

## Bases moléculaires de la phénylcétonurie en France : de l'invasion celte à la bataille de Poitiers

La phénylcétonurie est due à une altération du gène de la phénylalanine hydroxylase. De multiples mutations possibles sont à l'origine de l'hétérogénéité clinique de l'affection. Les techniques récentes de génétique moléculaire permettent, à partir de fragments biopsiques, de cloner l'ADN complémentaire de messagers anormaux ; grâce à la technique de l'amplification d'ADN par PCR (*polymerase chain reaction*), il est aussi possible, maintenant, de caractériser des mutations à partir de taches de sang séché de tests de Guthrie.

Stanislas Lyonnet  
Françoise Rey  
Catherine Caillaud  
Véronique Abadie  
Arnold Munnich  
Jean Rey

### ADRESSE

S. Lyonnet : interne des hôpitaux de Paris.  
F. Rey : chargé de recherches à l'Inserm.  
C. Caillaud : docteur ès sciences. V. Abadie : interne des hôpitaux de Paris. A. Munnich : chef de clinique. J. Rey : professeur de pédiatrie. Département de pédiatrie et unité de recherches de génétique médicale, Inserm U. 12, hôpital des Enfants-Malades, 149, rue de Sèvres, 75743 Paris Cedex 15, France.

La phénylcétonurie est la maladie héréditaire du métabolisme des acides aminés la plus fréquente. En rapport avec un déficit en phénylalanine hydroxylase (PAH), elle est responsable d'une arriération mentale sévère mais évitable. La fréquence du gène est de 1/100 en France, de sorte qu'il naît environ 60 à 80 enfants phénylcétonuriques par an dans notre pays. Heureusement, il existe un système performant de dépistage néonatal de cette affection par le test de Guthrie, effectué chez chaque nouveau-né en maternité. Les enfants atteints sont mis aussitôt à un régime restreint en phénylalanine pendant toute la période de croissance rapide du cerveau, c'est-à-dire six ou sept ans. Le dépistage néonatal systématique a permis de mettre en évidence des déficits complets en PAH et des formes variantes d'hyperphénylalaninémie permanente caractérisées par la persistance d'une activité résiduelle de l'enzyme. Il existe en France une très grande hétérogénéité phénotypique des déficits en phénylalanine hydroxylase, du déficit complet sévère au déficit modéré accom-

pagné d'une activité résiduelle non négligeable de l'enzyme. Ce très large spectre de phénotypes cliniques résulte vraisemblablement d'une hétérozygotie composite au locus de la PAH. C'est pour mieux comprendre les bases moléculaires de cette hétérogénéité que nous avons entrepris l'inventaire des mutations dans le gène de PAH en Europe du Sud et en Afrique du Nord.

L'originalité de ce projet tient au recrutement exceptionnel du Département de Pédiatrie de l'hôpital des Enfants-Malades qui a en charge près de 150 enfants phénylcétonuriques et qui entretient des relations extrêmement étroites avec les différents centres de dépistage français et étrangers. Il est ainsi envisageable de mener à bien une étude génétique approfondie de l'ensemble des mutations phénylcétonuriques en France. La cartographie précise du locus PAH devrait permettre de réaliser, à travers cette étude de mutations, une intéressante enquête sur l'origine génétique des populations françaises et sur les voies de propagation des mutations phénylcétonuriques en Europe du Sud. Cette recherche, déjà largement entamée, connaît actuelle-

ment un surcroît d'intérêt avec la découverte de deux mutations responsables de variants de la maladie.

### La phénylalanine hydroxylase et sa « famille » proche

La transformation de la phénylalanine en tyrosine est la première étape du catabolisme de cet acide aminé. Cette transformation dépend d'un système à trois composantes: la PAH, enzyme d'hydroxylation proprement dite, un cofacteur, la tétrahydrobioptérine (BH<sub>4</sub>, synthétisée *de novo* dans le foie à partir du GTP) et la dihydroptéridine réductase (DHPR) qui régénère le cofacteur oxydé au cours de la réaction d'hydroxylation.

La plupart des travaux concernant les caractéristiques biochimiques et structurales de la PAH ont été menés jusqu'en 1985 sur le rat. En effet, l'expression purement hépatique de cette enzyme, la décroissance rapide *post-mortem* de son activité dans le foie et la rareté du matériel de biopsie hépatique humain ont singulièrement compliqué l'étude de la PAH humaine.

C'est la séquence des acides aminés déduite de la séquence nucléotidique de l'ADNc de PAH qui a permis de connaître la structure primaire de l'enzyme humaine [1]. Cette protéine de 51,67 kDa (kilodalton) comporte 451 acides aminés. Il s'agit d'une phosphoprotéine qui contient dans sa région NH<sub>2</sub>-terminale (résidus 11 à 20) un site de phosphorylation (Gly-Arg-Lys-Leu-[P]Ser-Asp-Phe-Gly-Glu-Glu) identique à la séquence de phosphorylation de la PAH de rat à l'exception du premier acide aminé (glycine au lieu de sérine). Des études ultérieures devront montrer si, comme l'enzyme de rat [2], la PAH humaine est le substrat de protéines kinases AMPc-dépendantes (figure 1).

De nombreuses informations ont pu être tirées de comparaisons de séquences peptidiques entre la PAH humaine et la tyrosine hydroxylase de rat (TYH [3, 4]). Des données biochimiques suggéraient déjà que PAH et TYH sont structuralement et fonctionnellement apparentées. Toutes deux sont des hydroxylases d'acides aminés aromatiques, dépen-

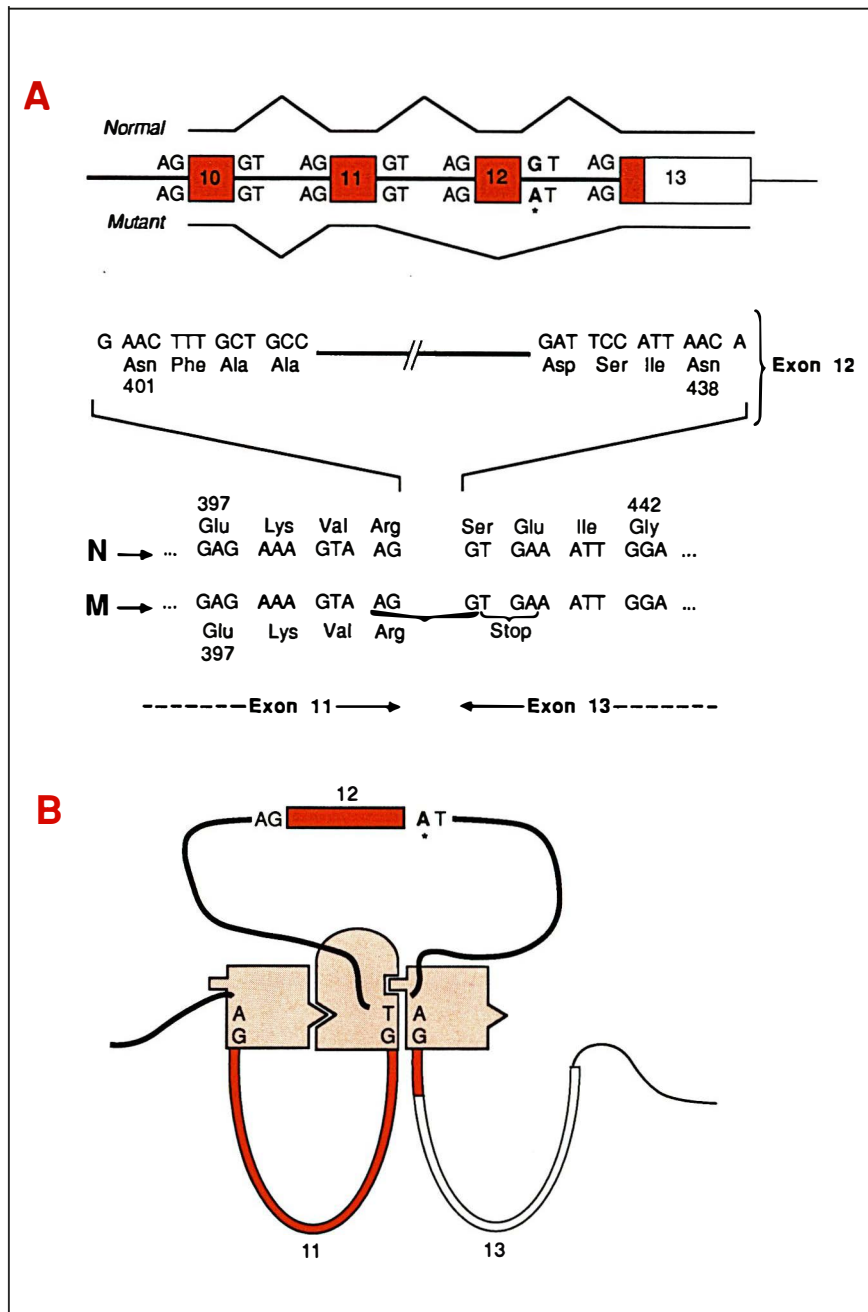


Figure 1. **A.** Représentation schématique des produits de l'excision-épissage normal ou aberrant (mutation GT → AT au site donneur d'épissage de l'intron 12) de l'ARN primitif du gène de PAH. Figurent aussi les séquences de l'ARNm de PAH normal (N) ou mutant (M) qui montrent la « déletion » de l'exon 12 ainsi que le décalage de phase de lecture qui résultent de la mutation intronique. **B.** Modèle pour l'excision anormale de l'exon 12 et de ses introns adjacents. Les « boîtes » représentent les structures hypothétiques impliquées dans le processus d'épissage. La partie codante des exons est en rouge, la partie 3' non codante en blanc. Les introns sont figurés par des traits épais et les séquences intergéniques par des traits minces. L'astérisque en 1-A et 1-B montre le A substitué au G normal (acide adénylique à la place d'un acide guanylique).

## RÉFÉRENCES

1. Kwok SCM, Ledley FD, DiLella AG, Robson KJH, Woo SLC. Nucleotide sequence of a full-length complementary DNA clone and amino acid sequence of human phenylalanine hydroxylase. *Biochemistry* 1985; 24: 446-561.
2. Donlon J, Kaufman S. Relationship between the multiple forms of rat hepatic phenylalanine hydroxylase and degree of phosphorylation. *J Biol Chem* 1980; 253: 2146-52.
3. Ledley FD, DiLella AG, Kwok SCM, Woo SLC. Homology between phenylalanine and tyrosine hydroxylase reveals common structural and functional domains. *Biochemistry* 1985; 24: 3389-94.
4. Grima B, Lamouroux A, Blanot F, Faucon Biguet N, Mallet J. Complete coding sequence of rat tyrosine hydroxylase mRNA. *Proc Natl Acad Sci USA* 1985; 82: 617-21.
5. DiLella AG, Kwok SCM, Ledley FD, Marvit J, Woo SLC. Molecular structure and polymorphic map of the human phenylalanine hydroxylase gene. *Biochemistry* 1986; 25: 743-9.
6. Ledley FD, Grenett HE, DiLella AG, Kwok SCM, Woo SLC. Gene transfer and expression of human phenylalanine hydroxylase. *Science* 1985; 228: 77-9.
7. Ledley FD, Grenett HE, Woo SLC. Biochemical characterization of recombinant human phenylalanine hydroxylase produced in *Escherichia Coli*. *J Biol Chem* 1987; 262: 2228-33.
8. Robson KJH, Chandra T, Mac Gillivray RTA, Woo SLC. Polysome immunoprecipitation of phenylalanine hydroxylase mRNA from rat liver and cloning its cDNA. *Proc Natl Acad Sci USA* 1982; 79: 4701-5.
9. Lidsky AS, Law ML, Morse HG, et al. Regional mapping of the phenylalanine hydroxylase gene and the phenylketonuria locus in the human genome. *Proc Natl Acad Sci USA* 1985; 82: 6621-5.

dantes de la tétrahydrobioptérine. De plus, elles partagent certains substrats, les acides aminés aromatiques pouvant jouer un rôle d'inhibiteur compétitif ou de substrat pour chacune d'elles. Enfin, il existe une réactivité croisée d'anticorps anti-PAH avec la TYH et vice versa.

La comparaison de leurs séquences nucléiques et en acides aminés a confirmé leur grande homologie : 48 % des acides aminés sont identiques et beaucoup de différences sont des substitutions conservatrices produisant seulement des altérations mineures des propriétés biochimiques [3, 4]. Les zones d'homologie entre PAH et TYH sont virtuellement limitées aux deux tiers carboxy-terminaux de la protéine, suggérant la présence dans cette région des fonctions catalytiques communes. A l'inverse, aucune homologie n'a été trouvée dans le tiers aminoterminal des deux protéines, zone qui contient le site phosphorylable. Ainsi, ces enzymes de la famille des hydroxylases semblent-elles contenir deux domaines fonctionnels : (1) un domaine COOH-terminal recouvrant l'essentiel de l'activité catalytique et qui contient tous les déterminants structuraux nécessaires à la réaction d'hydroxylation ; (2) un domaine plus court, NH<sub>2</sub>-terminal, présentant des déterminants qui contribuent majoritairement à la spécificité de substrat et à la régulation post-traductionnelle de l'activité enzymatique.

Il est intéressant de superposer ces derniers éléments aux informations fournies par l'analyse de l'organisation du gène de PAH [5]. Virtuellement, toute la zone d'homologie entre PAH et TYH (c'est-à-dire l'essentiel de l'ARNm, soit 1.698 bp sur 2.448 bp) est codée par les exons 6 à 13 du gène qui, nous le verrons, sont concentrées en 15 kb d'ADN génomique. A l'opposé, aucune homologie entre PAH et TYH n'est trouvée dans les 150 acides aminés NH<sub>2</sub>-terminaux codés par les cinq premiers exons qui comprennent seulement 567 bp mais qui sont disséminés sur 75 kb d'ADN en 5' du gène ! Il est donc possible que les parties aminoterminales de TYH et PAH soient apparues séparément au cours de l'évolution grâce au recrutement d'exons divers, portant des domaines

fonctionnels différents et responsables de la divergence de ces deux protéines (*exon shuffling*)\*. Les grands introns de cette région du gène de PAH pourraient être ainsi les vestiges de cette réorganisation génétique. Par exemple, un intron de 23,5 kb interrompt la séquence codante entre les acides aminés 117 et 118, c'est-à-dire exactement là où débutent les régions homologues des gènes PAH et TYH. Des éléments plus concluants devraient être tirés de l'étude de la structure du gène codant pour la tyrosine hydroxylase humaine.

Il y a eu de nombreuses controverses concernant la structure quaternaire de la PAH, c'est-à-dire l'assemblage de ses sous-unités. La protéine native est multimérique et semble comporter deux sous-unités distinctes. Il pouvait s'agir, entre autres, d'isozymes de PAH ou bien des formes phosphorylées et déphosphorylées d'une même protéine. Désormais, des expériences de transfection de l'ADNc pleine longueur de PAH dans un système d'expression eucaryote ont permis de montrer que toute l'information contenue dans ce seul ADNc de 2,4 kb est suffisante pour produire une protéine aux propriétés semblables à celles de la PAH native humaine [6, 7]. Ainsi, la PAH humaine est-elle bien un dimère constitué de deux sous-unités identiques codées par un seul et même gène et dont la différence de mobilité électrophorétique est due probablement à l'état variable de phosphorylation.

### Un très grand gène pour un transcrit unique

Le gène qui code pour la PAH est maintenant bien connu. L'équipe américaine de S.L.C. Woo en a rapporté la structure dès 1985 [5]. Le clonage d'un ADNc de PAH de rat par immunoprécipitation des polysomes avec un anticorps anti-PAH a été la première étape de ce travail [8]. Cette sonde a permis, grâce à une forte homologie de séquence entre les ADNc de rat et d'homme, d'obtenir par simple criblage d'une banque de foie humain un ADNc de PAH pleine longueur [1].

\* Voir *m/s, suppl. au n° 7, vol. 2, p. 15.*

Secondairement, l'organisation du gène de PAH humaine a été étudiée à partir d'une banque génomique de cosmides [5]. Le gène comporte 13 exons dans un grand domaine de 90 kb, la plupart d'entre eux étant concentrés en son extrémité 3'-terminale (8 exons dans une zone de 15 kb). Cette répartition inégale a peut-être, nous l'avons vu, une signification en terme d'évolution génétique. L'exon le plus long (exon 13, 892 bp) est le dernier ; il contient une petite fraction codante COOH-terminale et toute l'extrémité non codante 3'-terminale. La taille des introns est très variable, de 1 à 24 kb, et le rapport « séquences codantes/introns » est l'un des plus bas connu (2,4/85 kb) parmi les gènes eucaryotes.

L'analyse par *Southern blot* des clones génomiques de PAH et la localisation du gène sur le chromosome 12 grâce à des hybrides somatiques indiquent qu'il n'existe qu'une seule copie de ce gène et qu'il n'y a pas de pseudogène PAH dans le génome humain. L'hybridation *in situ* d'une sonde d'ADNc de PAH sur caryotype métaphasique a permis sa localisation plus précise en 12q22-24 [9].

Le gène PAH est transcrit en un ARN messager de 2,4 kb. La région codante de cet ARN messager est un cadre ouvert de lecture ininterrompu

de 1.356 bp débutant au premier AUG, en position 223, et finissant au codon stop UAA en position 1.579. Il code pour un peptide de 451 acides aminés (si l'on exclut la méthionine initiatrice).

La région 5'-non codante du premier exon est longue de 222 bp. Dans la description initiale de S.C.M. Kwok *et al.* [1], elle comporterait en position 105-114 une séquence oligonucléotidique (UCGUUACCGC) présentant une homologie partielle (80 %) avec l'extension 3'-terminale de l'ARN ribosomal de 18S. Elle est interprétée comme un possible site de reconnaissance pour la liaison du transcrit mature avec cet ARN ribosomal durant la traduction. Cette même région 5'-non codante contiendrait aussi deux séquences répétées inversées longues de 13 bp, situées en position 44-56 et 108-120 et présentant une homologie-inverse de 77 %. La structure en « épingle à cheveux » qui pourrait ainsi être réalisée dans l'extrémité 5' de l'ARNm serait impliquée dans sa stabilité post-transcriptionnelle.

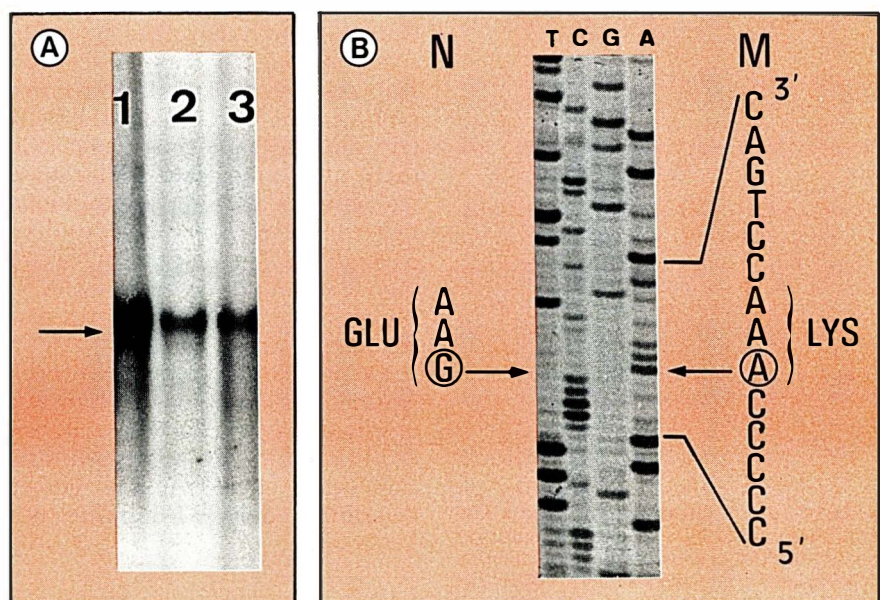
La région 3'-non codante est sans originalité. Longue de 851 bp, elle contient de multiples codons stop de traduction dans toutes les phases de lecture. Elle présente trois séquences consensus de polyadénylation (AATAAA) en position 2.198, 2.347 et 2.413. La distance entre la dernière

de ces séquences et la queue polyadénylée est de 13 bp et c'est elle qui semble être exclusivement utilisée dans le processus normal de coupure nucléolytique-polyadénylation comme le montre, en *Northern blot*, l'existence d'un seul ARNm (figure 2a [1]).

### Hétérogénéité génétique de la maladie et polymorphismes du gène

On connaît plusieurs déficits sur la voie du catabolisme de la phénylalanine. Les plus fréquents — plus de 98 % — sont liés à une anomalie de la PAH elle-même ; nous nous limiterons à ces déficits en PAH responsables de la phénylcétonurie (PKU), maladie récessive autosomique, jadis à l'origine de l'idiotie phénylpyruvique. Schématiquement, deux phénotypes cliniques peuvent être distingués [10, 11] : (a) la phénylcétonurie typique, caractérisée par une tolérance alimentaire de phénylalanine de l'ordre de 250 mg/jour et une activité *in vitro* de la PAH inférieure à 1 % par rapport aux témoins ; (b) les variants, définis par une tolérance supérieure à 500 mg/jour de phénylalanine et une activité résiduelle de l'enzyme allant de 2 à 6 % des témoins. Parmi ceux-là, on individualise parfois la forme « méditerranéenne » d'hyperphény-

Figure 2. **A. Analyse en Northern blot de l'ARNm du gène de PAH.** Cinq microgrammes d'ARN de foie humain normal poly A<sup>+</sup> (canal 1) ou total (canal 2) ont été déposés ainsi que cinq microgrammes d'ARN total de foie extraits d'un fragment de biopsie hépatique d'un patient homozygote pour l'haplotype 38 du gène et porteur d'une forme variante de phénylcétonurie (mutation « arabe ») (canal 3). L'hybridation a été faite avec une sonde ADNc de PAH pleine longueur. La flèche indique l'ARNm unique de PAH de 2,4 kb. **B. Analyse de la séquence des codons 278 à 282 de l'exon 7 du gène PAH du patient étudié** (mutation 280<sup>glu</sup> - <sup>lys</sup>). N = séquence normale ; M = séquence mutante. Les nucléotides normal et mutant au codon 280 sont cerclés (séquençage par la méthode de Sanger).



## RÉFÉRENCES

10. Rey F, Munnich A, Lyonnet S, Rey J. Classification et hétérogénéité des hyperphénylalaninémies liées à un déficit en phénylalanine hydroxylase. *Arch Fr Pediatr* 1987; 44: 639-42.
11. Güttler F. Hyperphenylalaninemia: diagnosis and classification of the various types of phenylalanine hydroxylase deficiency in childhood. *Acta Paediatr Scand* 1980; 280 (suppl): 1-80.
12. Woo SLC, Lidsky AS, Güttler F, Chandra T, Robson KJH. Cloned human phenylalanine hydroxylase gene allows prenatal diagnosis and carrier detection of classical phenylketonuria. *Nature* 1983; 306: 151-5.
13. Lidsky AS, Ledley FD, DiLella AG, et al. Extensive restriction site polymorphism at the human phenylalanine hydroxylase locus and application in prenatal diagnosis of phenylketonuria. *Am J Hum Genet* 1985; 37: 619-34.
14. Rey F, Berthelon M, Caillaud C, et al. Clinical and molecular heterogeneity of phenylalanine hydroxylase deficiencies in France. *Am J Hum Genet* 1988; in press.
15. DiLella AG, Marvit J, Lidsky AS, Güttler F, Woo SLC. Tight linkage between a splicing mutation and a specific DNA haplotype in phenylketonuria. *Nature* 1986; 322: 799-803.
16. Marvit J, DiLella AG, Brayton L, Ledley FD, Robson KJH, Woo SLC. GT to AT transition at a splice donor site causes skipping of the preceding exon in phenylketonuria. *Nucleic Acids Res* 1987; 15: 5613-28.
17. DiLella AG, Marvit J, Brayton K, Woo SLC. An amino-acid substitution involved in phenylketonuria is in linkage disequilibrium with DNA haplotype 2. *Nature* 1987; 327: 333-6.
18. Lichter-Konecki U, Konecki DS, DiLella AG, et al. Phenylalanine hydroxylase deficiency caused by a single base substitution in an exon of the human phenylalanine hydroxylase gene. *Biochemistry* 1988; 27: 2881-5.
- lalaninémie avec une activité résiduelle de l'ordre de 1 à 3 % mais caractérisée surtout par l'origine géographique des patients porteurs de ce phénotype variant. Utilisant une sonde ADNc pleine longueur, le groupe de S.L.C. Woo a conduit une étude complète des haplotypes de restriction au locus PAH. Ce gène est très polymorphe et huit RFLP\* principaux ont été décrits [12]. Il s'agit le plus souvent de polymorphismes de substitution, deux enzymes engendrant un système bi-allélique (PvuII et MspI). L'enzyme HindIII révèle un polymorphisme insertionnel à trois allèles. Près de 90 % de la population caucasienne est hétérozygote pour au moins un RFLP. Les RFLP de 66 allèles normaux et 66 allèles mutants dans 33 familles danoises ont été déterminés et 12 haplotypes ont été observés (Tableau I [13]). Bien que l'on retrouve neuf haplotypes différents parmi les allèles mutants, plus de 90 % d'entre eux sont associés à quatre haplotypes principaux (1 à 4) et près de 38 % d'entre eux sont porteurs de l'haplotype 3 qui est relativement rare chez les allèles normaux (3 %). Cet haplotype 3 est en déséquilibre de liaison avec une mutation ponctuelle au site donneur d'épissage de l'intron 12. De la même manière, l'haplotype 2 (présent dans 20 % des allèles mutants et seulement 5 % des allèles normaux de cette population) est associé à une mutation ponctuelle au codon 408 de l'ARNm. Nous reviendrons sur ces deux mutations mais il faut d'emblée remarquer qu'elles sont, toutes deux, responsables de phénylcétonurie typique avec une activité résiduelle enzymatique inférieure à 1 %. Ces deux mutations étant globalement responsables de 60 % des cas de phénylcétonurie en Europe du Nord, on conçoit dès lors fort bien la très grande similitude des phénotypes cliniques et biologiques observés dans cette population. Lors de l'étude menée par notre équipe dans la population française, originaire à la fois du Nord et du Sud de l'Europe, les haplotypes de restriction de 74 allèles mutants ont été déterminés dans 37 familles non
- apparentées (Tableau I [14]). Un total de 23 haplotypes a été observé: 18 pour les allèles normaux et 16 pour les allèles mutants. Huit haplotypes mutants n'avaient pas été observés jusqu'à présent dans les pedigrees caucasiens; ils sont associés à 20 % des gènes mutants. En dépit de ces nombreux haplotypes, plus de 80 % des allèles mutants sont confinés en six haplotypes: 1 à 4, 9 et 38 (cette numérotation arbitraire ne tient pas compte du nombre d'haplotypes effectivement décrits). Les haplotypes 1 à 4 représentent 66 % de ces allèles mutants au lieu de 50 % des allèles normaux ( $p = 0,05$ ); mais il est remarquable que ces mêmes haplotypes recouvrent 90 % de tous les allèles mutants au Danemark ( $p < 0,001$ ). Les fréquences individuelles des différents haplotypes au locus PAH sont grossièrement les mêmes parmi les allèles mutants et les allèles normaux à l'exception des haplotypes 2 (17,6 % contre 5,9 %,  $p < 0,05$ ) et 38 (8,1 % contre 0 %,  $p < 0,05$ ). Cependant, une distribution différente apparaît quand les allèles mutants sont séparés en allèles « typiques » et en « variants ». Plusieurs haplotypes se rencontrent exclusivement dans des formes typiques (comme les haplotypes 1 ou 3), d'autres uniquement chez des variants (100 % des allèles mutants porteurs de l'haplotype 38,  $p < 0,001$ ), alors qu'enfin, certains haplotypes sont présents dans les deux catégories de malades (Tableau I). C'est le cas pour l'haplotype 2, porteur (contrairement à ce qui est trouvé au Danemark) de différentes mutations aboutissant à des phénotypes soit typiques, soit de variants modérés. Ajoutons que le grand polymorphisme du gène PAH permet de proposer un diagnostic prénatal par analyse de l'ADN dans les familles qui en font la demande et pour lesquelles on a pu montrer, préalablement à la grossesse, le caractère informatif d'au moins un système RFLP au locus PAH [13].

### Les mutations celtiques responsables de phénylcétonurie typique

Depuis deux ans, trois mutations responsables de PKU typique ont été rapportées. S.L.C. Woo et al. ont

\* Voir m/s, suppl. au n° 7, vol. 3, p. 4.

Tableau I

DISTRIBUTION DES HAPLOTYPES RFLP AU LOCUS PAH DANS LA POPULATION DANOISE (COLONNES 1 ET 2 [13]) ET LA POPULATION FRANÇAISE (COLONNES 3 A 6 [14])

Les valeurs sont exprimées en % de n. Seuls les haplotypes portés par plus de 6 % des gènes normaux ou mutants dans une population ont été représentés ; ainsi n'apparaissent que 6 des 12 haplotypes et 8 des 23 haplotypes rapportés respectivement par DiLella *et al.* [13] et Rey *et al.* [14]. Les colonnes 5 et 6 donnent la répartition pour chaque haplotype entre allèles « typiques » et « variants » dans la population française étudiée. Cette analyse montre que certains haplotypes en France sont liés exclusivement à des formes typiques de PKU (3, 7 ou 16), certains (haplotype 38) seulement à des « variants », d'autres (1, 2 ou 4) aux deux phénotypes.

Haplotype	① N (n = 66)	② M (n = 66)	③ N (n = 68)	④ M (n = 74)	⑤ PKU (n = 51)	⑥ Variants (n = 23)
1	34,8	18,2	26,5	31,1	39,2	13,0
2	4,6	19,7	5,9	17,6	11,8	30,4
3	3	37,9	2,9	9,5	13,7	0
4	31,8	13,6	14,7	8,1	3,9	17,4
5	10,6	0	8,8	0	0	0
7	10,6	1,5	11,8	4,1	5,9	0
16	0	0	5,9	1,4	2	0
38	0	0	0	8,1	0	26,1

N = allèles normaux ; M = allèles mutants ; n = nombre total d'allèles étudiés.

décrit la première d'entre elles en 1986 [15]. Elle a été retrouvée associée à un haplotype particulier du gène — l'haplotype 3 — qui, nous l'avons vu, est en fort déséquilibre de liaison avec la maladie chez les Danois et plus généralement dans la population celte d'Europe du Nord. Une banque génomique de l'ADN d'un malade homozygote pour l'haplotype 3 a été criblée avec une sonde d'ADNc pleine longueur de PAH et les fragments contenant les zones transcrites du gène ont été séquencés. La séquence nucléotidique de ces exons est identique à la séquence normale et la seule modification identifiée siège sur la première base de l'intron 12 du gène PAH où une guanine est remplacée par une adénine. En dépit de sa situation intronique, cette mutation est hautement délétère puisqu'elle affecte le processus normal d'excision de l'intron 12 au cours de la maturation du transcrit primitif du gène. Elle altère, en effet, la séquence consensus « donneuse » d'épissage, GT, transformée en AT. Ce signal indispensable disparu, un processus d'excision-

épissage anormal se fait entre l'exon 11 et l'exon 13 qui emporte l'exon 12 avec les introns adjacents (*figure 1*)\*. L'ARNm de PAH est donc porteur d'une « délétion » complète de l'information contenue dans l'exon 12. De surcroît, l'épissage aberrant exon 11-exon 13 entraîne un décalage de la phase normale de lecture, ce qui fait apparaître au tout début de l'exon 13 un codon stop entraînant un arrêt de traduction. On conçoit bien que l'hydroxylase mutante produite soit totalement inactive, car elle est ainsi tronquée de 52 acides aminés dans sa partie COOH-terminale à la suite de l'excision de l'exon 12 et de la non-traduction de l'exon 13. Deux types de constatations ont permis de confirmer définitivement la responsabilité de cette mutation intronique dans l'inactivité de l'hydroxylase mutante [16]. Tout d'abord l'exon 12 est bien absent des ARNm de PAH mutants lors d'une expé-

\* Voir *m/s*, suppl. au n° 7, vol. 2, p. 4 et suppl. au n° 7, vol. 3, p. 8.

rience de protection à la nucléase S1. Enfin, un ADNc de l'ARNm mutant a été sous-cloné dans un vecteur d'expression sous le contrôle d'un promoteur fort. La transfection de cette construction dans un système cellulaire eucaryote entraîne la production d'une enzyme incapable de métaboliser la phénylalanine-C14 en tyrosine-C14 [16].

Cette mutation GT → AT (intron 12) a été recherchée chez de nombreux patients phénylcétonuriques et s'est trouvée dans tous les cas (23 gènes) associée à l'haplotype 3 du gène de PAH suggérant un « effet fondateur » pour la diffusion de cette mutation en Europe du Nord. Elle est globalement responsable d'environ 40 % des allèles mutés dans cette population [15].

Une deuxième mutation responsable du même phénotype a été rapportée par la même équipe en 1987 [17]. La méthodologie employée a été identique (clonage génomique et séquençage exon par exon) et a permis de découvrir une transition C → T sur la première base du codon 408 de l'exon 12 résultant en une substitu-

tion ponctuelle arginine (CGG) → tryptophane (TGG). Là encore, des expériences de transfection d'un ADNc porteur de cette mutation, dans une lignée cellulaire de mammifère en culture (cellules COS de rein de singe) ont montré l'inactivité du produit protéique mutant. Cette mutation, 408<sup>(Arg → Trp)</sup>, est en fort déséquilibre de liaison avec l'haplotype 2 du gène PAH. Elle a été retrouvée chez 20 % des phénylcétonuriques du Danemark, tous porteurs de l'haplotype 2 du gène. Ainsi, chez les Caucasiens d'Europe du Nord, ces deux mutations liées respectivement aux haplotypes 3 et 2 du gène PAH sont responsables de 60 % des gènes mutants.

Une autre mutation ponctuelle a été récemment décrite chez des patients allemands porteurs d'un fragment de restriction anormal engendré par l'enzyme MspI [18]. Elle entraîne le remplacement d'un codon leucine (CTG) par un codon proline (CCG) en position 311. Cette transition T → C dans l'exon 9 du gène crée de plus un nouveau site de restriction pour l'enzyme MspI. La mutation 311<sup>(Leu → Pro)</sup> produit une enzyme extrêmement instable qui est totalement inactive et ne donne aucun matériel à réaction croisée immunologiquement détectable (CRM-).

### Les cartes de Guthrie pour refaire l'histoire d'une mutation arabe

Nous venons de voir que, jusqu'à présent, seule la PKU « typique » avait été étudiée du point de vue moléculaire. Notre intérêt pour les formes atypiques vient naturellement du recrutement propre à la population française. Il nous a conduit à nous intéresser en priorité à la mutation associée à l'haplotype 38 et, d'autre part, à l'hétérogénéité moléculaire en Europe du Sud de l'haplotype 2, si univoque au Danemark.

Nous avons eu l'opportunité d'avoir en charge un enfant porteur d'un déficit en PAH, né de parents cousins germains, homozygote pour l'haplotype 38. Il est né en Algérie où aucun programme de dépistage systématique des hyperphénylalaninémies n'existe encore. A 13 mois, il était atteint d'un retard mental très sévère

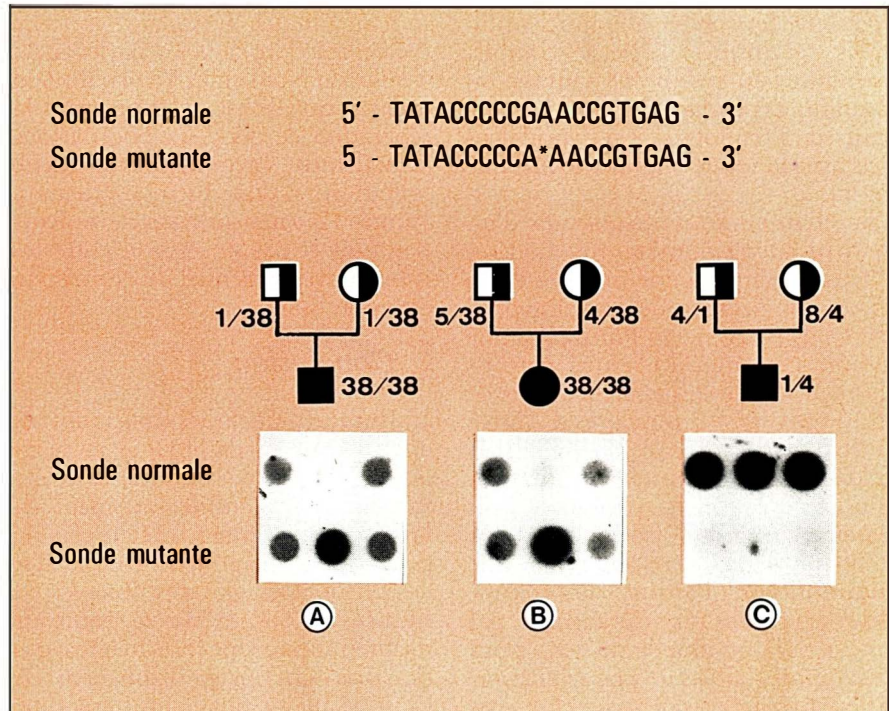
avec convulsions et microcéphalie. Ce patient, né d'un mariage consanguin et porteur à l'état homozygote de l'haplotype 38, a reçu, de ce fait, un seul et même gène mutant de ses deux parents. Dans l'hypothèse où le gène mutant est exprimé, un seul et même produit enzymatique défectueux ainsi qu'un ARN messager unique devraient être présents dans son foie [19]. Son étude directe devait donc éviter un fastidieux clonage et séquençage génomique. Une biopsie du foie à l'aiguille a été réalisée. Elle a permis de confirmer le déficit, de montrer l'existence d'une activité résiduelle de 2-3 % (forme variante) et de mener l'analyse moléculaire sur le fragment restant (20 mg). L'extraction immédiate des ARN de ce fragment biopsique a fourni 60 µg d'ARN totaux de bonne qualité [20]. Leur analyse par Northern-blot a montré la présence d'un ARNm de PAH de taille et d'abondance normales (*figure 2A*). Une banque d'ADNc a été construite à partir d'ADN poly A<sup>+</sup> purifiés. Le criblage de cette banque de 120 000 recombinants par une séquence normale de PAH a permis d'obtenir sept clones d'ADNc mutants dont deux clones représentatifs de la pleine longueur de l'ARNm. L'analyse de leurs séquences a montré une transition G → A au nucléotide 1065 de l'ARNm conduisant à la substitution glutamate → lysine au codon 280, dans l'exon 7 du gène (*figure 2* [21]). Bien que ce changement glutamate → lysine altère fortement la charge électrique de la protéine dans un domaine très conservé parmi différentes hydroxylases et même entre diverses espèces [3], le produit du gène mutant garde une certaine activité résiduelle. Cette transition G → A au codon 280 est compatible avec la désamination d'une 5-méthyl cytosine au sein d'un doublet CpG [22], la mutation ayant porté sur le brin non codant\*. Un tel mécanisme a déjà été évoqué par exemple à l'origine d'une mutation responsa-

\* Les méthylcytosines sont sensibles à une désamination qui les transforme en thymine. Après un cycle de réplication de l'ADN, une paire appariée C-G sera donc transformée sur l'un des doubles brins-fils en T-A.

## RÉFÉRENCES

- DiLella AG, Ledley FD, Rey F, Munnich A, Woo SLC. Detection of phenylalanine hydroxylase messenger RNA in liver biopsy samples from patients with phenylketonuria. *Lancet* 1985 ; i : 160-1.
- Munnich A, Daegelen D, Besmond C, Marie J, Dreyfus JC, Kahn A. Celle-free translation of messenger RNAs from human muscle biopsies : a miniaturized tool for investigation of neuromuscular diseases. *Pediatr Res* 1982 ; 16 : 335-9.
- Lyonnet S, Caillaud C, Rey F, *et al.* Molecular genetics of phenylketonuria in Mediterranean countries : characterization of the first mutation responsible for partial phenylalanine hydroxylase deficiency. (Soumis pour publication).
- Kazazian HH, Antonorakis SE, Youssoufian H, *et al.* Comparison of deficiency alleles of the  $\beta$ -globin and factor VIII:C genes : new lessons from a giant gene. *Cold Spring Harbor Symposia on Quantitative Biology* 1986 ; 51 : 371-9.
- Gitshier J, Wood WI, Shuman MA, Lawn RM. Identification of a missense mutation in the factor VIII gene of a mild hemophilic. *Science* 196 ; 232 : 1415-6.
- Saiki RK, Scharf S, Faloona F, *et al.* Enzymatic amplification of  $\beta$ -globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 1985 ; 230 : 1350-4.
- Lyonnet S, Caillaud C, Rey F, *et al.* Guthrie cards for detection of point mutations in phenylketonuria. *Lancet* 1988 ; ii : 507.
- Werner KF. Les origines. In : Favier J, ed. *Histoire de France. Vol. 1.* Paris : Fayard, 1984.
- Avigad S, Cohen BE, Woo SLC, Shiloh Y. A specific deletion within the phenylalanine hydroxylase gene is common to most Yeminite Jewish phenylketonuria patients. *Am J Hum Genet* 1987 ; 41 (suppl) : A 205.

Figure 3. **Analyse du génotype de trois familles nucléaires phénylcétonuriques sans histoire de consanguinité par amplification de l'ADN (méthode PCR) et hybridation avec un couple de sondes oligonucléotidiques normale ou mutante, pour la mutation 280<sup>glu</sup>-lys.** Cette figure présente deux familles où le patient est homozygote pour l'haplotype 38 du gène PAH, originaires d'Afrique du Nord (famille A) ou d'Autun, Saône-et-Loire, (famille B) et une troisième famille, originaire d'Afrique du Nord, où le propositus est porteur d'autres haplotypes du gène PAH (famille C). Les patients des familles A et B sont homozygotes pour la mutation (leur ADN amplifié n'hybride qu'avec la sonde mutante). Le patient C est porteur d'autres mutations au locus PAH.



ble d'une forme modérée d'hémophilie A [23].

Cette mutation a été recherchée parmi 72 gènes mutants et a été exclusivement retrouvée chez des patients atteints d'une forme variée de déficit en hydroxylase. Cette recherche a été menée par hybridation entre des sondes oligonucléotidiques (normales ou mutées) et l'ADN des patients amplifié par méthode PCR (*polymerase chain reaction* [24])\*\*. La sonde mutante n'a hybridé qu'avec l'ADN amplifié de six patients, tous porteurs de l'haplotype 38 du gène. Aucune hybridation n'a été détectée avec des allèles mutants d'autres haplotypes, ce qui suggère un fort déséquilibre de liaison entre la mutation 280<sup>(glu - lys)</sup> et l'haplotype 38 du gène de PAH [21]. Cette mutation a été retrouvée dans trois familles non apparentées, deux originaires d'Afrique du Nord et une dans une famille de Bourgogne (Autun) dont l'origine française a pu être formellement établie en remontant jusqu'en 1632 (figure 3). Pour apprécier la fréquence de cette

mutation dans la population d'Europe du Sud et d'Afrique du Nord, 33 autres patients (toutes formes confondues) ont été étudiés grâce à l'amplification de leur ADN extrait de leurs cartes de Guthrie souvent expédiées par la poste depuis le Maghreb [25]. Quatre taches de sang sont habituellement déposées sur ces buvards à la naissance et, en général, deux demeurent utilisables après le test de Guthrie. Approximativement 500 ng d'ADN ont pu être extraits d'une seule tache séchée (équivalente de 50 µl de sang frais), les plus vieux buvards utilisés ayant plus de deux ans. Ce « dépistage moléculaire de masse » a permis de découvrir deux nouveaux homozygotes pour cette mutation et deux hétérozygotes composites, tous d'origine arabe.

Du fait que le codon 280 est un point chaud de mutation, la mutation « arabe » et la mutation identique affectant cette vieille famille française de Bourgogne du Sud (figure 3, famille B) ont pu survenir indépendamment. A l'inverse, la mutation 280<sup>(glu - lys)</sup> a pu naître il y a très longtemps et être introduite en Afrique du Nord et en Europe du Sud par les envahisseurs sarrasins au VIII<sup>e</sup> siècle de notre ère. Cette dernière

hypothèse est la plus probable du fait que tous les patients rapportés sont porteurs du même haplotype de restriction au locus PAH. Il convient cependant de préciser que cette mutation au codon 280 ne suffit pas à créer l'haplotype 38 du gène PAH bien que ce dernier n'ait été trouvé chez aucun des gènes normaux de PAH étudiés jusqu'à présent. Enfin, l'absence de cet haplotype en Europe du Nord (et même au-dessus d'Autun...) peut être expliquée par les coups d'arrêt donnés à l'invasion sarrasine en France. Son axe principal, traversant l'Espagne, a été barré à Poitiers en 732 par Charles Martel. L'autre poussée a été précisément refoulée à Autun en 725 [26].

### Hétérogénéité de l'haplotype 2 en France : « la mutation des pêcheurs de morue » ?

Au Danemark, nous l'avons vu, l'haplotype 2 est en fort déséquilibre de liaison avec la mutation ponctuelle 408<sup>Arg - Trp</sup> [17]. De fait, nous avons pu mettre en évidence cette même mutation en France, chez des patients porteurs de PKU typique en

\*\* Voir *m/s*, suppl. au n° 7, vol. 2, p. 13.



utilisant des sondes oligonucléotidiques spécifiques d'allèles, complémentaires du codon 408 sauvage ou mutant. Cependant, les variants porteurs de l'haplotype 2 n'hybridaient pas avec la sonde oligonucléotidique mutante.

Par chance, nous disposons d'un patient atteint d'une forme modérée d'hyperphénylalaninémie, porteur à l'état homozygote de l'haplotype 2 du locus PAH, né au Portugal d'une union consanguine entre cousins germains. Une stratégie identique de clonage et séquençage d'ADNc mutant a été employée. Elle a permis d'identifier une mutation inattendue: la délétion en phase des trois bases d'un codon isoleucine. Cette délétion singulière entraîne une diminution de l'affinité de l'enzyme pour son substrat, caractérisée par une activité résiduelle de 4-5 % pour des apports faibles de phénylalanine. Elle est actuellement recherchée chez des malades portugais en utilisant leurs cartes de Guthrie expédiées par la poste grâce à nos collègues généticiens de Lisbonne (il s'agit de la première collaboration européenne concernant la génétique moléculaire de la PKU).

L'existence en Bretagne du Sud d'un petit « isolat » d'allèles mutants porteurs de l'haplotype 2 et n'ayant pas la mutation liée à cet haplotype décrite en Europe du Nord par S.L.C. Woo [17] est aussi actuellement en cours d'étude. La découverte de cette même mutation en Bretagne pourrait indiquer son trajet génétique le long de la côte Atlantique, « au gré des escales des marins pêcheurs ».

### La délétion yéménite

Pour être complet, signalons enfin l'existence d'une délétion de l'exon 3 du gène PAH, retrouvée systématiquement chez les Juifs Yéménites atteints de PKU typique. Elle est strictement confinée à cet isolat génétique. On ne connaît encore ni les bornes de cette délétion, ni son mécanisme moléculaire. Comme pour les mutations celtiques, la fréquence et le caractère univoque d'une mutation confinée à une population restreinte posent la question d'un effet fondateur ou d'un avantage sélectif accordé aux hétérozygotes par un allèle

mutant unique. L'existence de cette même délétion chez les patients du Yémen du Nord et du Yémen du Sud indique qu'elle est apparue avant la séparation de ces deux populations actuellement bien distinctes. Cette mutation est donc bien antérieure à l'année 1750 marquée par de violents pogroms et par la scission entre les tribus juives yéménites du Nord et du Sud.

### Conclusion

L'afflux en quelques mois de ces données sur la pathologie moléculaire de la phénylcétonurie et de ses variants usuels milite fortement en faveur d'une grande hétérogénéité génétique à l'origine du continuum des phénotypes mutants observés en Europe du Sud et sur le pourtour méditerranéen. Une forte hétérozygotie composite est probablement en cause, d'autant qu'aucun haplotype ne prédomine au sein de la cohorte des gènes mutants étudiés en France. Ce fait oppose enfin nos populations d'Europe du Sud à celles d'Europe du Nord, beaucoup plus homogènes, chez lesquelles quatre haplotypes sont associés à 90 % des gènes mutants et deux mutations à 60 % d'entre eux.

Nous montrons, par ailleurs, qu'il est possible de mener une large enquête génétique à la recherche d'une mutation particulière grâce à l'ADN extrait de cartes de Guthrie même anciennes. Cela suggère la possibilité de l'utilisation plus élargie de ces cartes de Guthrie pour l'amplification de l'ADN dans le cadre de la détection rétrospective voire prospective de n'importe quelle autre mutation du génome humain ■

### Remerciements

Le dépistage néonatal de la phénylcétonurie en France est organisé par l'Association Française pour le Dépistage et la Prévention des Maladies Métaboliques et des Handicaps de l'Enfant. Nous remercions Monique Poussière pour son travail sur le manuscrit.

### TIRÉS A PART

S. Lyonnet.

### Summary

Phenylketonuria (PKU) is an autosomal recessive human genetic disease caused by a deficiency in a liver-specific enzyme, phenylalanine hydroxylase (PAH). The incidence of the disease in Caucasians (1/17,000) and its severe consequences for brain development in untreated children warrant PKU screening in newborns using the Guthrie test. Mass-neonatal screening has led to the discovery of variants forms of the disease characterized by the persistence of a residual PAH activity. The restriction fragment length polymorphisms (RFLP) of normal and mutant alleles at the PAH locus were determined in French kindreds and the number of RFLP haplotypes observed is much higher than in the Northern European population previously reported. This feature illustrates the particular heterogeneity of the Southern European population. Several mutant haplotypes were present in typical PKU only, others were present in variants only and some were present in both. In fact, the broad spectrum of clinical phenotypes in PAH deficiencies in France can be accounted for by the combination of different mutations differing by the residual activity of the mutant gene product. In support of this, we provided the first evidence for a mutation in PAH mRNA resulting in a variant form of hyperphenylalaninemia. This point mutation, characterized by sequencing a mutant cDNA clone derived from a needle biopsy of the liver, is linked to an original RFLP haplotype at the PAH locus found in South Europe and North Africa. Guthrie cards were used for large-scale retrospective screening of this mutation in PAH deficient patients, using polymerase chain reaction amplification of the genomic DNA extracted from dried blood spots.