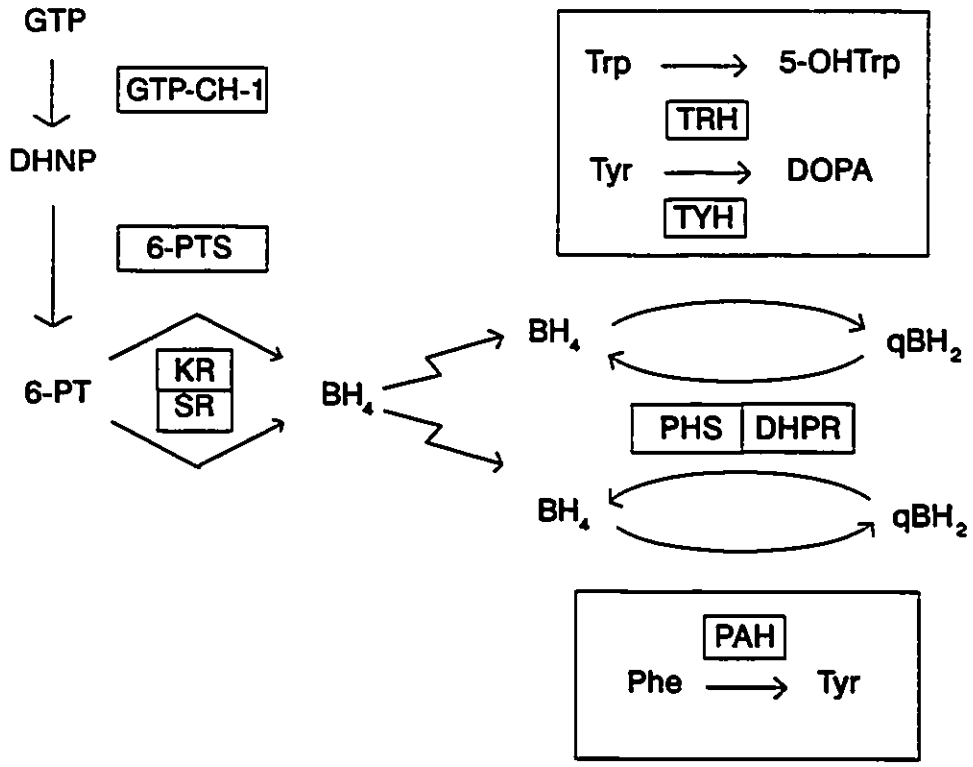
is accomplished using a microbiologic inhibition assay (Guthrie test), fluorometry or chromatography (Scriver et al. 1973; 1989). This test can be performed on liquid or dried blood spots on filter paper, the latter is more frequently used due to ease of handling and stability (Spieto et al. 1985). In Quebec, the newborn screening program has achieved a cooperation rate of greater than 97% (Grenier et al. 1980), and has a sensitivity of 100% (Lambert 1994).

The HPA screening program has implications on the molecular and population studies of HPA in that all chromosomes genotyped are acquired from affected individuals detected through the newborn screening program. Allele frequencies and distributions are therefore biased as only those chromosomes which cause a noticeable phenotype are detected. The requirement of a noticeable phenotype disqualifies all alleles causing milder phenotypes, polymorphisms and unique mutations found only in carrier. These alleles will remain undetected until, by chance, they are combined with a severe disease causing allele in a proband with an HPA phenotype.

Figure 1-1. The hydroxylation of L-phenylalanine and its dependence on BH_4 biosynthesis and regeneration. Taken from Scriver et al. 1989.

Abbreviations used:

GTP	Guanosine triphosphate	**DHPR**	Dihydropteridine reductase
GTP-CH-1	GTP-cyclohydrolase-1	**PHS**	pterin-4α-carbinolamine dehydratase
DHNP	Dihydroneopterin triphosphate	**Phe**	Phenylalanine
6-PTS	6-pyruvoyltetrabydropterin synthase	**Tyr**	Tyrosine
6-PT	6-pyruvoyltetrahydropterin	**PAH**	Phenylalanine hydroxylase
KR	2' ketotetrahydropterin reductase	**DOPA**	Dopamine
		TYH	Tyrosine hydroxylase
SR	Sepiapterin reductase	**5-OHTrp**	5-hydroxy tryptophan
BH4	Tetrahydrobiopterin	**TRH**	Tryptophan hydroxylase
qBH2	Quinoid dihydrobiopterin		



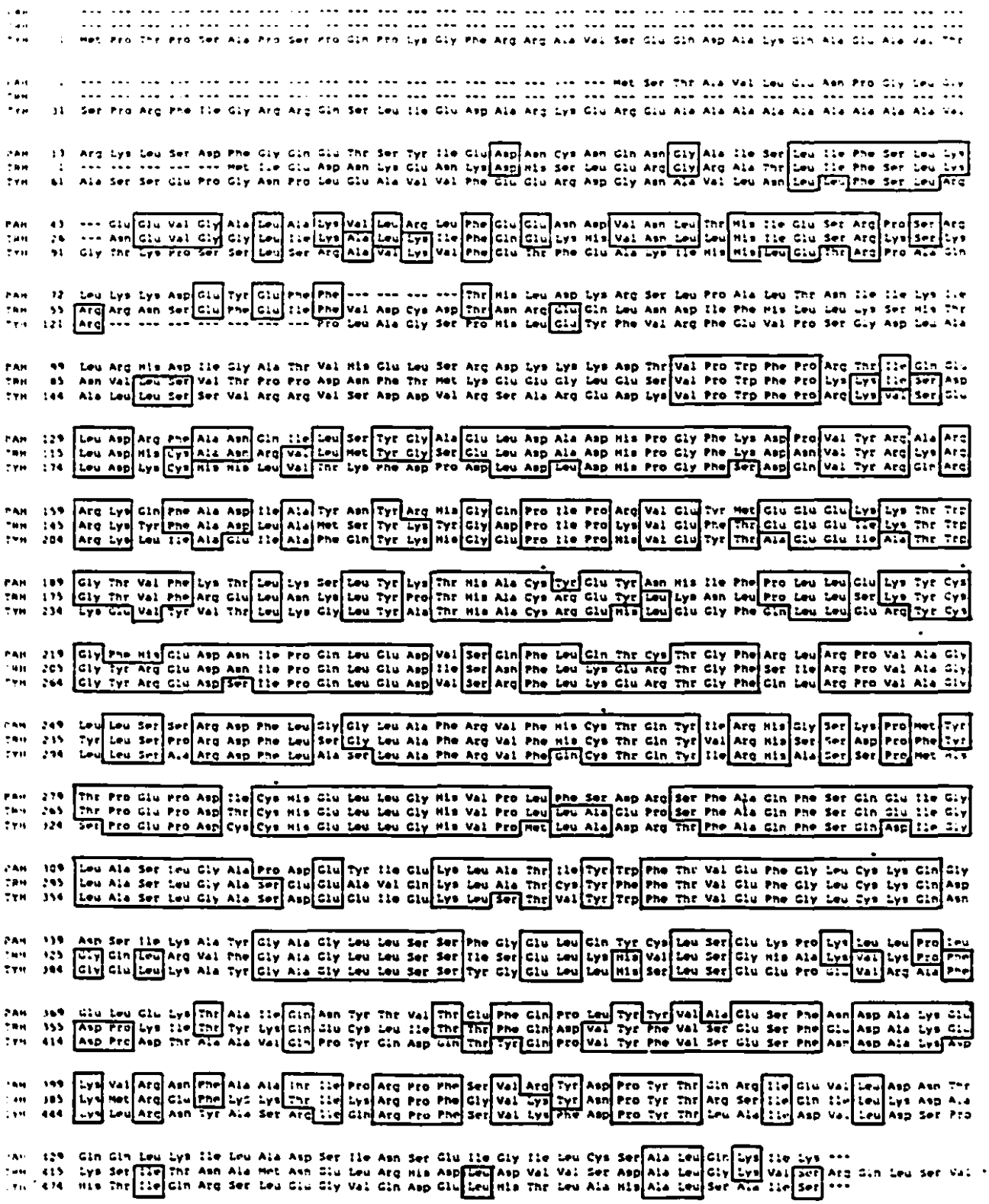
1.4.0 Human *PAH* gene structure:

Phenylalanine hydroxylase belongs to a family of related gene products, the aromatic L-amino acid hydroxylases. Other members of this family include tyrosine hydroxylase (TYH) and tryptophan hydroxylase (TPH) (Scriver et al. 1995). There are strong similarities in the primary sequences of these three proteins both within and between species (Figure 1-2) (Ledley et al. 1990; Grenett et al. 1987; Dahl et al. 1986). Intron/exon boundaries are conserved in the three genes (Stoll et al. 1991), and a single gene encodes both TPH and TYH in Drosophila (Neckameyer et al. 1992). From these findings it has been suggested that these three genes have evolved from a common ancestral gene by duplication and divergence (Grenett et al. 1987).

Using human/rodent cell hybrids containing different combinations of human chromosomes the human *PAH* cDNA was used to probe for the chromosome which harboured the *PAH* gene (Lidsky et al. 1984). This technique mapped *PAH* to chromosome 12 (GenBank U49897) (Lidsky et al. 1984; 1985). Further experiments used both deletion chromosome mapping and *in situ* hybridization to further localize the human *PAH* locus to region q22-24.1 (Lidsky et al. 1985).

The gene is approximately 90 kb in length and contains 13 exons separated by introns of varying size. The introns found in the 5′ region of the gene are large (up to 23 kb) and those in the 3′ region are much smaller in size (1 to 2.5 kb) (Figure 1-3) (Scriver et al. 1995). The locus codes for a 2.4 kb mature messenger RNA which is transcribed into a 452-amino-acid-protein.

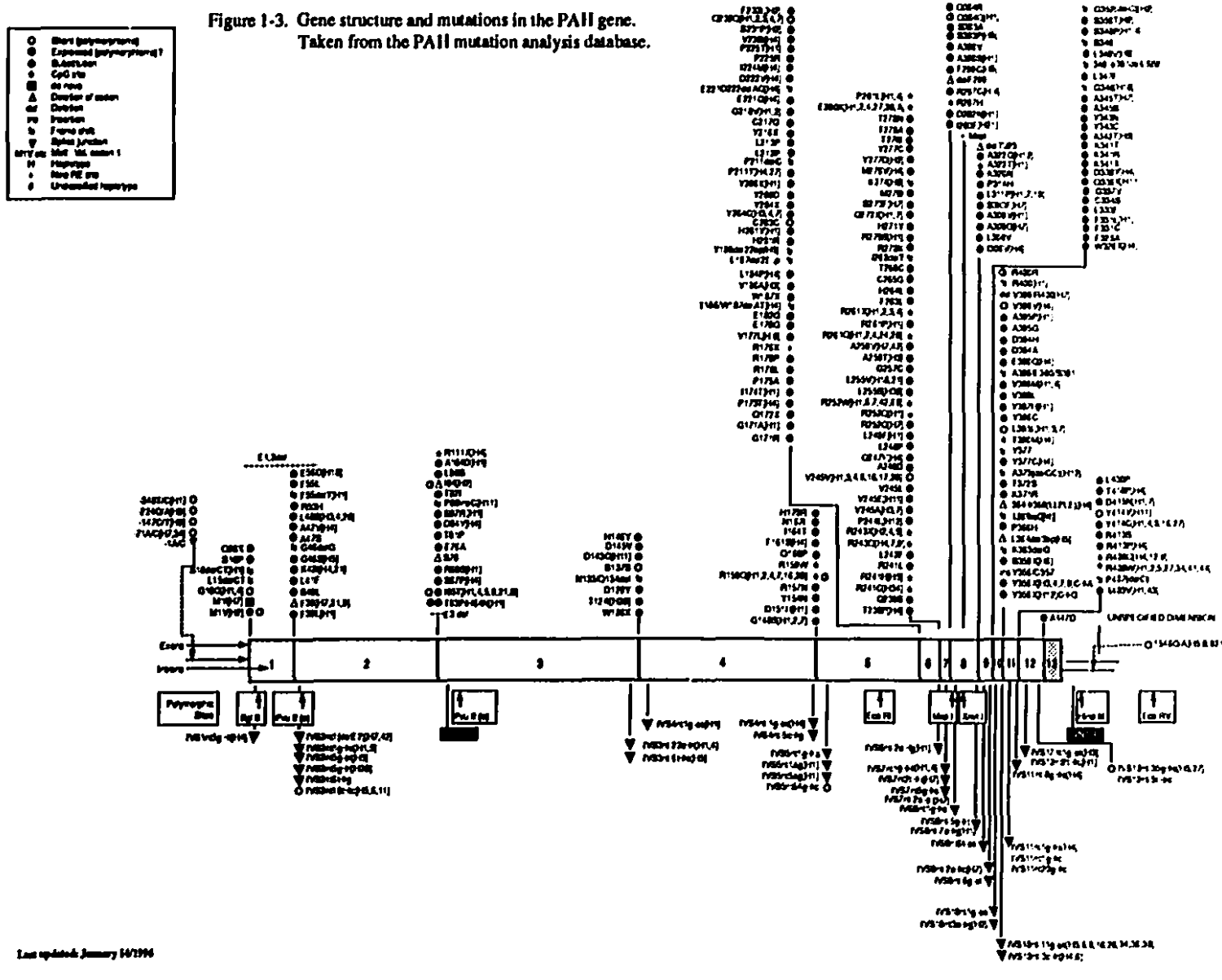
Figure 1-2. Alignment of amino acid sequences of human phenylalanine hydroxylase (PAH), rabbit tryptophan hydroxylase (TRH) and rat tyrosine hydroxylase (TYH). Boxes indicate conserved regions. Taken from Grenett et al. (1987).



The 5′ regulatory region of *PAH* is now well characterized (Konecki et al. 1992). Five potential CAP sites, a feature of housekeeping genes, have been reported upstream of the methionine translation initiation codon in exon one. Furthermore, sequences similar to GC, CACCC and CCAAT boxes, activator protein 2 (Ap-2) sites, partial glucocorticoid response elements (GREs), and partial cyclic AMP response elements (CREs) have also been identified (Konecki et al. 1992). This would imply transcription regulation by multiple transcription factors (Scriver et al. 1994). The use of a PAH-CAT fusion gene has revealed that the 5′ flanking region is required for stage and tissue specific expression the PAH gene in developing transgenic mice (Wang et al. 1992).

Seventeen non-disease producing restriction fragment length polymorphisms (RFLP) are present at the *PAH* locus. Ten of these sites (cut by the enzymes Bgl II, Pvu II, Eco RI, XmnI, Hind III, Eco RV, SphI, Dde I and Alu I) (Figure 1-3) are highly polymorphic making them useful in assigning *PAH* haplotypes to mutant and normal chromosomes; The 3′ untranslated region (UTR) contains a multiallelic variable number of tandem repeats (VNTR) which can by analyzed by PCR (Goltsov et al. 1992). There is both length (3,6,7,8,9,11,12 and 13 copies) and sequence variation within this 30 bp cassette (Byck et al. 1994). The 5′ region of the gene contains a multiallelic short tandem repeat (STR) (Figure 1-2) (Goltsov et al. 1993). Together these sites can define at least 384 different haplotypes, but due to the strong linkage disequillibrium between certain sites only about 70 of these haplotypes have been observed (Feingold et al. 1993). Still, these haplotypes reveal a very high degree of heterozygosity making them useful for the study of population structure, carrier testing and prenatal diagnosis

Figure 1-3. Gene structure and mutations in the PAH gene.
Taken from the PAH mutation analysis database.

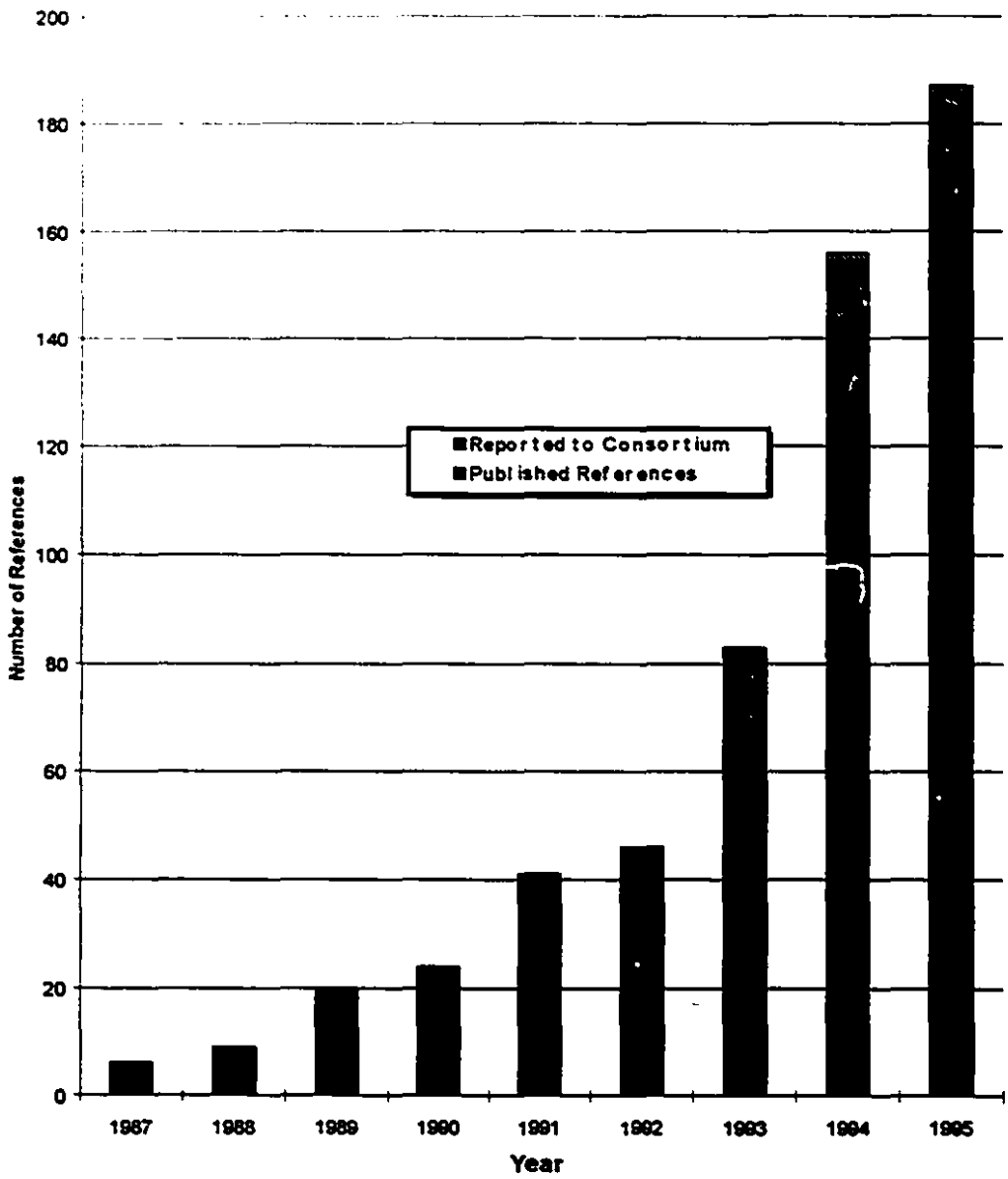


1.5.0 Mutations and mutation detection in the *PAH* gene:

In 1987 the first two *PAH* mutations were characterized (IVS12 nt1: Marrit et al 1987; R408W; DiLella et al. 1987). Since then, over 300 *PAH* mutations have been reported to the *PAH* mutation analysis database (Figure 1-4 *PAH* mutation analysis database). The large number of mutations reflects the great amount of interest in the *PAH* gene, the increase in membership of the *PAH* Consortium and the increased efficiency of mutation detection techniques (*PAH* mutation analysis database; Scriver et al. 1996a).

The efficiency of mutation detection depends on the method used and with the same method may vary between genes (Cotton 1993). Many mutation detection techniques have been used to find *PAH* mutations including haplotype analysis with subsequent sequencing or restriction digest, single strand conformation polymorphism (SSCP) (Rozen et al. 1994) and chemical cleavage methods (Forrest et al. 1991). The technique that has attained the highest efficiency of *PAH* mutation detection in any one population is denaturing gradient gel electrophoresis (DGGE). Based on the differential melting temperatures of hetero- versus homoduplex polymerase chain reaction (PCR) products (Myers et al. 1987) the sensitivity of DGGE depends on the quality of primer design. Per Guldberg has designed GC clamped primers which have detected greater than 95 % of *PAH* mutations in populations such as Sicily and Denmark (Guldberg et al. 1993a; 1993b). The increased efficiency of mutation detection has had a great impact on population studies, allowing regional comparisons in a single population, revealing 'genetic geography' or stratification (Eisensmith et al. 1992; Konecki et al. 1991a).

Figure 1-4. Number of *PAH* mutations reported annually, by publication and direct communication to the *PAH* Mutation Analysis Database. Many mutations are reported multiple times in different populations. Prepared with the assistance of Karan Manhas.



Of the 300+ *PAH* alleles only about 10% are "prevalent" (relative frequencies 5-30%) and the rest are "rare" (Scriver et al. 1996). The distribution of mutations is non-random in the *PAH* gene, with the majority of mutations being found in the 3' half of the gene between exons 5 and 12 (*PAH* Mutation Analysis database). Exon 7 has been ruled out as a mutational hot-spot (Dworniczak et al. 1992) and the apparent clustering of mutations reflects an ascertainment bias combined with the importance of the corresponding domain for proper protein function. This region of the gene encodes domains such as the putative active site and the BH_4 binding site, both important for normal PAH activity (Scriver et al. 1995).

Although missense point mutations are most prevalent at the *PAH* locus, nonsense mutations, splicing mutations, deletions (usually small) and insertions (with and without frameshifts) are also present (*PAH* mutation analysis database; Scriver et al. 1996a). 24 of the 48 CpG sites in the *PAH* gene harbour a missense mutation of which 19 are compatible with methylation-mediated deamination of a 5-methylcytosine (Byck et al. 1996). Since CpG dinucleotides are hypermutable (Cooper et al. 1988; Cooper and Krawczak 1993), 12 of these 19 mutations found on multiple haplotypes are compatible with a recurrent mutation mechanism. In analyzing polymorphic sites at the *PAH* locus the R408W and E280K mutations have been shown to be compatible with the recurrent mutation mechanism (Byck et al. 1994; Byck et al. 1996).

All *PAH* mutations are catalogued in the *PAH* mutation analysis database (Hoang et al. 1996). This database is maintained by the *PAH* Mutation Analysis Consortium (86 members in 28 countries), is updated from the Montreal Children's Hospital, and can by

found on the world wide web (http://www.mcgill.ca/pahdb) or in a stand alone, Fox Pro, IBM format. The database provides such information as mutations at the *PAH* locus, mutation/population associations, mutation/haplotype associations, RFLP information and publications. The electronic publication acquired from a submission to the *PAH* mutation analysis database will be regarded as an authentic publication and can be referenced as an "electronic publication" (Hoang et al. 1996).

1.6.0 Maintenance of disease producing alleles at an elevated aggregate frequency:

The aggregate frequency of PKU mutaions approaches 1 percent in most temperate-zone populations in the northern hemisphere (Scriver et al. 1994). The severe mental retardation associated with untreated PKU makes the disease a virtual genetic lethal, with two mutant alleles potentially being lost with every affected individual. With the loss of mutant alleles one would expect allele frequencies to decrease to an equilibrium maintained by mutation (Woolf et al.1975; Woolf 1978; 1986). Still, PKU remains one of the most common inborn errors of metabolism in man. Why this is so was asked a half century ago (Penrose 1946) and continues to be asked today (Scriver et al. 1996b).

Mutation rates of 1 in 5000 to 1 in 20 000 genes per generation would be required to maintain the incidence of PKU in many European populations (Ledley et al. 1985). These mutation rates are unusually high for eukaryotes. The *PAH* gene has not been found to be hypermutable when analyzed by various methods. Only 19 mutations are compatible with methylation-mediated deamination and the MUTPRED profile (Cooper and Krawczak 1993) of the *PAH* gene shows that there are not an abundance of hypermutable CpG sites within the gene (Byck et al. 1996; this study). In addition, a

calculation of the evolutionary distance between human and rat PAH has demonstrated PAH to have an estimated mutation rate of 7 accepted point mutations per 100 million years; an average value for proteins of higher eukaryotes (Ledley et al. 1985).

Founder effects and genetic drift have been documented in several populations (Avigad et al. 1990; John et al. 1989; Lyonnet et al. 1992; Eisensmith et al. 1992a). A founder effect originating in the capital of Yemen before 1762 can explain the elevated frequency of an exon 3 deletion in Yemenite Jews (Avigad et al. 1990). Likewise, the R408W mutation on haplotype one in French-Canadians seems to have Celtic origins (Treacy et al. 1993) and it is possible that the high relative frequency of the M1V allele in French-Canadians is due to a founder effect/genetic drift (John et al. 1989). Although these mechanisms can explain a small number of mutant alleles, the presence of numerous mutant *PAH* alleles in different populations implies that founder effects alone cannot account for the elevated frequency of PKU (Scriver et al. 1995). Also, random genetic drift can be discounted since it would have to affect a number of mutant *PAH* alleles concordantly (Scriver et al. 1995).

Since founder effect and drift may explain the high frequency of only some alleles (Scriver et al. 1996b), heterozygote advantage is the other possible mechanism that can account for the high relative frequency of the remaining mutant *PAH* alleles. In fact, heterozygote advantage could theoretically explain the presence of multiple *PAH* alleles, the maintenance of the high cumulative frequency of these alleles and the variation in frequencies between populations (selective agents and forces differ between populations) (John 1991). However, it is almost impossible to identify the exact selective agents that may effect the *PAH* locus since i) they may have been present in the past but may no

longer exist, ii) the effects of heterozygote selection may be too small to detect or define, or iii) the selective agent may be different between populations (Scriver et al. 1995). Nonetheless, a small number of selective agents have been hypothesized.

Ochratoxin A, an abortifacient mycotoxin found in contaminated grains and lentils, has been proposed to be one a selective agent allowing for PKU heterozygote advantage (Woolf 1986). A pregnant woman eating these contaminated foods may miscarry, but if she is heterozygous for PKU her elevated blood phenylalanine levels may protect the fetus. In the 1970's it was calculated that the reduction of fetal mortality in the PKU families of Ireland and west Scotland constituted a heterozygote advantage of 5%. Since an advantage of 1.3% would maintain the incidence of PKU in the population studied, this suggests that the frequency of PKU may be increasing in these areas (Woolf 1978; 1986).

Although it may be possible to account for the frequency and distribution of PKU alleles by implicating genetic hypermutability, genetic drift, founder effects and selection all mechanisms are likely to contribute at some time. The problem is then complicated with what weight to apportion to each (Flint et al. 1993), thus explanations for this matter remain difficult. To resolve this problem soon would be an advantage for at least two reasons; i) as human populations merge and fracture the genetic structures upon which the arguments rest will change and ii) it would allow both parents and patients to better cope with this disorder if they knew the ultimate cause (Scriver et al. 1996b).

1.7.0 Hyperphenylalaninemia and the Quebec population:

French Canadians are one of the best documented populations in the world (McEvedy et al. 1978). Settlement of New France began in the early 1600's, by 1640 approximately 160 French Europeans had immigrated and by 1750 settlers numbered about 15 000; An expansion that took place in relative genetic isolation. At the present time over 5 million people claim French-Canadian ancestry (Scriver et al. 1996b).

Not all Quebec residents are French Canadian. In the late 18th and up to the early 20th century continued migration saw English, Scots and Irish settlers come to Quebec. In the 20th century settlers from eastern then southern Europe and most recently Asia migrated to Canada. Today, approximately 1 million Quebec residents claim non-French Canadian ancestry (Scriver et al. 1996b).

In 1971 Quebec introduced universal screening for HPA, since then the test has gained 100 % sensitivity and has allowed the incidence of PKU in the province to be calculated at 1/24 985 live births (Lambert 1994). Molecular analysis of the *PAH* locus has demonstrated recurrent mutation, founder effect and genetic drift in Quebec. Using the rare disease-producing *PAH* alleles of Quebec it can be illustrated that the demographic history and the history of genes are contingent (Rozen et al. 1994; Scriver 1993).

Several novel mutations and mutation/haplotype associations have been found in Quebec. Molecular research of these rare alleles has illustrated the origins of the mutations as well as the origins of the population. R408W haplotype 2 ([H2]) and E280K [H1] are both prevalent in Europe (Eisensmith et al. 1992), but in Quebec these mutations occur about half the time on [H1] and [H2] respectively (John et al. 1990;

Treacy et al. 1993; Rozen et al. 1994; this study). It was hypothesized that R408W [H1] could have evolved via a single recombination between haplotypes 1 and 2 (or vice versa), or it could have originated by recurrence at the hypermutable CpG dinucleotide of this codon (John et al. 1990). To distinguish between the two hypotheses polymorphic markers 5′ and 3′ relative to codon 408 were analyzed on both Quebec and European R408W chromosomes (Byck et al. 1994). The 5′ STR markers were found to be different on the two haplotypes harbouring R408W. As well, the fifth cassette of the 3′ VNTR 8 allele differentiated mutant and normal [H1] chromosomes. It was then concluded that the occurrence of R408W on multiple haplotypes could be best explained by a recurrent mutation mechanism (Byck et al. 1994). The corresponding study was also performed for the E280K mutation which also occurs at a hypermutable CpG dinucleotide. It was also concluded that the occurrence of E280K on multiple haplotypes is compatible with recurrent mutation (Byck et al. 1996).

Genealogical studies were performed on Quebec families harbouring the novel R408W [H1] allele and the I65T mutation which was also identified first in Quebec (John et al. 1992), (Treacy et al. 1993). It was discovered in all Quebec families harbouring these mutations that an Irish or Scottish ('Celtic') ancestor could be identified (Treacy et al. 1993). Upon further studies of the Scotland, Ireland and northwestern European populations these alleles were found at high frequencies indicating that R408W [H1] and I65T are migration traces in Quebec (Scriver et al. 1996b). This is just one example illustrating that the history of the population and the history of genes are contingent in the Quebec population.

1.8.0 The questions of this thesis

My Masters thesis involves a systematic molecular characterization of the phenylalanine hydroxylase locus in the residents of Quebec. This study will attempt to characterize all PKU and variant-PKU causing alleles in the Quebec chromosomes available for study. This study is a continuation of three previous studies which have also attempted to characterize Quebec *PAH* alleles (John et al. 1990; Treacy et al. 1993; Rozen et al. 1994). These previous studies have fully and partially (one allele) genotyped several Quebec families, but were also unsuccessful in genotyping several families. The advantage of this study is the introduction of a denaturing gradient gel electrophoresis (DGGE) system of mutation scanning to the Quebec population. In visiting the John F. Kennedy laboratory in Denmark, I have learned this technique from its pioneer in the *PAH* gene, Per Guldberg. The J.F.K laboratory has used DGGE to reveal 99% of the *PAH* mutations in Denmark (Guldberg et al. 1993b) and in theory the technique should give close to the same results in the Quebec population. Therefore, I have used DGGE to fully genotype the partially and ungenotyped families from previous studies, and have added new PKU families to this study through a collaboration with the Centre Hospitalier l'Université de Laval, L'Hôspital Sainte-Justine and the Montreal Children's Hospital.

1.8.1 The significance of this thesis

This project will allow better regional comparisons of allele frequencies, and thus further elucidate the stratification of alleles within Quebec. This will assist in understanding how the history of each region within Quebec is unique and how the history of the genes in each region are contingent with the history of the population.

Furthermore, the events that may have contributed to the genetic structure of Quebec (founder effect, genetic drift, range expansion) can be more accurately hypothesized.

By revealing rare mutation/haplotype associations that are unique to Quebec, the origin of these mutations (recurrent mutation, recombination, gene conversion) can be hypothesized.

By introducing an efficient *PAH* mutation scanning technique a centre for *PAH* genotyping could evolve. To know the *PAH* genotype has clinical relevance since most *PAH* mutations are predictive of the phenotype and thus stringency of therapy.

Chapter 2
MATERIALS AND METHODS

2.1.0 Sources of DNA

DNA for this study was collected from three major centres which follow individuals with hyperphenylalaninemia (HPA); The Montreal Children's Hospital, Hôspital Sainte-Justine and Centre Hospitalier de l'Université de Laval. Informed consent and patient data were collected under the supervision of the clinic doctor. Ancestry of probands was determined from surname analysis and from parents when possible. No probands were known to be closely related to another and two were known to have consanguineous parents.

Two families not from Quebec, and therefore not included in the population portion of this study, were referred to me for genotypic analysis. The DNA of the first family, who resided in the United States and was of Chinese ancestry, was provided by Dr. Rima Rozen. The DNA of the second family, who resided in Iraq and was of Bedouin ancestry, was provided by Dr. Ahmed Teebe.

2.1.1 Clinical Diagnosis

The diagnosis of phenylketonuria (PKU) was defined by a blood phenylalanine of greater than 1000 mmol/L without a phenylalanine restricted diet and a dietary tolerance of less than 500 mg/day phenylalanine (see Scriver et al. 1995); These criteria may incorporate some patients with variant PKU-like genotypes. The clinic doctor at each centre classified patients as classical and variant PKU, or non-PKU HPA. Although some patients with non-PKU HPA were genotyped, only those with PKU were used in the population study.

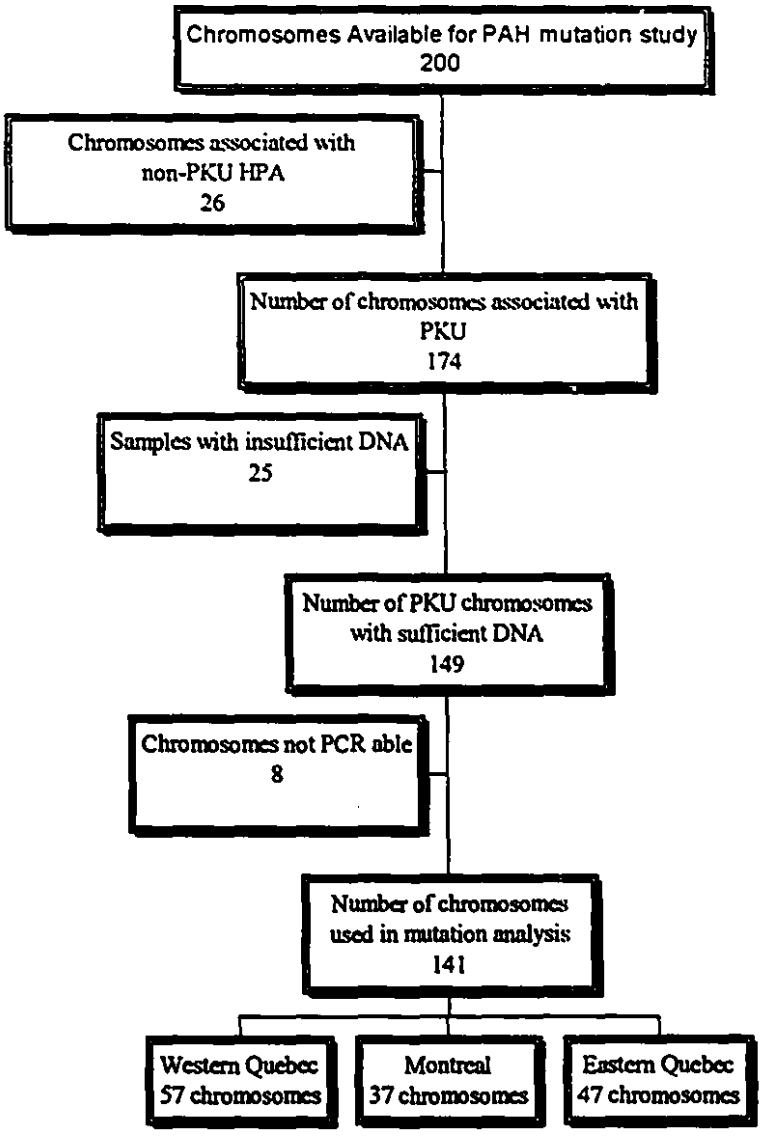
2.1.2 Regional division of the Quebec province

The province of Quebec was roughly divided into three regions using the catchment from which patients are drawn to each hospital. Patients who are followed by the Montreal Children's Hospital make up the Montreal region. Patients followed at L'Hôspital Sainte-Justine constitute the Western Quebec region. Likewise, patients of the Centre Hospitalier de l'Université de Laval make up the Eastern Quebec region.

2.1.3 Inventory of DNA

For all PKU families included in previous Quebec studies (John et al. 1992) (Treacy et al. 1993) (Rozen et al. 1994) DNA had been isolated and quantitated. With the help of Liem Hoang, a Foxpro database was designed to catalogue all HPA DNA samples available from these studies as well as new DNA samples obtained for this study. The database groups families into the three regions of Quebec and contains a unique family identifier number, family member for which DNA is available, and paternal, maternal mutations/haplotypes if available. This database is maintained on the molecular genetics floor of the Montreal Children's Hospital, it is not static but is updated and modified as new DNA samples are received or mutation/haplotypes are determined. To ensure patient confidentiality and privacy the database can only be accessed by myself and other selected researchers. A total of 141 chromosomes were used to determine relative frequencies of mutations and effectiveness of mutation scanning. This number was achieved after subtracting all chromosomes that could not undergo mutation analysis (eg. insufficient DNA, not PCRable etc.) from the total number of Quebec PKU chromosomes collected by three centers (Figure 2-1).

Figure 2-1. Quebec chromosomes available for DGGE analysis.



2.2.0 DNA analysis

2.2.1 Isolation of DNA

For samples obtained during this study, genomic DNAs were isolated from blood using a rapid procedure adapted from John et al., 1991. Whole blood was collected in vacutainer tubes containing EDTA as an anticoagulant. Blood was then transferred to a capped 50 ml conical tube using RSB (10mM Tris pH 7.6, 10mM Kcl, 10mM $MgCl_2$.). Nucleated cells were lysed using a non-ionic detergent (BRL's Nonidet p40) and then pelleted by centrifugation. White cell nuclei were then lysed by resuspending in a buffer containing SDS (Solution 2) (10mM Tris pH 7.6, 10mM Kcl, 10mM $MgCl_2$., 0.5M NaCl, 0.5% SDS, 8mM EDTA). Proteins were removed from the lysate using organic extractions. The DNA was then precipitated using ice cold 99% ethanol, dried and resuspended in sterile water. DNA concentrations and purity was determined spectrophotometrically (Maniatas et al., 1982).

2.2.2 Polymerase Chain Reaction

i) Polymerase chain reaction (PCR) for Denaturing Gradient Gel electrophoresis (DGGE)

All thirteen exons of the phenylalanine hydroxylase (*PAH*) gene were amplified simultaneously using PCR (Saiki et al. 1985, 1986) with a Perkin Elmer Cetus Thermocycler. The primers, obtained from the Sheldon Biotechnology Centre, were based on sequences provided by Dr. Savio Woo and were modified by adding a GC clamp (Sheffield et al. 1989) constructed by Per Guldberg (Guldberg et al. 1993).

Amplification was performed in 25 µl containing ~1 µg of DNA, 10 mM Tris-HCl (pH 8.0),

50 mM KCl, 1.5 mM $MgCl_2$, 0.02% (w/v) gelatin, 0.2 mM of each dNTP, 0.08 mM of each primer and 1 unit of Taq DNA polymerase (Gibco BRL).

Samples were heated to 94 °C for 5 minutes, followed by 40 cycles at 94 °C for 1 minute, 55 °C for 1 minute and 72 °C for 1 minute. Samples were then incubated at 72 °C for 7 minutes to extend PCR products and 98 °C for 5 minutes to create heteroduplexes. Amplification products were confirmed by running 10% of amplification products on a 1.4% agarose (Gibco BRL) gel and staining with ethidium bromide.

ii) Polymerase Chain Reaction for Direct Sequencing

Amplifications were performed in 100 µl containing ~1 µg genomic DNA, 10 mM Tris-HCL pH 8.0, 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% (w/v) gelatin, 0.2 mM of each dNTP, 0.4 mM of each primer and 2.5 units Taq polymerase (Gibco BRL).

Samples were heated at 97 °C for 7 minutes and then cooled to the annealing temperature (T_A) of the exonic primer (Figure 2-3) for 5 minutes at which time the Taq and PCR buffer was added. Thermo cycling was then performed; 92 °C for 1 minute, T_A for 1 minute and 72 °C for 1.5 minutes; 35 cycles. PCR products were then extended by heating to 72 °C for 5 minutes. Amplification products were confirmed by visualizing 10 µl on a 1.4% agarose gel.

Figure 2-2. *PAH* exonic primers and their annealing temperatures used in PCR amplifications. Primers are located in the flanking intronic regions.

1A	GCTTTACTGTGCGGAGATCACCAC	} T_A =56 °C
1B	CTTATGAAACCAGGAAGCAC	
2A	GAGTTCATGCTTGCTTTGTG	} T_A =56 °C
2B	CCAGATCCTGTGTTCTTTTC	
3A	CTTAGGTTTTCCTGTTCTGG	} T_A =52 °C
3B	CTTATGTTGCAAAATTCCTC	
4A	ATGTTCTGCCAATCT	} T_A =40 °C
4B	CAAGACACAGGCCATGGACT	
5A	TCATGGCTTTAGAGCCCCCA	} T_A =54 °C
5B	TCATGCTGGTATTTTCATCC	
6A	CACAGGTTCTGGTCCCCGAC	} T_A =60 °C
6B	CTCTCCTCTCCTCAATCCTC	
7A	CTCCTAGTGCCTCTGACTCA	} T_A =60 °C
7B	ACCAGCCAGCAAATGAACCC	
8A	CCTCATGTAGAAAGACTGAG	} T_A =48 °C
8B	CTCTTGCAGAGGGCAT	
9A	TCTGGCCACCCATCACCTTT	} T_A =58 °C
9B	CTATAGCACTCCACCATCCA	
10A	CCCAGTCAAGGTGACACATA	} T_A =56 °C
10B	ACAAATAGGGTTTCAACAAT	
11A	TGAGAGAAGGGGCACAAATG	} T_A =56 °C
11B	GTAGACATTGAGTCCACTCT	
12A	ATGCCACTGAGAACTCTCTT	} T_A =52 °C
12B	AGTCTTCGATTACTGAGAAA	
13A	GACACTTGAAGAGTTTTTGC	} T_A =52 °C
13B	TTTTCGGACTTTTTCTGATG	

2.2.3 Mutation Detection using Denaturing Gradient Gel Electrophoresis:

A) Identification of mutation containing exon

A modified form Denaturing Gradient Gel Electrophoresis (DGGE) (Myers et al., 1987); 'Broad Range' DGGE (Guldberg et al., 1993, 1994) was used to determine exons that harboured abnormal *PAH* sequences.

i) Gel Casting

The denaturing gradient gel was cast and run on a Protean II xi Vertical Electrophoresis apparatus (Bio Rad). The Protean II was first modified by removing the central cooling core. This modification allows both sides of the glass plates to be maintained at the same temperature (65 °C). 16 cm glass plates were first treated with Kodak Photo Flo 600 and assembled with 0.5mm spacers and sandwich clamps. First, 6.3 ml of 6%, 9.5M urea polyacrylamide gel was poured to create a constant denaturant at the bottom of the gel. A model 385 gradient former (Bio Rad) was then used to create the denaturing (urea) gradient. The gel was a 6% polyacrylamide gel with a vertically increasing concentration of urea (0 to 9.5M) from top to bottom. Once poured a 0.5mm, 25 well comb was inserted and the gel polymerizes for one hour.

ii) Running of Denaturing Gradient Gel

The denaturing gradient gel was run in a ~25 gallon tank containing 1X TAE. The TAE was circulated and kept at a constant temperature of 60 °C with a HAAKE E2 Thermo Circulator. 15 µl of each sample was run on the gel at 155 volts for 4 1/2 hours. The gel was then stained with EtBr and photographed under U.V. transillumination (300nm).

B) Identification of mutation through sequencing :

Once mutant exons were determined from DGGE analysis they were then amplified and sequenced using the dsDNA Cycle Sequencing System (GIBCO BRL).

i) DNA purification

To purify PCR reactions for dsDNA cycle sequencing (GIBCO BRL) the GlassMAX DNA isolation spin cartridge system (GIBCO BRL) was used. 90 µl of a 100 µl PCR reaction was mixed with NaI binding Solution. This mixture was then run through a GlassMAX spin cartridge by centrifugation (13000 rpm for 30 sec.) and washed four times with cold wash buffer. DNA was then eluted with 40 µl of 65 °C sterile water and 10% of the product was run on a 1.4% agarose gel. Purification yields and DNA concentration was performed by comparing 10% of the purification product to a ϕX HaeIII DNA ladder.

ii) Sequencing reactions and sequencing gel

PAH primers (Figure 2-3) (1 pmol) were end labelled with gamma ^{32}P ATP. These primers were then used in radioactive PCR reactions which use the purified PCR products (50 fmol) as a DNA template. The Sanger dideoxy chain termination procedure was used to terminate PCR reactions at each nucleotide (Sanger et al., 1977). Electrophoresis was performed by loading 3 µl of PCR products on 6% polyacrylamide gels (Sequa-Gel; National Diagnostics). Extended sequence information was obtained by long and short gel runs (4.5 hrs and 2.5hrs respectively) and by sequencing in both sense and antisense directions. The gels were dried for 2 hours at 85 °C on a Bio Rad gel dryer and autoradiographed with Kodak BioMax film overnight.

2.2.4 Mutation Confirmation

A) Allele specific oligonucleotide analysis

i) End labeling of Probe

Allele specific oligonucleotides were end labelled with γ^{32} pATP. 20 µM of probe (Table 2-1) was added to 7 units T4 Kinase (Gibco BRL) and 1.4 µl Kinase buffer (350 mM Tris-HCl (pH 7.6), 50 mM $MgCl_2$, 500 mM KCl, 5mM 2-mercaptoethanol). This mixture was then incubated at 37 °C for 1 hour. After incubation the probe was purified using a P4 Biogel column (Bio Rad). The sample was loaded onto the column and spun for five minutes at 3000 rpm. An aliquot of the elutant was then counted on a scintillation counter.

ii) Applying DNA to membrane

Amplified DNA was linked to a membrane by first adding 10 - 20 µl of an amplification product to 100 µl of denaturing solution (0.4 N NaOH, 25 mM EDTA Na_2) and vortexing for 30 seconds. Samples were applied to pure nitrocellulose membranes (Schleicher & Schuell) using a dot blot manifold (Bio Rad). Each well was first washed with 100 µl 20X SSPE (3 M NaCl, 20 mM NaH_2PO_4 pH7.4 and 20 mM EDTA) samples were then applied to the membrane and the wells were again washed with 20X SSPE. The DNA was cross linked to the membrane by exposure to UV light for 3 minutes and baking at 80 °C for 1 hour.

iii) Hybridization

For hybridization with allele specific oligonucleotide (ASO) (mutant probe), membranes were placed in sealable plastic bags and 10 ml of prehybridization buffer (5X SSPE, 0.5% SDS and 5X Denhardt's) and 20 million counts of labelled probe were added and placed at Tm -2°C for one hour. The membrane was then washed twice at room temperature with washing buffer (2X SSPe, 0.1% SDS) and once for 10 minutes at Tm -2°C with the same

buffer after which the membrane was autoradiographed with Kodak XAR film (Table 2-1).

Membranes were stripped by washing with a mild stripping solution (1 M NaCl, 0.5 M NaOH) until geiger counter readings were very low or undetectable. Membranes were then reprobed with a 'normal' probe.

B) Restriction Digests

Depending on PCR amplification success, usually 17 µl of a 100 ml PCR reaction was digested with 2 µl 10X react buffer and 1 ml (1-10 units) of restriction enzyme. Digests were incubated at 37 °C overnight or for three hours at 65 °C depending on the restriction enzyme. After digestion was complete 4 µl of 5X loading buffer (30% glycerol, 0.2% bromophenol blue and 0.2% xylene cyanol) was added and the entire reaction volume was run on either agarose or acrylamide gels depending on fragment size (Table 2-2).

C) Artificially created restriction site (ACRS)

Primers were designed to introduce a single base change that, in conjunction with a putative *PAH* mutation, would change a restriction site (Table 2-3). After amplification (see section 2.1 ii) with these primers 17µl of a 100µl PCR amplification was digested (see section 2.32). The entire 20 µl of digest was then visualized on an 8% polyacrylamide gel.

Table 2-1. The sequences and conditions used for dot blot analysis of single nucleotide changes.

ASO Identity	T_a (°C)	Wash Temp (°C)	Wash Time (min)	Sequence 5'- 3'
K42I Normal	52	50	10	CTCACTCAAAGAAGAAGTT
K42I Mutant	54	52	10	CTGATCTTCTCACTCATAGA

Table 2-2. Restriction enzymes and gel type used to confirm various mutation in the PAH gene.

MUTATION or POYYMORPHISM	ENZYME	GEL TYPE and %
F39L	*Mbo* II	Acrylamide (8%)
L48S	*Msc* I	Acrylamide (8%)
Ivs5 nt-54	*Bsm* I	Agarose (1.4%)
R157N	*Bsa WI*	Agarose (1.4%)
R158Q (ACRS)	*Msp* I	Acrylamide (8%)
Q232Q	*Dde* I	Acrylamide (8%)
V245V	*Alu* I	Acrylamide (8%)
V245A	*Bso FI*	Agarose (1.4%)
Ivs12 nt1 (ACRS)	*Rsa* I	Acrylamide (8%)
L385L	*Mnl* I	Acrylamide (8%)

Table 2-3. ACRS primers and restriction enzymes to perform ACRS mutation confirmation in the *PAH* gene. Primers that introduce base changes are used with normal *PAH* primers (see Figure 2-3).

MUTATION	ACRS PRIMERS 5'-3'	ENZYME
R158Q	ATCCTGTGTACCGTGCAAGC + PAH 5B	*Msp* I
Ivs12 nt1	PAH 12A + CTCGTAAGGTGTAAATTACGTA	*Rsa* I

2.3.0 Mutation nomenclature

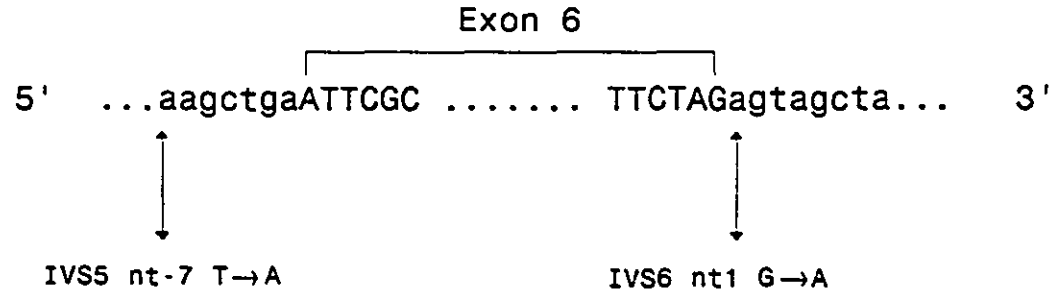
The nomenclature for designating mutations that has been used in this thesis and has been adopted by the *PAH* mutation analysis database generally follows that set out by Beaudet and Tsui (1993) with a few modifications.

The database has specified trivial and systematic nomenclature to designate missense mutations. The trivial name uses the single letter code and the amino acid position to describe the amino acid change and position of the mutation (ie K42I). The systematic name is more exact, using the nucleotide position in the cDNA and stating the specific change at this position (ie. c125 A→T).

Nonsense mutations are designated like missense mutations, except that X is used to indicate a termination codon.

Splicing mutations use the abbreviation IVS (intervening sequence) (Antonarakis and McKusick 1994) followed by the number of the intron in which the mutation is located. This is followed by the number of the nucleotide that is mutated. Since the entire intronic *PAH* sequence is not known, the nucleotide number is designated by counting from the first 5' or 3' flanking intronic nucleotide depending on which side of the exon the mutation falls. If the mutation is 5' to the exon then a minus (-) is used before the nucleotide number. Lastly the specific nucleotide change is stated, again using the one letter code and an arrow (indicating from Y → to Z) (Figure 2-3).

Figure 2-3. Two hypothetical splice mutations in introns 5 and 6 indicating the nomenclature that would be used for splice mutations 5' and 3' to an exon.



2..4.0 Predicted mutability of the Phenylalanine Hydroxylase gene

The predicted mutability profile of the *PAH* gene (cDNA) was obtained using the MUTPRED program (Cooper, Krawczak 1993). The cDNA sequence of the *PAH* gene was entered into the program and a plot of points is then given based on relative mutability calculations. These points were then imported into the Harvard Graphics program to produce a graphical representation of the realtive mutability of the *PAH* gene.

Chapter 3
RESULTS

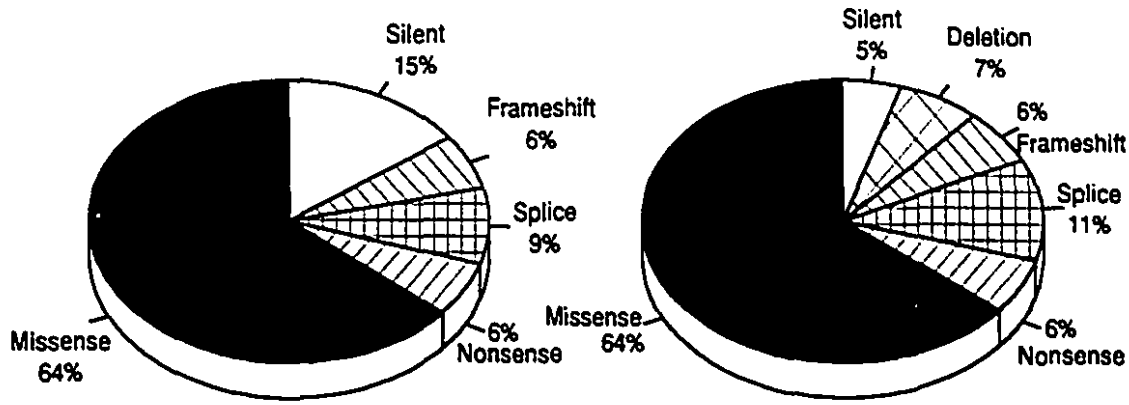
3.1.0 Types of mutations found on Quebec chromosomes

In the 300+ mutations of the phenylalalnine hydroxyalse (*PAH*) gene missense, nonsense, termination, splice, insertion, deletion, frameshift and silent mutations are all present. As in most genes these mutation types are found at varying frequencies (i.e. CFTR). Missense mutations make up the majority of mutations in the *PAH* gene (64%), they are followed by splice mutations (11%), deletions (3 base pairs +) (7%), nonsense mutations (6%) and silent polymorphisms (5%). While the number of mutations in Quebec is only one seventh that of global *PAH* mutations these 40 mutations are an accurate representation of the global spectrum of mutations (Figure 3-1). Quebec's majority of mutations are also missense mutations equaling the global frequency at 64%. Also equaling global frequencies are nonsense mutations and frameshift mutations (6%). Splice mutations are slightly less frequent (9%) while silent mutations are slightly more (15%). No large deletions have been identified in Quebec (Figure 3-1).

3.2.0 Mutation map of Quebec PAH mutations

The 46 disease producing and neutral mutations found in Quebec are randomly distributed throughout the gene (Figure 3-2). Two exons, 4 and 13, have no known *PAH* mutations in the Quebec population. Exon four harbors only 8 mutations on the global map of *PAH* mutations (Figure 1-3, Introduction), and has one CpG dinucleotide (Figure 3-9 section 3.3). Likewise, exon thirteen has only one mutation reported to the

Figure 3-1. Classification of global and Quebec *PAH* mutations.



Quebec Mutations

World Mutations

(Data From PAH Mutation Analysis Database)

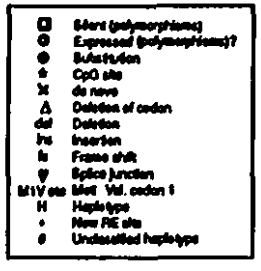
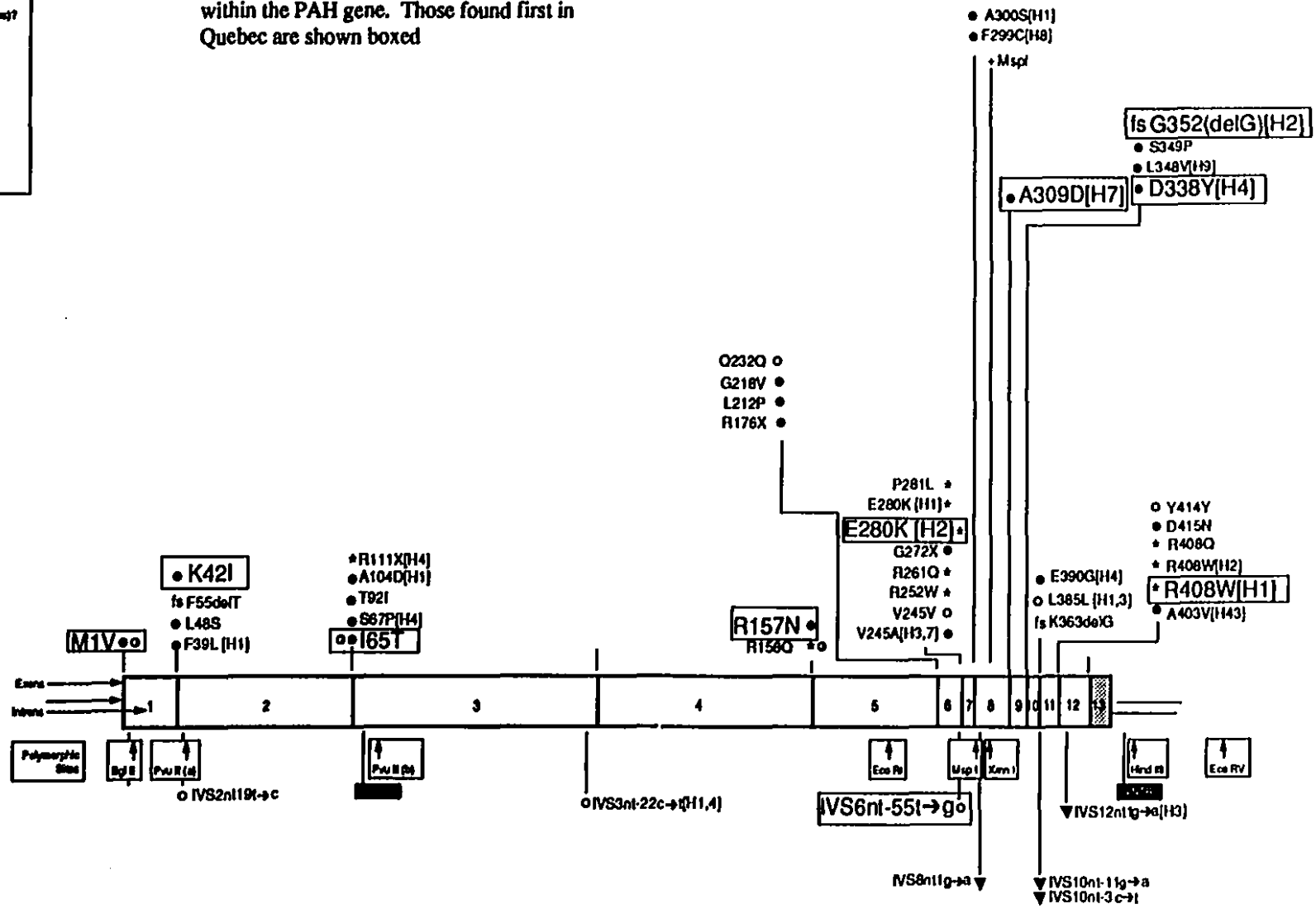


Figure 3-2. The location of Quebec PAH mutations within the PAH gene. Those found first in Quebec are shown boxed



consortium and has no CpG sites (Figure 3-9 section 3.3). The rest of the Quebec *PAH* mutations are evenly distributed throughout the gene. The four studies (John et al. 1992; Treacy et al. 1993; Rozen et al. 1994; current study) of Quebec *PAH* mutations has identified 10 novel mutations (boxed mutations Figure 3-2). Five of these mutations (K42I, R157N, E280K[H2], D338Y and Ivs6nt-55) remain unique to the Quebec population (*PAH* mutation analysis database 1996).

3.3.0 Mutation Analysis

Seventy one phenylketonuria (PKU) families were analyzed. All thirteen *PAH* exons were amplified using the polymerase chain reaction (PCR). Individual exons were then analyzed using 'Broad Range' Denaturing Gradient Gel Electrophoresis (DGGE). Exons which revealed heteroduplex formations were directly sequenced (see materials and methods sections 2.1.1 to 2.1.3). 38 different *PAH* mutations in 40 different mutation/haplotype combinations were identified in the Quebec population. Two mutations in the United States and one mutation in Kuwait were also identified.

Following is a summary of these findings which are all published in the *PAH* mutation analysis database (http//www.mcgill.ca/pahdb).

3.3.1 Novel Mutations:

Four mutations (K42I, R157N, IVS6 nt-55, R241C), three of which are believed to be disease causing (K42I, R157N, R241C), were previously unreported. No other changes were found on chromosomes which harboured novel mutations (except for silent polymorphisms on some chromosomes), furthermore these novel mutations were not

found on more than 100 normal chromosomes (data not shown). The novel mutations that were identified in the Quebec population (K42I, R157N, IVS6 nt-55) are presented here.

a. The K42I mutation

DGGE analysis identified a heteroduplex formation in exon two (figure 3-3). Upon sequencing an A to T transversion at nucleotide 125 was detected (figure 3-4). The substitution changes the codon from Lysine (AAA) to Isoleucine (ATA). This result was confirmed by Dot Blot analysis with an allele specific oligonucleotide (see table 2-1, section 2.1.3) (figure 3-5). K42I was found on the paternal allele of the proband, haplotype 21 (determined from previous study) and on only one chromosome in the Montreal region. The other mutation found in this proband was E280K; a mutation associated with classical PKU (*PAH* expression database) (Kayaalp et al. in progress). The compound heterozygosity of these two mutations produces a PKU phenotype. This mutation has been given the trivial name of K42I and the systematic name c125 A-T.

Figure 3-3. DGGE pattern for novel mutations.

GC-clamped amplifications of exons 2 and 5 were run on a denaturing gradient gel. The unique banding pattern of the novel K42I mutation (lane 2) was compared to normal (lane 1), F39L (lane 3), and L48S (lane 4) banding patterns. Likewise, in exon 5, the unique banding pattern of the novel R157N mutation (lane 6) was compared to normal (lane 5) and R158Q (lane 7) banding patterns.

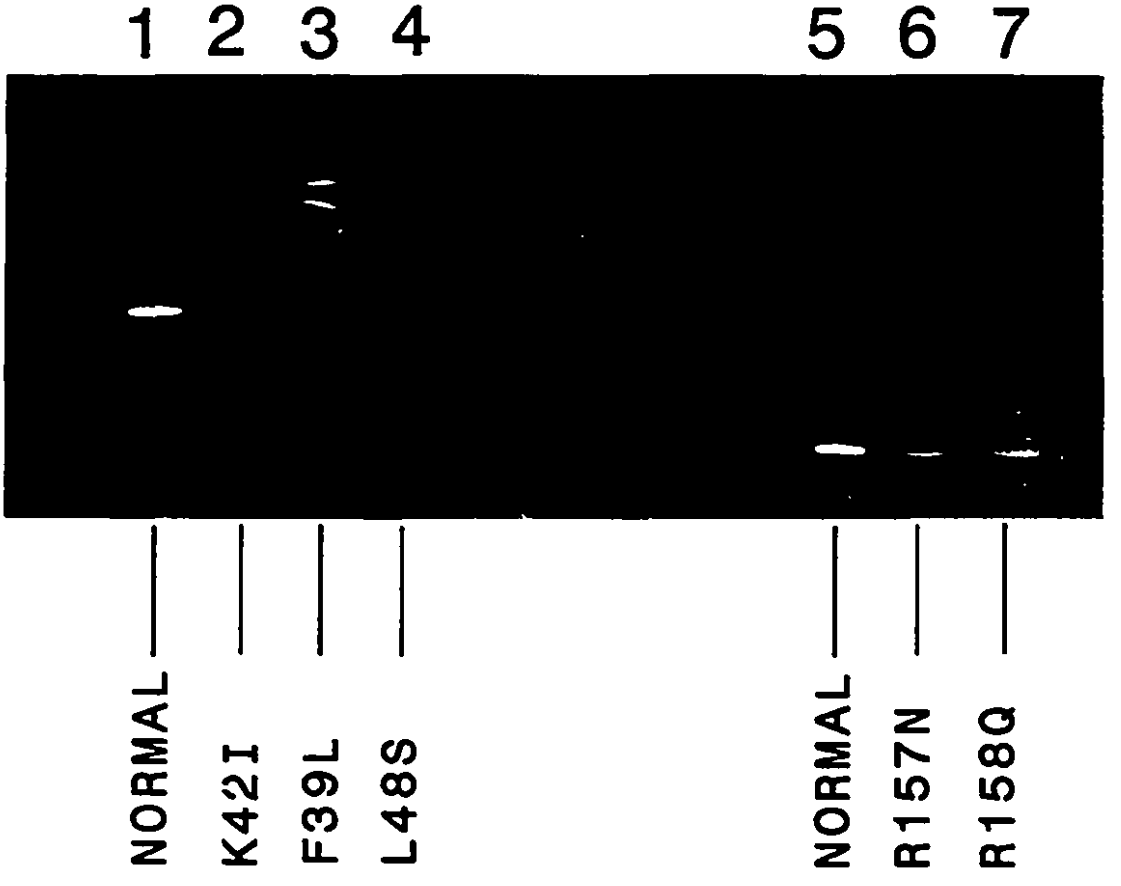

1 2 3 4 5 6 7
NORMAL
K42I
F39L
L48S
NORMAL
R157N
R158Q

Figure 3-4. Sequencing of the K42I mutation

Partial sequence ladders for *PAH* exon 2 from a proband heterozygous for the novel K42I mutation and a normal control carrying no mutation. The Lysine to Isoleucine change at codon 42 is indicated.

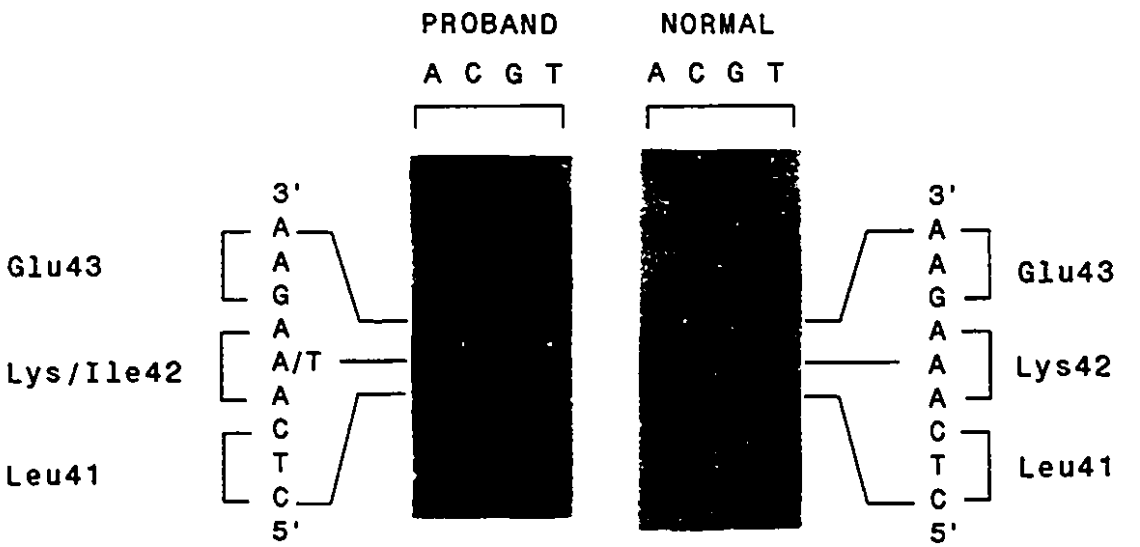
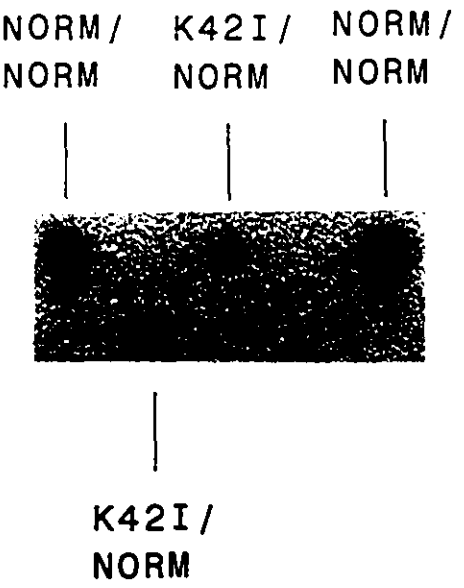

PROBAND
NORMAL
A C G T
A C G T
3'
A
A
G
A
A/T
A
C
T
C
5'
Glu43
Lys/Ile42
Leu41
3'
A
A
G
A
A
A
C
T
C
5'
Glu43
Lys42
Leu41

Figure 3-5. Dot Blot confirmation of the K421 mutation

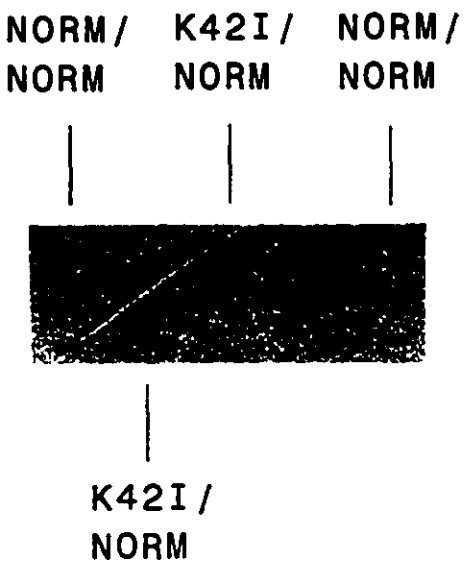
A) Amplified exon 2 fragments were dot blotted and screened with a normal probe. The probe bound to both normal exon two amplifications and to a proband heterozygous for a putative K42I mutation.

B) The blot was stripped and screened with a probe specific for the K42I mutation. The probe bound only to the amplified fragments from the proband who was heterozygous for the K42I mutation.

A


NORM/
NORM
K42I/
NORM
NORM/
NORM
K42I/
NORM

B

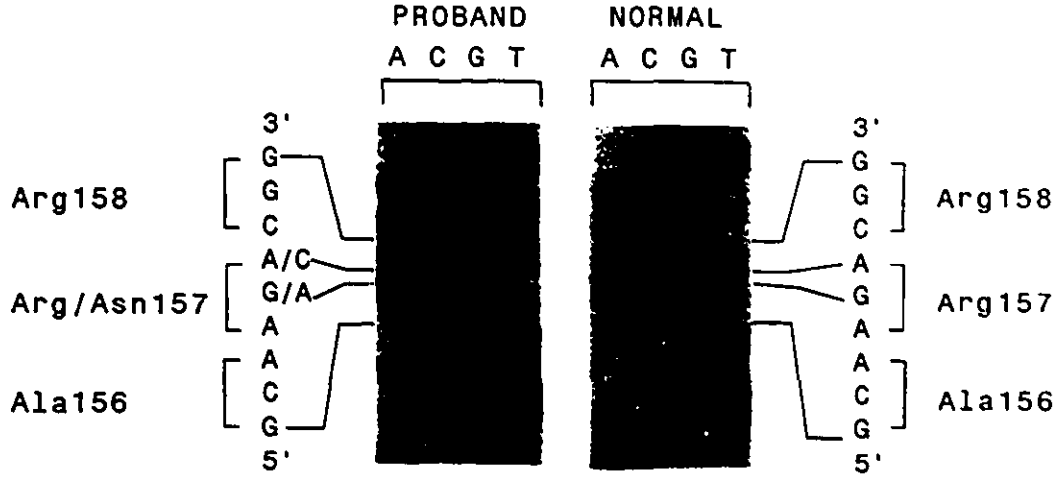

NORM/
NORM
K42I/
NORM
NORM/
NORM
K42I/
NORM

b. The R157N mutation

DGGE analysis revealed a heteroduplex formation in exon five (figure 3-3). Upon sequencing a rare dinucleotide (double base pair) change was found at nucleotides 470 and 471 in codon 157. The first nucleotide change in this codon is a G to A transition followed by an A to C transversion (figure 3-6). The two nucleotide changes in combination change the codon from an arginine (AGA) to an asparagine (AAC) and introduce a B*sa* WI restriction site in exon five. The mutation was confirmed using restriction analysis of amplified exon five with B*sa* WI (data not shown). This mutation was found on the paternal allele of the proband and on an unknown haplotype. It is a rare mutation being found on only one chromosome located in the Eastern Quebec region. The other allele carried by this proband was F299C; a null mutation associated with classical PKU (*PAH* expression database) (Kayaalp et al. in progress). The compound heterozygosity of these two mutations produces a PKU phenotype. To date, this is the only double base pair mutation in a single codon reported to the *PAH* mutation analysis database. The mutation has been designated a trivial name of R157N and the systematic name of c470/471 GA→AC.

Figure 3-6. Sequencing of the R157N mutation

Partial sequence ladders for *PAH* exon 5 from a proband heterozygous for the novel R157N mutation and a normal control carrying no mutation. The dinucleotide substitution and the arginine to asparagine change at codon 157 are indicated.


PROBAND
A C G T
NORMAL
A C G T
3'
G
G
C
A/C
G/A
A
A
C
G
5'
Arg158
Arg/Asn157
Ala156
3'
G
G
C
A
G
A
A
C
G
5'
Arg158
Arg157
Ala156

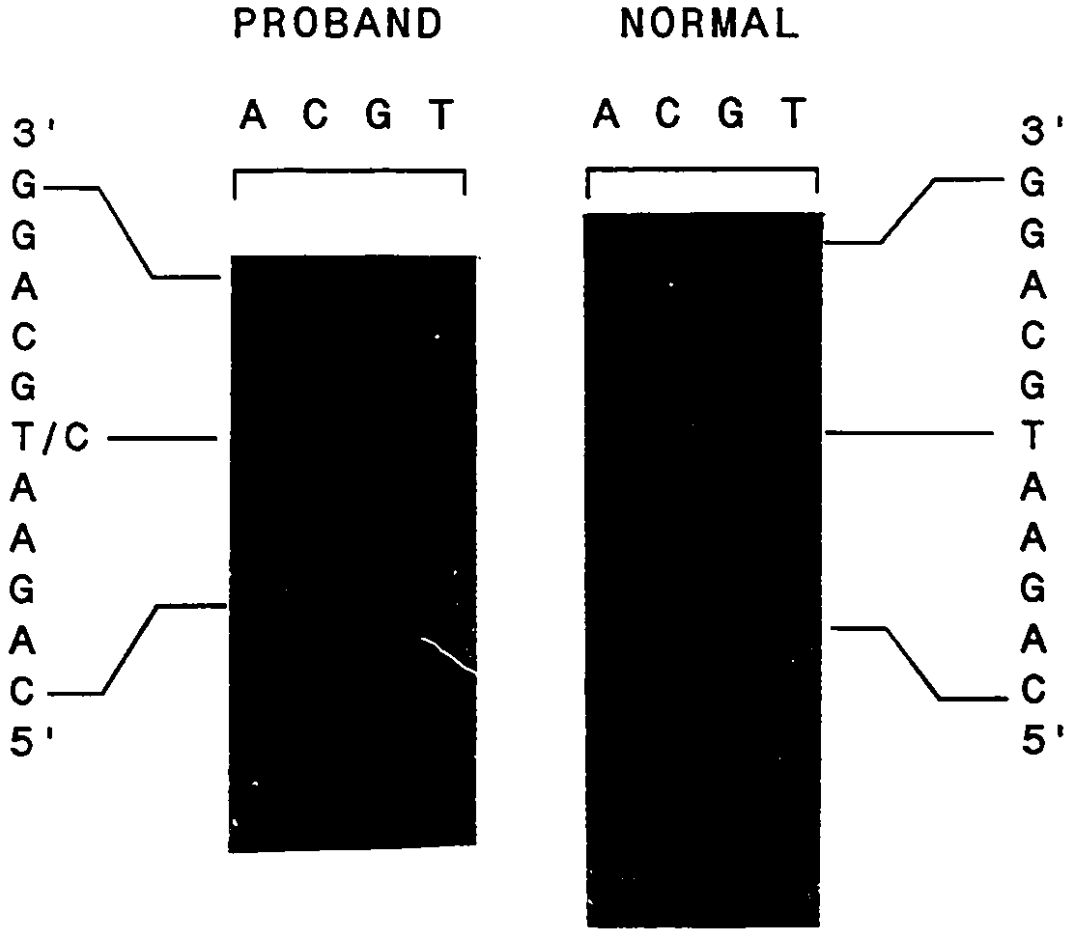
c. The IVS6 nt-55 mutation

A unique banding pattern was identified by DGGE analysis of exon six (data not shown). When sequenced a single nucleotide change was found in intron five, 55 nucleotides upstream from the first nucleotide of exon five (figure 3-7). The nucleotide change is an A to C transversion and introduces a B*sm* I restriction site. The mutation was confirmed using B*sm* I digestion of amplified exon 6 (data not shown). This mutation was found on three chromosomes, two of which harboured previously identified mutations. This, along with the fact that the mutation is deeply intronic leads to the belief that IVS6 nt-55 is a silent (neutral) mutation. Since the mutation was found on three chromosomes in 141 it has a frequency of 1.4% and is at a polymorphic frequency in Quebec.

3.3.2 Novel mutation/haplotype associations: Using the diallelic haplotyping system (previous studies), 3 novel mutation/haplotype associations (S67P [H1]; G218V [H2]; V245A [H7]) were found on Quebec chromosomes (Table 3-1). Each of these novel haplotypes could be easy formed by recombinations within and around the *PAH* locus (see discussion section section 4.2). However, other mutation mechanisms (recurrent mutation, gene conversion) cannot be ruled out.

Figure 3-7. Sequencing of the IVS6 nt-55 polymorphism.

Partial sequence ladders for *PAH* exon 6 from a proband heterozygous for the novel IVS6 nt-55 mutation and a normal control carrying no mutation. The intronic thymine to cytosine change is indicated.


PROBAND
NORMAL
A C G T
A C G T
3'
G
G
A
C
G
T/C
A
A
G
A
C
5'
3'
G
G
A
C
G
T
A
A
G
A
C
5'

3.3.2 Novel mutation / haplotype associations

Table 3-1. Mutations found in Quebec which have a novel mutation/haplotype association. All mutations are compared with other populations which have reported mutation/haplotype associations.

Mutation	Haplotype	Population/Location	Reference
S67P	4	Wales	Tyfield et al. 1993
	4	Italy	Dianzini et al. 1995
	4.3	S.W. England	Tyfield L. 1995 to Consortium
	1	Montreal	Carter K. 1996 to Consortium
G218V	1	Denmark	Guldberg et al. 1993
	1	France	Rey F. 1995 to Consortium
	1	Belgium	Michiels L. 1995 to Consortium
	2	Western Quebec	Carter K. 1996 to Consortium
V245A	3	Denmark	Guldberg et al. 1994
	3	N. Ireland	Zschoche et al. 1994
	3	Spain	Perez et al. 1994
	7	Western Quebec	Carter et al. 1995

3.3.3 Mutations found in Quebec and other populations:

Nineteen mutations that were previoulsy unreported in the Quebec population were found using the DGGE technique. 15 of these mutations are disease causing alleles of which 11 are rare (relative frequency < 1%) and 4 are common in Quebec (Table 3-2; Table 3-4). The other 4 mutations are polymorphic (relative frequency > 1%) neutral mutations which have no known phenotypic effects. These neutral polymorphisms are associated with only a small number of polymorphic haplotypes and are commonly found with various disease causing alleles (Table 3-3).

A) The disease causing *PAH* mutations found in Quebec and other populations

Table 3-2. Disease causing alleles were previously identified in other populations. These are listed with their diallelic haplotypes (if known) in both Quebec and the external populations.

Mutation	Quebec Region	Haplotype	Other Populations	Haplotype
F39L	Eastern Quebec	ND	Australia [a,b]	1.8
	Western Quebec	1	Denmark [c]	1
			Spain [d]	24.3
			New England [f]	ND
			Belgium [h]	1
			S.W. England [i]	1
			Czech Republic [j]	1.8
			W. Scotland [i]	ND.8
			N. Ireland [k]	ND.8
L48S	Western Quebec	4	Turkey [m,o]	4
	Montreal	ND	Germany [n,q,u]	3 / 4.3
			Sicily [c,p]	28.3
			Spain [d,s]	4.3
			Romania [r]	ND
			Belgium [h]	3 / 4
			Australia [b]	4.3
			Italy [t]	4
			Czech Republic [j]	4.3
			S.W. England [j]	4.3
			W. Scotland [i]	ND.8
			France [v]	ND.4
F55fs	Western Quebec	ND	Europe [w]	1
			Sicily [c,p]	1.8
			Poland [x]	1
			France [v]	1
			Italy [t]	1
			Germany [u]	ND
			Belgium [h]	ND
S67P	Montreal	1	Wales [y]	4
			Italy [t]	4
			S.W. England [i]	4.3
			N. Ireland [k]	ND.3
A104D	Montreal	ND	European [w]	1
			Denmark [c]	1
			Sicily [c]	ND
			The Netherlands [z]	ND
			France [v]	ND
R111X	Western Quebec	ND	Chinese [1]	4
			Denmark [c]	4
			Sicily [c,p]	4.3 / 28.3
			Japanese [2]	ND
			Mid West USA [3]	ND
			S.W. England [i]	5.8
R176X	Western Quebec	ND	Sicily [c]	ND
			Egypt [4]	ND
			Belgium [h]	ND

Mutation	Quebec Region	Haplotype	Other Populations	Haplotype
L212P	Eastern Quebec	ND	USA [u]	ND
G218V	Western Quebec	2	Denmark [c]	1
	Eastern Quebec	ND	France [v]	1
			Belgium [h]	1
			Wales [i]	ND.8
E221D222fs del AG	Eastern Quebec	ND	Denmark [c]	4
			Tartaria [12]	ND
			Netherlands [h]	ND
V245A	Western Quebec	7	Denmark [g]	3
	Eastern Quebec	ND	N. Ireland [l]	3.8
			Spain [d]	3.8
			France [v]	ND
			Germany [w]	ND
			Poland [5]	ND
			Belgium [h]	ND
Ivs8 nt1	Montreal	4	USA [6]	ND
			N. Ireland [k]	ND.3
K363fs	Montreal	ND	USA [7]	ND
			Germany [u]	ND
			Italy [t]	ND
Ivs10 nt-3	Western Quebec	ND	France [g]	4
			Spain [d]	6.7
			S.E. England [9]	ND
			Czech Republic [j]	4.3
			Chile [e]	6.7
			N. Ireland [k]	ND
			Belgium [h]	ND
			Poland [10]	ND
A403V	Western Quebec	ND	Sicily [o,p]	ND.8
			Spain [d]	ND
			Belgium [h]	1
			Italy [t]	1
			Czech Republic [j]	1.8
			France [v]	43
			Germany [w]	ND
			Poland [11]	ND

[a] Forrest et al. 1991
[b] Ramus et al. 1995
[c] Guldberg et a. 1993
[d] Perez et al. 1994
[e] Perez et al. 1995
[f] Guldberg et al. 1994 [a]
[g] Guldberg et al. 1994 [b]
[h] Michiel S 1995 to Consortium
[i] Tyfield L. 1995 to Consortium
[j] Kozak L. 1995 to Consortium
[k] Zschocke et al. 1995
[l] Zschoche et al. 1994
[m] Lichter-Konecki et al. 1990
[n] Konecki et al. 1991
[o] Kalaydjieva et al. 1993
[p] Romano et al. 1994
[q] Aulehla-Scholz 1994 to Consortium
[r] Popescu et al. 1994
[s] Martinez-Pardo et al. 1994
[t] Dianzini et al. 1995
[u] Guldberg P. 1995 to Consortium
[v] Rey 1995 to Consortium
[w] Horst et al. 1995
[x] Zygulska et al. 1994
[y] Tyfield et al. 1993
[z] Ploos Van Amstel 1994 to Consortium
[1] Huang et al. 1990
[2] Takarada et al. 1993
[3] Kaul et al. 1993
[4] Goltsov et al. 1994
[5] Zekanowski 1995 to Consortium
[6] Guldberg et al. 1995
[7] Guldberg P. 1993 to Consortium
[8] Abadie 1993 to Consortium
[9] Cadiou 1994 to Consortium
[10] Jaruzelska et al. 1995
[11] Zekanowski 1996 to Consortium
[12] Kuzmin et al. 1995

B) Polymorphic neutral mutations found in Quebec and other populations

Table 3-3. Polymorphic neutral mutations that have been previously identified in other populations. Polymorphisms are listed along with their diallelic haplotypes (if known) in both Quebec and the population in which they had been previously identified.

Mutation	Quebec Region	Haplotype	Other Populations	Haplotype
Ivs2 nt19	Montreal	ND	European [a]	11,5,6
			Egypt [b]	ND
			Belgium [c]	11,5
			Italy [d]	ND
Ivs3 nt-22	Montreal	ND	Belgium [c]	1,4
	Western Quebec	1,4.3,ND	Norway [e]	4.3
	Eastern Quebec	1,4.3	Italy [d]	ND
L385L	Montreal	1	Australia [f]	ND
	Western Quebec	1,3.8	Portugal [g]	ND
	Eastern Quebec	3.8,ND	Belgium [c]	1,3,7
			Poland [h]	ND
			Italian [d]	ND
Y414Y	Montreal	1	England [i]	ND
			Belgium [c]	1 i

[a] Lichter-Konecki et al. 1994
[b] Eisensmith R.C. 1994 to Consortium
[c] Michiels L. 1995 to Consortium
[d] Dianzani et al. 1995
[e] Apold J. 1995 to Consortium
[f] Forrest et al. 1991
[g] Leandro et al. 1993
[h] Zekanowski C. 1995 to Consortium
[i] Tyfield L. 1994 to Conortium

3.4.0 Relative Frequency of Quebec PKU mutations

Previous data (Rozen et al. 1994) and mine were combined to derive relative frequencies of *PAH* mutations in Quebec (Table 3-4). The relative frequency is obtained by dividing the number of independent chromosomes bearing a particular mutation by the number of independent mutant chromosomes in each region (see section 2.03, Figure 2-2). French Canadian *PAH* mutations have a different relative frequency from those found in France, the country of origin for French Canadian ancestors (Harris and Mathews 1989) (Table 3-4).

Table 3-4. Relative frequencies of *PAH* mutations in Probands with PKU or Variant.

Chromosome n	Province [a] 141	W. Quebec 57	Montreal 37	E. Quebec 47	France [b] 372
Mutation					
M1V	5.7	1.8	0	14.9	1.1
F39L	4.3	8.8	0	2.1	-
K42I	0.7	0	2.7	0	-
L48S	2.3	3.5	2.7	0	P
F55fs delT	0.7	1.8	0	0	P
F55L	0	0	0	0	P
Ivs2nt5	0	0	0	0	P
I65T	4.3	3.5	5.4	4.3	P
S67P	0.7	0	2.7	0	-
R68S	0	0	0	0	P
dell94	0	0	0	0	0.5
A104D	0.7	0	2.7	0	P
R111X	0.7	1.8	0	0	0
H146Y	0	0	0	0	P
R157N	0.7	0	0	2.1	0
R158Q	2.8	3.5	2.7	2.1	1.1
G171A	0	0	0	0	P
R176X	0.7	1.8	0	0	-
E178G	0	0	0	0	P
L212P	0.7	0	0	2.1	-
G218V	2.1	1.8	0	4.3	P
E221D222fs	0.7	0	0	2.1	-
G239S	0	0	0	0	P
R241H	0	0	0	0	P
R241C	0	0	0	0	P
R243Q	0	0	0	0	P
V245A	1.4	1.8	0	2.1	P
L249F	0	0	0	0	1.3
R252W	0.7	1.8	0	0	1.6
R252Q	0	0	0	0	P
A259V	0	0	0	0	0.5
R261Q	5.7	1.8	18.9	0	6.2
C265G	0	0	0	0	P
G272X	1.4	1.8	0	2.1	0.5
S273F	0	0	0	0	0.5
Y277D	0	0	0	0	0.3
E280K [H1]	0.7	0	2.7	0	4.8

Chromosome n	Province [a] 141	W. Quebec 57	Montreal 37	E. Quebec 47	France [b] 372
Mutation					
E280K [H2]	3.5	5.3	0	4.3	0
P281L	2.1	1.8	2.7	2.1	P
F299C	6.4	10.5	5.4	2.1	P
A300S	0.7	1.8	0	0	P
IVS8NT1	0.7	0	2.7	0	-
A309D	2.8	5.3	0	2.1	-
L311P	0	0	0	0	0.5
P314H	0	0	0	0	P
F331L	0	0	0	0	P
D338Y	1.4	3.5	0	0	-
LSSFfsdel11bp	0	0	0	0	P
L348V	1.4	3.5	0	0	P
S349P	5.7	1.8	2.7	12.8	P
S350T	0	0	0	0	P
1064delG	1.4	1.8	2.7	0	-
1054/55delG	0.7	0	2.7	0	-
IVS10nt-11	1.4	0	2.7	2.1	P
IVS10nt-3	0	0	0	0	P
A373fsdelGC	0	0	0	0	P
E390G	0	0	0	0	P
IVS11nt1	0	0	0	0	P
A403V	0.7	1.8	0	0	P
P407fs delC	0	0	0	0	P
R408W [H1]	7.1	5.3	10.8	6.4	0.3
R408W [H2]	4.3	1.8	10.8	2.1	5.6
R408Q	0	0	0	0	-
R413S	0	0	0	0	P
Y414C	0	0	0	0	P
IVS12nt1	17.9	15.8	10.8	17.0	5.6
Identified %	**92.5**	**96.4**	**94.6**	**89.3**	

Symbols: 0, searched for and absent;
-, not searched for;
P, present but realtive frequency unknown.

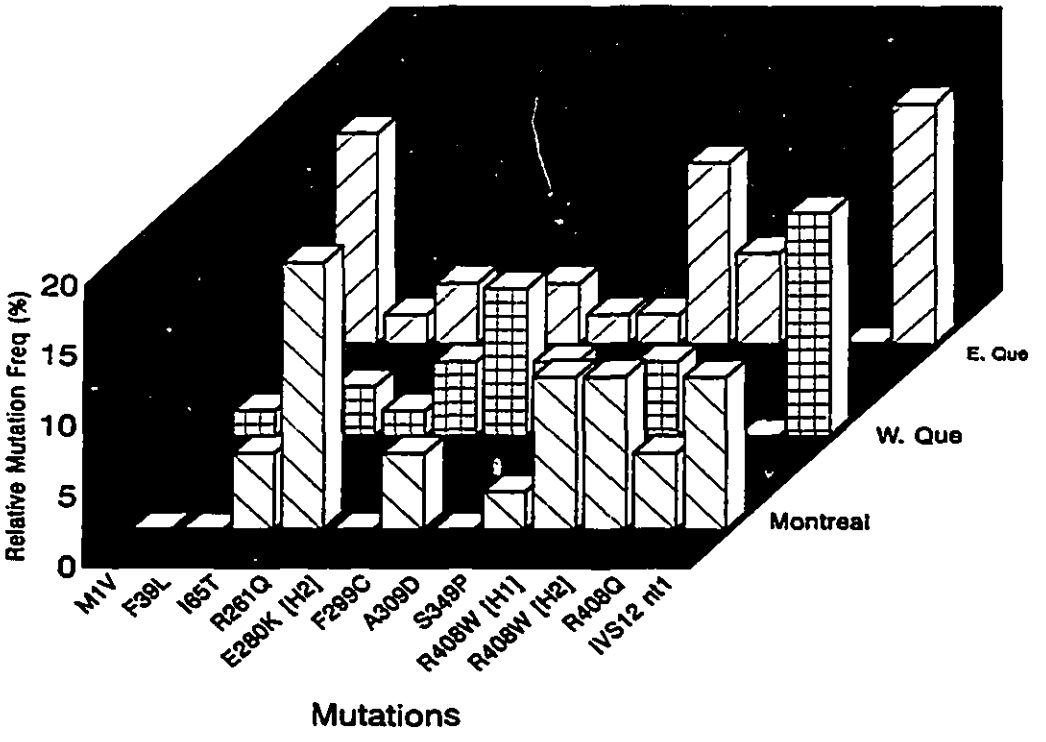
a) Quebec data includes data adapted from Rozen et al. 1994; new data provided by Carter K. (this study; unpublished material).

b) Data from (Abadie et al. 1993);(Benit et al. 1994);(Guldberg P. unpublished data) Rey F and Rey J have extended the findings in France (see Database); differences between Quebec and France still remain.

3.4.1 Stratification of Quebec PKU mutations

The distribution of both prevalent and rare alleles is non-random, both by geographic region and by population. When relative frequencies are shown by region or subpopulation the non-random distribution or stratification of mutations is apparent. This stratification becomes more obvious when the efficiency of mutation detection approaches 100%. Since mutation detection in Quebec has attained 92.5% (Table 3-4) mutations at an elevated relative frequency can be compared. In each of the three Quebec regions the four mutations with the highest relative frequencies were compared (Figure 3-8). Stratification is obvious as each region has a different set of *PAH* mutations which are at a raised frequency. The distribution of PKU mutations are different between French Canadians (East and West Quebec) and non-French Canadians (Montreal) and also between French Canadian probands from eastern and western regions of Quebec.

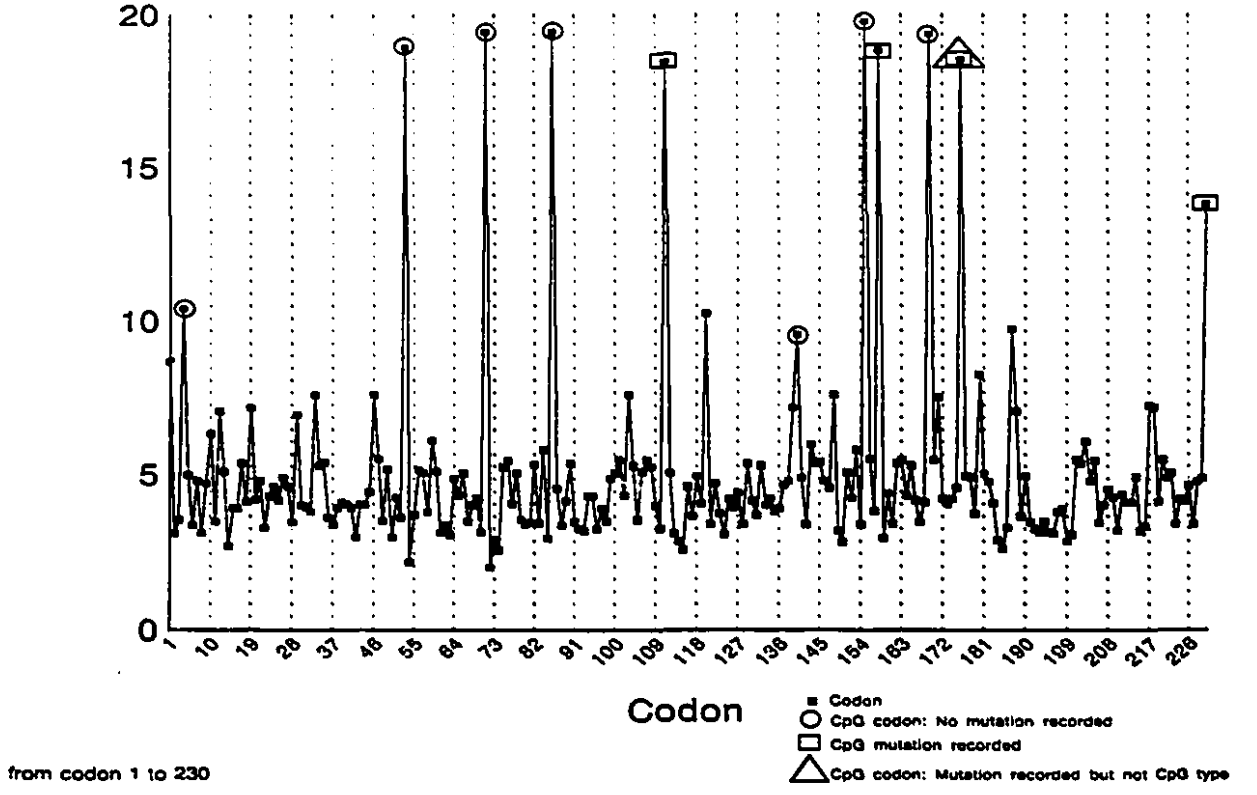
Figure 3-8. Relative frequency graph for the four most prevalent *PAH* mutations in each of three regions in Quebec.



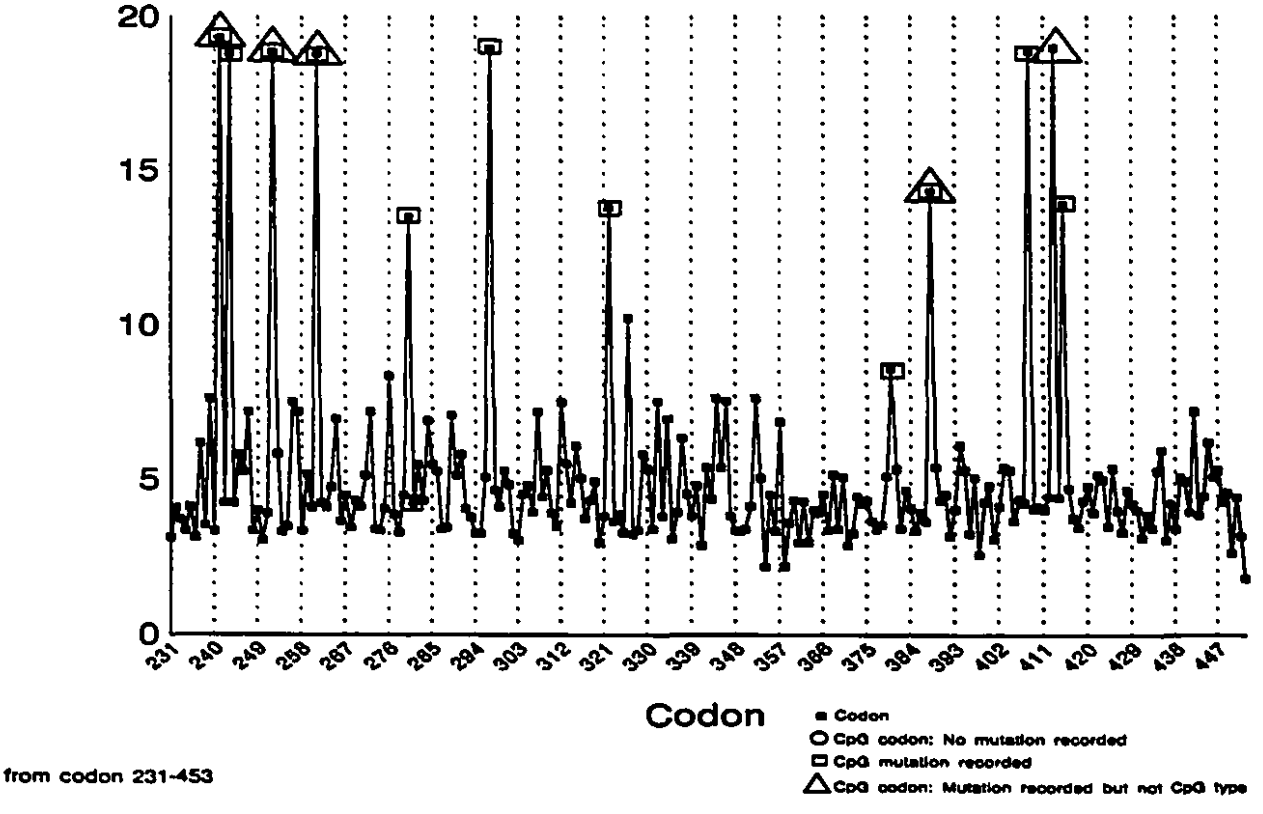
3.5.0 Mutability of the *PAH* gene

The mutatbility profile for the *PAH* gene was ascertained using the Mutpred program (Cooper and Krawczak, 1993) (Byck et al. 1996) (Figure 3-9). The profile shows 24 codons containing Cpg dinucleotides. Also indicated in the profile are the CpG sites which contain Cpg type mutations: CG>TG or CG>CA (16), CpG sites without mutations (7) and CpG codons without CpG type mutations (5). CpG sites are hypermutable (Cooper and Youssoufian 1988) and those *PAH* mutations which occur at high frequencies and at CpG dinucleotides could be recurrent.

Figure 3-9. Mutpred mutation predictability spread of *PAH* cDNA.



from codon 1 to 230



from codon 231-453

3.6.0 Mutation Detection in other populations

Mutation detection was performed on a family from the USA and a family from Kuwait. A novel mutation was discovered on one chromosome from the USA and a new mutation population association was discovered in the family from Kuwait. Below are the results of the mutation detection in each family.

3.6.1 Mutation detection on a Chinese family from the USA

All thirteen exons of the *PAH* gene were screened for mutations using DGGE. Two different heteroduplex formations were discovered in exon seven (data not shown). Upon sequencing one known (R241C) and one novel mutation was discovered (data not shown). One chromosome carried a single nucleotide change in codon 257 of exon seven. The G to A transition changes the codon from glycine (GGC) to aspartic acid (GAC), creates a Bst EII and abolishes a Hae III restriction site. The mutation was confirmed by restriction digest of amplified exon seven with Hae III (data not shown). G257D was found on haplotype four on a proband with Chinese ancestry living in the USA.

3.6.2 Mutation Detection in a Bedouin family

DNA from a Bedouin family was sent from Kuwait for mutation detection. It is the hope of researchers to find mutations which are at elevated frequency in this area and from this implement a screening program. Upon DGGE analysis it was apparent that each patient in this family was homozygous for a mutation in the *PAH* gene (data not

shown). Sequencing revealed an Ivs10 nt-11 mutation (data not shown). This mutation is the second most prevalent mutation in the *PAH* gene and is found in 25 populations on six different haplotypes (*PAH* mutation analysis database 1996). The fact that the affected children in this family were homozygous is not surprising due to the high inbreeding co-efficient of their pedigree.

Chapter 4
Discussion

4.1.0 Multiple mutations found:

Since the 1994 study (Rozen et al. 1994), 18 mutations were added to the Quebec PKU mutation spectrum. While a majority (10/18) of these mutations were found only once, in total they make up 12.2% of the Quebec PKU mutations. To date, 38 different PKU mutations in 40 different mutation/haplotype combinations have been identified in Quebec (see section 3.0), nine were first identified in Quebec and four remain unique to this population (see section 3.1). Approximately half the mutations reported to the *PAH* mutation analysis database are unique to a single population (155 mutations). Molecular and biochemical studies can determine if these mutations are disease causing. Genealogical studies combined with molecular haplotype studies can help to determine the origins of these mutations (Treacy et al. 1993). I will now discuss the two disease causing alleles unique to Quebec and identified in this thesis project.

a. The K42I mutation

Mutation scanning identified a c125 A-T transversion in codon 42 of the PAH gene (Figure 3-2 section 3.01). The substitution alters the codon for Lys^{42} (AAA) to Ile^{42} (ATA) and is confirmed by dot blot analysis (Figure 3-3 section 3.01). The systematic name of this mutation is c125 A→T; the trivial name is K42I.

The change is inherited with a variant PKU phenotype in one compound heterozygote (K42I/E280K). It was not found on any normal chromosomes (n >100) screened with DGGE, and no other changes (except for silent polymorphisms) were

found in the coding, or flanking intronic regions of the PAH gene. Together this suggests that K42I is functionally important and results in decreased PAH enzyme activity.

K42I occurs in a small stretch of conserved amino acids in human, rat and mouse PAH as well as rabbit tryptophan hyrdoxylase (TRP) (Figure 4-1) (Ledley et al. 1985, 1990; Dahl et al. 1986; Grennet et al. 1987). This region of the PAH enzyme is not generally considered to be a part of the catalytic core but instead contributes to substrate specificity and regulation (Dickson et al. 1994). For this reason K42I might not be expected to greatly decrease enzyme specific activity. However, it could cause decreased enzyme stability, as do the majority of PAH mutations this far studied by *in vitro* expression (PAH expression analysis database). *In vivo* ^{13}C-phenylalanine oxidation studies on the proband carrying this novel allele reveal oxidation levels 6% that of a normal control (Treacy et al. 1996). These oxidation levels are relatively high for a compound heterozygote in which the other allele is associated with very low activity (E280K <1% in vitro activity, in COS; *PAH* expression database). This would suggest that the K42I protein retains considerable enzymatic activity and would be associated with a mild PKU or non-PKU hyperphenylalaninemia (HPA) phenotype in the homoallelic state. We intend to measure *in vitro* enzyme activity for K42I in expression studies (E. Coli and Cos).

The K42I mutation has been identified on one chromosome in Montreal (see Table 3-4 section 3.), but has not been found in other populations. It is associated with a rare haplotype [H21] in Quebec. Its origin in the French Canadian population is as yet unknown.

b. The R157N mutation:

A new mutation was found in codon 157 of exon 5. The change is a dinucleotide substitution changing the first nucleotide (nt 470) of the codon from a G to an A and the second (nt 471) from an A to a C (Figure 3-4 section 3.01). R157N creates a *Bsa* WI restriction site which could be used for diagnostic or screening purposes. The systematic name of this mutation is c470-471 GA→AC; the trivial name is R157N.

The change was inherited with PKU in a heteroallelic genotype (R157N/F299C) and was not found on any normal chromosomes (n >100) screened with DGGE. No other changes (except for silent polymorphisms) were found in the coding, or flanking intronic regions of the mutant gene. These findings suggest that R157N is functionally important and leads to a decrease in enzyme activity.

R157N occurs in a region of high homology between human, rat and mouse PAH as well as rabbit TRH and rat tyrosine hydroxylase (TYH) (Figure 4-2) (Ledley et al. 1985, 1990; Dahl et al. 1986; Grennet et al. 1987). Amino acid 157 is suspected to be at the boundary of the catalytic core of PAH (Dickson et al. 1994) For this reason it is difficult to speculate what effect this mutation might have on enzymatic activity, substrate binding, regulation of enzyme activity or stability. Preliminary expression analysis in a human kidney cell line shows that the mutation is likely to result in ~5 % normal enzymatic activity (P. Waters and S. Hewson pers comm.). This level of enzyme activity shows that R157N is not a null mutation but is much lower PAH activity than mutations associated with non-PKU HPA (PAH expression analysis database). It is therefore likely that the R157N mutation would confer either a classical or variant PKU phenotype in a homozygous state.

The two individual nucleotide changes that make up the R157N mutation had not been previously reported to the *PAH* mutation analysis database. It is unknown if the dinucleotide change is the outcome of two separate events or is the result of a single event of unknown mechanism.

R157N was found on a single chromosome in the Western Quebec region. It is associated with an unknown haplotype and has not been found in any other populations. Its origin in the French Canadian population is as yet unknown.

Figure 4-1. Partial amino acid sequence for the human, mouse and rat phenylalanine hydroxylase (PAH) enzyme as well as the rabbit tryptophan hydroxylase (TRH) and rat tyrosine hydroxylase (TYH) enzymes. The numbers refer to the codons in the respective proteins. The boxed amino acids are conserved; the dotted box is the site of the K421 mutation. Adapted from Ledley et al. 1985, 1990; Dahl et al. 1986; Grennet et al. 1987.

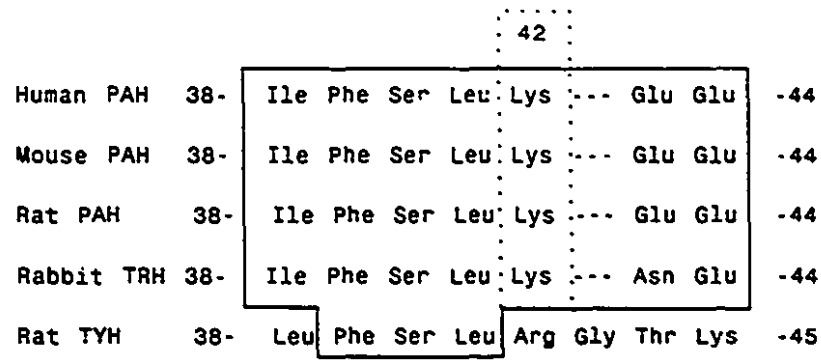
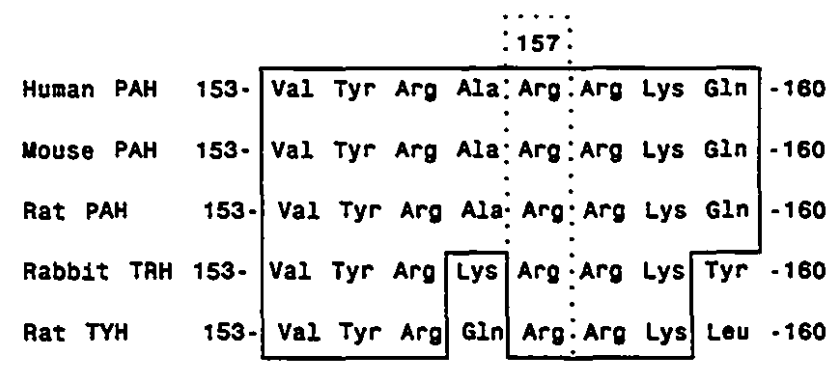


Figure 4-2. Partial amino acid sequence for the human, mouse and rat phenylalanine hydroxylase (PAH) enzyme as well as the rabbit tryptophan hydroxylase (TRH) and rat tyrosine hydroxylase (TYH) enzymes. The numbers refer to the codons in the respective proteins. The boxed amino acids are conserved; the dotted box is the site of the R157N mutation. Adapted from Ledley et al. 1985, 1990; Dahl et al. 1986; Grennet et al. 1987.



4.2.0 Silent Substitutions:

Seven silent substitutions were identified. Two of these (Q232Q, V245V) had already been described in Quebec and are frequent in French Canada (John et al. 1991). One silent substitution (IVS6 nt-55 t->g) is novel and has not been described in any other populations (see section $3.0_{(c)}$); four (IVS2 nt19, IVS3 nt-22, L385L, Y414Y) have been previously described in other populations (see section 3.4, Table 3-3).

Three of these polymorphisms were found on only a small number of chromosomes in Quebec. IVS2 nt19, IVS6 nt-55 and Y414Y were found on 2, 3 and 1 chromosomes respectively. Because of their low frequencies, these sites may not be useful for analysis of PKU chromosomes in French Canada; Frequency is not known in non-PKU families. The other four polymorphisms (Q232Q, V245V, IVS3 nt-22, L385L) are all frequent in Quebec, absolute frequencies are unknown but are listed here in rank order of approximate relative frequency. The four sites are associated with a small number of haplotypes both in Quebec and other populations which, when added to PCRable haplotyping sites can aid in haplotype determination (S. Byck in progress). In doing so, time-consuming Southern blotting could sometimes be avoided.

4.3.0 Some mutations are found on novel haplotypes:

Identical *PAH* mutations have previously been reported on two or more different haplotype [H] backgrounds. This study has identified four mutations which are present on novel haplotypes, bringing the total novel mutation/haplotype associations found in Quebec to six. The fact that one mutation occurs on two haplotypes can be explained by one of three mechanisms: i) two independent mutational events recurring at the same nucleotide, ii) gene conversion event or iii) a homologous recombination between the

two haplotypes during meiosis (Okano et al. 1990). I will discuss each of these mechanisms with regards to the novel mutation/haplotype associations found in Quebec.

A) The mutations:

The S67P mutation is associated with haplotype 1 in the Montreal region of Quebec (Table 3-1, section 3.02). Prior to this study, it had been reported to be in complete linkage disequilibrium with haploytpe 4 in Wales, Italy and South West England (see Table 3-1 section 3.02).

The G218V mutation was found to associated with haplotype 2 in the Western region of Quebec. It was previously reported on haplotype 1 in Denmark, France and Belgium (Table 3-1, section 3.02).

The V245A mutation is found on haploytpe 7 in Western Quebec. It is found on haplotype 3 in Denmark, Northern Ireland and Spain (Table 3-1, section 3.02)

B) Recurrent Mutation:

" Hot spots " for point mutations (eg. CpG dinucleotides) occur throughout many genes including *PAH* (Abadie et al. 1989) (Cooper et al. 1988). Two *PAH* mutations, both at CpG sites, are best explained by recurrent mutation (R408W: Byck et al. 1994); (E280K: Byck et al. 1996). *PAH* mutations not located at CpG sites have not been shown to be recurrent in any population. None of the mutations described here found on a novel mutation/haplotype association, occurs at a CpG site. Therefore, they are unlikely to be located at hypermutable sites. Furthermore, viewed on MUTPRED (Figure 3-9, section 3.3) the three codons which are located in the cDNA region have relatively low mutability profiles (F39L 4.09/20; G218V 7.18/20; V245A 5.8/20). Further studies using the highly polymorphic VNTR and STR systems and extended diallelic haplotypes could be used to test these mutations for uniqueness. However, taking the above information into account, it is unlikely that the presence of these four mutations on novel haplotypes is compatible with the recurrent mutation mechanism.

C) Gene conversion:

Gene conversion is a non-reciprocal homologous recombination whereby modifications in an acceptor gene are copied from a homologous donor sequence. While most human gene conversions are small (< 300 bp), they can be as long as 1.5 Kb (Smithies et al. 1986). In eukaryotes there has not been a consensus sequence found that in necessary for the initiation of gene conversions, however, in prokaryotes it has been suggested that a promoter specific 14 bp poly (dA.dT) tract is an important component of gene conversion (Schultes et al. 1991). If a promoter sequence was required in humans this would cause a very large gene conversion area for all three of the novel

mutation/haplotype associations in question, ranging from 20 Kb for S67P to 70 Kb for V245A. These large sizes of gene conversions make this mechanism unlikely. If however, the consensus sequence in humans could be found in intronic regions, each of these mutations could occur on novel haplotypes by relatively small gene conversion events. Present knowledge of gene conversion in humans cannot rule out the gene conversion mechanism to explain these novel mutation/haplotype associations.

D) Recombination:

The most likely mechanism for the explanation of novel mutation/haplotype associations is homologous recombination, which involves i) the pairing of two homologous DNA duplexes, ii) breaking two of the homologous DNA strands, iii) re-forming of phosphodiester bonds to join the two homologous strands and iv) breaking of the other two strands and joining them (Holliday 1964; Meselson and Radding 1975; Radding 1982; Szostak 1983). For three of the novel mutation/haplotype associations (S67P, G218V, V245A) a single crossover could transfer the mutant nucleotide from the common to the rare haplotype (Figure 4-3).

A mutation should be associated with two haplotypes in the source population when a recombination has occurred, assuming no admixture or gene flow (Okano et al. 1990). The S67P mutation has been found on only one haplotype 4 chromosome in each of Italy, Wales and South West England (see Table 3-1, section 3.1), with 110, 111 and 111 chromosomes studied in each population respectively. Since each population has been studied thoroughly and only one mutant chromosome found in each it is unlikely that the haplotype 4 to haplotype 1 recombination occurred in these populations.

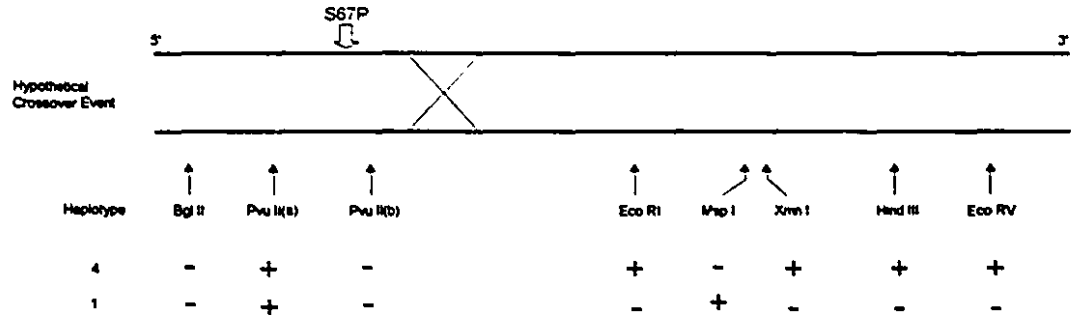
The G218V mutation has been identified on one [H1] chromosome in Denmark and Belgium (see Table 3-1, section 3.1), with 333 and 250 chromosomes studied in each population respectively. It is unlikely that a haplotype 1 to haplotype 2 recombination occurred in these populations. However, the relative frequency of G218V [H1] is not known in France and it is still possible that G218V [H2] originated there via a recombinational event.

The V245A mutation has bee found on five haplotype 3 chromosomes world wide. Each chromosome is in a population of sufficient size with low enough V245A frequencies to make the haplotype 3 to haplotype 7 recombination event unlikely in these populations (see Table 3-1, section 3.1). V245A has also been found on four chromosomes with unknown haplotypes in Germany (3) and Belgium (1). It is possible that this recombination event occurred in either of these populations and in turn migrated into Quebec.

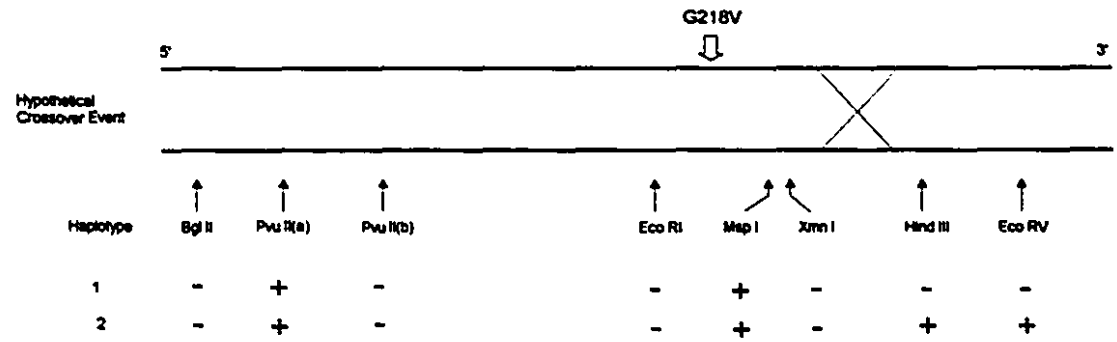
Figure 4-3. A single recombination could transfer a rare mutation from one polymorphic haplotype to another.

The filled and open boxes represent mutant and normal alleles respectively. #1 recombination for the S67P allele; #2 recombination for the G218V allele; #3 recombination for the V245A allele.

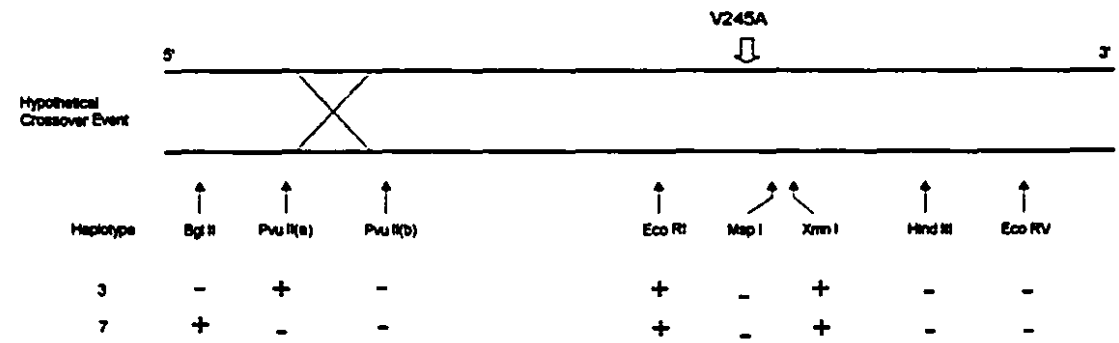
#1. A single crossover 5' to the Eco RI site could transfer the S67P mutation from haplotype 4 to haplotype 1.



#2. A single crossover 5' to the Hind III site could transfer the G218V mutation from haplotype 1 to haplotype 2.



#3. A single crossover 5' to the Pvu II(b) site could transfer the V245A mutation from haplotype 3 to haplotype 7.



4.4.0 Chromosomes with unknown mutations:

From the 141 chromosomes studied in Quebec there were seven chromosomes that no heteroduplex formation was detected by DGGE. There are two explanations for these mutations remaining unknown. i) it is possible that DGGE could not detect the nucleotide change because the change in melting temperature is not great enough. Due to the fact that DGGE is based on changes in melting temperature, it is possible that a nucleotide change could cause only a slight change in melting temperature causing the technique to fail. ii) The mutation could be located in a region of the gene that was not scanned for mutations. While the *PAH* PCR primers are generally located approximately 60 nucleotides into intronic regions of the gene, there are primers which are located only 10 nucleotides from the exon/intron boundry. For this reason it is possible that undetected mutations are located outside regions that were scanned, either deeply intronic (>60 nucleotides) or in introns not adequately scanned (>10 nucleotides). There have not been any disease causing mutations found in the promoter portion of the *PAH* gene but this is probably due to the under-sampling of this portion of the gene.

4.5.0 Mutation Diversity and Stratification

While mutation/haplotype associations which are unique to Quebec can be utilized to determine the origins of the mutation, the entire subset of Quebec alleles (both rare disease producing and polymorphic (neutral)) can be informative about their European counterparts. When classified by mutation type Quebec alleles resemble those for the world population (fig 3-8 section 3.2), on the other hand the spectrum of specific Quebec mutations differ from any single European population. For example the S67P mutation found in Western and Eastern Quebec is also found in Wales, Italy, South West

England and Northern Ireland. However, the R176X mutation, found in Western Quebec, has only been found in Sicily, Egypt and Belgium (see Table 3-2 section 3.3). The set of PKU alleles not only differs from that of France, the country of origin for French Canadian ancestors (Harris and Mathews 1989) (see Table 3-4 section 3.) but also from any European set of alleles.

In further subdividing the province of Quebec into eastern, western and Montreal regions, the demographic history of the province can be exposed using the frequencies of rare *PAH* alleles (Figure 3-6). While Quebec alleles are unique from Europe so is each region unique within the province. This 'genetic geography' or stratification of alleles becomes most apparent when studying the prevalent mutations in each of the three regions. When the four most frequent mutations are chosen from each region, there is only slight overlap with IVS12 nt1 and R408W [H1] being found in more than one region (Figure 3-6 section 3). Non-French Canadian alleles (Montreal) differ from those of French Canadian familes, and the French Canadian alleles from the Eastern and Western regions of the province also differ. This different pattern of mutations reveals a different pattern of founders and settlement by French Canadians in the eastern and western regions of the province as well as non-French Canadians in Montreal.

In summary, Quebec is a unique population whose structure allows the genetic history of the province to be easily documented. An increased level of mutation detection (93%) has shown that many Quebec alleles reflect evidence of founder effects, genetic drift, and range expansion in relative genetic isolation. In short, the history of the gene is the history of the population.

4.6.0 Applications of Mutation Detection and Genotyping:

This project has contributed directly and collaboratively to many other ongoing projects in this laboratory. i) From genotyping of Quebec PKU patients, a genotype/phenotype correlation study has been established. ii) By providing rare disease producing alleles, it has allowed for an *in vitro* expression system that will analyze mutant proteins which harbour unique Quebec mutations. iii) The genotyping of multiple Quebec families, has also assisted the study of a non-invasive method of measuring phenylalanine oxidation levels *in vivo*. iv) Lastly, I was involved in finding the PKU mutations in a family that requested pre-natal testing but had uninformative haplotypes. My role in each of these studies has been to provide and confirm individual mutations and family genotypes. Each of these projects are described below.

4..6.1 Genotype - Phenotype correlation study:

In analyzing HPA cases both in Quebec an around the world, Emre Kayaalp has produced one of the most extensive genotype-phenotype studies at the *PAH* locus to date. This study covers 392 individual cases, harbouring 177 different genotypes. and 78 different alleles (Kayaalp et al. in progress; abstract in press). The effects of 38 different alleles, in patient's either homoallelic or heteroallelic against a null allele, were evaluated with regards to clinical and metabolic phenotypes. Twenty seven cases could be easily classified in one of three phenotype categories; Classical PKU, variant PKU and non-PKU HPA. The remaining eleven mutations fell into two or even three phenotype categories when studied in different patients. Two such examples are I65T and Y414C. I65T (relative frequency in Quebec 4.3%) has an *in vitro* activity of 26 % that of normal

protein, but when classified in distinct patients was placed in all three phenotypic categories. Likewise, Y414C has an *in vitro* activity 50 % that of normal and also falls into all three phenotypic categories when found in different patients. Conclusions from this study are that HPA has features of a complex trait. While the majority of PAH mutations have predictive value for treatment and prognosis, identical PAH genotypes may act differently between patients. Furthermore, the *in vivo* properties of the PAH enzyme are not always concordant with *in vitro* findings.

4.6.2 *In vitro* expression of novel Quebec mutations:

Ten mutations were first identified in Quebec and four are thus far unique to the Quebec population (see section 3.1). Dr. Paula Waters is using two expression systems to characterize mutant PAH proteins which harbour Quebec alleles. The first expression system is in a human embryonic kidney cell line. This system is hoped to produce mutant proteins that have a similar unit-protein activity as would be found in patients harbouring this mutation. The second expression system is in *E. coli* as fusion protein of PAH with maltose binding protein. This prokaryotic system produces more pure PAH protein than it's eukaryotic counterpart and will thus allow more detailed biochemical studies of mutant proteins.

Mutations that are in the process of being expressed are K42I, D338Y, A104D and R157N. Initial studies have been completed on the A104D mutant. This mutant has been found to have 33 % of normal activity in the eukaryotic expression system, 52-59 % specific activity in the prokaryotic system compared to wild type and 8.4% normal activity in the *in vivo* system (Waters et al. abstract in press). From this data it was

concluded that the *in vitro* protein expression data is compatible with the *in vivo* oxidation data and metabolic phenotype of the patient homozygous for this mutation. The R157N mutation is hoped to be a useful mutation for detailed biochemical analysis of the PAH protein due to it's proximity to the hypothesized PAH catalytic core (see section $4.0_{(b)}$) and it's position beside the R158Q mutation. R158Q has been expressed in a Cos system and found to have considerable residual activity (PAH expression analysis database).

4.6.3 *In vivo* phenylalanine oxidation study:

When identical phenotypes are not observed with identical genotypes and one or more of the phenotypes does not correlate with the *in vitro* expression levels it is interesting to study the *in vivo* phenylalanine oxidation levels of individual patients. To measure the phenylalanine oxidation levels *in vivo*, ^{13}C labeled phenylalanine is used as a substrate and the resulting $^{13}CO_2$ is analyzed by isotope ratio mass spectrometry (Treacy et al. 1996; in progress). This is a non-invasive method of assessing genotype-phenotype correlations.

When percent phenylalanine oxidized is plotted against time the test proves to be very informative, with not only affected individuals being distinguished from normals but all three class of HPA (classical, variant, non-PKU HPA) being distinguishable. The differences between all cases of hyperphe with controls and carriers as well as the difference between controls and carriers are extremely significant ($p<.0001$). The difference between the PKU, variant and non-PKU HPA groups are also significant ($p<.05$) (Treacy et al. 1996).

As with the direct genotype-phenotype study, there are discrepancies between *in vivo* and *in vitro* data as expressed in a COS cell system. For example, a patient heterozygous for F299C/R158Q had zero oxidation of ^{13}C-phenylalanine as no $^{13}CO_2$ was detected. *In vitro*, the F299C allele has less than 3% normal activity but the R158Q allele has 10% activity. From these values one would expect to observe some phenylalanine oxidation *in vivo* (Treacy et al. 1996). While in vitro analysis measures only the unit-protein effect, the ^{13}C phenylalanine oxidation rate takes into account multiple loci and possible protein interactions. In doing so, hyperphenylalaninemia is seen as a complex trait with multiple loci and events determining the emergent phenotype.

4.6.4 Mutation scanning for pre-natal testing:

Most often, haplotype analysis and tests for specific alleles are used for informative pre-natal testing. On occasion, haplotypes are uninformative and mutations are unknown. In these cases mutations need to be determined in a reliable and time efficient manner. When such a case was referred to Dr. Rima Rozen I used the DGGE technique of mutation scanning to determine the affected exons (see section 3.6.2) and sequencing in the pre-natal testing lab determined the mutations. While DGGE has never attained 100% mutation detection efficiency in this or other populations and will never replace current screening techniques, in such rare cases as this, it has proven to be a useful diagnostic tool.

Bibliography

Abadie V, Lyonnet S, Maurin N, Berthelon M, Caillaud C, Giraud F, Mattei JF, Rey J, Rey F and Munnich A. 1989. CpG dinucleotides are mutation hot spots in phenylketonuria. Genomics 5:936

Antonarakis S.E. and McKusick V.A. 1994. Discussion on mutation nomenclature. Hum Mut 4:166

Avigad S, Cohen B.E., Bauer S, Schwartz G, Frydman M, et al. 1990. A single origin of phenylketonuria in Yemenite Jews. Nature 334:168-70

Beaudet A.L. and Tsui L-C. 1993. A suggested nomenclature for designating mutations. Hum Mut 2:245-48

Bickel H, Gerard J. and Hickmans E.M. 1953. Influence of phenylalanine intake on phenylketonuria. Lancet 265:812-813

Byck S, Morgan K, Tyfield L, Dworniczak B and Scriver C.R. 1994. Evidence for origin, by recurrent mutation, of the phenylalanine hydroxylase R408W mutation on two haplotypes in European and Quebec populations. Hum Molec Genet 3:1675-1677

Byck S, Tyfield L, Carter K and Scriver C.R. 1996. Prediction of multiple hypermutable codons in the human PAH gene: codon 280 contains recurrent mutations in Quebec and other populations. Hum Mut (in press)

Carter K.C., Rozen R, Byck S and Scriver C.R. 1995. Novel mutations and heterogeneity of the phenylalanine hydroxylase (PAH) gene on Quebec PKU chromosomes. Am J Hum Genet 57:A191-(915) (Abstract)

Cooper D.N. and Youssoufian H. 1988. The CpG dinucleotide and human genetic disease. Hum Genet 78:151-155

Cooper D.N. and Krawczak M. 1993. Human Gene Mutation. Oxford, U.K. Bios Scientific Publishers.

Cotton R.G.H. 1993. Current methods of mutation detction. Mut Res 285:125-44

Dahl H-H.M., and Mercer J.F.B. 1986. Isolation and sequence of a cDNA clone which contains the complete coding region of rat phenylalanine hydroxylase. Structural homology with typrosine hydroxlase, glucocorticoid regulation and use of alternate polyadenylation sites. J Biol Chem 261:4148

Danks D.M., Schleisinger P, Firgario F, Cotton R.G.H., Watson B.M., Rembold H and Hennings G. 1979. Malignant Hyperphenylalaninemia Clinical Features, Biochemical Findings, and Experience with Administration of Biopterins. Pediat Res 13:1150

Dianzani I, Giannattasio G, De Sanctis L, Alliaudi C, Lattanzio P, Dionisi Vici C, Burlina A, Burroni M, Sebastio G, Carnevale F, Guzzetta V, Marra E, Camaschella C and Ponzone A. 1995. Characterization of phenylketonuria alleles in the Italian population. Eur J Hum Genet 3:294-302

Dickson P.W., Jennings I.G. and Cotton R.G. 1994. Delineation of the catalytic core of phenylalanine hydroxylase and identification of glutamate 286 as a critical residue for pterin function. J Bio Chem 269:20369-75

DiLella A.G., Marrit J, Brayton K. and Woo S.L.C. 1987. An amino acid substitution involved in phenylketonuria is in linkage disequilibrium with DNA haplotype 2. Nautre 327:333-6

Dwornickzak B, Kalaydjieva L., Panoke S, Aulehla-Scholz C, Allen G. and Horst J. 1992. Analysis of exon 7 of the human phenylalanine hydroxylase gene: a mutation hot spot? Hum Mut 1:138-146

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 J.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 N, Jin L, Chakraborty R. and Woo S.L.C. 1992a. Multiple origins for phenylketonuria in Europe. Am J Hum Genet 51:1355

Eisensmith R.C. and Woo S.L.C. 1992. Molecular basis of phenylketonuria and related hyperphenylalaninemias: Mutations and polymorphisms in the human phenylalanine hydroxylase gene. Hum Mut 1:13-23

Feingold J, Guilloud-Bataille M, Feingold N, Rey F, Berthelon M. and Lyonnet S. 1993. Linkage disequilibrium in the human phenylalanine hydroxylase locus. Dev Brain Dysfunct 6:26-31

Flint J, Harding R.M., Clegg J.B. and Boyce A.J. 1993. Why are some genetic diseases common ? Hum Genet 91:91-117

Fölling A. 1934. Uber Susscheidung von Phenylbrenztraubensaure in den Harn als Stoffwechselanomalie in Ver-bindung mit Imbezillitat. Z. Physiol. Chem. 277:169

Forrest S.M., Dahl H.H., Howells D.W., Dianzani I, and Cotton R.G.H. 1991 Mutation detection in phenylketonuria by using chemical cleavage of mismatch: Importance of using probes from both normal and patient samples. Am J Hum Genet 49:175-183

Goltsov A.A., Eisensmith R.C., Konecki D.S., Lichter-Konecki U. and Woo S.L.C. 1992 Associatioins between mutations and a VNTR in the human phenylalanine hydroxylase gene. Am J Hum Genet 51:627-636

Goltsov A.A., Eisensmith R.C., Naughton E.R., Jin L, Chakraborty R. and 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 Molec Genet 2:577-581

Goltsov A, Kouzmine A, Eisensmith R.C., Effat L, Tempamy S, Rushdi S, Abdel-Meguid N and Woo S.L.C. 1994. Molecular basis of phenylketonuria in Egypt. Am J Hum Genet 55:#1288 (Abstract)

Grenett H.E., Ledley F.D., Reed L.L. and Woo S.L.C. 1987. Full length cDNA for rabbit tryptophan hydroxylase. Functional domains and evolution of the aromatic amino acid hydroxylases. Proc Natl Acad Sci USA 84:5530

Grenier A, Morrisette J. and Dussault J.H. 1980. Les maladies métaboliques héréditaires au Québec: le dépistage sanguin. L'Union Medicale de Canada 109:591-592

Guldberg P, Valentino R, Ceratto N, Bosco P, Ciuna M, Indelicato A, Mollica F, Meli C, Giovannini M, Riva E, Biasucci G, Friis Henriksen K and Guttler F 1993a. Mutational spectrum of phenylalanine hydroxylase deficiency in Sicily: implications for diagnosis of hyperphenylalaninemia in Southern Europe. Hum Mol Genet 2: 1703-1707

Guldberg P, Henriksen K.F. and Guttler F. 1993b. Molecular analysisi of phenylketonuria in Denmark: 99% of the mutations detected by denaturing gradient gel electrophoresis. Genomics 17:141-46

Guldberg P, and Guttler F. 1994a. Mutations in the phenylalanine hydroxylase gene: Method for their characterization. Acta Pediatr Suppl 407:27-33

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

Guldberg P, Friis Henriksen K and Guttler F 1994c. Constant Denaturant Gel Electrophoresis without formamide. Biotechniques 16:241

Guldberg P, Henriksen F.F., Thony B, Blau N, and Guttler F. 1994d. Molecular heterogeneity of nonphenylketonuria hyperphenylalaninemia in 25 Danish patients. Genomics 21:453-455

Guldberg P, Levy H.L., Koch R, Berlin C.M.J., Francois B, Henriksen K.F. and Guttler F. 1995. Mutation analysis in families with discordant phenotypes of phenylalanine hydroxylase deficiency. Inheritance and expression of the hyperphenylalaninaemias. J Inher Metab Dis 17:645-651

Harris R.C. and Matthews J.J. 1989. Historical Atlas of Canada. I. From the beginning to 1800, University of Toronto Press, Toronto

Hoang L, Byck S, Prevost L. and Scriver C.R. 1996. PAH mutation Analysis Consortium Database: a database for disease-producing and other allelic variation at the human PAH locus Nucl Acid Res 24:127-131

Holliday R, 1964. A mechanism for gene conversion in fungi. Genet Res 5:282

Holtzman C, Slazyk W.E., Corduo J.F. and Hannon W.H. 1986. Descriptive epidemiology of missed cases of phenylketonuria and congenital hypothyrodism. Pediatrics 78:553-558

Horst J, Eigel A, Kalaydjieva L and Dworniczak B. 1993. Phenylketonuria in Germany - Molecular heterogeneity and diagnostic implications. Dev Brain Dysfunct 6:32-38

Huang S.Z., Zhou X.D., Ren Z.R., Zeng Y.T. and Woo S.L.C. 1990. Prenatal detection of an arg- >ter mutation at codon 111 of the PAH gene using DNA amplification. Prenatal Diagnosis 10:289-293

Jaruzelska J, Abadie V, D'Aubenton-Carafa Y, Brody E, Munnich A and Marie J. 1995. In Vitro splicing deficiency induced by a C to T mutation at position -3 in the intron 10 acceptor site of the phenylalanine hydroxylase gene in a patient with phenylketonuria. J Biol Chem 270:20370-20375

John S.W.M., Rozen R, LaFramboise R, LaBerge C. and Scriver C.R. 1989. Novel PKU mutation on haplotype 2 in French-Canadians. Am J Med Genet 45:905

John S.W.M., Rozen R, Scriver C.R., Laframboise R. and Laberge C. 1990. Recurrent mutation, gene conversion, or recombination at the human phenylalanine hydroxylase locus: Evidence in French-Canadians and a catalog of mutations. Am J Hum Genet 46:970-974

John S.W., Weitzner G, Rozen R. and Scriver C.R. 1991. A rapid procedure for extracting genomic DNA from leukocytes. Nucl Acid Res 19:408

John S. W. M, Rozen R, Laframboise R, Laberge C, Scriver CR 1992 Five mutations at the PAH locus account for almost 90% of PKU mutations in French-Canadians from Eastern Quebec. Hum Mut 1:72-74

Kalaydjieba L, Dwornizcak B, Kremensky I, Radeva B and Horst J. 1993. Population genetics of phenylketonuria in Bulgaria. Dev Brain Dysfunct 6:39-45

Kaufman S. 1993. New tetrahydrobiopterin-dependent systems. Annu. Rev. Nutr. 13:261-86

Kaul R, Matalon R, Allen R, Fisch R.O., Michals K, Petrosky A and Sullivan D. 1994. Frequency of 12 mutations in 114 children with phenylketonuria in the midwest region in the USA. J Inher Metab Dis 17:356-358

Konecki D.S. and Lichter-Konecki. 1991a. The phenylketonuria locus: current knowledge about alleles and mutations of the phenylalanine hydroxylase gene in various populations. Hum. Genet. 87:377-388

Konecki D.S., Schlotter M, Trefz F.K. and Lichter-Konecki U. 1991. The identification of two missense mutations at the PAH gene locus in a Turkish patient with phenylketonuria. Hum Genet 87:389-393

Konecki D.S., Wang Y, Trefz F.K., Lichter-Konecki U. and Woo S.L.C. 1992. Structural characterization of the 5' regions of the human phenylalanine hydroxylase gene. Biochem 31:8363-8

Kuzmin A, Goltsov A, Eisensmith R.C., Baranovskaya S, Schwartz E. and Woo S.L.C. 1995. Molecular basis of phenylketonuria in seven populaitons from the Former Soviet Union. Am J Hum Genet 57:A166-(945) (Abstract)

Lambert D. 1994. The genetic epidemiology of hyperphenylalaninemia in Quebec. Masters thesis McGill University

Leandro P, Rivera I, Ribeiro V, Tavares De Almeida I. and Lechner C. 1993. Sequencing analysis of PAH genomic DNA reveals 4 novel mutations affecting exons 7 and 11 in Portuguese PKU population. Society for the Study of Inborn Errors of Metabolism. Sept. W02(Abstract)

Ledley F.D., Grennet H.E., Gunbar B.S. and Woo S.L.C. 1990. Mouse phenylalanine hydroxylase. Homology and divergence from human phenylalanine hydroxylase. Biochem J 267:399

Ledley F.D., DiLella A.G., Kwok S.C.M. and Woo S.L.C. 1985. Homology between phenylalanine and tyrosine hydroxylases reveals common structural and functional domains. Biochemistry 24:3389

Lichter-Konecki U, Schlotter M, Trefz F.K. and Konecki D.S. 1990. Identification of the new mutations at the phenylalanine hydroxylase gene locus. Vth International Congress Inborn Errors of Metabolism. W4.4(Abstract)

Lichter-Konecki U, Schlotter M. and 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 94:307-310

Lidsky A.S., Robson K, Chandra T, Barker P, Ruddle F. and Woo S.L.C. 1984. The PKU locus in man is on chromosome 12. Am J Hum Genet 78:347

Lidsky A.S., Law M.L., Morse H.G., Kao F.T. and Woo S.L.C. 1985. Regional mapping of the phenylalanine hydroxylase gene and the phenylketonuria locus in the human genome. Proc Natl Acad Sci USA 82:6221

Lyonnet S, Melle D, de Braekeleer M, Laframboise R, Rey F, John S.W.M., Berthelon M, Berthelot J, Journel H, Le Marec B, Parent P, de Parscau L, Saudubray J.M., Rozen R, Rey J, Munnich A. and Scriver C.R. 1992. Time and space clusters of the French-Canadian M1V phenylketonuria mutation in France. Am J Hum Genet 51:191

Maniatas T, Fritsch EF and Sambrook J, 1982. Molecular Cloning; A Laboratory Manual (1st Edition) Cold Spring Harbour, NY.

Martinez-Pardo M, Colmenares A.R., Garcia M.J., Perez B, Desviat L.R. and Ugarte M. 1994. Phenotype distribution in the Spanish phenylketonuria population and related genotypes. J. Inher Metab Dis 17:366-368

Marrit J., DiLella A.G., Brayton K, Ledly F.D., Robson K.J. and Woo S.L.C. 1987. GT to AT transition at a splice donor site causes skipping of the preceding exon in phenlketonuria. Nucl Acid Res 15:5613-28

McEvedy C. and Jones R. 1978. Atlas of World Population History. Penguin Reference Books

Meselson M, Radding C.M. 1975. A general model for genetic recombination. Proc Natl Acad Sci USA 72:358

Moncrieff A. 1968. For the Medical Research Council Working Party on Phenylketonuria. Present status of different mass screening procedures for phenylketonuria. British Medical Journal 4:7-13

Myers R.M., Maniatis T. and Lerman L.S. 1987. Detection and localization of single base changes by denaturing gradient gel electrophoresis. Methods Enzymol 155: 501-527.

Neckameyer W.S. and White K. 1992. A single locus encodes both phenylalanine hydroxylase and tryptophan hydroxylase activities in Drosophila. J Biol Chem 267:4199

Okano Y, Wang T, Eisensmith R.C., Guttler F and Woo S.LC. 1990. Recurrent mutation in the human phenylalanine hydroxylase. Am J Hum Gen. 16:919

Penrose L.S. 1935. Inheritance of phenylpyruvic amentia (Phenylketonuria) Lancet 2:192-94

Penrose L.S. 1946. Phenylketonuria. A problem in eugenics. Lancet 1:949-53

Perez B, Desviat L.R. and Ugarte M. 1995. Expression analysis of mutation P244L, which causes mild hyperphenylalaninemia. Hum Mut 5:188-190

Popescu A, Andrian T, Guttler F and Guldberg P. 1994. Genotype-phenotype correlation in 11 Romanian PKU families. J Inher Metab Dis 17:374-375

Radding C.M., Flary J, Wu A, Kahn R, DaGupta C. 1982. Three phases in homologous pairing: polymerisation of recA in single-stranded DNA, synapsis and polar strand exchange. Cold Spring Harbor Symp Quant Biol 47:821

Ramus S.J., Treacy E.P., and Cotton R.G.H. 1995. Characterization of phenylalanine hydroxylase alleles in untreated phenylketonuria patients from Victoria, Australia: Origin of alleles and haplotypes. Am J Hum Genet 56:1034-1041

Romano C, Cali F, Guldberg P, Guttler F, Indelicato A and Ceratto N. 1994. Association between haplotypes, Hind III-VNTR alleles and mutations at the PAH locus in Sicily. Acta Pediatr Japan 407:39-40

Rozen R, Mascisch A, Lambert M, Laframboise R, Scriver R 1994. Mutation profiles of phenylketonuria (PKU) in Quebec populations: Evidence of stratification and novel mutations. Am J Hum Genet 55:321-326

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N, 1985. Enzymatic amplification of b-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230:1350

Saiki RK, Bugawan TL, Horn T, Mullis KB, Erlich HA, 1986. Analysis of enzymatically amplified b-globin and HLA-dQa genomic DNA with allele specific oligonucleotide probes. Nature 324:163

Sanger F, Nicklen S, and Coulson A.R. 1977. Proc. Natl. Acad. Sci. USA 74:5463

Schultes N.P. and Szostak J.W. 1990. A poly (dA.dT) tract is a component of the recombination initiation site at the ARG4 locus in Saccharomyces cerevisiae. Molecular and Cellualr Biology 11:322

Scriver C.R. and Rosenberg L.E. 1973. Amino Acid Metabolism and Its Disorders. Saunders, Philadelphia. p290.

Scriver C.R. and Clow C.L. 1980a. Eptiome of human biochemical genetics. Part I. New Engl. J. Med. 303:1336-42

Scriver C.R. and Clow C.L. 1980b. Phenylketonuria: Eptome of human biochemical genetics. Part II. New Engl. J. Med. 303:1394-400

Scriver C.R., Kaufman S. and Woo S.L.C. 1989. The hyperphenylalaninemias. In: Scriver C.R., Beaudet A, Sly W, Valle D, eds. The Metabolic Basis of Inherited Disease. (6th ed.) New York: McGraw Hill: 495

Scriver C.R. 1993. Every disease has two histories. Mémories de la Société Royale du Canada. Serie VI: Tome: IV 19-32

Scriver C.R., Eisensmith R.C., Woo S.L.C. and Kaufman S. 1994. The hyperphenylalaninemias of man and mouse. Annu. Rev. Genet 28:141-65

Scriver, C. R., S. Kaufman, E. Eisensmith, and S.L.C. Woo. 1995. The Hyperphenylalaninemias. In The Metabolic and Molecular Bases of Inherited Disease. 7th edition. C.R. Scriver, A.L. Beaudet, W.S. Sly, and D. Valle, editors. McGraw Hill Book Co.

Scriver C.R., Hoang L, Byck S. and Prevost L. 1996a. PAH Gene Mutation Analysis Consortium Newsletter.

Scriver C.R., Byck S, Prevost L, Hoang L. and the PAH Muation Analysis Consortium. 1996b. The phenylalanine hydroxylase locus: a marker for the history of phenylketonuria and human genetic diversity. Ciba Foundation Symposium 197; Variation in the Human Genome. pg 73-96 John Wiley and Sons Ltd.

Smith I. and Wolff O.H. 1974. Natural history of phenylketonuria and influence of eary treatment. Lancet 2:540

Smithies O and Powers P.A. 1986. Gene conversions and their relation to homologous chromosome pairing. Phi Trans R Soc Lond B312:291

Spieto F.W., Hearn T.L., Gardner F.H. and Hannon W.H. 1985. Phenylalanine analyses of blood-spot control materials: Preparation of samples and evaluation of interlaboratory performance. Clin Chem 31:235

Stoll J and Goldman D. 1991. Isolation and structural characterization of the murine tryptophan hydroxylase gene. J Neurosci Res 28:457

Szostak J.W., Orr-Weaver T.L., Rothstein R.J. and Stahl F.W. 1983. The double strand break repair model for recombination. Cell 33:25

Takarada Y, Kalanin J, Yamashita K, Ohtsuka N, Kagawa S and Matsuoka A. 1993. Phenylketonuria mutant alleles in different populations -Missense mutation in exon-7 of phenylalanine hydroxylase gene. Clin Chem 39:2354-2355

Treacy E, Byck S, Clow C, Scriver CR 1993. 'Celtic' phenylketonuria chromosomes found? Evidence in two regions of Quebec province. Eur J Hum Genet 1:220-228

Tyfield L.A., Osborn M.J., King S.K., Jones M.M. and Holton J.B. 1993. Molecular basis of phenylketonuria in an English population. Dev Brain Dysfunct 6:60-67

Wang Y, DeMayo J.L., Hahn T.M., Finegold M.J., Konecki D.S. and Lichter-Konecki U. 1992. Tissue- and development-specific expression of the human phenylalanine hydroxylase/chloramphenicol acetyltransferase fusion gene in transgenic mice. J Biol Chem 267:15105-15110

Woolf L.I., McGean M.S., Woolf F.M. and Calahane S.F. 1975. Phenylketonuria as a balanced polymorphism: The nature of the heterozygote advantage. Ann Hum Genet 38:461

Woolf L.I. 1978. The high frequency of phenylketonuria in Ireland and western Scotland. J Inher Metab Dis 1:101

Woolf L.I. 1986. The heterozygote advantage of phenylketonuria. Am J Hum Genet 38:773-774

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

Zschocke J, Graham C.A., Stewart F.J., Carson D.J. and Nevin N.C. 1994. Non-phenylketonuria hyperphenlalaninaemia in Northern Ireland: Frequent mutation allows screening and early diagnosis. Hum Mut 4:114-118

Zygulska M, Eigel A, Pietrzyk J.J. and Horst J. 1994. Phenylketonuria in Southern Poland: A new splice mutation in intron 9 at the PAH locus. Hum Mut 4:297-299