Am. J. Hum. Genet. 51:627–636, 1992

Associations between Mutations and a VNTR in the Human Phenylalanine Hydroxylase Gene

Alexei A. Goltsov,* Randy C. Eisensmith,* David S. Konecki,† Uta Lichter-Konecki,† and Savio L. C. Woo*

*Howard Hughes Medical Institute, Department of Cell Biology, Baylor College of Medicine, Houston; and †Kinderklinik, Universität Heidelberg, Heidelberg

Summary

The *Hin*dIII RFLP in the human phenylalanine hydroxylase (PAH) gene is caused by the presence of an AT-rich (70%) minisatellite region. This region contains various multiples of 30-bp tandem repeats and is located 3 kb downstream of the final exon of the gene. PCR-mediated amplification of this region from haplotyped PAH chromosomes indicates that the previously reported 4.0-kb *Hin*dIII allele contains three of these repeats, while the 4.4-kb *Hin*dIII allele contains 12 of these repeats. The 4.2-kb *Hin*dIII fragment can contain six, seven, eight, or nine copies of this repeat. These variations permit more detailed analysis of mutant haplotypes 1, 5, 6, and, possibly, others. Kindred analysis in phenylketonuria families demonstrates Mendelian segregation of these VNTR alleles, as well as associations between these alleles and certain PAH mutations. The R261Q mutation, associated with haplotype 1, is associated almost exclusively with an allele containing eight repeats; the R408W mutation, when occurring on a haplotype 1 background, may also be associated with the eight-repeat VNTR allele. Other PAH mutations associated with haplotype 1, R252W and P281L, do not appear to segregate with specific VNTR alleles. The IVS-10 mutation, when associated with haplotype 6, is associated exclusively with an allele containing seven repeats. The combined use of this VNTR system and the existing RFLP haplotype system will increase the performance of prenatal diagnostic tests based on haplotype analysis. In addition, this VNTR may prove useful in studies concerning the origins and distributions of PAH mutations in different human populations.

Introduction

Classical phenylketonuria (PKU) is a relatively common autosomal recessive disorder. The frequency of PKU among populations varies significantly, from approximately 1/2,600 in Turkey (Özalp et al. 1986) to approximately 1/120,000 in Japan (Aoki and Wada 1988). Overall, the frequency among Caucasians is approximately 1/10,000 (Bickel et al. 1981), corresponding to a carrier frequency of about 1/50. PKU is primarily a consequence of a deficiency in hepatic phenylalanine hydroxylase (PAH) activity, and untreated PKU patients develop severe mental retardation. The isolation of a full-length human PAH cDNA (Kwok et al. 1985) permitted the identification of the structure of the PAH gene (DiLella et al. 1986*a*). Southern analysis of many individual genomic DNA samples digested by a battery of restriction enzymes demonstrated the presence of at least eight RFLPs associated with the PAH gene (Lidsky et al. 1985). The high degree of heterozygosity of these RFLPs among Caucasian PKU families greatly facilitated the implementation of prenatal diagnosis and genetic counseling based on haplotype analysis in PKU families (Chakraborty et al. 1987).

Southern analysis of genomic DNA digested with specific restriction endonucleases often demonstrates two alleles, reflecting the presence or absence of a particular restriction site. Direct sequence analysis has demonstrated that at least six of the eight RFLPs in the human PAH gene are the result of a single nucleotide

Received October 31, 1991; final revision received April 29, 1992.

Address for correspondence and reprints: Savio L. C. Woo, Ph.D., Howard Hughes Medical Institute and Department of Cell Biology, Baylor College of Medicine, Houston, TX 77030.

0002-9297/92/5103-0019$02.00

substitution that alters a restriction site (A. A. Goltsov, unpublished observations). The presence of more than two RFLP alleles is often due to variability in the number of repeated units between restriction sites. In humans such polymorphisms have been observed near the insulin gene (Bell et al. 1982), in the alpha-globin gene cluster (Proudfoot et al. 1982), in the Harvey-ras oncogene (Capon et al. 1983), in apolipoprotein B-100 (Huang and Breslow 1987), and in other genetic loci. For example, 14 different alleles of the APO B-100 3′-flanking region containing 25–52 repeats of a basic 15-bp unit were distinguished in a population study of 318 unrelated individuals (Ludwig et al. 1989). Similarly, three alleles have been reported for the *Hin*dIII polymorphism of the human PAH gene (4.0, 4.2, and 4.4 kb) (Woo et al. 1983), and possible differences in the number of 200 bp repeated units have been proposed to account for these three alleles (DiLella et al. 1986*a*).

In the present study, a number of PAH chromosomes, mostly of defined haplotypes, were examined to determine the exact cause of the length variations associated with the *Hin*dIII polymorphism. PCR-mediated amplification of this polymorphic segment revealed the presence of an AT-rich (70%) minisatellite region containing VNTRs (Jeffreys et al. 1985). Comparison of this region in Caucasian PKU families demonstrates the presence of at least six alleles that differ in the number of repeated units. These alleles are inherited in a Mendelian fashion and are often associated with specific PAH mutations.

Patients, Material, and Methods

Selection of Patient Samples

Genomic DNA obtained from leukocytes of 295 members of European PKU families was examined. Most samples had been haplotyped previously, although some samples were of unknown haplotype. The presence of incomplete families resulted in the examination of a greater number of mutant alleles than normal alleles. Samples were initially chosen to permit a detailed examination of the heterogeneity at the *Hin*dIII polymorphic site and, secondarily, to permit a detailed examination of the relationship between this polymorphism and specific PAH mutations. Thus, the sample population examined in this study is not a randomly distributed sample of all European populations or mutant PAH alleles but rather was selected to reflect the overall frequencies of the three *Hin*dIII alleles and RFLP haplotypes in the general European population. The frequencies of the various RFLP haplotypes and *Hin*dIII alleles among both normal and mutant chromosomes closely match those reported by Daiger et al. (1989) for the general European population, with the following exceptions: There was some overrepresentation of the rare 4.4-kb *Hin*dIII allele, to ensure that enough alleles were examined to successfully determine the presence or absence of VNTR size heterogeneity. Similarly, there was some overrepresentation of the 4.2-kb *Hin*dIII allele, to accurately determine the frequencies of the several different VNTR-size alleles contained on chromosomes of this type. There was also overrepresentation of mutant haplotype 6. This haplotype is extremely common in Turks, who had not been studied by Daiger et al. This was done to fully examine the association between the IVS-10 mutation and this polymorphic RFLP haplotype in different populations. There was some underrepresentation of the 4.0-kb *Hin*dIII allele and mutant RFLP haplotypes 2 and 3, since no chromosomes of these types demonstrated any size heterogeneity of VNTR alleles. Despite these exceptions, the sample population is, in general, representative both of the relative frequencies of the major mutant alleles and of the incidence of PKU within the different ethnic populations examined.

Plasmids

Plasmid pKSP23E12 contains a 12-kb *Eco*RI fragment isolated from a human genomic DNA library of a PKU patient homozygous for RFLP haplotype 3 (DiLella et al. 1986*b*). This *Eco*RI fragment contains the 3′ region of the PAH gene, including exons 9–13 and their flanking sequences (fig. 1). A 3.6-kb *Hin*dIII fragment and a 0.6-*Taq*I-*Hin*dIII fragment derived from pKSP23E12 were subcloned into pBluescript-M13+ for sequence analysis.

Oligonucleotides

All oligonucleotides were synthesized by American Synthesis, Inc. Primer 1 (5′-TTGGTTGGTAGATATTCAGG-3′) is complementary to the antisense sequence beginning 750 bp downstream of the terminal codon of the PAH gene. Primer 2 (5′-GCTTGAAACTTGAAAGTTGC-3′) is complementary to the antisense strand beginning 118 bp upstream from the first VNTR unit. Primer 3 (5′-GGAAACTTAAGAATCCCATC-3′) is complementary to the sense strand beginning 155 bp downstream of the last VNTR unit.

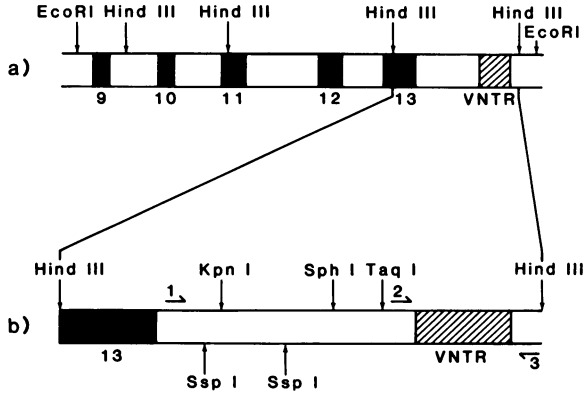

Figure 1 Restriction map of the 3′ end of the human PAH gene. Blackened and unblackened boxes represent exonic and intronic regions, respectively, and the hatched box represents the region containing a VNTR. *a,* Restriction map of the 12-kb *Eco*RI fragment from pKSP23E12. This fragment was originally isolated from a human genomic DNA library from a PKU patient homozygous for haplotype 3. *b,* 3.6-kb *Hin*dIII fragment containing the terminal portion of exon 13 of the PAH gene and the VNTR segment localized within the 0.6-kb *Taq*I-*Hin*dIII fragment. The relative positions of the primers used to amplify this region are indicated by the horizontal arrows.

PCR

PCR reactions were performed using an automated thermal cycler and *Taq* DNA polymerase (Perkin Elmer Cetus). Reaction mixtures contained 0.2–1 μg of genomic DNA, 0.5 μM of each primer, and 200 μM of each dNTP in 100 μl of 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl, and 0.01% gelatin. In reactions using primers 1 and 3, 40 cycles were performed, each consisting of 2 min of denaturation at 92°C, 2 min of annealing at 50°C, and 5 min of extension at 72°C. In reactions using primers 2 and 3, 35 cycles were used, each consisting of 40 s of denaturation at 92°C, 30 s of annealing at 55°C, and 40 s of extension at 72°C.

DNA Sequencing

Sequencing of double-stranded plasmid DNA and double-stranded amplified DNA was performed as described by Reichardt and Woo (1991).

Results

Identification of a VNTR within the Polymorphic HindIII *Fragment in the Human PAH Gene*

The restriction map of the 3′ end of the PAH gene is depicted in figure 1. According to initial genomic mapping, the *Hin*dIII polymorphic region was located inside a 4.0–4.4-kb fragment downstream of the constant *Hin*dIII site in exon 13 (DiLella et al. 1986*a*). More precise measurements of RFLP fragments in this region further reduced this length to 3.4–3.8 kb (Svensson et al. 1990). A previously constructed plasmid, pKSP23E12, contains a 3.6-kb *Hin*dIII fragment that includes the end of exon 13 and 3 kb of the 3′ flanking sequence. PCR amplification of this 3-kb fragment (primers 1 and 3; fig. 1*b*) from genomic DNA of individuals who are positive or negative for this polymorphism indicated that the polymorphic region is located between *Taq*I and *Hin*dIII restriction sites (data not shown). Sequencing of a 0.6-kb *Taq*I-*Hin*dIII fragment derived from pKSP23E12 gave the primary structure of this region (fig. 2). It contains a 240-bp AT-rich (70%) minisatellite beginning 133 bp downstream from the *Taq*I site. This sequence can be divided into eight 30-bp tandemly repeated units. This result suggests that the previously observed *Hin*dIII polymorphism may be caused by variations in the numbers of these repeated units.

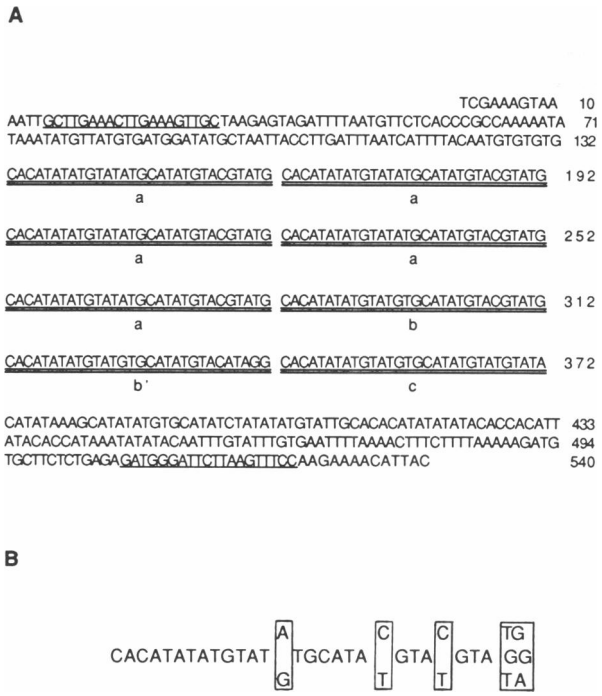

Figure 2 Sequence of the AT-rich minisatellite region present in pKSP23E12. *A,* Complete sequence of the *Taq*I-*Hin*dIII fragment from pKSP23E12, subcloned into pBluescript for sequence analysis. The PCR primer sequences are underlined, and VNTR units are double underlined. *B,* Consensus sequence for the repeated unit, derived from sequencing of 10 different VNTR alleles. Variable nucleotides are indicated by boxes.

Multiple VNTR Alleles and Their Mendelian Segregation

This AT-rich minisatellite region was amplified from genomic DNAs of patients and other members of PKU families, both to examine possible heterogeneity in the number of repeats present on chromosomes of various haplotypes in different human populations and to determine their Mendelian segregation. Electrophoretic resolution of the amplified products demonstrated DNA fragments of six discrete sizes—380, 470, 500, 530, 560, and 650 bp (fig. 3); direct sequencing of the PCR products indicated that these size differences reflected the presence of 3, 6, 7, 8, 9, or 12 copies of the repeated unit, respectively. The fragment containing three repeats corresponded to the 4.0-kb *Hin*dIII allele, and the fragment containing 12 repeats corresponded to the 4.4-kb *Hin*dIII allele. The small difference between intermediate fragments containing six, seven, eight, or nine repeats was apparently difficult to discern by previous Southern blot methods, and they were all classified as a single 4.2-kb *Hin*dIII allele. Thus, the *Hin*dIII polymorphism is the result of VNTRs in the 3′ region of the PAH gene.

Comparison of the length of PCR fragments in PKU families confirmed the Mendelian segregation of these VNTR alleles. Three pedigrees illustrating the inheritance of these VNTRs are shown in figure 4. In family A, both affected children inherited alleles with seven and eight repeats. Allele-specific oligonucleotide hybridization analysis of this family has previously shown that the father is a carrier of the R252W mutation and that the mother is a carrier of the P281L mutation (Okano et al. 1991). Since the patients are homozygous for mutant haplotype 1, both of these mutations are associated with the same RFLP haplotype but with different VNTR alleles. In family B, the R252W mutation is associated with haplotype 6 and the seven-repeat allele. In family C, both parents are carriers for the P281L mutation, associated with haplotype 1 and the seven-repeat allele.

Associations between VNTR Alleles and RFLP Haplotypes in the Human PAH Locus

The observation of several different VNTR alleles associated with a single haplotype (fig. 4, compare

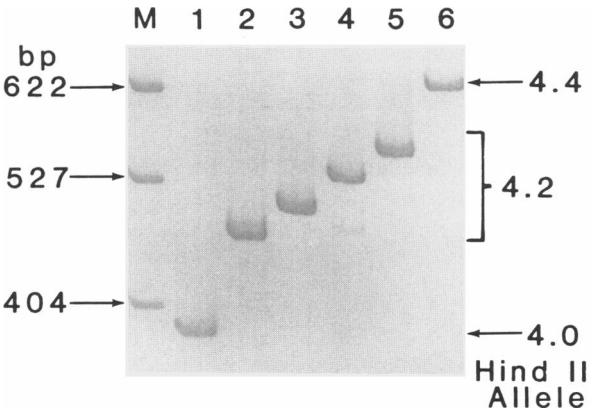

Figure 3 PCR-mediated amplification of the VNTR-containing region of the PAH gene. Primers 2 and 3 were used to specifically amplify the VNTR-containing region from genomic DNA of individuals of different RFLP haplotypes. Alleles differ by one (30 bp) or more VNTR units. From left to right, the number of repeats and the fragment lengths are as follows: 3, 380-bp fragment (lane 1); 6, 470-bp fragment (lane 2); 7, 500-bp fragment (lane 3); 8, 530-bp fragment (lane 4); 9, 560-bp fragment (lane 5); and 12, 650-bp fragment (lane 6). Lane M, *Msp*I digest of pBR322. The association of these VNTR alleles with the three *Hin*dIII alleles is indicated to the right.

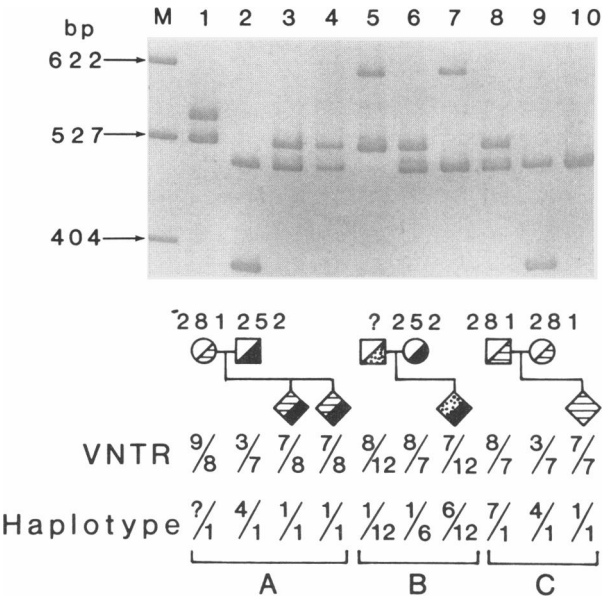

Figure 4 Mendelian inheritance of the 3′ PAH VNTR alleles in three PKU families. The pedigrees of three families are depicted below the corresponding lanes from an ethidium bromide–stained 4% polyacrylamide gel. The allele designations below each pedigree correspond to the number of repeats present in each PCR product. Haplotype and genotype information obtained from previous studies is also provided for comparison. A semi-blackened symbol indicates the presence of the R252W mutation, a semihatched symbol indicates the presence of the P281L mutation, the hatched symbol indicates the presence of two P281L mutations, and a semistippled symbol indicates the presence of an unknown mutation.

families A and C) led to a more detailed investigation of the relationship between previously reported RFLP haplotypes and these newly defined VNTR alleles. Table 1 displays the genotypes of eight of the more common PAH RFLP haplotypes examined in these studies. Table 2 presents a comparison of the numbers of normal and mutant chromosomes observed in a sample population drawn from European PKU families, segregated by RFLP haplotype and VNTR allele size. No heterogeneity of VNTR size was observed in those haplotypes bearing the 4.0-kb *Hin*dIII allele—e.g., haplotypes such as 2 or 4—that are characterized by three copies of the repeat. Similarly, no heterogeneity of VNTR size was observed for haplotypes bearing the 4.4-kb *Hin*dIII allele—e.g., haplotypes such as 12—that are characterized by 12 copies of the repeat. Regarding haplotypes containing intermediate-sized fragments, i.e., those haplotypes previously classed as having a 4.2-kb *Hin*dIII allele, heterogeneity of VNTR number was observed for haplotypes 1, 3, 5, 6, 7, and others of known or unknown haplotype. Although most haplotype 1 alleles contained eight copies of the VNTR, a significant number of these alleles contained seven copies, and one allele contained nine copies. Similarly, most haplotype 5 alleles had nine copies of the repeated unit, and most haplotype 6 alleles contained seven repeats. No size heterogeneity was observed for mutant haplotype 3 or mutant haplotype 7, each of which contained eight copies of the repeat. Although chromosomes of other haplotypes bearing the 4.2-kb *Hin*dIII allele were also polymorphic, too few of each class were available for analysis to conclusively establish differences in VNTR size.

VNTR Allele Frequencies

The relative frequencies of the VNTR alleles within a specific, nonrandom subset of European populations can be derived from the data presented in table 2. Since the 4.0- and 4.4-kb *Hin*dIII alleles examined in this study were not polymorphic for VNTR size, 100% of these *Hin*dIII alleles are associated with the 3- or 12-repeat VNTR alleles, respectively. The absolute frequency of these two VNTR alleles cannot be determined directly from our sample population, but it will directly reflect the combined frequencies of the RFLP haplotypes bearing either the 4.0- or 4.4-kb *Hin*dIII allele in any given population. Within chromosomes bearing the polymorphic 4.2-kb *Hin*dIII allele, the VNTR allele containing eight repeats is most prevalent (about 60%) among both normal and mutant chromosomes examined in this study. The VNTR allele containing seven repeats is the second most frequent allele associated with chromosomes bearing the 4.2-kb *Hin*dIII allele. This VNTR allele is present on about one-fifth of all normal chromosomes and on about one-third of all mutant chromosomes of this type. The VNTR allele containing nine repeats accounts for most of the remaining normal or mutant chromosomes in this class. Only one mutant allele bearing six VNTR units was observed in this study. Direct translation of the frequencies of these VNTR alleles to larger, more general populations will likely

Table 1

Selected RFLP Haplotypes at the PAH Locus

	Status of [a]							
Haplotype	*Bgl*II	*Pvu*II (a)	*Pvu*II (b)	*Eco*RI	*Msp*I	*Xmn*I	*Hin*dIII[b]	*Eco*RV
1	−	+	−	−	+	−	−	−
2	−	+	−	−	+	−	+	+
3	−	+	−	+	−	+	−	−
4	−	+	−	+	−	+	+	+
5	+	−	+	+	+	−	−	+
6	+	−	+	+	+	−	−	−
7	+	−	−	+	−	+	−	−
12	−	+	−	−	+	−	=	+

[a] A plus sign (+) denotes presence of restriction site; and a minus sign (−) denotes absence of restriction site.

[b] A plus sign (+) denotes presence of the 4.0-kb allele; a minus sign (−) denotes presence of the 4.2-kb allele; and the equals sign (=) denotes presence of the 4.4-kb allele.

Table 2

Associations among PAH RFLP Haplotypes and VNTR Alleles

Haplotype and No. of Repeated Units in VNTR	No. of Alleles Observed	
	Normal	Mutant
1:		
7	14	20
8	44	85
9	1	0
2:		
3	9	45
3:		
8	5	29
9	1	0
4:		
3	47	58
5:		
7	1	0
8	2	4
9	20	9
6:		
7	4	50
8	0	1
9	0	1
7:		
7	1	0
8	19	9
12:		
12	14	11
Other:		
3	10	7
6	0	1
7	5	10
8	19	18
9	9	2
12	3	2

result in some differences, due both to differences in the frequencies of the haplotypes bearing the 4.2-kb *Hin*dIII allele in different populations and to the purposefully nonrandom selection of patient samples in this study (see Selection of Patient Samples subsection above).

Associations between VNTR Alleles and PAH Mutations

The polymorphic nature of haplotype 1 led to an examination of the relationship between specific PAH mutations, RFLP haplotypes, and VNTR size (table 3). Previous studies have shown that the R261Q mutation is associated predominantly with haplotype 1 (Okano et al. 1990*b*), and this mutation was also associated primarily with a single VNTR allele (i.e., 8). Similarly, the IVS-10 mutation, when associated with haplotype 6 (Dasovich et al. 1991; Kalaydjieva et al. 1991*a*), was associated exclusively with a single VNTR allele (i.e., 7). In contrast, the P281L mutation, which is associated predominantly with haplotype 1 (Dworniczak et al. 1991; Okano et al. 1991), was associated with alleles containing seven or eight repeats, when present on this haplotype, and with an allele containing three repeats, when present on haplotype 4. In addition, the R252W mutation, when present on haplotype 1, was observed on two different VNTR alleles (i.e., 7 and 8) in the Italian population and on a different haplotype (i.e., 6) but on the same VNTR (i.e., 7) background in the Swedish population. Nearly all of the remaining mutations were associated with a single VNTR allele. This is primarily a consequence of their exclusive association with RFLP haplotypes that are not polymorphic for this VNTR (tables 2 and 3).

Discussion

A VNTR 3 kb from the 3′ end of the last exon of the PAH gene was shown to be responsible for the previously reported *Hin*dIII RFLP. Identification of this VNTR did not significantly increase, from the 69 presently documented (S. L. C. Woo, unpublished observation), the number of RFLP haplotypes observed at the human PAH locus, since the presence of this VNTR system created, at most, only four additional variants of some haplotypes bearing the 4.2-kb *Hin*dIII allele. However, both the apparently stable Mendelian segregation of VNTR alleles in PKU families and the variation in VNTR sizes associated with some 4.2-kb *Hin*dIII alleles led us to examine this system as a potentially useful marker in the PAH locus and permitted a more detailed examination of several haplotypes.

Reexamination of mutation/haplotype associations in light of this new VNTR allele system provided some additional insights regarding the origin of these mutations in different populations, especially those mutations associated with polymorphic haplotypes such as haplotype 1 (John et al. 1990; Okano et al. 1990*a*, 1990*b*; 1991; Dworniczak et al. 1991). The R261Q mutation appears to be the predominant mutation present on haplotype 1 alleles, especially in the Swiss (Okano et al. 1990*b*), Turkish (Konecki and Lichter-Konecki 1991), and German populations (Konecki and Lichter-Konecki 1991). The present distribution of the R261Q mutation could therefore be the result of founder effect and subsequent migrational

Table 3

Associations among PAH Mutations, RFLP Haplotypes, and VNTR Alleles in Different European Populations

Mutation	Haplotype[a]	No. of Repeated Units in VNTR	No. of Mutant Alleles Observed in						
			Denmark	Federal Republic of Germany	Italy	Sweden	Switzerland	Turkey	Other Populations
L48S	4	3	0	0	1	0	0	5	0
R158Q	4	3	2	0	0	0	2	4	5
E221K	3	8	0	0	0	0	0	1	0
R243X	4	3	0	0	1	0	0	0	1
R252W	1	7	0	0	1	0	0	0	0
	1	8	0	0	1	0	0	0	0
	6	7	0	0	0	1	0	0	0
R261Q	1	7	0	1	0	0	0	0	0
	1	8	3	11	2	3	15	7	2
E280K	1	8	0	0	0	0	0	0	1
	4	3	0	1	0	0	0	0	0
P281L	1	7	0	2	4	3	1	2	1
	1	8	0	3	2	0	0	0	0
	4	3	0	0	0	1	2	1	0
A322G	12	12	0	0	0	4	0	0	0
IVS-10nt546	ND	6	0	0	0	0	1	0	0
	6	7	4	3	7	0	1	33	1
	ND	7	2	0	4	0	0	0	1
	ND	8	0	0	1	0	0	0	1
R408W	1	8	0	0	0	2	0	0	0
	2	3	7	11	0	7	2	1	8
	ND	8	0	0	0	0	0	0	2
R408Q	12	12	0	0	0	3	0	0	0
IVS-12nt1	3	8	16	3	0	5	0	0	1

[a] ND = not determined.

effects. Alternatively, this distribution could be the result of recurrence of the R261Q mutation on haplotype 1 chromosomes in different populations, especially since this mutation involves a CpG dinucleotide, a known hot spot for mutation in the PAH gene (Abadie et al. 1989). We have demonstrated a strong association between the R261Q mutation and chromosomes with eight copies of the VNTR (43 of 44 R261Q alleles) in several European populations and also in Turkey (table 3). This observation suggests a common origin for this mutation, with subsequent dispersal into different populations. The distribution of frequencies of VNTR variants within the 4.2-kb *Hin*dIII allele (extrapolated from table 2) makes recurrent mutation less likely as an underlying mechanism, since this would imply that the R261Q mutation occurred repeatedly only on the same VNTR variant of haplotype 1 and not on other RFLP or VNTR haplotype backgrounds. Such an event seems relatively improbable when one considers the low (12.5%) frequency of normal haplotype 1 alleles in the Turkish population (Lichter-Konecki et al. 1989). It may therefore be more likely that founder effect and genetic drift were the primary mechanisms responsible for the present distribution of the R261Q mutation, although the influences of genetic isolation and drift preclude localization of this putative founding population. On the basis of this hypothesis, the presence of the R261Q mutation on a single haplotype 1 chromosome with seven repeats would then be the result of slippage (Schlötterer and Tautz 1992) or other rearrangements within the minisatellite region, rather than the result of recurrence.

In contrast, the association of the P281L mutation with RFLP haplotypes and VNTR alleles suggests that this mutation recurred at least once. This mutation is present on chromosomes of two highly unrelated haplotypes—i.e., 1 and 4 (see table 1)—in Germany

(Dworniczak et al. 1991), Sweden (E. Svensson, personal communication), Switzerland (Y. Okano, personal communication), and Turkey (D. S. Konecki, unpublished observation). Alternatively, one could argue that this mutation predated the divergence of haplotypes 1 and 4 and was redistributed to chromosomes of other VNTR haplotypes by genomic rearrangements such as slippage. This hypothesis is not supported by the data on the frequencies of the different VNTR alleles presented in tables 2 and 3. Thus, it seems more likely that the associations between the P281L mutation and these VNTR alleles are due to recurrent mutation and that the event producing the P281L mutation is more recent than the events producing this VNTR-allele system.

Haplotype 6 is associated with VNTR-size alleles containing seven, eight, and nine repeats. As has been reported independently by Dasovich et al. (1991) and Kalaydjieva et al. (1991*a*), the IVS-10 mutation is associated primarily with mutant haplotype 6, a common haplotype in Mediterranean regions. In our sample population, the IVS-10 mutation, when present on haplotype 6, was associated exclusively (49/49) with the VNTR allele containing seven repeats (tables 2 and 3). The strong degree of association observed between the IVS-10 mutation and the seven-repeat variant of mutant haplotype 6, similar to that observed between the R261Q mutation and the eight-repeat variant of haplotype 1, is again consistent with a relatively recent founder effect for this mutation, as proposed by Dasovich et al. (1991). In contrast, this finding is not supportive of a more ancient origin for this mutation, as Kalaydjieva et al. (1991*a*, 1991*b*) suggested on the basis of its presence on a few alleles of other RFLP haplotypes.

The nonexclusive associations between VNTR alleles and PAH mutations observed in the present study are caused, in two cases, by the presence of a given mutation on a single allele of another VNTR haplotype. We have proposed that these results are produced by the gain or loss of single VNTR units, possibly by slippage during DNA replication (Jeffreys et al. 1985; Schlötterer and Tautz 1992), rather than by recurrence of mutations on different VNTR-haplotype backgrounds. This hypothesis is based, in part, on the assumption that there is probably little or no selection from VNTR alleles but that there may be some adverse selective pressure on changes in the coding region of the gene. One piece of evidence supporting this assumption is the observation of small differences in the sequences of repeats, mostly between VNTR alleles of different sizes but sometimes within alleles of the same size (data not shown). The presence of sequence heterogeneity raises the possibility that RFLP haplotypes monomorphic for VNTR-size alleles may, in fact, be polymorphic with respect to VNTR sequence. This issue is very interesting, since sequence heterogeneity may significantly increase the PIC of this VNTR system, further increasing its utility as a diagnostic tool.

The association of multiple PAH mutations with VNTR alleles of the same repeat number suggests that most PAH mutations may have occurred more recently than the events producing the different VNTR alleles. Furthermore, the presence of VNTR alleles with seven or eight repeats on haplotypes that, independent of the *Hin*dIII RFLP, differ significantly from one another suggests that the origin of the VNTR system, and perhaps the divergence as well, may have predated the origin or divergence of certain RFLP haplotypes. The combined use of data concerning the frequencies and distributions of RFLP haplotypes, VNTR alleles, and PAH mutations in various ethnic populations may ultimately permit a more accurate determination of a relative time frame for the many mutational or other genetic events that have now been observed in the human PAH locus.

Such studies are, in turn, dependent on a more complete understanding of the mechanisms responsible for the origin and subsequent rearrangements of this VNTR region. Boerwinkle et al. (1989) have speculated that the high degree of variability and the multimodal frequency distribution of the 3′ ApoB-100 and other VNTR allele systems may be the result of unequal recombination resulting from either mismatching of repeat units or replication slippage. A similar trimodal distribution of VNTR alleles is observed in the present study. In contrast to some of these other systems, however, there was a discontinuous distribution of allele sizes at the PAH VNTR locus. Certain alleles were apparently absent. No unit less than three repeats was ever observed, suggesting that the three-repeat unit represents the minimal size of this system in the human genome. It is not completely clear what factors prevent contraction of this system to less than three repeats in man. However, sequence comparisons of limited numbers of three-repeat alleles demonstrate the presence, in each repeat, of unique sequences that may prevent the further loss of single units (data not shown). Also apparently absent were VNTR alleles containing 4, 5, 10, or 11 copies of the repeated unit. The frequency distribution of the

different-size alleles present in this VNTR system, along with the observation of unique sequences for each of the three repeats in the smallest allele, suggests that the initial event leading to the formation of larger-sized alleles was at least a duplication of the basic three-repeat unit, with subsequent divergence through the introduction of single base changes. More complex combinations of unequal crossing-over and gain or loss of single repeat units, without exchange of flanking markers (Wolff et al. 1988) and perhaps influenced by sequence heterogeneity, have probably also occurred in this region. Nevertheless, the apparently stable Mendelian segregation of these VNTR alleles in PKU kindreds suggests that the rate of these more complex rearrangement processes, which appears to be relatively high compared with the rate of mutation in coding regions of the PAH gene, may be sufficiently low to permit the use of this VNTR system in prenatal diagnosis by haplotype analysis.

Acknowledgments

This work was supported in part by NIH grant HD-17711 to S.L.C.W., who is also an Investigator with the Howard Hughes Medical Institute. This work was also supported in part by Deutsche Forschungsgemeinschaft grant Li.375/2-1/2-2 to U.L.-K. and by Fritz Thyssen Foundation grant 1990/1/51 to D.S.K. (grant was awarded to Professor Dr. H. J. Bremer). We would like to express our sincere gratitude to the many individuals who contributed samples from PKU patients and their families. We would like to thank Drs. Ranajit Chakraborty and Elizabeth Svensson and Mr. Jin Li for their critical reading of the manuscript and for insightful comments.

References

Abadie V, Lyonnet S, Maurin N, Bertelon M, Caillaud C, Giraud F, Mattel J-F, et al (1989) CpG dinucleotides are mutation hot spots in phenylketonuria. Genomics 5:936–939

Aoki K, Wada Y (1988) Outcome of the patients detected by newborn screening in Japan. Acta Paediatr 30:429–434

Bell IG, Selby JM, Rutter WJ (1982) The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences. Nature 295:31–35

Bickel H, Bachmann C, Beckers R (1981) Neonatal mass screening for metabolic disorders. Eur J Pediatr 137:133–139

Boerwinkle E, Xiong W, Fourest E, Chan L (1989) Rapid typing of tandemly repeated hypervariable loci by the polymerase chain reaction: application to the apolipoprotein B 3′ hypervariable region. Proc Natl Acad Sci USA 86:212–216

Capon DJ, Chen EY, Levinsson AD, Seeburg PH, Goeddel DV (1983) Complete nucleotide sequences of the T24 human bladder carcinoma oncogene and its normal homologue. Nature 302:33–37

Chakraborty R, Lidsky AS, Daiger SP, Gütter F, Sullivan S, DiLella AG, Woo SLC (1987) Polymorphic DNA haplotypes at the human phenylalanine hydroxylase locus and their relationship with phenylketonuria. Hum Genet 76:40–46

Daiger SP, Chakraborty R, Reed L, Fekete G, Schuler D, Berenssi G, Nasz I, et al (1989) Polymorphic DNA haplotypes at the phenylalanine hydroxylase (PAH) locus in European families with phenylketonuria (PKU). Am J Hum Genet 45:310–418

Dasovich M, Konecki D, Lichter-Konecki U, Eisensmith RC, Güttler F, Naughton E, Mullins C, Woo SLC (1991) Molecular characterization of a PKU allele prevalent in southern Europe and Ireland. Somatic Cell Mol Genet 17:303–309

DiLella AG, Kowk SCM, Ledley FD, Marvit J, Woo SLC (1986*a*) Molecular structure and polymorphic map of the human phenylalanine hydroxylase gene. Biochemistry 25:743–749

DiLella AG, Marvit J, Lidsky AS, Güttler F, Woo SLC (1986*b*) Tight linkage between a splicing mutation and a specific DNA haplotype in phenylketonuria. Nature 322:799–803

Dworniczak B, Grudda K, Strumper J, Bartholome K, Aulehla-Scholz C, Horst J (1991) Phenylalanine hydroxylase gene: novel missense mutation in exon 7 causing severe phenylketonuria. Genomics 9:193–199

Huang L-S, Breslow JL (1987) A unique AT-rich hypervariable minisatellite 3′ to the apoB gene defines a high information restriction fragment length polymorphism. J Biol Chem 262:8952–8955

Jeffreys AJ, Wilson V, Thein SL (1985) Hypervariable "minisatellite" regions in human DNA. Nature 314:67–73

John SWM, Rozen R, Scriver GR, Laframboise R, 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

Kalaydjieva L, Dworniczak B, Aulehla-Scholz C, Devoto M, Romeo G, Stuhrmann M, Horst J (1991*a*) Phenylketonuria mutation in southern Europeans. Lancet 1:865

Kalaydjieva L, Dworniczak B, Kremensky I, Horst J (1991*b*) Heterogeneity of haplotype 1 and 4 alleles in Bulgarian classical PKU. Am J Hum Genet 49 [Suppl]: A1027

Konecki DS, Lichter-Konecki U (1991) The phenylketonuria locus: current knowledge about alleles and mutations of the phenylalanine hydroxylase gene in various populations. Hum Genet 87:377–388

Kwok SCM, Ledley FD, DiLella AG, Robson KJH, Woo SLC (1985) Nucleotide sequence of a full-length complementary DNA clone and amino acid sequence of human phenylalanine hydroxylase. Biochemistry 24:556–561

Lichter-Konecki U, Schlotter M, Yaylak C, Özgüç M, Çoskun T, Özalp I, Wendel U, et al (1989) DNA haplotype analysis at the phenylalanine hydroxylase locus in the Turkish population. Hum Genet 81:373–376

Lidsky AS, Ledley FD, DiLella AG, Kwok SCM, Daiger SP, Robson KJH, Woo SLC (1985) Extensive restriction site polymorphism at the human phenylalanine hydroxylase locus and application in prenatal diagnosis of phenylketonuria. Am J Hum Genet 37:619–634

Ludwig EH, Freidl W, McCarthy BJ (1989) High-resolution analysis of a hypervariable region in the human apolipoprotein B gene. Am J Hum Genet 45:458–464

Okano Y, Wang T, Eisensmith RC, Güttler F, Woo SLC (1990*a*) Recurrent mutation in the human phenylalanine hydroxylase gene. Am J Hum Genet 46:919–924

Okano Y, Wang T, Eisensmith RC, Longhi R, Giovannini M, Cerone R, Romano C, et al (1991) Phenylketonuria missense mutations in the Mediterranean. Genomics 9: 96–103

Okano Y, Wang T, Eisensmith RC, Steinmann B, Gitzelmann R, Woo SLC (1990*b*) Missense mutations associated with RFLP haplotypes 1 and 4 of the human phenylalanine hydroxylase gene. Am J Hum Genet 46:18–25

Özalp I, Çoskun T, Ceyhan M, Tokol S, Oran O, Erdem G, Tekinalp G, et al (1986) Incidence of phenylketonuria and hyperphenylalaninemia in a sample of the newborn population. J Inherited Metab Dis 9 [Suppl 2]: 237–239

Proudfoot NJ, Gil A, Maniatis T (1982) The structure of the human beta-globin gene and closely linked, nearly identical pseudogene. Cell 31:553–563

Reichardt JKV, Woo SLC (1991) Molecular basis of galactosemia: mutations and polymorphisms in the gene encoding human galactose-1-phosphate uridylyltransferase. Proc Natl Acad Sci USA 88:2633–2637

Schlötterer C, Tautz D (1992) Slippage synthesis of simple sequence DNA. Nucleic Acids Res 20:211–215

Svensson E, Andersson B, Hagenfeldt L (1990) Two mutations within the coding sequence of the phenylalanine hydroxylase gene. Hum Genet 85:300–304

Wolff RK, Nakamura Y, White R (1988) Molecular characterization of a spontaneously generated new allele at a VNTR locus: no exchange of flanking DNA sequence. Genomics 3:347–351

Woo SLC, Lidsky A, Güttler F, Chandra T, Robson K (1983) Cloned human phenylalanine hydroxylase gene allows prenatal diagnosis and carrier detection of classical phenylketonuria. Nature 306:151–155