**European Journal of Human Genetics (2000) 8,** 683–696

www.nature.com/ejhg

ARTICLE

# Structural interpretation of mutations in phenylalanine hydroxylase protein aids in identifying genotype–phenotype correlations in phenylketonuria

Ian G Jennings[1], Richard GH Cotton[2] and Bostjan Kobe[1]

[1]*Structural Biology Laboratory, St Vincent's Institute of Medical Research;* [2]*Mutation Research Centre, Fitzroy, Victoria, Australia*

**Phenylalanine hydroxylase (PAH) is the enzyme that converts phenylalanine to tyrosine as a rate-limiting step in phenylalanine catabolism and protein and neurotransmitter biosynthesis. Over 300 mutations have been identified in the gene encoding PAH that result in a deficient enzyme activity and lead to the disorders hyperphenylalaninaemia and phenylketonuria. The determination of the crystal structure of PAH now allows the determination of the structural basis of mutations resulting in PAH deficiency. We present an analysis of the structural basis of 120 mutations with a 'classified' biochemical phenotype and/or available *in vitro* expression data. We find that the mutations can be grouped into five structural categories, based on the distinct expected structural and functional effects of the mutations in each category. Missense mutations and small amino acid deletions are found in three categories: 'active site mutations', 'dimer interface mutations', and 'domain structure mutations'. Nonsense mutations and splicing mutations form the category of 'proteins with truncations and large deletions'. The final category, 'fusion proteins', is caused by frameshift mutations. We show that the structural information helps formulate some rules that will help predict the likely effects of unclassified and newly discovered mutations: proteins with truncations and large deletions, fusion proteins and active site mutations generally cause severe phenotypes; domain structure mutations and dimer interface mutations spread over a range of phenotypes, but domain structure mutations in the catalytic domain are more likely to be severe than domain structure mutations in the regulatory domain or dimer interface mutations.** *European Journal of Human Genetics* (2000) **8**, 683–696.

**Keywords:** genotype–phenotype correlations; hyperphenylalaninaemia; phenylalanine; hydroxylase; phenylketonuria; X-ray crystal structure

## Introduction

Phenylalanine hydroxylase (PAH) is the enzyme that catalyses the conversion of phenylalanine to tyrosine, a rate-limiting step in phenylalanine catabolism and protein and neurotransmitter biosynthesis (see recent reviews[1–4]). For catalytic activity, the enzyme requires a cofactor tetrahydrobiopterin ($BH_4$), enzyme-bound iron and molecular oxygen. The activity is tightly regulated by a variety of mechanisms, including activation by the substrate phenylalanine, inhibition by the cofactor $BH_4$, and additional activation by phosphorylation. Phenylalanine is both an essential amino acid and toxic at pathophysiological levels. Consequently, defects in PAH result in the disorders hyperpheylalaninaemia (HPA) and phenylketonuria (PKU), characterised by elevated serum concentrations of phenylalanine[5,6] (Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/omim). Deficiency of PAH enzymatic activity is the most common cause of HPA, with 99% mutant alleles mapping to the *PAH* gene; the remainder map to the genes coding for other enzymes involved in $BH_4$ homeostasis. The disorder is transmitted in an autosomal

**Correspondence: Bostjan Kobe, Structural Biology Laboratory, St Vincent's Institute of Medical Research, 41 Victoria Parade, Fitzroy, Victoria 3065, Australia. Tel: +61 3 9288 2480; Fax: +61 3 9416 2676; E-mail: b.kobe@medicine.unimelb.edu.au**
**Received 18 February 2000; revised 18 April 2000; accepted 27 April 2000**

recessive pattern and is the most common inborn error of amino acid metabolism in the Caucasian population, affecting on average one in 10 000 people.

PAH deficiency is highly heterogenous and results in a broad spectrum of phenotypes. Severe deficiency of PAH activity is termed 'classical PKU', while the milder forms of the disorder have been categorised for diagnostic and therapeutic purposes into 'moderate PKU', 'mild PKU' and 'mild HPA (MHP)'. PKU patients require dietary control of phenylalanine uptake to prevent physical, neurological and cognitive abnormalities.

PAH deficiency is also highly heterogeneous genetically, with over 300 mutations identified[7] (PAH Mutation Analysis Consortium Database, http://www.mcgill.ca/pahdb). The heterogeneity of phenotypes may reflect this continuous spectrum of mutant enzyme activities, and additionally the plethora of possible combinations of mutations that result in over 10 000 genotypes corresponding to different compound mutations. A simple correlation between PAH genotype and the metabolic phenotype may not always exist;[8] however, recent studies suggested that the PAH genotype is indeed the main determinant of the metabolic phenotype in most patients with PAH deficiency,[9,10] (and references therein). One such analysis, applied to a large number of patients in several different medical centers, allowed the classification of 105 PAH mutations for which the prediction of the biochemical phenotype from diverse genotypes is possible.[10]

One of the requirements in the quest to elucidate the molecular basis of PKU is the knowledge of the three-dimensional structure of PAH. We have recently determined the crystal structure of a truncation mutant of rat PAH (lacking 24 C-terminal residues) with catalytic and regulatory properties of the wild-type protein.[11] Previously, the crystal structures of smaller fragments, lacking the regulatory domain, of human PAH[12,13] and a homologous enzyme, rat tyrosine hydroxylases (TyrOH)[14] have also been determined. Jointly, this structural information now allows a structural interpretation of any PAH mutation resulting in PAH deficiency. Rat PAH shares 92% identity and 96% similarity with human PAH and the structures of the catalytic domains are basically identical; the structure of rat PAH therefore represents an adequate model to interpret the structural effects of mutations.

A monomer of PAH comprises an N-terminal regulatory domain (residues 1–119), a catalytic domain (residues 120–427) and a C-terminal tetramerisation domain (residues 428–452) (Figure 1). The catalytic domain consists of 13 helices and nine β-strands; its helices form a basket-like arrangement that surrounds a deep pocket where the active site is located. The iron is bound at the floor of this pocket, coordinated by residues H285, H290 and E330. The core of the regulatory domain has an α/β sandwich structure, but the very N-terminal sequence (residues 1–33) reaches across the catalytic domain active site and acts as an autoregulatory sequence.[11] The domain interface of the regulatory and catalytic domains is formed by the association of the β-sheet of the regulatory domain mainly with the segments 118–131 and 409–422 of the catalytic domain. The phenylalanine regulatory site is proposed to be located in the vicinity of this interface.[11]

A PAH dimer is formed by the regulatory domain of one monomer contacting the catalytic domain of the other (Figure 1). The tetramerisation domain is a 40 Å-long α-helix, and the tetramer forms through the association of four helices into an antiparallel coiled coil.[13]

Site-directed mutagenesis in conjunction with a three-dimensional structure of a protein has emerged as one of the most powerful methods in understanding protein function. Here, we use the database of naturally occurring mutations in PAH jointly with the three-dimensional structure in an analogous way to elucidate their effects on protein function, which ultimately defines the metabolic phenotype of PKU. Because of the complex effects many mutations have on the PAH protein, we expect the analysis to be informative to variable degrees for different mutations, but be of predictive value with respect to the phenotype for at least a subset of mutations, and be able to identify the mutations for which simple correlations may not exist. Similar structural analyses have been performed in the past on different disease-associated proteins, although on smaller numbers of mutations (eg [15–21]). Similarly, structural interpretations of selected mutations in PAH have been reported recently.[4,12–14,22] However, the criteria for selection of mutations analysed in these studies have not been detailed, and the effects have been discussed without the structural knowledge of the entire protein. To gain more general insights into the molecular basis of PAH deficiency, we chose in this study to correlate the metabolic phenotype and expression studies with the structural knowledge for a larger unbiased set of mutations. We examine the structural environment of the corresponding residues and seek to explain its effect on activity in the most informative subset of 120 mutations that includes the 104 'classified' mutations,[10] and additionally all the mutations for which the activity of the corresponding expressed proteins has been tested *in vitro*[7,23] (Table 1). The structural analysis allows us to formulate some simple rules that can be used to predict the effects of newly identified and unclassified mutations.

An independent analysis of the structural effects of a larger number of PAH mutations has been reported very recently, after our study had been completed.[24] The authors of this study made no attempt to correlate their findings with disease phenotype, or to provide any general insights into the molecular basis of PKU and genetic diseases.

## Materials and methods

The crystal structure of phosphorylated rat PAH lacking 24 C-terminal residues (PAH–24;[11] Protein Data Bank (PDB; http://www.rcsb.org/pdb) code 1PHZ) has been used for most

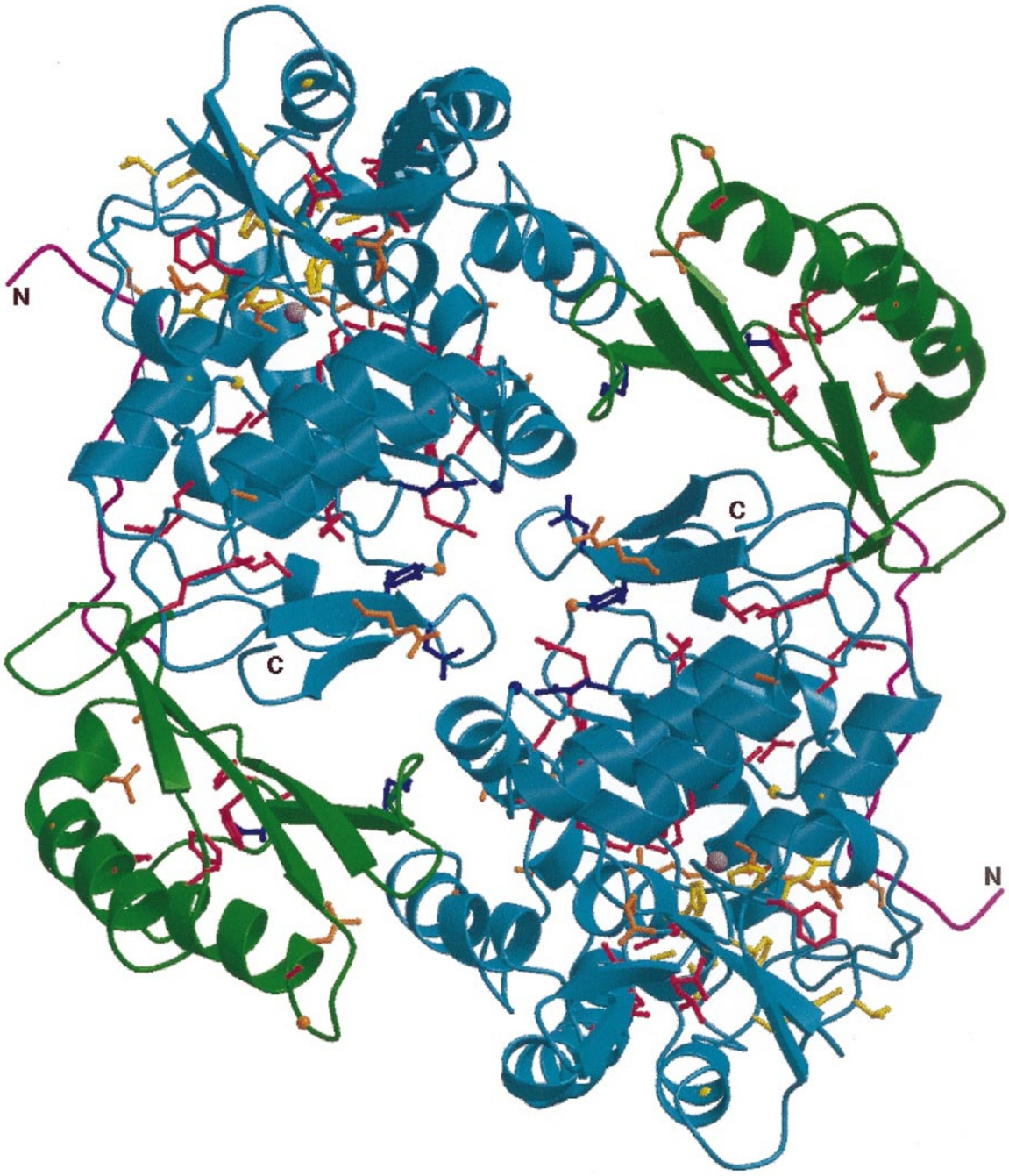


**Figure 1** Missense mutations in the three-dimensional structure of a dimer of PAH-24. The two monomers are related by a two-fold rotation around an axis perpendicular to the plane of the paper. The protein is shown in a ribbon representation, with the colours reflecting the domain organisation (N-terminal autoregulatory sequence, magenta; regulatory domain, green; catalytic domain, cyan). The tetramerisation domain is not present in this structure, but would be located at the C-terminus of PAH-24. The N-terminal 18 residues represent a mobile sequence and are not shown (the N-terminus corresponds to residue 19). Also mobile is the sequence 137–142, represented by a thin straight line. The iron atom is shown as a grey sphere. The side chains of residues affected by missense mutations causing PKU and listed in Table 1 are shown in a ball-and-stick representation. In the case of glycines, the C$\alpha$ atom is enlarged for better visibility. The colours reflect the mutation category: active site mutations, yellow; domain structure mutations, red (mostly buried residues) or orange (surface residues); dimer interface mutation, blue. The figure was generated with the programs BOBSCRIPT[48] and RASTER3D.[49]

**Table 1** Mutations associated with hyperphenylalaninaemia

| *Mutation* | *Metabolic phenotype* | *In vitro expression* | | | | *Structural environment* |
|---|---|---|---|---|---|---|
| | | *Host cells* | *Enzyme activity (% of wild type)* | *Immuno-reactivity (% of wild type)* | *Specific activity (% of wild type)* | |
| *Active site mutations* | | | | | | |
| G247V | Unclassified | COS | 4 | 56 | 7 | Loop Cβ1/Cα6 |
| R270S | Unclassified | COS | <1 | <1 | NA | Loop Cβ2/Cα7A |
| | | *E. coli*[a] | NA | NA | 0.8 | |
| | | TNT-T7[a] | NA | NA | 2 | |
| | | A293 a | NA | 2.2 | NA | |
| Y277D | Mild PKU | NA | NA | NA | NA | Loop Cβ2/Cα7A |
| E280K | Classic PKU | *E. coli* | 0.9 | ~100 | 0.9 | Loop Cβ2/Cα7A |
| | | COS | 1 to 3 | 1 to 3 | NA | |
| P281L | Classic PKU | COS | <1 | <1 | NA | Loop Cβ2/Cα7A |
| | | *E. coli* | NA | NA | NA | |
| | | *E. coli* (rat PAH) | NA | NA | 1 | |
| D282N | Classic PKU | *E. coli* (rat PAH) | NA | NA | ND | Loop Cβ2/Cα7A |
| H285Y | Classic PKU | NA | NA | NA | NA | Cα7A; iron binding |
| A322G | Unclassified | COS | 75 | 105 | 71 | Cα9; contacts autoregulatory sequence |
| F331L | Classic PKU | NA | NA | NA | NA | Loop Cα9/Cβ3 |
| G346R | Classic PKU | NA | NA | NA | NA | Cα10 |
| S349P | Classic PKU | *E. coli* | <0.2 | ~100 | <0.2 | Loop Cα10/Cα11 |
| | | COS | <1 | <1 | NA | |
| G352R | Classic PKU | NA | NA | NA | NA | Cα11 |
| T380M | MHP | NA | NA | NA | NA | Loop Cα12/Cβ6 |
| *Dimer interface mutations* | | | | | | |
| I65T | Moderate PKU | COS | 26 | 25 | NA | Rβ2; also close to domain interface |
| R68S | Mild PKU | NA | NA | NA | NA | Rβ2; also in the domain interface |
| G239S | Classic PKU | NA | NA | NA | NA | Loop Cα5/Cβ1 |
| R261P | Moderate PKU | NA | NA | NA | NA | Loop Cα6/Cβ2 |
| R261Q | Moderate PKU | *E. coli* | 47 | ~100 | 47 | Loop Cα6/Cβ2 |
| | | COS | 30 | 30 | NA | |
| Y414C | Mild PKU | COS | ~50 | ~50 | NA | Cβ7; also close to domain interface |
| D415N | MHP | NA | NA | NA | NA | Cβ7; also in the domain interface |
| *Domain structure mutations* | | | | | | |
| F39L | Moderate PKU | NA | NA | NA | NA | Rβ1; buried; also close to domain interface |
| G46S | Mild PKU | A293 | trace | 3 | NA | Loop Rβ1/Rα1; surface; close to dimer |
| | | *E. coli* | ND | NA | ND | interface |
| A47V | MHP | NA | NA | NA | NA | Rα1; mostly buried |
| L48S | Mild PKU | NA | NA | NA | NA | Rα1; partially buried; close to dimer interface |
| F55L | Classic PKU | NA | NA | NA | NA | Rα1; buried |
| T63P/H64N | Mild PKU | NA | NA | NA | NA | Loop Rα1/Rβ2; close to domain and dimer interfaces |
| D84Y | Classic PKU | NA | NA | NA | NA | Loop Rβ3/Rα2; surface |
| S87R | MHP | NA | NA | NA | NA | Loop Rβ3/Rα2; partially buried |
| T92I | MHP | NA | NA | NA | NA | Rα2; surface |
| delI94 | Unclassified | SW613–12A1 | 27 | NA | NA | Rα2; buried |
| I94S | Classic PKU | NA | NA | NA | NA | Rα2; buried |
| A104D | Mild PKU | A293 | 26 | 20 | NA | Loop Rα2/Rβ4; buried |
| | | *E. coli* | ND | NA | ND | |
| D143G | Unclassified | A293 | 15–33 | 100 | 15–33 | Loop Cα1/Cα2; surface |
| | | TNT–T7 | NA | NA | 16–102 | |
| | | *E. coli* | 33–52 | NA | 33–52 | |
| R155H | MHP | NA | NA | NA | NA | Cα2; surface |
| R157N | Unclassified | A293 | 5 | 5 | NA | Cα2; partially buried |
| R158Q | Classic PKU | COS | 10 | 100 | 10 | Cα2; partially buried |
| F161S | Unclassified | COS | 7 | 17 | NA | Cα2; buried |
| I164T | Mild PKU | NA | NA | NA | NA | Cα2; mostly buried |
| G171A | MHP | COS | 27 | NA | NA | Loop Cα2/Cα3; surface |
| I174T | Classic PKU | NA | NA | NA | NA | Loop Cα2/Cα3; buried |
| R176L | MHP | NA | NA | NA | NA | Loop Cα2/Cα3; surface |
| V177A | Mild PKU | NA | NA | NA | NA | Loop Cα2/Cα3; mostly buried |
| E178G | MHP | NA | NA | NA | NA | Loop Cα2/Cα3; surface |
| E178V | Unclassified | COS | 18 | NA | NA | Loop Cα2/Cα3; surface |

**Table 1** *Continued*

| *Mutation* | *Metabolic phenotype* | *In vitro expression* | | | | *Structural environment* |
|---|---|---|---|---|---|---|
| | | *Host cells* | *Enzyme activity (% of wild type)* | *Immuno-reactivity (% of wild type)* | *Specific activity (% of wild type)* | |
| V190A | MHP | NA | NA | NA | NA | Cα3; buried |
| L194P | Classic PKU | NA | NA | NA | NA | Cα3; buried |
| G218V | Unclassified | COS | 15 | NA | NA | Loop Cα4/Cα5; surface |
| V230I | MHP | NA | NA | NA | NA | Cα5; buried |
| S231P | Classic PKU | *E. coli* | <0.2 | NA | <0.2 | Cα5; mostly buried; close to dimer interface |
| R241C | MHP | COS | 25 | NA | NA | Cβ1; mostly buried; close to dimer interface |
| R241H | Mild PKU | COS | 23 | NA | NA | Cβ1; mostly buried; close to dimer interface |
| R243Q | Classic PKU | COS | <10 | <10 | NA | Cβ1; buried |
| P244L | Unclassified | COS | 70 | 100 | 70 | Cβ1; buried |
| V245A | MHP | NA | NA | NA | NA | Loop Cβ1/Cα6; mostly buried |
| A246V | Mild PKU | NA | NA | NA | NA | Loop Cβ1/Cα6; mostly buried |
| L311P | Classic PKU | COS | <1 | <1 | NA | Loop Cα8/Cα9; buried, close to domain interface |
| F327L | Classic PKU | NA | NA | NA | NA | Loop Cα9/Cβ3; buried |
| L333F | Unclassified | COS | 7 | NA | NA | Cβ3; buried |
| A342T | Classic PKU | NA | NA | NA | NA | Cβ4; buried |
| G344S | Mild PKU | NA | NA | NA | NA | Loop Cβ4/Cα10; buried |
| L348V | Moderate PKU | COS | 25–33 | NA | NA | Cα10; buried |
| delL364 | Unclassified | COS | ND | ~10 | NA | Cβ5; mostly exposed |
| V388M | Moderate PKU | COS | 43 | 100 | 43 | Cβ6; partially exposed |
| E390G | MHP | COS | 70 | NA | NA | Loop Cβ6/Cα13; surface |
| A395P | Classic PKU | NA | NA | NA | NA | Cα13; buried |
| A403V | MHP | COS | 32 | NA | NA | Cα13; partially buried |
| R408Q | Mild PKU | COS | 55 | 91 | 60 | Loop Cα13/Cβ7; mostly buried; close to domain interface |
| R408W | Classic PKU | COS | <3 | <3 | NA | Loop Cα13/Cβ7; mostly buried; close to domain interface |
| R413P | Unclassified | COS | <3 | ND | NA | Cβ7; surface; close to dimer interface |
| R413S | MHP | COS | 34 | NA | NA | Cβ7; surface; close to dimer interface |
| *Proteins with truncations and large deletions* | | | | | | |
| Q20X | Classic PKU | NA | NA | NA | NA | Segment preceding Rβ1 |
| R111X | Classic PKU | NA | NA | NA | NA | Loop Rb4/Rb5 |
| R176X | Classic PKU | NA | NA | NA | NA | Loop Cα2/Cα3 |
| W187X | Classic PKU | NA | NA | NA | NA | Cα3 |
| Y204X | Classic PKU | NA | NA | NA | NA | Cα4 |
| Y206X | Classic PKU | NA | NA | NA | NA | Cα4 |
| R243X | Classic PKU | *E. coli* | ND | ~100 | ND | Cβ1 |
| | | COS | <1 | <1 | NA | |
| R261X | Classic PKU | NA | NA | NA | NA | Loop Cα6/Cβ2 |
| R252G | Classic PKU | *E. coli*[a] | NA | NA | 5.2 | Cα6; buried; contacts N-term. regulatory |
| | | TNT–T7[a] | NA | NA | 0.7 | sequence; close to domain interface |
| | | A293[a] | NA | 5 | NA | |
| R252Q | Classic PKU | *E. coli*[a] | NA | NA | 11.4 | Cα6; buried; contacts N-term. regulatory |
| | | TNT–T7[a] | NA | NA | 3 | sequence; close to domain interface |
| | | A293[a] | NA | 2.3 | NA | |
| R252W | Classic PKU | *E. coli* | 0.5 | ~100 | 0.5 | Cα6; buried; contacts N-term. regulatory |
| | | COS | <1 | <1 | NA | sequence; close to domain interface |
| L255S | Unclassified | *E. coli*[a] | NA | NA | 0.03 | Cα6; buried |
| | | TNT–T7[a] | NA | NA | 1 | |
| | | A293[a] | NA | 10.8 | NA | |
| L255V | Unclassified | COS | 13 | 18 | NA | Cα6; buried |
| | | *E. coli*[a] | NA | NA | 17 | |
| | | TNT–T7[a] | TNT–T7a | NA | NA | |
| | | A293[a] | NA | 7.6 | NA | |
| A259T | Unclassified | *E. coli* | 0.3 | 100 | 0.3 | Loop Cα6/Cβ2; buried |
| | | *E. coli*[a] | NA | NA | 0.8 | |
| | | TNT–T7[a] | NA | NA | 8 | |
| | | A293[a] | NA | 2.6 | NA | |
| A259V | Classic PKU | *E. coli* | 0.2 | 100 | 0.2 | Loop Cα6/Cβ2; buried |
| | | COS | <1 | NA | NA | |
| | | *E. coli*[a] | NA | NA | 0.8 | |

**Table 1** *Continued*

| *Mutation* | *Metabolic phenotype* | *In vitro expression* | | | | *Structural environment* |
|---|---|---|---|---|---|---|
| | | *Host cells* | *Enzyme activity (% of wild type)* | *Immuno-reactivity (% of wild type)* | *Specific activity (% of wild type)* | |
| | | TNT–T7[a] | NA | NA | 3 | |
| | | A293[a] | NA | 5.8 | NA | |
| I269N | Classic PKU | NA | NA | NA | NA | Loop Cβ2/Cα7A; mostly buried |
| F299C | Classic PKU | *E. coli* | <0.2 | ~100 | <0.2 | Cα8; buried |
| | | COS | <3 | NA | NA | |
| A300S | MHP | NA | NA | NA | NA | Cα8; buried |
| I306V | MHP | NA | NA | NA | NA | Cα8; buried |
| G272X | Classic PKU | COS | ND | ND | NA | Loop Cβ7/Cα7A |
| | | *E. coli* | ND | NA | ND | |
| | | *E. coli*[a] | NA | NA | ND | |
| S295X | Classic PKU | NA | NA | NA | NA | Loop Cα7B/Cα8 |
| Q336X | Classic PKU | NA | NA | NA | NA | Cβ3 |
| Y356X | Classic PKU | NA | NA | NA | NA | Cα11 |
| S359X | Classic PKU | NA | NA | NA | NA | Cα13 |
| IVS12nt1G′A | Classic PKU | COS | <1 | <1 | NA | NA |
| *Fusion proteins* | | | | | | |
| F55fsdelT | Classic PKU | NA | NA | NA | NA | Rα1 |
| F89fsinsC | Classic PKU | NA | NA | NA | NA | Rα2 |
| L197fsdel22bp | Classic PKU | NA | NA | NA | NA | Cα3 |
| Y198fsdel22bp | Classic PKU | NA | NA | NA | NA | Cα3 |
| E221D222fsdel | Classic PKU | NA | NA | NA | NA | Loop Cα4/Cα5 |
| K274fsdel11bp | Classic PKU | NA | NA | NA | NA | Loop Cβ2/Cα7A |
| S310fsdel11bp | Classic PKU | NA | NA | NA | NA | Cα8 |
| A342fsdelG | Classic PKU | NA | NA | NA | NA | Cβ4 |
| G346fsdelG | Classic PKU | NA | NA | NA | NA | Cα10 |
| K363fsdelG | Classic PKU | NA | NA | NA | NA | Cβ5 |
| R367fsinsC | Classic PKU | NA | NA | NA | NA | Loop Cβ5/Cα12 |
| P407fsdelC | Classic PKU | NA | NA | NA | NA | Loop Cα13/Cβ7 |
| K452fsinsA | Classic PKU | NA | NA | NA | NA | Cα14 |

Methods used for *in vitro* expression studies vary for different mutants; different methods can yield different results. For a discussion of the methods used, see Waters *et al*.[23].

When the amount of immunoreactive protein is decreased as compared to the wild-type protein, this indicates reduced stability, through any of the processes such as defective synthesis or increased degradation. By contrast, normal levels of immunoreactive protein with decreased enzymatic activity reflect a stable protein with affected kinetic parameters.

Details on the *in vitro* expression data can be found in the PAH Mutation Analysis Database (http://www.mcgill.ca/pahdb) and in Waters *et al*[23] and Bjorgo *et al* (as indicated by superscript 'a').

Cell lines used: COS, monkey kidney cell line; A293, human embryonic kidney cell line; SW613–12A1, human colon adrenocarcinoma cell line. TNT–T7 is a rabbit reticulocyte lysate based cell-free expression system.

PAH enzymatic activity is assayed in lysates; both the natural cofactor tetrahydrobiopterin or a synthetic analogue (6-methyltetrahydropterin) have been used. PAH immunoreactive protein is quantified by western blots of cell lysates.

Specific activity denotes the activity of purified PAH protein normalised relative to the quantity of the protein.

For classic PKU mutations IVS1nt5G′T, IVS2nt5G′C, IVS7nt1G′A, IVS7nt5G′A, IVS8nt1G′A, IVS8nt–7A′G, IVS10nt–11G′A, IVS10nt–3C′T, IVS10nt–1G′A, and IVS11nt5G′A, moderate PKU mutation E6nt–96A′G and mild PKU mutation IVS4nt–5C′G in the 'Proteins with truncations and large deletions' category, no information is available either on *in vitro* expression or the structural environment, and have therefore not been included in Table 1.

analyses. The crystal structures of human PAH fragment, residues 117–424[12] (PDB code 1PAH) and rat TyrOH fragment, residues 156–498[14] (PDB code 1TOH) were also consulted. Structure analyses and modelling were performed using the graphics program 'O'[25] and the contacts were analysed using the program CONTACT, CCP4 program suite.[26]

Mutations are referred to by their 'trivial names', as registered in the PAH Mutation Analysis Consortium Database (http://www.mcgill.ca/pahdb). The corresponding recommended systematic names[27] can be found in the database. For consistency, amino acid residues are also referred to by the single letter code.

The phenotypes specified are metabolic phenotypes as classified by Guldberg *et al*.[10] These assignments were based on dietary phenylalanine tolerance (in patients with PKU) or pretreatment blood phenylalanine levels (in patients with MHP). To keep blood phenylalanine levels at 300 mmol/l, patients with classic PKU tolerate < 20 mg phenylalanine/kg body weight/d (equivalent to 250–350 mg/d); patients with moderate PKU tolerate 20–25 mg phenylalanine/kg body weight/d (equivalent to 350–400 mg/d); patients with mild

PKU tolerate 25–50 mg phenylalanine/kg body weight/d (equivalent to 400–600 mg/d). Patients with MHP have phenylalanine levels < 600 mmol/l on normal diet. For consistency, mutations analysed in this study that were not classified by Guldberg *et al*[10] are considered 'unclassified'. The phenotype of the 'unclassified' mutations may have been described in another study or in the PAH Mutation Analysis Consortium Database (http://www.mcgill.ca/pahdb), but the criteria for the classification may have been different from those of Guldberg *et al*.[10]

## Results

### Overview of mutations

As a step towards understanding the molecular basis of PKU, we present the interpretation of the effect of mutations on PAH activity, based on the available three-dimensional structural information of PAH,[11,13] of 120 mutations associated with PKU that include the 104 'classified' PKU mutations[10] and all the mutations where the corresponding mutant PAH protein has been expressed *in vitro*[7,23] (Table 1). Mutations in PAH associated with PKU occur in all 13 exons of the *PAH* gene and the flanking nucleotide sequence, and comprise missense, nonsense, insertion, deletion, frameshift and splicing mutations. Based on their apparent effects on the structure and function of PAH, we arranged the mutations in five structural mutation categories (Table 2). We discuss the structural effects together with the available *in vitro* expression data and evolutionary conservation, and correlate these data with the metabolic phenotype.

Mutations can affect the specific activity of the enzyme (through altered kinetic behaviour), decrease the PAH protein level without affecting the specific activity, or both. To predict the metabolic phenotype, it may not always be necessary to determine which of these effects is the dominant one; for example, a truncated protein lacking crucial active site residues cannot be active, irrespective of the quantity of protein actually expressed. The structural analysis, combined with an analysis of residue conservation in PAH sequences during evolution (Figure 2), tells us how important are the catalytic and structural roles the affected residues play in the enzyme; what we are assessing is the correlation between the anticipated effect of the mutation on the structure, and the metabolic phenotype. We can expect the correlation may be less evident for a subset of mutations, such as the ones affecting structurally important residues. The analysis should shed light on the likely effect of the mutations on thermodynamic stability of the protein; however, the simultaneous effects on the folding pathway (which may lead to incorrect association of polypeptides), and the intracellular stability (which involves interactions with chaperones, proteases and other factors), may be more difficult to assess. The same considerations underscore the previously observed discordances between genotype and metabolic phenotype,[8] (and references therein). Expression studies of mutant proteins can further illuminate these processes, and for this reason we discuss our structural interpretations in the light of the available expression data. Expression of the mutant proteins in different *in vitro* systems can yield different results.[22,23] We present all the available data, but note that mammalian expression systems will more closely reflect the true situation; similarly, careful analysis of the expressed proteins is necessary. Finally, unexpected effects of mutations in the coding region on the splicing process represent yet another complication.[28] In the absence of RT-PCR studies in PKU patients, little is presently known about such effects for most mutations, and we generally assume full-length transcripts are produced.

### Active site mutations

Because the exact roles in substrate binding and catalysis of many amino acid residues in the PAH active site are not yet known, we considered mutations affecting any residue lining the active site cavity in this category (yellow in Figures 1 and 3). Most of these residues are strictly conserved in PAH sequences (Figure 2). The iron, bound to the enzyme through the side chains of H285, H290, and E330, is essential for enzymatic activity. Accordingly, the mutation H285Y causes classic PKU, and the mutations H285S and H290S in the rat enzyme yielded enzymes with undetectable specific activities.[23]

The crystal structure of the catalytic domain of PAH with 7,8-dihydrobiopterin bound[29] identifies the putative residues involved in $BH_4$ binding. The G247V mutation affects one of these residues and results in severely decreased activity. The A322G mutant retains about 75% activity *in vitro*; this residue binds pterin through its main chain atoms, and is accordingly not conserved in *Drosophila* PAH. Mutations of some other pterin-binding residues in PAH (L249F, L249H, L255V, L255S) have been found in PKU patients. Mutations of

**Table 2** Distribution of mutation types and metabolic phenotypes in different structural mutation categories

| *Structural mutation category* | *Mutation types* | *Classified mutations* | *MHP* | *Mild PKU* | *Moderate PKU* | *Classic PKU* |
|---|---|---|---|---|---|---|
| Active site mutations | Missense | 10 | 1 | 1 | 0 | 8 |
| Dimer interface mutations | Missense | 7 | 1 | 2 | 3 | 1 |
| Domain structure mutations (regulatory domain/catalytic domain) | Missense, single amino acid deletions | 48 (11/37) | 16 (3/13) | 10 (4/6) | 3 (1/2) | 19 (3/16) |
| Proteins with truncations and large deletions | Nonsense, splicing | 26 | 0 | 1 | 1 | 24 |
| Fusion proteins | Frameshift | 13 | 0 | 0 | 0 | 13 |

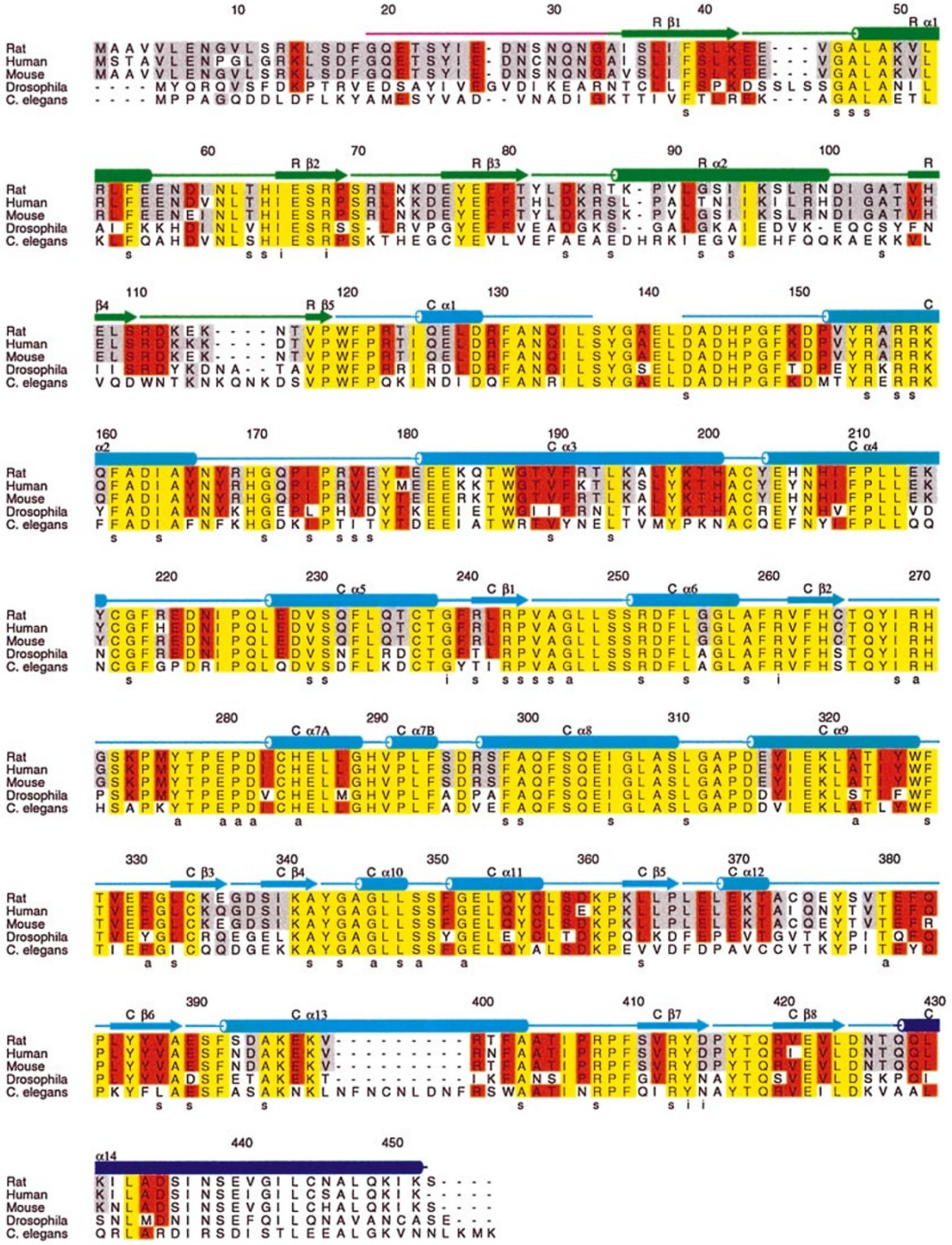


**Figure 2** Amino acid sequence alignment of rat, human, mouse, *Drosophila* and *C. elegans* PAH (SwissProt accession numbers P04176, P00439, P16631, P17276, P90925). Residues conserved in three, four or five proteins are highlighted in grey, orange and yellow, respectively. Elements of secondary structure as observed in the crystal structure of rat PAH-24[11] and their denominations are indicated above the PheOH sequence; the colours indicate the domain structure as shown in Figure 1 (N-terminal autoregulatory sequence, magenta; regulatory domain, green; catalytic domain, cyan). The secondary structure of the tetramerisation domain (blue) is indicated as observed in the crystal structure of the C-terminal fragment (residues 118–452) of human PAH.[13] The missense mutations associated with PKU are indicated below the sequences according to the structural mutation categories: 'a', active site mutations; 'i', dimer interface mutations; 's', domain structure mutations. The figure was generated with the program ALSCRIPT.[50]

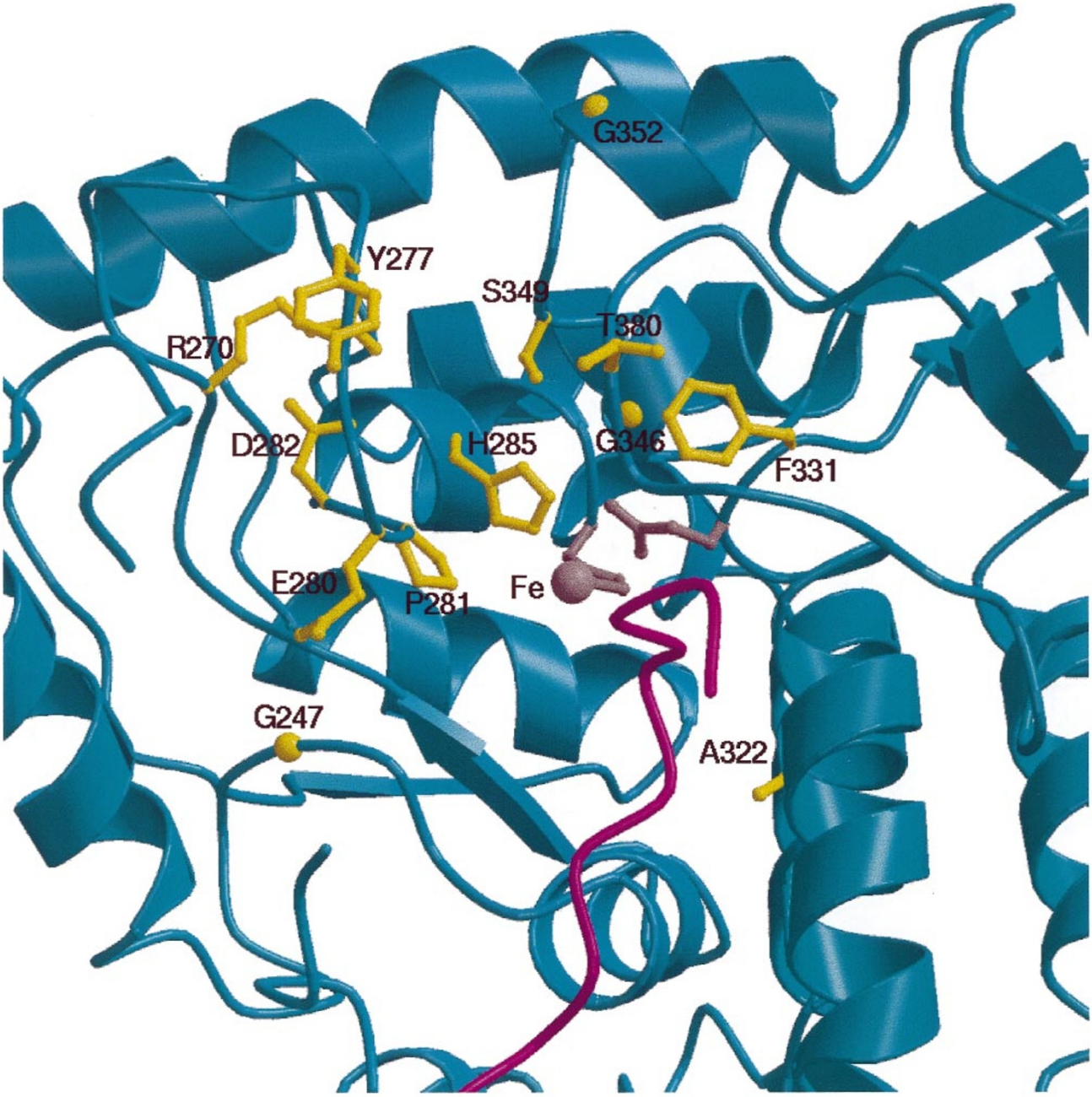


**Figure 3** Active site mutations in the three-dimensional structure of a monomer of PheOH-24. Residues affected by mutations (yellow) and the rest of the protein are drawn as in Figure 2. The side chains of H290 and E330 are also shown (grey). The view is related to the view in Figure 1 by a 90° rotation along the *y* axis.

another residue involved in pterin binding (E286A and E286Q), so far not observed in PKU patients, have been made in rat PAH and expressed in *E. coli*; the corresponding proteins have very low specific activities (0.5 and 0.02% of wild type, respectively), and E286A has a 70-fold increased $K_m$ for $BH_4$.[23]

Other mutations in the active site may affect phenylalanine substrate binding, the overall structure of the active site, or have more indirect effects by affecting protein folding and stability. Several mutations cluster in the Cβ2/Cα7A loop. The E280K mutation would be expected to affect severely the electrostatic properties of the active site crevice.

The more subtle modification E280A retains 30% specific activity in rat PAH,[23] and the corresponding mutation in TyrOH yields similar results,[30] suggesting this residue does not have a direct role in catalysis. The following residue, P281, protrudes in the active site and is likely an important determinant of the active site configuration. The results from *in vitro* expression suggest it may also be important for stability. However, while P281L and P281R mutants of rat PAH expressed in *E. coli* also have the activity reduced 100-fold, the P281 A mutant retains 83% activity.[23] It appears that bulky residues take up space required for substrate binding and catalysis (the $K_m$ for the P281L mutant of rat PAH is increased 23-fold compared with wild type), but may simultaneously affect folding and stability.

The charge elimination by the D282N mutation would also be expected to affect the conformation of the Cβ2/Cα7A loop, as this loop is held by a salt bridge between D282 and R270; this may explain the severe consequences of this apparently subtle mutation. R270 is similarly sensitive to modification; even the R270K mutation in rat PAH results in a complete loss of activity.[23] The D328S and R316K mutations in TyrOH, of the residues corresponding to D282 and R270 in PAH, respectively, affect amino acid substrate binding.[30] However, reduced stability is likely the most important attribute for the association of the R270S mutation with PKU *in vivo*, as underscored by the increased aggregation of the R270S mutant protein expressed in *E. coli* and increased degradation when expressed in eukaryotic cells.[22]

The remaining mutations in Table 1, Y277D, F331L, G346R, S349P, G352R, and T380M all map to a similar region of the active site crevice, opposite with respect to the iron binding site to where the pterin binding site is located. This is the most likely region for the phenylalanine binding site.[11,31] G352 and T380 are not conserved in bacterial PAH and in the homologue tryptophan hydroxylase (TrpOH), perhaps reflecting differences in substrate specificity. The hydroxyl group of S349 also hydrogen bonds to the side chain of the iron-ligating residue H285; this arrangement may be crucial for the proper configuration of the iron binding site. Also in this region is residue T278; T278A, T278N and T278I mutations are associated with PKU, while the T278V mutation of rat PAH reduces the activity twofold.[23]

**Domain structure mutations**

The mutations in this category affect either residues forming the hydrophobic cores of the protein domains, or residues that may in other ways be responsible for the structural integrity of the individual domains (red in Figure 1). Correlations between genotype and phenotype may be most difficult to make out for the mutations in this category, because they may affect not only the thermodynamic stability of a protein, but also processes that lead to the adoption of the folded structure, such as protein folding and possible non-specific association during folding, and consequently associations with cellular factors involved in this process, such as chaperones, and proteases. However, the analysis is useful if only to identify mutations in this class, because these mutations will be the most likely ones to show complex phenotypic behaviour. A few mutations in this class may also affect the specific activity by indirectly affecting the active site structure through structural perturbations.

Mutations R158Q, V245A, F299C, A300S, I306V, G344S and L348V affect residues very close to the active site. Some of the corresponding mutant proteins (R158Q, F299C) accordingly form stable proteins with reduced catalytic activities during *in vitro* expression. While the F299C mutant is essentially inactive when expressed in *E. coli*, the F299Y mutation only reduces the activity twofold in rat PAH expressed in *E. coli*,[23] suggesting a bulky side chain is required in this position. Mutants C265S, Q267E and V291T of rat PAH affecting structural residues very close to the active site have their activities reduced to 30, 11 and 57% of the wild type, respectively;[23] the activities appear to correlate with the magnitude of the change.

Most residues affected by the mutations in the domain structure category are entirely or mostly buried (red in Figure 1). Different types of mutations will have different effects. Mutations of large buried hydrophobic residues to smaller ones create cavities in the hydrophobic core and destabilise the protein, the larger the resulting cavity the larger the effect;[32] mutations in this group include F39L, F55L, V177A, V190A, V245A, L255V, F299C, I306V, F372L, and L348V. The mutations of small hydrophobic residues to larger ones may in most cases be even more devastating, because they would require the surrounding protein to adjust and create space for the accommodation of the bulkier side chains; the mutations in this group are A47V, V230I, A246V, A259V, L333F, V388M, and A403V. Mutations of buried hydrophobic residues to polar ones will have effects analogous to the ones described above, but additionally burial of polar side chains will result in a further decrease in stability; many PAH domain structure mutations fall in this group, eg L48S, I94S, A104D, F161S, I164T, I174T, L255S, A259T, A342T, I269N. Finally, there are a few deleterious mutations of buried polar residues, mainly arginines (R241C, R241H, R243Q, R252G, R252Q, R252W, R408Q, and R408W). Arginines have long aliphatic side chains with a positively charged head group. The polar atoms may form specific electrostatic and hydrogen-bonding interactions important for stability (eg R243 with D129, and R252 with D315), but the aliphatic portion of the side chain may also provide important hydrophobic stabilisation. Mutations to the conformationally restricted proline residues (eg L194P, A195P, S231P, L311P) represent a special group of mutations that may have severe folding effects.

The minority of domain structure mutations affect surface residues (orange in Figure 1). Because surface amino acids can

often freely be replaced with little effect on protein stability,[32] the PKU-associated mutations of surface residues probably reflect important structural or folding roles of the affected residues in the enzyme. Accordingly, some of these residues form electrostatic interactions that are affected by the substitutions (D84Y affects interaction with R86, R157N affects interaction with E183, R176L affects interaction with D229, and R413P and R413S affect interaction with E422). Glycines have conformational properties not accessible to other amino acids; conformational constraints imposed may therefore explain the effects of mutations G46S, G171A and G218V. Conversely, mutations that enhance conformational degrees of freedom (E178G and E390G) can increase the free energy of the unfolded state, also resulting in destabilisation.[32] Mutations of polar surface residues to hydrophobic ones destabilise the folded state of the protein (mutations T92I and E178V). The D143G mutation occurs in a loop at the edge of the active site crevice and may affect the active site structure besides stabilising the unfolded state. Finally, S87R and R155H mutations would lead to only minor structural effects, consistent with their mild phenotypes.

The domain structure mutation category includes two one-residue deletions delI94 and delL364. While amino acid substitutions often exhibit only local effects on the protein structure, deletions will in general have more severe effects, as they can only be accommodated by structural rearrangements of the surrounding sequence. Smaller effect would only be expected when deletions occur in longer solvent exposed loops. The effect of delI94 is likely less severe than the effect of delL364 because the former is located in the regulatory domain distant to the active site.

An extensive characterisation of the mutations R252G, R252Q, L255V, L255S, A259V and A259T recently showed that misfolding is the major cause for the metabolic defect for these mutations.[22] All mutant proteins showed reduced cellular stability when expressed in human embryonic kidney cells, and the folding defect further manifested itself through defective oligomeric assembly and increased proteolytic degradation in both prokaryotic and eukaryotic systems.

The evolutionary conservation of affected residues in PAH (Figure 2) and other aromatic amino acid hydroxylases correlates well with the structural observations; mutations of non-conserved residues lead to milder phenotypes, while mutations of highly conserved residues often lead to severe phenotypes.

**Dimer interface mutations**

Mutations in this category affect residues in the dimer interface (blue in Figure 1). Because extensive surface areas of individual PAH monomers are buried in the dimer interface, monomers are expected to be unstable on their own, and association is probably coupled with subunit folding. The nature of structural effects of mutations in the dimer interface will therefore be similar to the effects of mutations in the hydrophobic core of the protein domains, albeit the functional consequence of these effects may be less severe in the former. The dimer interface also probably participates in the conformational changes occurring upon activation by phenylalanine, and the interface mutations may interfere with the activation process. Most residues in the dimer interface affected by mutations are strictly conserved in PAH sequences (Figure 2).

*In vitro* expression studies indicate that I65T, R261Q and Y414C mutations result in decreased protein levels, suggesting folding and stability are affected. R261 is also located close to C237, chemical modification of which mimics phenylalanine activation.[33] Similarly, R68S and D415N affect residues in both the dimer and domain interfaces and may influence phenylalanine activation. R261 and D415 are accordingly not conserved in some PAH sequences that may not be regulated as the mammalian enzymes.

Most dimer interface mutations are associated with milder forms of PKU, suggesting alterations of the dimer interface have less severe consequences than alterations of the domain hydrophobic cores. The exception is the G239S mutation; the severe effect of this mutation may be a combination of steric hindrance introduced in the dimer interface and more restrictive conformational properties of the bulkier side chain that may interfere with folding.

**Proteins with truncations and large deletions**

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

The catalytic domain of PAH spans residues 120–427. The most C-terminal residue lining the catalytic cavity is Q383. If the proteins carrying nonsense mutations are expressed, the proteins terminating before residue Q383 would have an incompletely assembled active site; indeed, the corresponding mutations are all associated with classic PKU (Table 1). The R243X and G272X mutant proteins have been expressed *in vitro*; both are unstable in mammalian cells, but the former produces a stable protein with undetectable catalytic activity in *E. coli*.

Most splicing mutations occur in the introns; however, the mutation E6nt–96A→G that was originally referred to as Y204C actually affects RNA splicing *in vivo*.[34] Similarly, mutations G272X, P281L and R408Q result in transcripts with one or more exons skipped in addition to full-length transcripts, while Y356X yields no full-length transcripts in patients' lymphocytes.[28] This indicates that many mutations believed to affect the PAH enzyme may (also) affect splicing, and consequently patient mRNA should be analysed for all mutations. At present such analyses are available only for a few mutations. Structurally, at least exons 4–12 are absolutely required to produce an active enzyme.

The mutation IVS12nt1G→A results in a truncated protein lacking the C-terminal 52 residues (residues 401–452).[35] While deletions of the tetramerisation domain do not affect the catalytic and regulatory properties of PAH to a large extent,[36–38] residues 409–422 participate in the dimer and domain interfaces.[11] *In vitro* expression of the mutant protein comprising residues 1–400 suggests an unstable protein is produced.

The mutation E6nt–96A→G results in a 32-amino acid deletion, with residues 205–235 substituted for a single serine residue.[34] The affected region forms part of the dimer interface and would be expected to affect folding, stability and the integrity of the active site.

**Fusion proteins**

If expressed, insertion or deletion mutations introducing frameshifts will result in mutant proteins containing a truncated PAH sequence fused to an unrelated sequence at the C-terminus; such proteins are herein referred to as fusion proteins. Because an assembled active site is required for PAH activity, frameshift mutations at residues N-terminal to residue Q383 will result in inactive proteins. Indeed it is found that such mutant proteins are all associated with classic PKU. Mutations C-terminal to residue Q383 may indirectly affect the active site, interfere with folding, cause aggregation, or form unstable proteins. One of these mechanisms may be responsible for the classic PKU phenotype associated with frameshifts at residues P407 and K452. No *in vitro* expression data are available for any of the fusion proteins.

## Discussion

The synthesis of a functional protein from a gene is neither a one-step nor a single-route process. Processes of splicing, translation and protein folding compete with simultaneous degradation processes, all in the context of a complex cellular environment. The link between the final protein activity and the resulting disease phenotype represents yet another degree of complexity.[8] Nevertheless, the function of the protein will ultimately depend on the three-dimensional structure it adopts, and the nucleotide and the corresponding amino acid sequences are the basic determinants of the cellular fate and the activity of the protein. By examining the native three-dimensional structure the protein activity depends on, we should therefore unearth the mechanisms by which mutations perturb this structure and consequently the activity. Such an analysis represents only one piece in the genotype-to-phenotype puzzle, but especially when examined in the context of other data, such as *in vitro* expression studies, will help in determining the genotype–phenotype correlations. Likewise, site-directed mutagenesis when combined with the knowledge of the three-dimensional structures of proteins represents one of the most powerful methods to examine protein function.

The existence of some correlation between our structural interpretations and metabolic phenotype becomes apparent when the distribution of phenotypes is examined in the different structural mutation categories (Table 2). Proteins with truncations and large deletions, fusion proteins and active site mutations usually result in severe phenotypes. Domain structure mutations spread over a range of phenotypes, but mutations in the catalytic domain are more likely to be severe than mutations in the regulatory domain. Dimer interface mutations result in similar range of phenotypes as the domain structure mutations in the regulatory domain.

Our analysis allows the formulation of a few simple rules relating genotype with the metabolic phenotype.

(a) Nonsense mutations will generally lead to classic PKU, unless they are located at the extreme C-terminus of the protein and do not reduce protein expression.

(b) Splicing mutations will similarly result in classic PKU, with the possible exception of mutations that affect splicing only to a minor extent and/or result in expressed proteins missing only exons 1–3 and/or 13.

(c) Missense mutations and small amino acid deletions and insertions have to be analysed individually, assessing the magnitude of amino acid substitution, insertion or deletion.

(d) Active site mutations will often lead to classic PKU, although less severe effects are possible. Determination of three-dimensional structures of complexes of PAH with substrates or products and site-directed mutagenesis in combination with kinetic analyses will be necessary to further define the roles of important active site residues, and this in turn will be useful for the predictions of effects of mutations.

(e) Domain structure mutations will result in the broadest range of metabolic phenotypes, depending on the location of the affected amino acid; mutations in the catalytic domain will generally have more severe phenotypes than mutations in the regulatory domain, and mutations of buried residues will generally have more severe effects than mutations of surface residues.

(f) Whilst the mechanisms leading to loss of activity for dimer interface mutations are similar as for domain structure mutations, the associated phenotypes will generally be less severe for the former class.

Our analysis also shows that the largest percentage of mutations associated with PKU are domain structure mutations that generally affect the folding efficiency and thermodynamic stability of the protein, resulting in a reduced intracellular stability. Whilst it is usually possible to determine the effects such mutations may have on the native three-dimensional structure through an analysis like ours,

the effects on the folding process itself and the possibility of incorrect associations during folding, superimposed with the simultaneous processes in the cellular environment such as interactions with chaperones and degradation by proteases, make the prediction of the ultimate fate of the protein function in the cell difficult. For example, accelerated proteolytic turnover has been shown for mutations F39L, K42I, L48S, I65T, A104D and R157N.[39] Nevertheless, our results show that the analysis of the three-dimensional structure can in the most favourable cases give us reliable predictions of the metabolic phenotype, and in less favourable cases identify mutations for which such predictions may be difficult to make and the more laborious expression studies may be necessary. Ultimately, the protein function does depend on its three-dimensional structure, and that in turn is encoded by its sequence; it is the pathway between them that is complex, and a combination of methods will be necessary to get a grip on all the steps involved. This is underscored by the recent study of PAH mRNA from PKU patients, showing unexpected splicing effects by some nonsense and missense mutations.[28] Every one of these methods, however, yields a different piece of the puzzle and is therefore essential for the complete understanding of mechanisms leading to PKU.

There are some mutations in the PAH Mutation Analysis Consortium Database (http://www.mcgill.ca/pahdb), not included in our set of 120 mutations, that may not fit into any of our structural categories. These include mutations in the tetramerisation domain and the N-terminal autoregulatory sequence. The dimeric protein lacking the tetramerisation domain has similar catalytic and regulatory properties as wild type PAH,[36–38] but requires twofold higher phenylalanine concentration for activation. The two known mutations (L430P and A447D) may cause protein aggregation. The mutations in the N-terminal autoregulatory sequence that destabilised the interaction between the autoregulatory sequence and the catalytic domain could result in a constitutively active enzyme and lead to hypophenylalaninaemia, not HPA. The only known missense mutation in the N-terminal autoregulatory sequence known is S16P. Phosphorylation at S16 reduces the concentration of phenylalanine required for activation,[1–3] and the S16P mutation would be expected to affect the delicate balance of phenylalanine concentration and PAH activity *in vivo*.

The large number of different mutations resulting in defective enzyme activity points to PAH being an enzyme not very tolerant to mutations. In total 235 missense PAH mutations associated with PKU have been identified at the time of this study, affecting 170 residues. If only the catalytic domain of PAH is taken into account; mutations in 135 residues of the 308 residues are associated with PKU. Tolerance of protein sequences to mutations has been studied in a few systems, including T4 lysozyme,[40] HIV-1 protease[41] and lac repressor.[42] In these systems, some 40% of amino acid positions are sensitive to substitution. The percentage of sensitive residues in PAH as assessed by the PKU mutations is comparable, even though the scope of the natural 'mutagenesis' is likely to be less thorough. The high sensitivity of PAH to mutations likely reflects its special structural characteristics associated with its tight regulation, required as a result of its central role as a rate-limiting step of several metabolic pathways.

Further expression, structural and mechanistic studies will give us the opportunity to utilise one of the largest available mutation databases to improve our understanding of the resistance of enzymes to genetic alteration. On the other hand, the availability of structural information will help predict the likely effects of unclassified mutations associated with PKU, as well as new mutations that may be discovered. The type of analysis we performed could in principle be extended through computational modelling studies, but only when the available force fields are improved.[43] Deeper insight into the molecular effects of PAH mutations will finally be possible with a more detailed understanding of the substrate binding sites and the catalytic mechanism, detailed expression studies, analysis of transcripts, and structural studies of individual mutant proteins.

Because of the likely conservation of three-dimensional structures in the aromatic amino acid hydroxylase family, the structure of PAH may also provide a structural framework for understanding the inherited diseases associated with TyrOH and TrpOH, including the neuro-degenerative disorders infantile Parkinsonism and Segawa's syndrome.[44–47]

## Acknowledgements

*This work was supported by the Australian Research Council, Wellcome Trust (BK) and the National Health and Medical Research Council (NHMRC) (RGHC); BK is a Wellcome Senior Research Fellow in Medical Science in Australia, and RGHC is an NHMRC Fellow.*

## References

1 Kaufman S: The phenylalanine hydroxylating system. *Adv Enzymol* 1993; **67**: 77–264.
2 Hufton SE, Jennings IG, Cotton RGH: Structure and function of the aromatic amino acid hydroxylases. *Biochem J* 1995; **311**: 353–366.
3 Kappock TJ, Caradonna JP: Pterin-dependent amino acid hydroxylases. *Chem Rev* 1996; **96**: 2659–2756.
4 Flatmark T, Stevens RC: Structural insight into the aromatic amino acid hydroxylases and their disease-related mutant forms. *Chem Rev* 1999; **99**: 2137–2160.
5 Jervis JA: Studies on phenylpyruvic oligophrenia. The position of the metabolic error. *J Biol Chem* 1947; **169**: 651–656.
6 Scriver CR, Eisensmith RC, Woo SLC, Kaufman S: The hyperphenylalaninemias of man and mouse. *Annu Rev Genet* 1994; **28**: 141–165.
7 Nowacki PM, Byck S, Prevost L, Scriver CR: PAH Mutation Analysis Consortium Database 1997: prototype for relational locus-specific mutation databases. *Nucleic Acids Res* 1998; **26**: 220–225.
8 Scriver CR, Waters PJ: Monogenic traits are not simple. *Trends Genet* 1999; **15**: 267–272.
9 Kayaalp E, Treacy E, Waters PJ, Byck S, Nowacki P, Scriver CR: Human phenylalanine hydroxylase mutations and hyperphenylalaninemia phenotypes: a metanalysis of genotype–phenotype correlations. *Am J Hum Genet* 1997; **61**: 1309–1317.

10 Guldberg P, Rey F, Zschocke J *et al*: A European multicenter study of phenylalanine hydroxylase deficiency: classification of 105 mutations and a general system for genotype-based prediction of metabolic phenotype. *Am J Hum Genet* 1998; **63**: 71–79.
11 Kobe B, Jennings IG, House CM *et al*: Structural basis of autoregulation of phenylalanine hydroxylase. *Nat Struct Biol* 1999; **6**: 442–448.
12 Erlandsen H, Fusetti F, Martinez A, Hough E, Flatmark T, Stevens RC: Crystal structure of the catalytic domain of human phenylalanine hydroxylase reveals the structural basis for phenylketonuria. *Nat Struct Biol* 1997; **4**: 995–1000.
13 Fusetti F, Erlandsen H, Flatmark T, Stevens RC: Structure of tetrameric human phenylalanine hydroxylase and its implications for phenylketonuria. *J Biol Chem* 1998; **273**: 16962–16967.
14 Goodwill KE, Sabatier C, Marks C, Raag R, Fitzpatrick PF, Stevens RC: Crystal structure of tyrosine hydroxylase at 2.3 Å and its implications for inherited neurodegenerative diseases. *Nat Struct Biol* 1997; **4**: 578–585.
15 Brownlie PD, Lambert R, Louie GV *et al*: The three-dimensional structures of mutants of porphobilinogen deaminase: Toward an understanding of the structural basis of acute intermittent porphyria. *Protein Sci* 1994; **3**: 1644–1650.
16 Vihinen M, Veitre D, Maniar HS *et al*: Structural basis for chromosome C-linked agammaglobulinemia: a tyrosine kinase disease. *Proc Natl Acad Sci USA* 1994; **91**: 12803–12807.
17 Rayment I, Holden HM, Sellers JR, Fananapazir L, Epstein ND: Structural interpretation of the mutations in the beta-cardiac myosin that have been implicated in familial hypertrophic cardiomyopathy. *Proc Natl Acad Sci USA* 1995; **92**: 3864–3868.
18 Tuchman M, Morizono H, Reish O, Yuan X, Allewell NM: The molecular basis of ornithine transcarbamylase deficiency: modelling the human enzyme and the effects of mutations. *J Med Genet* 1995; **32**: 680–688.
19 Vihinen M, Zvelebil MJJM, Zhu Q *et al*: Structural basis for pleckstrin homology domain mutations in X-linked agammaglobulinemia. *Biochemistry* 1995; **34**: 1475–1481.
20 Saraiva MJM: Transthyretin mutations in health and disease. *Hum Mutat* 1995; **5**: 191–196.
21 Zurutuza L, Muller F, Gibrat JF *et al*: Correlations of genotype and phenotype in hypophosphatasia. *Hum Mol Genet* 1999; **8**: 1039–1046.
22 Bjorgo E, Knappskog PM, Martinez A, Stevens RC, Flatmark T: Partial characterization and three-dimensional-structural localization of eight mutations in exon 7 of the human phenylalanine hydroxylase gene associated with phenylketonuria. *Eur J Biochem* 1998; **257**: 1–10.
23 Waters PJ, Parniak MA, Nowacki P, Scriver CR: *In vitro* expression analysis of mutations in phenylalanine hydroxylase: linking genotype to phenotype and structure to function. *Hum Mutat* 1998; **11**: 4–17.
24 Erlandsen H, Stevens RC: The structural basis of phenylketonuria. *Mol Genet Metab* 1999; **68**: 103–125.
25 Jones TA, Bergdoll M, Kjeldgaard M: O: A macromolecule modelling environment. In Bugg CE, Ealick SE (eds). *Crystallographic and Modeling Methods in Molecular Design*. Springer-Verlag: New York, 1990, pp 189–195.
26 Dodson EJ, Winn M, Ralph A: Collaborative Computational Project, Number 4: providing programs for protein crystallography. *Methods Enzymol* 1997; **277**: 620–633.
27 Antonarakis SE, Nomenclature Working Group: Recommendations for a nomenclature system for human gene mutations. *Hum Mutat* 1998; **11**: 1–3.
28 Ellingsen S, Knappskog PM, Apold J, Eiken HG: Diverse PAH transcripts in lymphocytes of PKU patients with putative nonsense (G272X, Y356X) and missense (P281L, R408Q) mutations. *FEBS Lett* 1999; **457**: 505–508.
29 Erlandsen H, Bjorgo E, Flatmark T, Stevens RC: Crystal structure and site-specific mutagenesis of pterin-bound human phenylalanine hydroxylase. *Biochemistry* 2000; **39**: 2208–2217.
30 Daubner SC, Fitzpatrick PF: Site-directed mutants of charged residues in the active site of tyrosine hydroxylase. *Biochemistry* 1999; **38**: 4448–4454.
31 Teigen K, Froystein NÅ, Martinez A: The structural basis of the recognition of phenylalanine and pterin cofactors by phenylalanine hydroxylase. Implications for the catalytic mechanism. *J Mol Biol* 1999; **294**: 807–823.
32 Matthews BW: Structural and genetic analysis of protein stability. *Annu Rev Biochem* 1993; **62**: 139–160.
33 Gibbs BS, Benkovic SJ: Affinity labelling of the active site and the reactive sulfhydryl associated with activation of rat liver phenylalanine hydroxylase. *Biochemistry* 1991; **30**: 6795–6802.
34 Ellingsen S, Knappskog PM, Eiken HG: Phenylketonuria splice mutation (EXON6nt–96A→g) masquerading as missense mutation (Y204C). *Hum Mutat* 1997; **9**: 88–90.
35 Marvit J, DiLella AG, Brayton K, Ledley FD, Robson KJ, Woo SL: GT to AT transition at a splice donor site causes skipping of the preceding exon in phenylketonuria. *Nucleic Acids Res* 1987; **15**: 5613–5628.
36 Knappskog PM, Flatmark T, Aarden JM, Haavik J, Martinez A: Structure/function relationships in human phenylalanine hydroxylase. *Eur J Biochem* 1996; **242**: 813–821.
37 Kobe B, Jennings IG, House CM *et al*: Regulation and crystallization of phosphorylated and dephosphorylated forms of truncated dimeric phenylalanine hydroxylase. *Prot Sci* 1997; **6**: 1352–1357.
38 Hufton SE, Jennings IG, Cotton RGH: Structure/function analysis of the domains required for the multimerisation of phenylalanine hydroxylase. *Biochim Biophys Acta* 1998; **1382**: 295–304.
39 Waters PJ, Parniak MA, Akerman BR, Jones AO, Scriver CR: Missense mutations in the phenylalanine hydroxylase gene (*PAH*) can cause accelerated proteolytic turnover of PAH enzyme: A mechanism underlying phenylketonuria. *J Inher Metab Dis* 1999; **22**: 208–212.
40 Rennel D, Bouvier SE, Hardy LW, Poteete AR: Systematic mutation of bacteriophage T4 lysozyme. *J Mol Biol* 1991; **222**: 67–87.
41 Loeb DD, Swanstrom R, Everitt L, Manchester M, Stamper SE, Hutchison CA III: Complete mutagenesis of the HIV-1 protease. *Nature* 1989; **340**: 397–400.
42 Kleina LG, Miller JH: Genetic studies of the lac repressor. XIII. Extensive amino acid replacements generated by the use of natural and synthetic nonsense repressors. *J Mol Biol* 1990; **212**: 295–318.
43 Sternberg MJE, Bates PA, Kelley LA, MacCallum RM: Progress in protein structure prediction: assessment of CASP3. *Curr Opin Struct Biol* 1999; **9**: 368–373.
44 Knappskog PM, Flatmark T, Mallet J, Ludecke B, Bartholome K: Recessively inherited L-dopa-responsive dystonia caused by a point mutation (Q381K) in the tyrosine hydroxylase gene. *Hum Mol Genet* 1995; **4**: 1209–1212.
45 Ludecke B, Knappskog PM, Clayton PT *et al*: Recessive inherited L-DOPA-responsive parkinsonism in infancy caused by a point mutation (L205P) in the tyrosine hydroxylase gene. *Hum Mol Genet* 1996; **7**: 1023–1028.
46 Ludecke B, Dworniczak B, Bartholome K: A point mutation in the tyrosine hydroxylase gene associated with Segawa's syndrome. *Hum Genet* 1995; **95**: 123–125.
47 van den Heuvel LP, Luiten B, Smeitink JA *et al*: A common point mutation in the tyrosine hydroxylase gene in autosomal recessive L-DOPA-responsive dystonia in the Dutch population. *Hum Genet* 1998; **102**: 644–646.
48 Esnouf RM: An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J Mol Graphics* 1997; **15**: 133–138.
49 Merritt EA, Murphy MEP: Raster3D Version 2 – A program for photorealistic molecular graphics. *Acta Crystallogr* 1994; **D50**: 869–873.
50 Barton GJ: ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng* 1993; **6**: 37–40.