Biochemistry

 Volume 28, Number 1 January 10, 1989

Perspectives in Biochemistry

Molecular Basis and Population Genetics of Phenylketonuria†

Savio L. C. Woo

Howard Hughes Medical Institute, Department of Cell Biology and Institute of Molecular Genetics, Baylor College of Medicine, Houston, Texas 77030

Received September 13, 1988; Revised Manuscript Received November 1, 1988

Clinical and Biochemical Basis of Phenylketonuria

Phenylketonuria and Its Biochemical Lesion. Classical phenylketonuria (PKU) is caused by a deficiency of the hepatic enzyme phenylalanine hydroxylase (PAH) and is a typical example of inborn errors in amino acid metabolism. The disorder causes severe mental retardation in affected children who excrete large quantities of phenylpyruvate in the urine (Folling, 1934). A year after the discovery of the disease, Penrose (1935) observed that it was a genetic disorder transmitted as an autosomal recessive trait. Twelve years afterward, it was shown that the administration of phenylalanine to normal humans led to prompt elevation in serum tyrosine, but the response was absent in patients with PKU (Jervis, 1947). Subsequently, it was demonstrated that post mortem liver samples from normal individuals were able to convert phenylalanine to tyrosine in vitro, while those from PKU patients could not, thereby defining the bichemical basis of phenylketonuria (Jervis, 1953).

The Phenylalanine Hydroxylase System. The hydroxylation of phenylalanine to tyrosine is a complex biochemical reaction. In addition to PAH, a second protein fraction was required to convert phenylalanine to tyrosine (Mitoma, 1956); this was subsequently identified as the enzyme dihydropteridine reductase (Kaufman, 1957, 1959). Phenylalanine hydroxylase is a mixed-function monooxygenase that catalyzes the hydroxylation of phenylalanine to tyrosine by molecular oxygen in the presence of the cofactor tetrahydrobiopterin. During this reaction the cofactor is cooxidized to quinonoid dihydrobiopterin, which is subsequently reduced to the tetrahydro form by dihydropteridine reductase (Kaufman, 1976). Because of the complex nature of the enzymatic reaction, elevated serum phenylalanine levels can also be the result of enzymatic deficiency of the reductase and/or other enzymes involved in the biosynthesis of tetrahydrobiopterin (Kaufman et al., 1975, 1978; Leeming et al., 1976). This review deals only with hyperphenylalaninemic disorders due to PAH deficiency.

Dietary Treatment of PKU and Neonatal Screening. In the early 1950s, Bickel and collaborators placed young phenylketonuric patients on a diet consisting of a special protein hydrolysate from which phenylalanine was removed. They observed that not only was this treatment effective in the reduction of the patient's serum phenylalanine level and disappearance of phenylpyruvate from their urine, but also there was an apparent improvement in mental development and behavioral performance of the patients (Bickel et al., 1954). This report was the first to demonstrate that hereditary disorders in amino acid metabolism can be corrected by dietary management and, together with the dietary therapy of galactosemia, has served as a prototype for treatment of a variety of other inborn errors in metabolism.

The successful treatment of PKU provided a strong incentive to identify PKU in newborn patients, as the diet must be implemented as early in life as possible to be effective. While the ferric chloride test is useful in the diagnosis of PKU in children, it has only limited value in newborn screening as the elevation of the patient's serum phenylalanine preceded urinary excretion of phenylpyruvate by weeks (Armstrong & Binkley, 1956; Armstrong & Low, 1957). Nine years after the initial report on dietary correction of PKU, it was discovered that blood phenylalanine level can be easily measured in a semiquantitative manner by a bacterial inhibition assay (Guthrie & Susi, 1963). The method requires only a small volume of blood, is specific, is inexpensive, and is well suited to handle a large number of samples. It has since been adopted as a routine procedure to screen all newborns for PKU in most western countries. The first collective results of mass screening for PKU in eight western European countries were reported by Bickel et al. (1973), who observed a prevalence of about 1 in 8000 individuals. Subsequently, it was observed that the frequency of PKU varies considerably in various European countries, ranging from 1/3000 in Ireland to 1/28 000 in Belgium. The "Guthrie test" for PKU screening among newborns has also served as a prototype for implementing

† This work was supported in part by NIH Grant HD-17711. S.L.C.W. is an Investigator of the Howard Hughes Medical Institute.

0006-2960/89/0428-0001$01.50/0

newborn screening programs of other metabolic disorders.

Diet Discontinuation and Maternal PKU as a Public Health Problem

Diet discontinuation for patients after 6 years of age was a generally accepted practice in the 1960s and 1970s and was widely assumed to pose no substantial risk for treated PKU patients. A 5–10-year followup after diet discontinuation, however, suggested that these individuals suffered subtle changes in cerebral function and a decrease in IQ scores (Cabalska et al., 1977; Smith et al., 1978; Koch et al., 1982). More recent data based on a 10–15 year posttreatment followup under the auspices of the National Collaborative PKU project have confirmed this disturbing trend. These studies demonstrate an unequivocal deterioration of mean performance scores and severe deficits in specific areas, such as mathematical ability, approaching levels seen in untreated PKU (Koch, personal communication).

Many of the PKU patients who had been treated with the diet early in life have now reached child-bearing age. Most of these individuals had discontinued the diet years ago and have high serum phenylalanine levels. It has been observed that children born to severely hyperphenylalaninemic mothers suffered from mental retardatin, microcephaly, and/or other abnormalities including congenital heart disease. This novel syndrome has been termed "maternal PKU" (Perry et al., 1973), since it is apparent that normal fetal development can be disrupted by high level of phenylalanine transported across the placenta from the mother. This new clinical condition affects all infants born to PKU mothers regardless of the fact that most of these neonates are only carriers of PKU and bear a normal phenylalanine hydroxylase gene. Assuming each female PKU individual gave birth two two such infants, it has been estimated that it would only take one generation to create the same number of mentally retarded individuals as when there was no PKU screening in the population. Thus the success in dietary correction of PKU has created a new syndrome that is quickly becoming a public health problem.

Recent data suggest that lowering of the mother's serum phenylalanine level with dietary therapy before conception may prevent maternal PKU syndrome in the fetus (Levy & Waisbren, 1983). Nevertheless, it has been very difficult to resume the diet in adults and less than 25% of pregnant patients have successfully complied with the required dietary regimen (Levy, personal communication). Thus, the long-term solution to the problem of PKU may lie in a screening program for carriers so that the incidence can be reduced by family planning. The prerequisite to such a screening program is the ability to readily detect mutant alleles in carrier individuals without prior family history of PKU, and the application of molecular approaches toward this goal is the main theme of this review.

Molecular Basis and Genetic Analysis of PKU

Molecular Cloning of Rat and Human PAH cDNAs and Their Primary Structures. Phenylalanine hydroxylase (PAH) mRNA was purified from rat liver by polysome immunoprecipitation and used for the cloning of its cDNA. The authenticity of the cDNA clone was established by hybrid-selected translation (Robson et al., 1982) and confirmed by matching the nucleotide sequence with the partial amino acid sequence of the purified enzyme (Robson et al., 1984). By use of the cloned rat PAH cDNA as a specific hybridization probe, a human liver cDNA library comprised of 10^7 independent recombinants was screened. A full-length human PAH cDNA clone was obtained (designated phPAH247) and sequenced in its entirety. The clone contains an inserted DNA fragment of 2448 base pairs, including 19 bases of poly(A) at the 3′ end (Kwok et al., 1985). The first methionine codon occurs at nucleotide position 223, followed by an open reading frame of 1353 base pairs encoding 451 amino acids. The predicted amino acid sequence of the human enzyme was deduced from the nucleotide sequence of phPAH247 and shown to be 90% homologous with the amino acid sequence of the corresponding rat enzyme reported by Robson et al. (1984).

PAH Is a Homopolymer Encoded by a Single Genetic Locus in Man. Native PAH is a polymeric enzyme comprised of two 50-kDa subunit bands on SDS–polyacrylamide gels. It was not clear whether the enzyme is a heteropolymer or a homopolymer. This is a critical issue because if the enzyme is a heteropolymer, multiple genetic loci may be expected and genetic analysis could be complicated. The full-length human PAH cDNA was thus inserted into an eukaryotic expression vector containing the promoter and CAP site of the human metallothionein gene. The construct was cotransfected with pSV2neo DNA into mouse fibroblast cells (NIH3T3), which do not normally express PAH. The G-418 resistant mouse cells were cultured and shown to express PAH messenger RNA, immunoreactive PAH, and enzymatic activity characteristic of authentic human liver PAH (Ledley et al., 1985). Similar results were also obtained from analysis of a full-length rat PAH cDNA clone (Dahl & Mercer, 1986; Choo et al., 1986). More recently, the human PAH cDNA clone has been inserted behind a prokaryotic promoter and transferred into *Escherichia coli*. Bacterial extracts of the recombinants were shown to contain high levels of human PAH activity (Ledley et al., 1987). These data conclusively demonstrated that a single mRNA species contains all genetic information necessary to code for functional PAH. These observations support the notion that PAH is a homopolymer encoded by a single genetic locus. Thus the human PAH cDNA clone can be used as a probe to perform molecular analysis of the PAH gene and PKU locus.

The PKU Locus in Man Is on Chromosome 12. The PKU locus in man was originally studied by linkage analysis with other polymorphic human protein markers in PKU families with inconclusive results. It was originally assigned to chromosome 1 by moderate linkage data with the phosphoglucomutase locus PGM-1 (Berk & Saugstad, 1974) and the amylase loci AMY-1 and AMY-2 (Kamaryt et al., 1978). More extensive linkage studies using improved methods for obligate heterozygote determination among siblings in the PKU families, however, failed to establish genetic linkage between the PKU locus with a number of markers and was in disagreement with the previous assignment to human chromosome 1 (Paul et al., 1979).

Chromosomal assignments for human genetic loci can be made by using cloned genes as probes in molecular hybridization studies to genomic DNA isolated from human/rodent cell hybrids that contain different assortments of human chromosomes. A panel of mouse/human hybrid cell lines bearing an assortment of human chromosomes was analyzed by Southern blot using the human PAH cDNA clone as the hybridization probe. Results indicated that the PAH-hybridizing human DNA bands were consistent with only human chromosome 12, providing strong evidence that the human PAH locus is on chromosome 12 (Lidsky et al., 1984). Subsequently, the regional map position of PAH on human chromosome 12 was defined by deletion chromosome mapping as well as by in situ hybridization. It was observed that the hybridized grains were highly concentrated in the 12q22-q24.1 region (Lidsky et al., 1985a). Since the cDNA contains all

the genetic information necessary for expression of phenylalanine hydroxylase, which is the enzyme deficient in PKU, the PKU locus in man has also been mapped to 12q22-q24.1. More recently, the PAH locus has been mapped to chromosome 10 of the mouse (Ledley et al., 1988b).

Classical PKU Is Not Caused by Deletion of the Entire PAH Gene. Genomic DNAs were isolated from two PKU cell lines obtained from the Human Mutant Cell Repository (GM934 and GM2406) and from two normal individuals. The DNA preparations were digested with restriction enzymes followed by Southern hybridization using the human PAH cDNA clone as the specific probe. Identical hybridization signals were obtained from all four DNA preparations after digestion with a number of restriction enzymes (Woo et al., 1983). These results indicated not only that the PAH gene was present in the genome of cells derived from the PKU patients but also that the overall organization of the gene remained unchanged. Comparison of densitometer tracings of the gel lanes containing normal and PKU DNAs has shown that the hybridization signals generated by the PKU DNA samples were not the result of compound heterozygotes with deletions in nonoverlapping regions of the PAH gene in the two alleles present in each cell line. Consequently, it could be concluded that classical PKU, at least in these two cases, is not caused by deletion of the entire PAH gene. Subsequently, this observation has been extended to include several hundred PKU chromosomes. Deletion mutations were found to represent only a very minor fraction of total PKU alleles.

The Molecular Structure of the Human PAH Gene. In order to effectively study the molecular biology of PKU, the structural organization of the PAH gene needs to be established. Southern analysis of human genomic DNA using phPAH247 as the hybridization probe indicated that the chromosomal PAH gene was greater than 65 kb in length and contained multiple intervening sequences (Lidsky et al., 1985b). Due to the large size of this gene, a human genomic DNA library was constructed, using a cosmid vector (pCV107). The library was screened with the PAH cDNA probe, and the corresponding genomic sequences were isolated. Four overlapping PAH cosmid clones, spanning more than 125 kb of the genetic locus, were used for structural analysis of the gene. The structural gene is about 90 kb in length and contains 13 exons, with intron sizes ranging from 1 to 23 kb (DiLella et al., 1986a). The human PAH gene codes for a mature messenger RNA of approximately 2.4 kb and has one of the highest ratios of noncoding to coding DNA found among eukaryotic genes, possibly attributing to its polymorphic nature.

Extensive Restriction Fragment Length Polymorphisms (RFLP) in the Human PAH Locus. Genomic DNAs isolated from 20 random Caucasians were analyzed by Southern hybridization using the full-length human PAH cDNA clone and a battery of restriction enzymes. Enzymes that yielded polymorphic patterns in the PAH locus identified in this manner include *Bgl*II, *Pvu*II, *Eco*RI, *Xmn*I, *Msp*I, *Hin*dIII, and *Eco*RV (Woo et al., 1983; Lidsky et al., 1985b). The frequencies of these RFLP's among Caucasians are such that the observed heterozygosity in the population is about 90%. Using these enzymes to perform RFLP analysis in PKU families, it was possible to demonstrate that the segregation of the PKU alleles and the disease state are concordant. Thus the RFLP analysis of the human PAH locus can be applied for prenatal diagnosis of the hereditary disorder in most PKU families (Woo, 1984; Woo et al., 1984; Speer et al., 1986).

Prenatal Diagnosis of PKU. The polymorphism detected with *Hin*dIII is inherited in a Mendelian fashion and was used for prenatal diagnosis of PKU for the first time since the discovery of PKU by Folling in 1934 (Lidsky et al., 1985c). Both parents in a family at risk were heterozygous for the *Hin*dIII polymorphism containing the 4.2-kb and 4.0-kb alleles. DNA analysis revealed that the affected child in this family inherited the PAH gene containing the 4.0-kb *Hin*dIII fragment from both parents. The mutant genes are therefore associated with the 4.0-kb allele in this family. Analysis of DNA isolated from amniocytes revealed that the fetus was homozygous for the 4.0-kb fragment. The fetus has the same genotype as the affected child and was consequently diagnosed as having PKU. After delivery, this diagnosis was confirmed by the phenotype of the infant.

The extensive polymorphic nature of the human PAH locus permits the use of eight restriction enzymes for prenatal diagnosis of PKU. However, the polymorphisms generated by various enzymes tend to segregate as groups in PKU families. Consequently, it was determined that *Pvu*II, *Xmn*I, and *Eco*RV (detecting RFLP's in the 5′, middle, and 3′ regions of the gene) will establish disease status in about 85% of PKU families at risk and these are the useful enzymes for future prenatal diagnosis of PKU (Daiger et al., 1986; Ledley et al., 1988a).

RFLP Haplotypes at the PAH Locus and Their Association with PKU. The RFLP's identified in the PAH locus were assayed in 33 nuclear PKU families from Denmark. RFLP haplotype analysis of 66 normal chromosomes and 77 chromosomes bearing PKU mutation demonstrated two clusters of RFLP's: (1) *Bgl*II, *Pvu*IIa, and *Pvu*IIb at the 5′ end of the PAH gene and (2) *Eco*RI, *Msp*Ia, *Msp*Ib, *Xmn*I, *Hin*dIII, and *Eco*RV at the 3′ end. Having determined the exact map positions of the RFLP's by structural analysis of the PAH gene (DiLella et al., 1986a), a relationship between the physical distance and RFLP linkage map was established. The RFLP sites within each cluster have a significant tendency to segregate as a group ($p < 0.001$), a process referred to as "linkage disequilibrium". The RFLP sites in the 5′ and 3′ clusters, however, are randomly associated. The 5′ cluster is at significant ($p < 0.05$) linkage disequilibrium with PKU alleles. Furthermore, 93% of the PKU chromosomes and 77% of the normal PAH chromosomes analyzed in this population are of a single 5′ RFLP haplotype configuration, i.e., *Bgl*II, 3.6 kb/*Pvu*IIa, 6.0 kb/*Pvu*IIb, 11.0 kb). The data suggest two possibilities for the nature of PKU mutations in the population. If most PKU genes in the Danish population arose from a single mutation event, the mutation might have occurred on or beyond the 5′ end of the PAH gene. Another possibility is that the PKU genes in this population resulted from multiple independent mutations which occurred on chromosomes of the common normal haplotype background, and these mutations would not be restricted to a particular region of the PAH gene (Chakraborty et al., 1987).

Molecular and Genetic Analyses of Prevalent PKU Alleles. Prenatal diagnosis of PKU by RFLP analysis is applicable only to families with a previously affected child whose RFLP pattern serves as the reference for fetal DNA analyses. Its ability to release the incidence of PKU in a given population is thus limited, since the majority of PKU infants are born to couples without prior family PKU history. However, if PKU is caused by a limited number of mutations in the PAH gene, diagnosis and carrier detection could be achieved by direct analysis of the mutation sites in the gene within the population. We have observed the existence of multiple RFLP haplotypes in the human PAH gene from various European populations and a strong association of PKU alleles among distinct RFLP

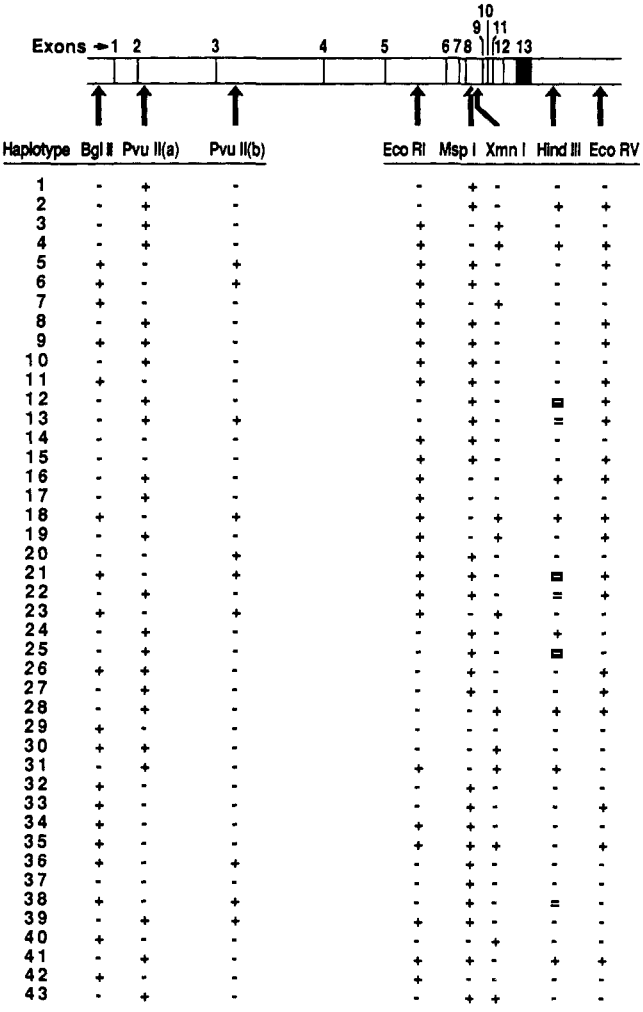

FIGURE 1: RFLP haplotypes at the human PAH locus. The molecular structure of the human PAH gene is shown schematically with its 13 exons encompassing about 90 kb of DNA. The heavy arrows correspond to the polymorphic restriction sites in and immediately flanking the gene. (+) and (-) symbols are used to designate the presence and absence of a polymorphic restriction site, respectively. (=) is used to designate a 4.4-kb *Hin*dIII allele. Contributing PKU centers in Europe include those in Denmark, Scotland, Switzerland, Germany, France, Italy, Hungary, and Czechoslovakia. Reprinted with permission from Woo (1988). Copyright 1988.

haplotypes in that population (Figure 1). Close associations between RFLP haplotypes of the β-globin locus and specific β-thalassemia mutations in different ethnic populations have previously been reported by other laboratories [for review, see Orkin and Kazazian (1984)]. The association of RFLP haplotypes and specific PKU mutations, if any, can be verified by isolation and sequencing of PKU genes. Having established the molecular structure and RFLP linkage map of the PAH gene, this important issue can be addressed by cloning the mutant alleles of the predominant RFLP haplotypes.

The PAH gene of a prevalent PKU haplotype (i.e., haplotype 3) was isolated by molecular cloning. Sequence analysis demonstrated a single base substitution in this gene (G to A) involving the 5′ donor splice site of intron 12 (DiLella et al., 1986b). In order to determine if this mutation is indeed the cause of PKU, we constructed a mini-PAH gene containing the entire PAH cDNA and intron 12, which was engineered behind a strong eukaryotic promoter. When the construct was transferred into cultured mammalian cells, authentic human PAH mRNA and enzymatic activity were detected. When the mutant DNA fragment was used to replace the corresponding normal gene fragment, however, no enzymatic activity could be detected in the transformed cells, while comparable levels of PAH mRNA were produced. These experiments proved that the G to A transition in the beginning of intron 12 of the PAH gene is indeed the cause of PKU and constitutes the first mutant PAH gene ever characterized. In addition, cells transformed with the normal gene accumulate immunoreactive protein in the cytoplasm, while those transformed with the mutant gene do not, suggesting the mutation results in the production of an unstable protein and a CRM^- phenotype (Marvit et al., 1987).

An oligonucleotide specific for the splicing mutation was synthesized and hybridized to cloned PAH DNA in order to test the feasibility of using this probe as a tool for direct analysis of the mutation. The mutant oligonucleotide probe specifically hybridized to a 2-kb *Pvu*II fragment of cloned DNA bearing the mutant donor splice site, while a corresponding normal oligonucleotide probe hybridized with only the normal sequence. Genomic DNAs were isolated from Danish PKU individuals with previously defined haplotypes, cut with *Pvu*II, and analyzed by hybridization using the mutant and normal oligonucleotide probes. These analyses have demonstrated that all haplotype 3 mutant alleles in the Danish population bear the same splicing mutation and none of the haplotype 3 mutant alleles bear that particular mutation (DiLella et al., 1986b). The absolute association between haplotype 3 mutant alleles and the splicing mutation must be the result of a recent mutational event on a haplotype 3 chromosome background, which was spread in the population before there had been sufficient time for transfer of the mutant fragment into chromosomes of other haplotypes by crossovers during meiosis. This is the first demonstration of linkage disequilibrium between a specific mutation and a particular haplotype in PKU.

Because the splicing mutation is not present in mutant alleles of other haplotypes, those alleles must bear other mutations and PKU is, therefore, a heterogeneous disorder at the gene level. This conclusion is confirmatory to the previous observation that the disorder is heterogeneous at the mRNA level, since some PKU patients have no detectable PAH mRNA in the liver, while others have ample hepatic PAH mRNA (DiLella et al., 1985). Indeed, different RFLP haplotypes have been shown to segregate independently in families with both PKU and hyperphenylalaninemic children (Ledley et al., 1986).

Subsequently, a mutant haplotype 2 allele was cloned from a PKU individual. Sequence characterization of the gene showed that there is a C to T transition in exon 12, resulting in the substitution of Arg^{408} to Trp^{408} in the enzyme. Site-directed mutagenesis using specific oligonucleotides was performed in order to create the specific mutant allele in an expression vector. When the normal and mutant constructs were introduced into cultured mammalian cells, only the former produced immunoreactive protein and active enzyme in the cytoplasm, while both produced similar levels of PAH mRNA. Thus, the Arg to Trp substitution is a PKU mutation, which also creates a CRM^- phenotype (DiLella et al., 1987). A specific oligonucleotide corresponding to the mutation sequence in the haplotype 2 allele was synthesized and used to determine if there is any association between mutation and haplotype. Results demonstrated there is a strong linkage disequilibrium, suggesting again that there was a recent mutation event on a normal haplotype 2 chromosome, which was then spread in the population (DiLella et al., 1987).

Most PKU Patients Are Genetic Compounds. Because of the linkage disequilibrium between specific mutations and

RFLP haplotypes in the PAH gene, it became possible to determine the genotypes of PKU patients with regard to specific mutant alleles they bear. Thus PKU patients who are homozygous for mutant haplotypes 2 or 3 must bear the same mutation alleles in both chromosomes, while those bearing one each of the mutant haplotype 2 and 3 alleles would represent compound heterozygotes bearing both mutant alleles. Furthermore, because the splicing mutation and the mis-sense mutation are not represented in the mutant alleles of other haplotypes, patients who are heterozygous for haplotypes 2 or 3 with any of the other haplotypes can also be classified as compound heterozygotes. By this analysis it could be estimated that about three-fourths of the PKU patients in the European population are genetic compounds.

Population Genetics and Evolutionary Origin of PKU. Mutation analysis of PKU by oligonucleotide hybridization was carried out with PKU kindreds from various Caucasian populations representing Western, Southern, and Central Europe. It was discovered that linkage disequilibrium between the splicing mutation and mutant haplotype 3, as well as that between the mis-sense mutation and mutant haplotype 2 alleles, was maintained throughout the European continent. The only exceptions were a few mutant haplotype 3 alleles in Italy that did not contain the splicing mutation (unpublished results). These data clearly demonstrated that the splicing mutation and the mis-sense mutation were the result of recent mutational events on haplotype 3 and 2 chromosome backgrounds, respectively, which were then spread throughout the European continent.

Because haplotype 3 and 2 chromosomes are relatively rare among normal PAH alleles in various European populations, it is puzzling to observe that the mutant alleles associated with these two haplotypes account for about 50% of all PKU alleles throughout Europe. While it is entirely possible for a mutational event to have taken place in a chromosome of rare haplotype by chance, followed by expansion of the mutation allele in the population (the "founder and drift" hypothesis), the probability of two such events involving rare chromosomes occurring in recent evolutionary history is rather unlikely. Alternatively, it is possible that these mutations might have occurred in an ancestral population that had better representation of haplotypes 2 and 3 among the normal PAH alleles, and the mutant alleles were then spread throughout Europe by some sort of mild selective advantage to account for their relatively high frequencies in the Caucasian population (the "heterozygote selection" hypothesis).

Having analyzed multiple polymorphic protein markers in various African, European, and Asian populations, L. Cavalli-Sforza and colleagues formulated a theory of human migration in the recent evolutionary history. They observed that there was a major vectorial human migrational pattern about 10 000–20 000 years ago, originating from today's Middle East areas toward northern Europe along a northwestern direction. They theorized that the migrational pattern was the result of the development of farming technology in the founder population, which made them more competitive than the early native Europeans (Menozzi et al., 1978). To superimpose onto this theory our observations of linkage disequilibrium between haplotypes and mutations in the human PAH gene, it might be suggested that the splicing and mis-sense mutations occcurred in the normal haplotype 3 and 2 PAH genes in the founding population, and the mutant alleles were then spread throughout the European continent with the farmers' migration pattern some 10 000–20 000 years ago. Finally, it may also be concluded that the high frequency of

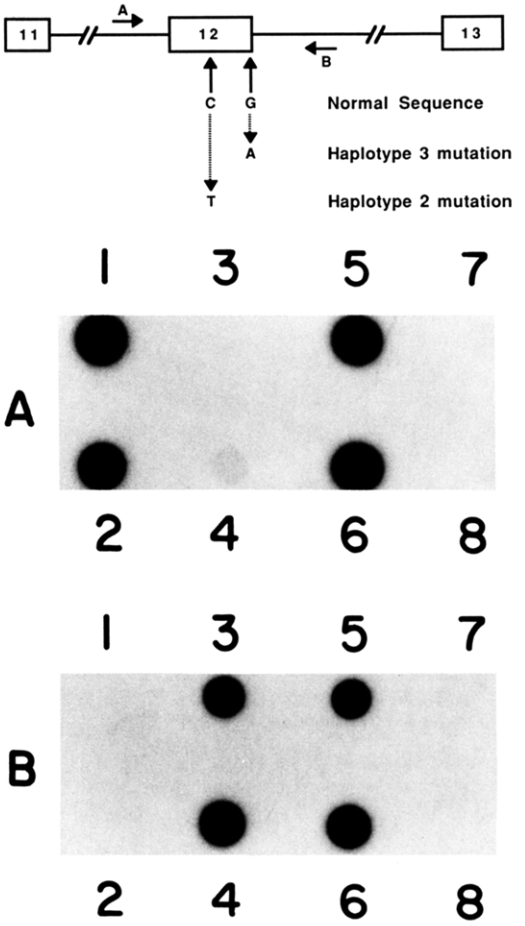

FIGURE 2: Upper panel: Schematic representation of the 245-bp DNA fragment containing exon 12 and flanking intronic sequences of the PAH gene. This segment contains both mutation sites for the mutant PAH alleles associated with haplotypes 2 and 3 that are prevalent among Caucasians of northern European ancestry. Lower panel: Dot-blot analysis of PCR-amplified genomic DNA. Panels A and B are autoradiographs of the membranes after hybridization with the mutant haplotype 2 and 3 oligonucleotide probes, respectively. Dots 1 and 2, two mutant haplotype 2 carriers; dots 3 and 4, two mutant haplotype 3 carriers; dots 5 and 6, two haplotype 2 and 3 compound heterozygotes; dots 7 and 8, two normal individuals. Reprinted with permission from DiLella et al. (1988). Copyright 1988.

PKU is not due to a high mutation rate of the human phenylalanine hydroxylase gene but rather to an expansion of a handful of mutant alleles in the population.

Carrier Detection of PKU by Direct Mutation Analysis. The splicing and mis-sense mutations associated with mutant haplotype 3 and 2 alleles account for roughly 50% of all PKU alleles in Europe. Thus, it is theoretically possible to use the specific mutant oligonucleotides to detect the presence of the corresponding mutant alleles in any given individual without prior family history of PKU. As additional mutant PAH alleles of other haplotypes become characterized at the gene level, they too can be identified in any individual carriers by hybridization with newly synthesized mutant oligonucleotides. Thus, it would be of great interest to determine the mutations in the PAH gene associated with the remaining prevalent haplotypes in order to enhance the accuracy rate to the 90% level, at which point the potential for carrier detection of PKU without a prior family history of the genetic disorder can be realized.

It should be further recognized that the identification of mutant alleles in genomic DNA of random individuals can be readily carried out by using the DNA amplification procedure of "polymerase chain reaction" (Saiki et al., 1985, 1986). By

utilization of two synthetic oligonucleotide primers complimentary to the sequences flanking a specific mutation site, the region of interest can be amplified experimentally by repeated cycles of DNA denaturation and polymerase repair reaction. After 20–30 cycles, the specific DNA region can be amplified in the neighborhood of 10^6-fold. Under these circumstances, the mutant alleles can be readily identified by using specific mutant oligonucleotides by dot-blot analysis. Indeed, identification of the splicing mutation and the mis-sense mutation in the PAH gene can be carried out in this manner (DiLella et al., 1987; Lichter-Konecki et al., 1988). Because the splicing mutation and the mis-sense mutation occur within 200 base pairs in the human PAH gene, a single pair of primers can be used to amplify a segment of genomic human DNA bearing both mutation sites for subsequent analysis (Figure 2A). By use of DNAs isolated from two normal individuals, two splicing mutant allele carriers, two mis-sense mutation allele carriers, and two patients who are compound heterozygotes for both mutant alleles, it was demonstrated that their respective genotypes can be readily established by dot-blot analysis of the enzymatically amplified DNA preparations (Figure 2B). Recently the polymerase chain reaction protocol has been refined in such a manner that only a drop of blood or a single hair is necessary for DNA amplification. The procedure has also been fully automated to analyze 50–100 samples at a time. With these new development and additional mutant PAH gene characterization, it can be expected that carrier detection of PKU in the population without prior PKU history will be technologically feasible in the near future.

Acknowledgments

I am most grateful to all my collaborators in Europe, for without their cooperation, the population genetic analysis of PKU would not have been possible. I thank Brenda Grossie for the typing of the manuscript.

References

Armstrong, M. D., & Low, N. L. (1957) *Proc. Soc. Exp. Biol. Med. 94*, 142–146.

Armstrong, M. D., & Binkley, E. L., Jr. (1956) *Proc. Soc. Exp. Biol. Med. 93*, 418–429.

Berk, K., & Saugstad, L. F. (1974) *Clin. Genet. 6*, 147.

Bickel, H., Gerrard, J., & Hickmans, E. M. (1954) *Acta Paediatr. Scand. 43*, 64–77.

Bickel, H., Beckers, R. G., Wambert, E., et al. (1973) *Acta Paediatr. Scand. 62*, 413–416.

Cabalska, B., Duszynska, N., Borzymowska, J., Zorska, K., Koslacz-Folga, A., & Bozkowa, K. (1977) *Eur. J. Pediatr. 126*, 126–253.

Chakraborty, R., Lidsky, A. S., Daiger, S. P., Güttler, F., Sullivan, S., DiLella, A. G., & Woo, S. L. C. (1987) *Hum. Genet. 76*, 40–46.

Choo, K. H., Filby, R. G., Jenning, I. G., Peterson, G., & Fowler, R. (1986) *DNA 5*, 529.

Dahl, H. H., & Mercer, J. F. B. (1986) *J. Biol. Chem. 261*, 4148.

Daiger, S., Lidsky, A. S., Chakraborty, R., Koch, R., Güttler, F., & Woo, S. L. C. (1986) *Lancet 1*, 229–232.

DiLella, A. G., Ledley, F. D., Rey, F., Munich, A., & Woo, S. L. C. (1985) *Lancet 19*, 160.

DiLella, A. G., Kwok, S. C. M., Ledley, F. D., Marvit, J., & Woo, S. L. C. (1986a) *Biochemistry 25*, 743–749.

DiLella, A. G., Marvit, J., Lidsky, A. S., Güttler, F., & Woo, S. L. C. (1986b) *Nature 322*, 799–803.

DiLella, A. G., Marvit, J., Brayton, K., & Woo, S. L. C. (1987) *Nature 327*, 333–336.

DiLella et al. (1988) *Lancet 1*, 497–499.

Fölling, A. (1934) *Z. Physiol. Chem. 227*, 169–176.

Guthrie, R., & Susi, A. (1963) *Pediatrics 32*, 338–343.

Güttler, F. (1980) *Acta Peadiatr. Scand., Suppl. 280*, 1–80.

Jervis, G. A. (1947) *J. Biol. Chem. 169*, 651–656.

Jervis, G. A. (1953) *Proc. Soc. Exp. Biol. Med. 82*, 514–515.

Kamaryt, J., Mrskos, A., Podhradska, D., Kolcova, V., Cabalska, B., Duczynska, N., & Borzymowska, J. (1978) *Hum. Genet. 43*, 205.

Kaufman, S. (1957) *J. Biol. Chem. 226*, 511–524.

Kaufman, S. (1959) *J. Biol. Chem. 234*, 2677–2682.

Kaufman, S. (1976) *Adv. Neurochem. 2*, 1–132.

Kaufman, S., Holtzman, N. A., Milstien, S., Butler, L. J., & Krumholz, A. (1975) *N. Engl. J. Med. 293*, 785–790.

Kaufman, S., Berlow, S., Summer, G. K., et al. (1978) *N. Engl. J. Med. 229*, 673–679.

Koch, R., Azen, C. G., Friedman, E. G., & Williamson, M. L. (1982) *Pediatrics 70*, 376–380.

Kwok, S. C. M., Ledley, F. D., DiLella, A. G., Robson, K. J. H., & Woo, S. L. C. (1985) *Biochemistry 24*, 556–561.

Ledley, F. D., Grenett, H. E., DiLella, A. G., Kwok, S. C. M., & Woo, S. L. C. (1985) *Science 228*, 77–79.

Ledley, F. D., Levy, H. L., & Woo, S. L. C. (1986) *N. Engl. J. Med. 314*, 1276–1280.

Ledley, F. D., Grenett, H. E., & Woo, S. L. C. (1987) *J. Biol. Chem. 262*, 2228–2233.

Ledley, F. D., Koch, R., Jew, K., Beaudet, A., O'Brien, W. E., Bartos, D. P., & Woo, S. L. C. (1988a) *J. Pediatr. 113*, 463–468.

Ledley, F. D., Ledbetter, S. A., Ledbetter, D. H., & Woo, S. L. C. (1988b) *Cytogenet. Cell Genet. 47*, 125–126.

Leeming, R. J., Blair, J. A., & Rey, F. (1976) *Lancet 1*, 99.

Levy, H. L., & Waisbren, S. E. (1983) *N. Engl. J. Med. 309*, 1269–1274.

Lichter-Konecki, U., Schlotter, M., Konecki, D. S., Labeit, S., Woo, S. L. C., & Trefz, F. K. (1988) *Am. J. Hum. Genet. 78*, 347–352.

Lidsky, A. S., Robson, K., Chandra, T., Barker, P., Ruddle, F., & Woo, S. L. C. (1984) *Am. J. Hum. Genet. 36*, 527–533.

Lidsky, A. S., Law, M. L., Morse, H. G., Kao, F. T., & Woo, S. L. C. (1985a) *Proc. Natl. Acad. Sci. U.S.A. 82*, 6221–6225.

Lidsky, A. S., Ledley, F. D., DiLella, A. G., Kwok, S. C. M., Daiger, S. P., Robson, K. J. H., & Woo, S. L. C. (1985b) *Am. J. Hum. Genet. 37*, 619–634.

Lidsky, A. S., Güttler, F., & Woo, S. L. C. (1985c) *Lancet 1*, 549–551.

Marvit, J., DiLella, A. G., Brayton, K., Ledley, F. D., Robson, K. J. M., & Woo, S. L. C. (1987) *Nucleic Acids Res. 15*, 5613–5628.

Menozzi, P., Piazza, A., & Cavalli-Sforza, L. (1978) *Science 201*, 786–792.

Mitoma, C. (1956) *Arch. Biochem. Biophys. 60*, 476–484.

Orkin, S. H., & Kazazian, H. H. (1984) *Annu. Rev. Genet. 18*, 131–172.

Paul, T. D., Brandt, I. K., Elisa, L. J., Jackson, C. E., Nance, C. S., & Nance, W. E. (1979) *Clin. Genet. 16*, 217.

Penrose, L. S. (1935) *Lancet 2*, 192–194.

Perry, T. L., Hansen, S., Tischler, B., Richards, F. M., & Sokol, M. (1973) *N. Engl. J. Med. 289*, 395–398.

Robson, K. J. H., Chandra, T., MacGillivray, R. T. A., & Woo, S. L. C. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4701–4705.

Robson, K. J. H., Beattie, W., James, R. J., Cotton, R. C. H., Morgan, F. J., & Woo, S. L. C. (1984) *Biochemistry 23*, 5671-5673.
Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., & Arnheim, H. (1985) *Science 230*, 1350-1354.
Saiki, R. K., Bugawan, T. L., Horn, G. T., Mullis, K. B., & Erlich, H. A. (1986) *Nature 324*, 163-166.
Scriver, C. R., & Clow, C. L. (1980a) *N. Engl. J. Med. 303*, 1336-1343.
Scriver, C. R., & Clow, C. L. (1980b) *N. Engl. J. Med. 303*, 1394-1440.
Smith, I., Lobascher, M. E., Stevenson, J. E., Wolff, O. H., Schmidt, H., Grubel-Kaiser, S., & Bickel, H. (1978) *Br. Med. J. 2*, 723-726.
Speer, A., Dahl, H. H., Reiss, D., Lober, G., Hanke, R., Cotton, R. G. H., & Coutelle, C. (1986) *Clin. Genet. 29*, 491.
Woo, S. L. C. (1984) *Pediatrics 74*, 412-423.
Woo, S. L. C. (1988) *Am. J. Hum. Genet. 43*, 781-783.
Woo, S. L. C., Lidsky, A. S. Güttler, F., Chandra, T., & Robson, K. J. H. (1983) *Nature 306*, 151-155.
Woo, S. L. C., Lidsky, A., Chandra, T., Güttler, F., & Robson, K. (1984) *JAMA, J. Am. Med. Assoc. 251*, 1998-2002.

Accelerated Publications

Characterization of a Partly Folded Protein by NMR Methods: Studies on the Molten Globule State of Guinea Pig α-Lactalbumin†

Jean Baum,‡ Christopher M. Dobson,* Philip A. Evans,§ and Claire Hanley
Inorganic Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QR, England
Received August 15, 1988; Revised Manuscript Received October 24, 1988

ABSTRACT: NMR spectroscopy has been used to investigate the structure of a partially folded state of a protein, the molten globule or A-state of α-lactalbumin. The ^{1}H NMR spectrum of this species differs substantially from those of both the native and fully unfolded states, reflecting the intermediate level of order. The resolution in the spectrum is limited by the widespread overlap and substantial line widths of many of the resonances. Methods have therefore been developed that exploit the well-resolved spectrum of the native protein to probe indirectly the A-state. A number of resonances of the A-state have been found to be substantially shifted from their positions in the spectrum of the unfolded state and have been identified through magnetization transfer with the native state, under conditions where the two states are interconverting. The most strongly perturbed residues in the A-state were found to be among those that form a hydrophobic core to the native structure. A number of amides were found to be highly protected from solvent exchange in the A-state. These have been identified through pH-jump experiments, which label them in the spectrum of the native protein. They were found to occur mainly in segments that are helical in the native structure. These results enable a model of the A-state to be proposed in which significant conformational freedom exists but where specific elements of native-like structure are preserved.

The properties of partially organized states of proteins, in which only a subset of the native folding interactions may be present, are of fundamental importance in relation to our understanding of the nature of protein folding. The cooperativity of folding is such that these species often exist only transiently in the course of folding (Tanford, 1968; Kim & Baldwin, 1982), so that their detailed characterization may be feasible only if some form of trapping is possible (Creighton, 1978; Ghelis, 1980; Kim, 1986; Roder, 1988). Recently, however, it has been reported that several proteins can exist, under certain conditions, in partially structured states which are stable at equilibrium (Kuwajima, 1977; Dolgikh et al., 1981; Brazhnikov et al., 1985; Denton et al., 1981; Ohgushi & Wada, 1983; States et al., 1987). A number of such species have been characterized as "molten globule states", and these appear to be compact states exhibiting a high level of secondary structure (Dolgikh et al., 1981, 1985; Ohgushi & Wada, 1983). This description has been established from extensive experimental studies, but in order to understand fully the nature of these species, and particularly the manner in which they are related to native structures, it will be necessary to characterize the behavior of individual residues within a molten globule structure. We have been investigating the application of NMR spectroscopy to this problem in the specific case of the molten globule state of α-lactalbumin.

α-Lactalbumins are globular proteins of molecular weight around 14000 produced in the lactating mammary gland (Hill & Brew, 1975; Hall & Campbell, 1986). They show close sequence homology with c-type lysozymes, and a recent X-ray crystallographic study of baboon α-lactalbumin confirms that in the native state they are also closely similar in conformation (Stuart et al., 1986; Phillips et al., 1987). The α-lactalbumins

† This work was supported by the U.K. Science and Engineering Research Council. J.B. acknowledges receipt of a University of California President's Fellowship and a Fulford Junior Research Fellowship, Somerville College, Oxford. P.A.E. acknowledges an SERC Postdoctoral Fellowship and C.H. a Dee Graduate Scholarship, St. Hugh's College. C.M.D. is a member of the Oxford Centre for Molecular Sciences.
‡ Present address: Department of Chemistry, Rutgers University, New Brunswick, NJ 08903.
§ Present address: Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, England.

0006-2960/89/0428-0007$01.50/0