The structure of full-length human phenylalanine hydroxylase in complex with tetrahydrobiopterin

Short title: Structure of BH_4-bound phenylalanine hydroxylase

Marte Innselset Flydal[a,1], Martín Alcorlo-Pagés[b,1], Fredrik Gullaksen Johannessen[a], Siseth Martínez-Caballero[a], Lars Skjærven[a], Rafael Fernandez-Leiro[c], Aurora Martinez[a,2] and Juan A. Hermoso[b,2]

[a]Department of Biomedicine, University of Bergen, 5009 Bergen, Norway

[b]Department of Crystallography and Structural Biology, Instituto de Química-Física "Rocasolano", Consejo Superior de Investigaciones Científicas, 28006 Madrid, Spain

[c]Spanish National Cancer Research Centre (CNIO), 28029 Madrid, Spain.

Classification: Biological Science (major), biochemistry (minor)

M.I.F. prepared PAH proteins, performed DSC and worked in crystallization and data collection, M.A.P worked with crystallization and data collection and carried out crystallographic determinations and cryo-electron microscopy. F.G.J performed activity experiments, S.M.C. worked in cryo-electron microscopy, L.S. performed the MD simulations and structural modeling, R.F.L. worked in cryo-EM determination. M.I.F., J.A.H. and A.M. designed research and wrote the paper. All authors edited the manuscript.

The authors declare no conflict of interest.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.wwpdb.org (PDB ID codes 6HYC and 6HPO). The EM maps have been deposited in the EMDB with accession s EMD-4605

[1]M.I.F. and M.A.P. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: xjuan@iqfr.csic.es or aurora.martinez@uib.no

This article contains supporting information.

ORCID IDs for authors

Marte I. Flydal	0000-0002-4070-8367
Martín Alcorlo-Pagés	0000-0001-8337-6203
Fredrik G. Johannessen	0000-0003-2739-9771
Siseth Martínez-Caballero	000-0003-4842-8447
Lars Skjærven	0000-0002-8176-4891
Rafael Fernandez-Leiro	0000-0002-7941-0357
Aurora Martinez	0000-0003-1643-6506
Juan A. Hermoso	0000-0002-1862-8950

Abstract

Phenylalanine hydroxylase (PAH) is a key enzyme in the catabolism of phenylalanine, and mutations in this enzyme cause phenylketonuria (PKU), a genetic disorder that leads to brain damage and mental retardation if untreated. Some patients benefit from supplementation with a synthetic formulation of the cofactor tetrahydrobiopterin (BH_4) that partly acts as a pharmacological chaperone. Here we present the first structures of full-length human PAH (hPAH) both unbound and complexed with BH_4 in the pre-catalytic state. Crystal structures, solved at 3.18Å resolution, show the interactions between the cofactor and PAH, explaining the negative regulation exerted by BH_4. BH_4 forms several H-bonds with the N-terminal autoregulatory tail but is far from the catalytic Fe^{II}. Upon BH_4 binding a polar and salt-bridge interaction network links the three PAH domains, explaining the stability conferred by BH_4. Importantly, BH_4 binding modulates the interaction between subunits, providing information about PAH allostery. Moreover, we also show that the cryo-EM structure of hPAH in absence of BH_4 reveals a highly dynamic conformation for the tetramers. Structural analyses of the hPAH:BH_4 subunits revealed that the substrate-induced movement of Tyr138 into the active site could be coupled to the displacement of BH_4 from the pre-catalytic towards the active conformation, a molecular mechanism that was supported by site directed mutagenesis and targeted MD simulations. Finally, comparison of the rat and human PAH structures show that hPAH is more dynamic, which is related to amino acid substitutions that enhance the flexibility of hPAH and may increase the susceptibility to PKU-associated mutations.

Keywords: human phenylalanine hydroxylase, phenylketonuria, X-ray crystallography, oligomerization, allosteric regulation

Significance Statement

The present crystal structure of phenylalanine hydroxylase (PAH) provides the first and long-awaited 3D-structure of the full-length human PAH, both unbound and complexed with the tetrahydrobiopterin (BH_4) cofactor.

The BH_4-bound state is physiologically relevant, keeping PAH stable and in a pre-catalytic state at low L-Phe concentration. Furthermore, a synthetic form of BH_4 (Kuvan®) is the only drug-based therapy for a subset of phenylketonuria patients.

We found two tetramer conformations in the same crystal, depending on the active site occupation by BH_4, which aid to understand the stabilization by BH_4 and the allosteric mechanisms in PAH. The structure also reveals the increased mobility of human- compared with rat PAH, in line with an increased predisposition to disease-associated mutations in human.

Introduction

Phenylalanine hydroxylase (PAH) catalyzes the tetrahydrobiopterin (BH_4)-dependent hydroxylation of L-phenylalanine (L-Phe) to L-tyrosine (L-Tyr), which is the rate-limiting step in the degradation of L-Phe and L-Tyr. In humans this reaction is vital for preventing the accumulation of neurotoxic amounts of L-Phe, which is the hallmark of the disease phenylketonuria (PKU) (1-3). Patients must follow a life-long protein-controlled diet to avoid severe neurological symptoms, but in a subset of the patients the diet can be alleviated by sapropterin dihydrochloride (Kuvan®), a synthetic form of the cofactor BH_4 (1) that shows multifactorial effects on PAH mutants, including their stabilization (4). As most of the mutations in *PAH* cause protein instability, leading to misfolding and loss-of-function (3), it is of medical interest to discover and optimize other compounds with pharmacological chaperone potential that can specifically prevent PAH misfolding and extend the range of mutations that can be rescued (5, 6). Structural information is essential in this endeavor.

Mammalian PAH is tetrameric and its 52 kDa subunits are composed of a regulatory domain (RD, residues 1-110) with an unstructured N-terminal tail (N-term; residues 1-29), a catalytic domain (CD, residues 111-410), and an oligomerization domain (OD, residues 411-452) responsible for dimerization and the subsequent tetramerization (Fig. S1). This oligomeric and multidomain arrangement allows the complex regulation of PAH activity, including *(i)* activation by the substrate L-Phe, which elicits a positive cooperative response on PAH ensuring that neurotoxic accumulation of L-Phe is avoided, and *(ii)* decreasing PAH activity at low substrate concentration to preserve a lower threshold value of L-Phe for protein synthesis (for reviews see (2, 7, 8)). At low, physiological concentration of L-Phe, PAH is complexed with the cofactor BH_4 (present in hepatocytes at equal concentrations as PAH subunits) that acts as a negative regulator of both the enzyme activity and activation by L-Phe (8, 9). BH_4 also stabilizes PAH and increases the steady-state levels of the enzyme *in vivo* (10, 11). While the activation by L-Phe induces a large conformational change (8), recently proposed to involve dimerization of RDs (12-14), BH_4 binding elicits a limited conformational change proposed to be mostly constrained to the N-term tail (15, 16). The importance of these N-terminal 29 residues in both the inhibitory regulation by BH_4 and the activation by L-Phe is evident, as PAH lacking this tail is not regulated by either BH_4 or L-Phe and is constitutively active (15).

Until now, the available mammalian PAH structures are truncated, lacking one or two of the domains (17-19) or full-length without ligands, the latter only of rat PAH (rPAH) (12, 14). Obtaining proper understanding of the structural defects associated to PKU mutations and development of specific stabilizing therapies would highly benefit from the availability of the structure of human PAH (hPAH), both with and without ligands. Thus, we aimed to solve the long-awaited full-length structures of hPAH both in the absence of ligands and with the cofactor BH_4 and have obtained two structures at 3.18 Å resolution from the same crystal, which present different occupancy by BH_4. The full-length structure in absence of BH_4 has been also obtained at medium resolution by cryoEM. These are the first structures of complete hPAH with all domains and the first structures of full-length mammalian PAH with ligand.

Results and discussion

Crystallization of hPAH with two tetramers in a single crystal. Both full-length hPAH and ΔN13-hPAH, the latter produced by cutting MBP-hPAH with Factor Xa protease (20, 21), were co-crystallized with BH_4 and produced crystals with similar appearance at the same conditions (*SI Materials and Methods*). However, data sets collected from hundreds of similar looking crystals showed both low resolution and severe anisotropy, suggesting that there was still heterogeneity in the conformation of tetrameric hPAH. Nevertheless, with ΔN13-hPAH and DTT we eventually obtained a crystal with a degree of anisotropy that allowed us to obtain the structural determination of hPAH at 3.18 Å resolution (Fig. 1, S2 and Table S1) (22). Comparison with the structure of the dimeric catalytic domain (hPAH-CD) – that we also solved at 1.67 Å resolution (Table S1) with previously published conditions (17) (Fig. S2B) – shows that the absence of Fe in our structure does not alter the arrangement of the iron-binding 2-His-1-carboxylate facial triad (Fig. 1C). Four independent monomers of PAH were found in the asymmetric unit belonging to two different tetramers, Tet1 and Tet2, each created as a dimer of dimers by a two-fold crystallographic axis. While Tet1 presents the BH_4 bound in two active sites (chains A, B) and the other two (chains C, D) without the cofactor, Tet2 shows BH_4 bound to all four active sites (Fig. 1). All four independent monomers in both tetramers present a similar overall structure for the backbone (RMSD ~0.6 Å for Cα superimposition), with hPAH arranged as a tetramer through the C-terminal oligomerization helix, following an approximate 222 symmetry between monomers, only the two-fold symmetry between dimers being strict (Fig.

1A). Each monomer presented electron density for all domains (residues 22-450) (Fig. 1A, S1). The tetramers can be roughly encased in a rectangle with a short edge along the dimer (~70 Å) and a long edge (~110 Å) between the two dimers, while differences in the long edges are observed between Tet1 and Tet2 (Fig. S3). The RD is essential for this specific oligomeric arrangement in hPAH as deduced by comparison with the human truncated form lacking RD (PDB 2PAH) (18) (Fig. S4).

The presence of unbound and BH_4-bound states – from here on referred to as apo and holo states, respectively – in the same crystal allowed a direct comparison between both conformations. Until now, investigations of how cofactor binding affects the conformation of the RD have come from modelling through superimposition of truncated structures (apo rPAH containing RD and CD, PDB 1PHZ) onto hPAH-CD with ligands (i.e. with BH_4, PDB 1J8U, and with both BH_4 and with substrate analog thienylalanine, THA, PDB 1MMK). These modelling analyses have predicted interactions of BH_4 with S23 at the flexible and unstructured N-term (residues 1-29) (15, 16, 23), a segment that has been termed an intrasteric autoregulatory sequence (IARS) (19). Comparing our apo and holo structures with each other and with the previous ligand-bound truncated forms reveals main events implicated in the negative regulation of PAH by BH_4 and derived enzyme stabilization, as well as in the conversion to the active BH_4-bound structure in the presence of substrate. As detailed below, the conformational events involve so far unidentified networks of residues, far beyond the IARS.

Structural effects of BH_4 binding associated to negative regulation of activation by L-Phe and enzyme stabilization. The BH_4 binding-site is flanked by the N-term (residues 21-32), the active-site lid (130-150), the Fe^{II}-coordinating residues, the β6-α7 loop (residues 245-251) and F254 (Fig. 1C). BH_4 is sandwiched between hydrophobic residues, F254 from one side and L248 and I25 from the other (Fig. 1C). Besides, some polar interactions are observed between BH_4 and residues from the IARS and the β6-α7 loop (Fig. 1C). The direct interaction of S23 in the N-term with BH_4, via H-bonds with both O1' and O2' in the dihydroxypropyl side chain, is observed for the first time in our BH_4-bound structure, and confirms previously predicted interactions with these hydroxyls in the natural (6*R*)-BH_4 configuration (15, 16, 23). These interactions appear related to the inhibitory effect of BH_4 on L-Phe activation of PAH, and explain why the cofactor

analogue 6-methyl-tetrahydropterin (6-MPH_4) is not inhibitory and (6S)-BH_4 is less inhibitory (24).

The activation of PAH by L-Phe requires a slow (s-to-min timescale with 1 mM L-Phe, depending on the temperature) global conformational change, and the activation rate is slower for the BH_4-preincubated than for the unbound enzyme (8-10, 23), associated to the negative regulation of PAH activation exerted by BH_4. Recent structural analyses have shown that full activation involves the shift and dimerization of the RDs, with concomitant movement of the N-term IARS that releases from the CD (12, 14). In apo hPAH, S23 H-bonds with Y377 in the CD, whereas in the holo structure S23 is also engaged in a larger H-bond network involving BH_4 and extends to residues in the CD and the RD (Fig. 1C). Structural comparison between apo- and holo-hPAH monomers reveals other BH_4-induced changes, beyond the N-term IARS (Fig. 2A). Surprisingly, the conformation of the N-term is similar for apo and holo hPAH (Fig. 2A), and also for apo rPAH (12), showing that binding of BH_4 does not induce large conformational changes on the N-term. Comparative analyses of apo- and holo-hPAH monomers by calculating differences in inter-residue distances between the apo and all holo monomers (*SI Materials and Methods*) revealed many distances that facilitate or abolish interactions upon BH_4 binding, most involving residues in the CD (Fig. S5 and Table S2). Out of the 35 largest distance differences, about half engage intra-CD residue pairs (19 pairs), most concerning residues around the BH_4 binding site (Table S2). These are followed by intra-RD residues (6 pairs), between RD and CD (5 pairs), between RD and OD (3 pairs), and intra-OD and between CD and OD (1 pair each). Thus, upon BH_4 binding changes are observed that link the CD with the OD, both directly and via the RD. Actually, the largest change (~6 Å) is the increased distance between K73 in the RD and L430 at the start of the tetramerization helix in the OD, indicating conformational changes affecting oligomerization and allostery (see below). In the active site, the most apparent change involves residues L248 and F254 in the BH_4 binding-site (Fig. 2A1 and Table S2). In the apo subunit, these two residues show a direct van der Waals interaction (~3 Å), observed for the first time, and partly occupy the BH_4 binding-site, whereas in the BH_4-bound active site they move to sandwich the cofactor (Fig. 2A). Interestingly, binding of BH_4 changes the interaction network between the N-term and the CD (Fig. 2A1). In the apo state residue D27 makes H-bond interactions with α7 residues (notably S251) while in the holo state makes a salt-bridge interaction with R252 and K113 instead. The salt-bridge interaction between E26 and K113

observed in apo state is lost in holo state (Fig. 2A1, S5). Interactions of BH_4 with the N-term and the CD, as well as the network of polar and salt-bridge interactions linking the RD (Y24, D27, K113) with the CD (D315, R252) (Fig. 1C), appear determinant for the regulatory inhibitory effect and the stabilization of BH_4. Differential scanning calorimetry (DSC) experiments at 200 °C/h with the BH_4 analog 7,8-dihydrobiopterin (BH_2) (Fig. S6A) support the stabilization effect of BH_4 beyond the CD. Hence, while apo-hPAH presents the expected unfolding transition temperatures at ~49 and 58 °C, previously associated to the RD and the CD, respectively (27), both transition temperatures increased ~2 °C in holo-hPAH (Fig. S6A).

Long-range structural effects of BH_4 binding. A different disposition of the monomers in Tet1 and Tet2 is observed depending on occupancy by BH_4 that changes the overall dimensions of the tetramer, the distances between monomers and the relative disposition (~5 Å gliding) of one dimer versus the other through the oligomerization helices (Fig. S3). Considering that all monomers presented a nearly identical backbone we reasoned that the observed differences should rely on different structural arrangements inside the dimers that would promote changes in the tetramer. Superposition of the holo monomers in apo-holo and holo-holo dimers shows a rotation of ~3º and displacements up to 3 Å between monomers. Thus, the changes observed inside the active site of a monomer upon BH_4 binding (Fig. 2A1) must also propagate to the dimeric partner (intra dimer interactions) and to the tetramer (inter dimer interactions). Structural analysis of Tet1 and Tet2 provides clues about such long-range effects of BH_4 binding (Fig. 2A). Changes are observed at the dimer interface between the apo-holo or holo-holo dimers (Fig. S7). It is worth to mention residue R297 that participates in an intra-dimer interaction network with residues E422, D415 (OD) and R71 (RD) only when BH_4 is bound (Fig. 2A2, 2A3, S7). Another relevant residue in this network is R413 that, in the apo monomer, can establish a salt-bridge interaction with the oligomerization helix (residue E440) of the dimeric partner (Fig. 2A2). This interaction is not seen between holo monomers (Fig. S7). As previously mentioned, the quaternary structure of hPAH is also affected by BH_4 binding, the most significative is a gliding of ~5 Å between dimers along the OD helices (Sup. movies 1 and 2) changing the network of polar interactions between the OD helices in the tetramer (Fig. 2A4, S8). Contrary to tyrosine hydroxylases or tryptophan hydroxylases the OD helices in mammalian PAH lack a proper Leu zipper and present abundance of polar residues inside the coiled-coil structure, as has already been noticed in the structure of rPAH (12). The lack of a classic leucine zipper explains why

PAH can dissociate to dimers as well as the conformational variability in the C-terminal helices.

To further investigate the dynamics of apo hPAH tetramer, i.e. with all four subunits in the apo state, we performed single particle cryo-electron microscopy (cryo-EM) studies on the same full-length construct but in the absence of cofactor (*SI Materials and methods* and Table S3). A consensus refinement of the hPAH tetramer yielded a cryo-EM map of resolution ranging from 5 to 7 Å due to the high mobility of the monomers within the tetramer (Fig. 2B). 3D classification allowed for the separation of different conformations in the dataset (Fig. S9), but these still harbored the same flexibility-related problems as the consensus refinement. In order to address these flexibility issues, we performed multibody refinement by masking individual dimers (Fig. 2B) (29). This procedure improved resolution to ~5 Å for most of the protein except for the RD (Fig. 2B) resulting in an excellent fit of the crystal structure onto the resulting maps (Fig. 2C). Moreover, variance analysis using principal component analysis showed that most of the orientational differences between dimers can be described by three main components (Fig. S10). Interestingly, the amplitude distribution along these components is monomodal, indicating that these are continuous movements in which dimers rotate around each other mainly as rigid bodies (Fig. S10C and Sup. movie 3). Thus, in absence of BH_4, hPAH dimers present a large range of motion inside the tetramer. While there are no noticeable changes in the internal structure of the CD, the RD seems to present extra flexibility.

In agreement with crystallographic and cryo-EM data, MD simulations (eight 100-ns long runs) of hPAH performed with and without BH_4 (*SI Materials and methods* and Fig. S11) show a highly dynamic tetrameric structure in both apo and holo states, with large-scale inter-dimer motions, but relatively small intra-dimer fluctuations. We calculated the fluctuation profile (root mean square fluctuation; RMSF) for each monomer in each of the simulations. Comparative analysis shows that BH_4 stabilizes significantly ($p < 0.05$) several regions in the CD, but increases flexibility in loop α14-β11 (residues 374-384) that contains Y377 residue (Fig. 1C, S12). This loop is important for embracing the substrate and moves about 5 Å closer to the active site upon substrate binding (30). Another loop located at the other side of the channel, i.e. the Y138 loop, termed the active site lid, also moves considerably (> 10 Å) towards the active site upon L-Phe binding (30) but, contrary to loop α14-β11, it loses mobility upon BH_4 binding (Fig. S12).

The significance of the PAH:BH_4 pre-catalytic complex and the role of Y138. In the holo hPAH subunits, BH_4 is bound placing its O4 atom, the site of hydroxylation of the pteridine ring during catalysis, at 3.8 Å from the catalytic iron (Fig. 3A), 0.7 Å longer than the required distance to form the Fe^{II}-OO-BH_4 adduct and generation of the Fe^{IV}=O hydroxylating intermediate (25, 30, 31). The cofactor bound hPAH structure presented in this work thus corresponds to a pre-catalytic conformation, which is of high significant physiological relevance, as PAH and BH_4 coexist in hepatocytes in stoichiometric amounts, forming a stable complex at normal low concentrations of L-Phe *in vivo* (≤ 54 μM) (10). BH_4-bound hPAH presents lower initial catalytic activity, and is activated by L-Phe at a slower rate compared with the ligand-free enzyme (8-10, 23) (Fig. S6B and see below). The physiological significance of the negative regulation of PAH activation by BH_4 is related to the maintenance of a low hydroxylating activity at low substrate concentrations to preserve a basal level of L-Phe required for protein synthesis (9, 10). Furthermore, the stabilization of PAH by BH_4 binding (Fig. S6A) contributes to maintaining the *in vivo* half-life of the enzyme, as supported by the decrease of steady state levels of PAH caused by BH_4 loss in cells and in liver (11, 32, 33). No atomic structural information is available on the large conformational changes associated to complete L-Phe activation, including the dimerization of the ACT domains (12, 14). Nevertheless, our structures provide novel insights into the substrate-induced conformational changes around the active site related to the transition from the pre-catalytic to the catalytic state upon entrance of the substrate at the active site. Superposition of holo-hPAH with the binary hPAH-CD:BH_4 and ternary hPAH-CD:BH_4:THA complexes reveals a possible role of Y138 in this transition by interacting with the O1' hydroxyl and contributing to the displacement of the cofactor 0.7 Å and 2.6 Å towards the iron and Glu286, respectively (Fig. 3A,B). In order to probe plausible transition pathways occurring upon substrate binding, we performed a 2 ns long targeted MD (TMD) simulation of Tet2 in complex with L-Phe and BH_4 (Sup. movie 4). With this simulation technique we gradually steer one catalytic domain of Tet2 to a target structure defined by the ternary hPAH-CD:BH_4:THA complex (PDB 1MMK). Of note, the binding site of L-Phe was solvent accessible and there was no clash of the bound substrate with residues in hPAH. At ~0.8 ns of the simulation Y138 has moved ~10 Å from its original position towards the active site and forms a transient H-bond with the carbonyl O of P279. After ~1.4 ns (Fig. 3C), a H-bond is formed between the side chain of T278 and the amino group of L-Phe, and the translation of

Y138 towards the active site is accelerated, and at the end of the simulation, a H-bond is formed between the phenol O of Y138 and the O1' of BH_4 in its pre-catalytic position (Fig. 3D). At that point the cofactor only maintains a single H-bond interaction with S23, which gets further weakened at the end of the simulation when S23 interacts with S251. To further investigate the role of the hydroxyl-group of Y138 in initiating the movement of BH_4 towards the catalytic binding mode, as well as the relevance of S23 and its interacting CD-residue Y377 (Figs. 1C, 3A-D), we prepared the hPAH mutants Y138F, S23A and Y377F, which were expressed and purified as tetramers, but at lower yield than WT. Other residues that interact with BH_4 in the pre-catalytic conformation (region 245-255, 286, 322 and 325), also interact with BH_4 in the catalytic conformation and, in addition, these sites are actually associated with severe destabilization of PAH (www.biopku.org), and thus not suitable for mutagenesis investigations. The initial reaction rates at 37 °C, comparative to the hPAH wild-type (WT), were measured for the four enzymes in 3 initial states, i.e. *L-Phe-activated* (preincubated with 1 mM L-Phe), *non-activated* (not preincubated with either substrate or cofactor) and *pre-catalytic BH_4-bound* (preincubated with 75 μM BH_4) (Figs. 3E, S6B). The concentration of substrate and cofactor were otherwise identical at the start of the reactions. Activated Y138F shows reduced activity, to 70% of WT, but more remarkable is the large reduction in activity for the non-activated and pre-catalytic states for this mutant (to 35-45% of WT). This reduction is not associated to different steady state kinetic parameters, as the Y138F mutant presents similar $K_m(BH_4)$, $S_{0.5}$(L-Phe) and hill coefficient (h) as WT-hPAH (Fig. S6C), supporting a role of the Y138 hydroxyl group in disengaging BH_4 from the pre-catalytic conformation in the full-length enzyme. The other two mutants S23A and Y377F were prepared to validate the effect of the H-bonding interactions centered at S23 to stabilize the enzyme-cofactor complex in the pre-catalytic state. Both mutants, and notably S23A, show reduced activity for the L-Phe-activated state and increased $S_{0.5}$(L-Phe) (Fig. S6C), indicating the participation of these residues in substrate activation, through the side-chain O in the case of S23. Moreover, the mutant S23A is expected to maintain the negative regulatory interactions with O1' and O2' of BH_4 through its backbone carbonyl and amide groups, explaining the low basal activity of the BH_4-bound state. On the other hand, as the function of Y377 appears related to its interaction with S23 for proper localization of this residue and the IARS through H-binding interactions (Fig. 3A-C), the rupture of the H-bond with S23 in the Y377F mutant results in an increased activity for the unactivated states (Fig. 3E, S6B).

Human and rat PAH present significant structural differences despite high-sequence homology. PAH for human and rat share a high sequence similarity (96%). Among the 452 residues of hPAH there are only 16 non-conservative- and 14 conservative mutations compared with rPAH (Fig. S13A). Conservative mutations are distributed among the three domains and non-conservative mutations are located in RD and CD (Fig. S13). Considering differences in size of RD and CD, mutations are strongly concentrated in the RD (8% of residues of RD are mutated between hPAH and rPAH *vs.* 2% of the CD) supporting differences in the regulatory properties of both enzymes (34).

Overall structures of the human and rat PAH monomers are very similar (RMSD values for the backbone ranging 0.8-1.1 Å) (Fig. 4A). Major differences are seen at the interface between RD and CD (loop β1-α1 (44-46)), in α2 (88-100) of the RD, residues 115 in the hinge region 110-118 (loop β4-β5), 143, 338 in the CD, and at the beginning of the OD helix (L427-K438) (Fig. 4, S14) that provoke differences in tetramer arrangement (Fig. S14). Regions with largest conformational changes in the RD are located around non-conservative mutations, i.e. C29 (S29 in rPAH), L88 (K88 in rPAH), K114 (E114 in rPAH), and D116 (N116 in rPAH) (Fig. 4B). In the RD/CD interface is worth to mention the salt-bridge in rPAH between E114 and K88 (K114 and L88 in hPAH), linking α2 with loop 110-118, that obviously is absent in hPAH, where a different interaction pattern is observed in this region. In the apo-state of hPAH salt-bridges are observed between D122 and both K114 and K85 and D116 and R99 (Fig. 4B). Consequently, both sides of the α2 helix in the RD, i.e. the C-terminal (K99) side in hPAH and the N-terminal (K85/K88) side in rPAH, are involved in interactions with the important loop 110-118, located at the end of the RD, at the boundary between the RD and the CD. Differences are also observed in the conformation of the OD helices, i.e. in rPAH the C-terminal helices of the four monomers adopt two distinct orientations with respect to the CD, differing by a tilt of ~10°, and with the monomers situated across the diagonal of the tetramer containing similarly positioned helices (12). Structural analysis reveals that the different conformations relate with a different pattern in polar interactions at the beginning of the OD helices depending on whether Q428 is making a H-bond with N73 (Fig. S14A) or inserted in the interface between the CD and the RD (Fig. S14B). In hPAH there are no such distinct orientations for the C-terminal OD helices (Fig. 4C,D) that are placed in between the two conformations observed in rPAH (Fig. 4A,B). In hPAH the

interaction network in this region is affected by the presence of the BH_4 cofactor in the active site. In the apo Q428 makes H-bond interactions with both the backbone atoms of the region connecting the OD helix with the protein core and with residue E78 (Fig. 4C), while in the holo Q428 interacts with S70 and, interestingly, K73 (N73 in rat) is salt-bridging with D425, thus contributing to the locked, inhibited structure in the holo-state of hPAH.

In order to study the effect of mutations between human and rat PAH we performed parallel MDs with our hPAH structures and existing rPAH structure(s) (*SI Materials and methods*) (Fig. S15). In agreement with our structural results, fluctuation analyses showed that, overall, hPAH is more dynamic, notably in the regions with significant structural differences (Fig. S16), such as RD, with several non-conservative mutations (Fig. S13B), and the area close to the OD helix.

Relevance for HPA and BH_4-responsive PKU. Increased understanding of the high vulnerability of human enzymes - notably those highly regulated, associated with central metabolic pathways - to disease-associated mutations points to evolutionary divergence from consensus amino acids that confer stabilization (35). This evolutionary mechanism also appears to describe the adaptation of hPAH towards a less stable and highly mobile enzyme (Fig. S16), most probably associated with the acquisition of the flexibility required for prompter response to changes in substrate and cofactor concentration along the evolution of the neuronal system (2). Flexibility is however often traded off with an increased susceptibility to mutations (35). Indeed, for the more than 900 hPAH mutations causing HPA/PKU (www.biopku.org), several studies have shown the relationship between hPAH protein stability and both remaining activity and allelic phenotype (3, 36). The enzyme destabilization caused by each mutation is actually highly predictive of patient phenotype and therapeutic response to BH_4 supplementation (3). As here reported, the stabilization by BH_4 is initiated at its binding-site in the CD but also affects the whole structure through networks that propagate to the RD and OD both inside the monomer (Fig. 2A, S5) and intra dimer (Fig. 2). HPA/PKU-associated mutations are located throughout the hPAH structure, but seem to be concentrated in hotspot regions with highly destabilizing mutations (235–281, corresponding to exon 7, and 282–330 to exons 8 and 9 in the *PAH* gene) (36). These regions include the flexible loop β6-α7 (244-250), whose conformation is modified (Fig. 2A1, S5) and rigidified (Fig. S12) upon BH_4 binding and presents a high mutation frequency with several mutations associated with moderate-to-severe (residues P244, V245,

A246, G247 and L249) or mild (L248) HPA/PKU phenotype (www.biopku.org and (36)). The hotspot regions also include mutations at R252, F254, and D315 (involved in changes from apo to holo conformation) (www.biopku.org and (36)). Together with S251, R252 appears essential for the stabilizing networks established with RD residues D27 in the IARS upon BH_4 binding (Fig. 2A1, S5). Likewise, the propagation of BH_4 binding from the CD to the OD through the RD towards the other subunits in the dimer also involves regions with high mutation frequency, such as the β2-β3 loop, which interacts with R297, D415 and E422 (Fig. 2A).

In conclusion, the identification of the networks that transmit the BH_4-bound state from the binding site at the CD to the other regions in the monomer, dimer and tetramer of hPAH, increments our understanding of the negative regulation of PAH by its cofactor and the stabilization caused by BH_4-binding, observed both for WT and mutant hPAH (4, 11). Thus, the structures of hPAH tetramers totally and partially bound with BH_4 presented here thus provide a rationale for BH_4-responsive PKU by sapropterin dihydrochloride, and pave the way to new stabilizing/chaperoning therapeutic approaches to address PKU.

Materials and Methods. Complete descriptions of the protein production; crystallization, crystallographic and electron microscopy structural determinations; of the computational methods; of activity measurements, and differential scanning calorimetry; and of the reagent preparations are in the *SI Materials and Methods*.

Accession numbers

The atomic coordinates for human PAH have been deposited in the Protein Data Bank under codes 6HYC and 6HPO, for the full-length hPAH and catalytic domain respectively. The EM maps have been deposited in the EMDB with accession s EMD-4605 (consensus refinement).

ACKNOWLEDGMENTS. The work in Spain was supported by grant from the Spanish Ministry of Economy and Competitiveness BFU2017-90030-P (to JAH), BFU2017-87316 (to RFL) and in Norway by grants from the programs Forny (248889/O30) and FRIMEDBIO (261826) from the Research Council of Norway (to AM), The Western Norway Regional Health Authorities (HV projects 911959 to MIF and 912246 to AM) and the K.G. Jebsen foundation (to AM). We thank Kay Diederichs and Pavel Afonine for discussions and help in data processing and refinement, the staff from ALBA and ESRF synchrotron facilities for support during

crystallographic data collection, the Midlands Regional Cryo-EM Facility at LIBSCB, major funder MRC, and Christos Savva for assistance during data acquisition. We thank the EM unit from CNIO and the cryo-EM unit from the CNB-CIB (CSIC) for support using the EM facilities and Peter Gimeson from Malvern Panalytical for help with DSC-experiments and discussions. We are very grateful to Professor Torgeir Flatmark for critical discussions on the manuscript.

References

1. Blau N, van Spronsen FJ, & Levy HL (2010) Phenylketonuria. *Lancet* 376(9750):1417-1427.

2. Flydal MI & Martinez A (2013) Phenylalanine hydroxylase: function, structure, and regulation. *IUBMB Life* 65(4):341-349.

3. Wettstein S*, et al.* (2015) Linking genotypes database with locus-specific database and genotype-phenotype correlation in phenylketonuria. *Eur J Hum Genet* 23(3):302-309.

4. Erlandsen H*, et al.* (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(48):16903-16908.

5. Pey AL*, et al.* (2008) Identification of pharmacological chaperones as potential therapeutic agents to treat phenylketonuria. *J Clin Invest* 118(8):2858-2867.

6. Santos-Sierra S*, et al.* (2012) Novel pharmacological chaperones that correct phenylketonuria in mice. *Hum Mol Genet* 21(8):1877-1887.

7. Fitzpatrick PF (2012) Allosteric regulation of phenylalanine hydroxylase. *Arch Biochem Biophys* 519(2):194-201.

8. Kaufman S (1993) The phenylalanine hydroxylating system. *Adv Enzymol Relat Areas Mol Biol* 67:77-264.

9. Xia T, Gray DW, & Shiman R (1994) Regulation of rat liver phenylalanine hydroxylase. III. Control of catalysis by (6R)-tetrahydrobiopterin and phenylalanine. *J Biol Chem* 269(40):24657-24665.

10. Mitnaul LJ & Shiman R (1995) Coordinate regulation of tetrahydrobiopterin turnover and phenylalanine hydroxylase activity in rat liver cells. *Proc Natl Acad Sci U S A* 92(3):885-889.

11. Thöny B, Ding Z, & Martinez A (2004) Tetrahydrobiopterin protects phenylalanine hydroxylase activity in vivo: implications for tetrahydrobiopterin-responsive hyperphenylalaninemia. *FEBS Lett* 577(3):507-511.

12. Arturo EC*, et al.* (2016) First structure of full-length mammalian phenylalanine hydroxylase reveals the architecture of an autoinhibited tetramer. *Proc Natl Acad Sci U S A* 113(9):2394-2399.

13. Jaffe EK (2017) New protein structures provide an updated understanding of phenylketonuria. *Mol Genet Metab* 121(4):289-296.

14. Meisburger SP*, et al.* (2016) Domain Movements upon Activation of Phenylalanine Hydroxylase Characterized by Crystallography and Chromatography-Coupled Small-Angle X-ray Scattering. *J Am Chem Soc* 138(20):6506-6516.

15. Jennings IG, Teh T, & Kobe B (2001) Essential role of the N-terminal autoregulatory sequence in the regulation of phenylalanine hydroxylase. *FEBS Lett* 488(3):196-200.

16. Pey AL, Thorolfsson M, Teigen K, Ugarte M, & Martinez A (2004) Thermodynamic characterization of the binding of tetrahydropterins to phenylalanine hydroxylase. *J Am Chem Soc* 126(42):13670-13678.

17. Erlandsen H*, et al.* (1997) Crystal structure of the catalytic domain of human phenylalanine hydroxylase reveals the structural basis for phenylketonuria. *Nat Struct Biol* 4(12):995-1000.

18. Fusetti F, Erlandsen H, Flatmark T, & Stevens RC (1998) Structure of tetrameric human phenylalanine hydroxylase and its implications for phenylketonuria. *J Biol Chem* 273(27):16962-16967.

19. Kobe B*, et al.* (1999) Structural basis of autoregulation of phenylalanine hydroxylase. *Nat Struct Biol* 6(5):442-448.

20. Martinez A*, et al.* (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 (Pt 2):589-597.

21. Døskeland AP, Martinez A, Knappskog PM, & Flatmark T (1996) 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 (Pt 2):409-414.

22. Martinez A, Olafsdottir S, Haavik J, & Flatmark T (1992) Inactivation of purified phenylalanine hydroxylase by dithiothreitol. *Biochem Biophys Res Commun* 182(1):92-98.

23. Solstad T, Stokka AJ, Andersen OA, & Flatmark T (2003) Studies on the regulatory properties of the pterin cofactor and dopamine bound at the active site of human phenylalanine hydroxylase. *Eur J Biochem* 270(5):981-990.

24. Haavik J, Doskeland AP, & Flatmark T (1986) Stereoselective effects in the interactions of pterin cofactors with rat-liver phenylalanine 4-monooxygenase. *Eur J Biochem* 160(1):1-8.

25. Andersen OA, Flatmark T, & Hough E (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(2):279-291.

26. Erlandsen H, Bjorgo E, Flatmark T, & Stevens RC (2000) Crystal structure and site-specific mutagenesis of pterin-bound human phenylalanine hydroxylase. *Biochemistry* 39(9):2208-2217.

27. Thorolfsson M*, et al.* (2002) L-phenylalanine binding and domain organization in human phenylalanine hydroxylase: a differential scanning calorimetry study. *Biochemistry* 41(24):7573-7585.

28. Zivanov J*, et al.* (2018) New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* 7.

29. Nakane T, Kimanius D, Lindahl E, & Scheres SH (2018) Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *Elife* 7.

30. Andersen OA, Flatmark T, & Hough E (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(5):1095-1108.

31. Panay AJ, Lee M, Krebs C, Bollinger JM, & Fitzpatrick PF (2011) Evidence for a high-spin Fe(IV) species in the catalytic cycle of a bacterial phenylalanine hydroxylase. *Biochemistry* 50(11):1928-1933.

32. Aguado C, Perez B, Ugarte M, & Desviat LR (2006) Analysis of the effect of tetrahydrobiopterin on PAH gene expression in hepatoma cells. *FEBS Lett* 580(7):1697-1701.

33. Gunasekera RS & Hyland K (2009) In vivo regulation of phenylalanine hydroxylase in the genetic mutant hph-1 mouse model. *Mol Genet Metab* 98(3):264-272.

34. Kowlessur D, Citron BA, & Kaufman S (1996) Recombinant human phenylalanine hydroxylase: novel regulatory and structural properties. *Arch Biochem Biophys* 333(1):85-95.

35. Medina-Carmona E*, et al.* (2017) Enhanced vulnerability of human proteins towards disease-associated inactivation through divergent evolution. *Hum Mol Genet* 26(18):3531-3544.

36. Pey AL, Stricher F, Serrano L, & Martinez A (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(5):1006-1024.

Figure Legends

Fig. 1. Crystal structure of human PAH. (**A**) Ribbon representation of hPAH in complex with BH_4 at the four active sites (Tet2). The tetramer is formed as a dimer of dimers by a two-fold symmetry (depicted). The equivalent monomers are represented with same colors. BH_4 cofactor is drawn as spheres. (**B**) The hPAH monomer with each domain colored differently. (**C**) Detailed view of the active site of hPAH in complex with BH_4 (green sticks). Relevant residues are represented as capped sticks and labeled. The Fe^{II} cation, not present in the structure, is represented as a transparent sphere for comparison purposes. Its location is obtained by direct superposition with the CD of human PAH (obtained in this work). Residues from this structure involved in the interaction with Fe^{II} cation are shown in gray sticks (H285, H290 and E330). Polar interactions represented as dotted lines.

Fig. 2. Long-range effects in hPAH upon BH_4 binding and cryo-EM of apo-hPAH. (**A**) Tet1 with monomers colored differently (apo in cyan and holo in ochre), relevant residues depicted as capped sticks and BH_4 as green spheres. (**A1**) Structural comparison of the active site in apo hPAH *vs* BH_4-bound hPAH. In the active site, L248 and F254 move their position to accommodate BH_4. In addition, the binding of the cofactor induces a change in the conformation of the β6-α7 loop that changes the interaction pattern between the N-term and the CD, that propagates to the RD and the start of the OD helix (see main text for details). Polar interactions represented as dotted lines. (**A2**,**A3**) Detailed view of the apo-holo dimer interface showing network of salt-bridge interactions. (**B**) Top and side view of the cryo-EM reconstruction of apo-hPAH tetramer (left) and multi-body reconstruction of two independent dimers (right). Map color represents the map local-resolution estimation (see color guide). Note that the multi-body refined map is a composite of two dimers refined independently. (**C**) Rigid body structural fitting of Tet2 dimer onto the cryo-EM multi-body reconstruction of two independent dimers.

Fig. 3. Transition from pre-catalytic to active cofactor binding. (A) Superposition of hPAH:BH_4 (Tet2 chain A; pink) with hPAH-CD:BH_4 (PDB 1J8U; grey) (RMSD 0.542 Å). In both structures, BH_4 is in the pre-catalytic position, with the O4 atom 3.8 Å from the Fe^{II} and with Y138 (not visible in Tet2) ~20 Å away. H-binding interactions of S23 with BH_4 and Y377 are shown as dashed lines. (B) Superposition of hPAH:BH_4 (Tet2 chain A; pink) with hPAH-CD:BH_4:THA (PDB 1MMK; grey), with BH_4 at the catalytic position, interacting with E286 and with the O4 atom 3.1 Å from Fe^{II}, as required for the generation of the Fe^{II}-O_2-BH_4 adduct. The movement of Y138, induced by substrate analogue (THA) binding, positions the hydroxyl group of Y138 2.9.1 Å away from O1' in the dihydroxypropyl side chain of BH_4. (C,D) Snapshots from the 2 ns TMD of following L-Phe binding to full-length hPAH (Tet2 chain A). At ~1.41 ns, a H-bond is formed between the side chain of T278 and the amino group of L-Phe, and the translation of Y138 is accelerated (C). At the end of the 2 ns-simulation the phenol O of Y138 a H-bonds with the O1' of BH_4 in its pre-catalytic position (D). (E) Initial reaction rates for WT, and mutants Y138F, S23A and Y377F of hPAH tested for the *L-Phe-activated* (preincubated with 1

mM L-Phe; black circles), *non-activated* (no ligand during preincubation; white circles) and *pre-catalytic BH_4-bound* (preincubated with 75 μM BH_4; red circles).

Fig. 4. Structural differences between human and rat PAH monomers. (**A**) Structural superimposition of apo hPAH (light blue), holo hPAH (orange) and the two conformers observed in apo rPAH (gray and black). Regions presenting most relevant changes are highlighted in boxes. The BH_4 cofactor is represented as spheres. (**B**) Detailed view of the structural changes observed in the N-terminal domain. (**C**) Detailed view of the interactions in the oligomerization helix of the apo hPAH. Relevant residues are drawn as capped sticks and labeled. (**D**) Detailed view of the interactions in the oligomerization helix of the holo hPAH.

Figure 1

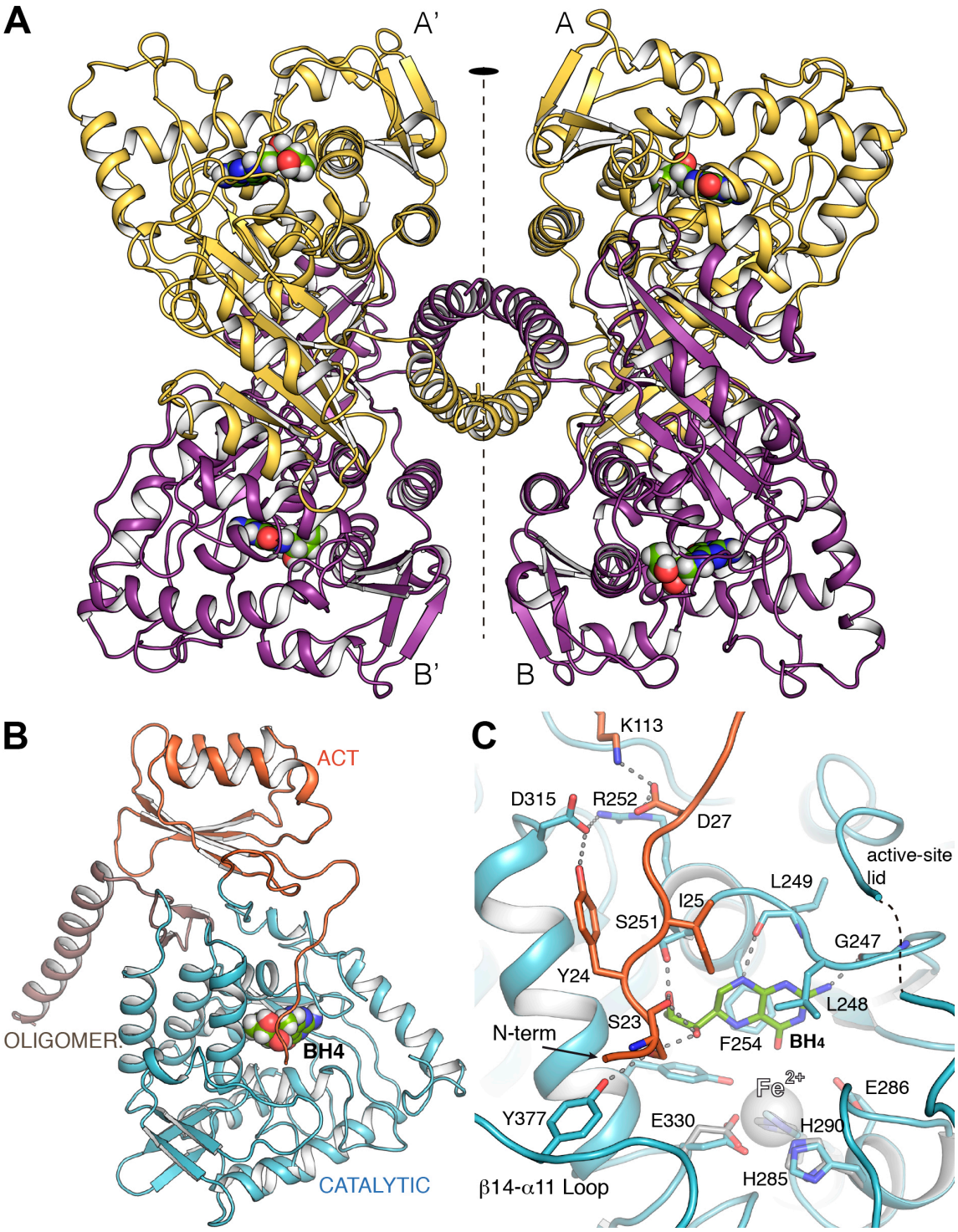

A
A'
A
B'
B
B
ACT
OLIGOMER
BH4
CATALYTIC
C
K113
D315
R252
D27
active-site
lid
L249
I25
S251
G247
Y24
L248
S23
N-term
F254
BH4
Fe2+
E286
Y377
E330
H290
H285
β14-α11 Loop

Figure 2

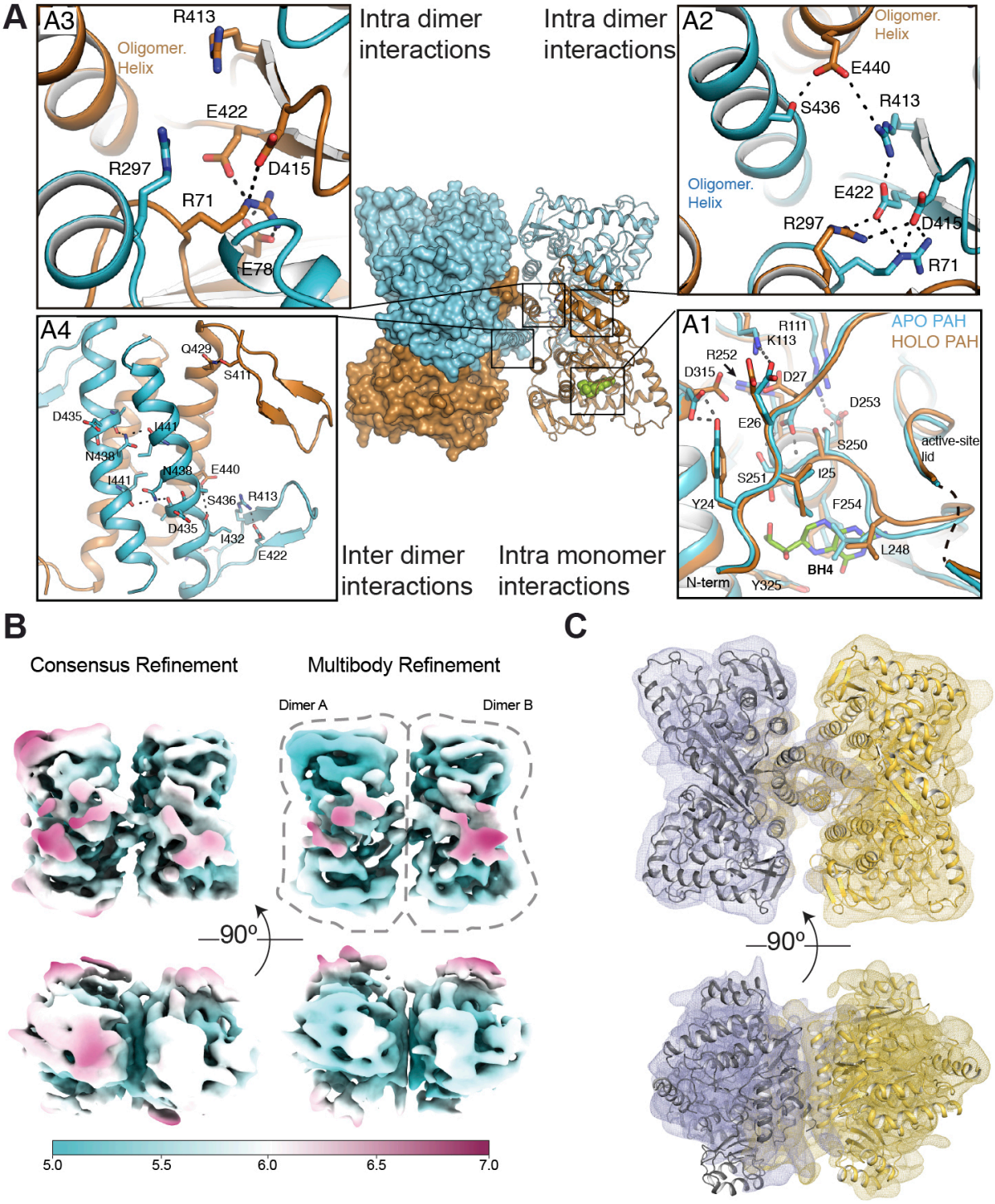

A
A3
R413
Oligomer. Helix
E422
R297
D415
R71
E78
Intra dimer interactions
Intra dimer interactions
A2
Oligomer. Helix
E440
S436
R413
Oligomer. Helix
E422
R297
D415
R71
A4
Q429
S411
D435
I441
N438
N438
E440
I441
S436
R413
D435
I432
E422
Inter dimer interactions
Intra monomer interactions
A1
R111
K113
APO PAH
HOLO PAH
R252
D315
D27
D253
E26
S250
active-site lid
S251
I25
Y24
F254
L248
BH4
N-term
Y325
B
Consensus Refinement
Multibody Refinement
Dimer A
Dimer B
90°
5.0
5.5
6.0
6.5
7.0
C
90°

Figure 3

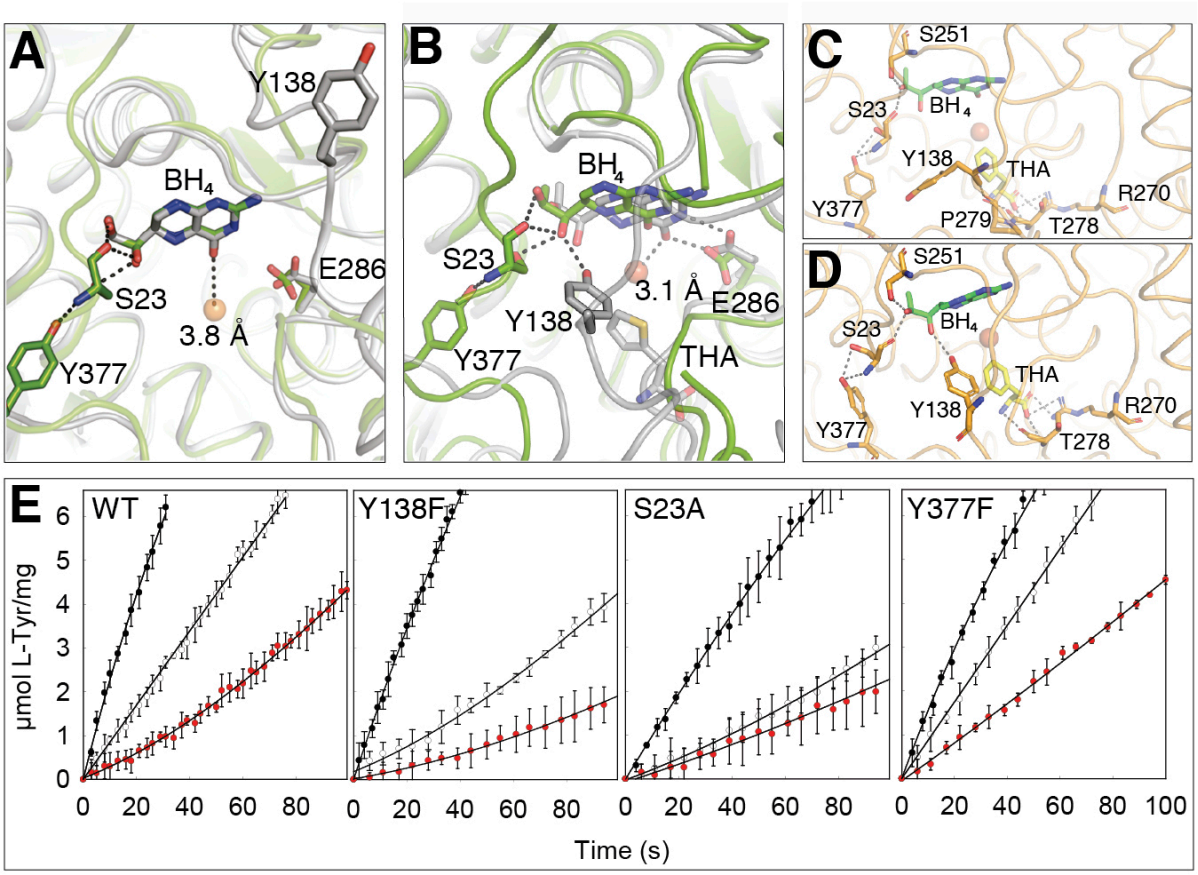
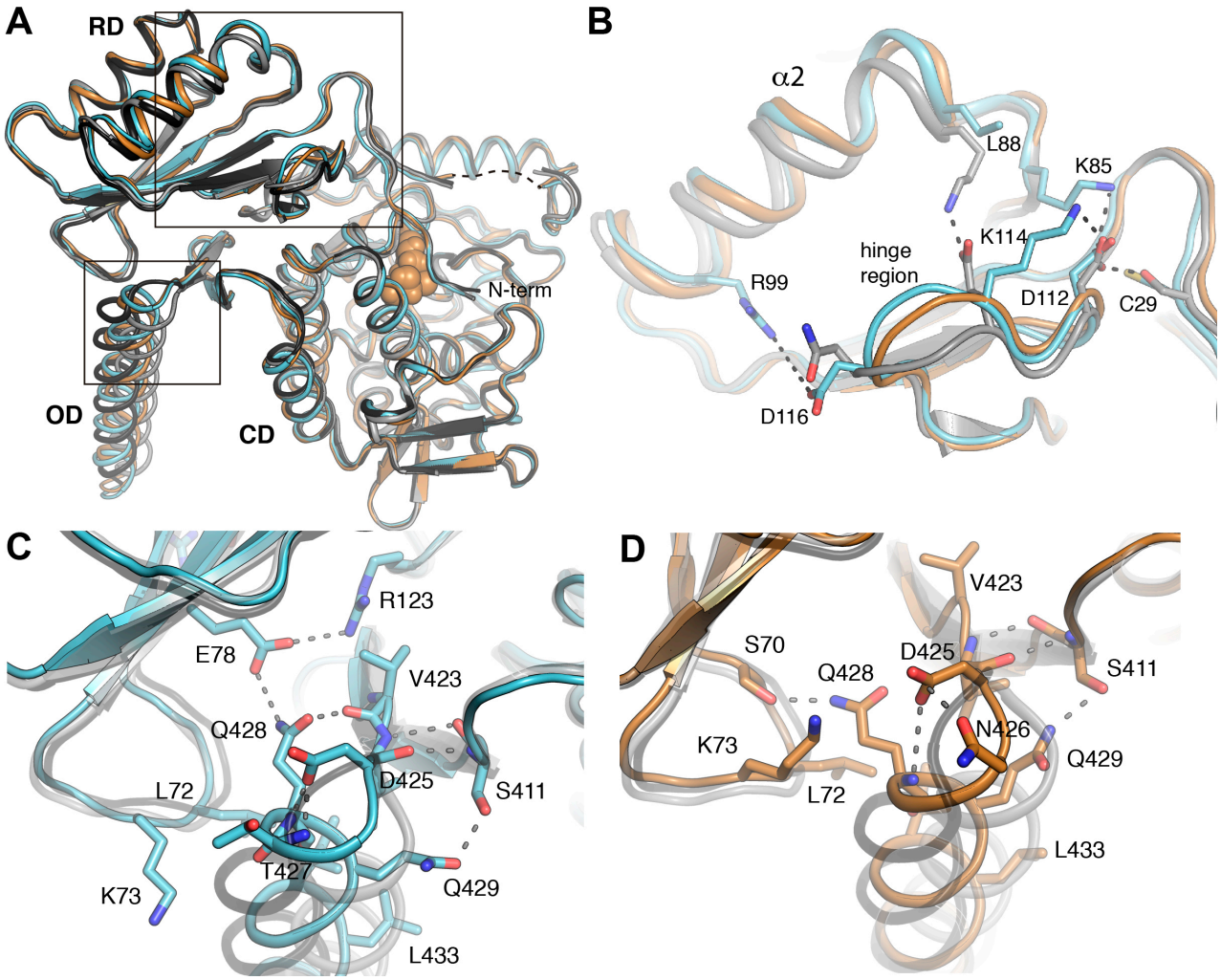

A
Y138
BH4
E286
S23
3.8 Å
Y377
B
BH4
S23
3.1 Å
E286
Y138
Y377
THA
C
S251
S23
BH4
Y138
THA
R270
Y377
P279
T278
D
S251
S23
BH4
THA
Y138
R270
Y377
T278
E
WT
Y138F
S23A
Y377F
µmol L-Tyr/mg
0
1
2
3
4
5
6
0 20 40 60 80
0 20 40 60 80
0 20 40 60 80
0 20 40 60 80 100
Time (s)

Figure 4


A
RD
N-term
OD
CD
B
α2
L88
K85
K114
hinge
region
R99
D112
C29
D116
C
R123
E78
V423
Q428
D425
S411
L72
T427
Q429
K73
L433
D
V423
S70
D425
S411
Q428
N426
K73
Q429
L72
L433