p-Coumaric Acid Decarboxylase from *Lactobacillus plantarum*: structural insights into the active site and decarboxylation catalytic mechanism

Héctor Rodríguez[1], Iván Angulo[2,4], Blanca de las Rivas[1], Nuria Campillo[3], Juan A. Páez[3], Rosario Muñoz[1] and José M. Mancheño[2*]

[1]Departamento de Microbiología, Instituto de Fermentaciones Industriales, CSIC, Juan de la Cierva 3, E-28006 Madrid, SPAIN

[2]Grupo de Cristalografía Macromolecular y Biología Estructural, Instituto Rocasolano, CSIC, Serrano 119, E-28006 Madrid, SPAIN

[3]Departamento de Quimioterapia, Instituto de Química Médica, CSIC, Juan de la Cierva 3, E-28006 Madrid, SPAIN

[4]Present address: The Scripps Research Institute, Scripps Florida, Jupiter, FL-33458, USA.

Structure of *p*-coumaric acid decarboxylase

Key words: catalytic mechanism; crystal structure; decarboxylase; phenolic acids; *p*-coumaric acid

Abbreviations used: JCSG, Joint Center for Structural Genomics; LpPDC, *Lactobacillus plantarum p*-coumaric acid decarboxylase; PAD, phenolic acid decarboxylase

*Correspondence to: José M. Mancheño, Grupo de Cristalografía Macromolecular y Biología Estructural, Inst. Rocasolano, CSIC, Serrano 119, 28006 Madrid, SPAIN. Tel: 34-915619400; Fax: 34-915642431.

E-mail: xjosemi@iqfr.csic.es

ABSTRACT

p-Coumaric acid decarboxylases (PDCs) catalyze the nonoxidative decarboxylation of hydroxycinnamic acids to generate the corresponding vinyl derivatives. Despite the biotechnological relevance of PDCs in food industry, their catalytic mechanism remains largely unknown. Here, we report insights into the structural basis of catalysis for the homodimeric PDC from *Lactobacillus plantarum* (LpPDC). The global fold of LpPDC is based on a flattened β-barrel surrounding an internal cavity. Crystallographic and functional analyses of single-point mutants of residues located within this cavity have permitted identifying a potential substrate-binding pocket and also to provide structural evidences for rearrangements of surface loops so that they can modulate the accessibility to the active site. Finally, combination of the structural and functional data with *in silico* results enables us to propose a two-step catalytic mechanism for decarboxylation of *p*-coumaric acid by PDCs where Glu71 is involved in proton transfer, and Tyr18 and Tyr20 are involved in the proper substrate orientation and in the release of the CO_2 product.

INTRODUCTION

Phenolic acids, mainly *p*-coumaric and ferulic acids, are covalently bound to polysaccharides in cell walls of higher plants.[1,2] Here, they act as crosslinkers between the lignins polymers and the hemicellulose and cellulose, probably contributing to the decrease in cell wall biodegradability[3] and regulating cell growth.[4] In this regard, lignins, the nature´s second most abundant organic plant substances in the terrestrial environment,[5] are mainly derived from the so called three classical monolignols,[6] namely, *p*-coumaric, coniferyl and sinapyl alcohols whose biosynthetic pathway involves the formation of *p*-coumaric acid from the amino acid phenylalanine.[7] As expected from the lignins composition, degradation of these two types of polymers by hemicellulases[8] and cinnamic acid esterases[9] releases a wide variety of phenolic acids, relevant molecules from biological[10] or biotechnological viewpoints.[11] In particular, lignins have attracted significant research attention because they represent a major obstacle in chemical pulping, forage digestibility, and processing of plant biomass to biofuels.[7] In this regard, it has been shown that phenolic acid catabolism is essential in the biodegradation of plant wastes[11] with the bacteria able to grow in this substrate as the sole source of carbon (*Pseudomonas* spp. and *Acinetobacter calcoaceticus*) metabolizing the phenolic acids ferulic and *p*-coumaric acids into vanillic and *p*-hydroxybenzoic acids, respectively.[12,13] Other microorganisms, such as *Saccharomyces cerevisiae*,[14] *Bacillus pumilus*,[15] *Bacillus subtilis*,[16] or *Lactobacillus plantarum*,[17] catalyze the nonoxidative decarboxylation of phenolic acids to produce 4-vinyl derivatives by means of phenolic acid decarboxylases (PADs).

Lactobacillus plantarum is a ubiquitous lactic acid bacterium in the plant kingdom that is frequently used as starter in the production of wine and many vegetable fermentations.[17,18] Expression of LpPDC by *L. plantarum* is highly regulated by phenolic acids such as caffeic, ferulic and *p*-coumaric acids,[17] which are effectively decarboxylated by the enzyme to their vinyl derivatives. Despite these volatile compounds are critical in providing the organoleptic properties of the wine and other fermented foods,[19] the structural basis of the catalytic mechanism of decarboxylation by LpPDC is largely unknown.

In this work, we provide insights into the structural basis of catalysis for LpPDC as a representative member of the PAD family of enzymes. With this aim, we present the crystal structure of wild-type LpPDC at 1.38 Å resolution, and also those from three single-point mutants whose activity is

either dramatically decreased (Y20F, R48Q) or completely lost (E71S). Analysis of the latter high

resolution crystal structures has permitted identifying a substrate-binding pocket within the unique

amphipathic cavity of LpPDC and also has revealed reorganizations of surface loops which may modulate

the access to the internal cavity. Finally, a catalytic mechanism is proposed based on the presented

structural, functional and *in silico* results.

MATERIALS AND METHODS

Protein mutagenesis, purification and crystallization

Wild-type his-tagged *p*-coumaric acid decarboxylase from *L. plantarum* (LpPDC) was produced recombinantly in *E. coli* as described previously.[20] Briefly, single point LpPDC mutants were constructed by a site-directed mutagenesis PCR technique using the plasmid pURI3-PDC as template. The mutagenic primers used to introduce the amino acid changes were: Y18F-5 (5´- CACACACTTTAT CT**TT**ACTTATGATAAC) and Y18F-3 (5´- GTTATCATAAGT**AA**AGATAAAG TGTGTG) for the Y18F change; Y20F-5 (5´-CACTTTATCTACACTT**T**TGATAACGGCTG) and Y20F-3 (5´-CAGCCGTTATCA**A**AAGTGTAGATAAAGTG) for Y20F change; Y20I-5 (5´-CACTTTAT CTACACT**AT**TGATAACGGCTG) and Y20I-3 (5´-CAGCCGTTATCA**AT**AGTGTAGATAAAGTG) for Y20I; Y20V-5 (5´-CACTTTATCTACACT**GT**TGATAACGGCTG) and Y20V-3 (5´-CAGCCGTTATCA**AC**AGTGTAGATAAAGTG) for Y20V; primers R48Q-5 (5´-GATGGTTGCCGG TC**AG**TGGGTCACTGATC) and R48Q-3 (5´-GATCAGTGACCCA**CT**GACCGGCAACCATC) for R48Q change, and finally, E71S-5 (5´-TCTTGGACT**AGC**CCAACTGGG) and E71S-3 (5´-CCCAGTTGG**GCT**AGTCCAAGA) for the E71S change (the nucleotide changes used to introduce the mutation are written in bold). The mutated LpPDC genes were sequenced to verify the expected nucleotide changes. All proteins were expressed and purified according to protocols described previously for the wild-type enzyme.[20] The mutant enzymes were crystallized at 291 K using the hanging-drop vapour diffusion method employing identical crystallization conditions as for the wild-type enzyme, namely, 20% (w/v) PEG4000, 12% (v/v) 2-propanol, 0.2 M sodium acetate, 0.1 M Tris-HCl pH 8.0 with 0.1 M barium chloride as additive.[20] Drops were prepared using 2 μl of protein solution (8-10 mg/ml), 2 μl of reservoir solution and 1 μl of additive.

Enzyme assay

PDC activity was assayed by the previously described high-performance liquid chromatography assay,[21] monitoring the formation of 4-vinyl-phenol from *p*-coumaric acid. Typically, the enzyme was

diluted with 25 mM phosphate buffer, pH 6.5, containing 4 mM *p*-coumaric acid, to a final concentration

of 0.1 μg/ml. The reaction mixture was incubated at 30°C for 20 min. when ethyl acetate was added for

reagents extraction.

Analytical ultracentrifugation

Equilibrium velocity ultracentrifugation experiments were performed at 10,500 rpm, 20°C, using

a Beckman XL-A ultracentrifuge with an An-50Ti rotor and standard double sector centerpiece cells.

Solvent density (1.005 mg/ml) and the partial specific volume of LpPDC (0.744) were calculated from the

buffer composition (0.1 M NaCl, 20 mM Tris-HCl) and from the predicted amino acid composition,

respectively, with SEDNTERP.[22] Centrifugation data were analyzed with the Beckman-Origin software.

Data collection, structure determination and refinement

Complete diffraction datasets were collected at the European Synchrotron Radiation Facility

(ESRF; Grenoble, France) in beamlines ID23-1, ID29 and BM16 for native, Y20F and R48Q LpPDCs,

respectively; dataset for E71S mutant was collected in-house on a Microstar rotating anode generator

(Bruker). All crystals were cryoprotected by a quick soak in optimized reservoir solution containing 15%

(v/v) MPD, and subsequently flash-cooled in a stream of liquid nitrogen at 100 K. Data were collected

using ADSC Q315 (ID23-1), ADSC Q315R (ID29) ADSC Q21R (BM16), and imaging plate mar345dtb

(in-house) detectors. Crystals of native and Y20F LpPDCs belong to the primitive tetragonal space group

$P4_3$ (unit cell parameters: $a = b = 43.18$ Å, $c = 232.49$ Å), those from R48Q mutant to the primitive

monoclinic space group $P2_1$ (unit cell parameters: $a = 39.24$ Å, $b = 94.92$ Å, $c = 106.72$ Å; $\beta = 100.51$),

and those from E71S mutant to the monoclinic space group $C2$ (unit cell parameters: $a = 108.81$, $b =$

52.75, $c = 82.09$; $\beta = 122.44$). Two protein molecules are found in the asymmetric unit of the tetragonal

crystals (53.2% (v/v) solvent content; $V_M = 2.63$ Å^3 Da^{-1}), four molecules for the R48Q monoclinic

crystals (47.1% (v/v) solvent content; $V_M = 2.33$ Å^3 Da^{-1}), and two molecules for the E71S monoclinic

crystals (52.9% (v/v) solvent content; $V_M = 2.61$ Å^3 Da^{-1}). Data collection and processing statistics are

shown in Table I.

Diffraction images were indexed and integrated with MOSFLM[23] and the data were scaled and truncated using SCALA and TRUNCATE from the CCP4 software suite.[24]

The structure of wild-type LpPDC was solved by molecular replacement (MR) using the atomic coordinates of the structure deposited under PDC entry 2GC9[25] as search model using the program MOLREP.[26] These last coordinates together with those of phenolic acid decarboxylase from *Bacillus subtilis* (PDB entry 2P8G[27]; see below) have been deposited in the PDB by the JCSG. In turn, the structures of single point mutants were solved by MR using the atomic coordinates of wild-type LpPDC. The models were refined with REFMAC5[28] applying non-crystallographic symmetry constraints. Atomic displacement parameters were refined in REFMAC by the TLS (translation, libration, screw) method. Automatic water molecule placement was done with ARP/wARP.[29] Manual rebuilding of the model was performed with the program O.[30] Data refinement statistics are shown in Table I. Stereochemical validation of the final models was checked using the program PROCHECK[31] or MOLPROBITY[32] for E71S mutant. Ramachandran plots were of good quality. Figures were prepared using PyMOL.[33]

The atomic coordinates and structure factors have been deposited at the Protein Data Bank under accession codes 2W2A (new crystal form of wild-type LpPDC), 2W2B (Y20F LpPDC), 2WSJ (E71S LpPDC), and 2W2F (R48Q LpPDC).

Docking studies

All calculations were performed on a Silicon Graphics Octane workstation (R12000, 300MHz) using the SYBYL 7.2 program suite.[34] The docking studies were achieved using the FlexiDock module of the SYBYL 7.2 suite of programs.[34]

We have used the crystal structure of R48Q LpPDC for docking studies since it adopts a closed conformational state (see below). For this purpose, this structure was edited (Gln48 was substituted by Arg48), protein hydrogen atoms were added, and partial charges were calculated using AMBER procedure. Positions of the hydrogen atoms were refined with the use of AMBER force field. To build the LpPDC:*o*-coumaric acid and LpPDC:*p*-coumaric acid complexes, the ligands were first manually positioned within the cavity close to the Tyr20 side chain taking into account the mutational studies. Subsequent energy minimization was performed using the AMBER99 force field and charges Gasteiger-

Hückel with geometry optimization of the ligands and the side chains of the enzyme. Energy minimizations were carried out using the conjugate gradient procedure until a gradient deviation of 0.01 kcal ml-1 Å^{-1} was attained. A distance-dependent dielectric constant was used in all the calculations.

These initial complexes were the input structure for docking studies using FlexiDock program that uses genetic algorithm[35] as quick method to generate conformations. This module analyses all possible ligand conformations within the active site and it takes into account both the receptor side chains and the ligand conformations. During the docking analysis, the protein was considered rigid except the residues involved in the putative binding site and the ligands which were considered flexible. The docking procedure consists of two steps: definition and refinement of a putative binding site, respectively. To achieve the first goal, four runs of FlexiDock were performed and the solutions were analysed and clustered into families. The representative conformer (with lower score of FlexiDock) from each family was minimized using the above mentioned conditions. For the second step, each complex obtained previously from the first step was subjected to four additional runs of FlexiDock. During this step possible hydrogen bonds were considered, defining donors and acceptor in the enzyme and in the ligand. The final solutions were analysed and clustered yielding different families. The representative conformer from each family was re-optimized. Analysis of the enzyme:ligand complex models generated after docking was based on the distances from the ligands to Tyr18, Tyr20, Arg48 and Glu71, hydrogen bond interactions, aromatic and hydrophobic interactions calculated with LPC program[36] and energy of binding and the difference accessible surface area (ΔASA) obtained with Structural Thermodynamics Calculations version 4.3.[37]

RESULTS AND DISCUSSION

Overall protein fold

p-Coumaric acid decarboxylase from *Lactobacillus plantarum* (LpPDC) is a cofactor-less homodimeric enzyme, with single-domain subunits composed of 178 amino acid residues. The atomic coordinates of LpPDC deposited in the Protein Data Bank by the JCSG (PDB code 2GC9[25]; unpublished results) have been used as search model for the resolution of a novel crystal form of LpPDC[20] which

permitted us to determine the structure of the enzyme at 1.38 Å resolution. The final model includes

residues 2-176 from each subunit (350 residues) and 566 water molecules (Table I). The approximate

dimensions of the dimeric assembly making up the asymmetric unit of the tetragonal crystal are 65 x 45 x

35 Å^3. Subunits within the dimer are virtually identical (rmsd, 0.07 Å for 175 Cα atoms).

The global structure of a single subunit of LpPDC is depicted in Fig. 1. The central folding motif

is based on a β-sandwich composed of two twisted and orthogonally stacked four-stranded anti-parallel β-

sheets plus one additional β-strand composed of 14 amino acid residues (β9; residues 129-142). This last

β-strand hydrogen bonds to both sheets, thereby making the sheets to be five-stranded *de facto* (Fig. 1a).

The final strand order is 9$_C$, 1, 2, 3, 4 for sheet 1 and, 5, 6, 7, 8, 9$_N$ for sheet 2 (9$_N$: residues 129-135; 9$_C$:

residues 136-142), respectively (Fig. 1b). The three dimensional arrangement of this highly bent β9 strand

is such that it closes one of the ends of the β-sandwich to form a continuously hydrogen-bonded β-barrel.

The opposite side of the barrel is capped by the connecting loops β2-β3 (residues 33-34) and β4-β5

(residues 51-55), respectively. Alternatively, the end of the β-barrel herein defined as the "top end", is

closed by the first 12 residue stretch of the polypeptide chain, which contains the single-turn helix α1

(residues 7-11), and also by the short loops β5-β6 (residues 62-63) and β7-β8 (residues 81-85).

Conversely, the "base" of the β-barrel is not closed in wild-type LpPDC but open to the external solvent,

with this opening providing the entrance to the internal cavity of the protein. The entrance to this cavity is

flanked by the loops β1-β2 (residues 21-22) and β3-β4 (residues 41-46) and by the short helix α2

(residues 96-100), and the N-terminal end of strand β9, and has rough dimensions of 8 Å by 8 Å.

The electron density of the 2F$_o$-F$_c$ map is of excellent quality for all of the side chains of the two

subunits of the asymmetric unit, with the exception of the sequence stretch comprising the residues Tyr20

to Trp24. Here, the electron density is poor in both subunits, suggesting that this region is flexible. As

above indicated, this region is located at the entrance of the cavity what would indicate that may be

functionally relevant. In fact, the analysis of the LpPDC mutants herein presented (see below) reveals that

the cavity can be completely shielded from the solvent as a result of structural rearrangements of this

sequence stretch in conjunction with the β3-β4 loop. These conformations permit defining open and

closed states in LpPDC which figuratively provide snapshots of two stages in the LpPDC catalytic

mechanism. Obviously, in this scenario both the entrance region and the internal cavity would be functionally relevant.

Comparison of the overall LpPDC protein structure

Analysis of multiple sequence alignments reveals that LpPDC exhibits high sequence similarity to several decarboxylases from bacterial sources and also to feruloyl decarboxylase from the protozoan *Trichomonas vaginalis,* the only close homolog from eukaryotic organisms. ClustalW2 sequence alignments of LpPDC with these proteins are reported in Fig. 1c. The level of sequence identity falls within the range of 88% (93% similarity) with phenolic acid decarboxylase (PAD) from *Lactobacillus brevis* (LbPAD), to 43% (61% similarity) with PAD from *Methylocella silvestris* (MsPAD), indicating that this group of proteins forms a distinct family of homologous enzymes (herein referred as PAD family). The highest level of sequence conservation corresponds to the β-strands that make up the β-barrel, in particular strands β1, β2, β3 and β4, with the highest variability being observed in helices α3 and α5, the connecting loop β9-α5 and the N- and C-terminal ends. This pattern of sequence conservation indicates that the β-barrel core is highly preserved within the PAD family. In this regard, the only members of the PAD family whose atomic coordinates have been deposited are LpPDC (PDB entry: 2GC9[25]) and PAD from *Bacillus subtilis* (BsPAD; PDB entry: 2P8G[27]). Both structures superpose almost perfectly with Cα rmsd of 1.0 Å over 162 aligned atoms, with the only significant differences being located at the N- and C-termini.

Not surprisingly, a structure-based similarity search with the DALI server[38] revealed BsPAD as the closest structural relative to LpPDC (Z-score= 29.6). Additionally, modest structural similarity is also detected with various β-barrel-containing proteins which bind hydrophobic ligands in an internal cavity located inside the β-barrel. Among these proteins, the most significant homologs are fatty acid-binding protein from toad liver (FABP, PDB entry 1P6P; Z-score= 7.2)[39] and cellular retinoic acid-binding protein 2 (CRABP-2, PDB entry 2G7B; Z-score= 6.7)[40] with rmsd of 1.7 and 1.9 Å for 73 and 63 Cα atoms, respectively (Fig.2). These two proteins are members of the fatty acid-binding proteins (FABP) family, which in turn belong to a superfamily of lipid-binding proteins.[41] All members of the FABP

family whose structure has been determined at high resolution show similar structural features:[42,43] they are composed of ten antiparallel β-strands that form a β-barrel, surrounding a central internal water-filled cavity which is the lipid binding site.[44] The structural similarity found with LpPDC together with its substrate specificity and the side chains that determine the interior of its cavity (see below) strongly supports that the substrate binding site of LpPDC is located within the internal cavity.

Quaternary structure

A dimeric assembly is identified within the asymmetric unit of the LpPDC tetragonal crystal (Fig. 3a). This dimer is maintained through interactions between residues that mainly belong to the β-sheet 2 of both subunits. The average value of the buried surface area at the interface between the subunits calculated by the PISA server[45] is large (~1170 Å^2) and contains both polar and apolar residues. The interactions identified involving hydrogen bonds and salt bridges are listed in Table II. Noteworthy, whereas polar side chains are mainly localized at the centre of the contacting interface, the apolar ones are preferentially found at the borders of this interface forming a species of "hydrophobic zipper" that would be highly conserved within the PAD family as indicated by the above sequence alignments.

In addition, the hydrodynamic behaviour of LpPDC has been analyzed by analytical ultracentrifugation techniques and gel-filtration chromatography. Results from sedimentation equilibrium experiments indicate that LpPDC exists in solution as a molecular species of 46.0 ± 2.0 kDa (Fig. 3b), which compares well with the theoretical mass expected for a dimer of the recombinantly expressed LpPDC (45.6 kDa). Additionally, results obtained from analytical gel-filtration chromatography on Superdex 200 10/300 GL, in conjunction with SDS-PAGE results (not shown) indicate that LpPDC behaves as a species of 45.2 ± 2.0 kDa (n=3) with subunits of 22 kDa. In summary, these results together with the crystallographic analysis demonstrate that LpPDC is a dimeric assembly both in solution and in the crystal state.

The relative orientation of the subunits within the dimer suggests that the accessibility of each cavity to the external solvent is not *prima facie* directly constrained by the accompanying subunit. In fact, the two openings are situated at opposing sides of the dimeric assembly (Fig. 3a) with the loops flanking the entrances not participating in intersubunit contacts.

Internal cavity of LpPDC

The β-barrel fold of the LpPDC subunits results in a large internal cavity which has an estimated solvent accessible volume of ~ 615 $Å^3$ using a probe radius of 1.4 Å with the CASTp algorithm.[46] The cavity has rough dimensions of 11 Å x 8 Å x 8 Å and a somewhat irregular surface. It has a unique connection to the external solvent flanked by the short loops β1-β2 and β3-β4, together with the single-turn helix α2 and the N-terminal end of strand β9. We define the entrance region as consisting of the above structural elements, and in particular, the region delimited by the amino acid side chains Asn22, Trp24, Met44, Glu99, Pro102, Leu129, Val131, and Glu133. According to the electron density map of wild-type LpPDC, the region comprising residues Tyr20 to Trp24 would be the one exhibiting the highest flexibility, what is consistent with a region of dynamic motion through which the substrate enters to the binding site.

The interior of the cavity is mainly determined by hydrophobic side chains contributed by the β-strands of the β-barrel (excluding strand β5), which are aromatic (Tyr18, Tyr20, Trp24, Tyr26, Trp28, Tyr38, Trp69, Phe94) (Fig. 4a) or aliphatic (Ile40, Val45, Val77, Leu79, Ile92, Val98) (Fig. 4b). Aromatic side chains are mainly clustered in one of the walls of the cavity and the aliphatic residues are mostly located in the opposite wall. Conversely, internal polar residues (Arg48, Glu71, Thr73, Thr75, Thr105) are also observed (Fig. 4c). They occupy a portion of the cavity that can be defined as polar because of the presence of an extensive network of polar interactions (Fig. 5). The carboxyl group of Glu71 makes a salt bridge with Arg48 (2.9 Å), and is at hydrogen bond distance to the hydroxyl group of Tyr38 (2.6 Å) and to the indole nitrogen of Trp69 (2.8 Å). In turn, the guanidinium moiety of the Arg48 side chain also participates in other polar interactions: the Nε atom forms a hydrogen bond with the hydroxyl group of Tyr38 (2.9 Å), and the NH1 and NH2 atoms are at hydrogen bond distances to the carbonyl oxygen atom of Thr105 (2.9 Å), and to the Oε1 atom of the Gln109 side chain (2.7 Å), respectively. Likewise, the hydroxyl group of Thr105 makes hydrogen bonds with the carbonyl oxygen atom of Pro102 (2.8 Å) and with an ordered water molecule (2.9 Å). Finally, another solvent molecule is at hydrogen bond distance to the hydroxyl groups of Thr73 and Thr75 (2.7 Å). Obviously, the formation

of this extensive network of polar interactions within an internal, predominantly hydrophobic cavity, contributes to the stabilization of these residues in such an environment. In this sense, it is worth to mention that the specific stabilizing role of solvent molecules within the cavity cannot be reliably deduced from the crystal structure since the inspection of the $2F_o-F_c$ electron density map of wild-type LpPDC indicates the presence of one molecule, whose precise identity remains undetermined. Here, it is observed a large and discontinuous stretch of poorly defined electron density which starts at the side chain hydroxyl group of Tyr20, located close to the entrance of the cavity, and penetrates deeply into this structure (Fig. 1 in Suplemental data). None of the components of the crystallization mother liquor or the optimized cryoprotectant solution properly fits into this electron density. Related to this, the UV-VIS absorbance spectrum of purified LpPDC exhibits a local maximum centred at 355 nm (not shown) revealing the presence of a non-dialyzable ligand tightly bound to LpPDC. As highly purified LpPDC is fully active against *p*-coumaric acid in our experimental conditions, it is probable that LpPDC with the bound molecule constitutes a minor fraction of the total crystallized protein.

As shown below, single point mutants of LpPDC affecting the two unique internal charged residues of the enzyme, namely, Arg48 and Glu71, and also the Tyr20, reveal a dramatic (Y20F; R48Q) or complete (E71S) loss of enzymatic activity without perturbing the structure of the protein. These results support the participation of the latter residues in the catalytic mechanism of LpPDC and thus indicate that both the polar portion of the cavity and the entrance region are functionally relevant. In other words, these residues form part of the substrate-binding pocket which would be located within the internal cavity of the β-barrel.

Mutational analysis of residues in the internal cavity and entrance loop

Currently the structural basis of the catalytic mechanism for members of the PAD family is poorly understood, and therefore our mutational analysis was initially based solely on the high resolution crystal structure of wild type LpPDC. In particular, we focused our attention on prominent features of the cavity, namely, the unique internal charged residues and also the putative flexible region at the entrance. Thus, our selected target residues were: Tyr20, Arg48 and Glu71. In particular, we prepared the following

mutants: Y20F, Y20I, Y20V, R48Q, and E71S. The catalytic activity against *p*-coumaric acid of these proteins was evaluated, and also they were subjected to crystallization trials.

Decarboxylase activity data against *p*-coumaric acid reveals that none of the designed point mutants is active in standard experiments (15 min. incubation at 37ºC), in conditions where 100% activity is observed for wild-type LpPDC. Longer incubation assays of 4 h. show 25% relative activity for Y20F and R48Q, 4% for Y20I and no activity for Y20V and E71S. Hence, it is evident that substitution of either Tyr20, Arg48 or Glu71 determines a dramatic or complete loss of decarboxylase activity which strongly supports the involvement of these residues in the catalytic mechanism of LpPDC.

To gain further insights into the structural basis of the decarboxylation mechanism of LpPDC we analyzed the structural consequences of the mutations. Suitable crystals could be prepared for three mutant variants: Y20F, R48Q and E71S. The structures have been refined to an R_{fac}/R_{free} of 19.7%/21.8% and 19.2%/21.0%, and 17.7%/25.9%, respectively (Table I). Three dimensional comparisons with wild-type LpPDC show that the three proteins are properly folded (rmsd values of 0.16, 0.70 Å, and 0.60 Å for 175 aligned Cα atoms for Y20F, R48Q and E71S, respectively), what reinforces the functional relevance of the mutated residues.

Markedly, the structures of the R48Q and E71S mutants revealed that the entrance region, particularly the β1-β2 and β3-β4 loops, adopt a distinct conformation which decreases the opening of the cavity dramatically (Fig. 6). In fact, this conformation indicates that the internal cavity of the enzyme is not readily accessible to any ligand, and therefore it can be operatively defined as "closed" (Fig. 6a) in contrast to the "open" conformation identified in wild-type LpPDC and Y20F mutant (Fig. 6b). In this sense, the crystal structure of the E71S mutant reveals conformational compatibility between subunits within the dimeric assembly since one E71S subunit is open and other subunit is closed. The putative switch between the open and closed conformations of LpPDC should involve important displacements of the Asn22, Trp24 and Met44 side chains, which would be the direct responsible for the closure of the entrance, together with minor rearrangements of the internal side chains Tyr20 and Tyr26. On the contrary, no significant changes are observed within the polar portion of the cavity. Not surprisingly, this conformational transition determines a significant decrease in the solvent accessible volume of the internal cavity from an initial estimated value of ~580 Å^3 (average value between wild-type and Y20F) to

a final value of ~290 Å^3 (average value between R48Q and E71S), which is similar to the value for the

van der Waals volume for *p*-coumaric acid (278 Å^3) as estimated with the program HOLES[47].

Three dimensional comparison of the polypeptide backbone of open (wild-type or Y20F) and

closed (R48Q or E71S) states of LpPDC reveals that the putative transition between them not only

involves movements of the β1-β2 and β3-β4 loops but also the entire molecule: the closure of the

entrance is accompanied by a subtle but global rearrangement of the polypeptide backbone which can be

interpreted as a contraction of the β-barrel, indicating some degree of plasticity of the LpPDC

architecture.

There are three important aspects of this conformational change that we have observed in the

crystalline state. First, the entrance regions of the mutant subunits are perfectly defined in the electron

density maps in contrast to the wild-type protein. Secondly, we can assume that the crystal packing

interactions do not affect the conformation of the entrance region as they are not engaged in crystal

contacts. In fact, the only exception to this is a crystal contact between subunits A within the monoclinic

crystal of the E71S mutant which has no significant structural consequences as deduced from the three

dimensional comparison with the subunit B where the entrance region does not participate in packing

interactions. In both cases, an identical closed state is observed which in turn perfectly superpose with the

closed state of the R48Q mutant. Thirdly, the crystallographic closed state of the enzyme is not a

sufficient condition for explaining the loss of decarboxylase activity in LpPDC since the R48Q mutant

shows some activity. Therefore, the most plausible interpretation of the two conformational states present

in our crystals is that they represent snapshots of two stages in the LpPDC catalytic mechanism.

Finally, the flexibility in the entrance loops suggests that the enzyme might undergo a

conformational change upon substrate binding. Similarly to the decarboxylation mechanisms proposed for

arylmalonate decarboxylase from *Bordetella bronchiseptica*[48] and α-amino-β-carboxymuconate-ε-

semialdehyde decarboxylase from *Pseudomonas aeruginosa,*[49] we assume that the entrance region closes

upon substrate binding to form a catalytically competent arrangement and also to exclude water

molecules from the substrate-binding pocket (see below).

Modeling of engaged *p*-coumaric acid

The presence of water molecules and also of some components of the mother liquor within the internal cavity of LpPDC in the crystal state (both wild-type and mutant variants) suggests that this structure is highly hydrated *in vivo*. It is probable that some water molecules must be displaced to accommodate the incoming *p*-coumaric acid. This assumption agrees well with the comparable estimated average values for the solvent accessible volume of the cavity in the closed sate of the enzyme (~290 $Å^3$) and the van der Waals volume of *p*-coumaric acid (~278 $Å^3$). Conversely, the side chains Tyr20, Arg48, and Glu71 are critical for the decarboxylation catalytic mechanism which reveals that the substrate binding pocket is located at the internal cavity. In this context, and considering that attempts to produce cocrystals of LpPDC (either native or mutant variants) with substrate analogues were unsuccessful, to determine the potential binding mode of *p*-coumaric acid to LpPDC we carried out modeling studies of docked *p*-coumaric acid into the cavity of the closed state of wild-type LpPDC, the most probable conformation for the substrate-bound enzyme according to the scenarios determined for arylmalonate decarboxylase from *B. bronchiseptica*[48] and α-amino-β-carboxymuconate-ε-semialdehyde decarboxylase from *P. aeruginosa*.[49] Therefore, we used the closed structure of the R48Q mutant (mutated back to the wild-type sequence) as the starting point for modeling studies.

With this aim, initial *in silico* complexes were built based on the above assumptions and subsequently used as input for the docking studies. The binding mode suggested by the lowest energy model suggests key contacts with putative active-site residues. We observe that the *p*-hydroxyl group of the substrate, which is a strictly required structural determinant for decarboxylation,[21,50] penetrates deeply into the cavity, reaching the polar portion (Fig. 7). Here, this group interacts with the functionally relevant residues Arg48 and Glu71; in particular, is at hydrogen bond distance to the NH1 (3.4 Å) and NH2 (2.9 Å) atoms of Arg48 and to the Oε1 (3.3 Å) atom of Glu71. On the other hand, the carboxyl moiety of the substrate can potentially form hydrogen bonds with the Tyr18 (3.2 Å) and Tyr20 (3.5 Å) side chain hydroxyl groups what may support the participation of the β1-β2 loop of LpPDC in the productive binding of the substrate, and therefore it is probable that the dynamic motion of this region would be necessary for catalysis. In addition, we also observe major hydrophobic contacts between the non-polar

regions of the substrate and the aliphatic/aromatic side chains of Tyr26, Tyr38, Ile40, Val45, Val77, Trp69 and Phe94 which would constitute an extended hydrophobic substrate binding pocket. In other words, the substrate binding pocket can be operatively described as formed by three distinct regions: a central hydrophobic pocket which contact both the aromatic ring and the aliphatic chain of the substrate, and two flanking regions participating in polar interactions, one involved in the specific recognition of the *p*-hydroxyl group, and the second one involved in interactions with the leaving carboxyl moiety of the substrate.

This binding mode reveals three important aspects. First, it suggests that the accessibility of the hydroxyl group of the substrate to the polar portion of the cavity where the functionally relevant residues Arg48 and Glu71 are located is crucial for the substrate specificity of the enzyme. In this sense, parallel *in silico* analyses with *o*-coumaric acid, which is not decarboxylated by LpPDC[17], suggested a similar binding mode, although in this case the *o*-hydroxyl group does not interact with Glu71 or Arg48. This result agrees with the observed interactions between the *p*-hydroxyl group and Arg48 and Glu71 being critical for the decarboxylation catalytic mechanism. Secondly, the hydrophobic pocket imposes a severe steric hindrance which restricts the maximum size of chemical groups present in the aromatic ring of potential substrates,[50] and therefore this region becomes a potential mutational target for the rational design of LpPDC variants with novel substrate specificities. Thirdly, Tyr18 directly interacts with the carboxyl group of *p*-coumaric acid, similarly to Tyr20. Based on this result, we generated the single point mutant Y18F and characterized its decarboxylase activity against *p*-coumaric acid. The obtained results indicated that Y18F behaves as the Y20F mutant, which agrees with the proposed binding mode, and also provides additional clues about potential catalytic mechanisms for LpPDC.

Proposed catalytic mechanism

Currently, the only available information on the chemistry employed by members of the PAD family proceeds from a previous work by Hashidoko and Tahara carried out on PAD from *Klebsiella oxytoca*[51] (KoPAD). Here, the authors proposed a chemical mechanism by KoPAD based upon the analysis of the stereochemical specificity of the proton transfer involved in the decarboxylation of *p*-coumaric acid.[51] On the other hand, in considering the potential mechanisms for members of the PAD

family one must take into account that the substrate must bear a *p*-hydroxyl group, a double bond between C7 and C8 atoms, and obviously the leaving carboxyl group. The relatively small size and rigidity of the *p*-coumaric acid, together with the structural and functional analysis herein presented for wild-type LpPDC and several mutant variants, has permitted reliable docking of the substrate into the cavity of the enzyme. These results indicate a substrate-binding mode where the *p*-hydroxyl group of the substrate must reach the polar portion of the cavity where it is engaged in a network of polar interactions with functionally essential residues and the leaving carboxyl group would interact with the side chain hydroxyl groups of Tyr18 Tyr20. In this scenario, it is remarkable that important features of the chemical mechanism proposed by Hashidoko and Tahara for decarboxylation of *p*-coumaric acid[51] are easily integrated in this binding mode suggesting a consistent two-step catalytic mechanism (Fig. 8) in which Glu71 would act as the catalytic acid-base, and Tyr18 and Tyr20 would play critical roles in proper orientation of the substrate and in the release of CO_2.

We have considered two assumptions in the proposed catalytic mechanism. First, the substrate binds the internal cavity of LpPDC with the carboxyl group in an unprotonated state, which agrees with the pH-dependence of the LpPDC decarboxylase activity,[21] and the recently estimated pK_a value of the carboxyl group of *p*-coumaric acid in aqueous solution[52] (~4.3). Secondly, the side chain carboxylic group of Glu71 is also unprotonated. Consequently, the negative charge on the Glu71 carboxylate group would be stabilized by the positively charged guanidinium moiety of the internal residue Arg48.

In the first step of the catalytic mechanism (Fig. 8), the phenolic *p*-hydroxyl group is deprotonated by the carboxylate anion of Glu71, promoting a first electron flow along the cinnamate structure. Subsequently, the C8 carbon appearing as the nucleophilic center on the quinone form accepts a proton from a proton donor to form a *para*-quinone methide intermediate. As pointed out previously[52], the requirement of a free *p*-hydroxyl group in the substrate suggests a mechanism involving this quinone intermediate rather than a 1,2-addition pathway.[53] Our structural results indicate that the most probable proton donor is a water molecule as no other potential donor from the enzyme is in the vicinity of the C8 carbon atom within the enzyme:substrate complex. Interestingly, ordered water molecules are identified in this region of the internal cavity of the closed state of R48Q, the only closed structure containing solely water molecules within the cavity. In this mechanism, Trp69 and Arg48 would interact with the Glu71

side chain and help to orientate it in the active site. In the second step, the *para*-quinone methide intermediate would promote a second electron flow from the carboxylate, liberating CO_2 and forming the C8/C9 double bond and the subsequent vinyl group. In this process, the hydroxyl groups of the Tyr18 and Tyr20 side chains would interact with the leaving carboxylate group, maintaining it in a fixed position, possibly contributing to the stereochemical specificity of the electron transfer that has been shown unambiguously for KoPAD.[52]

Finally, it should be pointed out that some aspects of the catalytic mechanism need to be further investigated, namely, the exact roles for the hydrophobic pocket protein residues and importantly the relation between the conformational switch between open and closed states of LpPDC and the chemical mechanism for decarboxylation. Clearly, trapping the substrate- and/or structural analogues-bound enzyme intermediates and characterizing their structures will be most informative with regard to the mechanism of LpPDC.

In summary, the results presented here provide insights into the structural basis for the catalytic mechanism of decarboxylation of phenolic acids by phenolic acid decarboxylases. As these enzymes have important biotechnological applications in food industry as catalysts for the production of vinyl-derivatives, our results open up novel biotechnological possibilities for the rational approach of designing enzymes *ad hoc* with altered specificities. The elucidation of the crystal structures of native and mutant LpPDCs allowed for the identification of the substrate-binding site located inside the β−barrel of the enzyme and revealed different conformations of flexible loops at the entrance of this cavity. This is consistent with a catalytic mechanism involving a conformational transition between open (substrate-free) and closed (substrate-bound) states. As according to the proposed catalytic mechanism, the hydrophobic pocket of the binding site harbors determinants of substrate specificity, novel specificities may likely be engineered by altering this region. In turn, the presence of flexible structural elements participating in the catalytic mechanism, which are known to provide de basis for functional diversity and enzyme evolvability.[54]

ACKNOWLEDGEMENTS

REFERENCES

1. Hartley RD, Ford CW. Phenolic constituents of plant-cell walls and wall biodegradability. In: Lewis NG, Paice MG, editors. Plant cell wall polymers: biogenesis and biodegradation. Washington, DC: American Chemical Society; 1989. p. 137-145.

2. de Vries RP, Michelsen B, Poulsen CH, Kroon P A, van den Heuvel RHH, Faulds CB, Williamson G, van den Hombergh JPTW, Visser J. The faeA genes from *Aspergillus niger* and *Aspergillus tubingensis* encode ferulic acid esterases involved in degradation of complex cell wall polysaccharides. Appl. Environ. Microbiol. 1997;63:4638-4644.

3. Borneman WS, Akin DE, van Eseltine WP. Effect of phenolic monomers on ruminal bacteria. Appl. Environ. Microbiol.1986;52:1331-1339.

4. Fry SC. Phenolic components of the primary-cell wall and their possible role in the hormonal-regulation of growth. Planta. 1979;146:343-351.

5. Freudenberg K. The constitution and biosynthesis of lignin. In Freudenberg K, Neish AC, editors. Constitution and Biosynthesis of Lignin. New York: Springer-Verlag; 1968. p 45-122.

6. Davin LB, Lewis NG. Lignin primary structures and dirigent sites. Curr. Opin. Biotechnol. 2005;16:407-415.

7. Vanholme R, Morreel K, Ralph J, Boerjan W. Lignin engineering. Curr. Opin. Plant Biol. 2008;11:278-285.

8. de Vries RP, Poulsen CH, Madrid S, Visser J. aguA, the gene encoding an extracellular alpha-glucuronidase from *Aspergillus tubingensis*, is specifically induced on xylose and not on glucuronic acid. J. Bacteriol. 1998;180:243-249.

9. Christov LP, Prior BA. Esterases of xylan-degrading microorganisms: production, properties, and significance. Enzyme Microb. Technol. 1993;15:460-475.

10. Lee Y-W, Jin S, Sim W-S, Nester EW. Genetic evidence for direct sensing of phenolic compounds by the VirA protein of *Agrobacterium tumefaciens*. Proc. Nat. Acad. Sci. USA 1995;92:12245-12249.

11. Barthelmebs L, Divies Ch, Cavin J-F. Knockout of the p-coumarate decarboxylase gene from *Lactobacillus plantarum* reveals the existence of two other inducible enzymatic activities involved in phenolic acid metabolism. Appl. Environ. Microbiol. 2000;66:3368-3375.

12. Venturi V, Zennaro F, Degrassi G, Okeke BC, Bruschi CV. Genetics of ferulic acid bioconversion to protocatechuic acid in plant-growth-promoting *Pseudomonas putida* WCS358. Microbiology 1998;144:965-973.

13. Segura A, Bünz PV, D´Argenio DA, Ornston LN. Genetic analysis of a chromosomal region containing vanA and vanB, genes required for conversion of either ferulate or vanillate to protocatechuate in *Acinetobacter*. J. Bacteriol. 1999;181:3494-3504.

14. Clausen M, Lamb CJ, Megnet R, Doerner PW. PAD1 encodes phenylacrylic acid decarboxylase which confers resistance to cinnamic acid in *Saccharomyces cerevisiae*. Gene 1994;142:107-112.

15. Zago A, Degrassi G, Bruschi, CV. Cloning, sequencing, and expression in *Escherichia coli* of the *Bacillus pumilus* gene for ferulic acid decarboxylase. Appl. Environ. Microbiol. 1995;61:44844-4486.

16. Cavin J-F, Dartois V, Diviès C. Gene cloning, transcriptional analysis, purification, and characterization of phenolic acid decarboxylase from *Bacillus subtilis*. Appl. Environ. Microbiol. 1998 ;64:1466-1471.

17. Cavin J-F, Barthelmebs L, Diviès C. Molecular characterization of an inducible p-coumaric acid decarboxylase from *Lactobacillus plantarum*: gene cloning, transcriptional analysis, overexpression in *Escherichia coli*, purification, and characterization. Appl. Environ. Microbiol. 1997;63:1939-1944.

18. Olsen EB, Russel JB, Henick-Kling T. Electrogenic L-malate transport by *Lactobacillus plantarum*: a basis for energy derivation from malolactic fermentation. J. Bacteriol. 1991;173:6199-6206.

19. Etièvant PX, Issanchou SN, Marie S, Ducruet V, Flanzy C. Sensory impact of volatile phenols on red wine aroma. Influence of carbonic maceration and time of storage. Sciences des Aliments. 1989 ;9:19-33.

20. Rodríguez H, de las Rivas B, Muñoz R, Mancheño, JM. Overexpression, purification, crystallization and preliminary structural studies of p-coumaric acid decarboxylase from *Lactobacillus plantarum*. Acta Crystallog. F. 2007;63:300-303.

21. Rodríguez H, Landete JM, Curiel JA, de las Rivas B, Mancheño JM, Muñoz R. Characterization of the p-coumaric acid decarboxylase from *Lactobacillus plantarum* CECT 748(T). J. Agric. Food Chem. 2008;56:3068-3072.

22. Laue TM, Shah BD, Ridgeway TM, Pelletier SL. Biochemistry and Polymer Science. London: Royal Society of Chemistry; 1992.

23. Leslie AGW. Recent changes to the MOSFLM package for processing film and plate data. Joint CCP4 and ESF-EAMCB Newsletter on Protein Crystallography. 1992 No. 26.

24. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr.1994;50:760-763.

25. Joint Center for Structural Genomics. Crystal structure of p-coumaric acid decarboxylase (NP_786857.1) from *Lactobacillus plantarum* at 1.70 Å resolution. PDB entry: 2GC9.

26. Vagin A, Tepliakov A. MOLREP: an automated program for molecular replacement. J. Appl. Crystallog. 1997;30:1022-1025.

27. Joint Center for Structural Genomics. Crystal structure of phenolic acid decarboxylase (2635953) from *Bacillus subtilis* at 1.36 Å resolution. PDB entry: 2P8G.

28. Murshudov GN, Vagin AA, Dobson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D. 1997;53:240-255.

29. Perrakis A, Morris R, Lamzib VS. Automated protein model building combined with iterative structure refinement. Nat. Struct. Biol. 1999;6:458-463.

30. Jones TA, Zou J-Y, Cowtan SW, Kjelgaard M. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A 1991;47:110-119.

31. Laskowski R, MacArthur M, Moss D, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. 1993;26:283–291.

32. Davis I W, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB 3rd, Snoeyink J, Richardson JS, Richardson DC. MolProbity: all-atoms contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. 2007;35:W375-383.

33. DeLano WL. The PyMOL Molecular Graphics System (Palo Alto, CA: DeLano Scientific, LLC). (http://www.pymol.org). 2002.

34. SYBYL 7.2 , Tripos Inc., 1699 South Hanley Rd., St. Louis, MO 63144, USA

35. Judson OP, Haydon D. The genetic code: what is it good for? An analysis of the effects of selection pressures on genetic codes. J. Mol. Evol. 1999;49:539-550.

36. Sobolev V, Sorokine A, Prilusky JE, Abola E, Edelman M. Automated analysis of interatomic contacts in proteins. Bioinformatics 1999;15:327-332.

37. Lavigne P, Bagu JR, Boyko R, Willard L, Holmes CFB, Sykes BD. Structure-based thermodynamic analysis of the dissociation of protein phosphatase-1 catalytic subunit and microcystin-LR docked complexes. Prot. Sci. 2000;9:252-264.

38. Holm L, Sander C. Searching protein structure databases has come of age. Proteins: Struct. Funct. Genet. 1994;19:165-173.

39. Di Pietro SM, Córsico B, Perduca M, Monaco HL, Santomé JA. Structural and biochemical characterization of toad liver fatty acid-binding protein. Biochemistry 2003;42:8192-8203.

40. Vasileiou C, Vaezeslami S, Crist RM, Rabago-Smith M, Geiger JH, Borhan B. Protein design: reengineering cellular retinoic acid binding protein II into a rhodopsin protein mimic. J. Am. Chem. Soc. 2007;129:61140-6148.

41. Zimmerman AW, Veerkamp JH. New insights into the structure and function of fatty acid-binding proteins. Cell. Mol. Life Sci. 2002;59:1096-1116.

42. Reese-Wagoner A, Thompson J, Banaszak LJ. Structural properties of the adypocyte lipid binding protein. Biochim. Biophys. Acta 1999;1441:106-116.

43. Thompson J, Reese-Wagoner A, Banaszak LJ. Liver fatty acid binding protein: species variation and the accommodation of different ligands. Biochim. Biophys. Acta 1999;1441:117-130.

44. Banaszak LJ, Winter N, Xu Z, Bernlohr DA, Cowan S, Jones TA. Lipid binding proteins: A family of fatty acid and retinoid binding proteins. Adv. Prot. Chem. 1994;45:89-151.

45. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. 2007;372:774-797.

46. Liang J, Edelsbrunner H, Woodward C. Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. Protein Sci. 1998;7:1884-1897.

47. Roux MV, Jiménez P, Vacas A, Cano FH, Apreda-Rojas MC, Ros F. A compact and lipophilic enantiomer showing a bilayer crystal structure – free energy of spontaneous resolution of the racemic compound into the crystalline enantiomers. Eur. J. Org. Chem. 2003; 2084-2091.

48. Kuettner EA, Keim A, Kircher M, Rosmus S, Sträter N. Active-site mobility revealed by the crystal structure of arylmalonate decarboxylase from *Bordetella bronchiseptica*. J. Mol. Biol. 2008;377:386-394.

49. Martytnowski D, Eyobo Y, Li T, Yang K, Liu A, Zhang H. Crystal structure of alpha-amino-beta-carboxymuconate-epsilon-semialdehyde decarboxylase: insight into the active site and catalytic mechanism of a novel decarboxylation reaction. Biochemistry 2006 ;45:10412-10421.

50. Cavin J-F, Barthelmebs L, Guzzo J, van Beeumen J, Samyn B, Travers J-F, Diviès C. Cloning and sequencing of the gene encoding alpha-acetolactate decarboxylase from *Leuconostoc oenos*. FEMS Microbiol. Lett. 1997;147:291-295.

51. Hashidoko Y, Tahara S. Stereochemically specific proton transfer in decarboxylation of 4-hydroxycinnamic acids by 4-hydroxycinnamate decarboxylase from *Klebsiella oxytoca*. Arch. Biochem. Biophys. 1998;359:225-230.

52. Ozkorucuklu SP, Beltrán JL, Fonrodona G, Barrón D, Alsancak G, Barbosa J. Determination of dissociation constants of some hydroxylated benzoic and cinnamic acids in water from mobility and spectroscopic data obtained by CE-DAD. J. Chem. Eng. Data 2009;54:807-811.

53. Gramatica P, Ranzi BM, Manitto P. Decarboxylation of cinnamic-acids by *Saccharomyces cerevisiae*. Bioorg. Chem. 1981;10:14-21.

54. Tokuriki N, Tawfik DS. Protein dynamism and evolvability. Science 2008;324:203-207.

55. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read NJ, Rice LM, Simonson T, Warre GL. Crystallography and NMR System: A new software Suite for macromolecular Structure Determination. Acta Cryst. 1998;D54:905-921.

FIGURE LEGENDS

Figure 1. Overall structure of LpPDC and amino acid sequence alignments. (a) Ribbon diagrams of a subunit of LpPDC (residues 2-176). Two orthogonal close-up views are shown. Strands making up the β-barrel are shown in blue, helices in yellow and loops in red. (b) Topology plot of a subunit of LpPDC. β-sheet 1 is indicated in blue and β-sheet 2 in red; helices are depicted as yellow circles. Numbers indicate the first and the last residues of the secondary structure element. (c) Amino acid sequence alignments of LpPDC and other phenolic acid decarboxylases. The regular secondary structure elements in *L. plantarum* LpPDC are indicated by green boxes (α-helix) and orange arrows (β-strand). **LpPDC**: *Lactobacillus plantarum* PDC (Q88RY7); **LbPAD**: *Lactobacillus brevis* PAD (Q03TU3); **PpPAD**: *Pediococcus pentosaceus* PAD (Q9F3X2); **LsPDC**: *Lactobacillus sakei* PDC (Q38UX6); **EfPDC**: *Enterococcus faecium* PDC (Q3Y2T7); **BcPAD**: *Bacillus coagulans* PAD (A2U705); **BlPAD**: *Bacillus licheniformis* PAD (Q65FC9); **BaPAD**: *Bacillus amyloliquefaciens* PAD (A7Z928); **BsPAD**: *Bacillus subtilis* PAD (007006); **BPAD:** *Bacillus sp.* PAD (Q8KNX7); **BpFDC**: *Bacillus pumilus* FDC (Q45361); **BpPAD**: *Bacillus pumilus* PAD (A8FAY2); **BmPAD:** *Bacillus mesentericus* PAD (Q9EXR7); **SePAD:** *Saccharopolyspora erythraea* PAD (A4FMV0); **CmPAD**: *Clavibacter michiganensis* PAD (A5CN25); **MpPUP**: *Mycobacterium paratuberculosis* Putative Uncharacterized Protein (Q743A0); **MaPAD**: *Mycobacterium avium* PAD (A0QB49); **MPAD**: *Methylobacterium* sp. PAD (B0UN01); **MsPAD**: *Methylocella silvestris* PAD (B1THI6); **KpPAD**: *Klebsiella pneumoniae* PAD (A6TFA1); **KoPAD**: *Klebsiella oxytoca* PAD (A6BMM9); **TvFED:** *Tricomonas vaginalis* Ferulate decarboxylase (A2F476); **EsPUP**: *Enterobacter sakazakii* Putative Uncharacterized Protein (A7MLR8); **EcPAD**: *Erwinia catovora* PAD (Q6DB32); **VcPAD**: *Vibrio cholerae* PAD (Q9KPX2); **LlPAD**: *Lactococcus lactis* PAD (A2RN76). In the alignment, positions identical in all sequences are marked with a blue background and white characters, and highly conserved positions are marked with a yellow background. Red characters in the LpPDC sequence correspond to residues facing the internal cavity of the enzyme.

Figure 2. Superposition of the LpPDC subunit with FABPs. The β-barrel of the wild-type LpPDC subunit is shown in blue as ribbon model and superimposed with: (a) the β-barrel of fatty-acid binding protein from toad liver (ribbon model in green; PDB entry 1P6P; Z-score= 7.2); and (b) the β-barrel of cellular retinoic acid-binding protein 2 (ribbon model in orange; PDB entry 2G7B; Z-score= 6.7).

Figure 3. Quaternary structure of LpPDC in the crystal and in solution. (a) Dimeric assembly of LpPDC. Each subunit is coloured according to secondary structure. Two orthogonal views are shown and the dimensions are indicated. The loop β1-β2 at the entrance of the internal cavity is indicated with a red circle. This loop is disordered in the electron density of the native LpPDC. (b) Analytical ultracentrifugation sedimentation equilibrium analysis of LpPDC (11.35 μM) at 16,000 rpm. Absorbance at 280 nm is plotted against the radial position from the centre of the rotor. The best fit to the data set (solid line curve) is a unique species of weight molecular mass of 46.0 kDa. Residuals from this fit are shown in the top panel.

Figure 4. Structural analysis of the internal cavity of LpPDC. The amino acid residues facing the internal cavity of LpPDC determined by the CASTp server[46] have been classified as aromatic (a), aliphatic (b), and polar (c). In the three panels, amino acid side chains are shown in *ball-and-stick* representation; green for aromatic, yellow for aliphatic and grey for polar residues, respectively.

Figure 5. Interactions between polar residues within the wild-type LpPDC cavity. Stereo view representation of the polar region of the LpPDC cavity. Representative side chains are shown as sticks; oxygen atoms are red and nitrogen atoms are blue. A water molecule is shown as a red ball. An $F_o - F_c$ omit map contoured at the 3σ level is shown in blue. Potential hydrogen bonds are depicted as dotted lines. The map has been calculated with the CNS[55] suite.

Figure 6. Open and closed conformational states of LpPDC. The entrance region of wild-type LpPDC is shown in yellow as ribbon model and superimposed with: (a) the R48Q mutant LpPDC structure (blue) and (b) the Y20F mutant LpPDC structure (green). The right panels show close-up views of the molecular surfaces of the R48Q (upper panel) and Y20F (lower panel), respectively. The same orientation is shown in the panels. Side chains of residues that constitute the entrance regions are shown as sticks (color code is as in the ribbon models).

Figure 7. *In silico* model for the complex of LpPDC and *p*-coumaric acid. Stereo view representation of the *p*-coumaric acid binding site of LpPDC (see text for further details). Residues potentially involved in hydrogen bonding interactions with *p*-coumaric acid (Tyr18, Tyr20, Arg48, and Glu71) are displayed in yellow as *stick* model. Potential hydrogen bonds (black broken lines) include interactions between the *p*-hydroxyl group of the substrate and functionally relevant residues Glu71 and Arg48, and also between the carboxyl moiety and hydroxyl groups from Tyr18 and Tyr20. Van der Waals interactions are significant in the binding of the substrate. Residues forming the hydrophobic pocket (Tyr26, Tyr38, Ile40, Val45, Trp69, V77 and Phe94) are depicted in grey.

Figure 8. Proposed catalytic mechanism of decarboxylation by LpPDC. The mechanistic proposal is based on inferences drawn from the crystallographic, functional and *in silico* results presented here and by analogy to the proposed chemical mechanism previously proposed for PAD from *Klebsiella oxytoca*.[51]

1
2
3
4
5
6 **Table I.** Data collection and refinement statistics.
7

	Wild-type	Y20F	R48Q	E71S
PDB code	2W2A	2W2B	2W2F	2WSJ
Data collection				
Beamline	ID23.1 (ESRF)	ID29 (ESRF)	BM16 (ESRF)	------------
Wavelength (Å)	0.976	0.978	0.979	1.5418
Space group	$P4_3$	$P4_3$	$P2_1$	$C2$
Unit cell parameters				
a, b, c (Å)	43.2, 43.2, 232.5	43.2, 43.2, 232.5	39.2, 94.9, 106.7	108.8, 52.75, 82.09
α, β, γ (º)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 100.5, 90.0	90.0, 122.4, 90.0
V_M (Å^3/Da)	2.6	2.6	2.3	2.6
Solvent content (% v/v)	53.2	53.2	47.1	52.9
Resolution range (Å)	46-1.40 (1.48-1.40)	58-1.40 (1.48-1.40)	52-1.70 (1.79-1.70)	34-2.21 (2.33-2.21)
Completeness (%)	100 (100)	91.5 (91.5)	97.9 (97.9)	65.0 (65.0)
Unique reflections	83204	76941	81732	12774
Multiplicity	7.2 (7.2)	6.7 (2.8)	2.8 (2.3)	4.6 (4.6)
R_{merge} (%)	11.2 (50.0)	9.0 (39.1)	4.9 (11.6)	6.6 (21.6)
Average I/σ(I)	13.3 (3.0)	15.6 (1.6)	16.7 (6.8)	17.5 (6.9)
Molecules per asymmetric unit	2	2	4	2
Wilson B-factor (Å^2)	10.4	13.5	13.0	29
Refinement				
Resolution range (Å)	40-1.40	58-1.40	45-1.70	23-2.24
R_{fac} (%)	19.0	19.9	19.2	17.7
R_{free} (%)	20.9	22.0	21.0	25.9
Number of atoms				
Protein	2918	2950	5800	2912
Water molecules	566	500	500	194
Ba^{2+}			2	1
Isopropanol		2		5
Acetate		2		
Geometry statistics				
r.m.s.d., bonds (Å)	0.006	0.006	0.007	0.009
r.m.s.d., angles (º)	1.02	1.07	1.051	1.20
Ramachandran plot				
Most favored	89.3	90.1	97.5	97.1
Allowed	10.7	9.9	2.5	2.9

Values in parentheses are for the highest resolution shell.

44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table II

Table II. Hydrogen bonding interactions and salt bridges identified in the interface between monomers.

Atoms subunit A	Atoms subunit B	Distance (Å)
Asp56 OD2	His89 NE2[b]	3.99 (4.00)[c]
Thr61 OG1[a]	Glu85 OE1	2.75 (2.70)
Lys66 NZ[a]	Asp80 OD1	2.86 (2.85)
Lys66 NZ[a]	His89 O	3.18 (3.15)
Lys66 NZ[a]	Ala78 O	3.02 (3.00)
Ser 68 OG	Thr91 OG1[a]	2.78 (2.78)
Gly74 O	Lys128 NZ[a]	3.51 (3.55)
Asp76 OD1	Lys128 NZ[b]	3.30 (3.35)
Asp76 OD2	Lys128 NZ[b]	2.85 (2.92)

[a]Hydrogen bond donor.
[b]The cation of the salt bridge.
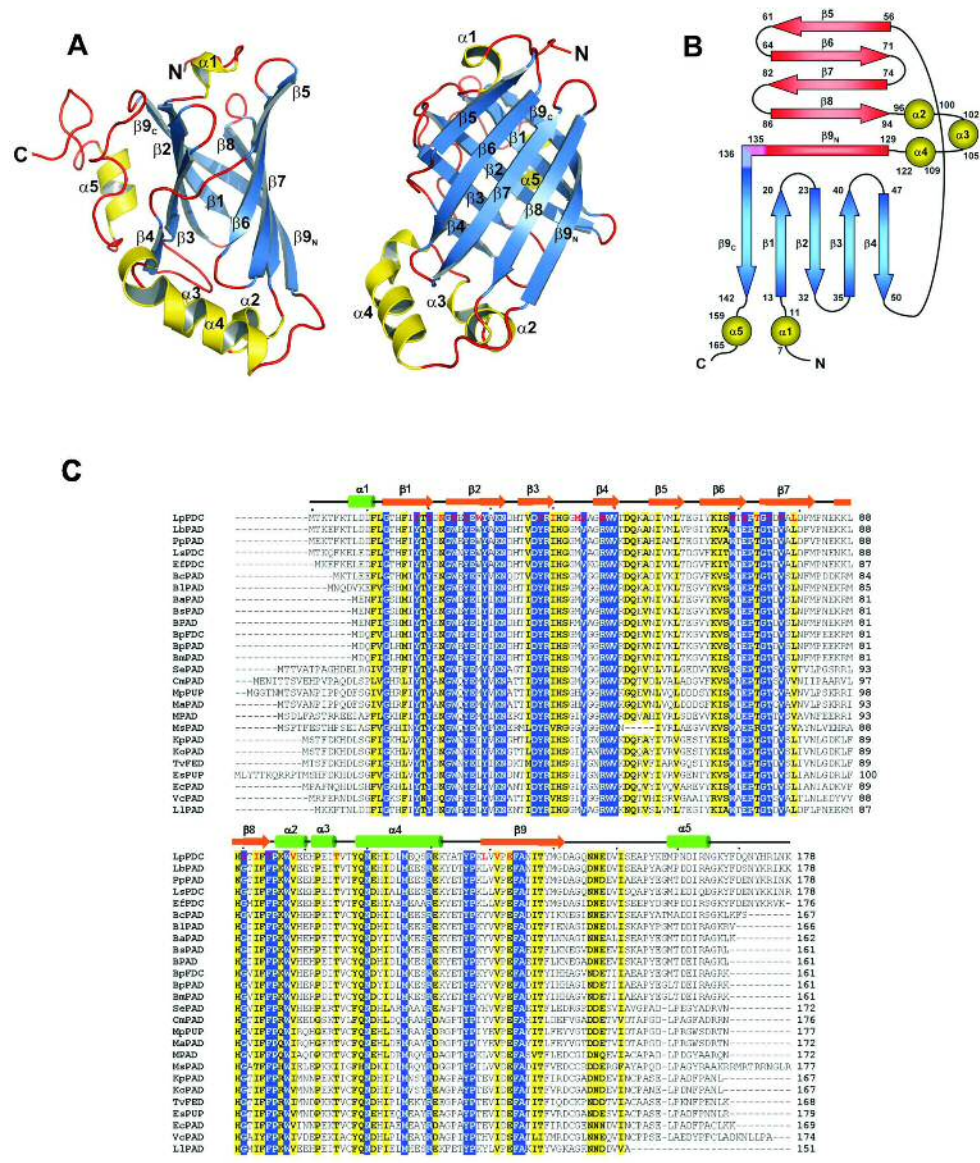[c]Distance between the symmetry-related residues is indicated in parenthesis.

Figure 1. Overall structure of LpPDC and amino acid sequence alignments. (a) Ribbon diagrams of a subunit of LpPDC (residues 2-176). Two orthogonal close-up views are shown. Strands making up the β-barrel are shown in blue, helices in yellow and loops in red. (b) Topology plot of a subunit of LpPDC. β-sheet 1 is indicated in blue and β-sheet 2 in red; helices are depicted as yellow circles. Numbers indicate the first and the last residues of the secondary structure element. (c) Amino acid sequence alignments of LpPDC and other phenolic acid decarboxylases. The regular secondary structure elements in L. plantarum LpPDC are indicated by green boxes (α-helix) and orange arrows (β-strand). LpPDC: Lactobacillus plantarum PDC (Q88RY7); LbPAD: Lactobacillus brevis PAD (Q03TU3); PpPAD: Pediococcus pentosaceus PAD (Q9F3X2); LsPDC: Lactobacillus sakei PDC (Q38UX6); EfPDC: Enterococcus faecium PDC (Q3Y2T7); BcPAD: Bacillus coagulans PAD (A2U705); BlPAD: Bacillus licheniformis PAD (Q65FC9); BaPAD: Bacillus amyloliquefaciens PAD (A7Z928); BsPAD: Bacillus subtilis PAD (007006); BPAD: Bacillus sp. PAD (Q8KNX7); BpFDC: Bacillus pumilus FDC (Q45361); BpPAD: Bacillus pumilus PAD (A8FAY2); BmPAD: Bacillus mesentericus PAD

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

(Q9EXR7); SePAD: Saccharopolyspora erythraea PAD (A4FMV0); CmPAD: Clavibacter michiganensis PAD (A5CN25); MpPUP: Mycobacterium paratuberculosis Putative Uncharacterized Protein (Q743A0); MaPAD: Mycobacterium avium PAD (A0QB49); MPAD: Methylobacterium sp. PAD (B0UN01); MsPAD: Methylocella silvestris PAD (B1THI6); KpPAD: Klebsiella pneumoniae PAD (A6TFA1); KoPAD: Klebsiella oxytoca PAD (A6BMM9); TvFED: Tricomonas vaginalis Ferulate decarboxylase (A2F476); EsPUP: Enterobacter sakazakii Putative Uncharacterized Protein (A7MLR8); EcPAD: Erwinia catovora PAD (Q6DB32); VcPAD: Vibrio cholerae PAD (Q9KPX2); LlPAD: Lactococcus lactis PAD (A2RN76). In the alignment, positions identical in all sequences are marked with a blue background and white characters, and highly conserved positions are marked with a yellow background. Red characters in the LpPDC sequence correspond to residues facing the internal cavity of the enzyme.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
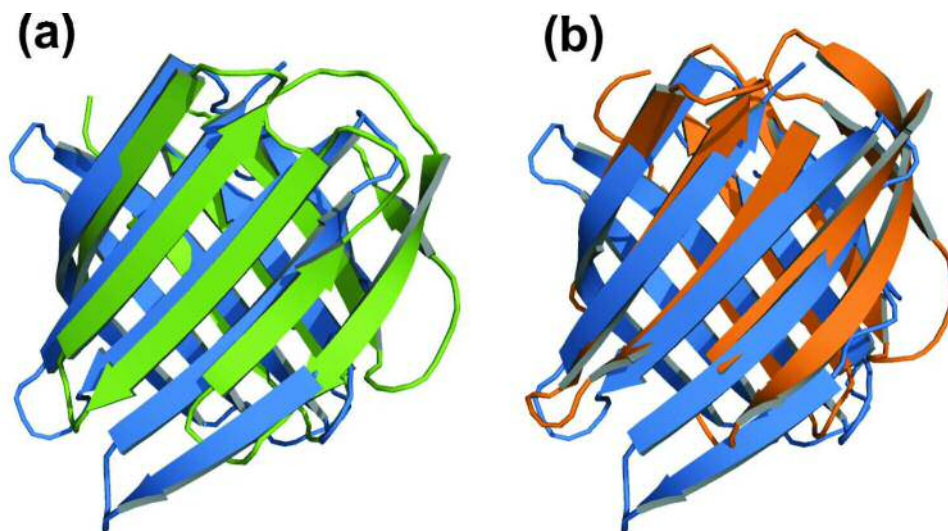42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 2. Superposition of the LpPDC subunit with FABPs. The ß-barrel of the wild-type LpPDC subunit is shown in blue as ribbon model and superimposed with: (a) the ß-barrel of fatty-acid binding protein from toad liver (ribbon model in green; PDB entry 1P6P; Z-score= 7.2); and (b) the ß-barrel of cellular retinoic acid-binding protein 2 (ribbon model in orange; PDB entry 2G7B; Z-score= 6.7).
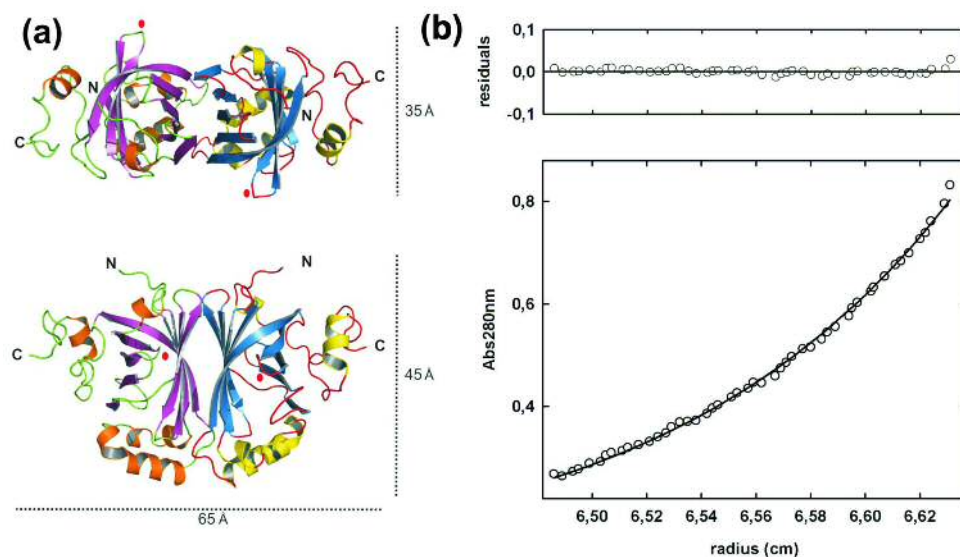
Figure 3. Quaternary structure of LpPDC in the crystal and in solution. (a) Dimeric assembly of LpPDC. Each subunit is coloured according to secondary structure. Two orthogonal views are shown and the dimensions are indicated. The loop ß1-ß2 at the entrance of the internal cavity is indicated with a red circle. This loop is disordered in the electron density of the native LpPDC. (b) Analytical ultracentrifugation sedimentation equilibrium analysis of LpPDC (11.35 µM) at 16,000 rpm. Absorbance at 280 nm is plotted against the radial position from the centre of the rotor. The best fit to the data set (solid line curve) is a unique species of weight molecular mass of 46.0 kDa. Residuals from this fit are shown in the top panel.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
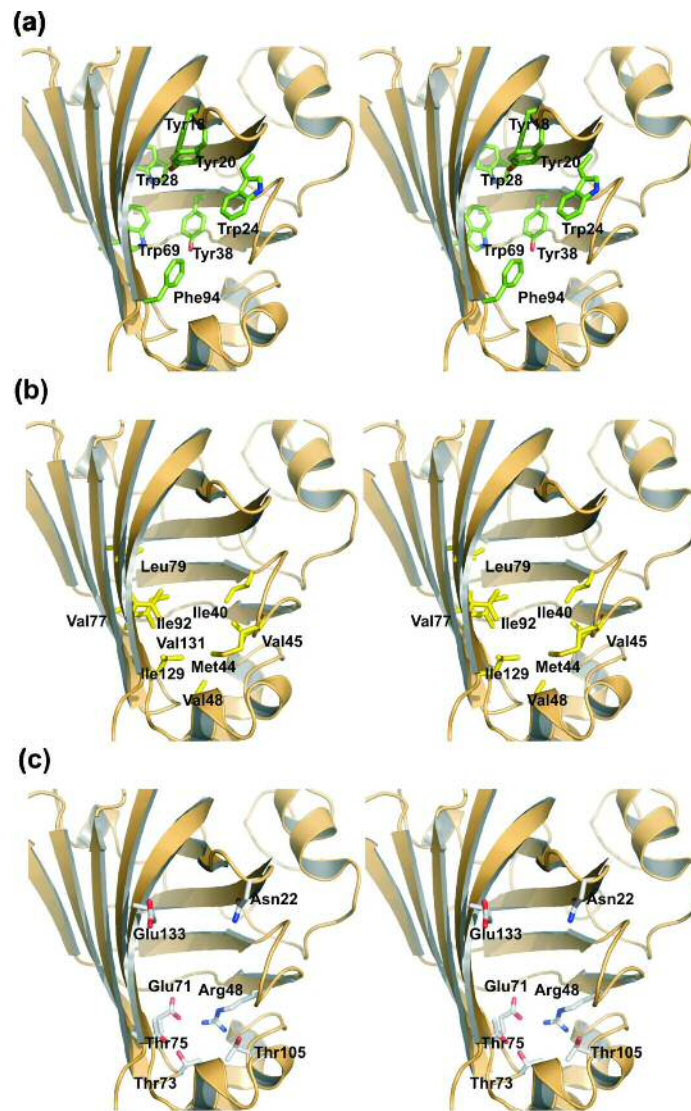46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 4. Structural analysis of the internal cavity of LpPDC. The amino acid residues facing the internal cavity of LpPDC determined by the CASTp server46 have been classified as aromatic (a), aliphatic (b), and polar (c). In the three panels, amino acid side chains are shown in ball-and-stick representation; green for aromatic, yellow for aliphatic and grey for polar residues, respectively.
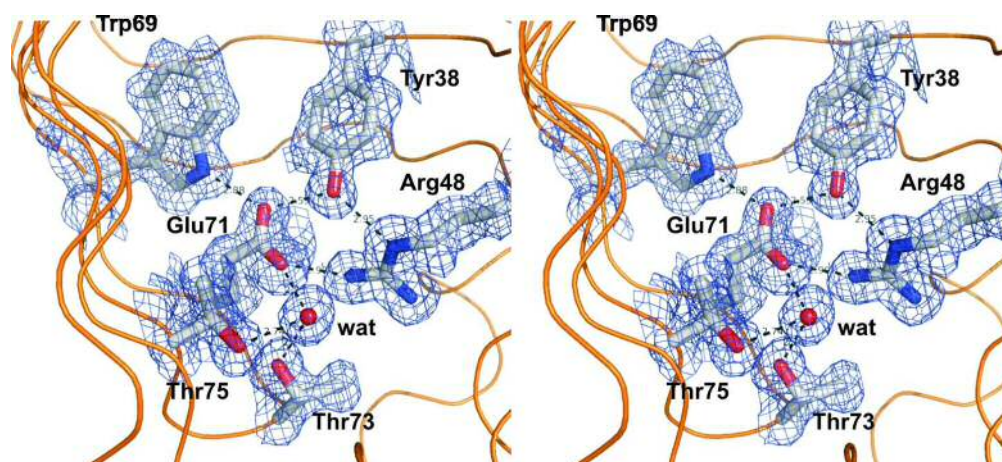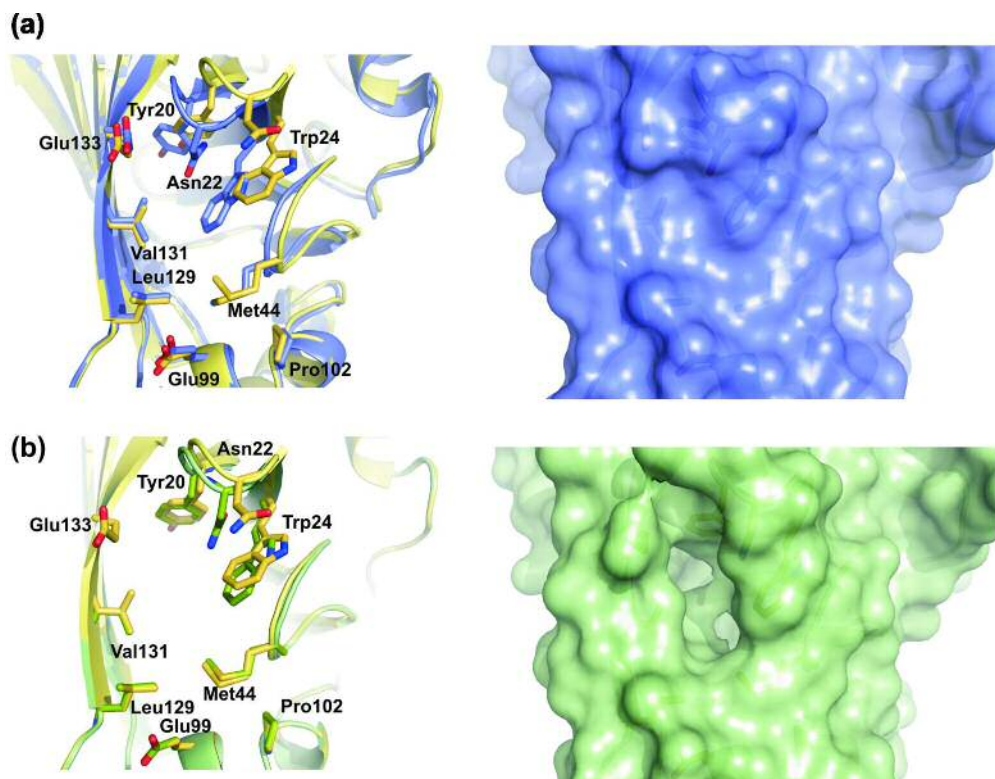
Figure 5. Interactions between polar residues within the wild-type LpPDC cavity. Stereo view representation of the polar region of the LpPDC cavity. Representative side chains are shown as sticks; oxygen atoms are red and nitrogen atoms are blue. A water molecule is shown as a red ball. An Fo – Fc omit map contoured at the 3-σ level is shown in blue. Potential hydrogen bonds are depicted as dotted lines. The map has been calculated with the CNS55 suite.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 6. Open and closed conformational states of LpPDC. The entrance region of wild-type LpPDC is shown in yellow as ribbon model and superimposed with: (a) the R48Q mutant LpPDC structure (blue) and (b) the Y20F mutant LpPDC structure (green). The right panels show close-up views of the molecular surfaces of the R48Q (upper panel) and Y20F (lower panel), respectively. The same orientation is shown in the panels. Side chains of residues that constitute the entrance regions are shown as sticks (color code is as in the ribbon models).
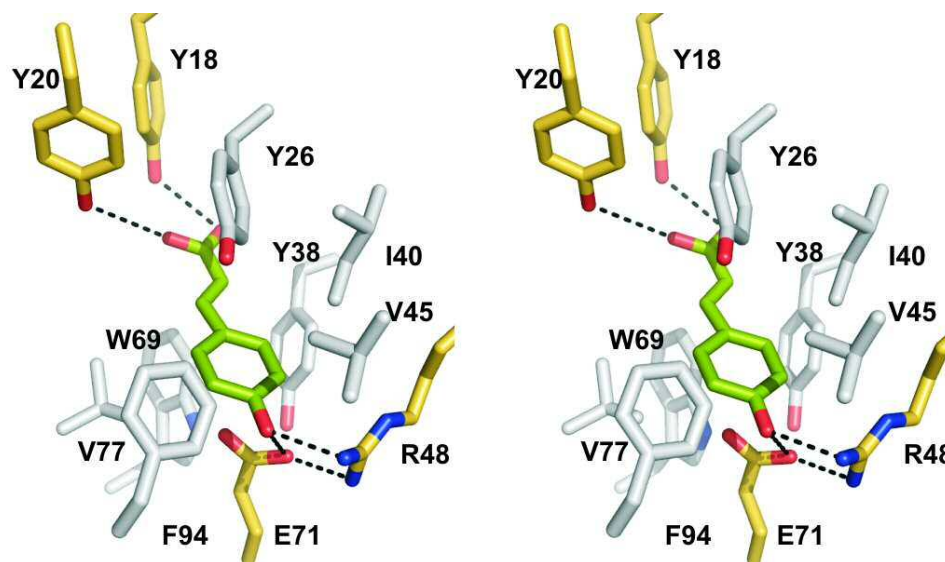
Figure 7. In silico model for the complex of LpPDC and p-coumaric acid. Stereo view representation of the p-coumaric acid binding site of LpPDC (see text for further details). Residues potentially involved in hydrogen bonding interactions with p-coumaric acid (Tyr18, Tyr20, Arg48, and Glu71) are displayed in yellow as stick model. Potential hydrogen bonds (black broken lines) include interactions between the p-hydroxyl group of the substrate and functionally relevant residues Glu71 and Arg48, and also between the carboxyl moiety and hydroxyl groups from Tyr18 and Tyr20. Van der Waals interactions are significant in the binding of the substrate. Residues forming the hydrophobic pocket (Tyr26, Tyr38, Ile40, Val45, Trp69, V77 and Phe94) are depicted in grey.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
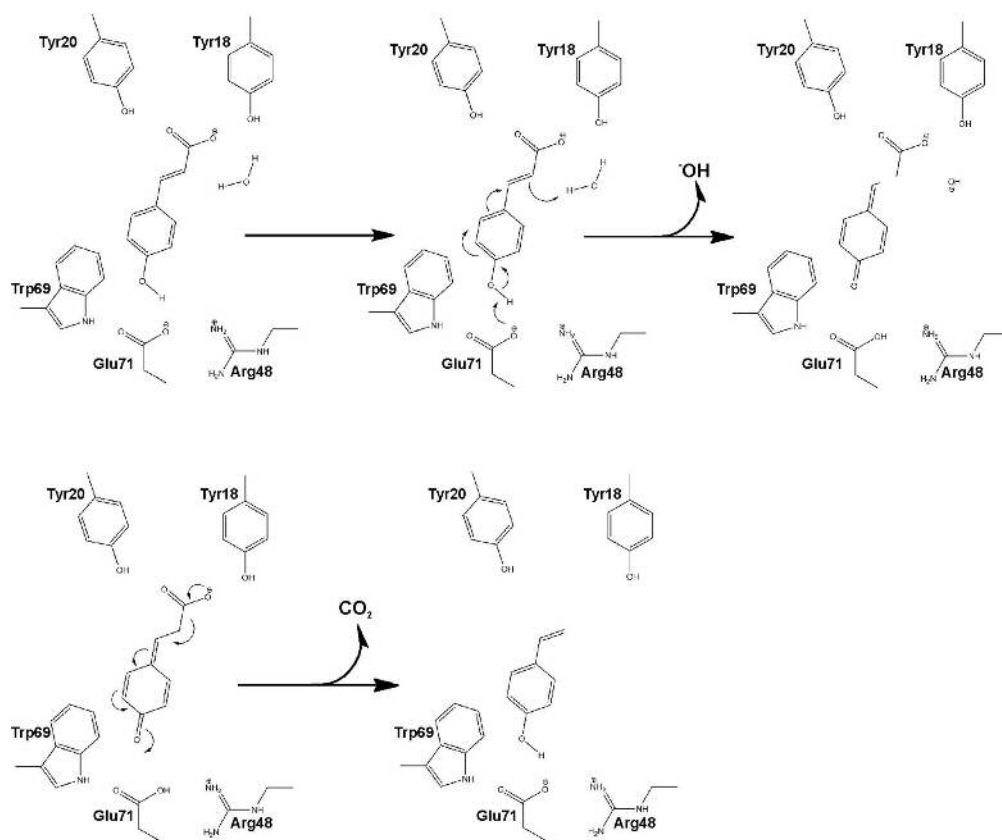51
52
53
54
55
56
57
58
59
60

Figure 8. Proposed catalytic mechanism of decarboxylation by LpPDC. The mechanistic proposal is based on inferences drawn from the crystallographic, functional and in silico results presented here and by analogy to the proposed chemical mechanism previously proposed for PAD from Klebsiella oxytoca.51