Dottorato in Biologia Computazionale e Bioinformatica - XXV ciclo

Università di Napoli Federico II

COMPUTATIONAL BIOLOGY AND BIOINFORMATICS AS TOOLS TOWARDS A BETTER UNDERSTANDING OF PHENYLKETONURIA

PhD Student: Carla Carluccio

Tutor: Prof. Francesco Salvatore

Co-Tutor: Prof.ssa Adriana Zagari

CONTENTS

Riassunto

La fenilalanina idrossilasi (PAH) è un enzima contenente ferro, che catalizza la conversione dell'aminoacido essenziale L-Phe in L-Tyr, utilizzando il cofattore 6R-L-eritro-tetraidrobiopterina (BH4) e ossigeno molecolare (Hufton et al. , 1995; Kaufman, 1993). Difetti nell'attività enzimatica, causati da mutazioni nel gene umano, portano ad una malattia ereditaria autosomica recessiva del metabolismo degli aminoacidi, nota come fenilchetonuria (PKU).

La funzione di una proteina è determinata dalla sua struttura 3D. Le mutazioni nel gene umano della PAH possono modificare la sequenza e la struttura di una proteina, alterando la sua funzione e causando quindi la malattia. Molti studi, infatti, hanno confermato una relazione diretta tra l'alterazione strutturale dell'enzima ed i fenotipi osservati. Ad oggi, più di 560 mutazioni del gene PAH sono stati identificati (vedi database Mutaztions PAH Analysis Consortium: http://www.pahdb.mcgill.ca/) e tali mutazioni sono localizzate in tutta la struttura 3D.

È stato dimostrato che PAH è un enzima tetramerico assemblato come dimero di dimeri. Ogni monomero adotta una struttura α/β ed è costituito da un dominio regolatorio N-terminale (RD) (residui 1-117), un dominio catalitico (residui 118-410), che comprende i siti di legame per il ferro, il substrato ed il cofattore, e un dominio di tetramerizzazione (residui 411-452) (Fusetti et al., 1998).

L'N-terminale RD di PAH contiene un dominio ACT (residui 33-111), che è costituito da un motivo strutturale composto da quattro β-strand e due α-eliche disposte in un βαββαβ fold. Il dominio ACT è stato identificato in alcune proteine che sono coinvolte nel metabolismo degli amminoacidi e delle purine, e che sono regolate da specifici aminoacidi (Aravind e Koonin, 1999). Generalmente, questi domini si associano in oligomeri di 2, 3 o 4 unità. Diversamente, nella struttura tetramerica della PAH i quattro domini ACT non si associano tra loro. Nelle proteine che contengono il domino ACT, i ligandi di solito si legano alle interfacce tra domini ACT e in prossimità di una specifica Gly (Gly46 in hPAH) (Grant, 2006).

La funzione regolatoria del dominio ACT di PAH è poco conosciuta e pertanto merita ulteriori studi. Perciò, in questo contesto si inserisce il mio lavoro di tesi.

I principali meccanismi regolatori della PAH includono l'attivazione da fenilalanina, inibizione da BH4, e un'ulteriore attivazione per fosforilazione (Hufton et al., 1995). In alcuni studi è stato proposto che la L-Phe si lega ad un sito allosterico sull'RD della PAH, diverso dal sito attivo, inducendo grandi cambiamenti conformazionali (Fitzpatrick, 2000; Kaufman, 1993;. Shiman et al, 1979).

Tuttavia, il putativo sito di legame allosterico per la L-Phe non è stato ancora identificato e il ruolo del dominio come modulo allosterico non è ancora chiaro.

Per investigare sulle basi strutturali dei difetti della proteina alla base della malattia, ho usato un approccio computazionale. L'approccio computazionale è molto efficace per capire come le variazioni genetiche possono modificare la struttura delle macromolecole biologiche e per far luce sulle relazioni struttura-funzione.

I risultati di questo lavoro di tesi contribuiscono a chiarire aspetti specifici sulla PHA e sulla PKU relativi a: (i) la stabilità conformazionale di alcuni mutanti di PAH che causano la malattia, (ii) le caratteristiche strutturali e dinamiche del dominio ACT isolato dell'enzima wild-type e di sei mutanti, (iii) le basi strutturali per la regolazione della PAH, attraverso l'identificazione *in silico* del putativo sito di legame allosterico per L-Phe.

Studi strutturali, effettuati tramite homology modelling, su selezionati mutanti patogeni di PAH hanno rivelato una diminuita stabilità conformazionale che può essere associata a possibili difetti di fold ed ad una ridotta attività dell'enzima, che causa la malattia.

Studi di dinamica molecolare sul dominio regolatore ACT isolato hanno evidenziato un comportamento dinamico specifico dell'enzima wild-type e dei sei mutanti. L'RD ha mostrato una transizione da una struttura iniziale aperta ad una struttura finale più chiusa, che porta alla riduzione dell'esposizione al solvente di una superficie idrofobica in cui il ligando L-Phe può legarsi al dominio. Inoltre, la correlazione, trovata tra l'accessibilità al solvente della superficie idrofobica e i moti globali intrinseci della proteina, suggerisce una relazione tra il legame con la L-Phe ed il comportamento dinamico globale dell'enzima. Nei mutanti, diversamente che nell'enzima wild-type, le simulazioni hanno rivelato cambiamenti nelle interazioni all'interno del core idrofobico dell'RD che possono influenzare significativamente la sua stabilità e, quindi, il suo ruolo regolativo.

Al fine di indagare sul sito di legame allosterico per la L-Phe, ho generato dei modelli per le interazioni L-Phe-PAH, attraverso studi di docking molecolare in combinazione a

simulazioni di dinamica molecolare sulla PAH dimerica. I modelli hanno confermato la mia ipotesi iniziale riguardo la posizione del sito di legame allosterico all'interfaccia dimerica RD/CD. Per la PAH umana, il legame con la L-Phe ha mostrato una riduzione della mobilità del dominio RD. L'analisi di questi modelli ha chiarito gli effetti strutturali del mutante G46S, che influenzano il corretto legame con la L-Phe.

In conclusione, questa tesi aggiunge ulteriori informazioni utili per una più profonda comprensione sia dei meccanismi di regolazione dell'enzima e sia dell'interconnessione tra le mutazioni e la malattia.

Summary

The phenylalanine hydroxylase (PAH) is an iron-containing enzyme, that catalyzes the conversion of the essential amino acid L-Phe into L-Tyr, utilizing the cofactor 6R-L-erythro-tetrahydrobiopterin (BH4) and dioxygen (Hufton et al., 1995; Kaufman, 1993).
Defects in the PAH enzymatic activity, caused by mutations in the human gene, result in an autosomal recessively inherited disorder of amino acid metabolism, known as phenylketonuria (PKU).
Many studies have confirmed a straightforward relation between 3D-structural alteration and the observed phenotypes. To date, more than 560 mutations in the *PAH* gene have been identified (see PAH Mutation Analysis Consortium database: http://www.pahdb.mcgill.ca/) and they are spread throughout the 3D-structure.
It has been shown that PAH is a tetrameric enzyme assembled as dimer of dimers. Each monomer adopts an α/β structure and is built up from an N-terminal regulatory domain (RD) (residues 1-117), a catalytic domain (residues 118-410), which includes binding sites for iron, substrate and cofactor, and a tetramerization domain (residues 411-452) (Fusetti et al., 1998).
The N-terminal RD of PAH contains an ACT domain (residues 33-111), which consists of a structural motif composed of four β-strands and two α-helices arranged in a βαββαβ fold. The ACT domain was identified in some proteins involved in amino acid and purine metabolism and regulated by specific amino acids (Aravind and Koonin, 1999). In these proteins, the ligands usually bind at the interfaces between ACT domains and in proximity of a specific Gly (Gly46 in hPAH) located in the loop between the first β-strand and the first α helix (Grant, 2006).
The regulatory function of the N-terminal ACT domain of PAH is poorly understood and hence deserves further investigation. The major regulatory mechanisms of the PAH include activation by phenylalanine, inhibition by BH4, and additional activation by phosphorylation (Hufton et al., 1995). In some studies it was proposed that L-Phe binds to an allosteric site on RD of PAH, different from the active site, inducing large conformational changes (Fitzpatrick, 2000; Kaufman, 1993; Shiman et al., 1979).
However, the putative allosteric L-Phe-binding site in PAH has not been identified yet and the role of the domain as an allosteric module is still unclear.

To get insight into the structural basis of the protein defects underlying the disease, I used a computational approach. The computational approach has been proven to be very powerful to understand how genetic variations modify the structure of biological macromolecules and to shed light on the structure-function relationships.

Here, my current results contribute to elucidate specific aspects of PAH and PKU related to: (i) the conformational stability of disease-causing PAH mutants, (ii) the structural and dynamical features of the isolated ACT domain of the wild-type enzyme and of six mutants, (iii) the structural basis for the regulation of PAH, through the identification *in silico* of the putative allosteric L-Phe-binding site.

Structural studies, by homology modelling, on selected disease-causing PAH mutants revealed a decreased conformational stability that can be associated to possible defects in the folding and to a reduced activity of the enzyme, leading to the disease.

MD studies on the isolated ACT regulatory domain highlighted a specific dynamic behaviour of the wild-type enzymes and of the six mutants. The RD showed a transition from an initial open to a final more closed structure, leading to the decrease of the solvent exposure of a hydrophobic surface where the L-Phe ligand can bind to the domain. Moreover, the correlation, found between the solvent accessibility of the hydrophobic surface and the global intrinsic motions of the protein, suggested a relationship between the L-Phe-binding and the overall dynamic behaviour of the enzyme. In the mutants, the simulations revealed changes in the interactions within the hydrophobic core of the RD that may significantly affect its stability and hence its regulatory role.

In order to investigate the allosteric regulatory binding site for L-Phe, I generated models for the L-Phe-PAH interactions, performing molecular docking studies combined with MD simulations on the dimeric PAH. The allosteric L-Phe-binding site was localized at the dimeric RD/CD interface, confirming my initial hypothesis on the site position. For the hPAH, the L-Phe-binding showed also to reduce the mobility of RD domain. The analysis of these models elucidated the structural effects of the disease-causing G46S mutant, which showed to significantly affect the correct binding of L-Phe.

In conclusion, this thesis adds further evidences toward a deeper understanding of the regulatory mechanisms of the enzyme and of the relationship between the mutations and the disease.

Abbreviations

3D	Three Dimensional
3PGDH	3-PhosphoGlycerate DeHydrogenase
AAAH	Aromatic Amino Acid Hydroxylase
ACT	from Aspartokinase, Chorismate mutase and prephenate dehydrogenase (TyrA), three proteins from the ACT family
ATP	Adenosine TriPhosphate
BH2	Dihydropterin
BH4	(6*R*)-L-*erythro*-5,6,7,8-tetrahydrobiopterin
CD	Catalytic Domain
DHPLC	Denaturant High Performance Liquid Chromatography
DSSP	Dictionary of Secondary Structure Protein
ED	Essential Dynamics
FMA	Functional Mode Analysis
GCK	GluCoKinase
H-bond	Hydrogen Bond
HPA	HyperPhenylAlaninemia
HPC	High Performance Computing
IARS	Intrinsic AutoRegulatory Sequence
MCM	Maximally Correlated Motion
MD	Molecular Dynamics
MHP	Mild HyperPhenylalaninemia
MODY2	Maturity Onset Diabetes of Young
MPKU	Mild PhenylKetonUria
MSA	Multiple Sequence Alignment

NMR	Nuclear magnetic resonance
PAH	PhenylAlanine Hydroxylase
hPAH	human PhenylAlanine Hydroxylase
rPAH	rat PhenylAlanine Hydroxylase
PDB	Protein Data Bank
PC	Principal Component
PCA	Principal Component Analysis
PDT	Prephenate DehydraTase
PKA	cAMP dependent protein kinase
PKU	classical PhenylKetonUria
RD	Regulatory Domain
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
SASA	Solvent Accessible Surface Area
SPC	Simple Point Charge
TH	Tyrosine Hydroxylase
THA	THienylAlanine
TPH	Tryptophan Hydroxylase
VDW	Van Der Waals
wt-hPAH	wild-type hPAH
wt-rPAH	wild-type rPAH

CHAPTER 1. *General introduction*

1.1 *The aromatic amino acid hydroxylases family*

The aromatic amino acid hydroxylases family consists of the non-heme iron-containing enzymes: phenylalanine hydroxylase (PAH, EC 1.14.16.1), tyrosine hydroxylase (TH, EC 1.14.16.2) and tryptophan hydroxylase (TPH, EC 1.14.16.4). These enzymes play an important role in the mammalian metabolism. The enzymes catalyze the hydroxylation of the respective aromatic amino acid L-phenylalanine (L-Phe), L-tyrosine (L-Tyr) and L-tryptophan (L-Trp) (Figure 1.1). They present a common catalytic mechanism involving the non-heme iron, the (6*R*)-L-*erythro*-5,6,7,8-tetrahydrobiopterin cofactor (BH4) and molecular oxygen (Flatmark and Stevens, 1999; Kappock and Caradonna, 1996). Dioxygen is used by the active site, containing a single reduced iron ion, to hydroxylate an unactivated aromatic substrate, concomitant with a two-electron oxidation of the cofactor to its quinonoid dihydropterin form. This form can also isomerize to L-*erythro*-7,8-dihydrobiopterin (BH2) which is converted back to BH4 by dihydrofolate reductase (DHFR) (Kappock and Caradonna, 1996) (Figure 1.1).

PAH, which is mainly found in liver, catalyzes the first and rate-limiting step in the catabolic pathway of L-Phe, by hydroxylating L-Phe into L-Tyr and resulting in the complete degradation of the amino acid (Flatmark and Stevens, 1999).

TH, which is found in the central nervous system and in the adrenal gland, catalyzes the first step in the formation of catecholamines by converting L-Tyr into L-3,4-dihydroxyphenylalanine (L-Dopa) (Fitzpatrick, 1999; Flatmark and Stevens, 1999). L-Dopa is decarboxylated by the aromatic L-amino acid decarboxylase and dopamine is the product. Dopamine can further be converted into norepinephrine and epinephrine, and these are all important hormones and/or neurotransmitters.

There are two isoforms of TPH: TPH1, mainly expressed in the pineal gland and the periphery, and TPH2, abundant in the brain (Walther and Bader, 2003). The reaction catalyzed by TPH is the first and rate-limiting step in the biosynthesis of serotonin and the initial and uncommitted step in the synthesis of melatonin, by hydroxylating L-Trp into 5-hydroxytrypthophan (5-HTP).

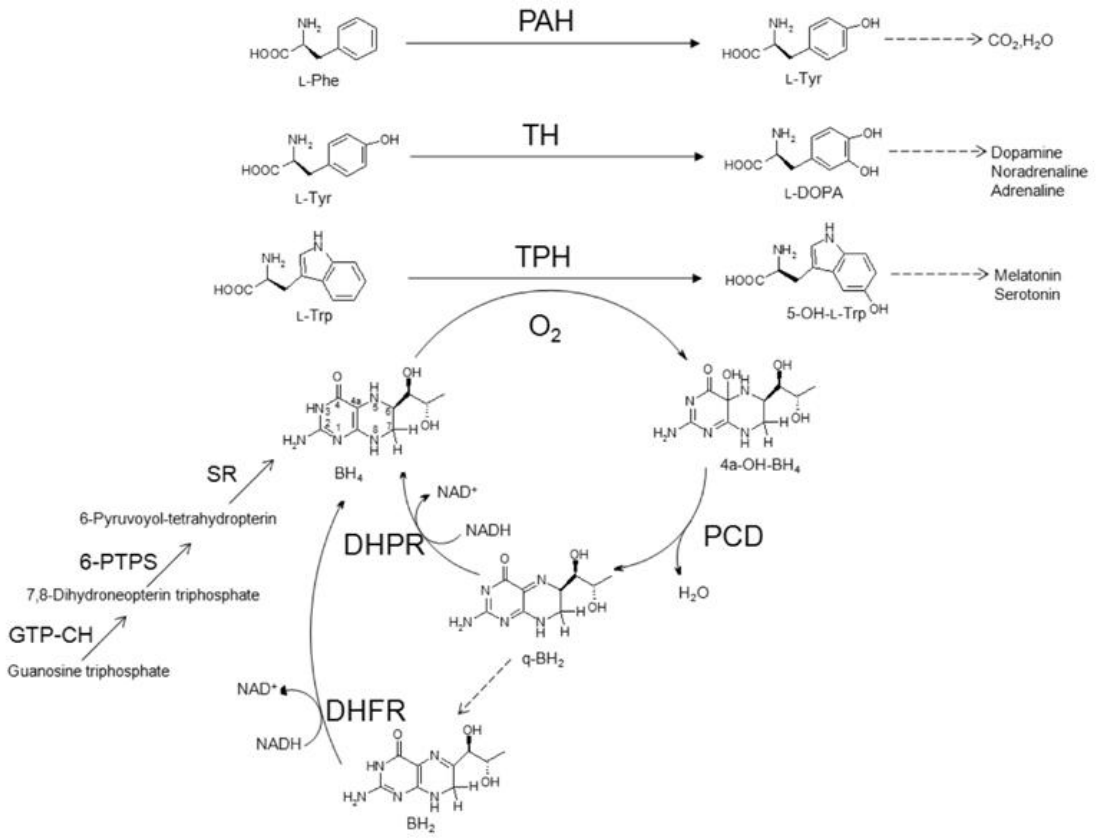

Figure 1.1: Scheme of the hydroxylation reactions catalyzed by the aromatic amino acid hydroxylases, using the (6*R*)-L-*erythro*-5,6,7,8-tetrahydrobiopterin cofactor (BH4) and molecular oxygen. The *de novo* BH4 biosynthetic pathway from guanosine triphosphate involves the enzymes GTP cyclohydrolase (GTP-CH), 6-pyruvoyl-tetrahydropterin synthase (6-PTPS) and sepiapterin reductase (SR).

The three enzymes share a high degree of sequence identity (65%) (Figure 1.2) and domain organization. In solution, as well as *in vivo*, all three hydroxylases form tetramers in which each monomer is composed of three structural and functional domains: an N-terminal domain with regulatory properties, a central catalytic domain where is located the active site, and a C-terminal tetramerization domain (Erlandsen and Stevens, 1999; Fusetti et al., 1998; Kobe et al., 1999). The latter domain, consisting of the last 20-23 amino acids, is responsible for the oligomeric state of the enzymes forming a domain-swapped tetrameric coiled-coil.

The active site, with the non-heme iron coordinated to a 2-His-1-carboxylate facial triad, is very similar in the three enzymes, that present 80% sequence similarity in the catalytic domains (Kappock and Caradonna, 1996). In contrast to the high degree of similarity found in the catalytic and oligomerization domains, the regulatory N-terminal

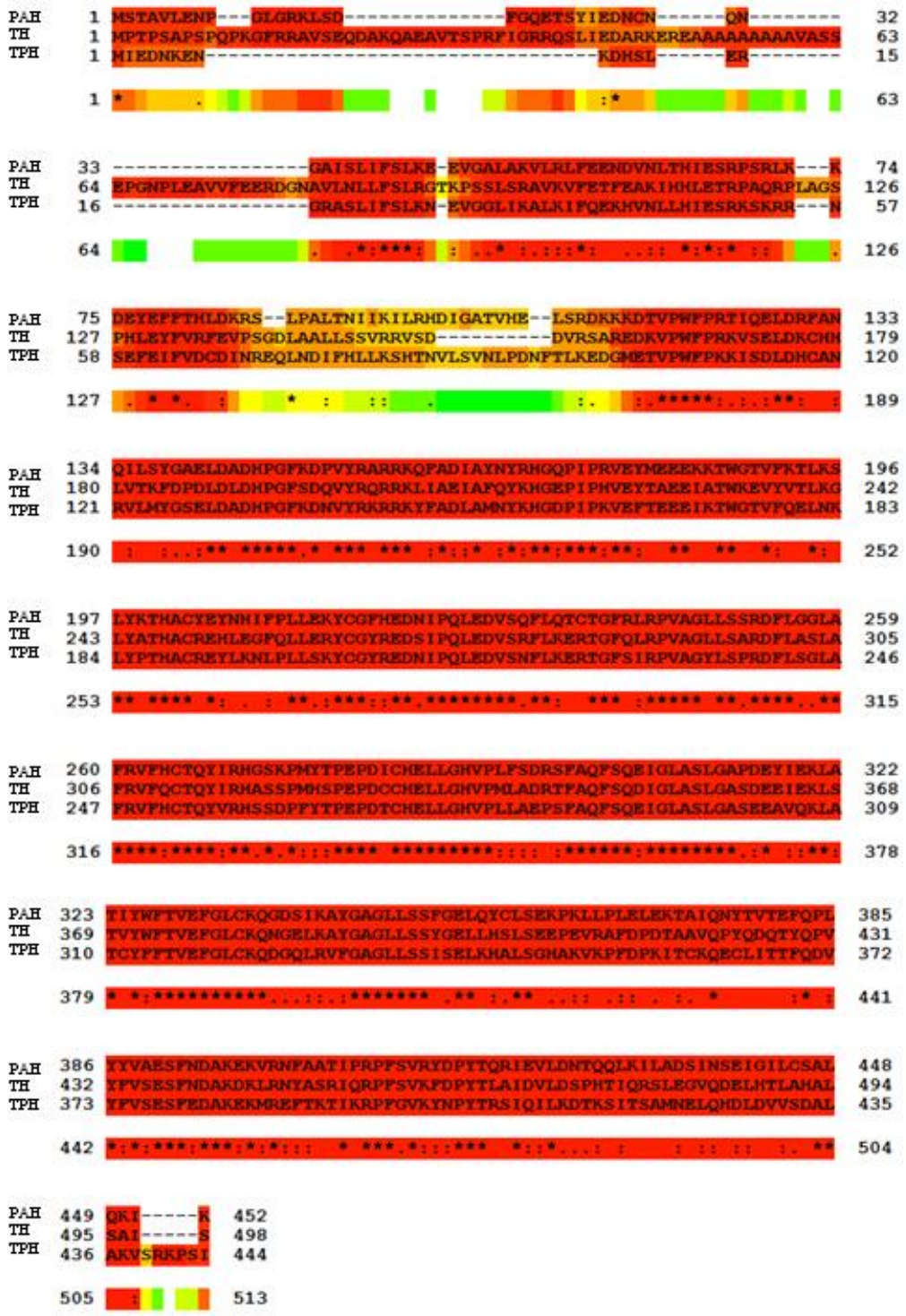

Figure 1.2: Multiple sequence alignment of human AAAHs. The sequences were aligned using T-Coffee (Notredame et al., 2000). The residue color scheme reflects the primary library support for the alignment of the considered residue on a scale between 0 (blue, poorly supported) and 9 (dark red, strongly supported). The asterisks below the sequences indicate identity, and two dots or one dot denote homolog properties, depending on the degree of similarity of the residues.

domain is highly variable in length and topology, reflecting different regulatory specific function of the three enzymes (Hufton et al., 1995).

1.2 *Human phenylalanine hydroxylase*

Human PAH (hPAH) is a catabolic enzyme that degrades the L-Phe from food and provides an endogenous source of L-Tyr to the organism by converting L-Phe, an essential amino acid that has to be included in the diet (Kalhan and Bier, 2008), into L-Tyr which therefore is a non-essential amino acid (Figure 1.1). Human PAH is present mainly in the liver, but is also expressed to some extent in the kidney and epidermis. The human *PAH* gene, mapped on chromosome 12q23.2 (Scriver, 2007; Woo et al., 1987), is composed of 171 kb organized into 13 exons. The full-lenght PAH cDNA (GenBank cDNA Reference Sequence U 49897) encodes a protein of 452 amino acids (about 52 kDa for each subunit).

1.2.1 *Human PAH three-dimensional structure*

Crystal structure information is available for several truncated forms of hPAH, with or without cofactor and substrate analogues, but the full-length structure has not been solved yet. In particular, the structure of the regulatory domain (RD) of hPAH is not available. Its closest homolog with known structural data (PDB code 1PHZ) is the RD from rat dimeric PAH (rPAH) (Kobe et al., 1999), which shares a 86% sequence identity of the RD with hPAH. A full-length composite model of human PAH can be constructed by superimposing the structure of rat PAH which includes the regulatory/catalytic domains and the structures of human dimeric and tetrameric PAH that include the catalytic/tetramerization domains (Figure 1.3 A,B).

It has been shown that hPAH is a tetrameric enzyme (Figure 1.3A) in equilibrium with a dimeric form (Bjorgo et al., 2001; Martínez et al., 1995). The hPAH tetramer ⇆ dimer equilibrium is dependent on pH (Kappock et al., 1995) and L-Phe concentration (Bjorgo et al., 2001; Martínez et al., 1995). A decrease of pH (to pH=6) and L-Phe concentration shifts the equilibrium towards the tetrameric form.

Each monomer adopts an α/β structure and is built up from an N-terminal RD (residues 1-117), a catalytic domain (residues 118-410), which includes binding sites for iron, substrate and cofactor, and a C-terminal oligomerization domain (residues 411-452) (Figure 1.3B).

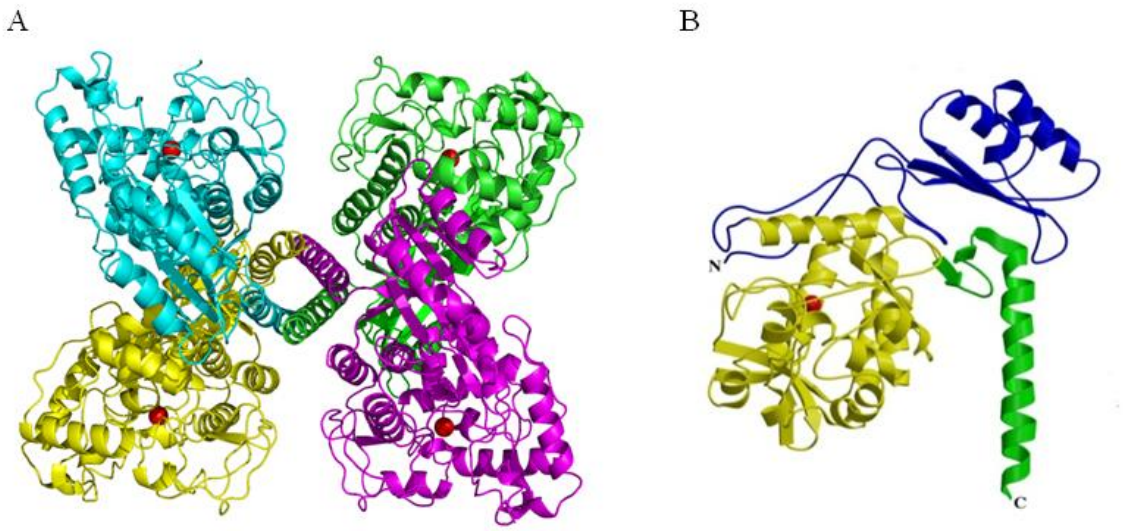

Figure 1.3: Composite model of the full-lenght (**A**) tetrameric and (**B**) monomeric PAH structures created by superimposing the regulatory/catalytic domain from rat crystal structure (1PHZ) and the catalytic/tetramerization domain from human crystal structure (2PAH). In (**A**) each monomer is colored separately, the iron is shown as a red sphere. In (**B**) RD is drawn in blue, the catalytic domain in yellow and the tetramerization domain in green.

The N-terminal RD of PAH contains an ACT domain (residues 33-111), consistng of a four-stranded antiparallel β-sheet flanked on one side by two short α-helices and on the other side by the catalytic domain (Kobe et al., 1999) (see section 1.3).

The catalytic domain has a basket-like arrangement of 14 α-helices and 8 β-strands. In this domain the active-site iron resides in an open and solvent accessible region (Erlandsen et al., 1997). There is also a narrow tunnel connected to the active-site pocket and this may be where the substrate is directed into the active site. The iron is located at the entrance of the pocket thus leaving space for binding of both the pterin and the substrate (Flatmark and Stevens, 1999). The iron is ligated in an octahedral manner by two histidine residues (His285 and His290), one glutamic acid residue (Glu330) and three water molecules (Figure 1.4).

The cofactor BH4 binds at the active site, in the vicinity of the iron ion, where it makes stacking interactions with Phe254 (Teigen et al., 1999; Teigen et al., 2004).

In the ternary crystal structures of substrate analogs (THA or NLE) the substrate was found to bind in the second coordination sphere of the iron with the aromatic ring stacking against the imidazole group of the iron-bound His285 (Andersen et al., 2002; Andersen et al., 2003).

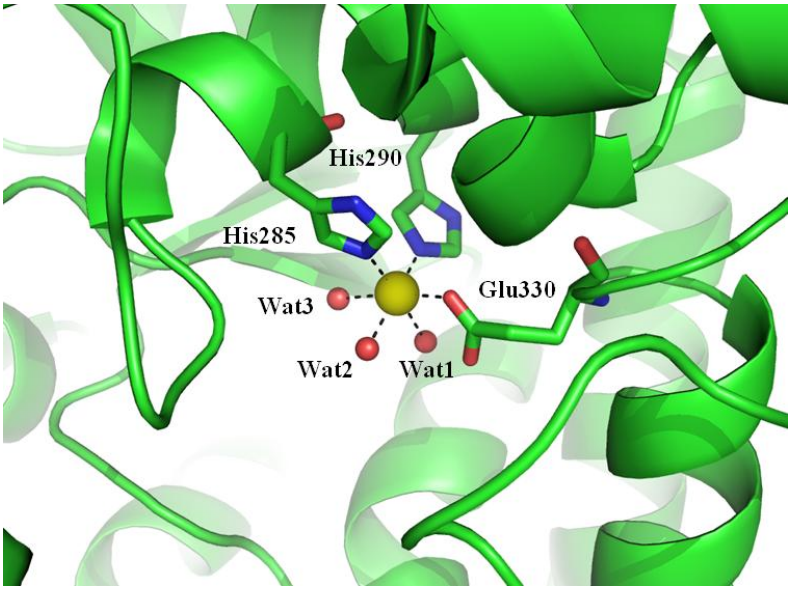

Figure 1.4: Close-up view of the iron binding site in hPAH. The iron ion is shown as a yellow sphere hexa-coordinated in an octahedral geometry to His285, His290, Glu330 (in stick) and three water molecules (Wat1, Wat2 and Wat3) represented as red spheres. The figure was prepared using the crystal structure of the binary cofactor complex (ΔN102/ΔC24-hPAH-Fe(II)·BH4) (PDB code: 1J8U) (Andersen et al., 2001) and the program Pymol (Schrödinger, 2009).

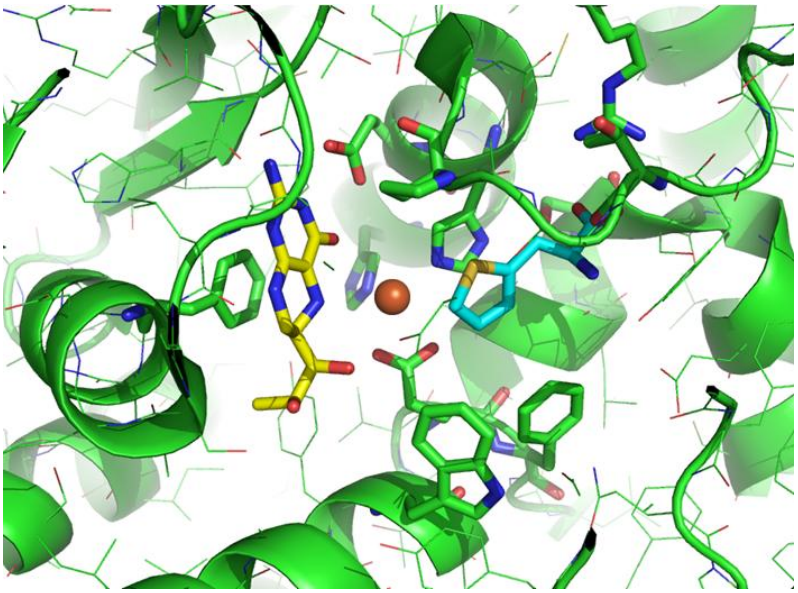

Figure 1.5: Close-up view of the BH4 and THA (analog of Phe) binding sites in hPAH. The iron ion is shown as a orange sphere, the BH4 in yellow stick and the THA in cyan stick. All residues involved in the binding of Fe, BH4 and THA (see text) are highlighted in green stick. The figure was prepared using the crystal structure of the ternary complex (PDB code: 1KW0) (Andersen et al., 2002) and the program Pymol (Schrödinger, 2009).

The substrate was found to interact through hydrogen bonds and hydrophobic interactions with some residue in the site (Andersen et al., 2002).

The short oligomerization domain includes a dimerization motif (411-427), consisting of two antiparallel β-strands, and a tetramerization motif (428-452), a single 40 Å long C-terminal α-helix responsible for forming the core of tetramer. The helix has a conserved heptad repeat of hydrophobic Leu-rich residues forming a leucine zipper that plays an important role in the tetramerization domain swapping mechanism. The tetramer core is formed by four α-helices, one from each subunit, tightly packed in an antiparallel coiled-coil motif (Fusetti et al., 1998).

The hPAH tetramer is not completely symmetrical, and is considered a dimer of two conformationally dimers (Fusetti et al., 1998). The two dimers within the tetramer adopt a different relative orientation, due to rotation around the hinge region Asp425-Gln429 (at Thr427) that connects the dimerization and tetramerization motifs, followed by an another rotation around Gly442 in the C-terminal tetramerization helix (Fusetti et al., 1998).

1.3 *Structure of N-terminal regulatory domain*

The RD has a flexible and quite instable structure, consistent with thermal denaturation studies that have shown that the RD unfolds at a lower temperature than the more compact catalytic domain (Thorolfsson et al., 2002). Structural information on this domain is just provided from the rat dimeric crystal PAH structures (PDB code 1PHZ, 2PHM) (Kobe et al., 1999). The rat dimeric crystal structure shows that the N-terminus of the RD extends across the active site opening. This first sequence stretch in the N-terminus (residues 1-33), different among the AAAHs, is called the intrinsic autoregulatory sequence (IARS) (Teigen and Martinez, 2003). The IARS partly controls the access of the solvent and substrates to the active site (Horne et al., 2002; Kobe et al., 1999) and includes the phosphorylation site (Ser16), which is involved in the activation of the enzyme (see section 1.4).

The structured part of RD (residues 34-110) is classified as an ACT domain, a structural motif composed of four β-strands and two α-helices arranged in a βαββαβ fold (Figure 1.5A). The ACT (from Aspartokinase, Chorismate mutase and prephenate dehydrogenase (TyrA), three proteins from the ACT family) domain is a small molecule-binding domain (SMBD) with a high sequence divergence and evolutionary

mobility (Anantharaman et al., 2001). The ACT domain was identified by sequence analysis of a set of proteins involved in amino acid and purine metabolism and regulated by specific amino acid (Aravind and Koonin, 1999).

In the available 3D structures, this domain was found either isolated or associated into oligomers of 2, 3 or 4 units (Figure 1.5B). Differently, in the tetrameric PAH structure the four ACT domains do not associate with each other. In many cases, the proteins containing the ACT domain are allosteric enzymes regulated *via* transmission of conformational changes upon ligand binding (Aravind and Koonin, 1999). In the ACT domain-containing proteins, the ligands usually bind at the interfaces between ACT domains (Figure 1.5B) and in proximity of a specific Gly (Gly46 in hPAH) located in the loop between the first β-strand and the first α helix (Grant, 2006).

The regulatory ACT domain is connected to the catalytic domain by a hinge bending region (residues 111-117), which is involved in domain movements occurring upon activation by the substrate (Li et al., 2010; Stokka et al., 2004). The regulatory ACT domain is also in contact with the catalytic domain of the adjacent chain in the dimer and with the oligomerization domain of its own chain.

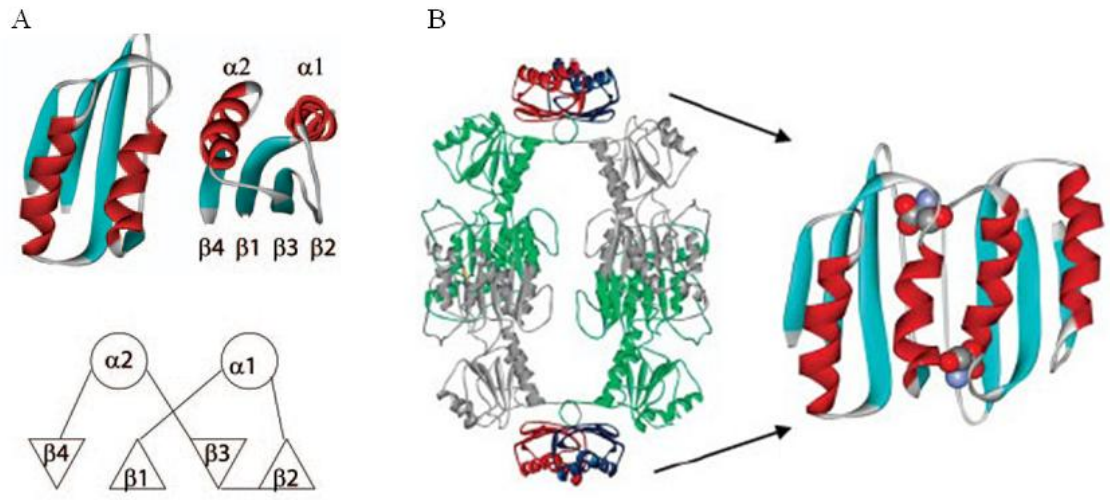

Figure 1.5: (**A**) A single ACT domain from *E. coli* D-3-phosphoglycerate dehydrogenase displaying the βαββαβ motif. In the schematic representation at the bottom, the triangles pointing up or down indicate the direction of the strand in the β sheet, with numbers starting from the N-terminus. (**B**) The *E. coli* D-3-phosphoglycerate dehydrogenase tetramer (PDB code 1PSD), with the subunits in green and gray and ACT domains colored blue and red. The arrows indicate the location of the ACT domains in the structure, with the ligand (L-Ser), in CPK format, bound at the dimer interface. Figure from Grant 2006.

1.4 *Regulatory function of PAH*

Human PAH is highly regulated in order to maintain the correct levels of L-Phe in the blood. The three main regulatory mechanisms of hPAH are i) activation by phenylalanine (Shiman and Gray, 1980; Shiman et al., 1990), ii) inhibition by BH4 (Døskeland et al., 1996), and iii) additional activation by phosphorylation (Shiman and Gray, 1980). Studies on truncated forms of hPAH partially or fully lacking the N-terminal domain have shown that the enzyme is already in an activated state (Knappskog et al., 1996a) and is not inhibited by the cofactor BH4 (Jennings et al., 2001), properties that are not exhibited by the full-length enzyme. Removal of a N-terminal fragment activates the enzyme, and a regulatory effect by this domain was then proposed (Iwaki et al., 1986).

1.4.1 *Activation by substrate*

The substrate L-Phe is an allosteric activator of the hPAH enzyme, inducing conformational changes (Kaufman, 1993). Preincubation with L-Phe results in a 5- to 6-fold activation of full-length hPAH (Bjorgo et al., 2001), a process caused by conformational changes in the tetrameric/dimeric enzyme (Flatmark and Stevens, 1999; Li et al., 2010; Stokka and Flatmark, 2003; Stokka et al., 2004). Movements of the RD with displacement of the IARS seem to be involved, resulting in an increased accessibility to the active site, as N-terminal truncated mutants are already in an activated state (Knappskog et al., 1996a; Stokka and Flatmark, 2003). This L-Phe activation results in a transition of the enzyme from a low activity, low affinity "T"-state to a high activity "R"-state (Shiman and Gray, 1980; Shiman et al., 1990). Furthermore, the full-length tetrameric form of hPAH binds the substrate L-Phe with positive cooperativity, that involves all four monomers (Bjorgo et al., 2001; Shiman and Gray, 1980; Shiman et al., 1990). Human PAH shows a non-hyperbolic (sigmoidal) dependence on substrate concentration for its equilibrium binding (Parniak and Kaufman, 1981) and steady-state kinetics (Kaufman, 1993). The substrate L-Phe binds with an affinity ($S_{0.5}$) of ~150 μM for eukaryotic PAH and a Hill coefficient (n_H) of ~2 (Kaufman, 1993; Teigen et al., 2007). The positive cooperative binding of L-Phe is physiologically important as a protection mechanism when the serum L-Phe concentration increases above normal levels, allowing L-Phe homeostasis in the blood by preventing HPA (Kappock and Caradonna, 1996). The activation and positive

cooperativity mechanism of the full-length hPAH enzyme is a complex process that is still not completely understood.

1.4.2 *Activation by phosphorylation*

Phosphorylation at Ser16 by the cyclic AMP dependent protein kinase A (PKA) is a post-translational modification that acts synergistically with L-Phe activation. The phosphorylated enzyme requires a lower L-Phe concentration for full catalytic activation, while L-Phe levels increase the rate of phosphorylation by PKA (Døskeland et al., 1996; Kaufman, 1993). In contrast, BH4 reduces the rate of phosphorylation, but this inhibition is prevented in the presence of L-Phe (Døskeland et al., 1996). Phosphorylation seems to induce a local conformational change at Ser16 in the IARS, which facilitates the access of L-Phe to the active site (Miranda et al., 2004; Miranda et al., 2002).

1.4.3 *Inhibition by cofactor*

The cofactor BH4 plays a role as co-substrate in the catalytic reaction, but has also a regulatory role. BH4 inhibits the enzyme by inducing a low-activity conformational state, inhibits L-Phe activation and reduces the rate of phosphorylation at Ser16 (Døskeland et al., 1996; Pey et al., 2004a; Pey et al., 2004b; Solstad et al., 2003; Teigen and Martinez, 2003). The cofactor BH4 inhibits hPAH in the absence of L-Phe by protecting the enzyme from unfolding and degradation (Mitnaul and Shiman, 1995), a stabilization process that plays a major role in the BH4-responsive PKU (Martínez et al., 2008). On the other side, recent studies demonstrate that the binding of BH4 to an activated form of PAH (in the presence of the substrate) results in a positive cooperativity indicated by a sigmoidal behaviour of data originating from continuous PAH activity assay (whereas the BH4-dependent kinetics of the non-activated enzyme followed the non-cooperative Michaelis-Menten model) (Gersting et al., 2010).

1.5 *Hyperphenylalaninemias*

Mutations in the human gene encoding the PAH enzyme result in the autosomal recessively inherited disease hyperphenylalaninemia (HPA). The resulting HPA phenotypes can range in severity as follow: mild hyperphenylalaninemia (MHP or HPA III) when plasma Phe levels of untreated subjects are between 120 and 600 μM/L; mild

phenylketonuria (MPKU or HPA II) when untreated plasma Phe levels are between 600 and 1200 μM/L; and classical phenylketonuria (PKU or HPA I) when untreated plasma Phe levels are > 1200 μM/L. The normal range of blood phenylalanine concentrations is 50-110 μM/L (Blau et al., 2010).

Impaired function of PAH results in the accumulation of high levels of blood plasma Phe and toxic concentrations in the brain (Erlandsen and Stevens, 1999). L-Tyr is the substrate for the biosynthesis of the thyroid hormone thyroxine, the neurotransmitter dopamine, the adrenal hormones and the pigment molecule melanin. Lack of L-Tyr and excess of L-Phe leads to various clinical manifestations such as mental retardation and decreased pigmentation. Additional symptoms include eczematous rash, autism, seizures, and motor deficits. Developmental problems, aberrant behaviour, and psychiatric symptoms often become apparent as the child grows (Blau et al., 2010).

Early diagnosis and prompt intervention has allowed most individuals with PKU to avoid severe mental disability (Blau et al., 2010). Nowadays, many countries include a test for HPA in neonatal screening programmes. To prevent mental retardation due to the build up of neurotoxic metabolites of L-Phe, patients with severe PKU must be treated with a low-L-Phe diet starting early in the life (Erlandsen and Stevens, 2001). This low-protein diet, supplemented with L-Phe-free mixture of amino acids, vitamins, minerals and other nutrients remains the most effective treatment for PKU (Blau et al., 2010; Giovannini et al., 2007). However, the diet is difficult to maintain and compliance if often poor, expecially in adolescents and pregnant women. About 1-2% of cases of HPA are due to mutations in the genes coding for enzymes involved in the biosynthesis or regeneration of the PAH cofactor BH4 (Thöny and Blau, 2006). BH4 stimulates PAH activity in about 20% of patients, and in those patients serves as a useful adjunct to the low-L-Phe diet, because it increases Phe tolerance and allows some dietary freedom (Blau et al., 2010). About 75% of PAH mutations, characterized by high residual enzyme activity, have been found to be associated with BH4-responsiveness, both *in vitro* (Erlandsen et al., 2004) and *in vivo* (Giovannini et al., 2007). The BH4-responsiveness in HPA has been the subject of numerous studies. Various mechanisms have been proposed to explain BH4-responsiveness, in particular: increased enzyme activity (Erlandsen et al., 2004; Muntau et al., 2002; Scriver, 2007), correction or compensation of the BH4 decreased affinity, protection toward catalytic inactivation and chaperone-like activity that, by stabilizing the protein, protects it from proteolytic

degradation (Erlandsen et al., 2004; Scriver, 2007). In addition, BH4 supplementation may restore the optimal concentration of BH4 cofactor in hepatocytes (Aguado et al., 2006). In some cases the addiction of BH4 influences the oligomerization equilibrium by shifting it toward a tetrameric quaternary structure (Cerreto et al., 2011).

Restoration of the function of the affected enzyme could be accomplished by a gene therapy approach (Blau et al., 2010; Cerreto et al., 2011). Enzyme substitution therapy with the plant derived phenylalanine ammonia lyase (PAL, E. C. 4.3.1.5) is an additional approach to treat PKU, that is currently in phase II clinical trials (ClinicalTrials.gov, 2010).

More than 560 disease-causing mutations in the *PAH* gene have been identified and registered in the Human PAH Mutation Knowledgebase (hPAHdb; *www.pahdb.mcgill.ca/*). The current set of PKU mutations consists of ~60% missense mutations; ~13% deletions; ~11% splice mutations; ~6% silent mutations; ~5% nonsense mutations and ~2% insertions. Mutations have been identified in all 13 exons and, due to the large number of mutations, most of the patients (approximately 75%) are heteroallelic for *PAH* mutations, being classified as compound heterozygous (Scriver and Waters, 1999).

Characterization analyses *in vitro* of several PAH missense mutations have identified at least three main groups of enzymatic phenotypes which differ in their kinetic behaviour and/or stability: (i) structurally stable mutations with altered kinetic properties, e.g. mutations at residues involved in substrate (L-Phe) or the pterin cofactor (BH4) binding; (ii) mutations with normal or almost normal kinetic properties, but reduced stability both *in vitro* and *in vivo*; and (iii) mutations affecting both kinetic and stability properties of the enzyme (Bjorgo et al., 1998; Gersting et al., 2008; Knappskog et al., 1996b; Waters, 2003; Waters et al., 2000).

Mutations are spread throughout the 3D structure, even though most of them fall in the catalytic domain. The mutated residues were classified on the basis of 3D-structural criteria into 5 diverse categories: 1) active site residues, 2) structural residues, 3) residues involving interdomain interactions within a monomer, 4) residues that interact with the N-terminal IARS, and 5) residues at the dimer or tetramer interface regions (Erlandsen, 2003).

Many studies have confirmed a straightforward relation between structural alteration and the observed phenotypes. The decrease in protein stability is the main molecular

pathogenic mechanism, so PKU is also considered a protein misfolded disease (Gregersen et al., 2001; Martínez et al., 2008).

1.6 *Aims of the thesis*

PKU is mainly caused by missense mutations in the phenylalanine hydroxylase gene. The effects of a specific mutation on protein function have been studied through a number of different approaches including *in vivo, in vitro* and *in silico* studies. All together these investigations have demonstrated that, in the majority of PKU cases, the loss of PAH function is due to decreased stability (Gersting et al., 2008; Knappskog et al., 1996a), increased susceptibility toward aggregation and degradation of PAH mutant proteins (Eiken et al., 1996; Gersting et al., 2008; Gjetting et al., 2001a), thermodynamic instability (Waters et al., 2000) and/or low folding efficiency (Pey et al., 2007). Structural analysis of wild-type enzyme and its mutants can contribute to a better understanding of the function of the enzyme and can provide a molecular explanation for the structural/functional effects induced by PAH mutations.

The computational approach has been proven to be very powerful to understand how genetic variations modify the structure of biological macromolecules and to shed light on the structure-function relationships. Therefore, in the framework of this doctorate, I used computational methods aimed to get insight into the structural basis of the protein defect underlying the disease.

PKU has been the subject of extensive research along many years, but there are still diverse aspects that demand a further study. In this perspective, the main goal of this thesis was to thoroughly elucidate some structural aspects. First, I carried out the mapping of some mutations to search for structural alterations (Chapter 2); then I studied the structural effects of the most frequent mutants falling in the ACT regulatory domain of PAH by molecular dynamics (MD) simulations in solution (Chapter 3). Finally, I investigated the open questions on the regulation function of the enzyme, through the identification of the putative allosteric Phe-binding site in the dimeric PAH model by molecular docking and subsequently by the analysis of the best scoring docking poses by MD simulations in solution (Chapter 4).

The study of the structure and the correlation between structure and function of proteins is important to understand how the different amino acid sequence will specify the folding and to understand how the structure assumed by the protein will determine the

biological function. In this context, to investigate the structural effects of the amino acidic replacements in the missense disease-causing PAH mutants, I examined *in silico* the alterations in the three-dimensional structure of two groups of mutations. First, a group of seven natural PAH variants (p.I65M, p.N223Y, p.R297L, p.F382L, p.K398N, p.A403V, and p.Q419R), that were, at the same time, biophysically-biochemically characterized *in vitro* (Cerreto et al., 2011). Then, a group of four novel natural PAH mutants (p.L321I, p.I406M, p.T418I, p.N426H), found in young HPA patients from Southern Italy (Chapter 2).

Phenylketonuria is considered a protein misfolding disease with loss of function (Gregersen et al., 2001). The misfolding of hPAH mutants does not directly affect the catalytic function, but impairs molecular motions involved in regulatory processes, substrate and cofactor binding and oligomerization assembly (Leandro et al., 2011). Particularly the N-terminal RD seems to play a crucial role in the instability and misfolding of the protein (Gersting et al., 2008). Therefore, I aimed to comprehensively study *in silico* the isolated ACT regulatory domain of the wild-type enzyme and of six selected disease-causing missense mutants (p.G46S, p.F39C, p.F39L, p.I65S, p.I65T and p.I65V) falling in such domain, by MD simulation in solution (Chapter 3). MD simulations are powerful tools for understanding internal motions in biological macromolecules. They are used to investigate motions that have particular functional implications and to reveal conformational changes as a function of time.

The N-terminal RD has a regulatory function, that is poorly understood and deserves further clarification. In particular the regulatory mechanism of the activation by the substrate phenylalanine is controversial. There is a debate regarding the ability of hPAH to bind L-Phe to an allosteric binding site, different from the active site, at the regulatory ACT domain (Fitzpatrick, 2000; Kaufman, 1993; Shiman and Gray, 1980). In order to investigate the allosteric regulatory binding site for L-Phe, I also aimed to extend the structural and dynamical studies on the dimeric hPAH and to perform a Phe-docking simulation of the dimer (Chapter 4). In this way, through the computational approach, I directly investigated the unclear regulation of the domain, by studing the motions and the flexibility of the enzyme upon ligand-binding.

Furthermore, this thesis contributes to clarify the relationship between the genetic mutations and the disease. All the performed simulations can inspire new experiments in the light of possible treatments of the disease.

CHAPTER 2. *Structural analysis of selected natural hPAH variants*

2.1 *Introduction*

HPAs represent the most common inherited disorder of amino acid metabolism, mostly caused by mutations in the *PAH* gene. About two-thirds of PKU mutations are caused by point mutations in which a single nucleotide is changed, resulting in a codon that codes for a different amino acid. A large proportion of all missense mutations, including many of the most common mutations found in patients, display stability and folding defects when they are expressed *in vitro* (Waters, 2003). These mutants are found to aggregate when expressed in prokaryote systems, where they appear as both soluble and insoluble aggregates (Leandro et al., 2010). The current understanding of PKU is related to a loss-of-function mechanism associated with increased degradation (Waters, 2006) and in some cases aggregation of the mutants.

Recombinant variant proteins containing prevalently missense mutations have been thoroughly characterized and analyzed *in vitro* and *in silico*. Each mutation exerts a distinct effect on the behavior of the enzyme and results in a specific biochemical phenotype. The effects of a specific mutation on protein structure and function have been studied by various approaches.

Several novel natural disease-causing mutants of hPAH were found (Daniele et al., 2008) in young patients of the "Dipartimento Clinico di Pediatria (DCP)" of the University of Naples "Federico II" and of the "Centro Screening Fenilchetonuria" of the hospital "S.S. Annunziata", Napoli. The resulting phenotypes of these mutations correspond to the mild form of HPA. Verification of clinical suspicion of HPA in these patients was carried out by the non-invasive molecular analysis of the *PAH* gene at the research institute CEINGE, Napoli. In our group, seven of these mutants have been deeply characterized, while for other four mutants found more recently we conducted a preliminary structural study.

2.1.1 *Study of seven selected natural disease-causing hPAH mutants*

A combined *in vitro-in silico* approach was used to analyzed the wild-type hPAH enzyme and seven selected natural disease-causing PAH mutants, in order to obtain further information about the molecular basis of HPAs and the mechanism of BH4 responsiveness (Paper I). The full-length wild-type and seven natural variants, namely p.I65M, p.N223Y, p.R297L, p.F382L, p.K398N, p.A403V, and p.Q419R were reproduced by *in vitro* mutagenesis. The activity, the oligomeric structure and the stability of the produced enzymes were determined to understand the impact of each mutation. In addition, the mechanism underlying BH4 responsiveness was investigated using p.A403V, which is one of the most frequent mutations in our geographic area, and is associated with a BH4-responsive mild phenotype. To complete this study, I performed the mapping of these mutations on the 3D hPAH structure to search for structural alterations caused by the amino-acidic replacement.

2.1.2 *Study of four novel natural disease-causing hPAH mutants*

Recently, four novel natural disease-causing mutants of hPAH, namely p.L321I, p.I406M, p.T418I and p.N426H, have been studied. Starting from wild-type human tetrameric phenylalanine hydroxylase structure (PDB code 2PAH), 3D structural models of the four novel mutants were generated, using template-based protein structure modelling to investigate the effects of the mutations on the structure.

2.2 *Experimental methods*

2.2.1 *PAH tetrameric structure*

No full-length structure, including all three domains, is available for PAH or any AAAHs in the PDB. There are various structures of catalytic domains with different combinations of pterin cofactors, Fe in diverse oxidation state and substrate analogs, mostly from human PAH. There are two structures (phosphorylated and dephosphorylated) of the regulatory domain in conjunction with the catalytic domain from rat PAH (Kobe et al., 1999). Only one structure of tetrameric hPAH with the catalytic domain and the entire oligomerization domain is available (Fusetti et al., 1998). Nevertheless, the available truncated structures have allowed the build up of composite full-length models, as e.g. in (Flatmark and Stevens, 1999; Thorolfsson et al., 2002). Our model (see Paper I) was constructed by combining various structures: tetrameric human PAH lacking residues 1-117 (PDB code 2PAH); dimeric truncated human PAH lacking residues 1-117 and 425-452 (PDB code 1KW0); and the dimeric rat PAH, from residue 19 to 427 (PDB code 1PHZ), which encompasses the regulatory domain. The composite monomeric model that included the three domains was built by superimposing the secondary structure elements of the respective catalytic domains of the above-mentioned structures. The structural superposition was performed using the program O.

2.2.2 *Homology modelling*

Evolution has produced diverse families of protein whose share similar 3D structure and similar sequences. In the post-genomic era, the structural genomics project aims to a systematic and large-scale effort towards structural characterization of all proteins encoded in the genomes. This aim will be achieved by a focused determination of selected protein structures by X-ray crystallography and NMR spectroscopy (one determined structure per sequence family), combined with accurate protein structure modelling techniques.

The homology modelling predicts the 3D structure of a given protein of unknown structure (target) on the basis of sequence similarity to one or more proteins of known structures (templates) (Bajorath et al., 1994; Blundell et al., 1987). This technique is based on the assumption that homologous proteins with similar sequences have similar 3D structures. The 3D structures of proteins in a family are more conserved than their

sequences (Lesk and Chothia, 1980). Moreover, even proteins that do not have sequence similarity can have similar structures.

The homology modelling is increasingly used because the number of unique structural folds that proteins adopt is limited (Zhang, 1997) and the number of experimentally determined new structures is increasing exponentially (Holm and Sander, 1996).

The structure-based modelling methods consist of four sequential steps (Sanchez and Sali, 1997) (Figure 2.1): template selection, target-template alignment, model building and model evaluation. If the model is not satisfactory, the steps can be repeated until a satisfactory model is obtained.

2.2.2.1 Template selection

The first step in homology modelling is to identify all protein structures related to the target sequence, and then to select those that will be used as templates. Templates can be found using the target sequence as a query for searching structure databases. The target sequence is compared with each of the database sequences independently, using programs as FASTA (Pearson, 1998) and BLAST (Altschul et al., 1990). To improve the sensitivity of the search, a multiple sequence comparisons can be performed. A widely used program for multiple sequences comparisons is PSI-BLAST (Altschul et al., 1997), which iteratively expands the set of homologs of the target sequence. For a given sequence, an initial set of homologs from a sequence database is collected, a weighted multiple alignment is made from the query sequence and its homologs. Generally, a higher overall sequence similarity between the target and the template sequence provides a better template. However, several other criteria should be considered when selecting the templates: the quality of the experimental template structure (the resolution and the R-factor of a crystallographic structure and the number of restraints per residue for an NMR structure), the choice of the template that contains the entire sequence to be modeled and similar conditions (e.g., solvent, pH, ligands and quaternary interactions). The use of several templates usually increases the model accuracy.

2.2.2.2 Target-Template Alignment

The alignment is very important in homology modelling. If the alignment is wrong, it will result in an inaccurate model. Therefore, it is sometimes important to use more

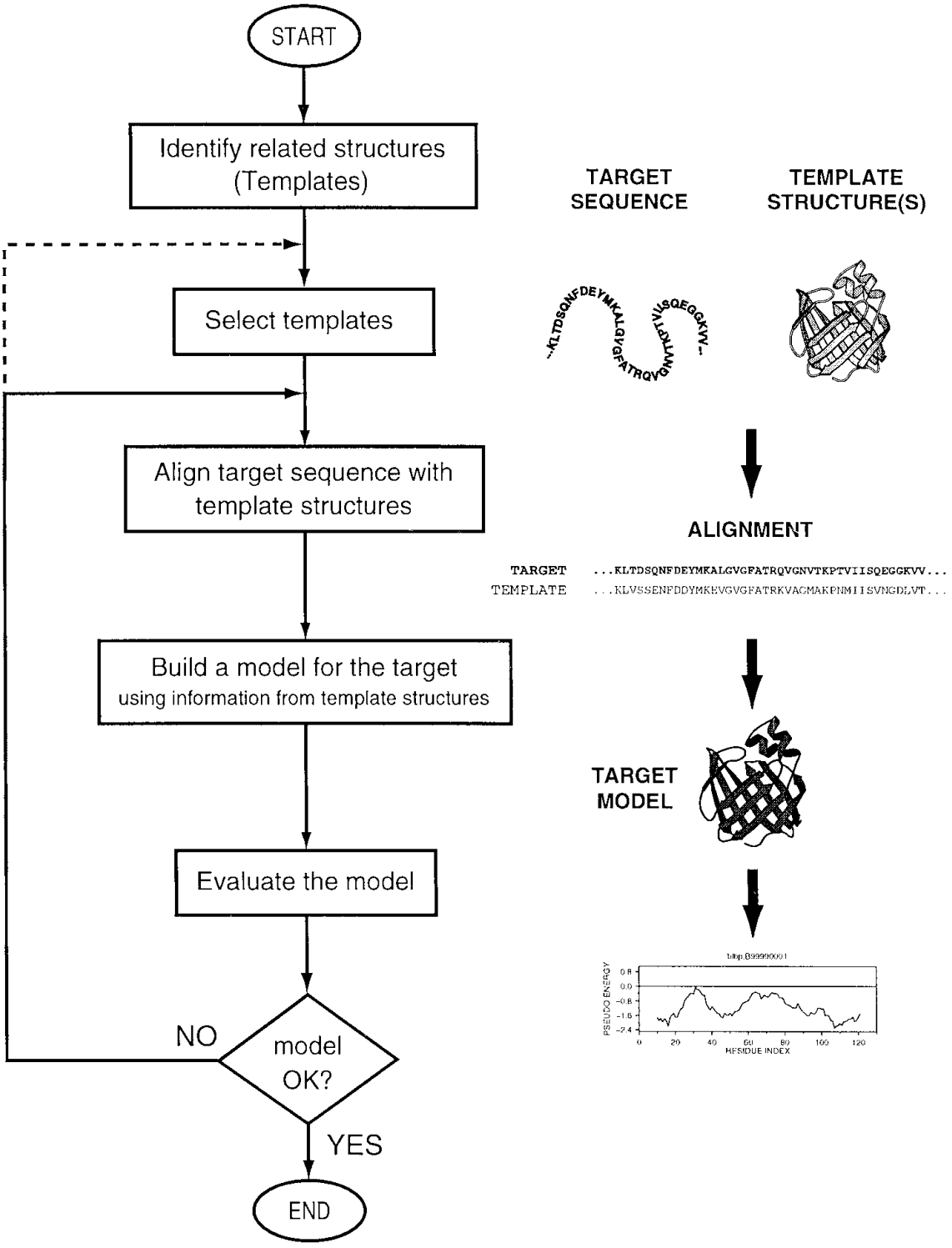

Figure 2.1: Scheme of the steps in the protein structure homology modelling. Figure from Martí-Renom 2000.

sequences and to build a multiple sequence alignment that provides more information. There is a great variety of protein sequence alignment methods. The frequently used programs for multiple sequence alignment are ClustalW, Blast, T-coffee, ect. The target and the template should be aligned taking structural information into account as much as possible. In principle, most sequence alignment and structure comparison methods can be coordinately used to obtain the best alignment.

2.2.2.3 Model building

The built alignment is used by software that, using the structure of the template as a reference, generate one or more 3D-models of the target protein.

The methods (Martì-Renom et al., 2000) for the construction of a model are:

- ***Rigid -bodies assembly*:** this method assembles a model from a small number of rigid bodies, obtained from the template. The frame-work is calculated by averaging the coordinates of the Cα atoms of structurally conserved regions in the template structure. The main chain atoms of each core region in the target model are generated by superposing the core segment from the template with the highest sequence similarity to the target on the framework. The loops are generated by scanning a database of all known protein structures to identify the structurally variable regions that fit the regions and have a compatible sequence. The side chains are modeled based on their intrinsic conformational preferences and on the conformation of the equivalent side chains in the template structures.
- ***Segment matching*:** this method is based on the finding that most hexapeptide segments of protein structure can be clustered into approximately 100 structural classes. Thus, comparative models can be constructed by using a subset of atomic positions from template structures as guiding positions, and by identifying in a representative database of all known protein structures and assembling hexapeptide segments that fit these guiding positions. This method can construct both main chain and side chain atoms and can also model insertions and deletions.
- ***Spatial restraints*:** this method generates many constraints or restraints on the structure of the target sequence, using the target as a guide. The restraints are generally obtained by assuming that the corresponding distances and angles between aligned residues in the template and the target structures are similar. In

addition, stereochemical restraints on bond lengths, bond angles, dihedral angles, and nonbonded atom-atom contacts are considered. The model is then derived by minimizing the violations of all the restraints.

Loop modelling is a mini-protein folding problem. The correct conformation of a given unstructured segment of a polypeptide chain is difficult to calculate. Structural variability, that results from substitutions, insertions and deletions of residues between members of the family, frequently occurs in the exposed loop regions. Thus, loops often determine the functional specificity of a given protein. Loop modelling methods include: *ab initio* methods (Bruccoleri and Karplus, 1987), based on a conformational search or enumeration of conformations in a given environment guided by a scoring or energy function; database search techniques (Chothia and Lesk, 1987), consisting in the finding of a main chain segment that fits the two stem regions of a loop; procedures that combine the first two basic methods (Martin et al., 1989).

As the loops, side chain conformations are predicted from similar structures and from steric or energetic considerations. Disulphide bridges are modeled using structural information from proteins in general and from equivalent disulfide bridges in related structures.

2.2.2.4 Model evaluation

The accuracy of 3D protein models is essential. The model can be evaluated as a whole as well as in individual regions. The first step in model evaluation is to assess if the model has the correct fold (Sanchez and Sali, 1998). Then, a basic requirement for a model is to have good stereochemistry (Martì-Renom et al., 2000). There are many model evaluation programs and servers. Some useful programs for evaluating stereochemistry are PROCHECK (Laskowski et al., 1993), AQUA (Laskowski et al., 1996), SQUID (Oldfield, 1992) and WHATCHECK (Hooft et al., 1996). The features of a model that are checked by these programs include bond lengths, bond angles, peptide bond and side-chain ring planarities, chirality, main-chain and side-chain torsion angles, and clashes between non-bonded pairs of atoms. In addition, many spatial features including packing, formation of a hydrophobic core, residue and atomic solvent accessibilities, spatial distribution of charged groups, distribution of atom-atom distances, atomic volumes and main-chain hydrogen bonds, are evaluated. Distributions

of these features have been compiled from high resolution protein structures, and large deviations from the most likely values have been interpreted as errors in the model.

2.2.3 *Building of the four novel hPAH mutants structural models*

The homology modelling was used to generate the 3D model structures of the four novel hPAH mutants p.L321I, p.I406M, p.T418I and p.N426H. The structure of the tetrameric hPAH (PDB code 2PAH) was selected as template. The template contains both the catalytic and the tetramerization domain, where the mutated residues under study are located. The sequences of the mutants and the sequence of the tetrameric wild-type template were aligned using Bodil software, a multi-platform package for biomolecular visualization and modelling (Lehtonen et al., 2004). Bodil aims to provide easy three-dimensional molecular graphics closely integrated with sequence viewing and sequence alignment editing. In particular, for each mutant, the mutated residue was replaced in the alignment file. The output file originating form Bodil is in a format readable by Modeller (Sali and Blundell, 1993), the program used to build mutant structural models. Modeller implements homology protein structure modelling by satisfaction of spatial restraints, that consists in extracting information about spatial relationships between residues of the template, and using this information as constraints for the construction phase of the model. It is assumed that the spatial restraints existing in the template are similar to those existing in the structure to predict; these restraints are used to create a mathematical function that describes the free energy of the protein. The optimal structure of the template will be obtained by varying its starting conformation so that the free energy becomes as low as possible. To create the models of the above-mentioned mutants, the input file of alignment originating from Bodil and the template structure were provided to Modeller. For each studied mutants, 200 models were generated. The program selects the most abundant conformers of the replaced residue and performs a simulated annealing procedure to optimize side chain conformations. The model that presented the best Modeller DOPE score was selected for further analysis. Structure analysis of the generated models was carried out by PDBSum server (Laskowski et al., 1997). For each models, the interatomic contacts, the Phi and Psi angles, and the hydrogen bonds were calculated using What If (Vriend, 1990), a molecular modelling package, used for the validation, that is specialized on working with proteins and the molecules in their environment like water, ligands, etc.

2.3 *Results*

2.3.1 *Mapping the seven natural disease-causing hPAH mutants*

A composite model of the full-length human enzyme was built and used as a template for structural analysis of the wild-type and the seven mutants. In Figure 2.2, the position of the seven mutations is drawn on the monomer structure. Mutations p.I65M and p.Q419R fall in the regulatory and tetramerization domain, respectively; mutations p.N223Y, p.R297L, p.F382L, p.K398N and p.A403V fall in the large catalytic domain (Figure 2.3). All mutations are far from the active and cofactor binding sites. All mutations, except for positions 65 and 403, are located on the surface of the molecule in hydrated regions.

Structural alterations were related to the residual enzyme activity, to the oligomeric equilibrium and to the thermal stability (it should be noted that in the denaturation process of hPAH upon heating, two transitions occur with melting temperatures T_{m1} and T_{m2}).

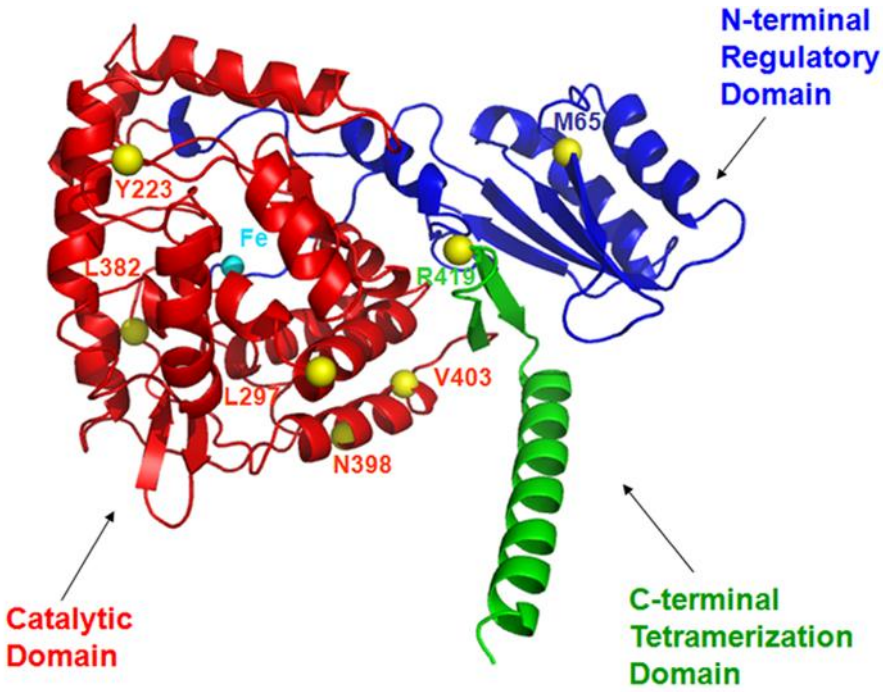

Figure 2.2: Full-length composite model of the PAH monomer structure and distribution of the seven selected HPA-related missense mutations studied. The N-terminal regulatory domain is shown in blue, the catalytic domain in red and the tetramerization domain in green. The Fe ion is shown as a cyan sphere at the center of the catalytic domain; the mutation sites are marked as yellow balls.

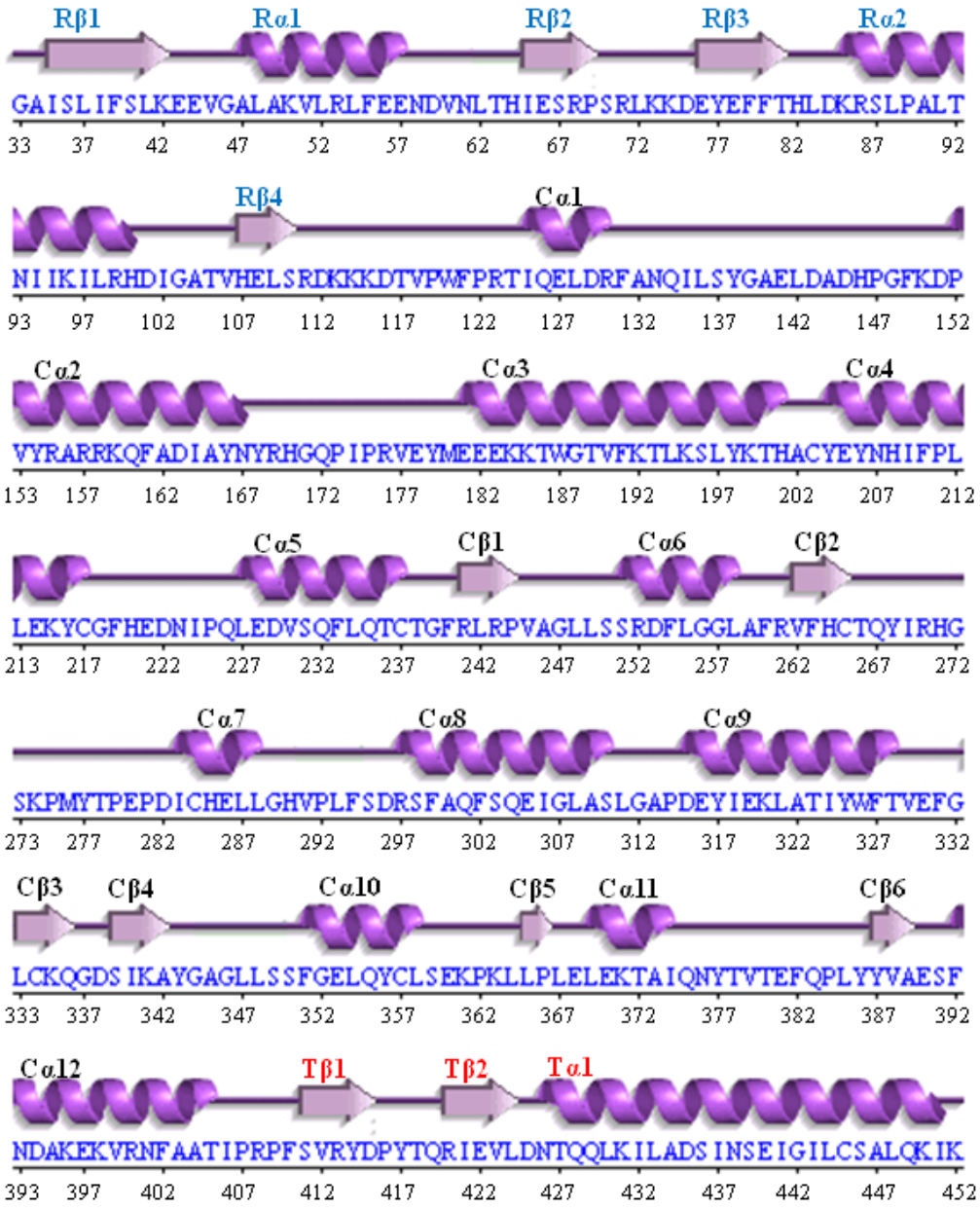

Figure 2.4: Primary and secondary structures of full-lengt hPAH model. Letters R, C and D are associated to the secondary structure nomenclature refer to Regulatory domain, Catalytic domain and Tetramerization. Figure from PDBsum server.

Ile65 is located in the Rβ2 strand in a hydrophobic environment of the regulatory domain. In the mutant enzyme p.I65M, the Ile change to a somewhat larger but differently shaped Met may distort the hydrophobic packing of the region. As a result, both the tertiary structure and the quaternary structure are perturbed. These data are in line with the dimer abundance (see Paper I).

Asn223 is located in the solvent exposed Cα4-Cα5 loop (Figure 2.4A). In mutant p.N223Y the structural perturbation is mainly due to tyrosine solvent exposure that may affect the function. It has been proposed that this residue is embodied in an intra-domain hinge bending region, that is involved in the conformational transition induced by substrate binding. Overall, its properties are similar to those of the wild-type enzyme, except that the dimeric form is favored over the tetramer.

Mutant p.R297L exhibits properties very similar to those of the wild-type enzyme. Interestingly, two other mutants reported in the database, e.g. p.R297C and p.R297H, which are also associated with a mild phenotype, indicate that mutations at this site are not critical.

Mutant p.F382L is the one that differs most from the wild-type enzyme. It occurs prevalently as a dimer (88.3%) (see Paper I). Moreover, the mutation (in the Cα11-Cβ6 loop) falls in a region rich in aromatic residues (Figure 2.4B). Phe forms various stacking aromatic interactions with the aromatic residues Tyr356 and Tyr277 in the wild-type enzyme. The replacement by Leu, despite the hydrophobic nature of this residue, breaks the aromatic network and results in marked destabilization. This is in line with larger ΔT_{m1} (19.6 °C) and ΔT_{m2} (14.0 °C) and with a lower inactivation temperature compared to the wild-type protein (see Paper I). This is the largest difference observed so far for a PAH mutant. This highlights that Phe in position 382 is crucial for the stability of this aromatic amino acid-rich region.

Lys398 is located in the Cα12 helix at the tetramer interface, although it is not directly engaged in inter-subunit interactions. The positively charged Lys is embedded in a negatively charged environment because of the presence of Asp394 and Glu390 (Figure 2.4C). The high dimer/tetramer ratio observed for mutant pK398N (75.0/17.2) is in line with tetramer destabilization. Also the structural stability was decreased. In fact, T_{m1} and T_{m2} were lower than that of the wild-type enzyme being 5.6 and 4.0 °C, respectively (see Paper I).

Q419 lies in the Tβ1-Tβ2 loop in a hydrated region at the dimer/tetramer interface. The substitution of Gln by an Arg generates unfavorable ionic interactions with a nearby Arg241 that destabilize the tetramer (Figure 2.4D). The thermal stability of p.Q419R is not significantly affected by the mutation.

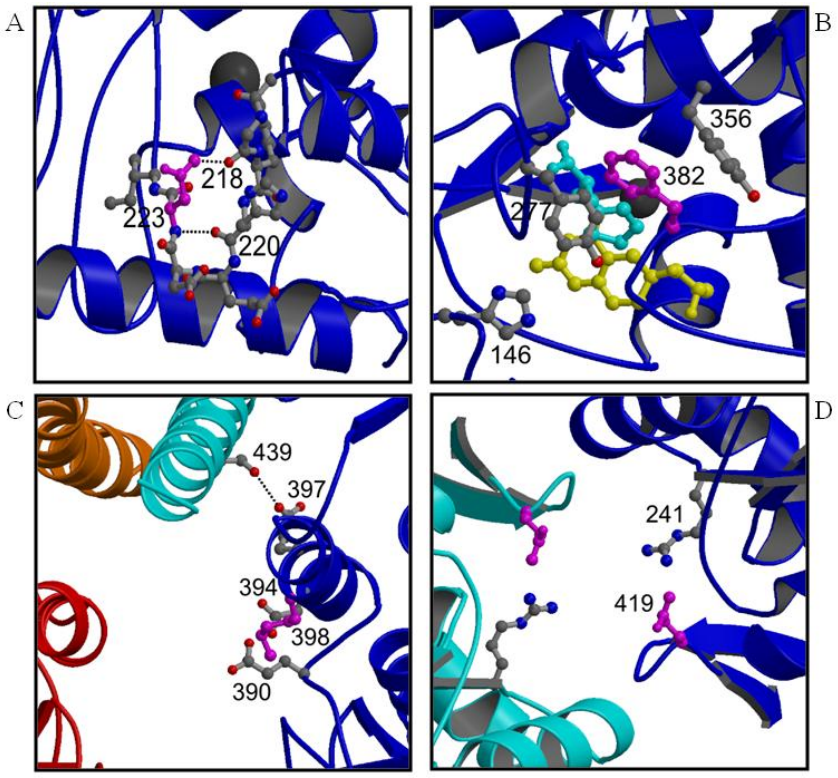

Figure 2.3: (**A**) Asn223 site. Segment 217–224 is drawn in ball-and-stick form. The stabilizing H-bonds displayed by Asn223 (magenta) in the β-turn 220–223 are represented by dotted lines. (**B**) Phe382 site. Phe382 (magenta) and nearby residues (Tyr356, Tyr277, His146) are highlighted in the substrate-bound structure (PDB code 1kw0). The substrate analogue, thienylalanine (cyan), and the BH4 cofactor (yellow) are also shown. (**C**) Lys398 site. The negatively-charged side-chains close to Lys398 (magenta) are shown. In the dimer, the H-bond between Glu397 and Ser439 is also drawn. (**D**) Gln419 site. Two Gln419 (magenta) residues at the dimer interface are shown with the positively-charged Arg241 residues nearby.

2.3.2 *Structural analysis of four novel hPAH mutants*

The structure models of four mutants, namely p.L321I, p.I406M, p.T418I and p.N426H, were generated *in silico* using the tetrameric hPAH structure (PDB code 2PAH) as template. Leu321 and Ile406 are located in the catalytic domain, in Cα9 helix and in the loop connecting Cα12 helix to Tβ1 strand, respectively; while Thr418 and Asn426 are located in the tetramerization domain, in the Tβ1-Tβ2 β-hairpin and in Tα1 helix respectively (Figure 2.3, 2.5).

A comparative analysis of the overall structure of each mutant model with the wild-type shows no significant changes. In particular, all mutated residues conserve secondary structure assignment of respective residues in the wild-type structure. Table 2.1 shows the main structural features of the four studied mutations.

Table 2.1: Features of the four studied PAH mutations.

Exon	Mutation	Other mutations described on the same position	Enzyme domain / Secondary structure position	Structural contacts in WT / Consequent alterations in mutated enzyme
9	p.L321I	None	Catalytic / α-helix H9 (315-326)	L321 is situated in a hydrophobic core involving helices H6, H8, H9 and H12 / Loss of a favorable interaction.
12	p.I406M	I406T	Catalytic / Loop preceding tetramerization domain (406-410)	I406 is situated in a hydrophobic core involving helices H6, H8, H9 and H12 / Mild structural alteration
12	p.T418I	T418P	Tetramerization / Loop (416-419) connecting two β-strands	T418 is situated in the loop connecting strands of β-sheet C in the dimerization motif / two hydrogen bonds are lost, one hydrogen bond weakens
12	p.N426H	None	Tetramerization / α-helix H13 (426-450)	N426 is situated in the α-helix H13 responsible for the formation of the tetramer / one hydrogen bond is lost

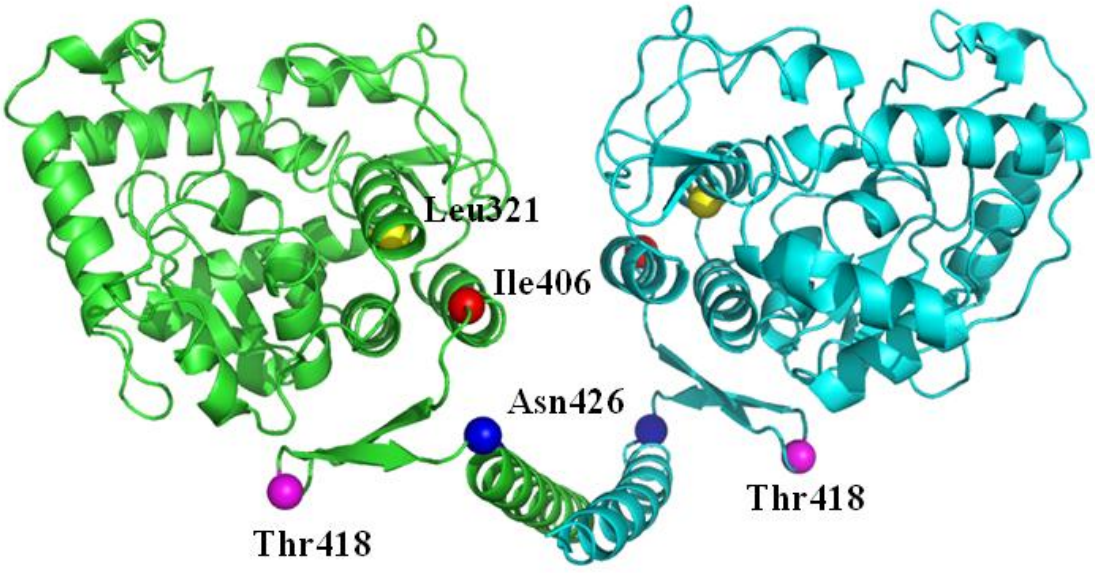

Figure 2.5: Dimeric model of the hPAH structure generated *in silico* and distribution of the four HPA-related missense mutations studied. The chain A in the dimer is shown in green and the chain B in cyan. The mutation sites are marked as balls in diverse color. Color code: Leu321 yellow, Ile406 red, Thr418 magenta, Asn426 blue.

Leu321 is located in Cα9 helix of the catalytic domain, in the hydrophobic core involving residues of Cα6, Cα8, Cα9 and Cα12 helices. In the mutant p.L321I some hydrophobic interactions are lost, others are formed; but in general there are no substantial changes in the hydrophobic asset of the core, where the mutated residue is located (Figure 2.6). In particular the number of interactions of residue 321 with Leu255 and Phe402 increases, while the number of interactions between residue 321 and Ile306 decreases. Leu is a γ-branched residue and has a major steric hindrance in the final part of the side chain, while Ile is a β-branched residue and has an increased steric hindrance in the first part of the side chain. Thus the Ile side chain, compared to Leu side chain, is located closer to residues Leu255, Ile306 and Phe402. Furthermore, in the mutant the interaction between Leu321 and Ile406, that occur in the wild-type enzyme, is lost. This result may be significant if we consider that Ile406 is located in the loop between catalytic and tetramerization domains. The interaction between Leu321 and Ile406 can contribute to the stability of the loop and can keep it in the proper orientation. So disrupting this interaction can destabilize the loop in the mutant structure.

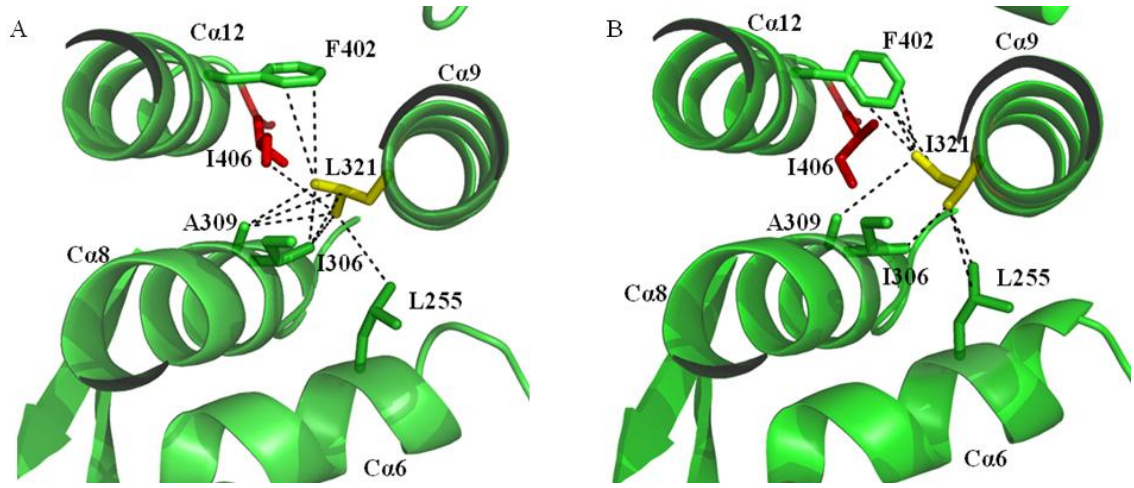

Figure 2.6: Hydrophobic interactions that involve atoms of residue 321 (**A**) in the wild-type structure (PDB code 2PAH) and (**B**) in the mutant p.L321I model structure. The residue 321 (in yellow) and the residues with which it interacts (including residue 406 in red) are shown as stick.

In p.I406M mutation, the substitution of an apolar residue into another apolar residue occurs. Ile406 is situated in the hydrophobic core of the catalytic domain, in the loop preceding the tetramerization domain. In the mutant I406M, some hydrophobic interactions of the residue 406 (with Tyr317, Leu321 and Phe402) change, others

interactions (with Ala309) are preserved. Moreover, three weak interactions with Ala313, Pro314 and Ala403 are lost in the mutant (Figure 2.7). Analyzing all these hydrophobic interactions, a rearrangement of the hydrophobic region is observed. Overall the number of favorable hydrophobic interactions decreases in the mutant; while the pattern of hydrogen bonds is conserved. The mutation causes a change of orientation of some close residues and leads to a reorganization of the region where is located.

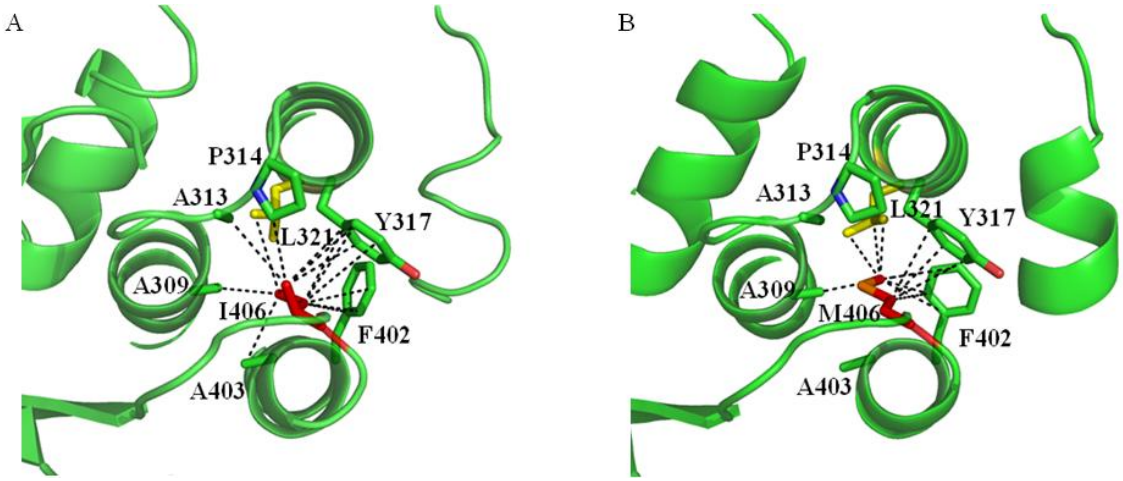

Figure 2.7: Hydrophobic interactions that involve atoms of residue 406 (**A**) in the wild-type structure (PDB code 2PAH) and (**B**) in the mutant p.I406M model structure. The residue 406 (in red) and the residues with which it interacts (including residue 321 in yellow) are shown as stick.

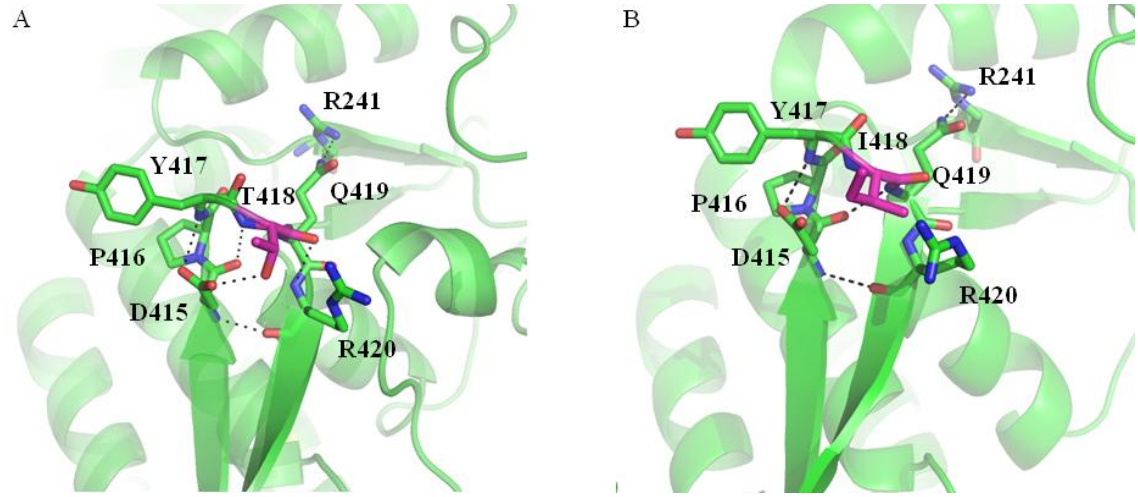

Figure 2.8: Hydrogen bonds in the loop of the dimerization motif (**A**) in the wild-type structure and (**B**) in the mutant p.T418I model structure. The loop residues are shown as stick, with the mutated residue 418 in magenta.

In the p.T418I mutant there is the substitution of a polar residue into an apolar residue. Thr418 is located in the tetramerization domain, in the loop (416-419) connecting the two β-strands forming the dimerization motif. The hydrogen bonds of all residues of the whole region have been examined (Figure 2.8). Both in wild-type and in mutant structures, the amino group of residue 415 binds the carbonyl oxygen of Arg420, the carbonyl oxygen binds the amino group of Gln419 and the OD1 oxygen binds the amino group of Tyr417. A hydrogen bond between side chains of residue 415 and Thr418 is lost in the mutant; this helps to stabilize the residues involved in the bond and its lack may cause a destabilization of the entire loop. Furthermore, a hydrogen bond between Gln419 and Arg241 (in the catalytic domain) is weaker in the mutant structure with respect to the wild-type (Figure 2.8). This bond can help to approximate the catalytic domain to the tetramerization domain, and a change in distance of this bond can affect the relative orientation of the two dimers leading to a destabilization of the tertiary and quaternary structure.

The p.N426H mutation substitutes a polar amino acid into a polar amino acid with a $pK_a \approx 6,5$. At physiological pH, relatively small shifts in pH will change the average charge of His. The alterations in the structure of this mutant can result from the presence of His that may have a positive charge on its side chain, in a region where there is an uncharged residue in the wild-type. Moreover, the alterations in the structure can result from the greater steric hindrance of the imidazole ring in the side chain. PAH is not a symmetric homotetramer but a dimer of dimers, and one of the main differences between subunits of the two diverse dimers lies in the orientation of the long C-terminal α-helix of the tetramerization domain. Particularly, the differences result from a change in backbone conformation at the hinge region that connects the β-hairpin to the helix, accompanied by another rotation of the helix. Asn426 is located in the region that differs in the two subunits constituting the two conformationally diverse dimers; for this reason, the substitution in position 426 on both A and B subunits of the dimeric hPAH structure model was analysed. In the chain A, Asn426 is the first residue of C-terminal α-helix (426-450), instead, in the chain B, is located in the loop preceding this α-helix (Figure 2.9). In chain A, a hydrogen bond with the amino group of L430 is conserved in the mutant, while the H-bond formed by the oxygen in the side chain of Asn426 with F410 is lost in the mutant, because of the lack of the acceptor of the hydrogen bond. The loss of this H-bond may have some effects on the stability of the tertiary and quaternary

structure of the enzyme, as it contributes to hold together the dimerization motif and the α-helix responsible for the formation of the tetramer. Moreover, the amino group of the side chain of N426 can bond the OD1 oxygen of D425; also this H-bond in the mutant is lost, because of the lack of the donor of the hydrogen bond. In the subunit B, the situation is quite different. In fact, in chain B of the wild-type structure, the carbonyl oxygen of N426 forms a hydrogen bond (3.13Å) with the amino group of Q429 (Figure 2.10), instead of the amino group of L430, as in chain A (Figure 2.9); moreover, in chain B there is no hydrogen bond involving N426 and F410, as in chain A. Also in chain B of the mutant structure, the carbonyl oxygen of H426 makes a hydrogen bond with the amino group of Q429 (Figure 2.10).

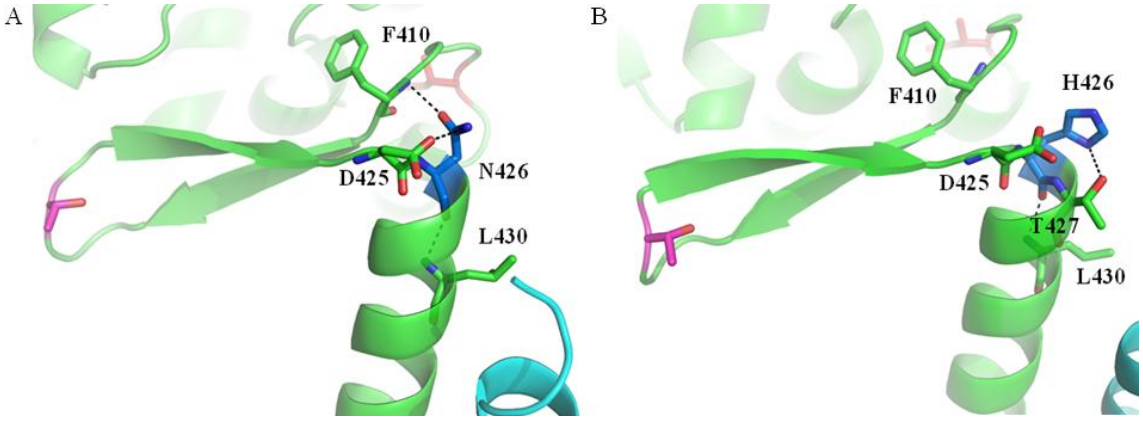

Figure 2.9: Hydrogen bonds of residue 426 in the chain A of (**A**) the wild-type structure and (**B**) the mutant p.N426H model structure. The residues involved in the H-bond are shown as stick with the mutated residue 426 in blue. In the figure are also drawn the other two mutated residues 418 (stick magenta), 406 (stick red) and part of the C-terminal helix of the adjacent subunits (chain B) in cyan.

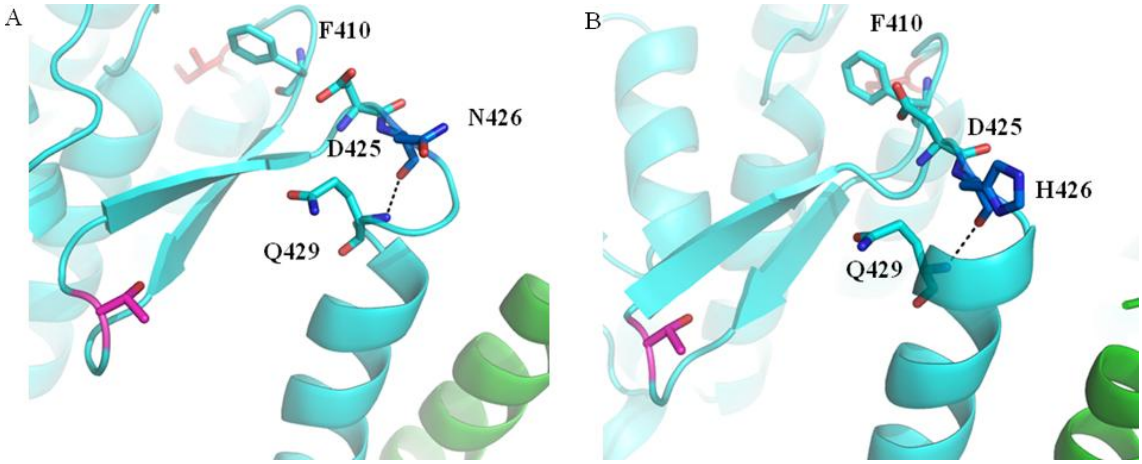

Figure 2.10: Hydrogen bonds of residue 426 in the chain B of (**A**) the wild-type structure and (**B**) the mutant p.N426H model structure. The residues involved in the H-bond are shown as stick with the mutated residue 426 in blue. In the figure are also drawn the other two mutated residues 418 (stick magenta), 406 (stick red) and part of the C-terminal helix of the adjacent subunits (chain A) in green.

2.4 Discussion

The aim of this structural analysis is to provide an explanation of the structural effects of the disease-causing studied hPAH missense mutations, aided by a three-dimensional visualization.

Previous studies on structure/function correlations showed that the disease-causing mutations are spread throughout the protein structure, with a certain number localised at dimer/tetramer interfaces. This finding highlights the relevance of interactions among subunits for the correct functioning of PAH (Andersen et al., 2002; Erlandsen et al., 1997). The structural examination of the two groups of hPAH mutants, here studied, revealed that the substituted residues are neither directly involved in the active site nor in the cofactor binding site. Overall, the analysis of the mutant models showed no dramatic changes in the structure. Accordingly all these mutations generate a mild phenotype. In details, the analysis showed structural differences in the regions where the mutated residues are located.

Mutations at positions 65, 321, 406 perturbs the hydrophobic packing of the regulatory and the catalytic domains; the mutations at positions 223 and 382 perturb the local structure of regions that are involved in substrate-induced molecular motions; whereas mutations at position 297, 398, 418, 419 and 426 disturb the dimer/tetramer interfaces. All the substitutions of the first group of seven mutants, that were also characterized *in vitro*, result in enzymes with decreased residual catalytic activity (to 38-69%) compared to the wild-type enzyme and affect the oligomerization state, the thermal stability and the folding of PAH protein.

In conclusion, the structural alterations do not directly affect the binding site of either the substrate or the coenzyme, rather they interfere with favorable interactions at the hydrophobic core, at the molecular surface and/or at its quaternary structure, and hence may destabilize structural integrity. The decreased conformational stability of the studied hPAH mutants can be associated to possible defects in the folding and to a reduced activity of the enzyme, leading to the disease.

CHAPTER 3. *Structural features of the regulatory ACT domain in the wild-type and in six hPAH mutants*

3.1 *Introduction*

The aim of this part of my thesis was to get further insights into the structural properties of the isolated human wild-type PAH RD and into the structural effects of the most frequent natural missense mutations on this domain. In PAH the RD has some unique biological, structural and physicochemical properties, that may explain why mutations in this domain affect the stability of the enzyme. First, the 3D structure of the rPAH RD has an overall high crystallographic B-factor and the 3D structure of the hPAH is not determined yet, consistent with the low stability of this domain. Then, the regulatory function of the RD is poorly understood and deserves further clarification.

The RD has been classified as an ACT domain homolog. The ACT domains are frequently found regulating metabolic enzymes in an allosteric manner. The allosteric effector is often an amino acid that binds to an interface formed by two ACT domains and in proximity of a specific Gly (Gly46 in hPAH) located in the loop between the first β-strand and the first α helix (Grant, 2006). Instead, no contacts are formed between the four ACT domains in the hPAH tetramer.

The major regulatory mechanisms of the PAH include activation by L-Phe (hereafter referred as to Phe) pre-incubation (Hufton et al., 1995). Tetrameric wild-type PAH displayed substrate activation and positive cooperativity for Phe binding, involving all three functional domains and all four subunits in the holoenzyme (Kaufman, 1993).

The literature contains conflicting models to explain these data. A possible explanation is that the activation mechanism results from the homotropic binding of Phe at the active site and the RD is involved in cooperativity, through its interactions with the catalytic and oligomerization domains (Liberles and Martinez, 2009; Liberles et al., 2005; Thorolfsson et al., 2003; Thorolfsson et al., 2002). In a number of studies it was instead proposed that Phe binds to an allosteric site on RD of PAH, different from the active site, inducing large conformational changes (Fitzpatrick, 2000; Kaufman, 1993; Shiman et al., 1979). In support of the latter model, it has been shown that when the RD is removed, the hPAH is no longer activated by preincubation with the substrate

(Knappskog et al., 1996a). Recently, NMR spectroscopy provided strong evidence for Phe-binding site in the isolated RD (Li et al., 2011). Moreover, it has been proved that the GAL (46-48) and IESRP (65-68) sequence motifs, conserved in the ACT domain family, are involved in Phe binding at the PAH RD (Gjetting et al., 2001b). A Phe bound structure of the protein prephenate dehydratase (PDT) from *Chlorobium tepidum TLS* is available (PDB code 2QMX), where the ligand-binding region in the ACT domain contains the GAL-IESRP motif (Tan et al., 2008). While Phe binds PDT at the dimer interface of two ACT domains, the GAL-IESRP motif in PAH lies at the interface between the RD and the catalytic domain of the adjacent subunit. Together with the structural similarities to the ACT regulatory domains of the above-mentioned allosteric enzymes, it was proposed that Phe binds PAH between the RD and the interacting catalytic domain, near the sequence binding motif (Hufton et al., 1995; Kobe et al., 1999).

In spite of all these evidences, the putative allosteric Phe-binding site in PAH has not been identified yet and the role of the domain as an allosteric module is still unclear. In this study, I analysed for the first time the intrinsic structural and dynamical properties of the isolated RD of the hPAH by molecular dynamics (MD) simulations in solution, in the light of the presence of a putative allosteric Phe-binding site. Moreover, I investigated the structural and dynamical effects induced by six selected missense disease-causing mutations found in young patients (p.G46S, p.F39C, p.F39L, p.I65S, p.I65T and p.I65V) (see PAH Mutation Analysis Consortium database: http://www.pahdb.mcgill.ca/) and located in the RD of the hPAH enzyme.

Most of the this work has been performed in a joint collaboration with Dr. Fraternali and Dr. Fornili at King's College, London.

3.2 *Experimental methods*

3.2.1 Molecular Modelling of the hPAH ACT domain and its mutants

Models of the ACT domain (residues 33-111) of the wt-hPAH enzyme and hPAH mutants were generated *in silico* with the program Modeller 9v8 (Sali and Blundell, 1993), using the rat wt-rPAH crystal structure as template (PDB code 1PHZ) (Kobe et al., 1999). The sequence identity percentage between the RD domain of wt-hPAH enzyme and the RD domain of wt-rPAH is 86%. The template and the target sequences were aligned using Bodil (Lehtonen et al., 2004). Subsequently, the alignment files were used as input in Modeller and 200 models of each studied forms were generated. For the mutants, the mutated residues were replaced in the alignment file. The program selects the most abundant conformers of the replaced residue and performs a simulated annealing procedure to optimize side chain conformations. The model that presented the best Modeller DOPE score was selected for the MD simulations.

The same procedure was used to generate the model of the rat PAH enzyme used to test the stability in wt-rPAH of an alternate conformation of the L2 region (Figure 3. 1B). In this case, two templates were used: the rat crystallographic structure of PAH for residues 33-57 and 67-111, and the ACT domain (fragment 339-408) of D-3-phosphoglycerate dehydrogenase from *Escherichia coli* (PDB code 1PSD) for residues 58-66.

Furthermore, a tetrameric model of wt-hPAH was built with Modeller 9v8, using as template the regulatory domain from rat crystal structure (PDB code 1PHZ) and the catalytic/tetramerization domain from human crystal structure (PDB code 2PAH).

3.2.2 Molecular Dynamics Simulations

The aim of computer simulations of molecular systems is to compute *macroscopic* behavior from *microscopic* interactions. MD simulations solve Newton' s equations of motion for a system of N interacting atoms:

$$m_i \frac{\partial^2 \boldsymbol{r}_i}{\partial t^2} = \boldsymbol{F}_i \qquad i = 1 \cdots N$$

Where $\mathbf{F}_i$ is the total force acting on the i-th atom, m_i is its mass, $\partial^2\mathbf{r}_i/\partial t^2$ is its acceleration, $\mathbf{r}_i$ is its position. The forces are the negative derivatives of a potential function V (r_1, r_2, . . . ,r_N):

$$\boldsymbol{F}_i = -\frac{\partial V}{\partial \boldsymbol{r}_i}$$

The equations are solved simultaneously in small time steps. In order to solve numerically the equations of motion, we need to be able to calculate the forces acting on the atoms, and these are usually derived from a potential energy that represents the complete set of 3N atomic coordinates. The potential energy function, also called force field, used in most of the standard molecular simulation programs for biological systems consist of:

$$V_{tot} = V_{bonds} + V_{angles} + V_{dihedrals} + V_{vdw} + V_{elect}$$

It takes form of the summation of different additive terms that correspond to bond distances stretching (V_{bonds}), bond angles bending (V_{angles}), bond dihedral or torsion angle ($V_{dihedrals}$), van der Waals potential (V_{vdw}) and electrostatic potential (V_{elect}). The first three terms are considered to be the intramolecular bonding interactions, each term involves a multiplet of atoms connected by chemical bonds. The other two terms represent the non-bonded interactions between atoms.

Molecular dynamics is a very useful and powerful tool. It can provide a wealth of detailed information on the structure and dynamics of peptides and proteins. However, it suffers certain limitations. First, the method is computationally very demanding and depending on the size of the system simulation; times are currently limited to hundreds of nanoseconds or a few microseconds at most. The possibility to observe certain properties is directly related to the quality of the force field and, whether or not it has been parameterized for the system simulated. The quality of the force field is especially critical in the simulation of proteins. Proteins are in general only marginally stable. The difference in free energy between the folded and unfolded form is in the order of 10-20 k_bT, which corresponds to the energy associated with the formation of a couple of hydrogen bonds in vacuum. The force field thus needs to be very accurate to discriminate between different conformations. However, it is questionable whether an empirical force-field can achieve the required accuracy especially when important effects such as polarization of the atoms by their environment is not taken into account by the electrostatic potential. The most used force fields in molecular dynamics are CHARMM, AMBER and GROMOS.

MD simulations are usually performed under periodic boundary conditions (PBC), to minimize boundary effects and to mimic the presence of the “bulk" environment. In this approach, the system is surrounded with replicas of itself in all directions, to yield an infinite periodic lattice of identical cells. When a particle moves in the central cell, its periodic image in every another cell moves accordingly. As one molecule leaves the central cell, its image enters from the opposite side without “filling" the cell boundary.
When Newton’s equations of motion are integrated the total energy is conserved (adiabatic system) and if the volume is held constant the simulation will generate a microcanonical ensemble (NVE). However, this is not always very convenient. Other statistical ensembles, such as canonical (NVT) and isothermal-isobaric (NPT) ensembles, better represent the conditions under which experiments are performed than the standard microcanonical ensemble. Moreover, with the automatic control of temperature and/or pressure, slow temperature drifts that are an unavoidable result of force truncation errors are corrected and also rapid transitions to new desired conditions of temperature and pressure are more easily accomplished.

3.2.3 MD Simulations of the PAH models and the template

MD simulations were performed with the GROMACS (Van Der Spoel et al., 2005) package using the GROMOS 43a1 force field (van Gunsteren et al., 1996). Each of the initial models generated *in silico* and the wt-rPAH crystal structure were solvated in boxes containing simple point charge (SPC) water molecules (Berendsen et al., 1981). Simulations were carried out with periodic boundary conditions and a 2-fs time step. All the bonds in the protein were frozen with the LINCS method (Hess et al., 1997), while SETTLE (Miyamoto and Kollman, 1992) was used for water molecules. The Berendsen algorithm (Berendsen et al., 1984) was applied for temperature and pressure coupling. The particle mesh Ewald method (Essmann et al., 1995) was used for the calculation of the electrostatic contribution to non-bonded interactions with a direct sums cut-off of 14 Å. The same cut-off was used for van der Waals interactions. The systems were energy minimized with 1000 steps of the steepest descent method and subsequently with 1000 steps of conjugate gradient method. After the minimization, harmonic positional restraints (with an initial force constant of 4.8 kcal/mol/$Å^2$) were imposed onto the protein heavy atoms and gradually turned off in 360 ps, while the temperature was increased from 200 to 300 K at constant volume. The system was then equilibrated for 2

ns without restraints in NPT conditions (T=300 K, p=1 bar). Finally, NPT MD simulations were performed for 50 ns. Two replicas were run for each system. MD trajectories were analysed using GROMACS analysis tools. Images were produced with the Visual Molecular Dynamics software (Humphrey et al., 1996) and PyMOL (Schrödinger, 2009).

3.2.4 Analysis tools

The displacement of the atoms or group of atoms along the trajectory could be estimated calculating the Root Mean Square Displacement (RMSD). The RMSD of a set of N atoms at time t, with respect to a reference conformation (e.g. initial conformation), is:

$$RMSD(t) = \sqrt{\frac{\sum_{i=0}^{N} \left|\vec{r}_i(t) - \vec{r}_i^{0}\right|^2}{N}}$$

Where $\left|\vec{r}_i(t) - \vec{r}_i^{0}\right|$ is the displacement of the i^{th} atom at time *t* from the reference position r_i^0. An increase of the RMSD indicates that the protein moves to a conformation different from the initial structure and thus suggests an incomplete sampling or a conformational change.

Similar to the previous analysis is the calculation of the fluctuations of an atom or a groups of atoms, this calculated according:

$$RMSF_i = \sqrt{\left\langle (\vec{r}_i - \langle \vec{r}_i \rangle)^2 \right\rangle}$$

Where $\vec{r}_i$ is the position vector of the i^{th} atom and the brackets stand for a temporal average. This analysis give clues about the most mobile regions of the protein.

The secondary structure content was measured with DSSP (Kabsch and Sander, 1983), while the solvent accessible surface area was calculated with the GROMACS utility g_sas (Eisenhaber et al., 1995).

Correlated motions were identified by performing Essential Dynamics (ED) (Amadei et al., 1993) on each MD trajectory. Principal components (PCs) were generated by diagonalizing the covariance matrix of C^{α} positions. Correlation webs, connecting C^{α} atom pairs with a correlation (per atom normalized covariance) higher than 0.3 were generated with Dynatraj (Barrett et al., 2004). A Combined ED analysis (van Aalten et al., 1995) was also performed on the trajectory obtained by concatenating the 16

trajectories obtained from all the simulated systems (2 replicas x (wt-hPAH + wt-rPAH + 6 hPAH mutants)).

To further investigate the accessible collective motions of the RD, we sampled alternative ensembles of structures for all the models with the tCONCOORD method (Seeliger et al., 2007). tCONCOORD allows a computationally efficient sampling of conformational transitions of a protein based on geometrical constraints as determined from the initial coordinates and interaction types. Under-wrapped hydrogen bonds are detected and modelled as unstable. Each model was first energy-minimized using the OPLS-AA force field (Jorgensen and Tirado-Rives, 1988) with 250 steps of the steepest descent algorithm. Ensembles of 1000 structures were then generated using standard tCONCOORD parameters (Seeliger et al., 2007). The similarity of the MD and tCONCOORD correlated motions from ED confirmed that all the relevant loop collective motions are already sampled by MD.

To investigate interactions with the solvent during the simulations, we generated the water density maps g(**r**) around each construct (De Simone et al., 2005; Lounnas and Pettitt, 1994; Lounnas et al., 1994). The maps were calculated at discrete points **r** defined by a 0.5-Å spaced rectangular grid around the solute. Each structure of the last 2 ns of the trajectory was superimposed to a reference to remove the overall roto-translational motion of the protein. The number density of the water oxygen atoms was then averaged at each grid point over the MD snapshots extracted every 0.1 ps and normalized by the bulk density evaluated in the 6-8 Å shell around the solute. The hydration sites were then identified as local maxima of the density map with g(**r**) > 1 and used to define the atomic hydration score S_{hyd}^{atom} as previously described (Fornili et al., 2012). Atoms with a high S_{hyd}^{atom} score are either surrounded by many maxima or close to a few maxima with a high density.

The relationship between the RD collective dynamics and the putative Phe binding site was investigated using the functional mode analysis (FMA) (Hub and de Groot, 2009). FMA aims to detect the collective motion that is maximally correlated with a particular protein functional property (Maximally Correlated Motion or MCM). The MCM is usually expressed as a linear combination of Principal Components (PCs) derived from ED. We selected the FMA approach where the MCM coefficients are determined by maximizing the Pearson correlation coefficient between the time evolution of the chosen functional property (solvent accessibility of the putative Phe binding site) and

the projection of the trajectory onto the MCM. For each MD simulation, the MCM was expanded in the essential space of the C^{α} atoms, composed of the first *n* PCs accounting for the 90% of the total fluctuation. The first 45 ns were used as model building set, while the remaining 5 ns were used as cross-validation.

3.3 *Results*

In silico models were generated for the ACT regulatory domain (residues 33-111) of wt-hPAH enzyme and of the six hPAH mutants, G46S, F39C, F39L, I65S, I65T and I65V. The model structures consist of a common α-β sandwich motif with a four-stranded antiparallel β-sheet flanked on one side by two short α-helices (Figure 3.1A-C). In particular, the loop-containing regions L42-V45 (L1), D59-H64 (L2), S70-D75 (L3) and H82-V90 (L4) are highlighted in Figure 3.1A and 3.1B. Two replicas of MD simulations were performed onto the generated models and the wt-rPAH template. During the 50-ns long simulations, the systems show no large deviations from the initial structures and at the end of the simulations all trajectories reach a plateau within 2-3 Å in the RMSD from the initial structure calculated over all C^{α} atoms.

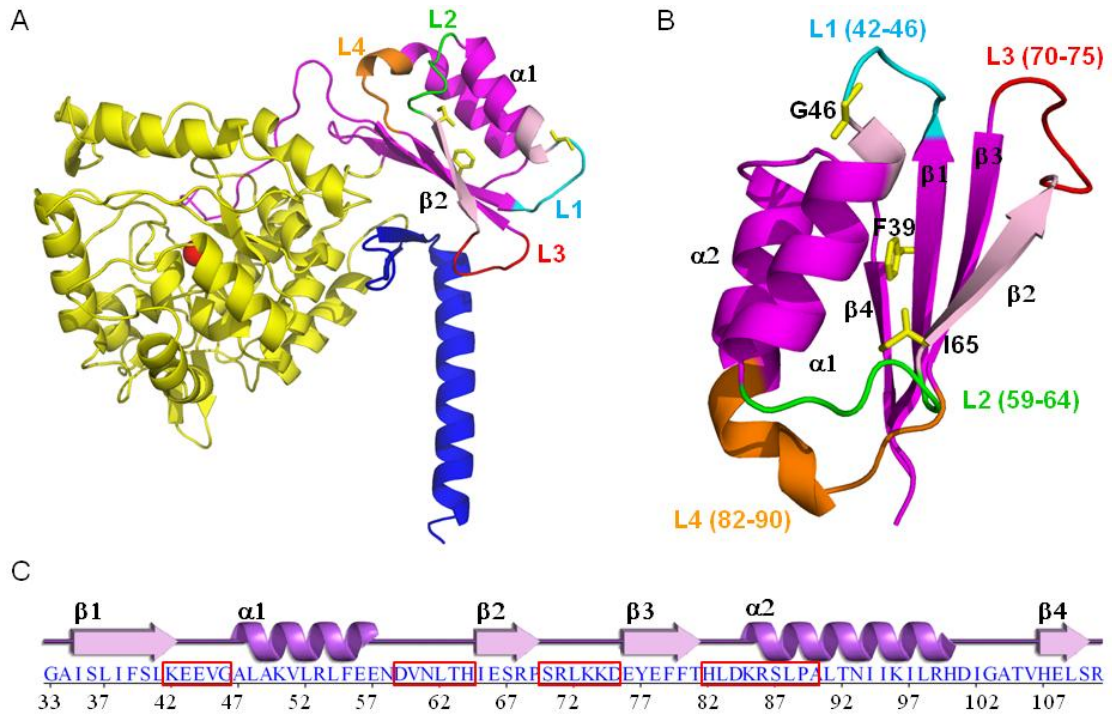

Figure 3.1: Structure representation of human PAH subunit and its RD. (**A**) Structural model of a PAH subunit created by superimposing the regulatory/catalytic domain from rat crystal structure (1PHZ) and the catalytic/tetramerization domain from human crystal structure (2PAH). RD is drawn in magenta, the catalytic domain in yellow and the tetramerization domain in blue. The iron is shown as a red sphere. (**B**) Starting models of the ACT domain (33-111 residues) of wt-hPAH with the residues G46, F39 and I65 drawn as stick in yellow. The two highly conserved GAL and IERSP motifs are highlighted in light magenta. Mobile loop-containing regions, L1 to L4, are drawn in different colours. (**C**) Primary and secondary structure of human RD represented in B. Loop-containing regions, L1 to L4, are highlighted in red boxes.

3.3.1 Mobility and conformational change of loop-containing regions

The dynamics of wt-hPAH is mostly characterized by the large motions of the L2, L3 and L4 regions (Figure 3.2A, B and 3.3) and a small fluctuation of L1. In particular, a large rearrangement is observed for the L3 loop, located near the dimerization interface in the dimer/tetramer structures. This motion is probably enhanced by the absence of the other subunits in the simulated system. In the rat dimeric PAH crystal structure, the L1 and L3 loops of RD interact with both the catalytic domain of the opposite subunit and the tetramerization domain of the same subunit. Differently, in the simulated systems, where the RD is isolated, L1 and L3 are free to move so that we can study their intrinsic flexibility.

L2 and L4 regions are located close to each other on the opposite side of the domain with respect to the L3 loop (Figure 3.1A, B). The L2 region corresponds to the loop/turn located between helix α1 and the strand β2, while the L4 region includes the β3-α2 connecting loop and the N-terminal part of helix α2 (Figure 3.1B, C).

As shown by Essential Dynamics (ED) analyses, the movements of the L2, L3 and L4 regions are represented by the first principal component (PC) of the motion in the human wild-type enzyme (Figure 3.2B). In particular, part of the L2 region (residues from 61 to 64) changes its conformation from turn to β-strand, resulting in the elongation of the β2 strand, while the L4 region undergoes the unwinding of the N-terminal part of the helix α2, leading the two regions closer to each other (Figure 3.2A-C). This is also highlighted by the formation of new H-bonds between the L2 and L4 regions observed during the dynamics (as shown in the WT panel of Figure 3.4). Indeed, His64 (L2) and Thr63 (L2) form H-bonds with His82 (L4), while Asn61 (L2) interacts with Asp84 (L4).

The conformational change of the L2 region from turn/bend to β-strand in the wt-hPAH is also highlighted in the secondary structure analysis by DSSP of the simulated trajectories (WT in Figure 3.5). This change is paralleled by the stabilization of the C-terminal part of the β3 strand (Figure 3.5) that interacts with the L2 region, forming a hydrogen-bond pattern typical of a β-sheet structure (Figure 3.6).

Similar features were also observed for most of the mutants (Figure 3.3, 3.4, 3.5). In particular, the conformational change of the L2 region from turn/bend to β-strand was found in all systems, except in the rat structure and in the F39C mutant (Figure 3.5). However, the closure of L2 and L4 and the formation of new H-bonds between these

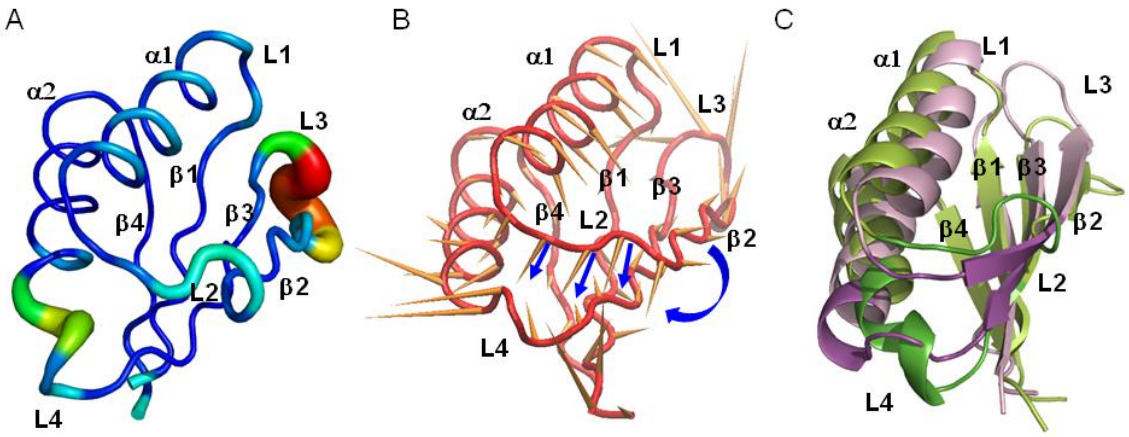

Figure 3.2: Mobility and conformational change of loop-containing regions. (**A**) Sausage representation of the Cα RMSF of the wt-hPAH structure, in which the thickness of the sausage is proportional to the fluctuation of Cα atoms from their average position. (**B**) Porcupine plot illustrating the motions represented by the first PC (accounting for 78% of the overall fluctuation of Cα atoms) from the simulation of the wt-hPAH. Each Cα atom has a cone attached pointing in the direction of motion described by PC1. Blue arrows highlight the motion corresponding to the transition of L2 to β-strand. (**C**) Initial (light green) and final (light pink) wt-hPAH structures with the L2 and L4 regions highlighted in green and magenta.

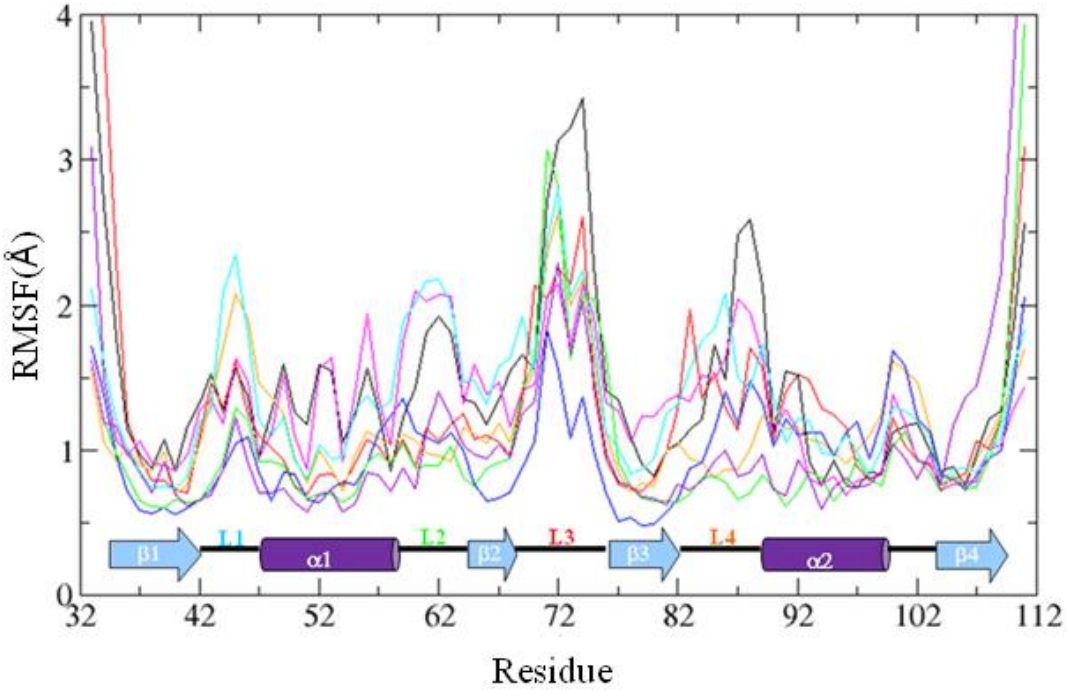

Figure 3.3: Root Mean Square Fluctuation (RMSF) of the C^{α} atoms during the simulation. Color code: wt-hPAH in black, wt-rPAH in red, G46S in orange, F39C in blue, F39L in green, I65T in violet, I65S in magenta, I65V in cyan.

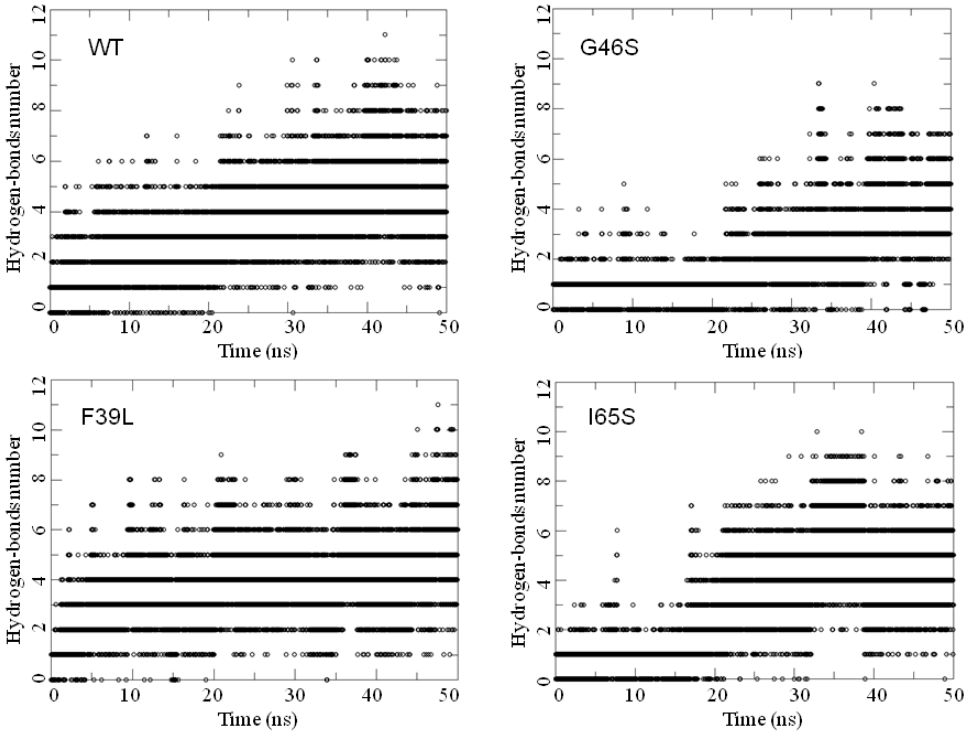

Figure 3.4: Variation of H-bonds number between the L2 and the L4 regions during the simulations. Number of H-bonds between the L2 and the L4 regions *versus* time in the wt-hPAH and in the G46S, F39L, I65S mutants. Hydrogen bonds are determined using a cutoff of 30 degrees on the Acceptor-Donor-Hydrogen angle and of 3.5 Å on the Acceptor-Donor distance.

two regions occur in all systems in at least one replica (examples for each mutated position in Figure 3.4). Moreover, in the rat structure and in the G46S, F39L, I65T, I65V mutants, also the L1 loop, located in close proximity to L3, shows a rearrangement (Figure 3.3).

In order to detect inter-simulation similarities and highlight possible differences among the conformational states of all the studied systems, a "combined essential dynamics" analysis was performed on the concatenated MD trajectories of all enzymes, for both replicas (see Materials and Methods). The first principal component (PC1) describes the transition to β strand of the L2 region (β2 elongation) and the getting closer of L2 and L4 regions, together with a reorientation of the α1 helix (Figure 3.7). The second principal component (PC2) describes the movements of all loops, with a smaller relative contribution from the β strands and α1 if compared to PC1, and a larger relative contribution from L1 and α2 (Figure 3.7B). The projection of the MD trajectories onto

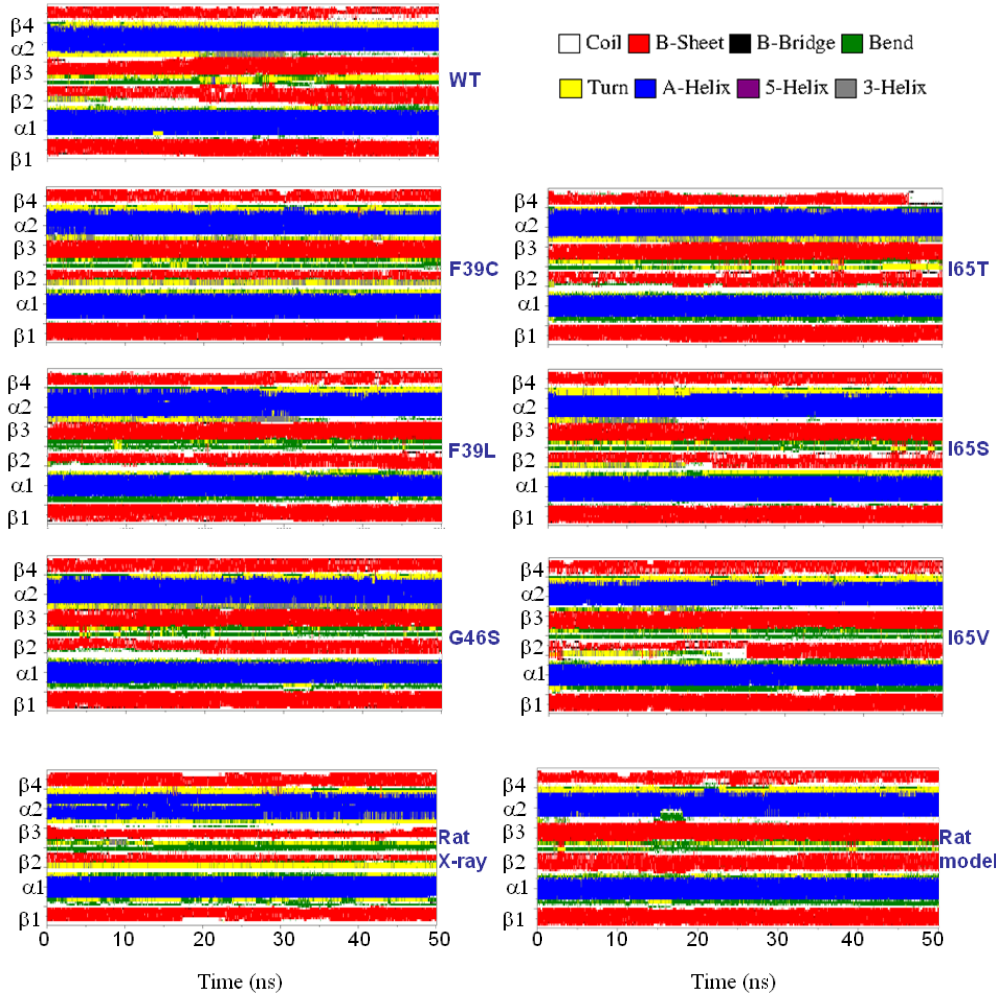

Figure 3.5: Secondary structure analysis. Time-evolution of the DSSP secondary structure for the human (WT) and rat (rat X-ray) wild-type enzymes, the human mutants, and the *in silico* model of the rat enzyme using as template the ACT domain of 3PGDH (rat model).

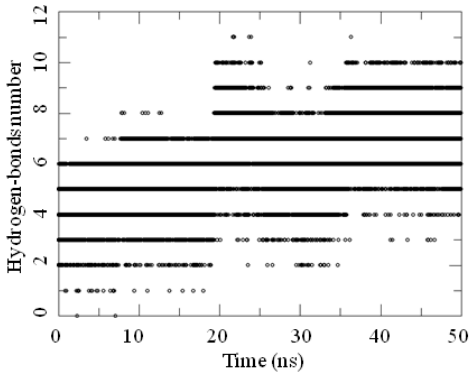

Figure 3.6: Number of H-bonds between the β2 strand (65-69 residues) and the β3 strand (76-81 residues) during the simulation in the wt-hPAH. Hydrogen bonds are determined using a cutoff of 30 degrees on the Acceptor – Donor – Hydrogen angle and of 3.5 Å on the Acceptor - Donor distance.

the (PC1, PC2) space shows that the human and rat wild-type forms explore different regions (Figure 3.7C). In particular, the human wild-type samples regions in the conformational space where the elongation of β2 occurs (PC1>0), while in the wt-rPAH L2 remains in a turn conformation (PC1<0) during both simulations. The regions at PC1>0 are sampled by all mutants in at least one replica, except for F39C. The largest differences between mutants and human wild-type are observed along the second PC. Indeed, the G46S, F39L, I65T and I65V mutants, which have a higher flexibility of α2 and of all loops (including L1) with respect to the β-sheet (Figure 3.3), explore large portions of the space with PC2>0, which seems inaccessible to the human wild-type and the F39C and I65S mutants (Figure 3.7C).

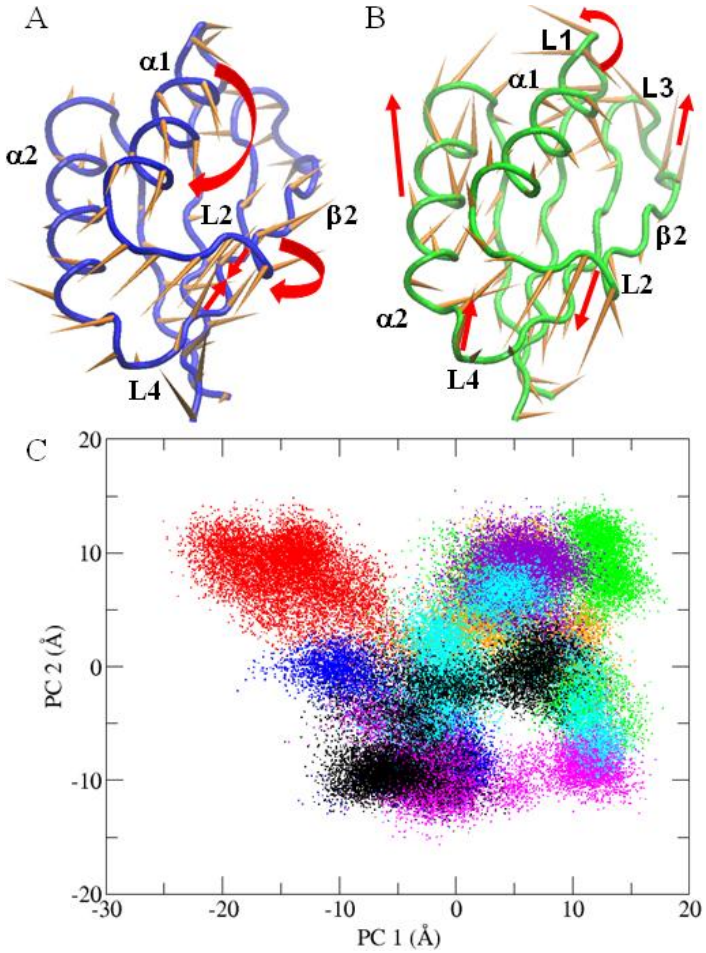

Figure 3.7: Combined essential dynamics analysis. (**A)** Porcupine plots illustrating the motions represented by the first and (**B**) the second PCs from the combined ED analysis. Each Cα atom has a cone pointing in the direction of motion described by each PC. Red arrows highlight the motions described in the text. (**C**) Projections of each trajectory along PC1 and PC2 from the combined ED. Color code: wt-hPAH in black, wt-rPAH in red, G46S in orange, F39C in blue, F39L in green, I65T in violet, I65S in magenta, I65V in cyan.

3.3.2 L2 conformation in the RD structure from wt-rPAH

In the rat RD crystal structure, the L2 region has a particular β-turn conformation (Asn61-His64 turn type IV, Leu62-Ile65 turn type I) and, unlike the human RD, does not undergo the conformational change of this region from turn to β-strand during the simulations (Figure 3.7C and 3.5). A screening of all the available structures of other ACT domains (PDB codes: 1PSD, 1RWU, 1TDJ, 1O8B, 1Q5Y, 1I1G, 1NH7, 1ZPV, 1U8S, 1Y7P, 2F1F, 2F06, 2CDQ, 2QMX) shows that the L2 region preferentially adopts the β-strand conformation, while the turn conformation of the L2 region is specific to the rat RD. The B-factor values of the atoms belonging to the L2 region in the rat PAH crystal structure (PDB code 1PHZ) are quite high, with values higher than the average B factor of the protein. This suggests that the L2 region is quite disordered in the crystal.

To test the stability of a hypothetical β conformation of the L2 region in wt-rPAH, we have generated a model of the RD of rat enzyme, using as template the ACT domain of D-3-phosphoglycerate dehydrogenase (3PGDH) from *Escherichia coli*. A superposition of the RD domain from rat crystal structure, rat modelled structure and 3PGDH crystal structure is shown in Figure 3.8. The overall fold is very similar, apart from the L2 region. Indeed, the model has the L2 region in a β-strand conformation as the template. The model was then subjected to 50-ns-long MD simulations and the β-strand conformation of the L2 region was found to be stable during the whole simulation (rat model in Figure 3.5).

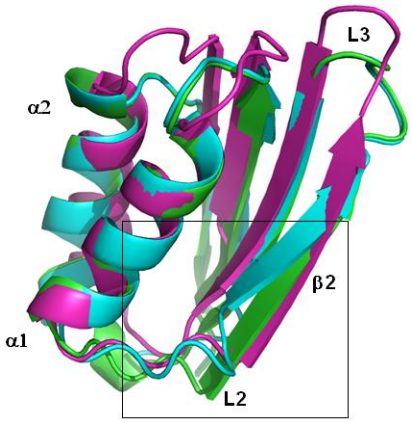

Figure 3.8: Superimposition of the ACT domains. Superimposition of the ACT domains of D-3-phosphoglycerate dehydrogenase from *Escherichia coli* (magenta), crystallographic wt-rPAH (cyan) and *in silico* model of wt-rPAH (green). A box encloses the L2 region discussed in the text.

3.3.3 Correlation between a hydrophobic surface and the motion in the RD

An extended hydrophobic surface between the helix α1 and the strand β2 was identified by analysing the water density distribution around the solutes during the MD simulations. Indeed, the water density maps show the presence of a reduced number of hydration sites near this surface in all systems (Figure 3.9A).

The variation of the solvent accessible surface area (SASA) of this region (residues Val45-Ala49, Val51, Leu52, Leu54, Phe55, Val60-Thr63, Ile65, Ser67, Tyr77 (Figure 3.9B)) was calculated during the simulations. In the wt-hPAH a significant decrease of the SASA was observed, while in the mutants simulations the SASA values generally showed smaller variations (Figure 3.9C).

To determine the correlation between the SASA variation of the identified hydrophobic surface and the protein collective motions, a functional mode analysis (FMA) was performed on all the simulations (see Materials and Methods). The collective motion that is maximally correlated with the SASA (MCM) was expanded in the space of the first n PCs that account for 90% of the overall fluctuation (essential space). For the wild-type enzyme, an overall MCM-SASA correlation coefficient of 0.82 was observed for the second replica (Figure 3.9E), with PC1 accounting for more than 40% of the variance of the SASA (% var in Figure 3.9D). This indicates that the reorganization of the hydrophobic patch during the simulation is mainly related with the motion described by PC1 (Figure 3.2C) and in particular with the restructuring of the L2 loop, which hosts some of the residues in the patch. The mutants showed correlation values ranging from 0.53 to 0.83, with only F39C and I65S having correlations as high as the human wild-type (Figure 3.9E). The MCM expansion of these two mutants was dominated by PC1 (Figure 3.9F), analogously to the human wild-type case. Although the F39C mutant does not show the elongation of the β2-strand, the PC1 of its trajectory describes a large movement of the L2 and L4 regions that approach each other, as in wt-hPAh and I65S mutant. This motion seems to modify the hydrophobic surface. For the rat wild-type and the other human mutants, no dominant contributions to the SASA variation were observed across the essential space (Figure 3.9F).

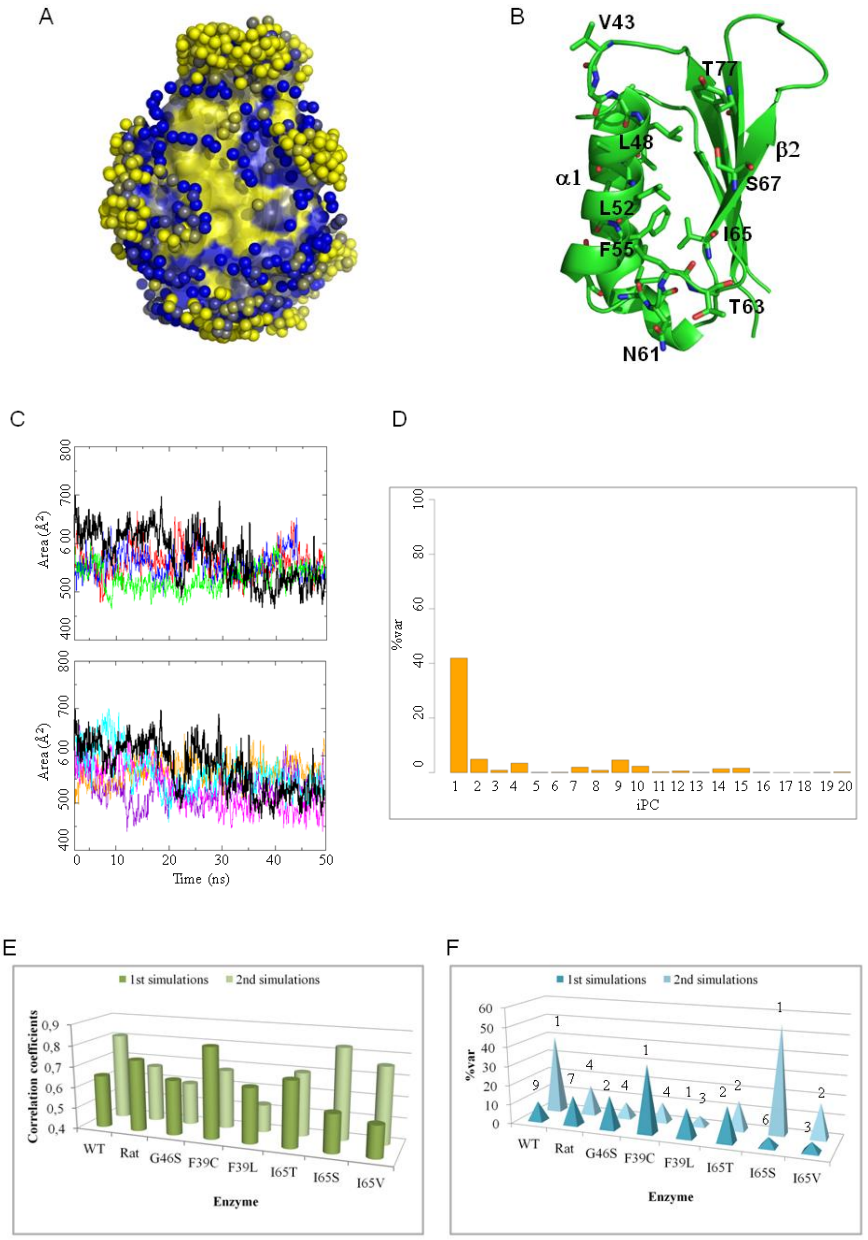

Figure 3.9: (**A**) Water distribution in the wt-hPAH. The protein surface atoms are coloured from yellow to blue according to the increasing hydration score S_{hyd}^{atom}. The local maxima of the water density map are represented as spheres, coloured from yellow to blue according to the increasing density value. Hydrophobic atoms (yellow surface) have low hydration scores and are poor in hydration sites. (**B**) The hydrophobic residues that belong to the hydrophobic surface between the helix α1 and the strand β2 are shown as stick in the wt-hPAH. (**C**) Variation of the SASA of the residues at the hydrophobic surface between the helix α1 and the strand β2 in the wt-hPAH (black) and the mutants (G46S in orange, F39C in blue, F39L in green, I65T in violet, I65S in magenta, I65V in cyan). (**D**) Relative contribution of the first 20 PCs in the MCM to the variance of the hydrophobic surface SASA in wt-hPAH (%var). (**E**) Plot of the Pearson correlation coefficient between the trajectory projection on the MCM and the SASA of the hydrophobic surface (C) for all the systems and replicas. (**F**) Plot of the maximum %var per PC for each system and replicas. In the plot, the PC that has the largest relative contribution is indicated for each replica.

3.3.4 Analysis of core interactions

The hydrophobic contacts formed by selected residues in the hydrophobic core of wt-hPAH (Leu37, Phe39, Leu41, Val51, Phe55, Ile65, Phe79, Leu91, Ile94, Leu98) were calculated on the first 5ns and on the last 5ns of the simulations by the minimum distance matrices (Figure 3.10). The side chains of these residues undergo a significant rearrangement during the simulation. In particular, the distances between the residues Leu37 (β1) and Phe39 (β1) on one side, and Leu91 (α2), Ile94 (α2) and Leu98 (α2) on the other, change due to the unwinding of the α2 N-terminal part. Moreover, the residue Ile65 (β2) changes its distance from Phe39 (β1), Phe55 (α1) and Leu91 (α2) due to the elongation of the strand β2.

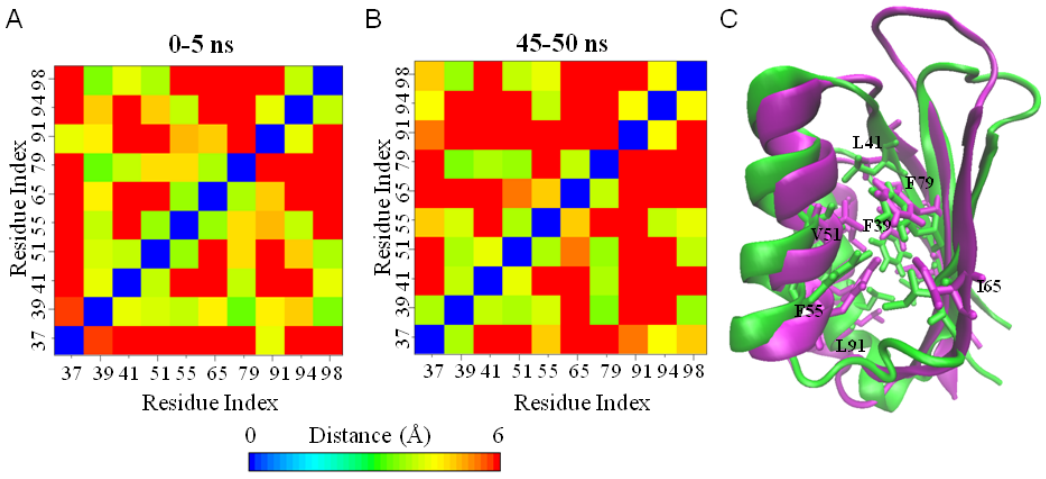

Figure 3.10: Average distances between selected residues of the hydrophobic core in the wt-hPAH. (**A**) Distance matrices calculated on selected residues (Leu37, Phe39, Leu41, Val51, Phe55, Ile65, Phe79, Leu91, Ile94, Leu98) of the hydrophobic core in the wt-hPAH. Average distances calculated over the first (0-5ns) and (**B**) the last (45-50ns) 5 ns of simulation are reported. (**C**) Superimposition of the structure at 0 (green) and 50 (magenta) ns of wt-hPAH with the hydrophobic core residues in stick. Selected residues are labelled.

In the F39C and F39L mutants, the replacement of Phe39 with either a polar Cys or a smaller hydrophobic residue such as Leu, induces a reorganization of the interactions in the core with respect to the wild-type enzyme. In particular, the residue 39 looses its contacts with other core hydrophobic residues, namely Leu41, Val51 and Phe55 in F39C, and Leu37, Phe55 and Leu98 in F39L (Figure 3.11). In the latter mutant, an increased distance between Ile65 and Phe79 is also observed, as a result of the L2 transition to β-strand. The perturbation of the core is particularly evident at the end of

the F39C simulation, where two innermost residues (Phe55 and Val60) become exposed to the solvent in response to the movements of the L2 loop and of the α1 helix (Figure 3.12A).

In the mutants in position 65, the substitution with either a polar or a smaller hydrophobic residue may distort the hydrophobic packing in the RD core. In I65T and I65V the side chains of some residues (Phe39, Val51, Phe55, Phe79, Ile94) of the hydrophobic core undergo a rearrangement during the simulation and the distances between such residues deviate with respect to the wild-type enzyme (Figure 3.11). Differently from I65T and I65V, in the I65S mutant the contacts of residue 65 with the rest of the hydrophobic core are lost (Figure 3.11). In all these mutants, the residue in position 65 becomes more exposed to solvent (Figure 3.12B) in response to the elongation of β2.

In the G46S mutant, the Gly residue of the GAL ligand-binding motif of the ACT domain family is replaced by a polar residue. In this mutant some changes occur due to a reorganization of the core side chains (Figure 3.11), that sample different orientations with respect to the wild-type enzyme during the simulation. The pattern of distances between the hydrophobic core residues, for this mutant, result similar to that of the I65T and I65V mutants.

The perturbation of the core introduced by the mutations is also evident from the comparison of the correlation webs of the different systems, where residues with dynamic correlations higher than 0.3 are connected by a line (Figure 3.13A, B). The wt-hPAH web (Figure 3.13A) shows a dense network of connections between the two helices on one side and the β sheet on the other, indicating that the two parts of the protein move concertedly in the main collective motions. All the mutant webs show a more disconnected pattern (see I65S in Figure 3.13B for an example), with a reduced number of links connecting helices with β-sheets (Figure 3.13C).

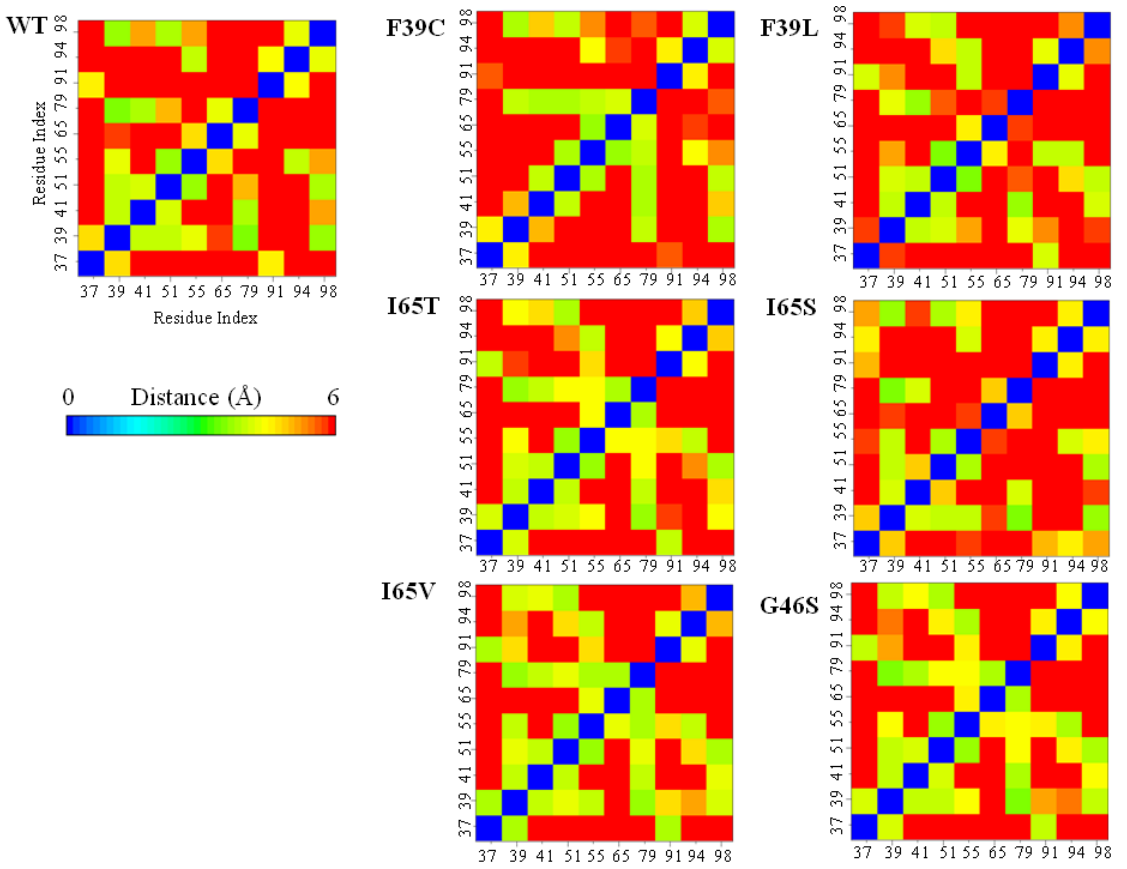

Figure 3.11: Average distances between selected residues of the hydrophobic core in the mutants. Distance matrices calculated on selected residues (Leu37, Phe39, Leu41, Val51, Phe55, Ile65, Phe79, Leu91, Ile94, Leu98) of the hydrophobic core in the wt-hPAH and in the mutants. Distances between pairs of residues are calculated as the minimum over all possible pairs of non-hydrogen atoms of the side chain. Average values calculated over the 50-ns simulation are reported.

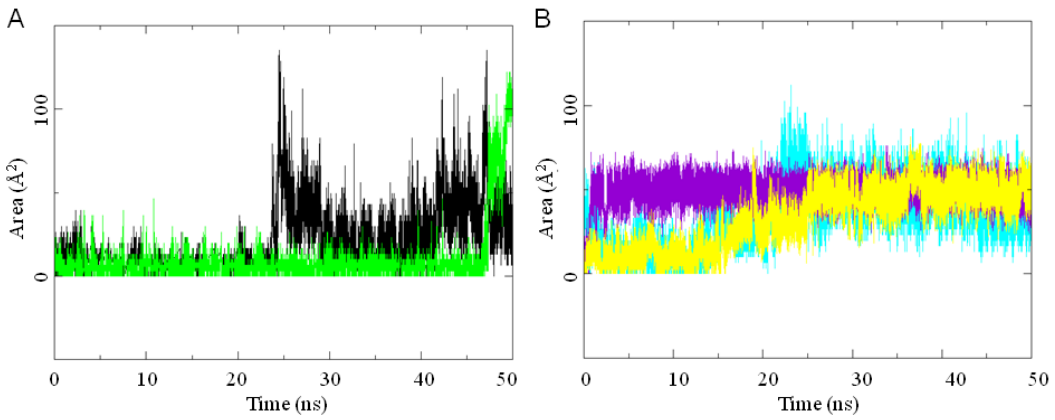

Figure 3.12: Time evolution of SASA of selected residues. (**A**) Time evolution of Phe55 (black) and Val60 (green) SASA in the F39C mutant. (**B**) Time evolution of the SASA of the residue in position 65 in the I65S (yellow), I65T (violet) and I65V (cyan) mutants.

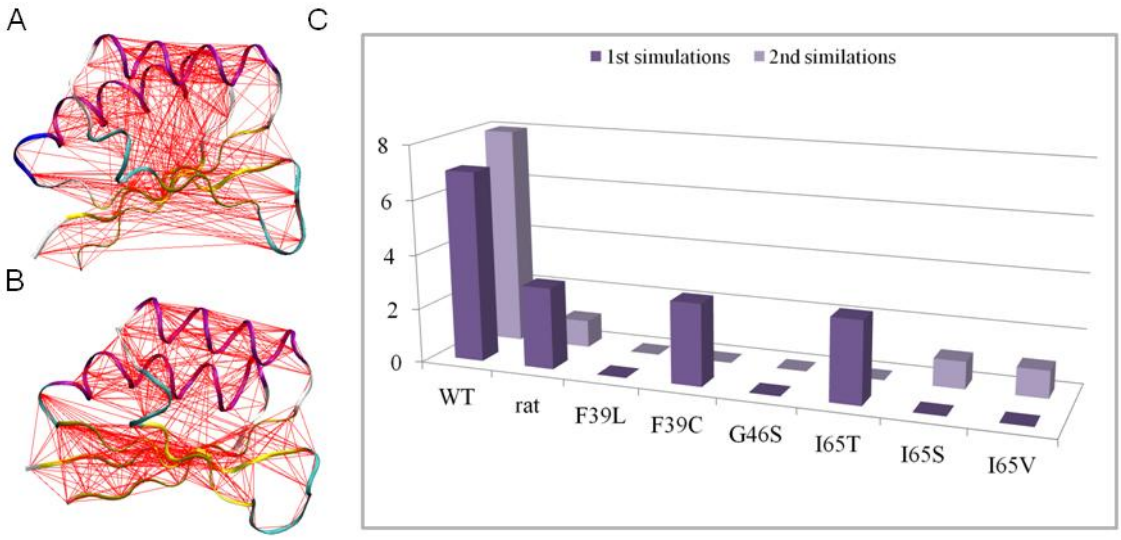

Figure 3.13: Correlation webs and number of links between helix and β-sheet residues. Correlation webs of wt-hPAH (**A**) and I65S (**B**) generated with Dynatraj. C^{α} atom pairs with dynamic correlation higher than 0.3 are connected with a line. (**C**) Total number of links connecting helix to β-strand residues in the correlation webs. Values from both replicas in each simulated system are reported.

3.4 *Discussion*

MD simulations were performed on the isolated ACT regulatory domain of both human and rat wild-type PAH and of six human disease-causing mutants. The analysis of the trajectories was shown to provide dynamical and structural insight on the isolated RD and on the effects of the mutations.

We have simulated the structured regions of the RD (residues 33-111), as the N-terminal part is partially unstructured (residues 1-32). The structured regions of the RD coincide with the structural motif known as ACT domain, consisting of a four-stranded antiparallel β-sheet, two short α-helices and four connecting loops.

Examination of the 50-ns-long MD trajectories highlights a specific dynamic behaviour of the wild-type enzymes and the mutants. The simulated systems, while preserving their overall secondary structure, showed significant rearrangements of the loop-containing regions.

In 10 out of 16 simulations, the L2 region, which in the rat crystallographic structure (and in the generated models) has a turn/bend conformation, undergoes a conformational change to a β-strand conformation that elongates the strand β2. The β2 elongation is associated with the formation of a typical H-bond pattern with the adjacent β3, leading to the stabilisation of the β-sheet. This suggests that the β-strand conformation of the L2 region is an accessible and intrinsically stable conformation of the RD. This finding is supported by the similarity with the known structure of the other ACT domain-containing proteins and by the stability of the L2 β-strand conformation during the 50-ns-long MD simulations of a model of the rat PAH, obtained using as template the ACT domain from 3PGDH with an elongated β2 strand.

The conformational change of the L2 region leads to the stabilization of the β2 strand formed by residues 65-69. These residues constitute the conserved motif IESRP involved in the putative Phe binding at the isolated PAH RD. In addition, the residues 60-63 in the L2 region, show a high degree of evolutionary conservation (Gjetting et al., 2001b). A large portion of mutations leading to hyperphenylalaninemia cluster within or around the β2 strand (Gjetting et al., 2001b), suggesting that these residues are relevant for the activity of the enzyme.

The principal motion in most of the studied systems involves a correlated conformational change in both the L2 and L4 regions. The movement of L2 towards L4 and the unwinding of the N-terminal part of the α2 helix promote the formation of H-

bonds between L2 and L4. Collectively, all these movements result in a transition from an initial open to a final more closed structure of the domain.

Our findings are in agreement with the hypothesis formulated by Li and co-workers on the basis of hydrogen/deuterium exchange experiments on the PAH (Li et al., 2010). These authors concluded that upon Phe-binding (i) the interactions between the two peptides 39-59 and 82-91 are partly disrupted, leading to a more open RD, and (ii) the interaction between the RD and the catalytic domain is altered. Interestingly, the 39-59 and 82-91 peptides partially and completely coincide with the dynamic L2 and L4 regions, respectively.

In addition, from the analysis of our trajectories we could identify a hydrophobic patch in the region between the helix α1 and the strand β2, corresponding to the ligand binding-site of the ACT protein family. This could be a possible site of interaction for the benzyl side chain of the Phe ligand. We found that the solvent accessibility of the hydrophobic surface is correlated to the global intrinsic motions of the protein and, particularly for the wild-type enzyme, to the motions of L2 and L4 regions, suggesting a relationship between the Phe-binding and the overall dynamic behaviour of the enzyme. The transition from an open to a closed form of the domain leads to the decrease of the solvent exposure of the hydrophobic surface where the Phe ligand could bind to the domain. This suggests that the closed form of the enzyme, with a smaller hydrophobic surface, is unbound, while the open form is able to bind Phe. This is in agreement with the regulatory model in which PAH switches from an inactive to an active form upon Phe binding (Thorolfsson et al., 2003; Thorolfsson et al., 2002).

The difference between the rat and human PAH RD dynamics might be related to the difference in the activity of the two enzymes. For the rat enzyme, pre-treatment with phenylalanine is reported to produce a 10- to 30-fold increment of the activity, while for the human enzyme this increment is only ~3 to 6-fold (Miranda et al., 2002; Thorolfsson et al., 2003). This is due to the higher levels of activity of the inactive human enzyme, which probably undergoes different conformational changes from the rat enzyme in the activation process.

To better compare the binding site of the ACT domain-containing proteins and the corresponding region in hPAH, we generated an *in silico* model of hPAH tetramer (see Materials and Methods) and superposed such model to Phe-bound PDT (PDB code 2QMX). The superposition reveals that a Phe-binding site of PDT is located at the

interface between the RD and the catalytic domain of the adjacent subunit in PAH. In particular, an α-helix (Cα4) from the catalytic domain in PAH is located in a similar position to the Phe-binding site region of the second ACT domain in PDT (Figure 3.14). Thus, some residues of Cα4 helix could participate in the Phe-binding in PAH, presumably through conformational changes in the catalytic domain and a reorientation of the two interacting domains upon Phe-binding.

To get insight into the structural role of HPA disease-causing mutations in the RD hydrophobic core, we selected F39C, F39L, I65T, I65S and I65V mutants. The residues 39 and 65 are the most frequently mutated positions and their resulting phenotypes can range from mild to severe PKU. In addition, we selected the G46S mutant, which is associated with a severe form of PKU (Eiken et al., 1996). The Gly46 residue, located between the β1 strand and the α1 helix, is a good candidate for Phe binding, since it belongs to the ligand binding site in other ACT domain-containing proteins.

In all mutants, the extent of correlation of motion between the helix and β-sheet residues is smaller than in the human wild-type enzyme. In G46S, F39L, I65T and I65V mutants, also the correlation between the variation of the hydrophobic surface and the movements of the L2 and L4 regions is smaller than in the wild-type enzyme. Together, these data suggest that changes in the interactions within the hydrophobic core of the RD may significantly affect its stability and hence its regulatory role. Indeed, changes in the RD stability could interfere with the cross-talks with the other domains and affect the conformational changes that are involved in cooperativity.

We hypothesise a mechanism for which alterations in the hydrophobic interactions of the RD may hamper the functionality of the whole enzyme.

The RD mutations can influence the correct binding of Phe to the domain, resulting in a poor Phe activation that ultimately reduces the enzyme activity. This is in line with experimental data on the mutants, that have been characterized *in vitro*, showing a reduced Phe activation and a reduced activity (Erlandsen et al., 2004; Gjetting et al., 2001a; Gjetting et al., 2001b; Pey et al., 2007).

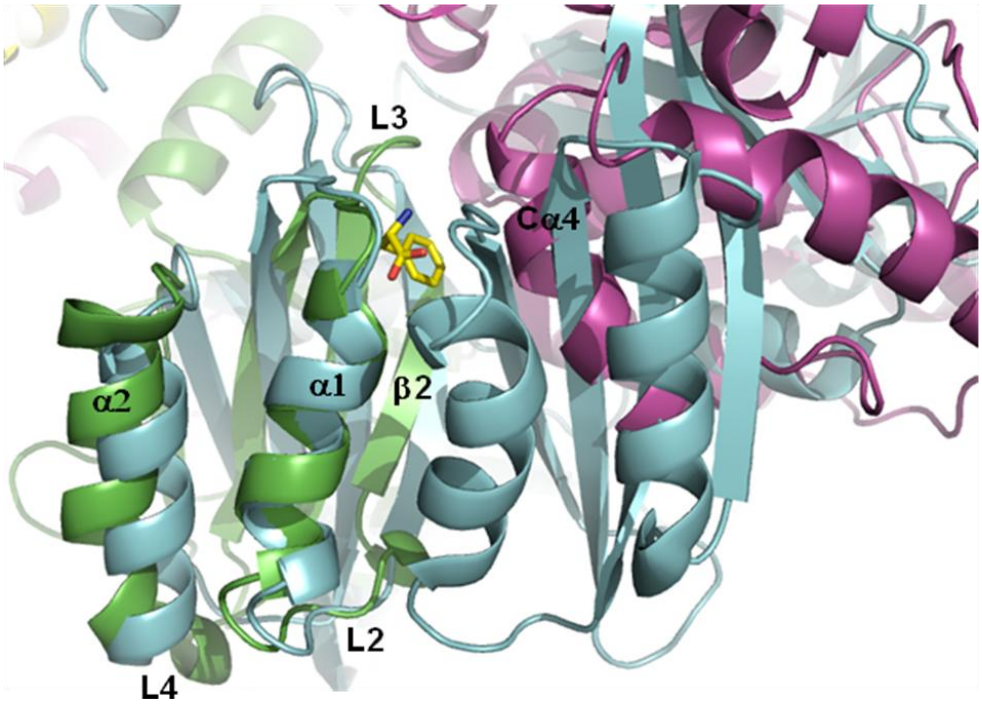

Figure 3.14: Superimposition of the Phe-bound prephenate dehydratase (PDT) from *Chlorobium tepidum TLS* (cyan) (PDB code 2QMX) and the tetrameric model of the human PAH (RD in green, catalytic domain in violet). The bound Phe ligand is shown in yellow as stick.

CHAPTER 4. *In silico identification of the allosteric Phe-binding site in the hPAH*

4.1 *Introduction*

The proposal that mammalian PAH has an allosteric binding site for Phe, in addition to the catalytic site, has been a controversial issue for many years.

Phe plays an important role in the regulation of the catalytic activity of tetrameric PAH which is activated several-fold by preincubation with the substrate (Shiman and Gray, 1980). Phe allosterically modulates the activity of PAH by positive cooperativity (Kappock and Caradonna, 1996; Kaufman, 1993; Knappskog et al., 1996a), but the mechanism is still unclear. The cooperativity reflects a transition from a low-activity and low-affinity ‘‘T’’ to a high-activity and high-affinity ‘‘R’’ state, accompanied by large conformational changes (Kappock et al., 1995; Shiman et al., 1979; Stokka and Flatmark, 2003; Thorolfsson et al., 2003). An *in vitro* binding of Phe to recombinant wild-type and some mutant forms of the isolated RD was demonstrated (Gjetting et al., 2001b). Recently, NMR spectroscopy of the isolated RD revealed chemical shift perturbations in the presence of Phe, that are consistent with its binding at a specific site (Li et al., 2011). However, isothermal titration calorimetry measurements, demonstrating that Phe binds to full-length hPAH tetramer with a stoichiometry of Phe:subunit = 1.2 : 1, do not support an effector binding site in addition to the active site (Thorolfsson et al., 2002).

By contrast, though low sequence identity occurs between the regulatory ACT domain of hPAH and its distant homolog prephenate dehydratase (PDT) (i.e. 20% over 80 residues) (Liberles and Martinez, 2009), it is noteworthy that both enzymes contain the sequence motifs GA(S)L/IESRP within the ACT domain. These motifs are involved in the binding of Phe to the isolated regulatory domain of hPAH (Gjetting et al., 2001b) and their role in Phe binding has been crystallographically proven for PDT (Tan et al., 2008).

However, in most cases ACT domains interact with other ACT domains, but this is not the case for the ACT domain in PAH. In PAH, the motifs are located at the interface between the interacting regulatory and catalytic domains. A structure of PDT from

Chlorobium tepidum TLS shows Phe bound to its regulatory ACT domain (PDB code: 2QMX) (Tan et al., 2008). Phe is coordinated by Asn206 (Asp in hPAH), Leu211, Tyr240, and Phe242, plus Asp224 (Asn in hPAH) and Leu225 from the adjacent ACT domain. A superimposition of hPAH and PDT reveals that an α-helix (Cα4, residues 204-216) from the catalytic domain in the adjacent chain in hPAH is located in a position similar to that in the second ACT domain in PDT.

In my previous MD studies on the isolated RD (manuscript in preparation), I found a correlation between the variation of the SASA of a hydrophobic region, corresponding to the ligand binding-site of the ACT proteins family, and the global intrinsic motions of the protein, suggesting a relationship between the Phe-binding and the dynamics of the enzyme (see sections 3.3.3 and 3.4).

In this context, I aimed to shed light on the open questions on the regulation role of the enzyme through the *in silico* identification of the putative allosteric Phe-binding site. According to our previous studies and by analogy with the other ACT-domain containing proteins, we proposed a putative hydrophobic region for Phe-binding at the dimeric interface between the RD and the CD (lined by the ACT-domain motifs GAL-ESRP) of two distinct subunits. Thus, I extended the structural and dynamical study to the hPAH, rPAH and G46S dimers. Furthermore, through docking simulations, I generated a model for the Phe-PAH interactions and elucidated the structural effects of the disease-causing G46S mutant (on the Phe-binding).

To address this complex study I considered, in explicit water solution, a total of 9 systems (3 unbound and 6 Phe-bound dimeric PAH forms), each consisting of about 20500 atoms. To run simulations of complex bio-systems of such dimensions (784 residues per dimer) long computation time was required. All the simulations on the unbound and Phe-bound forms were extremely demanding and computationally intensive, and they were beyond the capabilities of the resources available at my institution. For these reasons, I attended the "HPC-Europa2 Transnational Access programme" to accede to the modern fast supercomputer at EPCC (Edinburgh Parallel Computing Centre) in Edinburgh. Moreover, during the HPC-Europa2 project, I continued collaborating with the Dr. Fraternali's group at King's College in London.

4.2 *Experimental methods*

4.2.1 *Building of dimeric models*

Models of the dimer (residues 33-424 of two chains A and B within the dimer) of the wt-hPAH enzyme and G46S mutant were generated *in silico* with the program MODELLER 9v8 (Sali and Blundell, 1993) using as template the composite dimeric model built by superimposing the secondary structure elements of the respective catalytic domains of the dimeric rPAH (PDB code 1PHZ), the dimeric hPAH (PDB codes 1PAH) structures. The template was built using the dimeric rPAH structure for the N-terminal RD (residues 33-116), the dimeric hPAH structure for the catalytic domain, the dimerization motif (residues 117-424) and the position of Fe(III) ion and the three water molecules coordinated to Fe in the active site. The template and the target sequences were aligned using BODIL (Lehtonen et al., 2004). Subsequently, the alignment files were used as input in MODELLER and 200 models of each studied forms were generated. For the mutants, the mutated residues were replaced in the alignment file. The model that presented the best MODELLER DOPE score was selected for the MD simulations.

4.2.2 MD simulations of the studied dimers

MD simulations were performed with the GROMACS (Van Der Spoel et al., 2005) package using the GROMOS 43a1 force field (van Gunsteren et al., 1996) on the unbound and some selected Phe-bound systems. The initial dimeric models (each consisting of 784 residues) generated *in silico* and the dimeric wt-rPAH structure were solvated in dodecahedral boxes containing simple point charge (SPC) water molecules (Berendsen et al., 1981). To neutralize the system, some Na^+ ions were randomly placed far away from the surface of the enzymes. Van der Waals parameters and the charge for Fe(III) were fixed at zero. His285 and His290 coordinating the Fe(III) ion were assigned δ-hydrogen atoms. Constraints on the bond distances of the Fe with its coordinated residues were used. Simulations were carried out with periodic boundary conditions and a 2-fs time step. All the bonds in the protein were frozen with the LINCS method (Hess et al., 1997), while SETTLE (Miyamoto and Kollman, 1992) was used for water molecules. The Berendsen algorithm (Berendsen et al., 1984) was applied for temperature and pressure coupling. The particle mesh Ewald method (Essmann et al., 1995) was used for the calculation of the electrostatic contribution to non-bonded

interactions with a direct sums cut-off of 14 Å. A 14 Å cut-off was used for van der Waals interactions. The systems were energy minimized with 2000 steps of the steepest descent method and subsequently with 2000 steps of conjugate gradient method. After the minimization, harmonic positional restraints (with an initial force constant of 4.8 kcal/mol/$Å^2$) were imposed onto the protein heavy atoms and gradually turned off in 360 ps, while the temperature was increased from 200 to 300 K at constant volume. The system was then equilibrated for 2 ns without restraints in NPT conditions (T=300 K, p=1 bar). Finally, NPT MD simulations were performed for 40 ns for the unbound wt-hPAH, for 30ns for the unbound rat and mutant PAH, and from 5ns to 15ns for the complexes. Two replicas were run for each system, because we have noted, in our preliminary study, that replicas significantly improve the statistical description of the dynamics of the studied system on the nanosecond scale. We studied a total of 2x9 systems (3 unbound and 6 Phe-bound enzymes) in explicit water solution, each comprised of about 20500 atoms.

Long simulations are needed to investigate motions having functional implications and to reveal conformational changes as a function of time. To run long simulations of complex bio-systems, long computation time is required. The use of a modern fast supercomputer was essential to the success of this computationally intensive project. The work has been performed under the HPC-EUROPA2 project (project number: 228398) with the support of the European Commission - Capacities Area - Research Infrastructures. MD simulations were run on 128 processors of the HECToR (High End Computing Terascale Resources) supercomputer at EPCC in Edinburgh.

MD trajectories were analysed using GROMACS analysis tools.

4.2.3 Molecular docking

Molecular docking is a computer simulation procedure to predict the conformation of protein-protein and protein-ligand complexes. The docking simulation, investigating the conformational space, searches for the best fit between two or more molecules considering several parameters as: geometrical receptor-ligand complementarity, regarding atomic VDW radius and charge; receptor or ligand structure flexibility; interatomic interactions, such as hydrogen bonds and hydrophobic contacts. As result, the predicted orientations (poses) of a ligand in the target's binding site are generated.

The interactions receptor-ligand depend both on the specific interactions in the binding site as well as on the non-specific forces outside the binding pocket. The complexity of interactions between proteins and flexible target molecules, including other proteins, nucleic acids and small molecules, is often determined by the considerable flexibility of the protein binding sites and by the structural rearrangements that occur upon binding of the associated molecule. Protein flexibility and the dynamics of inter molecular interfaces can regulate binding affinity and specificity in molecular recognition.

Docking applications, considering the flexibility of the ligand and/or the receptor docking algorithms can be classified in two large groups: rigid-body and flexible docking. Rigid-body docking method does not take into account the flexibility of neither ligand nor receptor, limiting the specificity and accuracy of results, considering essentially geometrical complementarities between two molecules. This approach considers the protein-ligand system as a lock and key mechanism, with the ligand being the key that, upon correct positioning, opens the lock of the interaction. Flexible docking methods consider several possible conformations of ligand or receptor, as well as for both molecules at the same time, in order to accommodate upon the binding of the partner. In this second method, the docking is an induced fit, or hand-and-glove approach. However, the conformational flexibility results in very large degrees of freedom, which implies either that searching such a space would be very slow or very incomplete or both. The ways to reduce the degrees of freedom of the docking problem are: (i) to define local regions of space as interaction zones, and to look only at conformations in this region while keeping the rest of the system rigid; (ii) or alternatively, to keep the protein rigid and to allow only the generally much smaller ligand to sample conformation space.

The main issues of a docking approach therefore are: choosing whether the system would be considered absolutely rigid, semiflexible (e.g. flexible ligand and rigid protein) or flexible, deciding how the contribution of the solvent is included, the choice of the searching algorithm, the scoring function, and the evaluation of the results. The contemporary docking software programs have several methods for addressing the aforementioned issues and reaching the goals of wide input range, fast and reliable docking. There is a number of docking packages being used nowadays, ones more or less suited for specific types of interactions, mutually differing at the basis of approach to the docking problem. These docking software programs include for example GOLD

(Verdonk et al., 2003), GLIDE (Halgren, 2000), FlexX (Kramer et al., 1999), HADDOCK (Dominguez et al., 2003), ICM (Abagyan et al., 1994), DOCK (Lang et al., 2007) or AutoDock (Morris et al., 1998).

Usually docking algorithms predict several orientations (poses) for the ligand inside the binding site. Scoring functions, that are able to evaluate intermolecular binding affinity or binding free energy, are employed in order to optimize and rank results, obtaining the best orientation after the docking procedure. There are several scoring functions available. Scoring functions are mathematical approximating methods for estimating binding affinity, in order to lead optimization of virtual screening results, finding the highest affinity ligand against a target. Using as input the atomic coordinates of a binary complex, scoring functions are able to estimate free energy of binding or binding constant. Scoring functions can be classified in 3 different types: force-field based, knowledge based and empirical based scoring functions. Force-field based scoring functions use non-bonded terms of classical mechanics force fields. Knowledge-based scoring functions are based on statistical observations of intermolecular contacts identified from structure databases. Empirical scoring functions make use of several intermolecular interaction terms that are calibrated through a linear regression procedure, where theoretical values are fitted to be closest as possible to experimental data.

4.2.4 MOE (Molecular Operating Environment) software

Phe docking into the dimeric PAH enzymes has been performed with the MOE (Molecular Operating Environment) software (Chemical Computing Group Inc., 2012). MOE's Dock application searches for favorable binding configurations between small- to medium-sized ligands and a not-too-flexible macromolecular target. For each ligand, a number of configurations called *poses* are generated and scored. The score can be calculated as either a free energy of binding, which takes into account solvation and entropy, or the enthalpic term of the free energy of binding. The final top-scoring poses, along with their scores and conformation energies, are written to a database where they are ready for further analysis.

The Dock application was designed for flexibility and customizability. The algorithm is arranged as a series of self-contained stages. At most stages, there is a choice from among multiple different methods.

The Dock workflow is divided into five sequential stages (Figure 4.1): receptor and ligand preparation, placement, rescoring (1), refinement, rescoring (2).

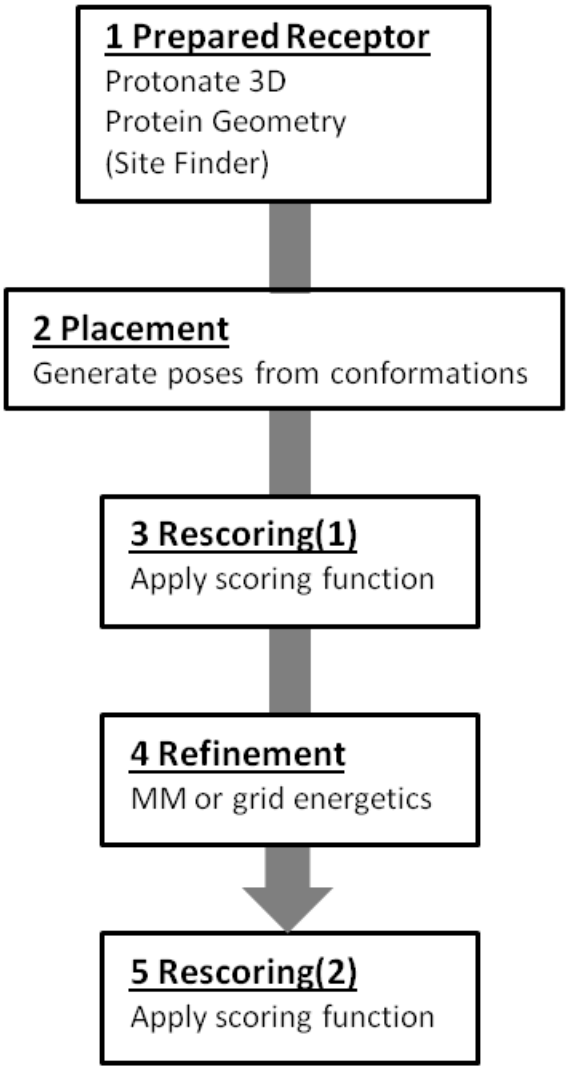

Figure 4.1: Dock workflow of the MOE's dock application (Chemical Computing Group Inc., 2012).

The purpose of the first step is to correct structures and to prepare macromolecular data for further computational analysis. X-ray crystal structures with missing or poorly resolved atomic data need to be modeled and fixed before subsequent computational analyses. The bond orders, charges and protonation states of ligands and cofactors should be inspected and corrected. Once the backbone and side chains are complete and any errors in the ligand and cofactor structures have been corrected, a typical structure

preparation workflow includes addition of hydrogens and calculation of partial charges followed by tethered energy minimization (molecular mechanics refinement) to relieve bad crystallographic contacts or other bad geometries. A sophisticated algorithm for calculating protonation state and hydrogen placement, designed especially for macromolecular systems, is provided by PROTONATE3D that optimizes the hydrogen orientations to maximize H-bond networks and minimize the overall self-energy.

SITE FINDER is used to calculate possible active sites in the receptor from the 3D atomic coordinates. The SITE FINDER methodology is based upon Alpha Shapes which are a generalization of convex hulls developed in (Edelsbrunner et al., 1995).

The Placement phase generates a collection of poses from the pool of ligand conformations using one of the placement methods. To each of the generated poses a score is assigned by the placement method. Triangle Matcher method was used to generate 30 poses for each docking simulation. This placement method generates poses by aligning ligand triplets of atoms on triplets of alpha spheres in a very systematic way.

Poses generated by the placement methodology can be rescored using one of the available methods. In the procedure I used, the London dG scoring function has been chosen for the first step of rescoring, it estimates the free energy of binding of the ligand from a given pose. For the second step of rescoring the GBVI/WSA dG scoring has been used; it is a forcefield-based scoring function which estimates the free energy of binding of the ligand from a given pose. For both scoring functions, lower scores indicate more favorable poses.

Poses resulting from the placement stage can be refined using either the explicit molecular mechanics forcefield method or the grid-based energetics method. In the used Forcefield Refinement**,** energy minimization of the system is carried out using the conventional molecular mechanics setup.

4.2.5 Phe docking into putative allosteric site on the dimeric PAH forms

Phe docking has been performed into one putative allosteric site (between the RD of chain A and the CD of chain B) on different conformations of the unbound dimeric enzymes. As template for the position of the Phe and for the identification of the Phe-binding site, the Phe bound structure of the protein prephenate dehydratase (PDT) has been used. Complexes were generated with MOE software using different

conformations of the unbound dimeric enzymes: (i) the initial models at 0 ns, (ii) the minimum energy structures extracted from the previous MD simulations on the unbound form, (iii) the representative structures of the first most-populated cluster of protein configurations sampled by the simulations of the unbound form. In addition, (iv) complexes were generated superimposing the initial dimeric models and the Phe in the PDT structure. Furthermore, to verify the presence of a site in the putative region that we identified for the Phe-binding, MOE's SITE FINDER (Chemical Computing Group Inc., 2012) was used and some complexes were generated using the site found by such program. The latter complexes resulted very similar to the complexes generated using as template the PDT binding site and hence they were not subjected to MD simulations.

4.2.6 Analysis Tools

Cluster analysis was performed using the GROMOS method (Daura et al., 1999). Structures of the enzymes were taken from the MD trajectory at 100 ps intervals. The protein configurations sampled by the simulations are assigned to classes defined by similarity. The criterion of similarity for two structures was the backbone atom positional RMSD $\leq$ 0.12 nm. For each cluster a structure represents the class of structurally similar configurations.

To obtain the minimum energy structure of the unbound PAH forms for docking simulations, potential energy of all structures, produced during MD simulations of the three PAH forms, was calculated using g_energy program in GROMACS. This program extract the energy components from the energy files. Subsequently, the minimum energy structure was extracted from the files obtained from g_energy using a home-made script.

The electrostatic potential surfaces of the monomeric and dimeric forms of hPAH and PDT were calculated with PyMOL (Schrödinger, 2009). PyMOL can call upon the Adaptive Poisson-Boltzmann Solver (APBS) under a Unix environment to perform quantitative electrostatic calculations. As a first step, the PDB2PQR server (Dolinsky et al., 2004) was used to generate a PQR file. PDB2PQR is a software package that automates many of the common tasks of preparing structures for, by among others, electrostatics calculations, providing a utility for converting protein files in PDB format to PQR format. The PQR file contain the same information as the PDB file and, in addition, the hydrogen atoms, information about titration states (protonation), and

charges and radius parameters for each atom. PyMOL provides both the execution of APBS and the visualization of the resulting electrostatic potentials.

The calculation of the evolutionary conservation in the structure of hPAH and PDT was made with the ConSurf server (Ashkenazy et al., 2010; Glaser et al., 2003; Landau et al., 2005). The ConSurf server is a bioinformatics tool for estimating the evolutionary conservation of amino/nucleic acid positions in a protein/DNA/RNA molecule based on the phylogenetic relations between homologous sequences. Given the amino or nucleic acid sequence (can be extracted from the 3D structure), ConSurf carries out a search for close homologous sequences using BLAST (or PSI-BLAST) (Altschul et al., 1990; Altschul et al., 1997). A multiple sequence alignment (MSA) of the homologous sequences is constructed using T-COFFEE (Notredame et al., 2000). The MSA is then used to build a phylogenetic tree using the neighbor-joining algorithm as implemented in the Rate4Site program (Pupko et al., 2002). Position-specific conservation scores are computed using the empirical Bayesian or ML algorithms (Mayrose et al., 2004; Pupko et al., 2002). The continuous conservation scores are divided into a discrete scale of nine grades for visualization, from the most variable positions (grade 1) colored turquoise, through intermediately conserved positions (grade 5) colored white, to the most conserved positions (grade 9) colored maroon. The conservation scores are projected onto the protein/nucleotide sequence, on the structure and on the MSA.

4.3 *Results*

In silico models were generated for the dimer (residues 33-424 of chain A and chain B) of wt-hPAH enzyme and of the G46S hPAH mutant, using as template the dimeric wt-rPAH structure. In Figure 4.2 the wiring diagram of the generated dimeric hPAH models is shown. Two replicas of MD simulations were performed onto the generated models and the dimeric wt-rPAH template. Overall, during the simulations, the systems showed no large deviations from the initial structures and at the end of the simulations all trajectories reached a plateau (hPAH in 40 ns, rPAH and G46S in 30 ns) within 4-5 Å in the RMSD from the initial structures calculated over C^{α} atoms.

4.3.1 Mobility of the unbound dimeric PAH forms

The dynamics of wt-hPAH is mostly characterized by the large motions of the RD, the hinge region Arg111-Thr117 and the segments 135-151 (L5), 336-339 (L6), 360-371(L7), 376-384 (L8) in the CD (Figures 4.3, 4.4, 4.5).

During all simulations, in at least one chain of the dimers, a high RD mobility was observed, confirming the high flexibility found in our previous MD simulations on the isolated RD (Figures 4.4, 4.5A). Particularly, in the chain A of wt-hPAH, the two α-helices and the loop-containing regions (from L1 to L4) show higher mobility with respect to the β-sheet (Figures 4.4, 4.5A).

A large rearrangement is also observed for the loop Arg111-Thr117 (Figures 4.4, 4.5B), connecting the RD to the CD (Figure 4.3). The loop represent the center of hinge-bending motions of the two domains.

Moreover, in the CD dynamics, the segments L5, L6, L7, L8 showed large motions (Figures 4.4, 4.5C). The segment L5 is a long loop located at the surface and connecting two α-helices (Cα1 and Cα2) of the CD (Figure 4.3). The segment L6 consists of a β-hairpin (between Cβ3 and Cβ4) located at the surface of the protein (Figure 4.3). The segment L7 contains both loop and secondary structure elements in a surface region (Figure 4.3). The L7 and L8 segments are divided by a short helix (Cα11) on the protein surface that results quite flexible as well. The L8 segment is a surface loop (Figure 4.3) that closes over the active site cavity when both BH4 and the substrate are bound.

As shown by Essential Dynamics (ED) analyses, the movements of RD, of the hinge 111-117 segment and of the abovementioned CD loops are represented by the first PC of the motion in the human wild-type enzyme (Figures 4.5A-C).

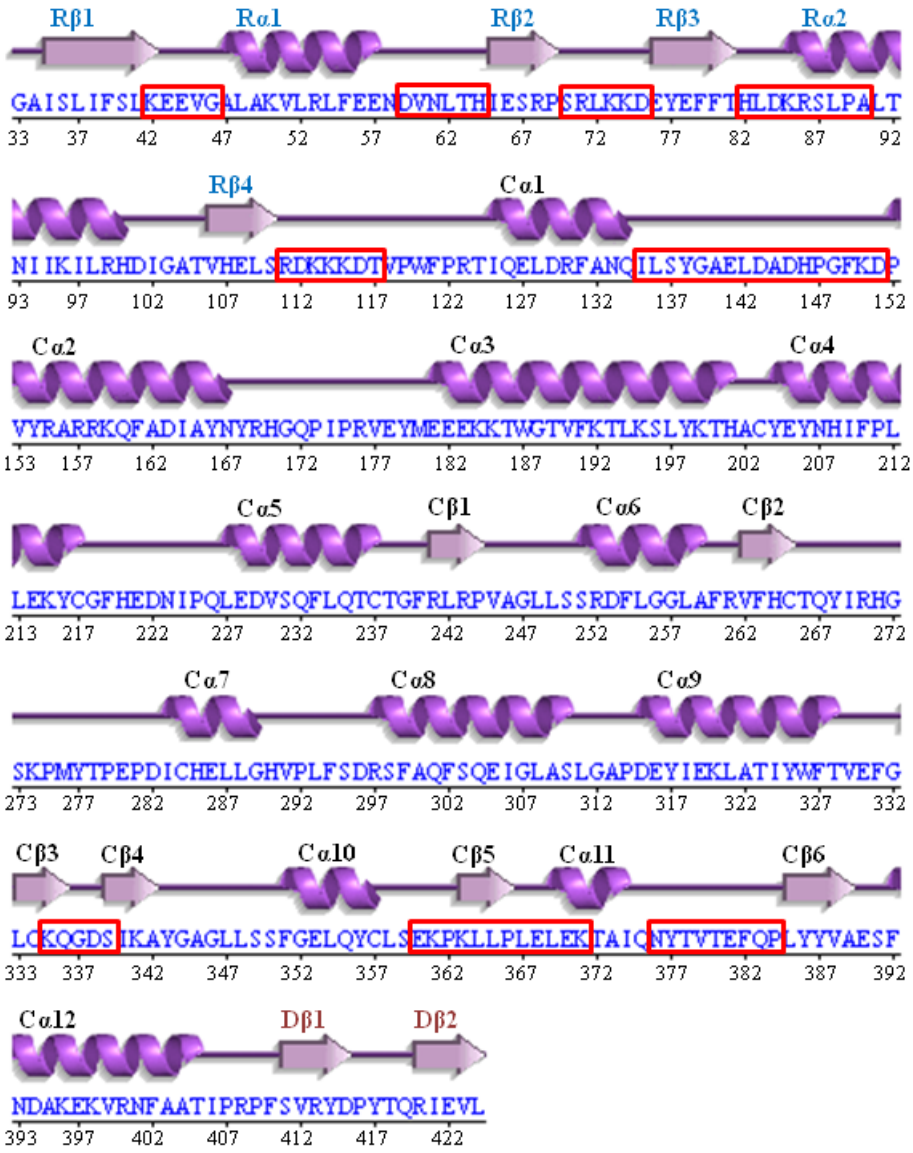

Figure 4.2: Primary and secondary structure of a subunit belonging to the *in silico* generated hPAH dimer model. Letters R, C and D associated to the secondary structure nomenclature refer to Regulatory domain, Catalytic domain and Dimerization motif. Flexible loop-containing regions (L1 to L4 in the RD, hinge 111-117, L5 to L8 in CD) are highlighted in red boxes. Figure from PDBsum server (Laskowski et al., 1997).

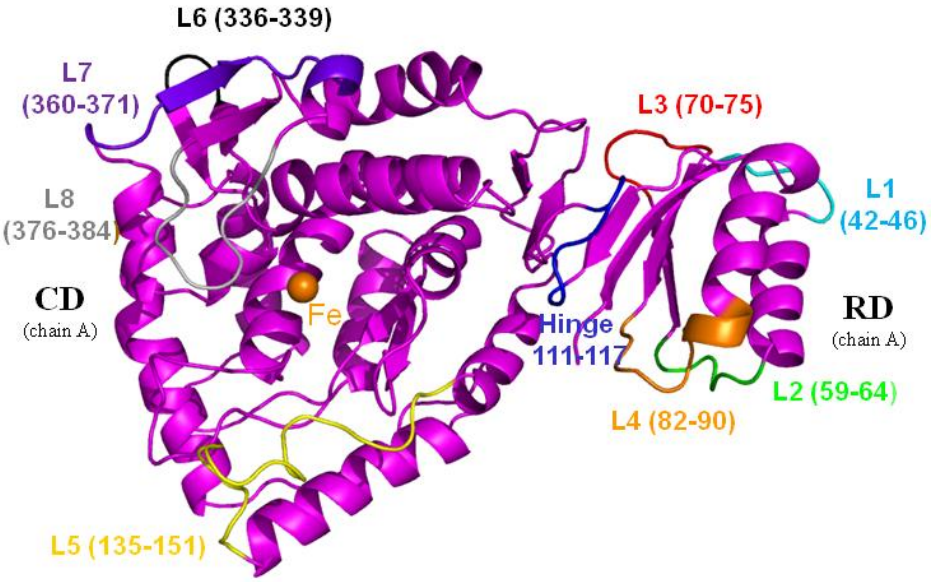

Figure 4.3: Starting model of a subunit of wt-hPAH dimer. The iron is shown as an orange sphere. Mobile loop-containing regions L1 to L4 in RD, the hinge region between RD and CD, regions L5 to L8 are drawn in different colours.

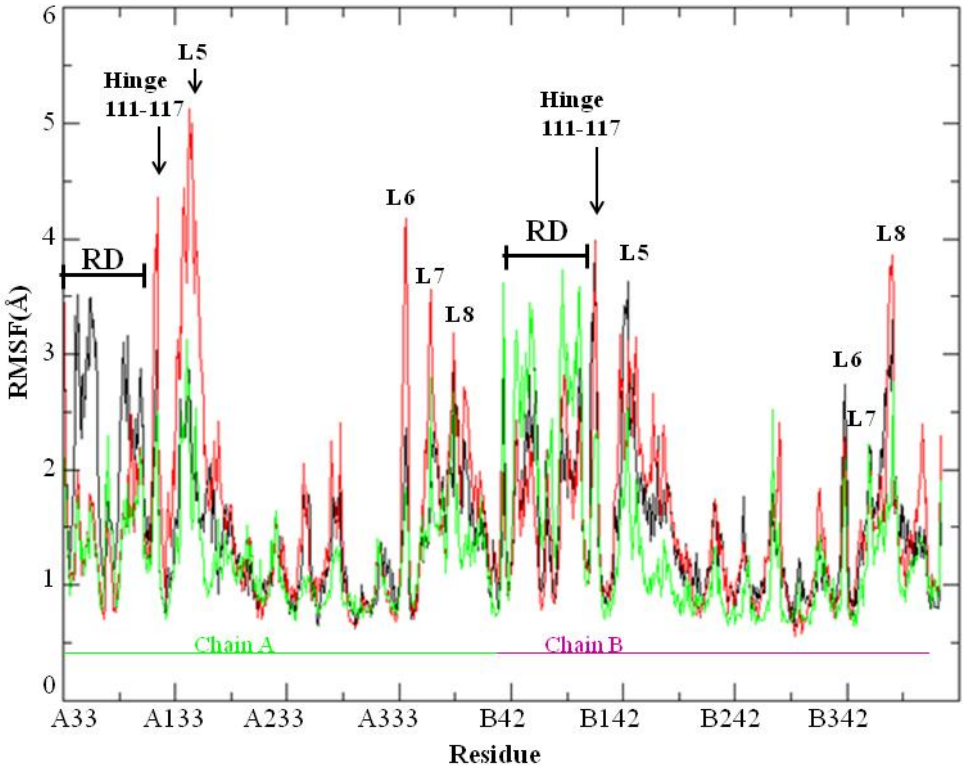

Figure 4.4: Root Mean Square Fluctuation (RMSF) of the C^{α} atoms during the simulations of the unbound PAH forms. Color code: wt-hPAH in black, wt-rPAH in red, G46S in green.

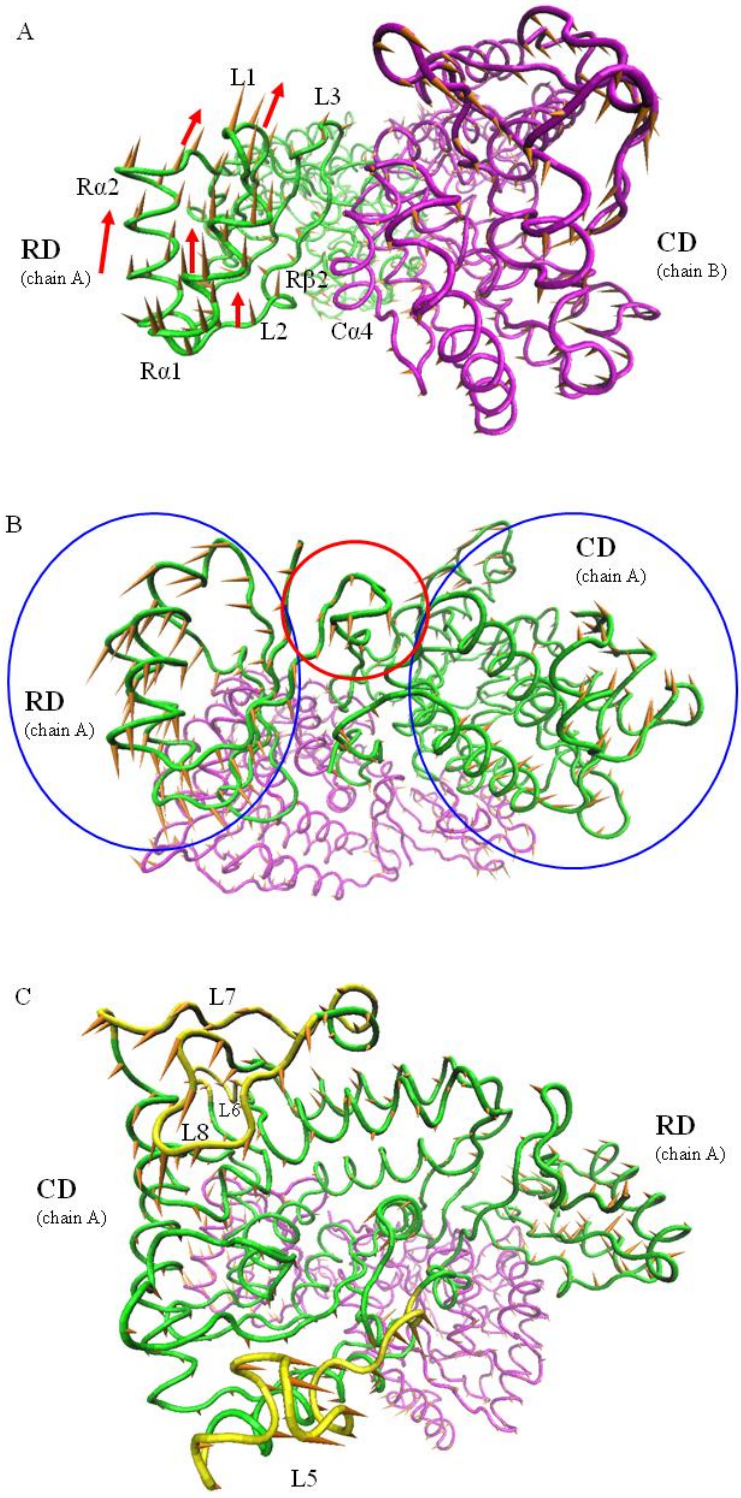

Figure 4.5: Porcupine plot illustrating the motions represented by the first PC from the simulation of the wt-hPAH. Each Cα atom has a cone attached pointing in the direction of motion described by the eigenvector corresponding to that atom. Chain A is drawn in green, chain B in magenta. Three different orientations show the movements of (**A**) the RD, where red arrows highlight its motion, of (**B**) the hinge region 111-117 highlighted in a red circle, and of (**C**) the surface loop-containing regions L5 to L8 in the CD drawn in yellow.

The principal movement of the RD and the hinge region results in the RD shift with respect to the CD of the adjacent subunit within the dimer (Figure 4.5A), with changes in the distances between residues at the dimeric interface (Figure 4.6A). The contacts formed by selected residues in the putative Phe-binding site of wt-hPAH (Glu43 (A), Gly 46 (A), Leu48 (A), Ala49 (A), Leu52 (A), Tyr77 (A), His208 (B), Pro211 (B), Leu212 (B), Lys215 (B)) were calculated on the first 5ns and on the last 5ns of the simulations by the minimum distance matrices (Figure 4.6B). The distances of residues Leu48 (A), Ala49 (A), Leu52 (A), Tyr77 (A) from His208 (B) and the distances of Leu52 (A) from Pro211 (B) and Lys215 (B) decrease (Figure 4.6B). Thus, regions involved in the RD/CD interface and close to the putative Phe-binding site show relevant differences from the initial structures.

The conformational stability of residues within 6 Å from the catalytic iron was assessed measuring during simulations the RMSD of C^{α} atoms from their initial coordinates. These residues can be functionally important as they can play an important role in biological activity of the molecule. All these residues were found to be very stable in all systems.

Similar features were also observed for the rat dimer and the G46S mutant (Figure 4.4). During each simulation, the diverse rearrangement of the hinge 111-117, exploring the conformational space, drives the orientation of the RD respect to the CD within the dimer in all MD runs. However, compared to the wt enzymes, the mutant shows lower flexibility of the RD (Figure 4.4), where the mutated residue is located.

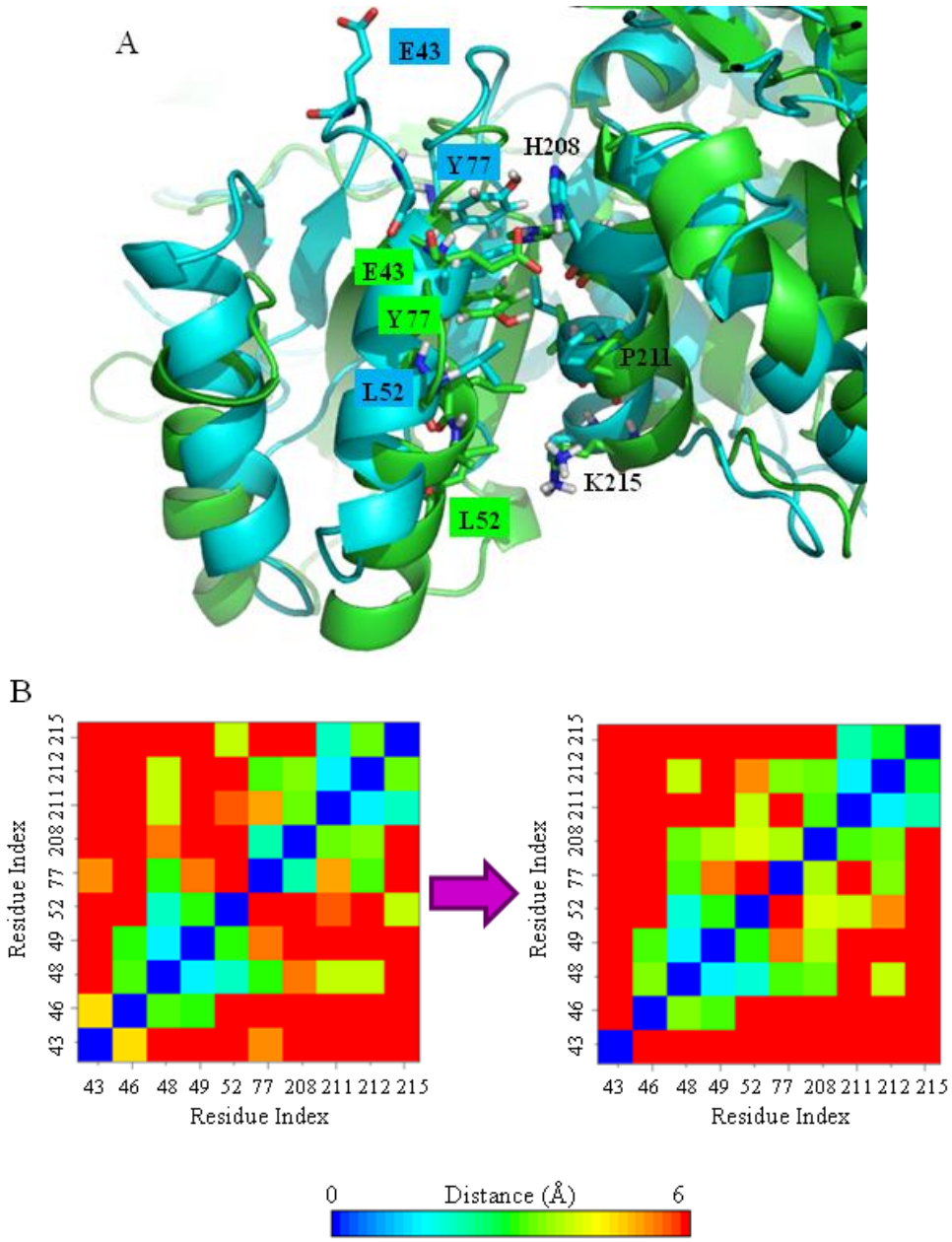

Figure 4.6: (**A**) Superimposition of the ribbon structure at 0 (green) and 50 (cyan) ns of wt-hPAH with selected residues in the putative Phe-binding site as stick. Some residues are labeled. (**B**) Distance matrices calculated on selected residues in the putative Phe-binding site of wt-hPAH (Glu43 (A), Gly 46 (A), Leu48 (A), Ala49 (A), Leu52 (A), Tyr77 (A), His208 (B), Pro211 (B), Leu212 (B), Lys215 (B)). Average distances calculated over (rigth) the first (0-5ns) and (left) the last (45-50ns) 5 ns of simulation are reported. Distances between pairs of residues are calculated as the minimum over all possible pairs of non-hydrogen atoms of the side chain.

4.3.2 Docking poses of the Phe-bound dimeric PAH forms

Taking into account (i) the experimental information about the residues in the RD that are most affected by the Phe-binding (Gjetting et al., 2001b), (ii) the data obtained by our previous MD simulations (Chapter 3) and (iii) the information about the position of the allosteric Phe-binding site in the ACT-containing PDT protein (Tan et al., 2008), we generated models of the PAH/Phe complex by molecular docking.

The docking procedure was driven with the MOE software (Chemical Computing Group Inc., 2012) using diverse conformations of the unbound hPAH, rPAH and G46S mutant dimers (see section 4.2.5). A detailed scheme summarizing all the generated complexes is shown in Table 4.1. Althougth the Phe-bound structure of the PDT has been used as template for the position of the Phe in the dimers, in all the docking simulations Phe has been placed in a region slightly different from that of the template, i.e. a region more external of the protein. However, the binding site of all the generated poses is located in proximity of the L1 loop and β2 strand in the RD at the interface with the Cα4 helix formed by residues 204-216.

Because of the great number of all the generated complexes and the similarity of some poses, only some selected poses were subsequently subjected to MD simulations. These poses were chosen on the basis of three criteria: (i) the good agreement between the position of the Phe docked and the position of the Phe in the template structure; (ii) the analysis of the protein-ligand interactions in order to compare the poses; (iii) the choice, for each performed docking simulation, of the pose with the lowest MOE's score. The poses that satisfied the three criteria for each PAH forms resulted in: (i) the complex formed by the initial conformation of the unbound forms (complex1), (ii) the complex formed by the minimum energy conformation extracted from the MD simulations on the unbound forms (complex2).

In figure 4.7 the Phe-PAH interactions in the binding site of the selected complexes are shown.

Table 4.1: Scheme summarizing all the generated complexes.

Complex generated by MOE	**PAH structure used in docking simulations**	**Ligand**
Complex1	Structure of PAH dimeric forms at 0 ns	Phe
Complex2	Structure at minimum energy extracted from the MD simulations of the unbound PAH dimeric forms	Phe
Complex3	Representative structure of the first most-populated cluster sampled by the simulations of the unbound forms	Phe
Complex generated by superposition	**Structures used**	**Ligand**
Complex4	Structure of PAH dimeric forms at 0 ns + structure of PDT	Phe

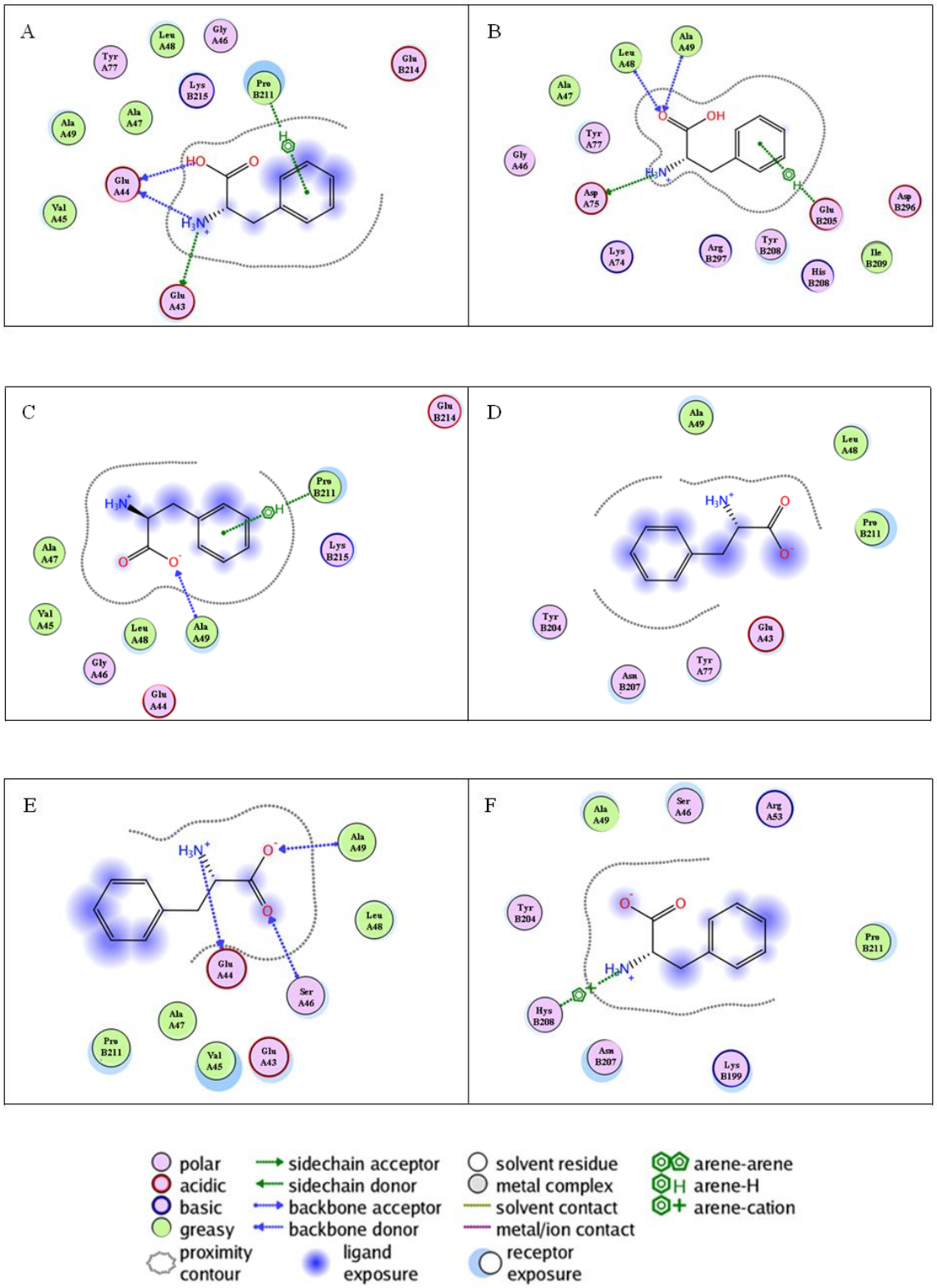

Figure 4.7: Phe-PAH interactions diagrams of the selected docked complexes. Interactions in: (**A**) hPAH complex1, (**B**) hPAH complex2, (**C**) rPAH complex1, (**D**) rPAH complex2, (**E**) G46S complex1, (**F**) G46S complex2. Figure from the MOE's ligand interaction tool (Chemical Computing Group Inc., 2012).

4.3.3 Analysis of the Phe-binding site in hPAH and PDT

To investigate the diversity of the Phe position in the hPAH complexes from the template, the electrostatic surface and the evolutionary conservation of residue positions of the binding site in hPAH and PDT has been calculated.

The electrostatic potential surfaces of the monomers and dimers were calculated with PyMOL (Schrödinger, 2009). The resulting electrostatic potential is mapped onto the molecular surface in Figure 4.8, where the proteins are shown as an electrostatic surface colored blue in the positive regions and red in the negative regions. The analysis showed that the Phe-binding site in both enzymes is positively charged. However, in the PDT the positive charge is more localized and buried in the site (Figure 4.8A), whereas in the hPAH the positive charge is also extended on the surface (Figure 4.8B).

The calculation of the evolutionary conservation in the structure of hPAH and PDT was made by the ConSurf server (Ashkenazy et al., 2010; Glaser et al., 2003; Landau et al., 2005). The ConSurf's conservation score at a site corresponds to the site's evolutionary rate: positions that evolve slowly are commonly referred as "conserved", whereas positions that evolve rapidly are referred as "variable". Usually, conserved patches suggest functions as in enzymatic activity or in ligand binding or, alternatively, in protein-protein interactions. In Figure 4.9, where the nine-color conservation scores are projected onto the 3D structures, we can see the conservation pattern of the monomer of the enzymes, which shows that ConSurf clearly detects the conservation among the amino acids facing the Phe-binding region. In fact, the analysis revealed, as expected, that the binding regions of these proteins are highly conserved. In particular, the glycines in the loop between the first β-strand and the first α-helix were assigned the highest conservation scores (Figure 4.9). However, in the PDT the conservation is higher only for the residues interacting with Phe (Figure 4.9), whereas in the hPAH a higher number of residues in the same region has a good conservation score (Figure 4.9). Thus, the PDT seems to have a more centered conserved Phe-binding site than hPAH.

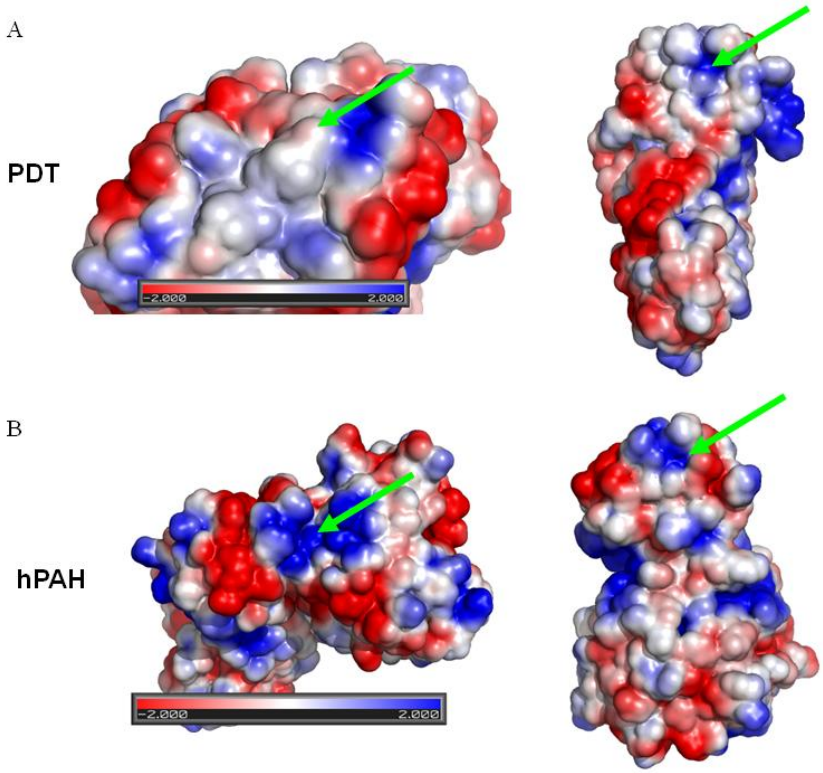

Figure 4.8: The calculated electrostatic potential is mapped onto the molecular surface of (**A**) the dimeric (left) and monomeric (right) PDT, and (**B**) the dimeric (left) and monomeric (right) hPAH. The Phe-binding sites are indicated by green arrows. Figures from Pymol.

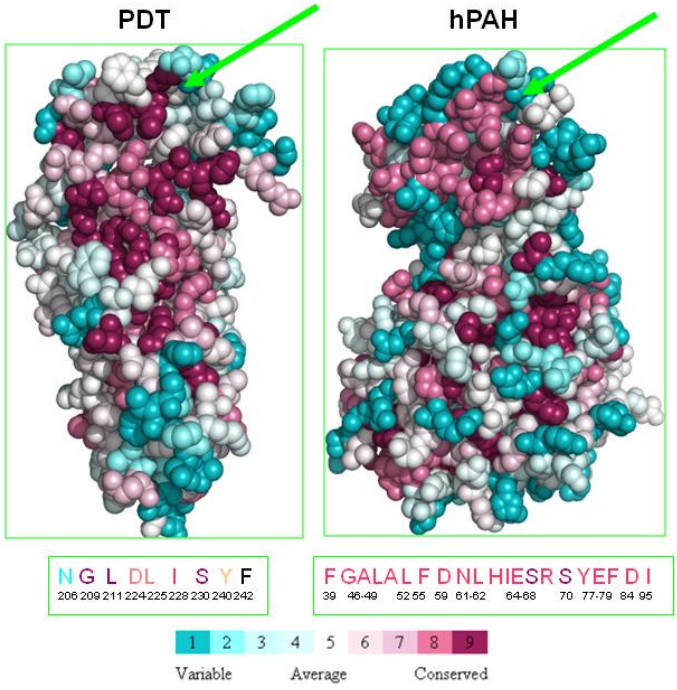

Figure 4.9: Conservation pattern of the PDT and hPAH monomers. The Phe-binding sites are indicated by green arrows and the residues of the sites are drawn in green boxes according to the nine-color conservation scores. Figures from the Consurf server.

4.3.4 MD simulations of the complexes

The selected complexes were refined by performing 10ns-long MD simulations. This helped in incorporating protein as well as ligand flexibility and also to check the stability of the complexes obtained by docking studies. Furthermore, it was interesting to investigate whether the binding pocket residues, identified by docking, persisted and if new contacts were formed throughout the MD simulations.

For all three PAH forms, the trajectories of the complexes2 were more stable over the simulations than the trajectories of complexes1 and reached a plateau after 2-3ns (Figure 4.10), indicating their higher stability and reliability.

In both wt-hPAH complexes, the RD of the chain A, which binds Phe, has less mobility than in the unbound enzyme (Figure 4.11), indicating that Phe-binding reduces the movements of this domain. Instead, in both complexes of the wt-rPAH and of the G46S mutant, the mobility of RD in the chain A is similar to that of the unbound forms.

During the simulations, in the wt-hPAH complex1 Phe explored diverse conformations and moved deeper into the ACT-domain core (Figure 4.12A); while in the complex2 the Phe position does not substantially change (Figure 4.12B). In Figure 4.13A and 4.13B superpositions of the structures of wt-hPAH complex1 and complex2 are shown at 0ns and 10ns respectively. These superpositions show that the positions and the orientations of Phe in the two complexes are quite different. In the wt-hPAH complex1 at the end of the simulation, the amino group forms H-bonds with the residues Glu44, Tyr77 from chain A and His208 from chain B (Figure 4.14A). Whereas, in the wt-hPAH complex2, the carboxylic group forms H-bonds with residues Ala47, Leu48, Lys49, Lys74, Tyr 77 from chain A (Figure 4.14B).

For the wt-rPAH, during the dynamics, the Phe in both complexes slightly shifted towards the N-terminal part of the Cα4 helix and its aromatic ring was pointing inwards of the enzyme (Figures 4.15A, B). Also for the rat form, the superpositions of the two complexes (Figures 4.16A, B) show that the positions and the orientations of Phe in the two complexes are quite different, with the polar moiety pointing in diverse directions (towards the enzyme in complex1 and towards the surface in complex2). In the wt-rPAH complex1 at the end of the simulation, the amino group forms a H-bond with the residue Glu43 from chain A and the carboxylic group forms H-bonds with residues Glu44, Gly46, Ala47, Leu48 from chain A (Figure 4.17A). Whereas, in the wt-rPAH

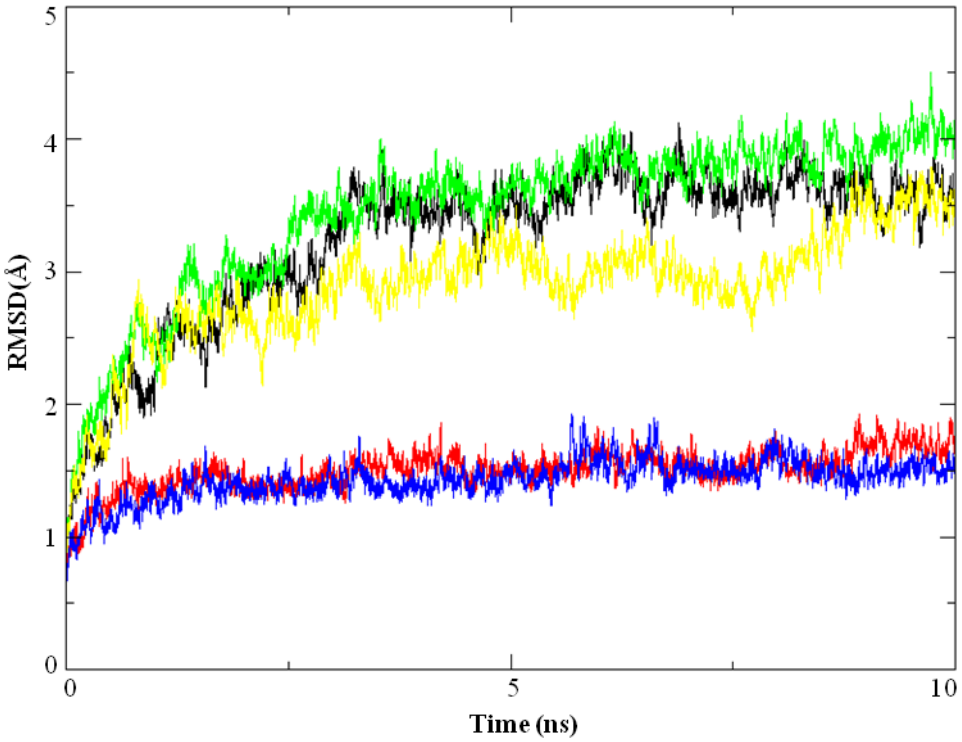

Figure 4.10: Root Mean Square Deviation (RMSD) of the C^{α} atoms during the simulations of all analysed complexes. Color code: hPAH complex1 in black, hPAH complex2 in red, rPAH complex1 in green, rPAH complex2 in blue, G46S complex1 in yellow.

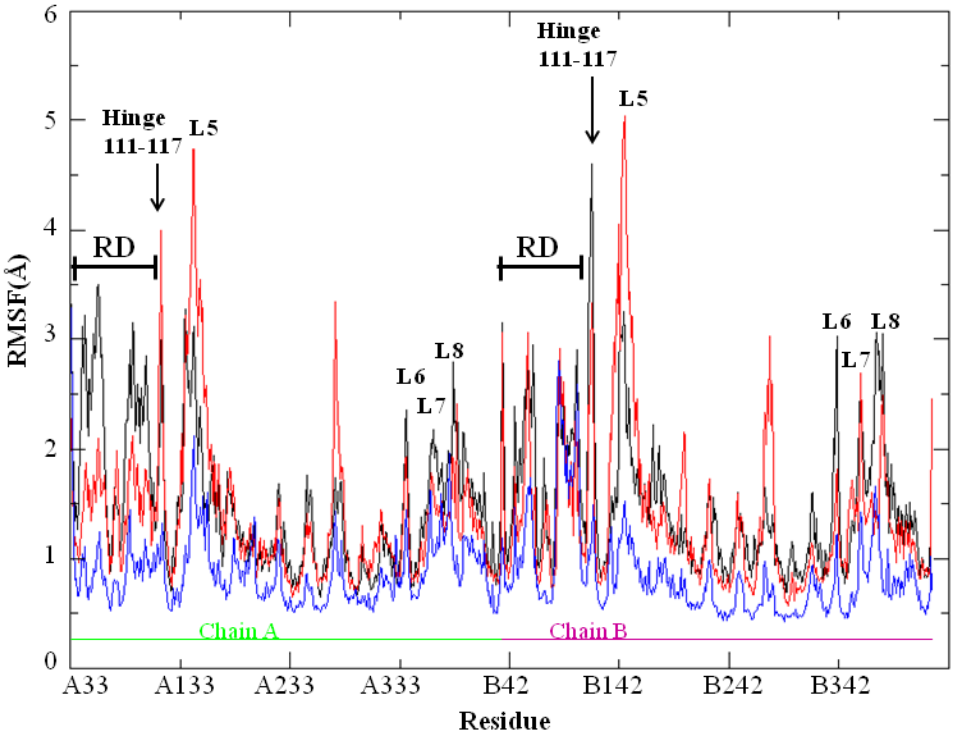

Figure 4.11: Root Mean Square Fluctuation (RMSF) of the C^{α} atoms during the simulations of the unbound hPAH in black, the hPAH complex1 in red and the hPAH complex2 in blue.

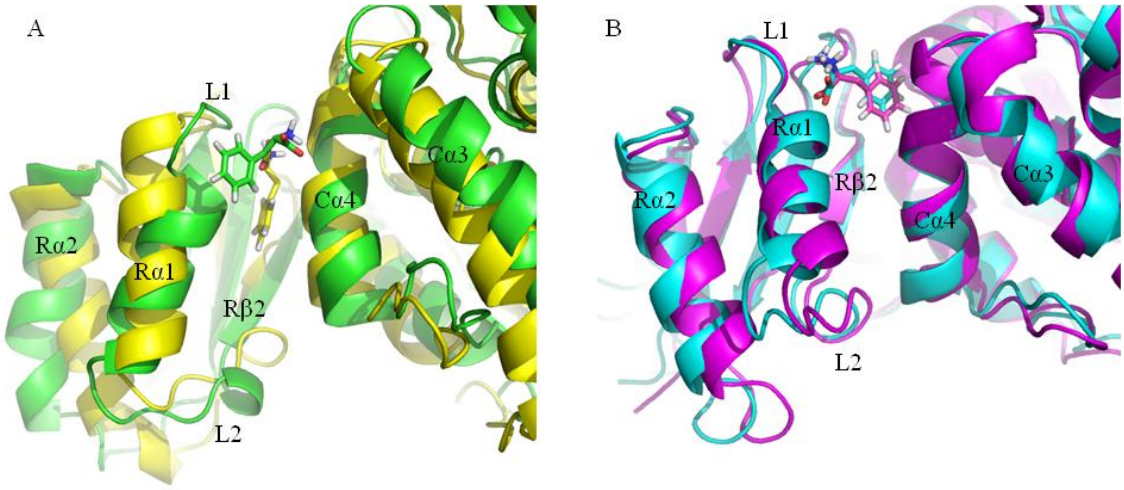

Figure 4.12: (**A**) Superimposition of the structure at 0 (green) and 10 (yellow) ns of hPAH complex1. (**B**) Superimposition of the structure at 0 (cyan) and 10 (magenta) ns of hPAH complex2. Phe is shown as stick.

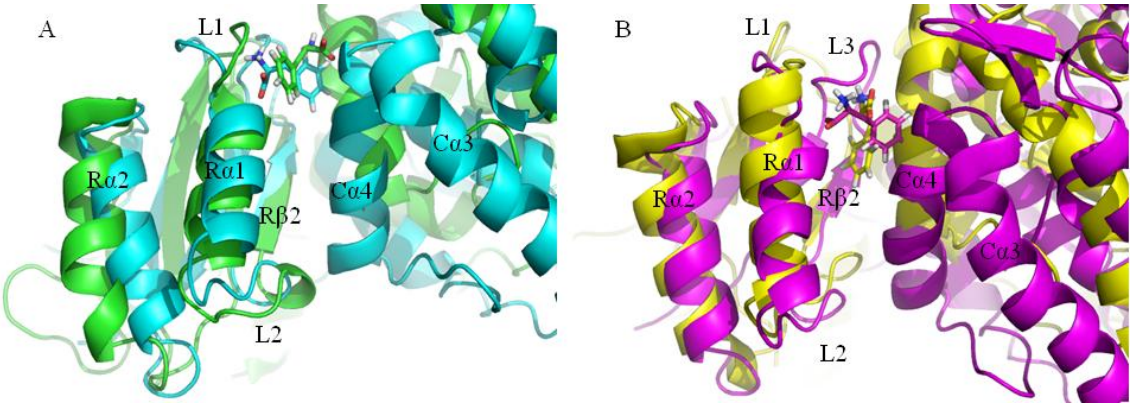

Figure 4.13: (**A**) Superimposition of the structure of hPAH complex1 (green) and complex2 (cyan) at 0 ns. (**B**) Superimposition of the structure of hPAH complex1 (yellow) and complex2 (magenta) at 10 ns. Phe is shown as stick.

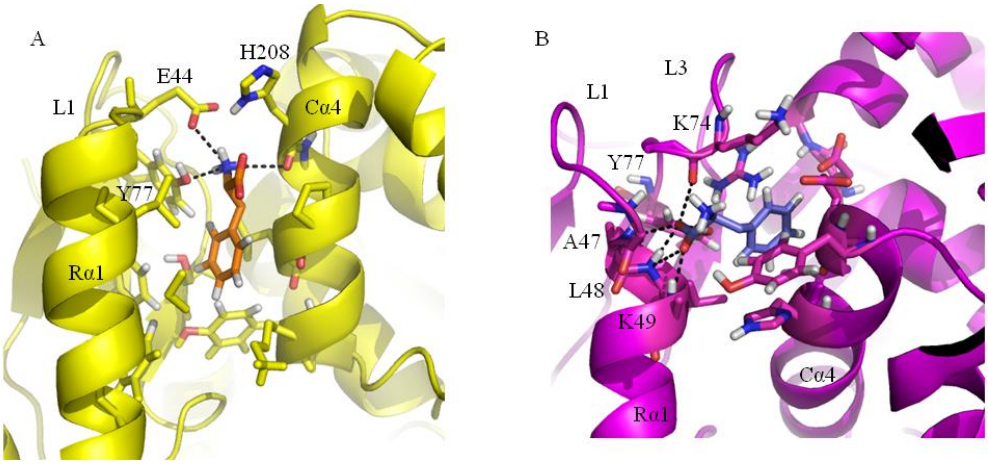

Figure 4.14: (**A**) H-bonds of Phe, shown as orange stick, in hPAH complex1 at 10 ns. (**B**) H-bonds of Phe, shown as blue stick, in hPAH complex2 at 10 ns (see text for details).

complex2, the amino group forms a H-bond with the residue Lys199 from chain B and the carboxylic group forms a H-bond with residues Asn207 from chain B (Figure 4.17B).

For the G46S mutant, the dynamics of Phe are very different from that of the other systems. In fact, only in 1 out of 4 simulations (two replicas per each complex) the Phe remained in the putative allosteric site. In the other simulations, Phe explored regions of the space far away from the mutant after 1-5 ns. At the end of one simulation of the complex1, the Phe was located closer to the CD than before MD (Figure 4.18A), forming H-bonds between the amino group and the residues Glu43 (from chain A), His208 (from chain B) and between the carboxyl group and the mutate residue Ser46 (Figure 4.18B).

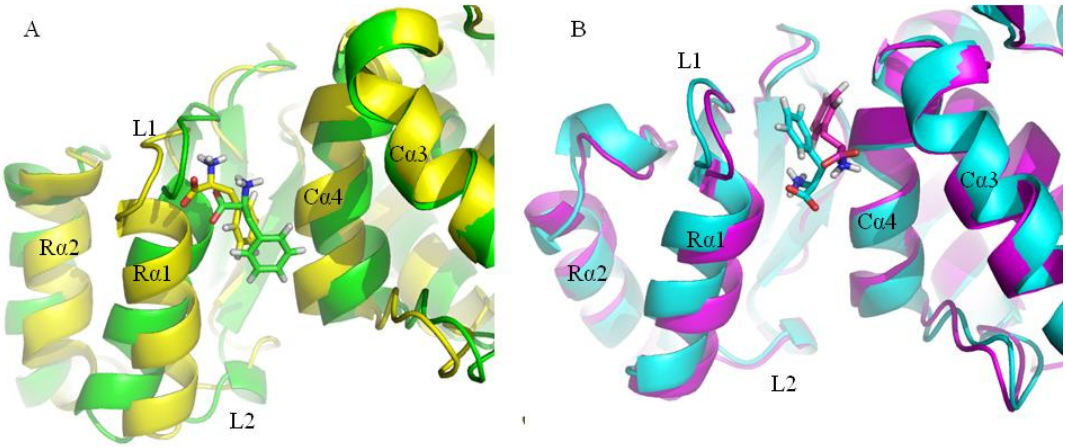

Figure 4.15: (**A**) Superimposition of the structure at 0 (green) and 10 (yellow) ns of rPAH complex1. (**B**) Superimposition of the structure at 0 (cyan) and 10 (magenta) ns of rPAH complex2. Phe is shown as stick.

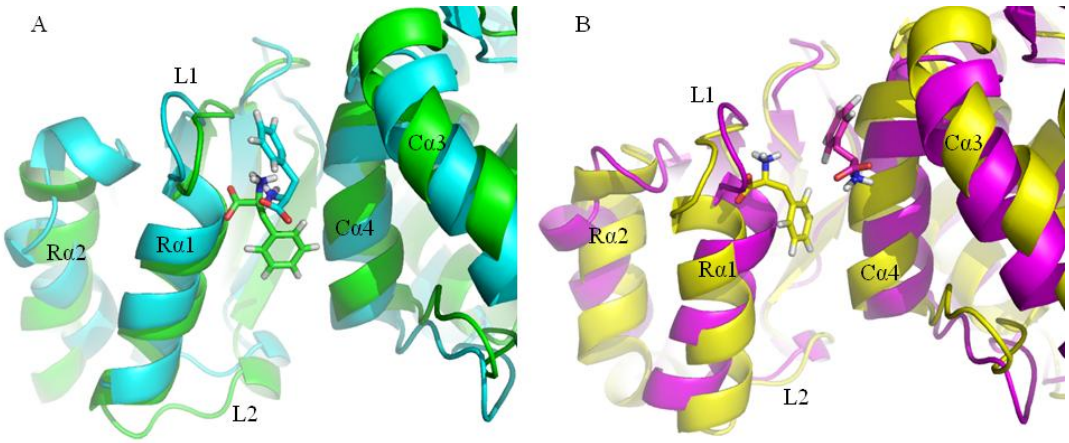

Figure 4.16: (**A**) Superimposition of the structure of rPAH complex1 (green) and complex2 (cyan) at 0 ns. (**B**) Superimposition of the structure of rPAH complex1 (yellow) and complex2 (magenta) at 10 ns. Phe is shown as stick.

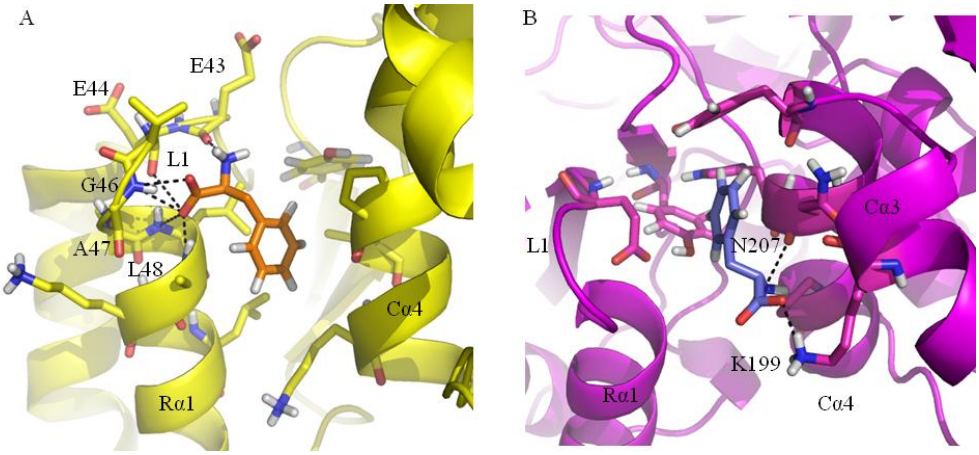

Figure 4.17: (**A**) H-bonds of Phe, shown as orange stick, in rPAH complex1 at 10 ns. (**B**) H-bonds of Phe, shown as blue stick, in rPAH complex2 at 10 ns (see text for details).

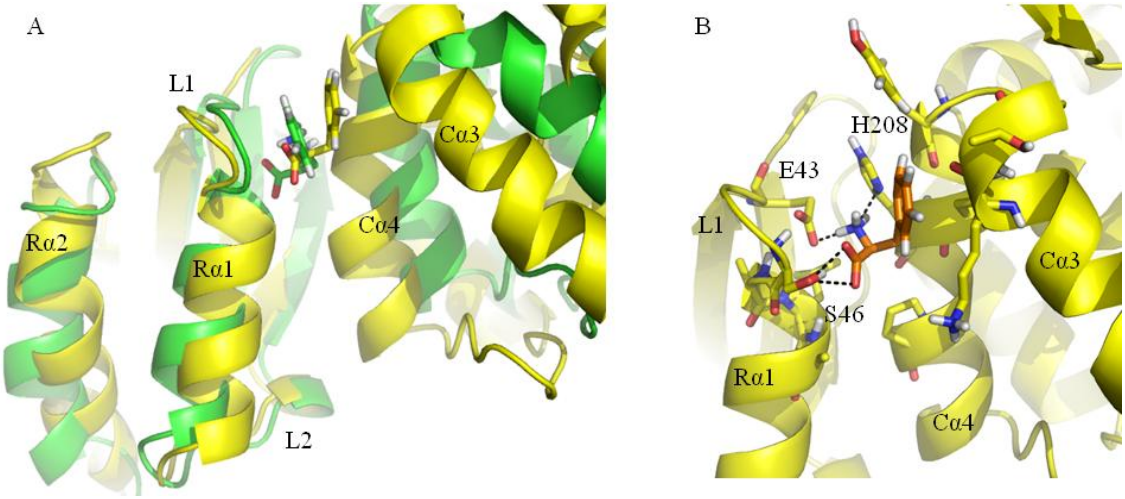

Figure 4.18: (**A**) Superimposition of the structure at 0 (green) and 10 (yellow) ns of G46S complex1. Phe is shown as stick. (**B**) H-bonds of Phe, shown as orange stick, in G46S complex1 at 10 ns (see text for details).

4.4 *Discussion*

As an identification of the proposed allosteric Phe-binding site has been lacking by now, I used computational methods, i.e. homology modelling, molecular docking and MD simulations, to achieve this goal.

MD simulations were performed on the *in silico* generated unbound and Phe-bound dimers of human and rat wild-type PAH and of human disease-causing G46S mutant. The analysis of the trajectories provided dynamical and structural insight on the unbound dimeric forms, on the complexes and on the effects of the G46S mutation.

The dynamics of the unbound dimeric PAH forms are mostly characterized by the large motions of the RD, the hinge region Arg111-Thr117 and the surface segments L5, L6, L7, L8 in the CD.

Once again, MD simulations of the unbound PAH dimeric forms highlighted the flexibility of the RD and, in particular, the high mobility of the loop containing regions from L1 to L4 in line with our previous MD simulations on the isolated ACT domain. In the dimer, L1 and L3 loops showed a large rearrangement although they form interactions with the CD of the other subunit.

As expected, very high flexibility was also detected for the hinge region Arg111-Thr117 connecting the RD to the CD from the same subunit. My results indicate that the dynamics of the enzyme is characterized by conformational changes including the domain movements around this hinge.

Furthermore, in the CD the simulated systems showed significant rearrangements of the loop-containing surface regions L5, L6, L7 and L8. These findings from MD simulations, linking dynamics of the loop-containing regions in the CD, are in good agreement with experimental data on allosteric regulation of PAH (Li et al., 2010).

Li and co-workers showed, by an increase in exchange kinetics of hydrogen/deuterium exchange, that these regions are highly flexible, solvent-accessible and, in addition, they are influenced by the presence of Phe.

The L5 region contains the 136-143 segment that is usually not seen in crystal structures of the AAAHs in the absence of the substrate, consistent with its flexibility. The L5 loop is in close proximity to the active site and was found to be highly flexible also in the MD simulations performed by Fuchs and co-workers (Fuchs et al., 2012). These authors observed a loop movement towards the active site in the oxidized state of PAH

(containing an artificial disulfide bond between two vicinal Cys residues), thereby partially blocking access for the substrate (Fuchs et al., 2012). Moreover, with respect to the binary cofactor complex crystal structure (ΔN102/ΔC24-hPAH-Fe(II)·BH4), in the ternary complex, hinge-bending motions, upon Phe-binding, result in a displacement of L5 flexible surface loop (Andersen et al., 2002; Andersen et al., 2003) and a relocation of the side chain of Tyr138 from a surface position to a partially buried position within a hydrophobic core at the active site. A similar movement of the side chain of Tyr138 was also observed in my MD simulations; it leads to a decrease in the accessibility of the catalytic site.

Also the L8 dynamic loop closes over the Phe when both BH4 and Phe are bound to the catalytic site in the crystal structure (Li et al., 2010).

Thus, this dynamic behavior of the surface loop-containing regions (L5 to L8) from my MD simulations agrees with data from literature and confirms the role of the movements of L5 and L8 loops in the catalytic mechanism.

MD simulations of the Phe-bound PAH dimeric forms were performed on selected complexes generated with MOE, starting from different conformations of the unbound enzymes (see sections 4.2.5 and 4.3.2). In all the docking simulations, Phe has been placed in a region slightly different from that of the Phe-binding site of PDT used as template, i.e. a region more external of the protein. To better compare the binding site of the ACT domain-containing protein PDT and the corresponding region in hPAH, the electrostatic surfaces and the residues conservation of the binding site were evaluated. Evaluation of the electrostatic properties of biomolecules has become an important tool, as many molecules are charged and electrostatic forces are relatively long range, the electrostatic properties have the potential to accelerate association of molecules, and are thus an important contributor to molecular recognition, protein-ligand, protein-membrane, and protein-protein binding. Another important tool is the calculation of the evolutionary conservation of residues positions. Usually, conserved patches suggest functions as in enzymatic activity or in ligand binding or, alternatively, in protein-protein interactions. These analysis reveled a more centered positive charge and residue conservation in the binding site of PDT than hPAH, indicating a higher specificity for the PDT Phe-binding site.

However, during MD simulations, the trajectories of the selected Phe-bound hPAH and rPAH reached a plateau, indicating their stability and reliability. Although the Phe position slightly differs in the different analysed systems, the binding site over the simulations remained localized at the dimeric RD/CD interface and some residues from L1, α1 (which includes the GAL sequence motif) and Cα4 were involved in the binding. These results confirm my starting hypothesis on the Phe-binding site.

For the hPAH, I also found a decrease in the mobility of the RD where the Phe is bound in the complexes. This interesting finding indicates that Phe-binding reduces the movements and probably enhances the stability of this domain.

Moreover, during the simulations, the complexes2 resulted more stable than the complexes1, and Phe in complex2 moved much less than in complex1. This suggests that starting from a more stabile protein structure, as that used in complex2, improves the reliability of the docking simulation predictions.

For the G46S mutant, during the MD simulations, the dissociation of most of the complexes occurred. Furthermore, in the only stable complex, Phe formed a H-bond with the side chain of the mutated residue, i.e. Ser46. These results suggest that the mutation affects the correct binding of Phe, as we predicted in a previous study (see section 3.4). The substitution of the Gly in the conserved GAL sequence motif with Ser, having a polar side chain that obstructs the allosteric site, leads to an unproductive Phe binding. The G46S mutant has a reduced stability both *in vitro* and *in vivo* that results in misfolded protein and in a severe form of PKU (Goodwill et al., 1997). Phe binding to the wt-hPAH showed a stabilizing effect, and a loss of this ability in the G46S mutant may contribute to explain its low stability and hence its disease-causing nature.

In conclusion, this work paves the way for more detailed theoretical investigation aimed at validating my models and sets the basis for site-directed mutagenesis experimental work on the proposed Phe binding-site and on the interface of the PAH dimeric structure.

CHAPTER 5. *Conclusions*

Phenylketonuria (PKU) is an autosomal recessive disease that leads to severe mental retardation in humans if left untreated. In this classical inborn error of metabolism, the gene primarily affected is the *PAH* gene, which results in a protein with reduced enzyme activity that is not sufficient to hydroxylate phenylalanine to tyrosine. PKU is considered a protein misfolding disease with loss of function. In fact, a large number of mutations result in a reduced folding efficiency and a reduced thermal/chemical stability of the enzyme.

Understanding the background of diseases is crucial to medical research, with implications in diagnosis, treatment and drug development. As molecular approaches to this challenge are costly and time consuming, computational approaches offer an efficient alternative. Thus, in my thesis, to get insight into the structural basis of the PAH defects underlying the disease, I used computational methods.

Here, my current results contribute to elucidate specific aspects of PAH and PKU related to: (i) the conformational stability of disease-causing PAH mutants (Chapter 2), (ii) the structural and dynamical features of the isolated ACT domain of the wild-type enzyme and of six mutants (Chapter 3), (iii) the structural basis for the regulation of PAH, through the identification *in silico* of the putative allosteric Phe-binding site (Chapter 4).

The studies described in the Chapter 2, showing a decrease in the PAH conformational stability of the PKU-causing mutations, revealed the relevance of conformational flexibility for PAH activity (Paper I). The decreased conformational stability of the studied hPAH mutants can be associated to possible defects in the folding and a reduced activity of the enzyme, leading to the disease.

MD studies on the isolated ACT regulatory domain (Chapter 3) highlighted a specific dynamic behaviour of the wild-type enzymes and of the six mutants. Collectively, all the movements result in a transition from an initial open to a final more closed structure, leading to the decrease of the solvent exposure of the hydrophobic surface, where the Phe ligand can bind to the domain. Moreover, the correlation, found between the solvent accessibility of the hydrophobic surface and the global intrinsic motions of the protein, suggested a relationship between the Phe-binding and the overall dynamic

behaviour of the enzyme. In the mutants, the simulations revealed changes in the interactions within the hydrophobic core of the RD that may significantly affect its stability and hence its regulatory role.

Molecular docking studies combined with MD simulations on the dimeric PAH (Chapter 4) generated models for the Phe-PAH interactions. The allosteric Phe-binding site was localized at the dimeric RD/CD interface where some residues from L1, α1 (which includes the GAL sequence motif) and Cα4 are involved in the binding. For the hPAH, the Phe-binding showed also to reduce the mobility of RD domain. The analysis of these models elucidated the structural effects of the disease-causing G46S mutant which showed to affect the correct binding of Phe.

In conclusion, this thesis adds further evidences toward a deeper understanding of the regulatory mechanisms of the enzyme and of the relationship between the mutations and the disease.

Appendix I. *Structural analysis of selected glucokinase (GCK) mutants causing MODY2*

Mutations in human genes can change the sequence and structure of a protein, impair its function, and lead to a disease. The function of most genes is determined by the 3D structure of their protein product. Interpreting the effects of mutations in a protein structure provides more detailed information about the environment and role of a mutated residue and can explain the impaired function and thus the disease.

Genetic diseases due to loss-of-function mutations, as the PKU, are very common. An additional example, among this diseases, is the type 2 Maturity Onset Diabetes of the Young (MODY2).

MODY2 is a monogenic autosomal disease characterized by a primary defect in insulin secretion and hyperglycemia. It results from glucokinase (GCK) gene mutations that impair enzyme activity. GCK catalyzes the ATP-dependent phosphorylation of glucose in the first, rate-limiting step of glycolysis in pancreatic β-cells. The 3D structure consists of two domains, namely small and large domain, separated by the catalytic site. The GCK switches from a super-open inactive conformation to a closed active conformation upon ligand binding. A huge conformational transition occurs through a large rotation of the small domain.

GCK function and its involvement in the MODY2 disease is one of the main field studied by the group of Prof. Sacchetti and her co-workers at the research institute CEINGE. During my thesis work, in collaboration with the afore-mentioned group, I carried out a structural study, by homology modelling, on the effects of some disease-causing GCK mutations (Paper II).

In the attempt to understand how detected mutations could contribute to MODY2 insurgence, we searched for structure/function correlations of the disease-causing GCK mutants. To this aim, I evaluated the impact of studied mutations (p.E40D, p.E70D, p.F150Y, p.V154L, p.A188T, p.G223S, p.A259T, p.G261R, p.D278E, p.Y289C, p.R303W, p.E395_R397del, p.K420E, p.A449T) on the enzyme 3D-structure.

All the mutations considered are buried in the enzyme, except for p.E70D which is located on the surface of the protein, and all mutations are far from the active and ATP binding sites. Apart from p.V154L and p.F150Y that undergo large movements during

the conformational transition from the super-open to the closed active form, all the mutant structures exhibit local structural alterations that well correlate to kinetic parameters and thermal inactivation data (see Paper II). In all cases, the amino acid replacement either provokes the loss of stabilizing interactions or generates unfavorable interactions that can destabilize the whole structure of the enzyme.

In conclusion, this our study, carried out by integrating DHPLC, sequencing, bioinformatics and functional analysis, provides new information about the structure-function relationship of human GCK mutations and MODY2.

Appendix II. *List of publications, communications, research activity in scientific institutions abroad and fellowships*

Publications:

- M. Cerreto, P. Cavaliere, C. Carluccio, F. Amato, A. Zagari, A. Daniele, F. Salvatore (2011). *Natural phenylalanine hydroxylase variants that confer a mild phenotype affect the enzyme's conformational stability and oligomerization equilibrium.* Biochimica et Biophisica Acta. Molecular Basis of Disease, vol. 1812; p. 1435-1445, ISSN: 0925-4439, doi: 10.1016/j.bbadis.2011.07.012.
- M. Capuano, C. M. Garcia-Herrero, N. Tinto, C. Carluccio, V. Capobianco, I. Coto, A. Cola, D. Iafusco, A. Franzese, A. Zagari, M. A. Navas, L. Sacchetti (2012). *GCK Mutation and their Characterization in MODY2 Children of Southern Italy*. PLoS ONE. 12;7(6):e38906. doi: 10.1371/journal.pone.0038906.

Manuscript in preparation:

- C. Carluccio, F. Fraternali, F. Salvatore, A.Fornili, A. Zagari. *Structural features of the regulatory ACT domain in phenylalanine hydroxylase.*

Proceedings:

- M. Capuano, C. M. Garcia-Herrero, A. Zagari, C. Carluccio, N. Tinto, V. Capobianco, I. Coto, A. Cola, A. Galderisi, A. Franzese, D. Iafusco, M.-A. Navas and L. Sacchetti (2011). Epidemiology of GCK mutations in diabetic children from South Italy. In: The FEBS Journal (suppl. 1) Abstracts of the 36th FEBS Congress, vol. 278 pag 176.
- F. Fraternali, C. Carluccio, A. Fornili, A. Zagari, F. Salvatore (2010). Molecular Dynamics Simulations of Phenylalanine Hydroxylase Natural Mutants Leading to Hyperphenylalaninemia. In: "Biological Wikis", Proceedings of the Joint NETTAB-BBCC 2010 workshops. Napoli, Italy, November 29 - December 1, 2010, Roma: ARACNE editrice srl, p. 127, ISBN/ISSN: 9788854836587.

Communications:

- M. Cerreto, P. Cavaliere, C. Carluccio, A. Zagari, A. Daniele, F. Salvatore (2010). Characterization of the biochemical, biophysical and structural

phenotypes of seven PAH natural variants identified in patients affected by Hyperphenylalaninemias. In: 42° SIBioC National Congress. Roma, October 05-08, 2010, Roma, p. 524-524.

- P. Cavaliere, M. Cerreto, C. Carluccio, A. Cappuccio, A. Daniele, A. Zagari, F. Salvatore (2010). Natural phenylalanine hydroxylase mutants provoke conformational instability and oligomerization defects of the enzyme. In: The EMBO meeting. Barcelona, September 04-07, 2010, Barcelona, p.119.
- C. Carluccio, A. Fornili, A. Zagari, F. Salvatore, F. Fraternali (2011). Molecular Dynamics Simulations of Phenylalanine Hydroxylase Natural Mutants Leading to Hyperphenylalaninemia. In: 3DSIG: Structural Bioinformatics and Computational Biophysics. Vienna, July 15-16, 2011.

Research activity in Scientific Institutions abroad:

- From July to October 2010, as visiting student in the Structural Bioinformatics group at Randall Division of King's College London, to develop the project: "*Structural features of the regulatory ACT domain in phenylalanine hydroxylase*". Host researcher: Dr. Franca Fraternali.
- From September to December 2011, as visiting student in the Structural Bioinformatics group at Randall Division of King's College London, to develop the project: "*Structural features of the regulatory ACT domain in phenylalanine hydroxylase*". Host researcher: Dr. Franca Fraternali.
- From 2nd to 6th May 2012, as recipient of a short-term fellowship at Edinburgh Parallel Computing Centre, University of Edinburgh, to develop the project: *"In silico identification of the allosteric Phe-binding site in the human phenylalanine hydroxylase"*.
- From May to August 2012, as recipient of a short-term fellowship at Randall Division of King's College London, to development the project: *"In silico identification of the allosteric Phe-binding site in the human phenylalanine hydroxylase"*. Host researcher: Dr. Franca Fraternali.

Fellowships:

- From May to August 2012 short-term fellowship of the HPC-Europa2 Transnational Access Programme.

Acknowledgments

I would like to thank my supervisors, Professor Francesco Salvatore and Professor Adriana Zagari, for their valuable guidance throughout these years.
I would like to express my gratitude to Professor Francesco Salvatore for giving me the chance to work and study on the PKU project at CEINGE.
I am also deeply grateful to Professor Adriana Zagari for her scientific advice, support, encouragement and useful critiques and suggestions.

Most of the results described in this thesis would not have been obtained without a close collaboration with Dr. Fraternali's group at King's College, London. I owe a great deal of appreciation and gratitude to Dr. Franca Fraternali for hosting me in her lab and for making me fell as part of her group. She taught me many things and was an inspiration guide.
My special thanks go to Dr. Arianna Fornili, a friend and tutor who taught me everything about molecular dynamics simulations.

I would like to thank my coworkers from CEINGE and from King's College: Paola, Vincenzo, Luca, Flavia, Grace, Alessandro, Pietro, Nesrine. All of you, for our fruitful discussions, for the pleasant companionship and for the amazing atmosphere both in the work and out.

Finally, I wish to thank my parents and my sister for their support and encouragement throughout my study.

References

Abagyan, R., M. Totrov, D. Kuznetsov, 1994. Icm - a New Method for Protein Modeling and Design - Applications to Docking and Structure Prediction from the Distorted Native Conformation. Journal of Computational Chemistry 15, 488-506.

Aguado, C., B. Pérez, M. Ugarte, L.R. Desviat, 2006. Analysis of the effect of tetrahydrobiopterin on PAH gene expression in hepatoma cells. FEBS Lett. 580, 1697-1701.

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, D.J. Lipman, 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403-10.

Altschul, S.F., T.L. Madden, A.A. Schaffer, J.Z. Zhang, W. Miller, D.J. Lipman, 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389-402.

Amadei, A., A.B. Linssen, H.J. Berendsen, 1993. Essential dynamics of proteins. Proteins 17, 412-425.

Anantharaman, V., E.V. Koonin, L. Aravind, 2001. Regulatory potential, phyletic distribution and evolution of ancient, intracellular small-molecule-binding domains. J. Mol. Biol. 307, 1271-1292.

Andersen, O.A., T. Flatmark, E. Hough, 2001. High resolution crystal structures of the catalytic domain of human phenylalanine hydroxylase in its catalytically active Fe(II) form and binary complex with tetrahydrobiopterin. J. Mol. Biol. 314, 279-291.

Andersen, O.A., T. Flatmark, E. Hough, 2002. Crystal structure of the ternary complex of the catalytic domain of human phenylalanine hydroxylase with tetrahydrobiopterin and 3-(2-thienyl)-L-alanine, and its implications for the mechanism of catalysis and substrate activation. J.Mol.Biol. 320, 1095-1108.

Andersen, O.A., A.J. Stokka, T. Flatmark, E. Hough, 2003. 2.0 Å resolution crystal structures of the ternary complexes of human phenylalanine hydroxylase catalytic domain with tetrahydrobiopterin and 3-(2-thienyl)-L-alanine or L-norleucine: substrate specificity and molecular motions related to substrate binding. J. Mol. Biol. 333, 747-757.

Aravind, L., E.V. Koonin, 1999. Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches. J Mol Biol 287, 1023-1040.

Ashkenazy, H., E. Erez, E. Martz, T. Pupko, N. Ben-Tal, 2010. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucl. Acids Res., DOI: 10.1093/nar/gkq399; PMID: 20478830

Bajorath, J., R. Stenkamp, A. Aruffo, 1994. Knowledge-based model building of proteins: Concepts and examples. Protein Sci. 2, 1798-810.

Barrett, C.P., B.A. Hall, M.E. Noble, 2004. Dynamite: a simple way to gain insight into protein motions. Acta Crystallogr D Biol Crystallogr 60(Pt 12 Pt 1), 2280-7.

Berendsen, H.J.C., J.P.M. Postma, W.F. van Gunsteren, J. Hermans, 1981. Interaction models for water in relation to protein hydration. In Intermolecular Forces (Pullman, B., ed.), 331-342.

Berendsen, H.J.C., J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, 1984. Molecular dynamics with coupling to an external bath. J Chem Phys 81, 3684-3690.

Bjorgo, E., R.M. de Carvalho, T. Flatmark, 2001. A comparison of kinetic and regulatory properties of the tetrameric and dimeric forms of wild-type and Thr427-->Pro mutant human phenylalanine hydroxylase: contribution of the flexible hinge region Asp425-Gln429 to the tetramerization and cooperative substrate binding. Eur J Biochem 268, 997-1005.

Bjorgo, E., P.M. Knappskog, A. Martinez, R.C. Stevens, T. Flatmark, 1998. 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 257, 1-10.

Blau, N., F.J. van Spronsen, H.L. Levy, 2010. Phenylketonuria. Lancet 376, 1417-27.

Blundell, T.L., B.L. Sibanda, M.J.E. Sternberg, J.M. Thornton, 1987. Knowledge-based prediction of protein structures and the design of novel molecules. Nature 326, 347-52.

Bruccoleri, R.E., M. Karplus, 1987. Prediction of the folding of short polypeptide segments by uniform conformational sampling. Biopolymers 26, 137-68.

Cerreto, M., P. Cavaliere, C. Carluccio, F. Amato, A. Zagari, A. Daniele, F. Salvatore, 2011. Natural phenylalanine hydroxylase variants that confer a mild phenotype affect the enzyme's conformational stability and oligomerization equilibrium. Biochimica et Biophysica Acta 1812, 1435-1445.

Chemical Computing Group Inc., 2012. Molecular Operating Environment (MOE), 2012.10. 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7.

Chothia, C., A.M. Lesk, 1987. Canonical structures for the hypervariable regions of immunoglobulins. J. Mol. Biol. 196, 901-17.

ClinicalTrials.gov, 2010. Dose-Finding Study to Evaluate the Safety, Efficacy, and Tolerability of Multiple Doses of rAvPAL-PEG in Subjects With Phenylketonuria (PKU). Retrieved January 18,2011, from http://www.clinicaltrials.gov/ct2/show/NCT00925054?term=pal-001&rank=1.

Daniele, A., G. Cardillo, C. Pennino, M.T. Carbone, D. Scognamiglio, L. Esposito, A. Correra, G. Castaldo, A. Zagari, F. Salvatore, 2008. Five human phenylalanine hydroxylase proteins identified in mild hyperphenylalaninemia patients are disease-causing variants. Biochim. Biophys. Acta. 1782, 378-384.

Daura, X., K. Gademann, B. Jaun, D. Seebach, W.F. van Gunsteren, A.E. Mark, 1999. Peptide folding: when simulation meets experiment. Angew. Chem. Int. Ed. Engl. 38, 236-240.

De Simone, A., G.G. Dodson, C.S. Verma, A. Zagari, F. Fraternali, 2005. Prion and water: tight and dynamical hydration sites have a key role in structural stability. Proc. Natl. Acad. Sci. USA 102, 7535-7540.

Dolinsky, T., J. Nielsen, J. McCammon, N. Baker, 2004. PDB2PQR: an automated pipeline for the setup, execution, and analysis of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Research 32, W665-W667.

Dominguez, C., R. Boelens, A.M.J.J. Bonvin, 2003. HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. Journal of the American Chemical Society 125, 1731-1737.

Døskeland, A.P., A. Martínez, P.M. Knappskog, T. Flatmark, 1996. Phosphorylation of recombinant human phenylalanine hydroxylase: effect on catalytic activity, substrate activation and protection against nonspecific cleavage of the fusion protein by restriction protease. Biochem. J. 313, 409-414.

Edelsbrunner, H., M. Facello, R. Fu, J. Liang, 1995. Measuring Proteins and Voids in Proteins. Proceedings of the 28th Annual Hawaii International Conference on Systems Science., 256-264.

Eiken, H.G., P.M. Knappskog, J. Apold, T. Flatmark, 1996. PKU mutation G46S is associated with increased aggregation and degradation of the phenylalanine hydroxylase enzyme. Hum Mutat 7, 228-38.

Eisenhaber, F., P. Lijnzaad, P. Argos, C. Sander, M. Scharf, 1995. The double cube lattice method: efficient approaches to numerical integrations of surface area and volume and to dot surface contouring of molecular assemblies. J. Comput. Chem. 16, 273–284.

Erlandsen, H., 2003. Structural studies on phenylalanine hydroxylase and implications toward understanding and treating phenylketonuria. Pediatrics 112, 1557-65.

Erlandsen, H., R.C. Stevens, 1999. The structural basis of phenylketonuria. Mol Genet Metab 68, 103-25.

Erlandsen, H., R.C. Stevens, 2001. A structural hypothesis for BH4 responsiveness in patients with mild forms of hyperphenylalaninaemia and phenylketonuria. J Inherit Metab Dis 24, 213-30.

Erlandsen, H., F. Fusetti, A. Martínez, E. Hough, T. Flatmark, R.C. Stevens, 1997. Crystal structure of the catalytic domain of human phenylalanine hydroxylase reveals the structural basis for phenylketonuria. Nat. Struct. Biol. 4, 995-1000.

Erlandsen, H., A.L. Pey, A. Gamez, B. Perez, L.R. Desviat, C. Aguado, R. Koch, S. Surendran, S. Tyring, R. Matalon, C.R. Scriver, M. Ugarte, A. Martinez, R.C. Stevens, 2004. Correction of kinetic and stability defects by tetrahydrobiopterin in phenylketonuria patients with certain phenylalanine hydroxylase mutations. Proc Natl Acad Sci U S A 101, 16903-8.

Essmann, U., L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, 1995. A smooth particle mesh Ewald method. J Chem Phys 103, 8577-8593.

Fitzpatrick, P.F., 1999. Tetrahydropterin-dependent amino acid hydroxylases. Annu Rev Biochem 68, 355-81.

Fitzpatrick, P.F., 2000. The aromatic amino acid hydroxylases. Adv Enzymol Relat Areas Mol Biol 74, 235-294.

Flatmark, T., R.C. Stevens, 1999. Structural insight into the aromatic amino acid hydroxylases and their disease-related mutant forms. Chem Rev 99, 2137-2160.

Fornili, A., F. Autore, N. Chakroun, P. Martinez, F. Fraternali, 2012. Protein-Water Interactions in MD Simulations: POPS/POPSCOMP Solvent Accessibility Analysis, Solvation Forces and Hydration Sites. Methods Mol. Biol 819, 375–392.

Fuchs, J.E., R.G. Huber, S. von Grafenstein, H.G. Wallnoefer, G.M. Spitzer, D. Fuchs, K.R. Liedl, 2012. Dynamic regulation of phenylalanine hydroxylase by simulated redox manipulation. PLOs one 7, e53005.

Fusetti, F., H. Erlandsen, T. Flatmark, R.C. Stevens, 1998. Structure of Tetrameric Human Phenylalanine Hydroxylase and Its Implications for Phenylketonuria. The Journal of Biological Chemistry 273, 16962-16967.

Gersting, S., K. Kemter, M. Staudigl, D. Messing, M. Danecka, F. Lagler, C. Sommerhoff, A. Roscher, A. Muntau, 2008. Loss of function in phenylketonuria is caused by impaired molecular motions and conformational instability. Am. J. Hum. Genet. 83, 5-17.

Gersting, S.W., M. Staudigl, M.S. Truger, D.D. Messing, M.K. Danecka, C.P. Sommerhoff, K.F. Kemter, A.C. Muntau, 2010. Activation of phenylalanine hydroxylase induces positive cooperativity toward the natural cofactor. J Biol Chem 285, 30686-97.

Giovannini, M., E. Verduci, E. Salvatici, L. Fiori, E. Riva, 2007. Phenylketonuria: dietary and therapeutic challenges. J. Inherit. Metab. Dis. 30, 145-152.

Gjetting, T., M. Petersen, P. Guldberg, F. Guttler, 2001a. In vitro expression of 34 naturally occurring mutant variants of phenylalanine hydroxylase: correlation with metabolic phenotypes and susceptibility toward protein aggregation. Mol Genet Metab 72, 132-43.

Gjetting, T., M. Petersen, P. Guldberg, F. Guttler, 2001b. Missense mutations in the N-terminal domain of human phenylalanine hydroxylase interfere with binding of regulatory phenylalanine. Am J Hum Genet 68, 1353-60.

Glaser, F., T. Pupko, I. Paz, R.E. Bell, D. Bechor, E. Martz, N. Ben-Tal, 2003. ConSurf: Identification of Functional Regions in Proteins by Surface-Mapping of Phylogenetic Information. Bioinformatics 19, 163-164.

Goodwill, K.E., C. Sabatier, C. Marks, R. Raag, P.F. Fitzpatrick, R.C. Stevens, 1997. Crystal structure of tyrosine hydroxylase at 2.3 Å and its implications for inherited neurodegenerative diseases. Nat. Struct. Biol. 4, 578-585.

Grant, G.A., 2006. The ACT domain: a small molecule binding domain and its role as a common regulatory element. Jurnal of biological chemistry 281, 33825-33829.

Gregersen, N., P. Bross, B.S. Andresen, C.B. Pedersen, T.J. Corydon, L. Bolund, 2001. The role of chaperoneassisted folding and quality control in inborn errors of metabolism: protein folding disorders. J. Inherit. Metab. Dis. 24, 189-212.

Halgren, T., 2000. Flexible docking of ligands to receptor sites with glide. Abstracts of Papers of the American Chemical Society 220, U168-U168.

Hess, B., H. Bekker, H. Berendsen, J. Fraaije, 1997. LINCS: a linear constraint solver for molecular simulations. Journal of computational chemistry 18, 1463-1472.

Holm, L., C. Sander, 1996. Mapping the protein universe. Science 273, 595-602.

Hooft, R., C. Sander, G. Vriend, 1996. Verification of protein structures: sidechain planarity. J. Appl. Crystallogr. 29, 714-16.

Horne, J., I. Jennings, T. Trazel, P. Gooley, B. Kobe, 2002. Structural characterization of the N-terminal autoregulatory sequence of phenylalanine hydroxylase. Protein Science 11, 2041-2047.

Hub, J.S., B.L. de Groot, 2009. Detection of functional modes in protein dynamics. Plos Comp. Biol. 5, 1-13.

Hufton, S., I. Jennings, R. Cotton, 1995. Structure and function of the aromatic amino acid hydroxylases. Biochem J 311, 353-366.

Humphrey, W., A. Dalke, K. Schulten, 1996. VMD -- Visual Molecular Dynamics. Journal of Molecular Graphics 14, 33-38.

Iwaki, M., R.S. Phillips, S. Kaufman, 1986. Proteolytic modification of the amino-terminal and carboxylterminal regions of rat hepatic phenylalanine hydroxylase. J. Biol. Chem. 261, 2051-2056.

Jennings, I.G., T. Teh, B. Kobe, 2001. Essential role of the N-terminal autoregulatory sequence in the regulation of phenylalanine hydroxylase. FEBS Lett. 488, 196-200.

Jorgensen, W., J. Tirado-Rives, 1988. The Opls Potential Functions for Proteins - Energy Minimizations for Crystals of Cyclic-Peptides and Crambin. J. Am. Chem. Soc. 110, 1657-1666.

Kabsch, W., C. Sander, 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22, 2577-2637.

Kalhan, S.C., D.M. Bier, 2008. Protein and amino acid metabolism in the human newborn. Annu. Rev. Nutr. 28, 389-410.

Kappock, T., J. Caradonna, 1996. Pterin-dependent amino acid hydroxylase. Chem Rev 96, 2659-2756.

Kappock, T.J., P.C. Harkins, S. Friedenberg, J.P. Caradonna, 1995. Spectroscopic and kinetic properties of unphosphorylated rat hepatic phenylalanine hydroxylase expressed in Escherichia coli. Comparison of resting and activated states. J. Biol. Chem. 270, 30532-30544.

Kaufman, S., 1993. The phenylalanine hydroxylating system. Adv Enzymol 70, 77-264.

Knappskog, P.M., T. Flatmark, J.M. Aarden, J. Haavik, A. Martinez, 1996a. Structure/function relationships in human phenylalanine hydroxylase. Effect of terminal deletions on the oligomerization, activation and cooperativity of substrate binding to the enzyme. Eur J Biochem 242, 813-21.

Knappskog, P.M., H.G. Eiken, A. Martinez, O. Bruland, J. Apold, T. Flatmark, 1996b. PKU mutation (D143G) associated with an apparent high residual enzyme activity: expression of a kinetic variant form of phenylalanine hydroxylase in three different systems. Hum Mutat 8, 236-46.

Kobe, B., I.G. Jennings, C.M. House, B.J. Michell, K.E. Goodwill, B.D. Santarsiero, R.C. Stevens, R.G. Cotton, B.E. Kemp, 1999. Structural basis of autoregulation of phenylalanine hydroxylase. Nat Struct Biol 6, 442-8.

Kramer, B., G. Metz, M. Rarey, T. Lengauer, 1999. Ligand docking and screening with FlexX. Medicinal Chemistry Research 9, 463-478.

Landau, M., I. Mayrose, Y. Rosenberg, F. Glaser, E. Martz, T. Pupko, N. Ben-Tal, 2005. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. Nucl. Acids Res. 33, W299-W302.

Lang, P.T., D. Moustakas, S. Brozell, N. Carrascal, S. Mukherjee, S. Pegg, K.D.S. Raha, R. Rizzo, D. Case, B. Shoichet, I.D. Kuntz, 2007. DOCK, 6.1 ed. University of California, San Francisco.

Laskowski, R., M. McArthur, D. Moss, J. Thornton, 1993. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystalogr. 26, 283-91.

Laskowski, R., J. Rullmann, M. MacArthur, R. Kaptein, J. Thornton, 1996. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR 8, 477-86.

Laskowski, R.A., E.G. Hutchinson, A.D. Michie, A.C. Wallace, M.L. Jones, J.M. Thornton, 1997. PDBsum: A Web-based database of summaries and analyses of all PDB structures. Trends Biochem. Sci. 22, 488-490.

Leandro, J., N. Simonsen, J. Saraste, P. Leandro, T. Flatmark, 2010. Phenylketonuria as a protein misfolding disease: The mutation pG46S in phenylalanine hydroxylase promotes self-association and fibril formation. Biochim. Biophys. Acta. 1812, 106-120.

Leandro, J., N. Simonsen, J. Saraste, P. Leandro, T. Flatmark, 2011. Phenylketonuria as a protein misfolding disease: The mutation pG46S in phenylalanine hydroxylase promotes self-association and fibril formation. Biochim Biophys Acta 1812, 106-20.

Lehtonen, J., D. Still, V. Rantanen, J. Ekholm, D. Björklund, Z. Iftikhar, M. Huhtala, S. Repo, A. Jussila, J. Jaakkola, O. Pentikäinen, T. Nyrönen, T. Salminen, M. Gyllenberg, M. Johnson, 2004. BODIL: a molecular modeling environment for structure-function analysis and drug design. J Comput Aided Mol Des. 18, 401-19.

Lesk, A.M., C. Chothia, 1980. How different amino acid sequences determine similar protein structures: The structure and evolutionary dynamics of the globins. J. Mol. Biol. 136, 225-70.

Li, J., L.J. Dangott, P.F. Fitzpatrick, 2010. Regulation of phenylalanine hydroxylase: conformational changes upon phenylalanine binding detected by hydrogen/deuterium exchange and mass spectrometry. Biochemistry 49, 3327-3335.

Li, J., U. Ilangovan, S.C. Daubner, A.P. Hinck, P.F. Fitzpatrick, 2011. Direct evidence for a phenylalanine site in the regulatory domain of phenylalanine hydroxylase. Archives of Biochemistry and Biophysics 505, 250-255.

Liberles, J.S., A. Martinez, 2009. Searching distant homologs of the regulatory ACT domain in phenylalanine hydroxylase. Amino Acid 36, 235-249.

Liberles, J.S., M. Thorolfsson, A. Martinez, 2005. Allosteric machanism in ACT domain containing enzymes involved in amino acid metabolism. Amino Acid 28, 1-12.

Lounnas, V., B.M. Pettitt, 1994. Distribution function implied dynamics versus residence times and correlations: solvation shells of myoglobin. Proteins 18, 148-160.

Lounnas, V., B.M. Pettitt, G.N. Phillips Jr., 1994. A global model of the protein-solvent interface. Biophys. J. 66, 601-614.

Martì-Renom, M.A., A.C. Stuart, A. Fiser, R. Sanchez, F. Melo, A. Sali, 2000. Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. 29, 291-325.

Martin, A.C.R., J.C. Cheetham, A.R. Rees, 1989. Modeling antibody hypervariable loops: a combined algorithm. Proc. Natl. Acad. Sci. USA 86, 9268-72.

Martínez, A., A.C. Calvo, K. Teigen, A.L. Pey, 2008. Rescuing proteins of low kinetic stability by chaperones and natural ligands phenylketonuria, a case study. Prog. Mol. Biol. Transl. Sci. 83, 89-134.

Martínez, A., P.M. Knappskog, S. Olafsdottir, A.P. Døskeland, H.G. Eiken, R.M. Svebak, M. Bozzini, J. Apold, T. Flatmark, 1995. Expression of recombinant human phenylalanine hydroxylase as fusion protein in Escherichia coli circumvents proteolytic degradation by host cell proteases. Isolation and characterization of the wild-type enzyme. Biochem. J. 306, 589-597.

Mayrose, I., D. Graur, N. Ben-Tal, T. Pupko, 2004. Mol. Biol. Evol. 21, 1781-1791.

Miranda, F.F., M. Thórólfsson, K. Teigen, J.M. Sanchez-Ruiz, A. Martínez, 2004. Structural and stability effects of phosphorylation: Localized structural changes in phenylalanine hydroxylase. Protein Sci. 13, 1219-1226.

Miranda, F.F., K. Teigen, M. Thorolfsson, R.M. Svebak, P.M. Knappskog, T. Flatmark, A. Martinez, 2002. Phosphorylation and mutations of Ser(16) in human phenylalanine hydroxylase. Kinetic and structural effects. J Biol Chem 277, 40937-43.

Mitnaul, L.J., R. Shiman, 1995. Coordinate regulation of tetrahydrobiopterin turnover and phenylalanine hydroxylase activity in rat liver cells. Proc. Natl. Acad. Sci. U. S. A. 92, 885-889.

Miyamoto, S., P.A. Kollman, 1992. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. J Comput Chem 13, 952-962.

Morris, G.M., D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, 1998. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. Journal of Computational Chemistry 19, 1639-1662.

Muntau, A.C., W. Röschinger, M. Habich, H. Demmelmair, B. Hoffmann, C.P. Sommerhoff, A.A. Roscher, 2002. Tetrahydrobiopterin as an alternative treatment for mild phenylketonuria. N. Engl. J. Med. 347, 2122-2132.

Notredame, C., D.G. Higgins, J. Heringa, 2000. T-Coffee: A novel method for multiple sequence alignments. J. Mol. Biol. 302, 205-217.

Oldfield, T., 1992. Squid: a program for the analysis and display of data from crystallography and molecular dynamics. J. Mol. Graphics 10, 247-52.

Parniak, M.A., S. Kaufman, 1981. Rat liver phenylalanine hydroxylase. Activation by sulfhydryl modification. J. Biol. Chem. 256, 6876-6882.

Pearson, W.R., 1998. Empirical statistical estimates for sequence similarity searches. J. Mol. Biol. 276, 71-84.

Pey, A., F. Stricher, L. Serrano, A. Martinez, 2007. Predicted effects of missense mutations on native-state stability account for phenotypic outcome in phenylketonuria, a paradigm of misfolding diseases. Am J Hum Genet 81, 1006-24.

Pey, A.L., M. Thórólfsson, K. Teigen, M. Ugarte, A. Martínez, 2004a. Thermodynamic characterization of the binding of tetrahydropterins to phenylalanine hydroxylase. J. Am. Chem. Soc. 126, 13670-13678.

Pey, A.L., B. Pérez, L.R. Desviat, M.A. Martínez, C. Aguado, H. Erlandsen, A. Gámez, R.C. Stevens, M. Thórólfsson, M. Ugarte, A. Martínez, 2004b. Mechanisms underlying responsiveness to tetrahydrobiopterin in mild phenylketonuria mutations. Hum. Mutat. 24, 388-399.

Pupko, T., R.E. Bell, I. Mayrose, F. Glaser, N. Ben-Tal, 2002. Bioinformatics 18, S71-77.

Sali, A., T.L. Blundell, 1993. Comparative protein modeling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779-815.

Sanchez, R., A. Sali, 1997. Advances in comparative protein-structure modeling. Curr. Opin. Struct. Biol. 7, 206-14.

Sanchez, R., A. Sali, 1998. Large-scale protein structure modeling of the Saccharomyces cerevisiae genome. Proc. Natl. Acad. Sci. USA : 95, 13597-602.

Schrödinger, L., 2009. The PyMOL Molecular Graphics System, Version 1.2r0.

Scriver, C.R., 2007. The PAH gene, phenylketonuria, and a paradigm shift. Hum. Mutat. 28, 831-845.

Scriver, C.R., P.J. Waters, 1999. Monogenic traits are not simple: lessons from phenylketonuria. Trends Genet. 15, 267-272.

Seeliger, D., J. Haas, B.L. de Groot, 2007. Geometry-based sampling of conformational transitions in proteins. Structure 15, 1482-92.

Shiman, R., D.W. Gray, 1980. Substrate activation of phenylalanine hydroxylase. A kinetic characterization. J Biol Chem 255, 4793-4800.

Shiman, R., D.W. Gray, A. Pater, 1979. A simple purification of phenylalanine hydroxylase by substrate-induced hydrophobic chromatography. Biol Chem 254, 11300-11306.

Shiman, R., S.H. Jones, D.W. Gray, 1990. Mechanism of phenylalanine regulation of phenylalanine hydroxylase. J Biol Chem 265, 11633-42.

Solstad, T., R.N. Carvalho, O.A. Andersen, D. Waidelich, T. Flatmark, 2003. Deamidation of labile asparagine residues in the autoregulatory sequence of human phenylalanine hydroxylase. Eur J Biochem 270, 929-38.

Stokka, A.J., T. Flatmark, 2003. Substrate-induced conformational transition in human phenylalanine hydroxylase as studied by surface plasmon resonance analyses: the effect of terminal deletions, substrate analogues and phosphorylation. Biochem J 369, 509-18.

Stokka, A.J., R.N. Carvalho, J.F. Barroso, T. Flatmark, 2004. Probing the role of crystallographically defined/predicted hinge-bending regions in the substrate-induced global conformational transition and catalytic activation of human phenylalanine hydroxylase by single-site mutagenesis. J Biol Chem 279, 26571-80.

Tan, K., H. Li, R. Zhang, M. Gu, S.T. Clancy, A. Joachimiak, 2008. Structures of open (R) and close (T) states of prephenate dehydratase (PDT)-Implication of allosteric regulation by L-phenylalanine. Jurnal of Structural Biology 162, 94-107.

Teigen, K., A. Martinez, 2003. Probing cofactor specificity in phenylalanine hydroxylase by molecular dynamics simulations. J Biomol Struct Dyn 20, 733-40.

Teigen, K., N.Å. Froystein, A. Martínez, 1999. The structural basis of the recognition of phenylalanine and pterin cofactors by phenylalanine hydroxylase: implications for the catalytic mechanism. J. Mol. Biol. 294, 807-823.

Teigen, K., J.A. McKinney, J. Haavik, A. Martinez, 2007. Selectivity and affinity determinants for ligand binding to the aromatic amino acid hydroxylases. Curr Med Chem 14, 455-67.

Teigen, K., K.K. Dao, J.A. McKinney, A.C. Gorren, B. Mayer, N.Å. Frøystein, J. Haavik, A. Martínez, 2004. Tetrahydrobiopterin binding to aromatic amino acid hydroxylases. Ligand recognition and specificity. J. Med. Chem. 47, 5962-5971.

Thöny, B., N. Blau, 2006. Mutations in the BH4-metabolizing genes GTP cyclohydrolase I, 6-pyruvoyltetrahydropterin synthase, sepiapterin reductase, carbinolamine-4a-dehydratase, and dihydropteridine reductase. Hum. Mutat. 27, 870-878.

Thorolfsson, M., K. Teigen, A. Martinez, 2003. Activation of phenylalanine hydroxylase: effect of substitutions at Arg68 and Cys237. Biochemistry 42, 3419-28.

Thorolfsson, M., B. Ibarra-Molero, P. Fojan, S.B. Petersen, J.M. Sanchez-Ruiz, A. Martinez, 2002. L-phenylalanine binding and domain organization in human phenylalanine hydroxylase: a differential scanning calorimetry study. Biochemistry 41, 7573-85.

van Aalten, D.M., A. Amadei, A.B. Linssen, V.G. Eijsink, G. Vriend, H.J. Berendsen, 1995. The essential dynamics of thermolysin: confirmation of the hinge-bending motion and comparison of simulations in vacuum and water. Proteins 22, 45-54.

Van Der Spoel, D., E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J.C. Berendsen, 2005. GROMACS: fast, flexible, and free. Journal of computational chemistry 26, 1701-1718.

van Gunsteren, W.F., S.R. Billeter, A.A. Eising, P.H. Hünenberger, P. Krüger, A.E. Mark, W.R.P. Scott, I.G. Tironi, 1996. Biomolecular Simulation: The GROMOS96 manual and userguide. Hochschuleverlag AG an der ETH Zürich, Zürich, Switzerland.

Verdonk, M.L., J.C. Cole, M.J. Hartshorn, C.W. Murray, R.D. Taylor, 2003. Improved protein-ligand docking using GOLD. Proteins-Structure Function and Genetics 52, 609-623.

Vriend, G., 1990. WHAT IF: A molecular modeling and drug design program. J. Mol. Graph. 8, 52-56.

Walther, D.J., M. Bader, 2003. A unique central tryptophan hydroxylase isoform. Biochem. Pharmacol. 66, 1673-1680.

Waters, P.J., 2003. How PAH gene mutations cause hyper-phenylalaninemia and why mechanism matters: insights from in vitro expression. Hum Mutat 21, 357-69.

Waters, P.J., 2006. Molecular bases of phenylketonuria: insights from functional studies in vitro. In: Blau N (ed) PKU and BH4: advances in phenylketonuria and tetrahydrobiopterin SPS Verlagsgesellschaft, Heilbronn, pp 277-310.

Waters, P.J., M.A. Parniak, B.R. Akerman, C.R. Scriver, 2000. Characterization of phenylketonuria missense substitutions, distant from the phenylalanine hydroxylase active site, illustrates a paradigm for mechanism and potential modulation of phenotype. Mol Genet Metab 69, 101-10.

Woo, S.L., A.G. DiLella, J. Marvit, F.D. Ledley, 1987. Molecular basis of phenylketonuria and recombinant DNA strategies for its therapy. Enzyme 38, 207-13.

Zhang, Z.T., 1997. Relations of the numbers of protein sequences, families and folds. Protein Eng. 10, 757-61.

Paper I

Biochimica et Biophysica Acta 1812 (2011) 1435–1445

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbadis

Natural phenylalanine hydroxylase variants that confer a mild phenotype affect the enzyme's conformational stability and oligomerization equilibrium

Monica Cerreto [a,1], Paola Cavaliere [a,b,1], Carla Carluccio [a,c], Felice Amato [a], Adriana Zagari [a,b], Aurora Daniele [a,d,e,*], Francesco Salvatore [a,c,e,**]

[a] CEINGE–Biotecnologie Avanzate Scarl, Naples, Italy
[b] Dipartimento di Scienze Biologiche, Università di Napoli "Federico II", Naples, Italy
[c] Dipartimento di Biochimica e Biotecnologie Mediche, Università di Napoli "Federico II", Naples, Italy
[d] Dipartimento di Scienze Ambientali, Seconda Università di Napoli, Caserta, Italy
[e] IRCCS – Fondazione SDN, Naples, Italy

ARTICLE INFO

Article history:
Received 8 March 2011
Received in revised form 19 July 2011
Accepted 20 July 2011
Available online 27 July 2011

Keywords:
BH4 responsiveness
Number and brightness
PAH conformational stability
PAH oligomerization equilibrium
Hyperphenylalaninemia
Phenylalanine hydroxylase

ABSTRACT

Hyperphenylalaninemias are genetic diseases prevalently caused by mutations in the phenylalanine hydroxylase (*PAH*) gene. The wild-type PAH enzyme is a homotetramer regulated by its substrate, cofactor and phosphorylation. We reproduced a full-length wild-type protein and seven natural full-length PAH variants, p.I65M, p.N223Y, p.R297L, p.F382L, p.K398N, p.A403V, and p.Q419R, and analyzed their biochemical and biophysical behavior. All mutants exhibited reduced enzymatic activity, namely from 38% to 69% of wild-type activity. Biophysical characterization was performed by size-exclusion chromatography, light scattering and circular dichroism. In the purified wild-type PAH, we identified the monomer in equilibrium with the dimer and tetramer. In most mutants, the equilibrium shifted toward the dimer and most tended to form aggregates. All PAH variants displayed different biophysical behaviors due to loss of secondary structure and thermal destabilization. Specifically, p.F382L was highly unstable at physiological temperature. Moreover, using confocal microscopy with the number and brightness technique, we studied the effect of BH4 addition directly in living human cells expressing wild-type PAH or p.A403V, a mild mutant associated with BH4 responsiveness in vivo. Our results demonstrate that BH4 addition promotes re-establishment of the oligomerization equilibrium, thus indicating that the dimer-to-tetramer shift in pA403V plays a key role in BH4 responsiveness. In conclusion, we show that the oligomerization process and conformational stability are altered by mutations that could affect the physiological behavior of the enzyme. This endorses the hypothesis that oligomerization and folding defects of PAH variants are the most common causes of HPAs, particularly as regards mild human phenotypes.

1. Introduction

Phenylketonuria (PKU, Online Mendelian Inheritance in Man database: 261600) and its hyperphenylalaninemia (HPA) variants represent the most common inherited disorder of amino acid metabolism transmitted by an autosomal recessive mode [1]. The primary cause of HPA is a dysfunction of phenylalanine hydroxylase (PAH: EC 1.14.16.1). In the liver, this enzyme metabolizes L-Phenylalanine (L-Phe) to L-Tyrosine (L-Tyr) using (6R)-L-erythro-5,6,7,8-tetrahydrobiopterin (BH4) as coenzyme. More rarely, forms of HPAs can also be caused by a lack of the BH4 cofactor due to defective cofactor biosynthesis and regeneration [2]. Untreated PKU patients present an abnormal phenotype, with growth failure, microcephaly, seizures and permanent neurologic damage because their body fluids contain elevated levels of L-Phe and its neurotoxic metabolites (phenylpyruvate, phenylacetate and phenylactate) [1,3].

The major manifestations of the disease, which start at birth, are prevented by severe dietary restriction of L-Phe [4] and, in selected cases, by BH4 supplementation, which reduces L-Phe levels and increases L-Phe tolerance in HPA patients [5]. Several new therapeutic strategies based on the biochemistry and pathogenetic features of PKU are currently being investigated [6,7].

Abbreviations: BH4, (6R)-L-erythro-5,6,7,8-tetrahydrobiopterin; CD, circular dichroism; Δ13-PAH, PAH missing residues 1–13; DSC, differential scanning calorimetry; Ek, enterokinase; HPA, hyperphenylalaninemia; IPTG, isopropylthio-β-D-galactoside; L-Phe, phenylalanine; LS, light scattering; MALS, multiangle light scattering; MBP, maltose-binding protein; N&B, number and brightness; PKU, phenylketonuria; SEC, size-exclusion chromatography

* Correspondence to: A. Daniele, CEINGE–Biotecnologie Avanzate Scarl, Via Gaetano Salvatore 486, 80145 Naples, Italy.
** Correspondence to: F. Salvatore, CEINGE–Biotecnologie Avanzate Scarl and Dipartimento di Biochimica e Biotecnologie Mediche, Università di Napoli "Federico II", Via Sergio Pansini 5, Ed.19, 80131 Naples, Italy. Tel.: +39 0817463133; fax: +39 0817463650.
E-mail address: salvator@unina.it (F. Salvatore).
[1] These authors contributed equally to this paper.

0925-4439/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbadis.2011.07.012

The human *PAH* gene, mapped on 12q23.2, consists of 13 exons encompassing 171 kb. The full-length *PAH* cDNA encodes a protein of about 52 kDa (452 amino acids) that in the mature form is assembled as a homotetramer. Each subunit consists of three functional domains: a flexible N-terminal regulatory domain (residues 1–142); a catalytic domain (residues 143–410) that includes binding sites for iron, substrate and cofactor; and a C-terminal oligomerization domain (residues 411–452) with dimerization (residues 411–426) and tetramerization motifs (residues 427–452) [8] (Fig. 1). The crystal structure of PAH has revealed that the tetrameric oligomers are dimers of dimers in which the interaction between the two dimers is mediated by the C-terminal 'arm' [8].

The tetrameric and dimeric forms of PAH are in equilibrium [9,10] and have different catalytic properties because the tetramer, but not the dimer, demonstrates a positive kinetic cooperativity with respect to L-Phe [11]. In particular, substrate activation results in conformational changes involving the tertiary as well as the quaternary structure [12] and drives the tetramer–dimer equilibrium toward the tetrameric form [13]. Further regulation mechanisms require binding of BH4 [12] and phosphorylation [14].

The clinical and metabolic features associated to HPA are complex [15] and various genotype–phenotype correlation studies have been performed [16–18]. In addition, at molecular level, about 600 mutations have been identified, most of which are point mutations scattered throughout the whole *PAH* gene (http://www.pahdb.mcgill.ca), although frequencies differ among populations and geographic areas [16–18].

Recombinant variant proteins containing prevalently missense mutations have been analyzed and characterized in vitro; each mutation exerts a distinct effect on the behavior of the PAH protein, which in many cases allows the prediction of the biochemical phenotype [19–21]. The effects of a specific mutation on protein functions have been studied with different approaches, i.e., enzyme assay, in vivo isotopic studies and in vitro expression. Moreover, the crystal structures of human and rat PAH [8,22] provided the structural basis of HPA [23]. It is now established that, in most HPA cases, the loss of PAH function is due to decreased stability [21,24], increased susceptibility toward aggregation and degradation of PAH mutant proteins [19,24,25], thermodynamic instability [26] and/or folding efficiency [27]. About 75% of PAH mutations, characterized by high residual activity, have been found to be associated with BH4 responsiveness, both in vitro [28–30] and in vivo [4,5]. It has recently been speculated that pharmacological doses of BH4 may augment the conformational stability and the amount of functional PAH [31].

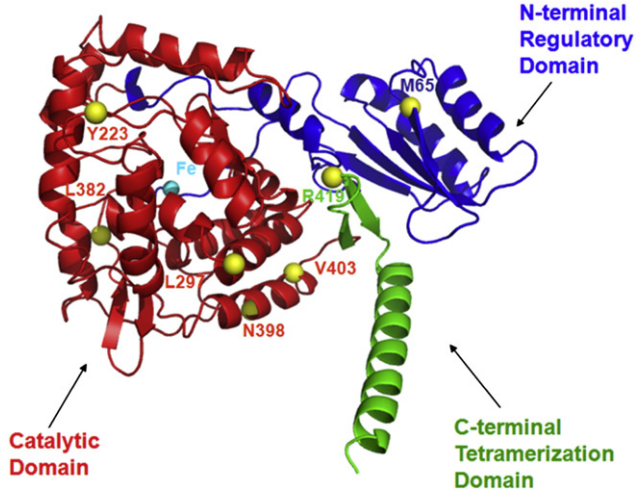

Fig. 1. Full-length composite model of the PAH monomer structure and distribution of the HPA-related missense mutations studied in this paper. The N-terminal regulatory domain is shown in blue, the catalytic domain in red and the tetramerization domain in green. The Fe ion is shown as a cyan sphere at the center of the catalytic domain; the mutation sites are marked as yellow balls.

We previously carried out a molecular analysis of the *PAH* gene in HPA patients from Southern Italy, and identified and characterized several novel mutations [32–34]. In the present study, we analyzed the wild-type PAH protein and some PAH natural mutants to obtain further information about the molecular basis of HPAs and the mechanism of BH4 responsiveness. To this purpose, we reproduced, by in vitro mutagenesis, the wild-type and seven natural variants of PAH: p.I65M, p.N223Y, p.R297L, p.F382L, p.K398N, p.A403V, and p.Q419R, and characterized the biophysical–biochemical properties of the above-mentioned purified full-length proteins in order to understand the impact of each mutation on the activity, oligomeric structure and stability of PAH. In addition, we investigated the mechanism underlying BH4 responsiveness using p.A403V, which is one of the most frequent mutations in our geographic area [32], and is associated with a BH4-responsive mild phenotype [35].

2. Material and methods

2.1. Construction of PAH expression plasmid and site-directed mutagenesis

PAH wild-type and mutant constructs were obtained by modifying the pMAL Xa PAH plasmid, kindly provided by Drs CR. Scriver and P. Waters (McGill University-Montreal Children's Hospital Research Institute, Montreal, Canada), to digest the fusion protein MBP-PAH with enterokinase (Ek) [9]. The Ek site was inserted downstream the sequence encoding the factor Xa site, by site-directed mutagenesis using the primers 5′gatgacgatgacaagtctactgcggtcctgg3′ and 5′ gccctaactcccttccctactgctactgttc3′ and the Quick Site-directed mutagenesis kit (Stratagene, CA, USA). GFP was amplified by PCR from pEGFP-N1 (Clonetech, CA, USA) using the primers: 5′acattaggtaccagccatggtgagcaagggcgag3′ and 3′ gtagatgaattctccaccacccttgtacagctcg 5′ and cloned into pcDNA3 (Invitrogen, Carlsbad, CA) using Kpn1 and EcoR1 sites. PAH cDNA was amplified from pMAL Xa Ek PAH by using the primers: 5′catcattgaattctccactgcggtcctgg3′ and 3′ cgtaatgcggccgcttactttattttctggag 5′ and cloned downstream GFP, using EcoR1 and Not1 sites. Sequence analysis confirmed the correct frame.

Mutations were introduced into pMAL Xa Ek PAH and pCDNA3-GFP-PAH using mutagenic primers and the Quick Site-directed mutagenesis kit (Table 1). The modified plasmid and the mutant clones were sequenced to verify the introduction of each single mutation. Commercial MBP was obtained from New England Biolabs (Ipswich, MA, USA).

2.2. Expression, purification and N-terminal sequencing of the PAH protein

pMAL Xa Ek PAH expression plasmids were transformed into *Escherichia coli* BL21 cells and the colonies were selected using Luria-Broth plates with ampicillin (0.1 mg mL^{-1}). Bacteria were grown to 2×10^8 cells/mL (A_{600nm}~0.5) and overexpression of wild-type and variant MBP-PAH fusion proteins was induced with 1 mM isopropylthio-β-D-galactoside (IPTG) for 16 h at 37 °C. Cells were harvested by centrifugation and treated according to the instruction manual of the pMAL protein fusion and purification system (New England Biolabs). The fusion protein was digested overnight at room temperature with Ek (New England Biolabs) to obtain wild-type and mutant forms of PAH without the MBP tag, using 0.5 ng of enzyme for each 50 μg fusion protein. The mixture was then applied to a hydroxyapatite column (1 cm × 10 cm, Bio-Rad Laboratories, CA, USA) and subjected to several washes with 20 mM sodium phosphate, 200 mM NaCl (pH 7.2). Proteins were then eluted with 0.5 M Na phosphate, pH 7.2 and subsequently MBP was isolated from the

Table 1
The seven mutant forms of PAH reproduced in vitro as recombinant products and analyzed. The primer pairs used to produce the pMAL Xa Ek PAH plasmid.

Mutant forms of PAH	Mutagenic primers (forward)	Mutagenic primers (reverse)
Ek site	5′GATGACGATGACAAGTCTACTGCGGTCCTGG3′	5′GCCCTAACTCCCTTCCCTACTGCTACTGTTC3′
p.I65M	5′ACCTGACCCACATGGAATCTAGACCTTCTC3′	5′GAGAAGGTCTAGATTCCATGTGGGTCAGGT3′
p.N223Y	5′CATGAAGATTACATTCCCCAGCTGGAAGAC3′	5′GTCTTCCAGCTGGGGAATGTAATCTTCATG3′
p.R297L	5′TGTTTTCAGATCTCAGCTTTGCCCAGTTTT3′	5′AAAACTGGGCAAAGCTGAGATCTGAAAACA3′
p.F382L	5′ACTGTCACGGAGTTGCAGCCCCTGTATTAC3′	5′GTAATACAGGGGCTGCAACTCCGTGACAGT3′
p.K398N	5′GCCAAGGAGAACGTAAGGAACTTTGCTGCC3′	5′GGCAGCAAAGTTCCTTACGTTCTCCTTGGC3′
p.A403V	5′GAAAGTAAGGAACTTTGTTGCCACAA3′	5′CCGAGGTATTGTGGCAACAAAGTTCC3′
p.Q419R	5′ATACACCCGAAGGATTGAGG3′	5′CCTCAATCCTTCGGGTGTAT3′

cleavage mixture by re-binding to amylose resin. PAH was finally purified by size-exclusion chromatography with a HiLoad 16/60 Superdex 200 column (GE Healthcare, UK). Protein concentrations were determined spectrophotometrically with the use of the absorption coefficient A280 or the dye-binding Bradford assay (Bio-Rad Laboratories). The N-terminal amino acid sequence of PAH was determined by Edman degradation on an automated Procise H49 sequencer (Applied Biosystems, MA, USA).

It has been reported [36] that, during the various steps of recombinant protein production, some labile Asn residues are deamidated: however, any possible deamidation of these residues, which would in the activity assay produce substrate inhibition, would most likely be the same in the presence of wild-type and of the variant forms of PAH we have employed.

2.3. Concentration and storage of proteins

The concentration of proteins was determined both by the micro BCA protein assay and spectrophotometrically. The final protein concentration was measured based on optical density at 280 nm using, for the extinction coefficient, a value of 49,780, 66,350 and 115,630 M^{-1} cm^{-1} for PAH, MBP and the fused wild-type MBP-PAH respectively, deduced from the amino acid composition of the protein. We used a SpeedVac concentrator (Thermo Scientific, Rockford, IL, USA) to concentrate the protein solutions because ultra-filtration led to considerable sample loss. All protein samples, if not utilized within few days, were stored at −20 °C in their elution buffer (see below).

2.4. PAH activity and thermal inactivation assay

Enzymatic activity analysis was carried out on a whole equilibrium mixture of oligomeric species of purified full-length wild-type and mutant PAH (see Supplementary Table S1). The activity was assayed in the presence of 0.25 μCi L-[^{14}C] Phe (Amersham, Buckinghamshire, UK; 460 μCi/mmol); 0.25 mM cold L-Phe, 1.3 units of beef liver catalase; 0.8 mM BH4; 250 mM Tris HCl pH 7.8, 1 μg of protein; final volume: 100 μL. The reaction was conducted at 25 °C with an incubation time of 1 h [32–34]. After centrifugation at room temperature and maximum speed, an aliquot of each supernatant was applied on a thin-layer chromatography system and the amount of [^{14}C] radio-labeled L-Phe converted to [^{14}C] radio-labeled L-Tyr was measured. The mean PAH activities were calculated from three sets of experiments. The residual activities of mutant PAH enzymes were expressed as a percentage of wild-type enzyme activity. The same assay was used for the thermal inactivation analysis of the wild-type protein and of the p.F382L mutant. Thermal inactivation of both proteins was determined by measuring the decay of tyrosine production as a function of temperature in the 32–58 °C range.

2.5. Electrophoresis and immunoblotting

SDS/PAGE was performed at 100 V (2 h) in a 10% (w/v) polyacrylamide gel. The purification of all proteins was verified after staining with Colloidal Coomassie. The absence of MBP contamination in all preparations was determined by western blot analyses. Immunoblotting was performed using affinity-purified rabbit anti-hPAH [32–34] and mouse anti-MBP (New England Biolabs) as primary antibodies. The enhanced chemiluminescence system from Amersham (GE Healthcare) was used for immunodetection.

2.6. Size-exclusion chromatography with multiangle light scattering

Size-exclusion chromatography (SEC) was performed using Akta FPLC chromatography equipment (GE Healthcare) on a Superdex 200 10/300 GL column. The column was previously calibrated for molecular mass using a Gel Filtration High Molecular Weight calibration kit (GE Healthcare) constituted by ovalbumin (44 kDa), conalbumin (75 kDa) and aldolase (158 kDa). Before each run, the column was equilibrated with at least two column volumes of elution buffer and each run was performed at both 4 °C and room temperature (20–23 °C). The elution buffer was Tris–HCl 20 mM pH 7.4, NaCl 100 mM (buffer A) and, in a few cases, Na-phosphate 20 mM pH 7.0, KF 150 mM (buffer B).

The eluent flow was 0.5 mL min^{-1}. Only freshly prepared or recently defrosted proteins were loaded on the column because storage at 4 °C for more than one week led to sample aging. A total of 0.1 ÷ 0.2 mg of proteins were loaded. At room temperature, the elution profile was monitored at three UV absorption wavelengths (280 nm, 215 nm and 258 nm). At 4 °C, UV absorption was measured only at 280 nm. The chromatograms were deconvoluted by Unicorn software, supplied with the FPLC device, to calculate the percentage of single species. All chromatograms were normalized in the (0.0 ÷ 1.2) range.

At room temperature, size-exclusion chromatography with multiangle light scattering (MALS) experiments was carried out with the AKTA system coupled to a three-angle (45°, 90° and 135°) Wyatt Minidown EOS light scattering instrument linked to a Wyatt Optilab refractometer (Wyatt Technology Corp., Santa Barbara, CA). These experiments served to determine the sizes of oligomers of wild-type and mutant PAH. Molecular masses, polydispersity and root mean square radius calculations were made by ASTRA software, which is supplied with the light scattering device, using a dn/dc value of 0.185 mL g^{-1}.

2.7. Circular dichroism

Circular dichroism (CD) analysis was performed on commercial MBP, on the fused wild-type MBP-PAH tetrameric sample and on the most abundant species of cleaved PAH proteins after SEC purification. In the case of the cleaved wild-type PAH, CD experiments were performed with both the tetrameric form and the whole equilibrium mixture. A 2-day dialysis at 4 °C against Tris–HCl 20 mM pH 7.4 (buffer C) was conducted to eliminate NaCl before CD experiments. All CD spectra were recorded with a Jasco J-810 spectropolarimeter equipped with a Peltier temperature control system (model PTC-423-S). The spectropolarimeter was calibrated with an aqueous solution of D-10-(+)-camphorsulfonic acid at 290 nm. The molar ellipticity per mean residue, $[\theta]$ in deg cm^2 $dmol^{-1}$, was calculated from the equation $[\theta] = [\theta]_{obs}$ mrw $(10\ l\ C)^{-1}$, where $[\theta]_{obs}$ is the ellipticity measured in degrees, mrw is the mean residue molecular weight (111.5 Da), C is the protein concentration in

g mL^{-1}, and l is the optical path length of the cell in cm. Far-UV (ultraviolet) measurements (190–260 nm) were carried out at 20 °C using a 0.2 cm optical path length cell and a protein concentration in the range 0.10–0.15 mg mL^{-1}. CD spectra, recorded with a time constant of 4 s, a 1-nm bandwidth, and a scan rate of 20 nm min^{-1}, were signal averaged over at least three scans. The baseline was corrected by subtracting the buffer spectrum. Thermal denaturation curves were recorded over the 20–80 °C temperature range and by monitoring the CD signal at 222 nm. After preliminary trials within the scan rate interval of 0.5 ÷ 2.0 °C min^{-1}, all curves were recorded using a 0.2-cm path length cell and a scan rate of 1.0 °C min^{-1}. Irreversibility of the denaturation was verified by recording, after cooling to 20 °C, the spectra of samples subjected to a first temperature scan. All denaturation curves were normalized in an interval of 0.0 ÷ 1.0. The T_m values were determined from the first derivatives of the denaturation curves after noise reduction by using the standard analysis program provided with the instrument.

2.8. Cell culture

HeLa cells were grown in D-MEM high glucose medium (Invitrogen) containing 10% heat inactivated fetal bovine serum (Invitrogen), 1% penicillin/streptomycin, at 37 °C in 5% CO_2. Transfections of both wild-type PAH and the p.A403V mutant were carried out using the ProFection Mammalian Transfection System (Promega, WI, USA) in accordance with the manufacturer's protocol. Cells were plated on 35-mm glass bottom dishes (MatTek, MA) for imaging.

2.9. Confocal microscopy and number and brightness analysis

Constructs containing GFP-PAH fusion proteins were transfected in HeLa cells and experiments were started 18 h after transfection. To evaluate BH4 responsiveness, transfected cells were treated with 60 μM BH4 in aqueous solution. Analyses were performed 6 h later. Number and brightness (N&B) analysis was performed according to Gratton and colleagues [37,38] using a Zeiss LSM 510 META confocal microscope. Cells grown for 4 days on bottom glass dishes were imaged in vivo in CO_2-independent medium (150 mM NaCl, 5 mM KCl, 1 mM $CaCl_2$, 1 mM $MgCl_2$, and 20 mM Hepes, pH 7.4). Fifty frame time series were acquired with a LSM 510 META equipped with a plan apo 63× oil-immersion (NA 1.4) objective lens using the following settings: 488 nm Argon laser, 25% of output power, 1.5% transmission, 505–550 nm emission, gain equal to 850, offset 0.1, and digital gain 1. Scanning parameters were: 512 × 512 frame window, 25.61 μs/pixel dwell time, no average, zoom 6×, ROI (x,y) 256 × 64, and pinhole corresponding to 1 μm optical slice. Data from each cell were analyzed with the simFCS software (Globals Software, East Villa Grove, IL 61956, USA). Briefly, data analysis was performed pixel-by-pixel over time to calculate variance and average intensity. Correction was applied to take into account the analog detection of fluorescence by the photomultiplier tubes of the confocal microscope. Briefly, the correction parameters S, offset and sigma0 were determined, for each experiment, by plotting the measured average intensity (<I>) vs average variance (<Var>) of 50 frame time series acquired using same settings as above, but setting 4 different values of laser transmission percentages. Moreover, filters and beam splitters were configured to obtain reflection images to detect the defined amount of light originating directly from the laser. The obtained plots were linearly interpolated and the equation of straight line ($R \geq 0.99$) was used to extract the parameters S and offset based on the following equation: $<Var> = S * <I> + offset * S$. The parameter sigma0 was estimated from time-series acquired with laser off, as the half maximum width of the histogram peak of the "dark"-counts. Its value was constantly lower than 0.1, and consequently was approximated to zero in all the calculations. Brightness (B) was calculated pixel by pixel from the following equation $B = V(x,y)/(S * Ix,y)$, where I = Im-offset (Im: measured average fluorescence).

The number of molecule subunits constituting each oligomer was calculated with the following calibration procedure: the brightness of monomeric GFP and of GFP-based molecules made by two or three GFP moieties fused in tandem in expression vector pcDNA3.1 and expressed in HeLa cells were measured with the same experimental settings described above. The measured B values for monomeric, dimeric and trimeric GFP and the corresponding number of subunits were linearly interpolated to obtain the experimental equation describing the dependence of brightness on the number of GFP subunits: $B = k \times nsub + c$. This equation was used to calculate the unknown number of GFP subunits corresponding to the measured brightness. The following formula $B = sum(B_i^2\ n_i)/sum(B_i\ n_i)$ provides the brightness of multiple species (B_i) whose number is n_i in a single pixel, is combined in the measured brightness (B). This formula is used to interpret non integer subunit values obtained by the above procedure.

2.10. Structural analysis

The effect of the mutations on the 3D structure was investigated by analyzing the structural environment of each substituted residue in the PAH crystal structures deposited in the Protein Data Bank (PDB). Since no crystal structure of any full-length enzyme is available, various models were used: tetrameric human PAH lacking residues 1–117 (pdbcode 2pah); dimeric truncated human PAH lacking residues 1–117 and 425–452 (PDB code 1kw0); and the dimeric rat PAH, from residue 19 to 427 (pdbcode 1phz), which encompasses the regulatory domain. A composite monomeric model that included the three domains was built by superimposing the secondary structure elements of the respective catalytic domains of the rat (1phz) and human structures (2pah and 1kw0). The structural superposition was performed using the program O [39].

3. Results

3.1. Enzymatic activity analysis of wild-type and mutant forms of PAH

We measured the enzymatic activity of the equilibrium mixture of the wild-type and mutant forms of PAH to investigate the effect of the mutations on catalytic activity. As shown in Fig. 2, the residual activities of the mutant PAH enzymes, expressed as a percentage of wild-type enzyme activity, were: 38 ± 3% (p.I65M), 48 ± 14% (p.N223Y), 60 ± 12% (p.R297L), 69 ± 11% (p.F382L), 58 ± 9% (p.K398N), 43 ± 4% (p.A403V) and 63 ± 5% (p.Q419R). In our conditions, we found in multiple experiments that isolated oligomeric species are not stable and evolve toward the equilibrium mixture (see

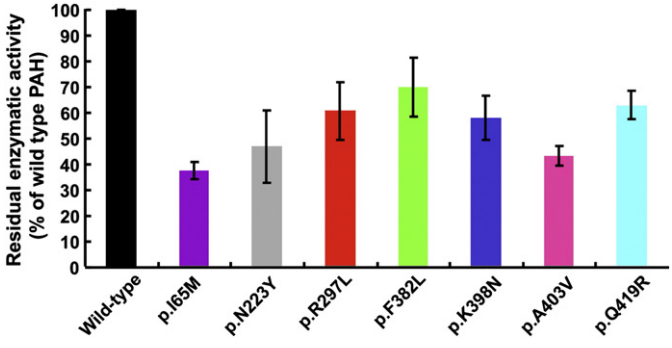

Fig. 2. Enzymatic activity of the PAH mutants expressed as percentage of wild-type PAH activity. The activity was measured by the amount of [^{14}C] Tyr produced by [^{14}C] Phe as analyzed by thin layer chromatography. Mean values and the standard deviation derived from three sets of independent experiments for each construct are shown expressed as percentage of wild-type PAH activity. The experimental assay conditions are reported in Section 2.4.

also below under Section 3.2); hence the activity of the isolated tetramers and dimers was not possible to be measured with accuracy.

3.2. Oligomerization equilibrium of wild-type and mutant forms of PAH

The oligomerization equilibrium for the fused MBP-PAH protein, the purified cleaved wild-type protein, and the missense mutants was analyzed by SEC (Fig. 3a,b). SEC runs were carried out at both 4 °C and room temperature. The elution profile was essentially the same at both temperatures, therefore only the data at room temperature are reported. Unless specified otherwise, SEC measurements were carried at a constant pH of 7.4 (buffer A). The SEC profiles revealed that wild-type and most of the mutant forms of PAH elute as tetramers and dimers although monomers and aggregates are also present, depending on the mutation of the protein. Fig. 3a shows the SEC profiles of proteins in which the tetramer is the most abundant species, i.e., the fused protein, the cleaved wild-type PAH and mutant p.R297L. Fig. 3b shows the profiles of the wild-type PAH (as reference) and of mutants p.I65M, p.N223Y, p.F382L, p.K398N, p.A403V and p.Q419R in which dimer is the dominant species. The peaks corresponding to tetramer, dimer and monomer species elute at 11.5 mL, 13.1 mL and 14.7 mL respectively (Fig. 3a,b).

The presence of the PAH protein was confirmed in each of the fractions obtained by SEC (Fig. 3a) after separation by SDS/PAGE and Western blot analysis using anti-PAH serum (Fig. 3c). This result demonstrates the existence of a monomeric form of the PAH enzyme. This form was probably underestimated in previous studies because it was masked by MBP. Western blot analysis also showed that samples did not contain MBP after cleavage (Fig. 3c).

The percentage of PAH oligomeric species is reported in Fig. 4 (see also Supplementary Table S1). The tetramer and dimer species of the wild-type PAH were separately collected and stored for less than one week at 4 °C. Subsequently, SEC analysis of tetramers and dimers showed that the tetramer shifts back toward the dimer and the monomer, while the dimer is fairly stable, since about 90% of the protein applied was recovered in the dimeric form (data not shown), according to literature data [21].

All chromatographic runs were done in conjunction with MALS. The masses resulting from MALS measurements in wild-type PAH mutants were 203 kDa (within 1%) and 102 kDa (within 3%) for tetramer species and dimer species, respectively. Because of the small quantities of monomer, we were not able to calculate its molecular mass. A representative profile of wild-type PAH obtained from MALS is shown in Fig. 5.

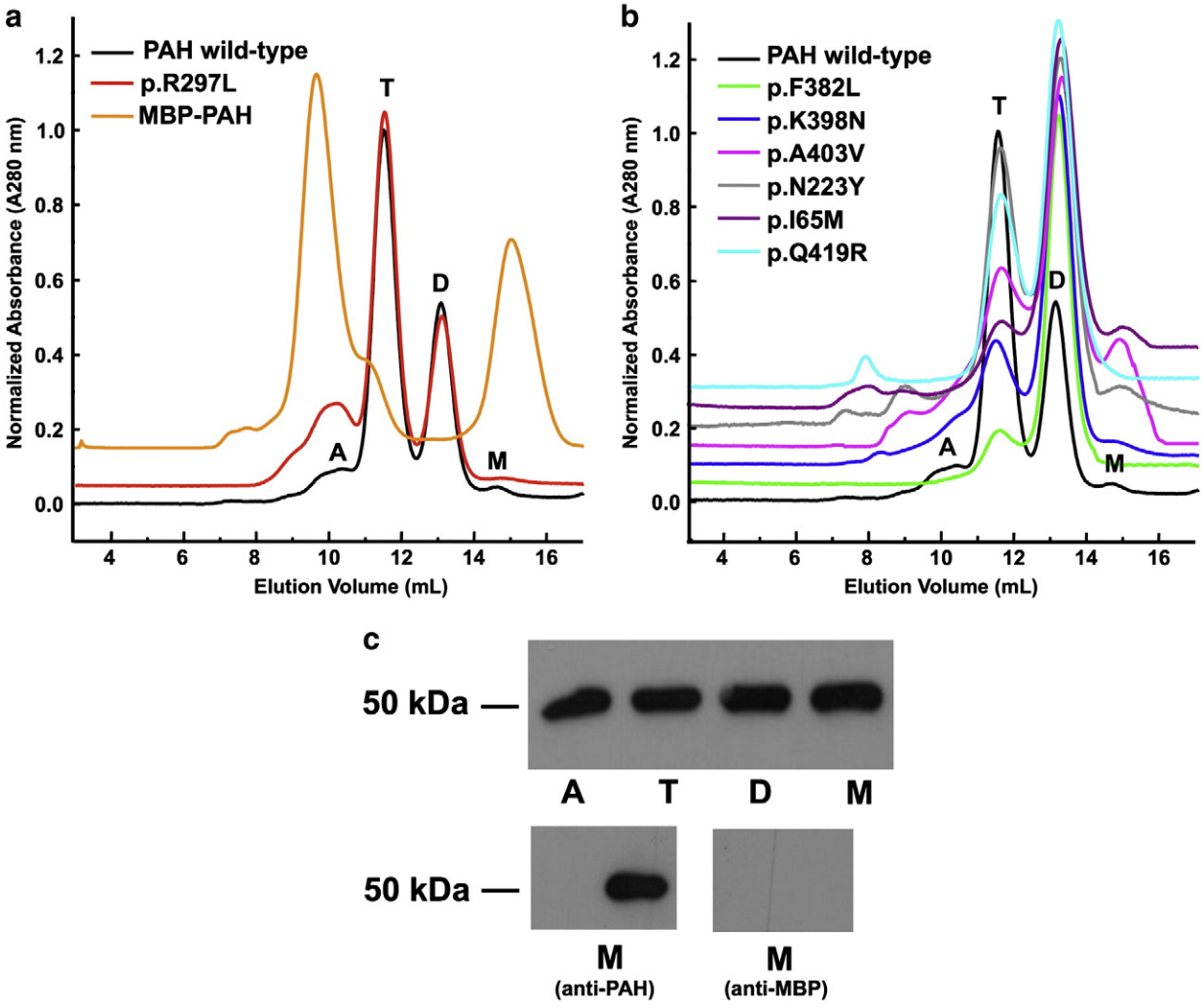

Fig. 3. Chromatographic profiles of the PAH natural variants and immunoblotting of wild-type PAH. a) Size exclusion chromatographic (SEC) runs carried out in buffer A (see text) and normalized in the 0.0 ÷ 1.2 interval. All curves are shifted along the ordinate axis by 0.05n, n integer ($1 \le n \le 5$), with respect to the wild-type PAH curve. b) SEC runs of wild-type and other mutant recombinant PAH proteins. All curves are shifted along the ordinate axis by 0.05n, n integer ($1 \le n \le 5$), with respect to the wild-type PAH curve, to better visualize the single curves. c) Western blot analysis using the anti-PAH antibody of each of the fractions in panel a obtained from wild-type (depicted as a black curve in a); peak A) aggregated forms, possibly multimeric; peak T) tetramer; peak D) dimer; peak M) monomer. Western blot analysis of the wild-type PAH monomer (peak M of panel a) with anti-PAH or anti-MBP antibodies.

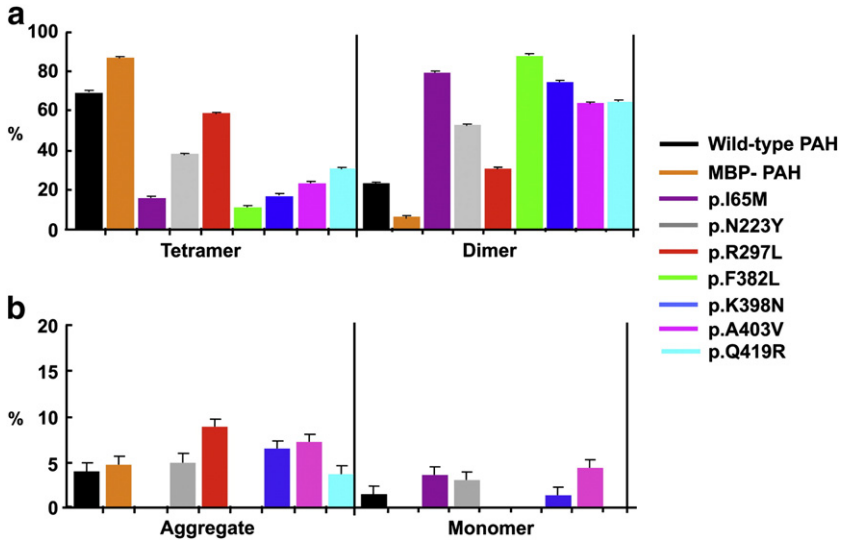

Fig. 4. Percentage of the molecular species of the recombinant wild-type PAH protein and its mutants. The histograms indicate a) the percentage of tetramers and dimers, and b) the percentage of aggregate and monomers of recombinant wild-type PAH, MBP-PAH, p.I65M, p.N223Y, p.R297L, p.F382L, p.K398N, p.A403V and p.Q419R. All data with standard deviations were calculated on at least triplicate runs carried out at 20 °C.

3.3. Thermal stability and inactivation

The dominant oligomeric species of wild-type PAH and mutants, collected after gel filtration and dialyzed against buffer C, were analyzed by CD. The CD spectra of all samples were recorded in the far-UV region (190–250 nm), in buffer C, pH 7.4, at 20 °C. Fig. 6a shows the normalized spectra for wild-type PAH and selected mutants, together with the spectrum for the fused wild-type MBP-PAH. In the spectra of Fig. 6a, the maximum centered at 195 nm and the two broad minima centered at 208 and 222 nm are indicative of the presence of both α and β secondary structure elements. The spectra profiles do not differ significantly from each other, but the molar ellipticity of all the mutants is much lower than that of wild-type enzyme. Therefore, these spectra provide evidence for local unfolding of the mutants.

We also recorded the thermal denaturation curves in the 20÷80 °C temperature range by monitoring the molar ellipticity at 222 nm (Fig. 6b) of the most abundant oligomeric species of cleaved PAH proteins (Figs. 3a,b; 4, Supplementary Table S1). The far-UV CD spectra of denatured species of all samples at 80 °C are typical of random coil (data not shown). The spectrum of cooled samples after denaturation is virtually identical to that of the denatured protein at 80 °C (data not shown), thereby confirming the irreversibility of the unfolding process, as previously reported [24,40]. In the case of the wild-type PAH, the denaturation curve of the equilibrium mixture was also recorded; for comparison purposes, this curve and that of the tetramer are shown in Fig. 6c. It is clear that the denaturation process depends on the oligomeric species.

The sigmoidal denaturation curve obtained with the fused wild-type MBP-PAH protein shows two inflection points, with a plateau, corresponding to the two melting temperatures of the two proteins. Visual inspection of the denaturation curves (Fig. 6b) of the cleaved PAH proteins shows a sigmoidal shape with a single inflection point. Differently, a detailed analysis of the curves through the first derivatives, revealed two midpoints corresponding to two transitions (data not shown), according to literature data [40]. As shown in Table 2, the melting temperatures of the PAH proteins range from 33 to 57 °C. The thermal denaturation curve of mutant p.F382L is significantly shifted to lower temperatures with respect to the wild-type PAH (Fig. 6b). Therefore, we carried out the thermal inactivation assay of the wild-type PAH and the p.F382L mutant at temperatures between 32 and 58 °C (Fig. 6d). Consistently, the p.F382L mutant was much more temperature sensitive than the wild-type enzyme. In fact, under these conditions a lower amount of Tyr was produced (Fig. 6d), and the half-inactivation temperature was close to 44 °C, which well agrees with the value of $T_{m2} = 43.6$ °C obtained by CD. These findings show that the stability of specific mutants, such as p.F382L, can markedly diverge from that of the wild-type protein.

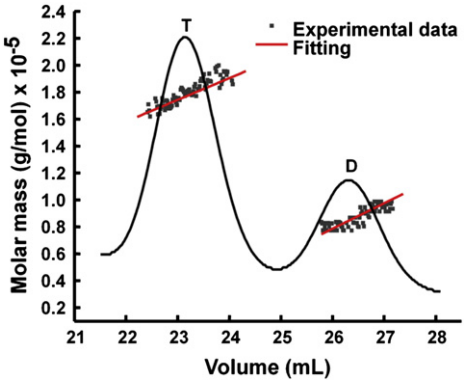

Fig. 5. SEC-MALS analysis of wild-type PAH. Representative profile of wild-type PAH obtained from light scattering and refractive index detectors. The molecular mass is shown. The experimental data are represented by squares and the fitting of experimental data is shown by a red line.

3.4. N&B analysis

Confocal microscopy observation of cells expressing GFP-wild-type PAH and GFP-A403V fusion proteins revealed a cytoplasmic localization of PAH and its mutant p.A403V, as expected (Fig. 7a,b). No fluorescence was detected in the cell nuclei. To estimate the number of subunits constituting wild-type PAH and the p.A403V mutant at steady state in live cells, the recently developed N&B technique was performed on cells expressing the fusion proteins treated and not with BH4 for 6 h. The "brightness" of the GFP-wild-type PAH fusion protein was 1.091 and 1.099 upon BH4 treatment, which, at the t-test, was not significantly different from values obtained without BH4 treatment. On the contrary, in cells expressing the GFP-A403V fusion protein there was a statistically significant ($p<0.001$) shift of brightness from 1.047 without BH4 to 1.094 upon its addition.

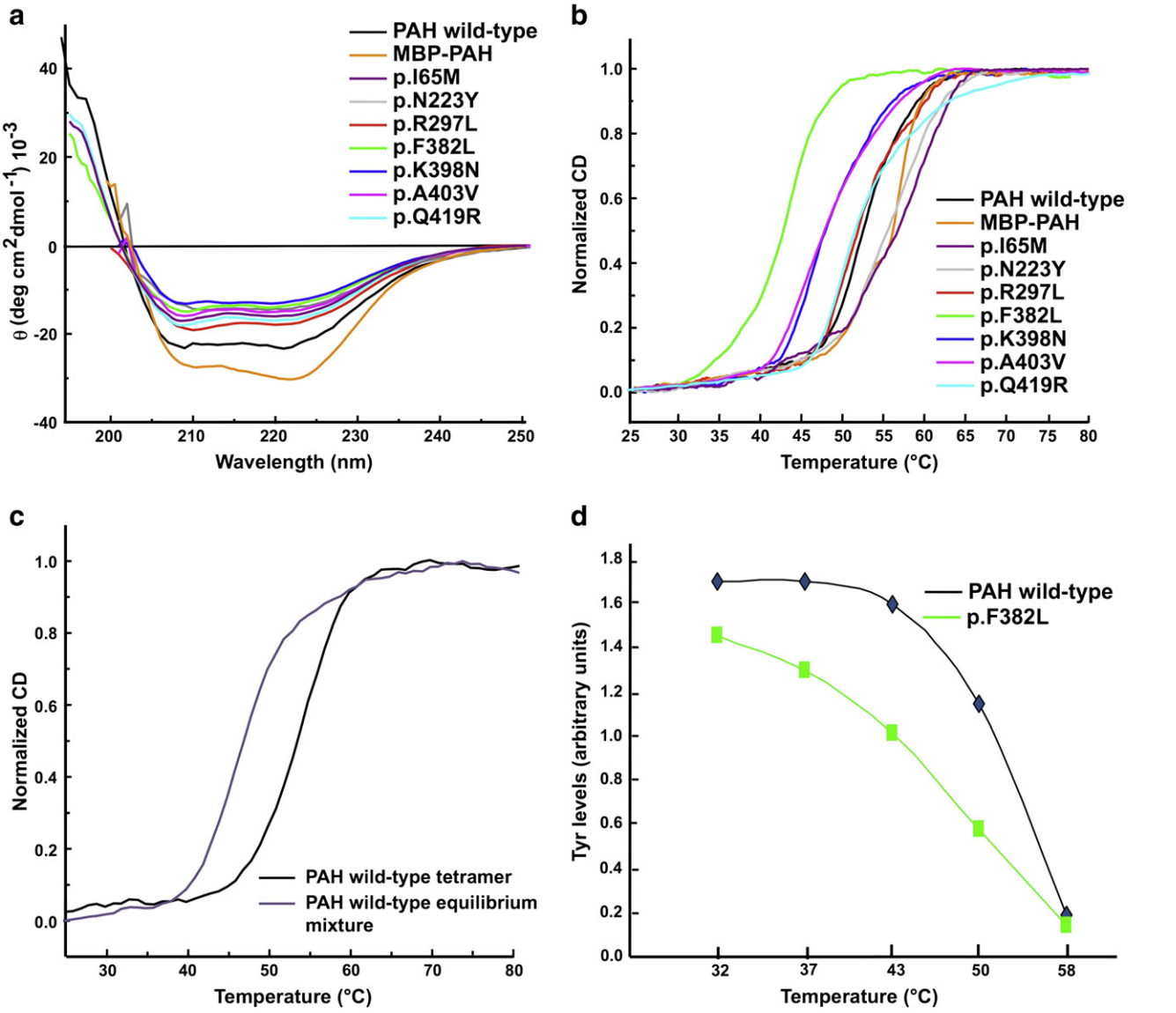

Fig. 6. CD measurements. a) Far-UV spectra recorded in buffer C (see text) at 20 °C. b) CD-monitored thermal unfolding curves in the 25–80 °C temperature range. The change of the molar ellipticity at 222 nm was monitored versus the temperature. The most abundant oligomeric form was considered for each PAH protein. c) Thermal denaturation of the tetrameric form and of the equilibrium mixture of cleaved wild-type PAH. d) Thermal inactivation assay of wild-type and F382L PAH. Proteins were incubated at increasing temperatures and the amount of produced tyrosine was determined.

These brightness values were used to calculate the number of subunits that constitute each PAH oligomer at steady state, namely wild-type and p.A403V before and after BH4 administration (see Material and methods for details). Briefly, a calibration curve (number of subunits vs brightness) was constructed using reference GFP-based molecules constituted by 1, 2 or 3 GFP monomers and measuring their brightness in the same experimental conditions used above. The reference curve obtained (which was perfectly linear) was used to extrapolate the unknown number of subunits from the brightness of GFP-PAHs. The wild-type PAH conserved an average number of subunits (see Material and methods for definition of "average number of subunits") equal to approximately 3.1, which reflects the presence of tetramers and dimers in equilibrium. The addition of BH4 did not affect this number. Differently, exposure of the A403V-PAH mutant form to BH4 caused the subunit number to increase from 2.4 to 3.1, this latter being the same as that of the wild-type enzyme (Fig. 7c).

Table 2
Melting temperatures inferred by first derivative of unfolding curves.[a]

Forms of PAH	$T_{m1} \pm 1.0$ (°C)	$T_{m2} \pm 1.0$ (°C)
Wild-type PAH	52.2	57.6
MBP-PAH	52.0	57.5
p. I65M	53.0	59.2
p.N223Y	54.2	57.6
p.R297L	50.8	59.0
p.F382L	32.6	43.6
p.K398N	46.4	53.6
p.A403V	46.0	55.6
p.Q419R	50.1	57.3

All experiments were conducted in buffer C (see Material and methods).
[a] See also Fig. 6b.

4. Discussion

4.1. The oligomerization equilibrium is shifted toward the dimeric form in PAH mutants

To characterize the oligomeric state of PAH in solution, we performed SEC with on-line detection of LS and refractive index signals. This approach allowed us to measure the molecular mass of

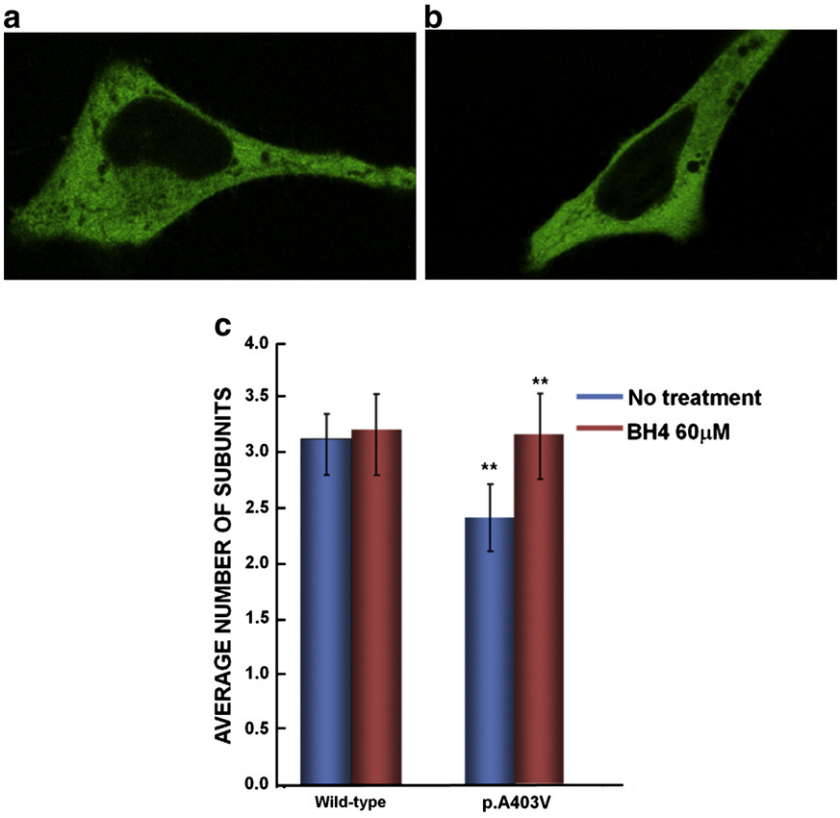

Fig. 7. N&B analysis. a) Confocal microscopy images of cells expressing GFP-wild-type PAH protein and b) GFP-p.A403V mutant protein. c) Histograms showing the oligomeric state of both GFP-wild-type PAH and GFP-p.A403V mutant in untreated (blue bars) and BH4-treated (red bars) cells. Asterisks above the bars indicate statistical significance for the p.A403V mutant protein.

the eluted peaks independently of their shape and/or amino acid composition and elution volume. The operative conditions used in the SEC experiments were the same for all samples. No substrate, cofactor or any other compound was added and only a buffer at physiological pH (buffer A; pH 7.4) was used because a lower pH and the addition of L-Phe can affect the oligomeric equilibrium [10].

In line with previous data [10,21,24], the SEC profile of the fused wild-type MBP-PAH protein revealed that tetramer was the dominant species (88.4%), whereas the dimeric form was less abundant and the monomeric form was absent. Differently, all three species were obtained in the cleaved full-length wild-type enzyme (see Fig. 4 and Supplementary Table S1 for the relative percentages). The observation of three distinct peaks demonstrates that self-association of monomers or dimers is slow on the time scale of separation. Previous studies showed that both fused and truncated/cleaved forms of wild-type and mutant enzymes consist of an equilibrium mixture of dimeric and tetrameric species in a ratio that depends on the protein nature and on the experimental conditions used [10,24].

Only in a few cases in which proteins were expressed in a coupled in vitro transcription–translation eukaryotic system [41,42] was the monomeric form also detected. Here, we report evidence that a monomer species occurs in the equilibrium mixture of PAH albeit in a low amount. The presence of a monomer species may have been underestimated in previous studies for several reasons: i) because of its low abundance, ii) because the fused protein was analyzed [24] or iii) because the cleaved wild-type PAH was analyzed in the presence of MBP which co-elutes with the PAH monomer [10]. Western blot analysis (Fig. 3c) and re-chromatography of the isolated tetramer of wild-type PAH, which regenerate the dimer and the monomer peak (data not shown), indicate that the monomer contributes to the equilibrium mixture. The equilibrium mixture of the PAH enzyme can be described as follows: $4M \leftrightarrows 2D \leftrightarrows T$.

The oligomerization process of the mutants is clearly disturbed by the mutation, except in the case of mutant p.R297L that has a SEC profile similar to that of the wild-type enzyme although it lacks a monomer (Fig. 3a). For all the other mutants (p.I65M, p.N223Y, p.F382L, p.K398N, p.A403V and p.Q419R), the equilibrium was shifted toward the dimeric enzyme (Fig. 3b). In any event, all mutants present a reduced activity of the enzyme because of the diverse composition of the equilibrium mixture. This confirms that oligomerization defects of PAH variants are one of the most common causes of HPAs.

4.2. Thermal stability and defect at the secondary structure level of PAH mutants

The CD spectrum of wild-type PAH is characterized by typical fingerprints of α/β proteins (Fig. 6a), in line with the known 3D structure [43]. Despite the similarity of CD spectra profiles in the wild-type and mutant forms of PAH, the latter present a partial loss of secondary structure with respect the wild-type folded structure, their molar ellipticity being significantly decreased (Fig. 6a). This loss occurs for all the mutants except for mutant p.R297L in which the loss is less pronounced. Therefore, the CD spectra have provided evidence for a folding defect at secondary structure level. To characterize further all the PAH mutants, we estimated the α-helical content using the equation of Richardson and Makhatadze [44]. The calculations showed a progressive decrease of the α-helical content of the PAH mutants (see Supplementary Table S2).

Thermal denaturation of all the PAH proteins proved to be an irreversible process as previously reported [40]. Therefore, the thermodynamic features of the enzyme cannot be further characterized. As expected for the wild-type MBP-PAH fused protein, the two steps in the sigmoidal curve correspond to two independent unfolding processes for the two proteins. Indeed, the two melting

temperatures (Table 2) correspond to the independent structural collapse of PAH (52.0 °C) and MBP (57.5 °C), respectively (see also Supplementary data for details).

Earlier differential scanning calorimetry, fluorescence and CD experiments indicated that, upon heating, two partially overlapping transitions occur with an apparent melting temperature T_{m1} of about 46 °C and a T_{m2} of about 54 °C, respectively [14,40]. Apart from rare exceptions, two transitions were identified also in the mutants. Their melting temperatures diverged from that of wild-type PAH within a range from −6 to +2 °C [24,45]. On the other hand, unfolding experiments conducted on the truncated PAH (aa 112–452) revealed only one cooperative transition centered at about 54 °C [40]. Therefore, the first transition was interpreted as a denaturation of the N-terminal regulatory domain and the second, partially overlapping transition as a denaturation of the catalytic-tetramerization domain [40].

Unfolding experiments were carried out for wild-type PAH and mutants in buffer C and the ellipticity at 222 nm was monitored as a function of temperature. Two overlapping transitions occurred for the wild-type PAH and all the mutants. For the wild-type PAH, the T_{m1} and T_{m2} were 52.2 °C and 57.6 °C, slightly different from previous reports [24,40,45]. This discrepancy is due to the species studied. If the denaturation process is followed using only the PAH tetramer, as in our case, the melting temperatures are $T_{m1} = 52.2$ °C and $T_{m2} = 57.6$ °C. On the other hand, if all the species present in the oligomerization equilibrium are considered, the T_{m1} well agrees with that reported previously ($T_{m1} = 46.1$ °C) (Fig. 6c) [24,40,45]. To verify that the unfolding process depends only on the oligomerization equilibrium, we carried out unfolding experiments in the same buffer and scan-rate conditions used in previous studies (buffer B, scan-rate 0.7 °C/min) [40,45], and obtained a T_{m1} and T_{m2} of 52.0 °C and 54.8 °C, respectively. Indeed, a small destabilization has been reported for most of the previously described mutants, as shown by a decrease in ΔT_m of about a couple of degrees. There are only a few instances where ΔT_m decreases down to ~6 °C [45]. For example, T_{m1} and T_{m2} decreased by as much as 42.6 °C and 52.0 °C and 40.7 °C and 50.2 °C, respectively in the non-natural mutants p.N223D and p.T427P with respect to wild-type enzyme T_{m1} and T_{m2} [45].

The melting temperatures of mutants p.I65M and p.N223Y indicate a very marginal stability gain, being only a couple of degrees higher with respect to the wild-type enzyme (Table 2). Furthermore, the denaturation curves of p.R297L and p.Q419R are similar, and the denaturation slopes differ only slightly from that of the wild-type enzyme, which indicates that the thermal stability is hardly affected. For mutants p.K398N and p.A403V, however, the T_{m1} is about 6° lower and the T_{m2} about 2–3° lower with respect to the wild-type, thereby indicating a loss of stability. Interestingly, the T_{m1} (33.0 °C) and T_{m2} (43.6 °C) of mutant p.F382L are about 20 °C and 14 °C lower with respect to those of the wild-type enzyme, which indicates that this enzyme variant has a high thermal structural instability.

The results of thermal denaturation experiments of the p.F382L mutant suggest that the replacement of L-Phe by L-Leu at position 382 leads to temperature-dependent functional alterations. These results were confirmed by the thermal inactivation assay performed on the p.F382L mutant and, for comparison, on the wild-type PAH. Indeed, the half-inactivation temperature obtained for this mutant (~44 °C) well agrees with the T_{m2} of 43.6 °C obtained by CD (Fig. 6b and d, Table 2).

4.3. Mapping the mutations to search for structural alterations

For the present study, we built a composite model of the full-length human enzyme and used it as a template for structural analysis of the wild-type and mutant proteins. Mutations p.I65M and p.Q419R fall in the regulatory and tetramerization domain, respectively; mutations p.N223Y, p.R297L, p.F382L, p.K398N and p.A403V fall in the large catalytic domain. All mutations are far from the active and cofactor binding sites. All, except mutations at positions 65 and 403, are located on the surface of the molecule in hydrated regions. In Fig. 1, the position of the seven mutations is drawn on the monomer structure. There are four previously identified variants at codon 65 (p.I65T, p.I65V, p.I65N, and p.I65S). The PAH database contains in vitro expression data only for p.I65T, namely, increased proteolytic degradation in pulse-chase experiments, increased aggregation and decreased solubility. The residual activity of the I65T/V mutants, tested in different in vitro systems, was about 21–29% (http://www.pahdb.mcgill.ca).

Ile65 is located in a hydrophobic environment of the regulatory domain. In the mutant enzyme p.I65M, the Ile change to a somewhat larger but differently shaped Met may distort the hydrophobic packing of the region [30]. In addition, Ile 65 makes a favorable interaction with the catalytic domain of another subunit within the dimer (the Ile65 oxygen atom is hydrogen-bonded to the hydroxyl group of Tyr 216). As a result, both the tertiary structure and the quaternary structure are perturbed. These data are in line with the dimer abundance (Fig. 4 and Supplementary Table S1).

Asn223 is located in a solvent exposed loop [46]. In mutant p.N223Y the structural perturbation is mainly due to tyrosine solvent exposure that may affect the function. It has been proposed that this residue is embodied in an intra-domain hinge bending region [45] that is involved in the conformational transition induced by substrate binding. Overall, its properties are similar to those of the wild-type enzyme, except that the dimeric form is favored over the tetramer. Mutant p.R297L exhibits properties very similar to those of the wild-type enzyme. Interestingly, two other mutants reported in the database, i.e., p.R297C and p.R297H, which are also associated with a mild phenotype, indicate that mutations at this site are not critical [32].

Mutant p.F382L is the one that differs most from the wild-type enzyme. It occurs prevalently as a dimer (88.3%). Moreover, the mutation falls in a region rich in aromatic residues. L-Phe forms various stacking aromatic interactions with the aromatic residues Tyr356 and Tyr277 in the wild-type enzyme [33]. The replacement by Leu, despite the hydrophobic nature of this residue, breaks the aromatic network and results in marked destabilization. This is in line with larger ΔT_{m1} (19.6 °C) and ΔT_{m2} (14.0 °C) and with a lower inactivation temperature compared to the wild-type protein (Fig. 6d). This is the largest difference observed so far for a PAH mutant. This highlights that L-Phe in position 382 is crucial for the stability of this aromatic amino acid-rich region.

Lys398 is located at the tetramer interface, although it is not directly engaged in intersubunit interactions. The positively charged Lys is embedded in a negatively charged environment because of the presence of Asp394 and Glu390 [43]. The high dimer/tetramer ratio observed for mutant pK398N (75.0/17.2, Fig. 4 and Supplementary Table S1) is in line with tetramer destabilization. Also the structural stability was decreased. In fact, T_{m1} and T_{m2} were lower than that of the wild-type enzyme being 5.6 and 4.0 °C, respectively (Table 2). Q419 lies in a hydrated region at the dimer/tetramer interface. The substitution of Gln by an Arg generates unfavorable ionic interactions with a nearby Arg241 that destabilize the tetramer. The thermal stability of p.Q419R is not significantly affected by the mutation.

4.4. BH4 effect and the dimer–tetramer p.A403V equilibrium

In this study, we investigated the effect of BH4 on p.A403V that, in our geographic area, is the most frequent mutant in mild, BH4-responsive patients [32,33]. To this aim, we used the N&B technique that has been successfully employed to discern minor changes in molecular composition [37,38]. We monitored the fluorescent molecular forms of GFP-wild-type PAH and the GFP-A403V mutant after transient transfection of HeLa cells. The addition of BH4 to the

wild-type enzyme did not cause any change. Differently, BH4 addition to cells expressing the p.A403V mutant promoted a shift from dimeric to tetrameric molecular forms.

The BH4 responsiveness in HPA has been the subject of numerous studies. Various mechanisms have been proposed to explain BH4-responsiveness, namely: increased enzyme activity [1,30,47,48], correction or compensation of the BH4 decreased affinity, protection toward catalytic inactivation and chaperone-like activity that, by stabilizing the protein, protects it from proteolytic degradation [1,15,30]. In addition, BH4 supplementation may restore the optimal concentration of BH4 cofactor in hepatocytes [28,46]. In our study, the N&B technique revealed that the addition of BH4 to HeLa living cells influences the oligomerization equilibrium of the p.A403V mutant by shifting it toward a tetrameric quaternary structure.

4.5. Concluding remarks

About 600 mutations associated with the PKU/HPA phenotype have been identified in the PAH gene. The resulting phenotypes range from mild to severe HPA. This study demonstrates that our seven gene variants reduce the enzymatic activity to 38–69% of that of wild-type activity and affect the oligomerization state, thermal stability and folding of PAH protein. Perturbation of these biophysical and biochemical features may, therefore, be considered major disease-causing alterations, also in mild phenotypic forms of HPAs. In addition to the well recognized effect of BH4 on the stability of PAH mutants [30], we provide the first evidence, obtained with the N&B technique, that in cellulo exposure to BH4 influences the oligomerization equilibrium of the p.A403V mutant. Therefore, we surmise that, in this case, a shift from dimer to tetramer plays a major role in BH4 responsiveness.

Acknowledgements

This study was supported by grants from Regione Campania (Convenzione CEINGE-Regione Campania, G.R. 27/12/2007), from Ministero dell'Istruzione, dell'Università e della Ricerca-Rome PS35-126/IND, from IRCCS – Fondazione SDN, and from Ministero Salute, Rome, Italy.

We thank Simona Monti and Nina Dathan for their suggestions and advices regarding SEC. Pompea del Vecchio and Vincenzo Granata are acknowledged for their comments regarding CD measurements. We are grateful to Jean Ann Gilder (Scientific Communication srl) for revising and editing the text, and to Vittorio Lucignano for graphic editorial help in figure composition.

Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.bbadis.2011.07.012.

References

[1] C.R. Scriver, The PAH gene, phenylketonuria, and a paradigm shift, Hum. Mutat. 28 (2007) 831–845.
[2] N. Blau, B. Thony, R.G.H. Cotton, K. Hyland, Disorders of tetrahydrobiopterin and related biogenic amines, in: C.R. Scriver, A.L. Beaudet, W.S. Sly, D. Valle, B. Vogelstein (Eds.), The Metabolic and Molecular Bases of Inherited Disease, McGraw_Hill, New York, 2001, pp. 1725–1776.
[3] R.A. Williams, C.D. Mamotte, J.R. Burnett, Phenylketonuria: an inborn error of phenylalanine metabolism, Clin. Biochem. Rev. 29 (2008) 31–41.
[4] M. Giovannini, E. Verduci, E. Salvatici, L. Fiori, E. Riva, Phenylketonuria: dietary and therapeutic challenges, J. Inherit. Metab. Dis. 30 (2007) 145–152.
[5] H.L. Levy, A. Milanowski, A. Chakrapani, M. Cleary, P. Lee, F.K. Trefz, C.B. Whitley, F. Feillet, A.S. Feigenbaum, J.D. Bebchuk, H. Christ-Schmidt, A. Dorenbaum, Efficacy of sapropterin dihydrochloride (tetrahydrobiopterin, 6R-BH4) for reduction of phenylalanine concentration in patients with phenylketonuria: a phase III randomised placebo-controlled study, Lancet 370 (2007) 504–510.
[6] M. Cerreto, R. Nistico, D. Ombrone, M. Ruoppolo, A. Usiello, A. Daniele, L. Pastore, F. Salvatore, Complete reversal of metabolic and neurological symptoms in PKU mice after PAH-HD-Ad vector treatment, Hum. Gene Ther. 11 (2009) 1391.
[7] F.J. Van Spronsen, G.M. Enns, Future treatment strategies in phenylketonuria, Mol. Genet. Metab. 99 (2010) 90–95.
[8] F. Fusetti, H. Erlandsen, T. Flatmark, R.C. Stevens, Structure of tetrameric human phenylalanine hydroxylase and its implications for phenylketonuria, J. Biol. Chem. 273 (1998) 16962–16967.
[9] A.P. Doskeland, A. Martinez, P.M. Knappskog, T. Flatmark, Phosphorylation of recombinant human phenylalanine hydroxylase: effect on catalytic activity, substrate activation and protection against non-specific cleavage of the fusion protein by restriction protease, Biochem. J. 313 (1996) 409–414.
[10] A. Martinez, P.M. Knappskog, S. Olafsdottir, A.P. Doskeland, H.G. Eiken, R.M. Svebak, M. Bozzini, J. Apold, T. Flatmark, Expression of recombinant human phenylalanine hydroxylase as fusion protein in *Escherichia coli* circumvents proteolytic degradation by host cell proteases. Isolation and characterization of the wild-type enzyme, Biochem. J. 306 (Pt 2) (1995) 589–597.
[11] S. Kaufman, The phenylalanine hydroxylating system, Adv. Enzymol. Relat. Areas Mol. Biol. 67 (1993) 77–264.
[12] O.A. Andersen, A.J. Stokka, T. Flatmark, E. Hough, 2.0 Å resolution crystal structures of the ternary complexes of human phenylalanine hydroxylase catalytic domain with tetrahydrobiopterin and 3-(2-thienyl)-L-alanine or L-norleucine: substrate specificity and molecular motions related to substrate binding, J. Mol. Biol. 333 (2003) 747–757.
[13] A. Doskeland, T. Ljones, T. Skotland, T. Flatmark, Phenylalanine 4-monooxygenase from bovine and rat liver: some physical and chemical properties, Neurochem. Res. 7 (1982) 407–421.
[14] F.F. Miranda, M. Thòròlfsson, K. Teigen, J.M. Sanchez-Ruiz, A. Martinez, Structural and stability effects of phosphorylation: localized structural changes in phenylalanine hydroxylase, Protein Sci. 13 (2004) 1219–1226.
[15] C.R. Scriver, P.J. Waters, Monogenic traits are not simple: lessons from phenylketonuria, Trends Genet. 15 (1999) 267–272.
[16] S. Giannattasio, I. Dianzani, P. Lattanzio, M. Spada, V. Romano, F. Cali, G. Andria, A. Ponzone, E. Marra, A. Piazza, Genetic heterogeneity in five Italian regions: analysis of PAH mutations and minihaplotypes, Hum. Hered. 52 (2001) 154–159.
[17] P. Guldberg, F. Rey, J. Zschocke, V. Romano, B. Francois, L. Michiels, K. Ullrich, G.F. Hoffmann, P. Burgard, H. Schmidt, C. Meli, E. Riva, I. Dianzani, A. Ponzone, J. Rey, F. Guttler, 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. 63 (1998) 71–79.
[18] J. Zschocke, Phenylketonuria mutations in Europe, Hum. Mutat. 21 (2003) 345–356.
[19] H.G. Eiken, P.M. Knappskog, J. Apold, T. Flatmark, PKU mutation G46S is associated with increased aggregation and degradation of the phenylalanine hydroxylase enzyme, Hum. Mutat. 7 (1996) 228–238.
[20] M. Thórólfsson, K. Teigen, A. Martinez, Activation of phenylalanine hydroxylase: effect of substitutions at Arg68 and Cys 237, Biochem. 42 (2003) 3419–3428.
[21] P.M. Knappskog, T. Flatmark, J.M. Aarden, J. Haavik, A. Martinez, Structure/function relationships in human phenylalanine hydroxylase. Effect of terminal deletions on the oligomerization, activation and cooperativity of substrate binding to the enzyme, Eur. J. Biochem. 242 (1996) 813–821.
[22] B. Kobe, I.G. Jennings, C.M. House, B.J. Michell, K.E. Goodwill, B.D. Santarsiero, R.C. Stevens, R.G. Cotton, B.E. Kemp, Structural basis of autoregulation of phenylalanine hydroxylase, Nat. Struct. Biol. 6 (1999) 442–448.
[23] P.J. Waters, M.A. Parniak, B.R. Akerman, A.O. Jones, C.R. Scriver, Missense mutations in the phenylalanine hydroxylase gene (PAH) can cause accelerated proteolytic turnover of PAH enzyme: a mechanism underlying phenylketonuria, J. Inherit. Metab. Dis. 22 (1999) 208–212.
[24] S.W. Gersting, K.F. Kemter, M. Staudigl, D.D. Messing, M.K. Danecka, F.B. Lagler, C.P. Sommerhoff, A.A. Roscher, A.C. Muntau, Loss of function in phenylketonuria is caused by impaired molecular motions and conformational instability, Am. J. Hum. Genet. 83 (2008) 5–17.
[25] T. Gjetting, M. Petersen, P. Guldberg, F. Guttler, In vitro expression of 34 naturally occurring mutant variants of phenylalanine hydroxylase: correlation with metabolic phenotypes and susceptibility toward protein aggregation, Mol. Genet. Metab. 72 (2001) 132–143.
[26] P.J. Waters, M.A. Parniak, B.R. Akerman, C.R. Scriver, Characterization of phenylketonuria missense substitutions, distant from the phenylalanine hydroxylase active site, illustrates a paradigm for mechanism and potential modulation of phenotype, Mol. Genet. Metab. 69 (2000) 101–110.
[27] A.L. Pey, F. Stricher, L. Serrano, A. Martinez, Predicted effects of missense mutations on native-state stability account for phenotypic outcome in phenylketonuria, a paradigm of misfolding diseases, Am. J. Hum. Genet. 81 (2007) 1006–1024.
[28] C. Aguado, B. Perez, M. Ugarte, L.R. Desviat, Analysis of the effect of tetrahydrobiopterin on PAH gene expression in hepatoma cells, FEBS Lett. 580 (2006) 1697–1701.
[29] M. Staudigl, S.W. Gersting, M.K. Danecka, D.D. Messing, M. Woidly, D. Pinkas, K.F. Kemter, N. Blau, A.C. Muntau, The interplay between genotype, metabolic state and cofactor treatment governs phenylalanine hydroxylase function and drug response, Hum. Mol. Genet. 20 (2011) 2628–2641.
[30] H. Erlandsen, A.L. Pey, A. Gamez, B. Perez, L.R. Desviat, C. Aguado, R. Koch, S. Surendran, S. Tyring, R. Matalon, C.R. Scriver, M. Ugarte, A. Martinez, R.C. Stevens, Correction of kinetic and stability defects by tetrahydrobiopterin in phenylketonuria patients with certain phenylalanine hydroxylase mutations, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 16903–16908.
[31] A.C. Muntau, S.W. Gersting, Phenylketonuria as a model for protein misfolding diseases and for the development of next generation orphan drugs for patients with inborn errors of metabolism, J. Inherit. Metab. Dis. 33 (2010) 649–658.

[32] A. Daniele, G. Cardillo, C. Pennino, M.T. Carbone, D. Scognamiglio, A. Correra, A. Pignero, G. Castaldo, F. Salvatore, Molecular epidemiology of phenylalanine hydroxylase deficiency in Southern Italy: a 96% detection rate with ten novel mutations, Ann. Hum. Genet. 71 (2007) 185–193.
[33] A. Daniele, G. Cardillo, C. Pennino, M.T. Carbone, D. Scognamiglio, L. Esposito, A. Correra, G. Castaldo, A. Zagari, F. Salvatore, Five human phenylalanine hydroxylase proteins identified in mild hyperphenylalaninemia patients are disease-causing variants, Biochim. Biophys. Acta. 1782 (2008) 378–384.
[34] A. Daniele, I. Scala, G. Cardillo, C. Pennino, C. Ungaro, M. Sibilio, G. Parenti, L. Esposito, A. Zagari, G. Andria, F. Salvatore, Functional and structural characterization of novel mutations and genotype–phenotype correlation in 51 phenylalanine hydroxylase deficient families from Southern Italy, FEBS J. 276 (2009) 2048–2059.
[35] T. Bardelli, M.A. Donati, S. Gasperini, F. Ciani, F. Belli, N. Blau, A. Morrone, E. Zammarchi, Two novel genetic lesions and a common BH4-responsive mutation of the PAH gene in Italian patients with hyperphenylalaninemia, Mol. Genet. Metab. 77 (2002) 260–266.
[36] R.N. Carvalho, T. Solstad, E. Bjørgo, J.F. Barroso, T. Flatmark, Deamidations in recombinant human phenylalanine hydroxylase. Identification of labile asparagine residues and functional characterization of Asn–>Asp mutant forms, J. Biol. Chem. 278 (2003) 15142–15152.
[37] M.A. Digman, C.M. Brown, P. Sengupta, P.W. Wiseman, A.F. Horwitz, E. Gratton, Measuring fast dynamics in solutions and cells with a laser scanning microscope, Biophys. J. 89 (2010) 1317–1327.
[38] R.B. Dalal, M.A. Digman, A.F. Horwitz, V. Vetri, E. Gratton, Determination of particle number and brightness using a laser scanning confocal microscope operating in the analog mode, Microsc. Res. Technol. 71 (2008) 69–81.
[39] T.A. Jones, J.Y. Zou, S.W. Cowan, M. Kjeldgaard, Improved methods for building protein models in electron density maps and the location of errors in these models, Acta Crystallogr.A. 47 (1991) 110–119.
[40] M. Thòròlfsson, B. Ibarra-Molero, P. Fojan, S.B. Peterson, J.M. Sanchez-Ruiz, A. Martinez, L-Phenylalanine binding and domain organization in human phenylalanine hydroxylase: a differential scanning calorimetry study, Biochem. 41 (2002) 7573–7585.
[41] E. Bjørgo, P.M. Knappskog, A. Martinez, R.C. Stevens, T. Flatmark, 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. 257 (1998) 1–10.
[42] A.L. Pey, B. Perez, L.R. Desviat, M.A. Martinez, C. Aguado, H. Erlandsen, A. Gamez, R.C. Stevens, M. Thórólfsson, M. Ugarde, A. Martinez, Mechanisms underlying responsiveness to tetrahydrobiopterin in mild phenylketonuria mutations, Hum. Mutat. 24 (2004) 388–399.
[43] H. Erlandsen, F. Fusetti, A. Martinez, E. Hough, T. Flatmark, R.C. Stevens, Crystal structure of the catalytic domain of human phenylalanine hydroxylase reveals the structural basis for phenylketonuria, Nat. Struct. Biol. 4 (1997) 995–1000.
[44] J.M. Richardson, G.I. Makhatadze, Temperature dependence of thermodynamics of helix-coil transition, J. Mol. Biol. 335 (2004) 1029–1037.
[45] A.J. Stokka, R.N. Carvalho, J.F. Barroso, T. Flatmark, Probing the role of crystallographically defined/predicted hinge-bending regions in the substrate-induced global conformational transition and catalytic activation of human phenylalanine hydroxylase by single-site mutagenesis, J. Biol. Chem. 279 (2004) 26571–26580.
[46] S. Kure, K. Sato, K. Fujii, Y. Aoki, Y. Suzuki, S. Kato, Y. Matsubara, Wild-type phenylalanine hydroxylase activity is enhanced by tetrahydrobiopterin supplementation in vivo: an implication for therapeutic basis of tetrahydrobiopterin-responsive phenylalanine hydroxylase deficiency, Mol. Genet. Metab. 83 (2004) 150–156.
[47] N. Blau, H. Erlandsen, The metabolic and molecular bases of tetrahydrobiopterin-responsive phenylalanine hydroxylase deficiency, Mol. Genet. Metab. 82 (2004) 101–111.
[48] A.C. Muntau, W. Roschinger, M. Habich, H.H. Demmelmair, B. Hoffmann, C.P. Sommhoff, A.A. Roscher, Tetrahydrobiopterin as an alternative treatment for mild phenylketonuria, N. Engl. J. Med. 347 (2002) 2122–2132.

Further reading

[49] V. Novokahtny, K. Ingham, Thermodynamics of maltose binding protein unfolding, Protein Sci. 6 (1997) 141–146.

Paper II

OPEN ACCESS Freely available online

PLoS one

Glucokinase (GCK) Mutations and Their Characterization in MODY2 Children of Southern Italy

Marina Capuano[1,2☯], Carmen Maria Garcia-Herrero[3,4☯], Nadia Tinto[1,2], Carla Carluccio[2,5], Valentina Capobianco[6], Iolanda Coto[1,2], Arturo Cola[1,2], Dario Iafusco[7], Adriana Franzese[8], Adriana Zagari[2,5], Maria Angeles Navas[3,4], Lucia Sacchetti[1,2]*

1 Department of Biochemistry and Medical Biotechnology, University of Naples "Federico II", Naples, Italy, **2** Centro di Ingegneria Genetica (CEINGE) Advanced Biotechnology, s. c. a r. l., Naples, Italy, **3** Department of Biochemistry and Molecular Biology III, Faculty of Medicine, Complutense University of Madrid, Madrid, Spain, **4** Centro de Investigación Biomédica en Red de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), Barcelona, Spain, **5** Department of Biological Science, University of Naples "Federico II", Naples, Italy, **6** Fondazione SDN-IRCSS (Istituto di Diagnostica Nucleare-Istituto di Ricerca e Cura a Carattere Scientifico), Naples, Italy, **7** Department of Pediatrics, Second University of Naples, Naples, Italy, **8** Department of Pediatrics, University of Naples "Federico II", Naples, Italy

Abstract

Type 2 Maturity Onset Diabetes of the Young (MODY2) is a monogenic autosomal disease characterized by a primary defect in insulin secretion and hyperglycemia. It results from GCK gene mutations that impair enzyme activity. Between 2006 and 2010, we investigated GCK mutations in 66 diabetic children from southern Italy with suspected MODY2. Denaturing High Performance Liquid Chromatography (DHPLC) and sequence analysis revealed 19 GCK mutations in 28 children, six of which were novel: p.Glu40Asp, p.Val154Leu, p.Arg447Glyfs, p.Lys458_Cys461del, p.Glu395_Arg397del and c.580-2A>T. We evaluated the effect of these 19 mutations using bioinformatic tools such as Polymorphism Phenotyping (Polyphen), Sorting Intolerant From Tolerant (SIFT) and *in silico* modelling. We also conducted a functional study to evaluate the pathogenic significance of seven mutations that are among the most severe mutations found in our population, and have never been characterized: p.Glu70Asp, p.His137Asp, p.Phe150Tyr, p.Val154Leu, p.Gly162Asp, p.Arg303Trp and p.Arg392Ser. These seven mutations, by altering one or more kinetic parameters, reduced enzyme catalytic activity by >40%. All mutations except p.Glu70Asp displayed thermal-instability, indeed >50% of enzyme activity was lost at 50°C/30 min. Thus, these seven mutations play a pathogenic role in MODY2 insurgence. In conclusion, this report revealed six novel GCK mutations and sheds some light on the structure-function relationship of human GCK mutations and MODY2.

Citation: Capuano M, Garcia-Herrero CM, Tinto N, Carluccio C, Capobianco V, et al. (2012) Glucokinase (GCK) Mutations and Their Characterization in MODY2 Children of Southern Italy. PLoS ONE 7(6): e38906. doi:10.1371/journal.pone.0038906

Editor: Pal Bela Szecsi, Lund University Hospital, Sweden

Received March 12, 2012; **Accepted** May 14, 2012; **Published** June 20, 2012

Funding: Work was supported by grants from: CEINGE-Regione Campania, Italy (DGRC 1901/2009), the Instituto de Salud Carlos III, Spain (PI1000424) and the Dirección General de Universidades e investigación de la Consejería de Educación de la Comunidad de Madrid-Universidad Complutense de Madrid, Spain (CCG10-UCM/BIO-4728). C.M. García-Herrero was supported by a predoctoral FPU fellowship from the Ministerio de Ciencia e Innovación of Spain. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: sacchett@unina.it

☯ These authors contributed equally to this work.

Introduction

Maturity Onset Diabetes of the Young (MODY), a monogenic diabetes inherited in an autosomal dominant mode, accounts for 1–2% of all diabetes forms in Europe [1,2]. It is a clinically heterogeneous group of diseases caused by at least eight gene defects in the pancreatic β-cell that impair insulin secretion [3]. MODY is characterized by onset before 25 years of age; patients usually lack auto-antibodies, and clinical manifestations go from slight non-ketotic hyperglycemia to severe complicated hyperglycemia [4]. Heterozygous mutations in the *glucokinase* (*GCK*) gene produce two distinct diseases, MODY type 2 (MODY2) (MIM:125851) and persistent hyperinsulinemia of infancy (MIM:601820) [5,6]. Persistent hyperinsulinemia of infancy is associated with hyperactive GCK variants, while MODY2 is associated with GCK mutations that impair its activity [7]. GCK (hexokinase IV) catalyzes the ATP-dependent phosphorylation of glucose in the first, rate-limiting step of glycolysis in pancreatic β-cells [1]. The crystal structure determination of the enzyme by Kamata et al. [8] revealed that the enzyme exists in at least two distinct forms, the super-open ligand-free form and the closed active form that is bound to glucose and ATP. The molecule consists of two domains, namely a small and large domain separated by the glucose site. In detail, amino acid residues 1–64 and 206–439 belong to the large domain, amino acid residues 72–201 and 445–465 to the small domain, and amino acid residues 65–71, 202–205 and 440–444 to the three loops connecting the domains [8]. The GCK protein switches from an inactive conformation to a close active conformation upon ligand binding. A huge conformational transition occurs through a large rotation of the small domain [8].

The heterozygous loss-of-function GCK mutations causative of MODY2 diabetes include missense, nonsense, splicing, small deletions/insertions/duplications variants, and result in stable fasting hyperglycemia from birth (>5.5 mol/L) and rare microvascular complications [1]. Over 644 *GCK* mutations have been

described, and a study of the mutational mechanisms for a number of these has shed light on GCK regulation [9].

The molecular diagnosis of MODY2 is important: to classify the type of diabetes correctly, to predict prognosis, and to initiate screening of family members [1]. It is particularly important to identify MODY2 in diabetic pregnant patients in order to start "ad hoc" treatment [10]. However, the identification of *GCK* mutations by molecular analysis will not always reveal whether a variant is really pathogenic or how serious the diabetic phenotype could be. Therefore, in this five-year study, we applied an integrated approach to investigate the effect of *GCK* mutations in the diabetic phenotype in children from southern Italy. First, we used DHPLC and mutation sequencing to identify variants, then a computational approach to predict the effect of the variants identified, and finally a functional study to determine the effect of seven candidate variants on enzyme activity and on enzyme thermo-stability.

Results

Among the 66 enrolled patients with suspected MODY2, 28/66 were diagnosed as MODY2 based on mutations in the *GCK* gene (MODY2+) and 38/66 were MODY2-negative (MODY2−). All mutated patients were unrelated, except two pairs of siblings (patient identification: MD19/20: two sisters and MD69/70: brother/sister). The mean age at diagnosis (± Standard Deviation: SD) was lower, albeit not significantly, in MODY2+ than in MODY2− patients (105±45 months vs 113±45 months). Mean triglyceride levels did not differ between the two groups (0.6±0.2 and 0.7±0.3 mmol/L, respectively). Mean Fasting Plasma Glucose (FPG) and glycosylated hemoglobin (HbA1c) concentrations were significantly higher ($p<0.003$ and $p<0.001$, respectively) in MODY2+ than in MODY2− patients (6.7±0.8 mmol/L and 6.2±0.3% vs 6.1±0.7 mmol/L and 5.5±0.5%, respectively). The Body Mass Index zeta-score (BMIz-score) of enrolled children at first admission was always in the reference range for the children's age (reference range: −1.5/+1.5), except in one patient (BMIz-score: 2.5). Two/28 MODY2+ patients were positive for only one type-1 diabetes auto antibody (glutamate decarboxylase: 8 and 18 U/mL). The patients were untreated until diagnosis.

Identification of *GCK* Mutations

We identified 19 different *GCK* mutations in 28/66 patients (Table 1). Six of them were novel: two missense mutations (p.Val154Leu and p.Glu40Asp), one splice site variant (c.580-2A>T) and three deletions (p.Lys458_Cys461del, p.Arg447Glyfs, and p.Glu395_Arg397del). We also found 13 previously reported mutations: 11 missense mutations (p.Arg303Trp, p.Gly261Arg, p.Phe150Tyr, p.Ala259Thr, p.Glu70Asp, p.Lys420Glu, p.Ala188Thr, p.Tyr289Cys, p.Asp278Glu, p.Gly223Ser, and p.Ala449Thr) and two splice site variants (c.864-1G>A and c.1019+5G>A). All the detected mutations were always present in the children's mother or father and were absent from 200 chromosomes of our euglycemic controls. Seven mutations were each present in two unrelated families: p.Arg303Trp, p.Lys458_Cys461del, p.Arg447Glyfs, p.Gly261Arg, p.Phe150Tyr, p.Lys420Glu and p.Ala259Thr, the latter was also identified in two siblings. The splice variant c.580-2A>T was present in two siblings. Table 1 shows, for each mutation (except the three splice variants c.864-1G>A, c.580-2A>T and c.1019+5G>A, the frameshift mutation p.Arg447Glyfs and the p.Lys458_Cys461del), the nucleotide position, the amino acid change (if present), the effect on the protein predicted by bioinformatics.

Bioinformatics Study of the GCK Variants

All substitutions, except p.Val154Leu and p.Glu40Asp, were predicted to be deleterious mutations by online prediction tools (Table 1, column 4). Overall, the theoretical structural models of the mutants we obtained *in silico* (Table 1, column 5), preserve the overall protein fold. In detail, however, all mutations produced local conformational alterations − in some cases, such as p.Gly162Asp, dramatic perturbations – that well account for the functional alterations we found (Table 2). **p.Arg303Trp**-Arg303 is a highly conserved residue located in the α8 helix within the large domain. Mutation p.Arg303Trp disrupts two salt-bridges between the side chain of Arg303 and the side chain of Glu300, located in the same helix. These salt-bridges may be essential for the stability of the helix and their loss could destabilize the helix structure. **p.Val154Leu**-The p.Val154Leu mutation (Figure 1) does not cause any significant change in the local structure. Indeed, Val154 is located in the β-sheet that encompasses the small domain hydrophobic core. Its substitution with a leucine residue does not affect the hydrophobic interactions. Nevertheless, Val154 is involved in the large conformational variation from the super-open to close form upon glucose binding (Figure 1A). **p.Gly261Arg**-Gly261 is located in the loop connecting the β-sheet and the α6 helix in the large domain. p.Gly261Arg is a dramatic mutation because the small, flexible hydrophobic Gly residue is replaced by a very large Arg residue that bears a positive net charge. This substitution causes a local re-arrangement that involves the nearby residues such as Leu266 and Leu270. This process results in destabilization of the local structure.

p.Phe150Tyr-Phe150 is a highly conserved residue located in the β-sheet that encompasses the small domain hydrophobic core. The p.Phe150Tyr mutation (Figure 1B, C) introduces a polar residue inside the hydrophobic core in a region rich in phenylalanine. This replacement can influence the stability of the β-sheet thereby altering the structure of the domain. **p.Ala259Thr-**Ala259 residue is located in the large domain near the glucose-binding cleft. Introduction of a larger and polar side chain of threonine in the p.Ala259Thr mutant could influence the hydrogen bond network in that area. In fact, Thr259 can compete with other residues in the formation of hydrogen bonds with two water molecules. See our previous study for a description of the **p.Glu70Asp** variant [11]. **p.Lys420Glu-**Mutation p.Lys420Glu involves an inversion from a positively charged lysine to a negatively charged glutamic acid. The substitution affects the interaction of Lys420 with surrounding residues. Indeed, Lys420 is located in the α12 helix (Figure S1) in the large domain and forms a salt bridge with Glu440 located in a connecting loop between the two domains. The loss of this bond, caused by the mutation, could destabilize the region. **p.Ala188Thr-**The p.Ala188Thr mutation alters a highly conserved amino acid. Ala188 is located in the α4 helix within the small domain. The mutation leads to the substitution of a hydrophobic residue, alanine, with a polar residue, tyrosine, on the hydrophobic side of an amphipathic helix. Thr188 can form hydrogen bonds through the hydroxyl group with the side chains of the Ser127 and Asp124 residues on the α3 helix that is on the surface of the enzyme. The introduction of the threonine can cause a different arrangement of such side chain. **p.Tyr289Cys-**Tyr289 is located in the α7 helix in the large domain. The substitution of the bulk tyrosine in position 289 with the smaller cysteine side chain leads to the formation of a cavity. The mutant disrupts a favorable interaction between Tyr289 and Met381 of the α11 helix that may be important to keep helices together. It is noteworthy that, in our model, Cys289 is not bound to the nearby Cys230 because the distance (4.1 Å) between the two sulfur atoms too long for a disulfide bond formation. **p.As-**

Table 1. GCK mutations detected in MODY2 children from South Italy.

Patient code	GCK Exon[a]	cDNA mutation[b]	Polyphen/SIFT prediction[c]	Aminoacid change[d]/Domain localization/Secondary structure	Effect on protein 3D-structure	Reference
MD01/92	8	c.907C>T	1/deleterious	p.Arg303Trp/Large domain/α8 helix	Disruption of two salt-bridges	[25]
MD05/71	10	c.1373_1385del		p.Lys458_Cys461del/Small domain/ α13 helix		NOVEL
MD10	4	c.460G>T	2/tolerated	p.Val154Leu/Small domain/β-strand 6	Severe perturbation during the transition from the super-open to closed form. Mild local structural alteration	NOVEL
MD12/89	10	c.1339delC		p.Arg447Glyfs/Small domain/α13 helix		NOVEL
MD16/75	7	c.781G>T	1/deleterious	p.Gly261Arg/Large domain/loop	Replacement of a small and flexible hydrophobic residue with a large positive residue	[26]
MD18/94	4	c.449T>A	1/deleterious	p.Phe150Tyr/Small domain/β-strand 5	Introduction of a polar residue in the hydrophobic core. Perturbation of the β-sheet	[27]
MD19[e]/20[e]/79	7	c.775G>A	3/deleterious	p.Ala259Thr/Large domain/loop	Influence on the hydrogen bond network near the glucose-binding cleft	[28]
MD25	2	c.210A>C	1/deleterious	p.Glu70Asp/Connection/Loop spatially near α13 helix	Weakness of salt bridge interaction with K458 (α13 helix)	[11]
MD26/65	10	c.1258A>G	2/deleterious	p.Lys420Glu/Large domain/α12 helix	Loss of a salt bridge between K420 and E440	[29]
MD38	Intron 7	c.864-1G>A		IVS7-1G>A		[9]
MD57	5	c.562G>A	1/deleterious	p.Ala188Thr/Small domain/α4 helix	Substitution of a hydrophobic residue with a polar residue	[30]
MD59	8	c.866A>G	1/deleterious	p.Tyr289Cys/Large domain/α7 helix	Disruption of a favorable interaction between Y289 and M381	[9]
MD68	7	c.834G>A	1/deleterious	p.Asp278Glu/Large domain/α6 helix	Mild structural alteration	[9]
MD69[e]/70 [e]	Intron 5	c.580-2A>T		IVS5-2A>T		NOVEL
MD80	6	c.667G>A	3/deleterious	p.Gly223Ser/Large domain/β-strand 9	Perturbation of the β-strand	[31]
MD86	Intron 8	c.1019+5G>A		IVS8+5G>A		[27]
MD90	2	c.120G>C	2/tolerated	p.Glu40Asp/Large domain/α2 helix	Mild structural alteration	NOVEL
MD91	10	c.1345G>C	3/deleterious	p.Ala449Thr/Small domain in the closed form/α13 helix	Introduction of a larger and polar side chain	[9]
MD95	9	c.1182_1191del		p.Glu395_Arg397del/Large domain/ last residue of α11 helix and following loop	Destabilization of the C-terminal region of the helix and shortening of loop	NOVEL

[a]GenBank: accession number (AH005826). [b]The reference cDNA sequence was obtained from GenBank (NM_000162) and +1 corresponds to the A of the ATG translation initiation codon. [c]Polyphen prediction: probably damaging (1), benign (2), possibly damaging (3). SIFT score: <0.05 deleterious variant, ≥0.05 tolerated variant. [d]Swissprot accession number: P35557. [e]Sibling pairs (MD19/20: two sisters; MD69/70: brother/sister).
doi:10.1371/journal.pone.0038906.t001

p278Glu-The p.Asp278Glu mutation affects a highly conserved amino acid. Asp278 is located on the polar side of the α6 helix in the large domain. The replacement of Asp with a Glu residue in the mutant does not seem to cause significant changes in the enzyme structure. **p.Gly223Ser-**Gly223 is a highly conserved residue located in the β-sheet of the large domain hydrophobic core. The substitution with a serine involves the serine side-chain hydrogen bonding to Cys 233, close to Gly223 in the structure of GCK and could perturb the β-strand. **p.Glu40Asp-**The pGlu40Asp mutation substitutes a conserved glutamate residue with aspartate. Glu40 is located on the polar side of the α2 helix in the large domain. These amino acids are acidic residues but the side chain of the aspartate is shorter than the side chain of the glutamate. Nevertheless, the mutation does not seem to cause significant changes in the structure of the enzyme. **p.Ala449Thr-**Ala449 is a highly conserved residue located on the C-terminal α13 helix. This helix is part of the small domain in the closed form, but it lies between the two domains in the super-open form. In the closed form, the small domain has a three-layer architecture and the α13 helix is in the inner layer. At the domain interface, the α13 helix makes favorable interactions with the α5 helix of the large domain. Because the conformational features of this region are essential for the super-open (inactive form)/closed (active form) conversion, the introduction of a larger and polar side chain of the threonine in the p.Ala449Thr mutant could influence the enzymatic activity and/or stability. **p.Glu395_Arg397del–**The p.Glu395_Arg397 deletion causes the elimination of the last residue, Glu395, of the α11 helix in the large domain and two residues of the following loop, Ser396 and Arg397. In the wild-type enzyme, Glu395 is involved, through the backbone N atom, in the formation of a hydrogen bond of the α11 helix. Its deletion in the mutant causes the lack of this bond and could destabilize the C-terminal region of the helix. Moreover, also the backbone N atom of Arg397 interacts with an oxygen atom of Arg394 thereby

Table 2. Kinetic constants of human recombinant wild type-GCK and mutant β-cell GST-GCK fusion proteins.

Preparation	Protein yield (mg/L of culture)	Kcat (s^{-1})	$S_{0.5}$ for glucose (mmol/L)	nH	Km for ATP (mmol/L)	I_a
Wild-type GST-GCK	3.12±0.91	44.10±7.50	7.20±0.40	1.40±0.08	0.40±0.00	1.0±0.07
GST-GCK(Glu70Asp]	2.35±0.55	27.40±4.80*	20.10±3.20†	1.20±0.07	0.25±0.05	0.25±0.06**
GST-GCK(His137Asp]	2.82±0.47	17.20±1.20**	18.10±3.10*	1.10±0.03**	0.23±0.03*	0.26±0.09**
GST-GCK(Phe150Tyr]	3.07±0.55	9.92±2.38*	101.40±6.20**	1.06±0.05*	3.11±0.48*	0.014±0.003*
GST-GCK(Val154Leu]	2.77±0.48	46.60±4.40	26.00±3.40*	1.54±0.01	1.62±0.16*	0.099±0.036**
GST-GCK(Gly162Asp]	2.78±0.58	Undetectable activity				
GST-GCK(Arg303Trp]	1.58±0.26	14.60±1.38**	4.62±0.10**	1.52±0.02	0.29±0.01*	0.59±0.09*
GST-GCK(Arg392Ser]	2.44±0.74	41.30±11.30	11.90±1.70*	1.30±0.03	0.63±0.03*	0.60±0.25

Data represent means ± SEM of 3 separate enzyme expressions each tested in duplicate. Note that the Hill coefficient (nH) and the relative activity index (I_a) are unit less. Kcat: GCK catalytic constant; $S_{0.5}$: affinity constant for glucose; nH: Hill coefficient; Km for ATP: affinity constant for ATP; I_a: GCK activity index. (*)$p<0.05$, *t* test; (**)$p<0.005$, *t* test.

doi:10.1371/journal.pone.0038906.t002

stabilizing the helix. In the mutant, the loop becomes shorter and the side chain of Arg394, which in the wild-type enzyme is directed towards the loop, changes the orientation because of steric interactions. This change disrupts interactions between the side chain of Arg394 and some residues of the loop. Although the interaction between the side chains of Arg394 and Ser433 in the loop following the α12 helix is conserved, the binding appears to be weakened compared with the wild-type enzyme.

Kinetic Analysis and Thermo-stability of Recombinant GCK Mutants

To select representative mutations for a functional study to evaluate their pathogenic significance, we considered all the mutations found in our population in the last decade (11 and present work) in terms of severity and location. All the selected mutations were mapped in the whole GCK structure: p.His137Asp, pPhe150Tyr, p.Val154Leu, p.Gly162Asp, in the small domain, Arg303Trp and p.Arg392Ser, in the large domain and p.Glu70Asp in a loop connecting the two domains. Three among these mutations, (p.Gly162Asp, p.His137Asp and p.Arg392Ser) were the most severe and none have been produced or characterized. Figure 2 shows the position of selected mutated residues in the structure model of the GCK enzyme. All variants displayed reduced enzyme activity versus the wild-type GST-GCK, as shown by the I_a index. The kinetic characteristics of GST-GCK-wt and GST-GCK-mutants, determined *in vitro* enzymatic assays, are shown in Table 2. Mutations p.Gly162Asp, p.Phe150Tyr and p.Val154Leu, localized in the core of the molecule, not far from the substrate binding site, produced the strongest effect. The p.Gly162Asp change is very deleterious since no enzymatic activity was detected in the mutant protein, and mutations p.Phe150Tyr and p.Val154Leu reduced enzyme activity to below 10% of the wild-type enzyme. In contrast, mutations p.Glu70Asp, p.His137Asp, p.Arg303Trp and p.Arg392Ser, localized in more peripheral positions, retained at least 25% of wild-type activity. We also evaluated the protein stability of wild-type and of mutant GST-GCK and the time-course of thermal inactivation at different temperatures (Figure 3A and B respectively). Wild-type GCK activity remained practically constant under temperatures up to 50°C, but fell abruptly at 55°C. In contrast, the enzyme activity of all mutants, except GST-GCK (p.Glu70Asp), rapidly decreased by more than 50% at 50°C (Figure 3A). The time-course analysis of GCK thermal inactivation indicated that the mutants rendered the enzyme thermo-unstable (more than 50% of GCK activity was lost within 30 min at 50°C), whereas 50% of wild-type GCK activity was present after 30 min of incubation at the critical temperature (Figure 3B).

Discussion

A correct diagnostic approach to the MODY2 patient is important to avoid useless, expensive analyses and misclassification of the diabetes. In this study, we characterized the *GCK* gene by DHPLC followed by sequence analysis in 66 patients with suspected MODY2 enrolled between 2006 and 2010. We identified 19 GCK mutations in 28 children accounting for 42.4% of suspected MODY2 children. Our data are similar to those we obtained between 2001 and 2005 [11] and are also in agreement with the high prevalence of MODY2 (up to 61% of MODY forms) found in Italy and in southern Europe [12,13]. Among the 19 GCK variants described herein, six are new (p.Lys458_Cys461del, p.Val154Leu, p.Arg447Glyfs, IVS5-2A>T, p.Glu40Asp and p.Glu395_Arg397del) and 13 previously reported. Glucokinase missense mutations are the most frequent causes of MODY2 [9]. In our study, the missense mutations (13/19: two new and 11 known), and deletions (n = 3) were distributed throughout the protein: six/16 (38%) in the small domain, nine/16 (56%) in the large domain and one/16 (6%) in the connecting region of the protein. These findings are in agreement with a study in which no hot spot mutations were reported [9].

In the attempt to understand how the detected mutations could contribute to MODY2 insurgence, we searched for structure/function correlations of the disease-causing mutated proteins. First, we evaluated the impact of each mutation (except splice variants and deletions) on the enzyme 3D-structure and then, for seven selected mutations, we carried out functional studies. All the mutations considered here are buried in the enzyme, except for p.Glu70Asp that is located on the surface of the protein, and all mutations are far from the active and ATP binding sites (Table 1). Apart from p.Val154Leu and p.Phe150Tyr that undergo large movements during the conformational transition from the super-open to the closed active form, all the mutant structures exhibit local structural alterations (Table 1) that well correlate to kinetic parameters and thermal inactivation data (Table 2, Figure 3). In all cases, the amino-acid replacement either provokes the loss of stabilizing interactions or generates unfavorable interactions that may destabilize the enzyme.

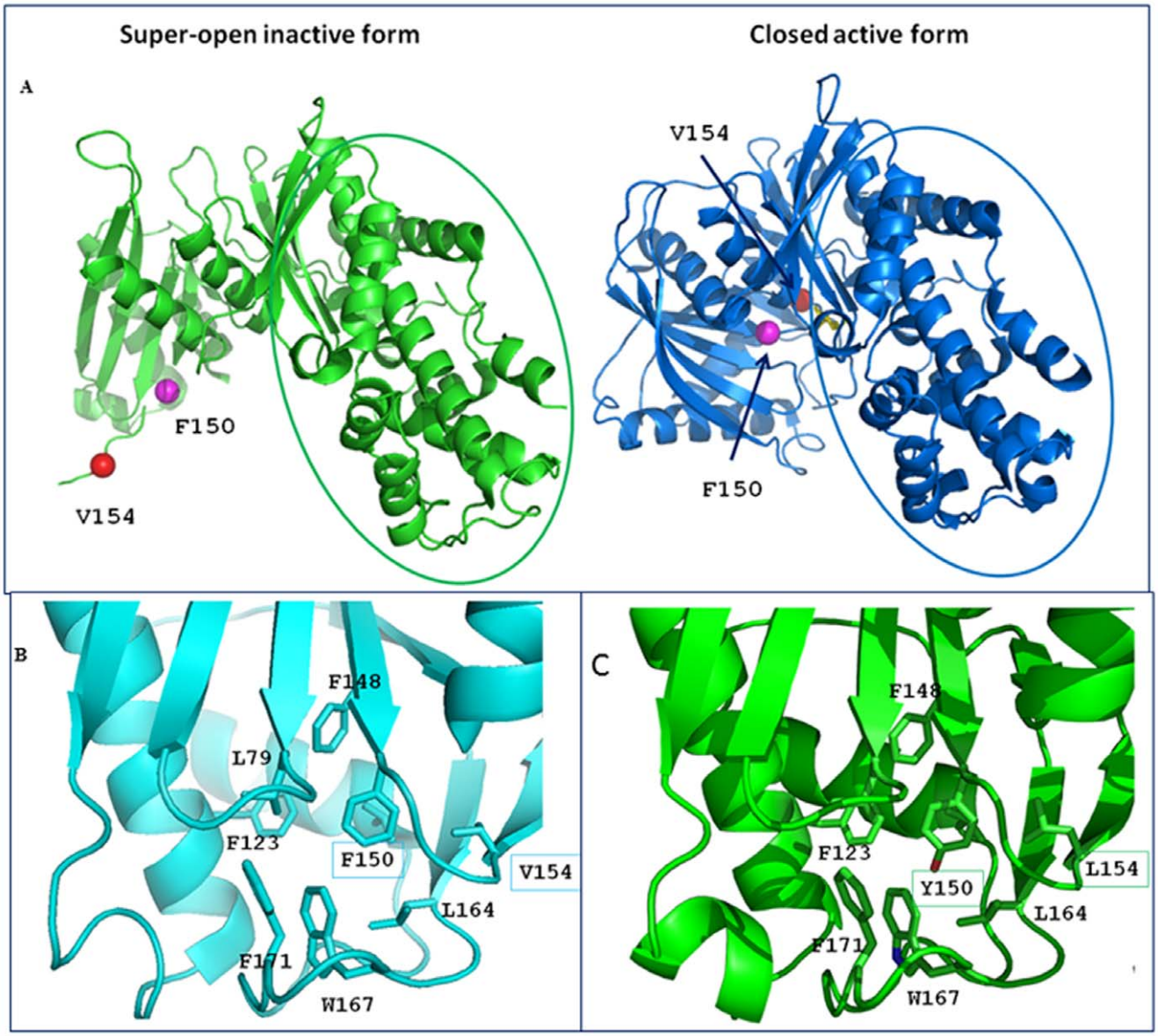

Figure 1. View of the mutations p.Phe150Tyr and p.Val154Leu in the whole structures and in their local environment. A: The structures of GCK in the super-open inactive form (PDB code code: 1v4t) and in the closed active form (PDB code code: 1v4s) are shown on the left and on the right, respectively. In the closed form (right) the sugar is shown as a yellow stick. In both panels, the large domain is in the same orientation and is circled. It is clear that the small domain undergoes a large conformational variation from the super-open to the closed form. Specifically the region embodying the residues Phe150 (blue sphere) and Val154 (red sphere) dramatically changes its orientation. **B**: Close-up view of the wild-type closed structure showing the hydrophobic core rich in aromatic/hydrophobic amino acids. Leu79, Phe123, Phe148, Phe150, Val154, Leu164, Trp167 and Phe171 form an intricate network of stabilizing hydrophobic interactions. **C**: Structure of GCK closed structure containing the both the Tyr150 and Leu154 mutated residues. Introduction of the oxydryl group (red) of Tyr150 within the hydrophobic core disrupts the interactions present in the wt-enzyme. The replacement of Val154 by leucine produces only small changes in the closed form.
doi:10.1371/journal.pone.0038906.g001

We evaluated the effects on GCK kinetics and thermo-stability of seven GCK variants distributed in different functional domains of the enzyme: three variants described in the present study (p.Phe150Tyr, p.Val154Leu, p.Arg303Trp) and four (p.Glu70Asp, p.His137Asp, p.Gly162Asp and p.Arg392Ser) previously reported [11]. The GST-GCK (p.Glu70Asp) mutant showed a GCK activity index significantly lower than the wild type (mean $I_a = 0.25$). This effect is due to discrete defects in the kinetic constants, in particular, glucose affinity was significantly reduced ($S_{0.5}$ for glucose: 20.1 ± 3.2 vs 7.2 ± 0.4 mM). Variant p.Glu70Lys displayed a similarly decreased activity index (mean $I_a = 0.31$) without significant thermo-instability [14,15]. The substitution of histidine with aspartic acid (p.His137Asp) affected GCK function (mean $I_a = 0.26$). As shown by molecular modelling, this mutation introduces a negative charge in the region that may affect the charge distribution of the protein [11]. The previously reported mutation p.His137Arg, in which His is replaced by the positively charged arginine, did not affect enzyme activity (mean $I_a = 0.92$) [15]. This finding shows that a negatively charged amino acid is not tolerated at this site. p.Phe150Tyr and the novel p.Val154Leu mutations dramatically reduced GCK activity (mean $I_a = 0.014$ and 0.099, respectively)

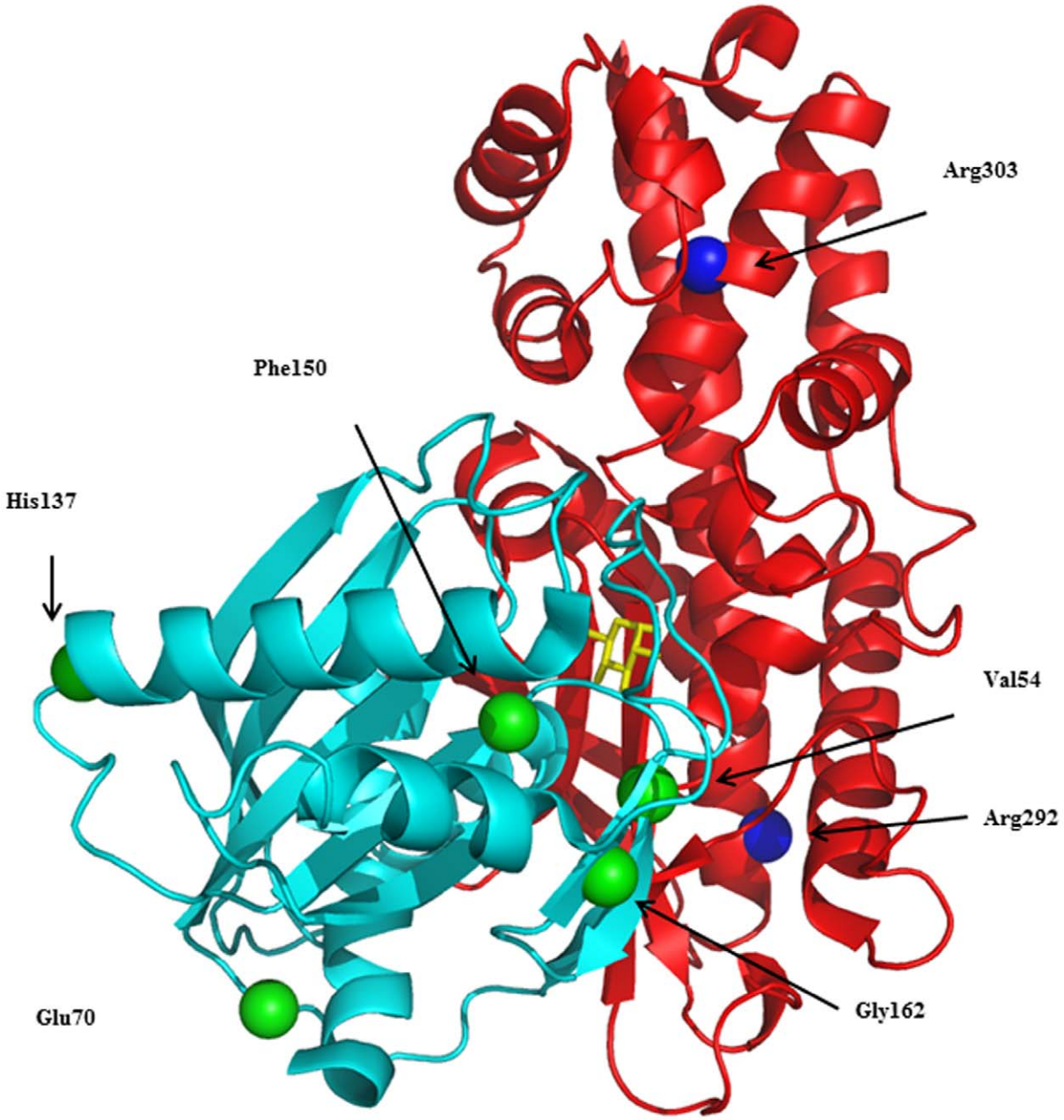

Figure 2. Distribution of the selected GCK mutations. The structure of GCK in the active closed form is shown (PDB code: 1v4s); the small and large domains are drawn in cyan and red, respectively; glucose appears as a yellow stick. Mutation sites are shown as green or blue spheres.
doi:10.1371/journal.pone.0038906.g002

and thermo-stability (<40% at 50°C/30 min than in wild-type). Both mutations displayed high $S_{0.5}$ and ATP Km values, which indicate a greatly reduced affinity for glucose and ATP. In particular, the cooperativity of p.Phe150Tyr was significantly reduced (nH = 1.06 vs 1.40, *t* test, $p = 0.038$). It is noteworthy that Val154 is one of residues that undergo large movements during the transition from the super-open to the closed form. Even though the replacement of Val by Leu provokes only small local perturbations, the latter may be relevant during the conformational transition induced by ATP and glucose binding. The substitution of glycine 162 by aspartic acid (p.Gly162Asp) affected the hydrophobic core of the enzyme and, as predicted by molecular modelling [11], markedly alters the structure and dynamics of the domain. This is one of the most dramatic mutations we identified as it introduces a net negative charge into a hydrophobic environment [11]. In fact, the enzyme activity of this mutant was undetectable. Although any reduction in the I_a below 30% of the wild type would have little further effect on the fasting plasma glucose [15], it is noteworthy that fasting glycemia was higher in the child with this mutation (10.0 mmol/L) [11] than in other MODY2 children (mean 6.7±0.8 mmol/L). Arg 303 is located in the α8 helix and molecular modelling indicates that the Arg-to-Trp change at this position would destabilize the helix structure. Mutation p.Arg303Trp caused a reduction of the activity index (mean I_a = 0.59), mainly due to a decrease in the catalytic constant, being

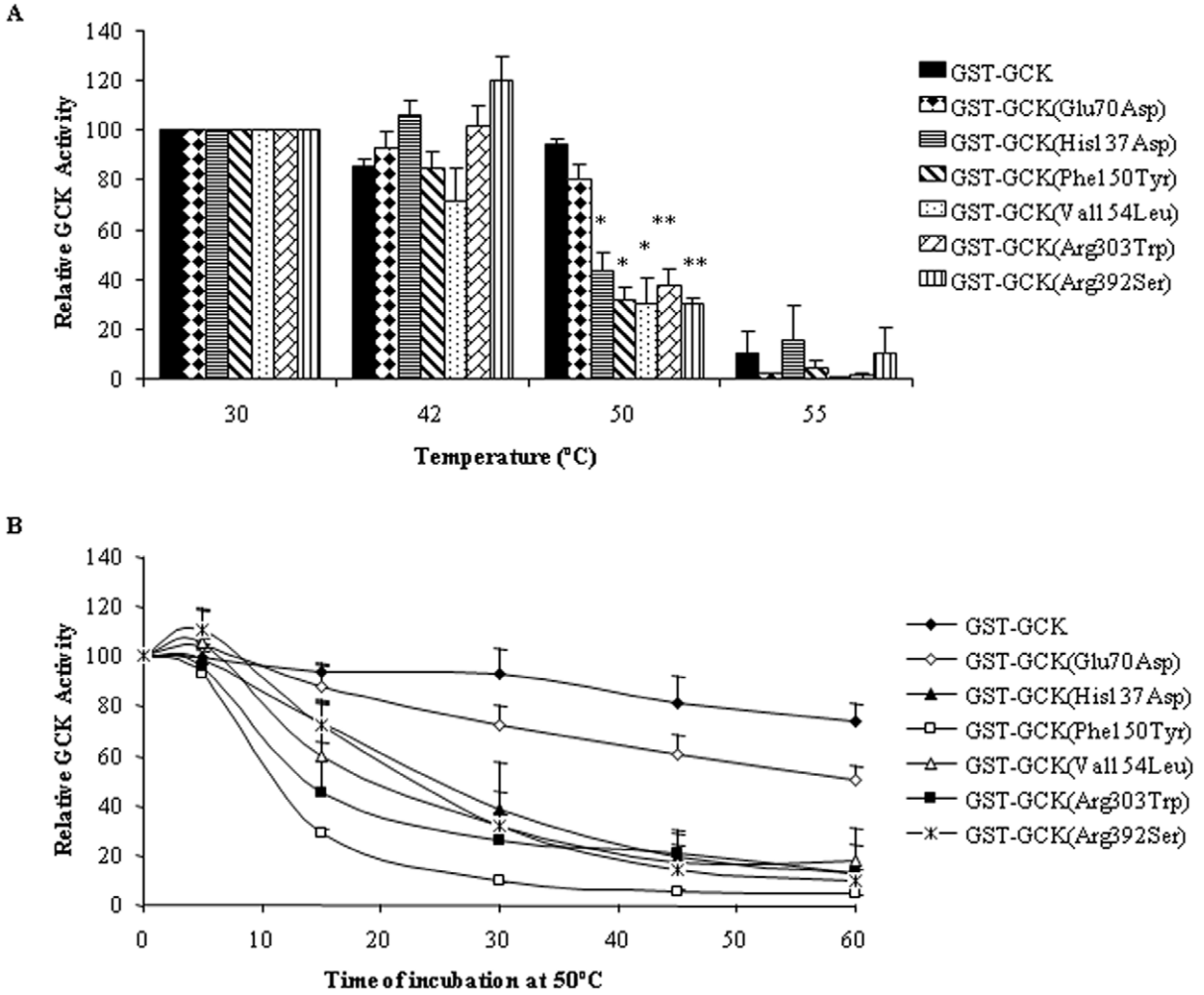

Figure 3. Effect of temperature on the stability of GST-GCK mutants. Stock enzyme solutions were diluted to 250 μg/ml in storage buffer containing 30% glycerol, 50 mM glucose, 10 mM glutathione, 5 mM DTT, 200 mM KCl and 50 mMTris/HCl, pH 8.0. Panel A: The enzyme solutions were incubated for 30 min at temperatures ranging from 30 to 55°C and then assayed at 30°C as described in the Methods section. Panel B: The enzyme solutions were incubated for periods of time from 5 to 60 min at 50°C. Results are means and SEM of three independent enzyme preparations for each mutant except for GST-GCK (Phe150Tyr) which corresponds to two independent enzyme preparations. (*) $p \leq 0.03$, (†) $p < 0.008$.
doi:10.1371/journal.pone.0038906.g003

both glucose and ATP affinities modified. Our results concord with the finding that other mutations in the same helix (p.Leu309Pro and p.Arg308Trp) increased thermal inactivation and/or modified glucose affinity [16,17]. Taken together, these results suggest that the α8 helix plays a role in the regulation of substrate affinity and protein stability. Mutation p.Arg392Ser is located in the protein periphery and, like the neighboring mutation p.Arg397Leu [17], it only slightly affected GCK activity. Thus, our mutants reduced the enzyme's catalytic activity by altering one or more kinetic parameters. Moreover, all mutants, except GCK-Glu70Asp, displayed thermal-instability, which has been implicated in hyperglycemia in MODY2 patients [17,18]. Nevertheless, additional defects caused by these mutations on other mechanisms of GCK control, such as post-translational regulation, interaction with other protein partners or organelles, cannot be excluded. Globally our evaluation of enzyme activity indicates that all seven mutants play a pathogenic role in MODY2 insurgence. In addition although altered glucose and ATP binding, and thermal stability appear to be the major causes of the disease, in a few cases mutations affected cooperativity and molecular motions, and hence impaired enzyme activity.

Although the in vitro functional evaluation of a GCK-mutant is a useful method with which to predict the effect exerted by a DNA change on enzyme activity, it is not a practical approach to the diagnosis of MODY. In our experience, taking into account all the experimental data concerning the seven mutants expressed, we found that the mutations induced structural alterations predicted by modeling that were in good agreement with kinetics/thermostability analyses. Therefore, this approach could be a reliable surrogate to predict the pathogenic role of a GCK variant.

In conclusion, in this five-year update of GCK mutations in MODY2 children from southern Italy, we have identified six new GCK variants thereby expanding the MODY2 mutation panel. Furthermore, our study, carried out by integrating DHPLC, sequencing, bioinformatics and functional analysis, provides new information about the structure-function relationship of human glucokinase mutations and MODY2.

Materials and Methods

Ethics Approval

The study was conducted according to the Helsinki II declaration and it was approved by the Ethics Committee of the School of Medicine Federico II, Naples, Italy.

Written informed consent to the study was obtained from each adult subject and from both parents of children.

Subjects

Between 2006 and 2010, 720 hyperglycemic children were monitored at the Departments of Pediatrics of the University of Naples "Federico II" and of the Second University of Naples, Italy. Sixty-six patients (mean age 109 months, 53% girls) who had fasting hyperglycemia (>5.5 mmol/L), HbA1c <7.0% and a family history of diabetes for at least two consecutive generations were enrolled in the study as suspected MODY2 individuals. The autoimmune markers of type-1 diabetes (glutamate decarboxylase, protein tyrosine phosphatase-like protein and insulin antibodies) were evaluated; the presence of more than one antibody was considered an exclusion criterion. One hundred unrelated euglycemic controls (mean age 363 months, 69% girls) were recruited at CEINGE (Advanced Biotechnologies, s.c.a.r.l. Naples) also between 2006–2010.

Patients, their mother and father, and controls provided 2 blood samples for biochemical testing and for DNA extraction. BMIz-score (Center for Disease Control normative, http://www.cdc.gov), family history of diabetes and/or other diseases, birth weight and age at admission were recorded for each patient. FPG, triglycerides (evaluated with routine methods) and HbA1c measure with HPLC (HLC-723 G7 TOSOH Bioscience Tokyo, Japan) were evaluated in each suspected MODY2 child. Genomic DNA from patients, their mother and father, and controls was extracted with the Nucleon BACC 2 kit (Amersham Biosciences Europe, Milan, Italy).

GCK Gene Analysis

The 10 exons, including their flanking intronic regions, of the *GCK* gene were amplified using primers and PCR conditions described elsewhere [11]. The amplified fragments were denatured at 95°C for 10 min and then renatured for 10 min to generate heteroduplices. Each fragment was run on the DHPLC WAVE DNA fragment analysis system (Transgenomic, Inc. Omaha, NE) using DHPLC conditions reported in Table S1. Any additional or abnormal peak shape observed in the chromatogram was further sequenced (ABI PRISM 3730; Biosystems Foster City, CA, USA). All sequences were analysed in comparison with the wild-type reference sequence (NM_000162, http://www.ncbi.nlm.nih.gov/nuccore/NM_000162) by the ABI Seqscape software v.2.5 (Applied Biosystems). The mutations and variants were then numbered according to the Human Genome Variation Society (http://www.hgvs.org/).

Analysis of *GCK* Mutations

We first evaluated each detected GCK mutation by bioinformatics (Polyphen and SIFT algorithms and Modeller programs). We also examined the enzyme function of seven mutations among the most severe found in our population in the last decade and never characterized. The mutations for the latter study mapped in the whole enzyme structure (small, large and loops connecting domains).

Bioinformatic Analysis

Two free online prediction programs, Polyphen (http://genetics.bwh.harvard.edu/pph/) and SIFT (http://sift.jcvi.org/www/SIFT_seq_submit2.html), [1] were used to predict the effect of GCK mutations on the protein. The Polyphen program [19] is a tool for the prediction of the impact of an amino acid substitution on the structure of a human protein using straight forward physical and comparative considerations. The SIFT program [20] tests a query sequence and uses multiple alignment information to predict tolerated and deleterious substitutions for every position of the query sequence. SIFT is a multistep procedure that searches for similar sequences, chooses closely related sequences that may share a function similar to that of the query sequence, obtains the alignment of these chosen sequences, and calculates normalized probabilities for all possible substitutions from the alignment. Models of all mutants of GCK were generated *in silico* with the Modeller 9v8 program using the active form of the GCK crystal structure as template [2.3 Å, Brookhaven Protein Database code (PDB code) 1v4s]. The inactive super-open form (3.4 Å, PDB code 1v4t) was also considered for comparison.

After alignment with using BODIL [21], the query and template sequences were used as input in the Modeller program [22] and 20 models were generated. The most abundant conformers of the replaced residue were selected and a simulated annealing procedure was carried out to optimize side chain conformations. The model that presented the best Modeller Discrete Optimized Protein Energy (DOPE) score was selected to be used in subsequent analyses. The initial models of selected mutants were energy minimized with the GROningen MAchine for Chemical Simulations (GROMACS) package using the GROningen MOlecular Simulation (GROMOS) 43a1 force field, to regularize the protein structure geometry. The molecules were solvated in boxes containing simple point charge water molecules. The energy minimization was obtained by 1000 steps of the steepest descent method and subsequently by 1000 steps of the conjugate gradient method. The notation used for secondary structure was taken from Kamata et al. [8].

Production of Recombinant Wild-type and Mutant Glutathionyl-S-transferase-glucokinase (GST-GCK)

Recombinant human wild-type beta cell GCK fused to GST (GST-GCK) was prepared as described previously [23]. Mutations were introduced into the GST-GCK construct by PCR using the QuikChange® II Site Directed Mutagenesis Kit (Stratagene, La Jolla, CA, USA); the oligonucleotides are reported in Table S2. Mutant constructs were checked by sequencing and digestion with specific restriction enzymes (Table S2). Fusion proteins from *E. coli* were expressed and purified as described previously [24]. Fusion proteins were stored at a concentration of about 1 mg/ml at –80°C in 30% glycerol, 50 mM glucose, 10 mM glutathione, 5 mM DTT, 200 mM KCl and 50 mM Tris/HCl, pH = 8.0. Protein concentrations were determined with the Bio-Rad Protein Assay (Bio-Rad Laboratories, GmbH München Germany), and bovine serum albumin as standard. GST-GCK purification was verified by Coomassie-stained SDS-PAGE as a single band of 75 kDa.

Enzymatic Assay

GCK activity was measured spectrophotometrically on a UVIKONxl spectrophotometer (Secoman, France), using a $NADP^+$-coupled assay with glucose-6-phosphate dehydrogenase. Determination of kinetic parameters and thermal inactivation were performed as previously described [23]. The Km for ATP was measured at a glucose concentration near the relative $S_{0.5}$

values (7, 15, 8.5, 100, 20, 5 and 11 mM respectively for wild type, p.Glu70Asp, p.His137Asp, p.Phe150Tyr, p.Val154Leu, p.Gly162Asp, p.Arg303Trp and p.Arg392Ser). Thermal inactivation experiments were assayed at a glucose concentration of 100 mM. Results are reported as means ± standard error of the mean (SEM) of three independent enzyme preparations assayed at least in duplicate.

Statistical Analysis

Variables are reported as mean±SD (continuous variables) or mean ±SEM (categorical variables). We used the Student *t* test to compare variables; significance was set at $p<0.05$. The SPSS statistical software was used.

Supporting Information

Figure S1 Close-up view of the p.Lys420Glu mutation at the inter-domain interface. The small and large domains are drawn in cyan and red, respectively. Helix 13 is shown in orange. Lys420 (red stick) forms a salt-bridge with Glu440 (yellow stick) which is located in a loop connecting the two domains.
(DOCX)

Table S1 DHPLC conditions to detect GCK variants.
(DOCX)

Table S2 Primers used for site-directed mutagenesis in GCK cDNA.
(DOCX)

Acknowledgments

We thank our diabetic patients and their families for their kind cooperation in this study.

Writing assistance: Jean Ann Gilder (Scientific Communication srl) provided writing assistance.

Provenance and peer review: Not commissioned; externally peer reviewed.

Author Contributions

Conceived and designed the experiments: LS. Performed the experiments: MC CMGH. Analyzed the data: LS MC CMGH NT MAN AZ. Contributed reagents/materials/analysis tools: DI AF CC IC VC AC. Wrote the paper: LS MC NT AZ MAN.

References

1. Ellard S, Bellanné-Chantelot C, Hattersley AT, European Molecular Genetics Quality Network (EMQN) MODY group (2008) Best practice guidelines for the molecular genetic diagnosis of maturity-onset diabetes of the young. Diabetologia 51: 546–553.
2. Incani M, Cambuli VM, Cavalot F, Congiu T, Paderi M, et al. (2010) Clinical application of best practice guidelines for the genetic diagnosis of MODY2 and MODY3. Diabet Med 27: 1331–1333.
3. Naylor R and Philipson LH (2011) Who should have genetic testing for MODY? Clin Endocrinol (Oxf) 75: 422–426.
4. Nyunt O, Wu JY, McGown IN, Harris M, Huynh T, et al. (2009) Investigating maturity onset diabetes of the young. Clin Biochem Rev 30: 67–74.
5. Iynedjian PB (2009) Molecular physiology of mammalian glucokinase. Cell Mol Life Sci 66: 27–42.
6. Massa ML, Gagliardino JJ, Francini F (2011) Liver glucokinase: An overview on the regulatory mechanisms of its activity. IUBMB Life 63: 1–6.
7. Hussain K (2010) Mutations in pancreatic ß-cell Glucokinase as a cause of hyperinsulinaemic hypoglycaemia and neonatal diabetes mellitus. Rev Endocr Metab. Disord 11: 179–183.
8. Kamata K, Mitsuya M, Nishimura T, Eiki J, Nagata Y (2004) Structural Basis for Allosteric Regulation of the Monomeric Allosteric Enzyme Human Glucokinase. Structure 12: 429–438.
9. Osbak KK, Colclough K, Saint-Martin C, Beer NL, Bellanné-Chantelot C, et al. (2009) Update on mutations in glucokinase (GCK), which cause maturity-onset diabetes of the young, permanent neonatal diabetes, and hyperinsulinemic hypoglycemia. Hum Mutat 30: 1512–1526.
10. Colom C, Corcoy R (2010) Maturity onset diabetes of the young and pregnancy. Best Pract Res Clin Endocrinol Metab 24: 605–615.
11. Tinto N, Zagari A, Capuano M, De Simone A, Capobianco V, et al. (2008) Glucokinase gene mutations: structural and genotype-phenotype analyses in MODY children from South Italy. PLoS One 3 e1870.
12. Mantovani V, Salardi S, Cerreta V, Bastia D, Cenci M, et al. (2003) Identification of eight novel glucokinase mutations in Italian children with maturity-onset diabetes of the young. Hum Mutat 22: 338.
13. Pruhova S, Dusatkova P, Sumnik Z, Kolouskova S, Pedersen O, et al. (2010) Glucokinase diabetes in 103 families from a country-based study in the Czech Republic: geographically restricted distribution of two prevalent GCK mutations. Pediatr Diabetes 11: 529–535.
14. Burke CV, Buettger CW, Davis EA, McClane SJ, Matschinsky FM, et al. (1999) Cell-biological assessment of human glucokinase mutants causing maturity-onset diabetes of the young type 2 (MODY-2) or glucokinase-linked hyperinsulinaemia (GK-HI).Biochem J 342: 345–352.
15. Davis EA, Cuesta-Muñoz A, Raoul M, Buettger C, Sweet I, et al. (1999) Mutants of glucokinase cause hypoglycaemia- and hyperglycaemia syndromes and their analysis illuminates fundamental quantitative concepts of glucose homeostasis. Diabetologia 42: 1175–1186.
16. Gidh-Jain M, Takeda J, Xu LZ, Lange AJ, Vionnet N, et al. (1993) Glucokinase mutations associated with non-insulin-dependent (type 2) diabetes mellitus have decreased enzymatic activity: implications for structure/function relationships. Proc Natl Acad Sci U S A 90: 1932–1936.
17. García-Herrero CM, Galán M, Vincent O, Flández B, Gargallo M, et al. (2007) Functional analysis of human glucokinase gene mutations causing MODY2: exploring the regulatory mechanisms of glucokinase activity. Diabetologia 50: 325–333.
18. Kesavan P, Wang L, Davis E, Cuesta A, Sweet I, et al. (1997) Structural instability of mutant beta-cell glucokinase: implications for the molecular pathogenesis of maturity-onset diabetes of the young (type-2). Biochem J 322: 57–63.
19. Zou M, Baitei EY, Alzahrani AS, Parhar RS, Al-Mohanna FA, et al. (2011) Mutation prediction by PolyPhen or functional assay, a detailed comparison of CYP27B1 missense mutations. Endocrine 40: 14–20.
20. Ng PC, Henikoff S (2003) SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res 31: 3812–3814.
21. Lehtonen JV, Still DJ, Rantanen VV, Ekholm J, Björklund D, et al. (2004) BODIL: a molecular modeling environment for structure-function analysis and drug design. J Comput Aided Mol Des 18: 401–419.
22. Eswar N, Marti-Renom MA, Webb B, Madhusudhan MS, Eramian D, et al. (2006) Comparative Protein Structure Modeling With MODELLER. Current Protocols in Bioinformatics, John Wiley & Sons Inc. Supplement 15 5.6.1–5.6.30.
23. Galán M, Vincent O, Roncero I, Azriel S, Boix-Pallares P, et al. (2006) Effects of novel maturity-onset diabetes of the young (MODY)-associated mutations on glucokinase activity and protein stability. Biochem J 393: 389–396.
24. Liang Y, Kesavan P, Wang LQ, Niswender K, Tanizawa Y, et al. (1995) Variable effects of maturity-onset-diabetes-of-youth (MODY)-associated glucokinase mutations on substrate interactions and stability of the enzyme. Biochem J 309: 167–173.
25. Weng J, Ekelund M, Lehto M, Li H, Ekberg G, et al. (2002) Screening for MODY mutations, GAD antibodies, and type 1 diabetes-associated HLA genotypes in women with gestational diabetes mellitus. Diabetes Care 25: 68–71.
26. Stoffel M, Froguel PH, Takeda J, Zouali H, Vionnet N, et al. (1992) Human glucokinase gene: isolation, characterization, and identification of two missense mutations linked to early-onset non-insulin-dependent (type 2) diabetes mellitus. Proc Natl Acad Sci U S A 89: 7698–7702.
27. Lorini R, Klersy C, d'Annunzio G, Massa O, Minuto N, et al. (2009) Maturity-onset diabetes of the young in children with incidental hyperglycemia: a multicenter Italian study of 172 families. Diabetes Care 32: 1864–1866.
28. Matyka KA, Beards F, Appleton M, Ellard S, Hattersley A, et al. (1998) Genetic testing for maturity onset diabetes of the young in childhood hyperglycaemia. Arch Dis Child 78: 552–554.
29. Estalella I, Rica I, Perez de Nanclares G, Bilbao JR, Vazquez JA, et al. (2007) Mutations in GCK and HNF-1alpha explain the majority of cases with clinical diagnosis of MODY in Spain. Clin Endocrinol Oxf 67: 538–546.
30. Shimada F, Makino H, Hashimoto N, Taira M, Seino S, et al. (1993) Type 2 (non-insulin-dependent) diabetes mellitus associated with a mutation of the glucokinase gene in a Japanese family. Diabetologia 36: 433–437.
31. Massa O, Meschi F, Cuesta-Munoz A, Caumo A, Cerutti F, et al. (2001) High prevalence of glucokinase mutations in Italian children with MODY. Influence on glucose tolerance, first-phase insulin response, insulin sensitivity and BMI. Diabetologia 44: 898–905.