Structural basis for independent origins of new catalytic machineries in plant AAAD proteins

Michael P. Torrens-Spence[1], Ying-Chih Chiang[2], Tyler Smith[1,4], Maria A. Vicent[1,3], Yi Wang[2], and Jing-Ke Weng[1,4]*

[1]Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA.

[2]Department of Physics, the Chinese University of Hong Kong, Shatin, N.T., Hong Kong.

[3]Department of Biology, Williams College, Williamstown, Massachusetts 01267, USA.

[4]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

*Corresponding Author:

Jing-Ke Weng
Whitehead Institute for Biomedical Research
455 Main Street, Cambridge, MA, 02142, USA
Email: wengj@wi.mit.edu
Tel.: +1 617 324 4921

Abstract

A remarkable feature of rapidly evolving specialized metabolic systems is the origin of new catalytic machineries from progenitor enzymes catalyzing alternative reactions. Divergent evolution of the pyridoxal 5'-phosphate (PLP)-dependent aromatic amino acid decarboxylases (AAADs) in plants not only resulted in paralogous AAADs with specific substrate preference, but also repeatedly yielded evolutionarily new aldehyde synthases with additional oxidative deamination activity. The molecular mechanisms underlying such evolutionary development were unknown. Here, we report the structural and biochemical characterization of a number of functionally distinct plant AAAD-family enzymes. We resolve how independently evolved aldehyde synthases employ disparate molecular strategies to construct an aldehyde-producing catalytic machinery. This work highlights the pliability of an ancestral enzyme fold that furnishes access to new catalytic mechanisms with only a few mutations.

Introduction

Plants are sessile organisms that produce a dazzling array of specialized metabolites as their unique adaptive strategy to cope with a multitude of abiotic and biotic stresses. Underlying plants' remarkable chemodiversity is their rapidly evolving metabolic systems that support specialized metabolite biosynthesis[1]. Genome-wide analysis of metabolic evolution across major green plant lineages has revealed pervasive and progressive expansion of discrete specialized metabolic enzyme families, where new catalysts continued to emerge predominantly through gene duplication events followed by subfunctionalization or neofunctionalization[2]. To arrive at evolutionarily new enzymatic activities, most enzymes explore mutations that alter substrate specificity and/or product diversity without changes in the ancestral catalytic machineries. In the meantime, rarer cases have also be described where adaptive mutations occur in a progenitor protein fold—either non-catalytic or catalytic—that gave rise to new catalytic machineries and novel chemistry[3,4]. Understanding the structural and mechanistic bases for these cases provides key insights into the evolutionary transitions between functional disparate enzymes that share common ancestry.

Aromatic amino acid decarboxylases (AAADs) are an ancient group of pyridoxal 5'-phosphate (PLP)-dependent enzymes with primary functions associated with amino acid metabolism. Mammals possess a single AAAD, DOPA decarboxylase (DDC), responsible for the synthesis of several key monoamine neurotransmitters, such as dopamine and serotonin, from their respective amino acid precursors[5]. In contrast, the AAAD family in plants has undergone extensive radiation to yield diverse paralogous enzymes with functional variation in both substrate preference and catalytic mechanism[6] (Fig. 1a-b, Fig. S1-3 and Supplementary Note 1). While tryptophan decarboxylases (TDCs) and tyrosine decarboxylases (TyDCs) are canonical

AAADs that supply the aromatic arylalkylamine precursors for indole alkaloid and benzylisoquinoline alkaloid (BIA) biosynthesis respectively, phenylacetaldehyde synthases (PAASs) and 4-hydroxyphenylacetaldehyde (4HPAA) synthases (4HPAASs) are evolutionarily derived from progenitor AAADs, and catalyze decarboxylation-oxidative deamination of phenylalanine and tyrosine respectively to produce the corresponding phenolic aldehydes[7–10]. The neofunctionalization of PAASs and 4HPAASs is key to the emergence of numerous important specialized metabolic traits in plants, e.g. the volatile floral scents such as phenylacetaldehyde, phenethyl alcohol and phenethyl acetate, and the medicinally relevant tyrosol-derived metabolites such as salidroside[9,10]. Despite their importance to many aromatic-amino-acid-derived specialized metabolic pathways in plants, no crystal structures have been reported for any of the plant AAAD family proteins to date. The molecular mechanisms underlying how diverse plant AAAD family enzymes acquired their unique catalytic functions remain unknown.

Results

The crystal structures of four divergent plant AAAD proteins

To understand the structural basis for the functional divergence of the AAAD family in plants, we determined the x-ray crystal structures of *Catharanthus roseus* TDC (*Cr*TDC), *Papaver somniferum* TyDC (*Ps*TyDC), *Arabidopsis thaliana* PAAS (*At*PAAS), and *Rhodiola rosea* 4HPAAS (*Rr*4HPAAS), presenting the first set of structures for plant AAADs (Fig. 1C, Table S1). All four enzymes purify and pack in the crystal lattice as homodimers, which share highly similar overall structure of the type II PLP-dependent decarboxylase fold (Table S2)[11]. As represented by the L-tryptophan-bound *Cr*TDC structure, each monomer is composed of three structural segments (Supplemental note 2 and Fig. S4a-b). The N-terminal segment (*Cr*TDC$^{1\text{-}119}$) comprises three antiparallel helices that interlock with the reciprocal helices of the other monomer to form the primary hydrophobic dimer interface (Fig. 1d and S5a-b). The middle (*Cr*TDC$^{120\text{-}386}$) and the C-terminal (*Cr*TDC$^{387\text{-}497}$) segments harbor the enzyme active site at the dimer interface, featuring a conserved Lys^{319} with its ζ-amino group covalently linked to the coenzyme PLP to form an internal aldimine in its resting state (Fig. 1e and Fig. S5c). The phosphate moiety of the lysine-pyridoxal-5'-phosphate (LLP) is coordinated by Thr^{167}, Ser^{168}, and Thr^{369} through hydrogen bonding, while the pyridine-ring nitrogen forms a salt bridge with the side chain of a conserved Asp^{287}, supporting its role in stabilizing the carbanionic intermediate of the PLP external aldimine[12] (Fig. 2a).

The substrate-binding pocket of *Cr*TDC is composed of a set of conserved residues, as well as three variable residues (Ala^{103}, Thr^{369} and Gly^{370}) that seem to have undergone sequence divergence across different AAAD clades (Fig. 2b-c, Fig. S6). Gly^{370} is of particular interest as this residue is exclusively conserved as a glycine within the TDC clade, but substituted to a serine or threonine within the TyDC clade (Fig. S6)[13]. Structural comparison of the L-tryptophan-bound *Cr*TDC and the L-tyrosine-bound *Ps*TyDC indicates that the G370S substitution dictates the size and shape of the substrate-binding pocket to favor indolic or phenolic amino acid binding in *Cr*TDC and *Ps*TyDC, respectively (Fig. 2d). To test this, we expressed the wild-type or the G370S mutant of *Cr*TDC in *Saccharomyces cerevisiae*, and profiled the ectopically produced aromatic monoamines. The transgenic yeast expressing the wild-type *Cr*TDC predominantly produces tryptamine, whereas yeast expressing the *Cr*TDC^{G370S} mutant produces increased level of tyramine at the expense of tryptamine, suggesting a broadened substrate specificity towards tyrosine due to the G370S mutation (Fig. 2e-f and Fig.

S7). Together, these results suggest that the phylogenetically restricted sequence variants at position 370 were likely selected to serve the respective biochemical functions of the TyDC and TDC clades after they separated from the basal AAAD clade.

The catalytic cycle of canonical type II PLP-dependent AAAD

PLP is a versatile coenzyme employed by various enzyme classes to catalyze a selection of reactions, including elimination, addition, racemization, aldol cleavage, in addition to amino acid decarboxylation[14]. The active site of canonical AAADs constrains the reactivity of PLP to specifically drive decarboxylation chemistry[14]. Despite the availability of several animal DDC crystal structures, several aspects of the full catalytic cycle of the type II PLP-dependent decarboxylases remain speculative[5,15–17]. Comparative analysis of the four paralogous plant AAAD enzymes shed light on the previously unresolved catalytic mechanisms of the canonical AAAD reaction as well as the derived aldehyde synthase activities through parallel molecular evolution. Evident from the L-tryptophan-bound *Cr*TDC, the L-tyrosine-bound *Ps*TyDC, and the L-phenylalanine-bound *At*PAAS structures, the aromatic amino acid substrate first enters the active site, and is oriented to present its labile α-carbon-carboxyl bond perpendicular to the pyridine ring of the internal aldimine LLP (Fig. S8). This triggers the subsequent transaldimination that conjugates the α-amino group of the incoming substrate to PLP through a Schiff base to yield the external aldimine, captured by one of the active sites of the L-tyrosine-bound *Ps*TyDC structure (Fig. 3a-b and Fig. S9). The resultant external aldimine subsequently loses the α-carboxyl group as CO_2 to generate a quinoid intermediate (Fig. 3c, reaction step 1). Through an as yet unknown mechanism, the nucleophilic carbanion of the quinoid intermediate is protonated to yield the monoamine product and a regenerated LLP (Fig. 3b, reaction steps 2-3)[12].

Structural features of two catalytic loops dictate divergent catalytic fates

Although mechanistic and structural studies of various AAADs have identified a large loop region that harbors a highly conserved tyrosine residue essential for catalysis[16], this loop is invisible in the electron density map of all the animal DDC structures solved to date. Our crystallographic datasets of various plant AAAD proteins provide electron-density support for this large loop in different conformations relative to the active site (Fig. 3d-e and Supplementary Note 3). In the *Cr*TDC structure, for instance, the large loop (*Cr*TDC$^{342\text{-}361\text{-}A}$) of monomer A adopts an open conformation, revealing a solvent-exposed active site. On the contrary, this loop undergoes a crankshaft rotation in the *Ps*TyDC structure to close the entrance, and seals the active site from solvent access. In many previously characterized plant aldehyde synthases, a Tyr-to-Phe substitution on this large loop is associated with the conversion of the canonical decarboxylation activity to the decarboxylation-deamination activity that yields arylalkyladehyde products[18]. These observations implicate that the conserved tyrosine on the large loop serves as the catalytic acid responsible for the reprotonation of the α-carbon in the production of arylalkylamine. In *At*PAAS, the naturally occurring Y338F substitution abolishes the catalytic acid, instead enabling molecular oxygen to attack the carbanion of the quinoid intermediate to generate a peroxide intermediate, which subsequently decomposes into the arylalkyladehyde, carbon dioxide, ammonia, and hydrogen peroxide products (Fig. 3c, reaction steps 4-5). In addition to precluding carbanion protonation, the Tyr-to-Phe substitution may also disrupt the ionic interactions necessary for proper loop closure and active-site desolvation. Substitutions at this location occur naturally as valine, leucine, isoleucine, methionine and phenylalanine within the basal and TyDC clades, and are found in AAAD sequences beyond land plants (Fig. S2 and Fig. S6). Evidently, substitution of the conserved large-loop tyrosine with hydrophobic residues imparts aldehyde synthase chemistry as the characterized *Rosa hybrid*, *Petunia hybrida,* and *Cicer arietinum* PAASs possess phenylalanine, valine, and isoleucine at this position, respectively[18–20].

In our *Ps*TyDC structure, the closed conformation of the large loop places the *p*-hydroxy group of Tyr$^{350\text{-}A}$ in close vicinity with the Cα of the L-tyrosine substrate as well as the imidazole ring of a neighboring His$^{205\text{-}B}$ situated on a small loop from the opposite monomer (*Cr*TDC$^{202\text{-}205\text{-}B}$) (Fig. 3f). In this closed conformation, the histidine imidazole forms a pi-stacking interaction with the aromatic PLP-pyridine ring. Such pi stacking between PLP and an active-site aromatic residue is a common feature among AAADs as well as the broader α-aspartate aminotransferase superfamily[21,22]. Conversely, in the *Cr*TDC structure where the large loop is in open conformation, this small loop (*Cr*TDC$^{202\text{-}205\text{-}B}$) rotates away from LLP and the bound substrate (Fig. 3e). The vicinity of this conserved histidine to the active site and its dynamic nature suggest its potential catalytic role in carbanion protonation; however, this histidine can not function directly as the catalytic acid, because the pH dependence of AAADs suggests that the deprotonated histidine is the active form[15,23,24]. Similar Tyr-His side-chain interaction has been described in several other catalytic systems involving proton transfer[25]. We therefore propose a catalytic mechanism where the large-loop Tyr$^{350\text{-}A}$ and small-loop His$^{205\text{-}B}$ act cooperatively to facilitate the protonation of the quinonoid intermediate, resulting in the canonical AAAD reaction outcome (Fig. 3c). To test this, we produced and characterized the *Ps*TyDC^{H205N} mutant. Interestingly, the *Ps*TyDC^{H205N} mutant displays both the canonical AAAD activity and the aldehyde synthase activity in vitro (Fig. 4a). Furthermore, transgenic *S. cerevisiae* expressing the *Ps*TyDC^{H205N} mutant produces both tyramine and tyrosol (Fig. S10-11). These results suggest that carbanion protonation is assisted by the conserved small-loop histidine. Although decarboxylation can still occur when this histidine is mutated, a significant fraction of the intermediate undergoes the alternative oxidative deamination of the aldehyde synthase chemistry.

Molecular dynamics simulations reveal the dynamic nature of two catalytic loops

Considering the crystallographic observation of the alternative conformations of the two loops that harbor the potential key catalytic residues Tyr$^{348\text{-A}}$ and His$^{203\text{-B}}$ (numbering according to *Cr*TDC), we sought to examine the flexibility and possible cooperativity of the two loops by molecular dynamics (MD) simulation in *Cr*TDC. We began with 36 sets of 100-ns simulations on six *Cr*TDC systems with LLP and the substrate L-tryptophan in different protonation states (Fig. S12 and Supplementary Note 4). These simulations revealed considerable flexibility of both loops, with one simulation capturing a dramatic closing motion of the open large loop. Upon extending this simulation to 550 ns (Fig. 3g, Video S1), the large loop was found to reach a semi-closed state characterized by a minimal C_α RMSD of 4.3Å with respect to the modeled *Cr*TDC closed-state structure (Supplementary Note 5). The catalytic Tyr$^{348\text{-B}}$ was found to form stacking interactions with His$^{203\text{-A}}$, with a minimal distance of approximatley 2 Å between the two residues (Fig. S13). Interestingly, a short helix (residues 346-350) that unfolded at the beginning of this simulation appeared to ‘unlock’ the large loop from its open state. To further examine the correlation between its structure and large loop conformation, we artificially unfolded this short helix and initiated 72 sets of 50-ns simulations from the resulting structure. Structures similar to that revealed by Fig. 3g were observed in all six systems (Fig. S14c) with minimal C_α RMSD ranging from ~5 to 8 Å with respect to the modeled *Cr*TDC closed-state structure. These results suggest that the initial closing motion of the large loop is independent of coenzyme and substrate protonation states and that the unfolding of the aforementioned short helix can significantly accelerate such motion. This is further substantiated by three sets of 600-ns simulations of *Cr*TDC in an apo state with neither PLP nor L-tryptophan present (Fig. S15 and Supplementary Note 4). We should note, however, that the fully closed state of *Cr*TDC was not achieved in our sub-microsecond simulations. For instance, in the trajectory shown in Fig. 3g and Video S1, L-tryptophan left the active site at around t=526 ns, shortly after which the simulation was terminated. Overall, while the transition from the semi-closed to the fully closed state can be expected to occur beyond the sub-microsecond timescale, the MD results support our hypothesis that Tyr$^{348\text{-}B}$ can form close contact with His$^{203\text{-}A}$, readying the latter residue to direct and stabilize proton transfer from the former residue to the C_α quinoid intermediate (Fig. S13b).

Aldehyde synthases from Myrtle evolved via an alternative molecular mechanism

The rapid expansion of the plant genomic resources in recent years enabled us to mine AAAD paralogs with potential new functions, bearing in mind those key residues involved in substrate selectivity and catalytic mechanism. A query of all AAAD sequences within the 1KP transcriptomes[26] identified a number of myrtle-family AAAD homologs, derived from the TyDC clade, with the conserved small-loop histidine substituted to asparagine (Fig. 1a). Since the same mutation impairs the canonical AAAD activity in the context of *Ps*TyDC (Fig. 4a), we hypothesized that this unusual clade of Myrtle AAAD proteins might have evolved new functions. To test this, we produced recombinant protein for a representative enzyme from *Eucalyptus grandis* (MG786260). Biochemical characterization of this *E. grandis* enzyme (*Eg*PAAS) demonstrates that it is indeed an aldehyde synthase with apparent substrate specificity towards phenylalanine and no detectable decarboxylase activity (Fig. 4b and Fig. S16). Interestingly, phenylacetaldehyde is reportedly a major metabolite in the essential oil and honey of plants from the myrtle family[27–29], consistent with these newly discovered PAAS enzymes in myrtle. Together with the previously characterized aldehyde synthases harboring the large-loop Tyr-to-Phe substitution, the discovery of aldehyde synthase chemistry derived from the naturally occurring small-loop His-to-Asn substitution in myrtle AAADs suggests that nature has independently exploited two alternative molecular strategies to arrive at the evolutionarily new aldehyde synthase.

Metabolic engineering of a shortened BIA biosynthetic pathway in yeast with 4HPAAS

In addition to their roles in floral volatile and tyrosol anabolism, amino-acid-derived arylalkyladehydes have also been demonstrated in indole-3-acetic acid (IAA, auxin) and BIA biosynthesis[30,31]. In particular, 4HPAA, the product of *Rr*4HPAAS or the *Ps*TyDC^{Y350F} mutant, is a key intermediate in the biosynthesis of BIAs, which include the narcotic analgesics codeine and morphine[31]. In the currently proposed BIA pathway, 4HPAA is generated from L-tyrosine through two consecutive transamination and decarboxylation steps catalyzed by an L-tyrosine aminotransferase (TAT) and an unidentified 4-hydroxyphenylpyruvate decarboxylase (4HPPDC), respectively[31]. 4HPAA is then stereoselectively condensed with dopamine by the norcoclaurine synthase (NCS) to produce the BIA intermediate (S)-norcoclaurine. To demonstrate the utility of 4HPAAS in reconstituting BIA biosynthesis, we engineered yeast strains to produce (S)-norcoclaurine through a shortened biosynthetic route, where 4HPAA is biosynthesized directly from L-tyrosine by either *Rr*4HPAAS or *Ps*TyDC^{Y350F}, bypassing the necessity for TAT and 4HPPDC (Fig. 4c-d). The control experiment using the wild-type *Ps*TyDC yielded only tyramine (Fig. 4c-d, Fig. S18). This engineering excise thus demonstrates how minimal amino acid substitutions in key metabolic enzymes can dictate the fate of specialized metabolic pathway evolution, and the structure-function understanding of plant

AAADs can provide new tools for engineering heterologous production of high-value aromatic-amino-acid derivatives in microbial hosts.

Discussion

Co-opting progenitor enzymes to synthesize novel and adaptive metabolites is a universal mechanism underscoring metabolic evolution[32]. Most specialized metabolic enzymes present in extant plants evolved through the recruitment of malleable ancestral enzyme folds followed by neofunctionalization of substrate specificity, product diversity, or, in much rarer cases, alternative catalytic mechanisms[33–36]. The functional divergence of the plant AAAD family illustrates all these evolutionary mechanisms. For example, early radiation of the substrate-permissive basal AAADs gave rise to the bifurcation of indolic- and phenolic-selective TDC and TyDC clades by fixing specific mutations at the substrate-binding pocket. Moreover, tinkering the ancestral AAAD catalytic machinery by mutating either the catalytic acid tyrosine situated at the large loop or its neighboring histidine that helps facilitate the proton transfer led to the emergence of the new aromatic-acetaldehyde-forming activities. We note that laboratory mutation of the catalytic tyrosine to phenylalanine in *Ps*TDC fully converted it to an aldehyde synthase, whereas the mutation of the catalytic histidine to asparagine in *Ps*TDC, which mimicks the natural substitution observed in *Eg*PAAS, turned the mutant enzyme into a bifunctional enzyme that harbors both decarboxylase and aldehyde synthase activities. This suggests that additional adaptive mutations must have followed the initial His-to-Asn mutation in Myrtle aldehyde synthases to polish the newly acquired aldehyde synthase activity by repressing the ancestral decarboxylase activity.

While the aldehyde synthase activity has occurred repeatedly in plant AAADs, it is also noteworthy that the insect alpha-methyldopa resistant (AMD-r) protein, an AAAD-family enzyme responsible for producing 3,4-dihydroxyphenylacetaldehyde necessary for insect soft cuticle formation[37], also contains the small-loop His-to-Asn substitution reminiscent to that of *Eg*PAAS, illustrating repeated parallel evolution of aldehyde synthase chemistry via alternative molecular mechanisms in both plants and animals. Applying the structure-function understanding of the AAAD-family proteins gained from this study further enabled the engineering of a shorten pathway for the production of the BIA presorsor (S)-norcoclaurine in yeast. Not only does this exercise represents an alternative metabolic engineering strategy for microbial production of pharmaceutically relevant alkaloids, the engineered pathway may also impinge on the unresolved plant endogenous benzylisoquinoline alkaloid biosynthesis, e.g. in opium poppy. Collectively, our work illuminates how the structural and mechanistic pliability of the AAAD fold enabled extensive functional divergence within this enzyme family that directs specific aromatic amino acid precursors into a diverse array of specialized metabolic pathways.

Methods

Reagents

L-tryptophan, tryptamine, L-tyrosine, tyramine, tyrosol, L-phenylalanine, phenylacetaldehyde, phenylethyl alcohol, L-3,4-dihydroxyphenylalanine, dopamine, (S)-norcoclaurine, PLP, and sodium borohydride were purchased from Sigma-Aldrich. 4-hydroxyphenylacetaldehyde was purchased from Santa Cruz Biotechnology.

Multiple sequence alignment and phylogenetic analysis

ClustalW2 was used to generate the protein multiple sequence alignments with default settings[38].

The phylogeny was inferred using the Maximum Likelihood method based on the JTT matrix-based model. The bootstrap consensus unrooted trees were inferred from 500 replicates to represent the phylogeny of the analyzed enzyme families. The phylogenetic analysis encompases AAAD family sequences from all Phytozome V12 species in addition previously characterized AAAD proteins and select bacterial sequences[39]. Initial trees for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. All phylogenetic analyses were conducted in MEGA7[40]. ESPript 3.0 was used to display the multiple sequence alignment[41]. Conservation of the active site residues between various AAAD clades was displayed using WebLogo[42].

Plant materials

E. grandis seeds were purchased from Horizon Herbs. Seeds were stratified at 4 °C for three days, and germinated in potting soil. Plants were grown under a 16-h-light/8-h-dark photoperiod at 23 °C in a local greenhouse.

cDNA production and molecular cloning

Leaf tissue of seventy-day-old *E. grandis* plants were harvested for total RNA extraction using the Qiagen's RNeasy Mini Kit (Qiagen). Total RNA for *A. thaliana*, *P. somniferum, C. roseus* and *R. rosea* were produced as previously reported [43,44]. First-strand cDNAs of were synthesized by RT-PCR using total RNA sample as template. The coding sequences (CDS) of candidate genes were amplified from cDNAs by PCR using gene-specific primers (Table S3). PpDDC (NP_744697.1), PsNCS2 (AKH61498), and BvTyH (AJD87473) were synthesized as gBlocks (IDT) with *S. cerevisiae* codon optimization. Gibson assembly was used to ligate the *Cr*TDC, *Ps*TyDC, *At*PAAS and *Rr*4HPAAS PCR amplicons into pHis8-4, a bacterial expression vector containing an N-terminal 8xHis tag followed by a tobacco etch virus (TEV) cleavage site for recombinant protein production. *Eg*PAAS was alternatively cloned though Gibson assembly into pTYB12, a commercially available N-terminal intein/chitin domain fusion vector designed for affinity chromatography purification. Ectopic expression of various AAADs in *S. cerevisiae* was achieved through the use of p423TEF, a 2 μ plasmid with the HIS3 auxotrophic growth marker for constitutive expression[45]. Golden Gate assembly was used generate the multi gene vectors designed for (S)-norcoclaurine production in *S. cerevisiae*[46]. PCR amplicons or gBlocks were ligated into the entry vector pYTK001 and subsequently assembled into 2 μ, pTDH3, tTDH1, *HIS3* multigene vectors for constitutive expression in *S. cerevisiae*[46]. A second multigene vector, containing the *S. cerevisiae* tyrosine metabolism feedback resistant mutants ARO4^{K229L} and ARO7^{G141S}, was additionally used to boost tyrosine flux in (S)-norcoclaurine producing *S. cerevisiae* lines. This vector was built using 2 μ, pTDH3, tTDH1 and LEU2 as previously described[44].

Recombinant protein production and purification

BL21(DE3) *E. coli* containing the pHis8-4 or pTYB12 AAAD constructs were grown in terrific broth (TB) at 37 °C to OD^{600} of 0.9 and induced with 0.15 mM isopropyl-β-D-thiogalactoside (IPTG). The cultures were cooled to 18 °C and and shaken for an additional 20 hours. Cells were harvested by centrifugation, washed with phosphate buffered saline (PBS) (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4 and 1.8 mM KH_2PO_4), resuspended in 150 mL of lysis buffer (50 mM Tris pH 8.0, 0.5 M NaCl, 20 mM imidazole, and 0.5 mM dithiothreitol (DTT)), and lysed with five passes through a M-110L microfluidizer (Microfluidics). The resulting crude protein lysate from the pHis8-4 cultures were clarified by centrifugation prior to Qiagen Ni-NTA gravity flow chromatographic purification. After loading the clarified lysate, His-tagged recombinant protein-bound Ni-NTA resin was washed with 20 column volumes of lysis buffer, and eluted with 1 column volume of elution buffer (50 mM Tris pH 8.0, 0.5 M NaCl, 250 mM imidazole and 0.5mM DTT). 1 mg of His-tagged TEV protease was added to the eluted protein, followed by dialysis at 4 °C for 16 h in dialysis buffer (50 mM Tris pH 8.0, 0.1 M NaCl, 20 mM imidazole and 2 mM DTT). After dialysis, the protein solutions were passed through Ni-NTA resin to remove uncleaved protein and His-tagged TEV. The *Eg*PAAS frozen cell pellet was homogenized and in a smilary manner abet in a imidazole free lysis buffer. The resulting crude protein lysate was applied to a column packed with chitin beads, washed with 1 L of buffer and subsequently hydrolyzed under reducing conditions as per the manufacturer's instructions. Recombinant proteins were further purified by gel filtration on a fast protein liquid chromatography (FPLC) system (GE Healthcare Life Sciences). The principle peaks were collected, verified for molecular weight by SDS-PAGE, stored in storage buffer (20 mM Tris pH 8.0, 25 mM NaCl, 200 µM PLP and 0.5 mM DTT) at a protein concentration of 10 mg/mL and flash frozen for subsequent investigation.

Protein Crystallization and Structural Determination

Crystals for the various plant AAADs were grown at 4 °C by a hanging-drop vapor diffusion method with the drop containing 0.9 µl of protein sample and 0.9 µl of reservoir solution at a reservoir solution volume of 500 µl. The crystallization buffer for the *At*PAAS contained 0.16 M ammonium sulfate 0.8M HEPES:NaOH pH 7.5 and 20%w/v PEG 3350. Crystals were soaked in well solution containing 15 mM L-phenylalanine for six hours and cryogenized with an additional 10% weight/volume ethylene glycol. *Ps*TyDC crystals were formed in 1.2 M ammonium Sulfate 0.1 Bis Tris pH 5.0 and 1% w/v PEG 3350. Crystals were soaked in the presence of 4mM L-tyrosine four twelve hours 12 hours and cryoprotected with an additional

25% weight/volume ethylene glycol. 0.22M calcium chloride and 12% w/v PEG 3350 formed the *Cr*TDC crystals which were subsequently soaked with 10mM L-tryptophan for 16 hours and then cryogenized with an additional 18% weight/volume ethylene glycol. Finally, to form the *Rr*4HPAAS crystals, protein solution was mixed with a reservoir buffer of 0.21 M potassium thiocyanate and 22% w/v PEG 3350. Ligand soaks for this crystal proved unsuccessful and ultimately the crystals were cryoprotected with an additional 13% weight/volume PEG 3350 in the absence of ligand. The structure of *Ps*TyDC was determined first though molecular replacement with the insect DDC structure[47] using Molrep[48].The resulting model was iteratively refined with Refmac 5.2[49] and the manually refined with Coot 0.7.1[50]. Solvent molecules were added and refined with Refmac 5.2 and Coot 0.7.1.

Enzyme assays

The *in vitro* decarboxylation and aldehyde synthase activities of the wild type *Ps*TyDC, *Ps*TyDC^{H204N} and *Ps*TyDC^{Y350} enzymes were assayed in 100 µL of 50 mM Tris, pH 8.0 containing 20 µg of recombinant enzyme, 100 µM PLP and 0.5 mM L-tyrosine. Reactions were incubated at 30 degrees for various time points and subsequently stopped with 200 µL methanol. After clarification, the soluble fraction was analyzed by LC-MS-UV. Chromatographic separation and absorption at 280 nm were performed by an Ultimate 3000 liquid chromatography system (Dionex), equipped with a 150 mm C18 Column (Kinetex 2.6 µm silica core shell C18 100 Å pore, Phenomenex) and coupled to an UltiMate 3000 diode-array detector (DAD) in-line UV-Vis spectrophotometer (Dionex). Compounds were separated through the use of an isocratic mobile phase as previously described[44]. The reduction of aldehyde products was achieved by addition ethanol containing saturating sodium borohydride. The *Eg*PAAS enzyme assays were started by adding 2 µg of recombinant protein into 200 µL of 50mM tris pH 8.0 and 2mM L-phenylalanine. Reactions were incubated for various time points at 30 degrees and reactions were stopped with equal vol of 0.8 formic acid, extracted with 150ul of ethyl acetate and analyzed by gas chromatography–mass spectrometry (GC-MS) as previously described against an analytical phenylacetaldehyde standard[44]. The initial substrate selectivity was measuring through the detection of the hydrogen peroxide co-product using Pierce Quantitative Peroxide Assay Kit (Pierce) and a standard curve of hydrogen peroxide. Reactions were conducted as described using reaction mixtures containing 0.5 mM amino acid substrate concentrations. Triplicate reactions were stopped after five minutes of incubation at 30 degrees with an equal volume of 0.8 M formic acid and measured by absorbance at 595 nm.

Metabolomic profiling by LC-HRAM-MS

Individual 2-micron TEF-promoter expression plasmids containing various AAAD family enzymes were generated to test the *in vivo* activity of AAAD family enzymes. *S. cerevisiae* lines were additionally transformed with various multi gene vectors to assay ectopic (S)-norcoclaurine production. 15 mL cultures of transgenic *S. cerevisiae* BY4743 strains were grown in 50ml mini bioreactor tubes for 24 hours with shaking at 30 °C. The culture were subsequently pelleted, washed and flash frozen as previously described[44]. Frozen yeast pellets were suspended in 50% methanol (500 µL per 100 mg fresh weight) and subsequently disrupted using a tissuelyser (Qiagen) and zirconia-silicate beads. The clarified extracts were analyzed by LC-HRAM-MS as previously described[44]. The raw data was processed and analyzed using MZmine2[51].

Molecular dynamics simulation and analysis

All simulations were performed using GROMACS 5.1.4[52] and the CHARMM force field[53]. The non-standard residue LLP was parameterized using Gaussian[54] and the Force Field Toolkit (FFTK)[55] implemented in VMD [56] based on the initial parameters provided by the CGenFF program[57–60]. A number of *Cr*TDC residues buried deeply within the protein or at the monomer-monomer interface were modeled in their neutral forms based on PROPKA[61,62] calculation results: $\text{Asp}^{268\text{-A/B}}$, $\text{Asp}^{287\text{-A/B}}$, $\text{Asp}^{397\text{-A/B}}$, $\text{Lys}^{208\text{-A/B}}$, and $\text{Glu}^{169\text{-A}}$. All the histidines were kept neutral, with a proton placed on the ε-nitrogen except for His^{203} and His^{318}, for which the proton was placed on the δ-nitrogen to optimize hydrogen bond network. All simulation systems were constructed as a dimer solvated in a dodecahedron water box with 0.1 M NaCl (Fig. S12) and a total number of atoms of ~124,000. Prior to the production runs listed in Table S4, all systems were subjected to energy minimization followed by a 100-ps NVT and a 100-ps NPT runs with the protein heavy atoms constrained. In all simulations the van der Waals interactions were smoothly switched off from 10 Å to 12 Å. The electrostatic interactions were computed with the Particle-Mesh-Ewald (PME) method[63] with a grid spacing of 1.5 Å, a PME order of 6, and a cutoff of 12 Å in real space. The system temperature was kept at 300 K using the velocity-rescaling thermostat[64], and the system pressure was kept at 1 bar with the Parrinello-Rahman barostat[65,66]. All bonds involving hydrogens were constrained using LINCS[67,68] to allow an integration timestep of 2 fs. The helix-unfolding simulation was performed using the metadynamics method[69] as implemented in PLUMED[70]. A 10-ns metadynamics simulation was performed on System 1 by placing gaussian potentials (height=35 kJ/mol, sigma=0.35 rad) every 500 steps on the collective variables, which were chosen as the backbone dihedral angles Ψ and Φ of residues 346 to 350. We should note that this simulation was not intended for an accurate free energy calculation and instead was only used to generate an unfolded structure of the short helix (residues 346-350). The resulting unfolded large loop structure was then used in all systems, each of which was subjected to 12 replicas of 50-ns MD simulations listed in Table S4. Clustering analysis was performed using gmx cluster over all simulated trajectories with a

RMSD cutoff of 3.5 Å. 3D occupancy maps were created at a resolution of 1 $Å^3$ using the VMD VOLMAP plugin. DSSP calculations[71] were performed with gmx do_dssp implemented in GROMACS. An output of H (α-helix), I (π-helix) or G (3_{10}-helix) was considered as a helix and the corresponding residue was assigned a helical content of 1; otherwise a helical content of 0 was assigned. Clustering and occupancy analysis as well as the average helical content calculations were performed on the combined trajectories of all simulation replicas for a given *Cr*TDC system. The two monomers of a *Cr*TDC dimer were treated equivalently in these analyses. All simulation figures were made using VMD.

Accession codes

The sequences of *P. somniferum*, *C. roseus*, *E. grandis* genes reported in this article are deposited into NCBI GenBank under the following accession numbers: *Ps*TyDC (MG748690), *CrTDC* (MG748691), *Eg*PAAS (MG786260).

Acknowledgments

This work was supported by the Pew Scholar Program in the Biomedical Sciences (J.K.W.), the Searle Scholars Program (J.K.W.), the National Science Foundation (CHE-1709616, J.K.W.) and direct grants from the Chinese University of Hong Kong (Y.W.). This work is based on research conducted at the Northeastern Collaborative Access Team (NE-CAT) beamlines, which are funded by the National Institute of General Medical Sciences from the National Institutes of Health (P41 GM103403). The Pilatus 6M detector on NE-CAT 24-ID-C beam line is funded by a NIH-ORIP HEI grant (S10 RR029205). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-

06CH11357.

Author Contributions

M.P.T.S. and J.K.W. conceived the idea and designed the research. M.P.T.S., T.S, and M.A.V. performed cloning and phylogenetic analysis. M.P.T.S. conducted enzyme purification, enzyme assays, crystallization, and structural elucidation. Y.C and Y. W performed the molecular dynamic simulations and data interpretation. M.P.T.S., Y.C, Y.W., and J.K.W. wrote the paper.

Figures

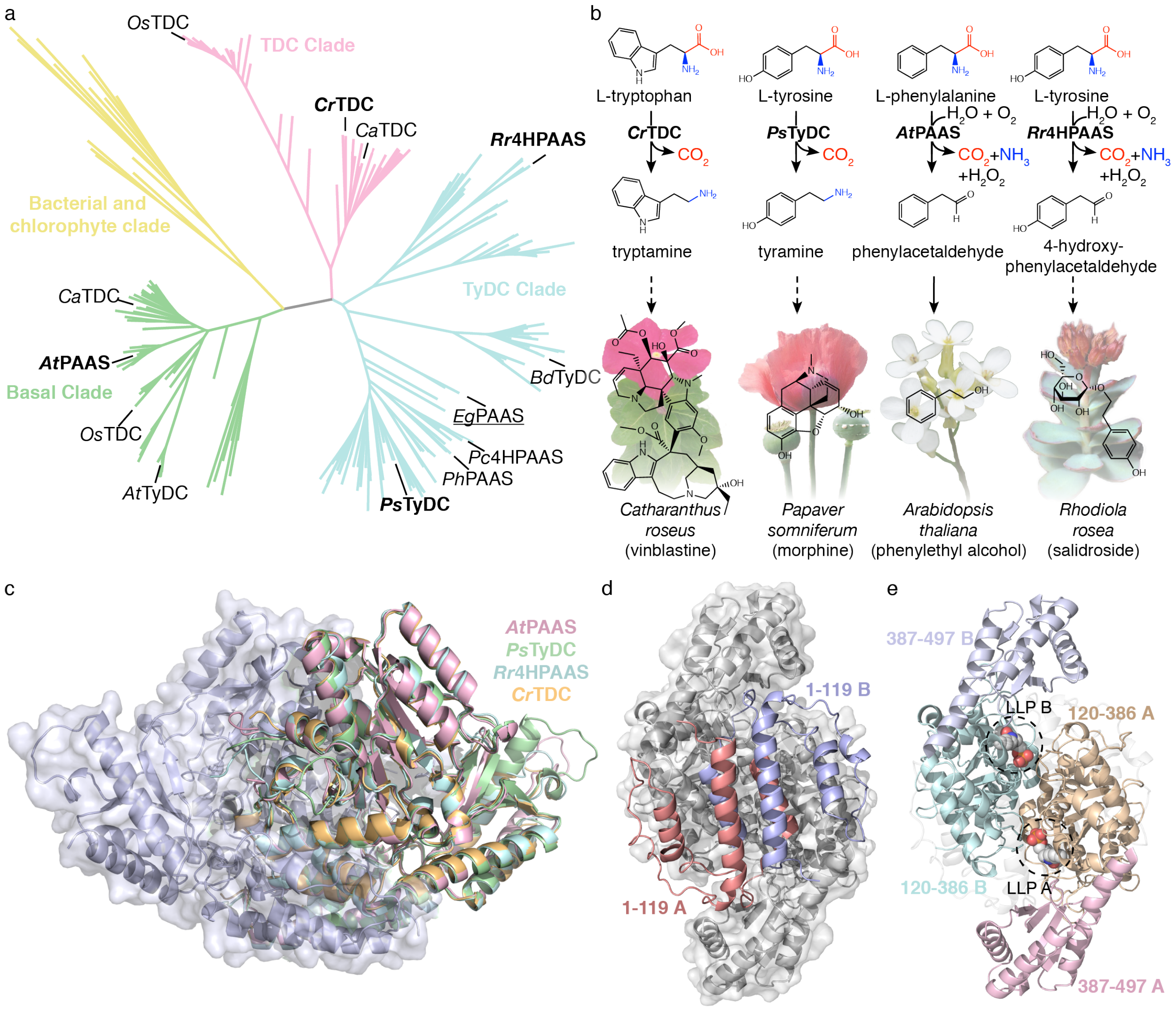


Fig. 1 | Phylogeny, functions, and structures of plant AAADs (a) A simplified maximum likelihood (ML) phylogenetic tree of bacteria, chlorophyte, and plant AAADs. A fully annotated tree is shown in Fig. S1. Nomenclature of the bacterial/chlorophyte (yellow), basal (green), TyDC (blue) and TDC (pink) clades is based on phylogenetic patterns as well as the general conservation of a substrate-specifying residue at the active site[13]. Functionally characterized enzymes are labeled at the tree branches. The four AAADs proteins with their crystal structures resolved in this study are denoted in bold. The *Eg*PAAS identified and characterized in this study is underlined. (**b**) Biochemical functions of four representative plant AAADs in the context of their native specialized metabolic pathways. The dotted arrows indicate multiple catalytic steps. (**c**) An overlay of the α2-dimer architectures of the *Cr*TDC (orange), *Ps*TyDC (green), *At*PAAS (pink), and *Rr*4HPAAS (cyan) structures. The left monomer of *Ps*TyDC is displayed in gray cartoon and surface representations. (**d**) Top-down view of the N-terminal segments (*Cr*TDC$^{1-119}$) from the two monomer labeled in salmon and blue, respectively. The remainder of the homodimer is displayed in gray. (**e**) The association of the *Cr*TDC middle (beige) and C-terminal (pink) segments of one monomer with the middle (teal) and C-terminal (light blue) segments of the another monomer. The N-terminal segments are displayed in transparent gray cartoons, and the LLPs are circled and displayed as spheres.

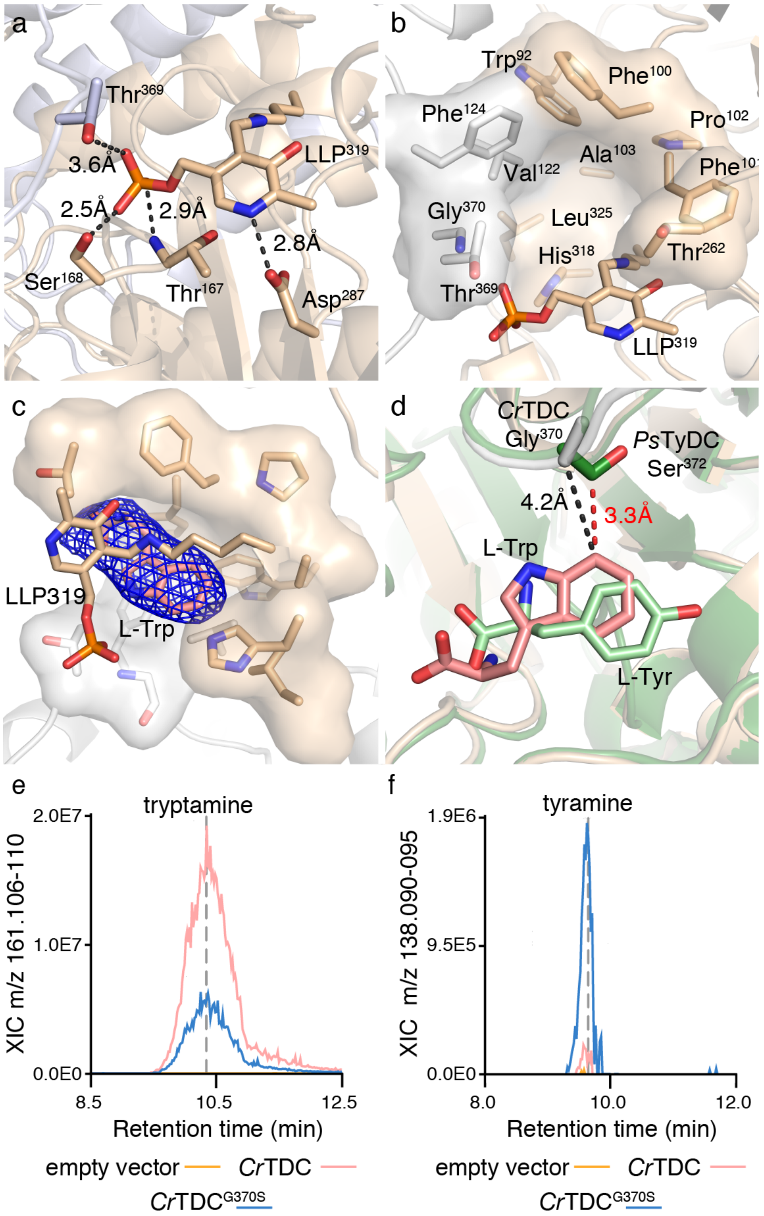


Figure 2 | Active-site composition and the structural basis for aromatic amino acid substrate selectivity (**a**) Coordination of the LLP prosthetic in *Cr*TDC with chain-A residues colored in beige and chain-B residues colored in white. (**b**) The *Cr*TDC active-site entrance composed by residues from both chain A and B, colored in beige and white, respectively. The pocket is composed of conserved nonpolar residues (Pro^{102}, Val^{122} and Leu^{325}), aromatic residues (Trp^{92}, Phe^{100}, Phe^{101}, Phe^{124}, His^{318}), and a polar residue (Thr^{262}). Additionally, the active site contains three nonconserved residues (Ala^{103}, Thr^{369} and Gly^{370}), which vary across different AAAD clades. (**c**) The interaction of the L-tryptophan substrate with the chain A (beige) and chain B (white) active-site residues. The $|2Fo - Fc|$ electron density map for L-tryptophan is contoured at 2 σ. (**d**) Superimposition of the substrate-complexed *Cr*TDC and *Ps*TyDC structures. *Cr*TDC Chain-A and Chain-B are displayed in beige and white respectively, while *Ps*TyDC is displayed in green. The L-tryptophan ligand from the *Cr*TDC structure is rendered in pink while the *Ps*TyDC L-tyrosine ligand is colored in light green. The intermolecular distances between the L-tryptophan and the *Cr*TDC Gly^{370} and *Ps*TyDC Ser^{372} are displayed in black and red dotted lines, respectively. (**e**) Relative in vivo tryptophan decarboxylase activity of the wild-type *Cr*TDC and *Cr*TDC^{G370S} enzymes as examined in transgenic yeast. (**f**) Relative in vivo tyrosine decarboxylase activity of the wild type *Cr*TDC and *Cr*TDC^{G370S} enzymes in transgenic yeast.

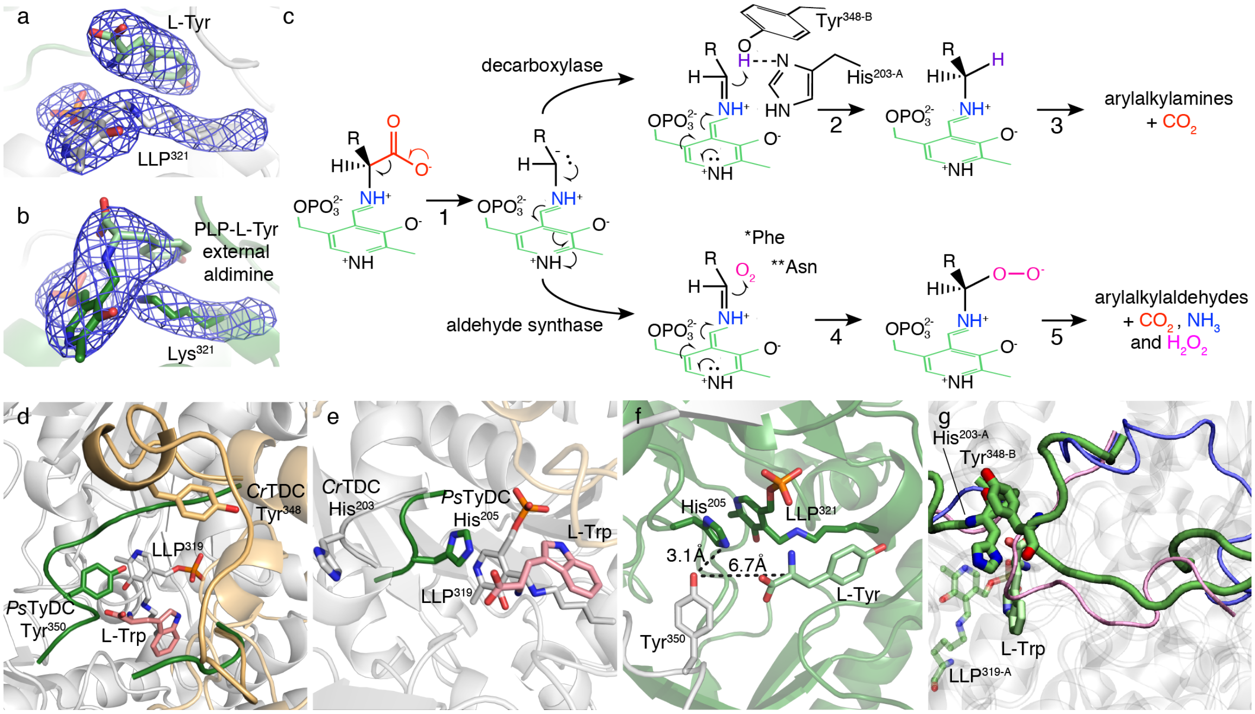


Figure 3 | Catalytic mechanisms and conformational changes of AAAD proteins. The PLP-Lys^{321} internal aldimine (**a**) and PLP-L-tyrosine external aldimine (**b**) captured by the two active sites of the *Ps*TyDC homodimer. The Chain A and Chain B are colored in green in gray, respectively, and the $|2Fo - Fc|$ electron density map is contoured at 2 σ. (**c**) The proposed alternative PLP-mediated catalytic mechanisms of decarboxylase and aldehyde synthase in AAAD proteins. The external aldimine loses the α-carboxyl group as CO_2 to generate a quinoid intermediate stabilized by the delocalization of the paired electrons (1). In a canonical decarboxylase, the carbanion is subsequently protonated in the presence of the acidic Tyr$^{348\text{-B}}$ facilitated by its neighboring His$^{203\text{-A}}$ (2), followed by the regeneration of the imine complex and the release of the arylalkylamine and carbon dioxide product (3). In the evolutionarily new aldehyde synthase, the proton transfer is impaired when either Tyr$^{348\text{-B}}$ or His$^{203\text{-A}}$ is mutated to alternative residues (e.g. Tyr-to-Phe mutation as in *At*PAAS denoted by an asterisk or His-to-Asn mutation as in *Eg*PAAS denoted by two asterisks), enabling the concomitant attack of molecular oxygen (4) to produce a peroxide intermediate. This peroxide intermediate further decomposes into the arylalkyladehyde, ammonia, and hydrogen peroxide products (5). R represents the side chain of an aromatic amino acid substrate. (**d**) An overlay of the *Ps*TyDC large loop in closed conformation (green) upon the *Cr*TDC structure with the large loop in open conformation (beige). (**e**) The open conformation *Cr*TDC α2-dimer (shown in beige and white cartoon) with the superimposition of the closed conformation *Ps*TyDC small loop (green). (**f**) The closed conformation *Ps*TyDC active site displaying the catalytic machinery necessary for carbanion protonation. Chain A is colored in white, chain B is colored in green ,and the L-tyrosine substrate is displayed in light green. (**g**) A snapshot from the MD simulation of *Cr*TDC System 1 at t=398 ns, exhibiting a semi-closed loop conformation. The open and closed conformations of the loops observed from the crystal structures are shown in blue and pink tubes, respectively.

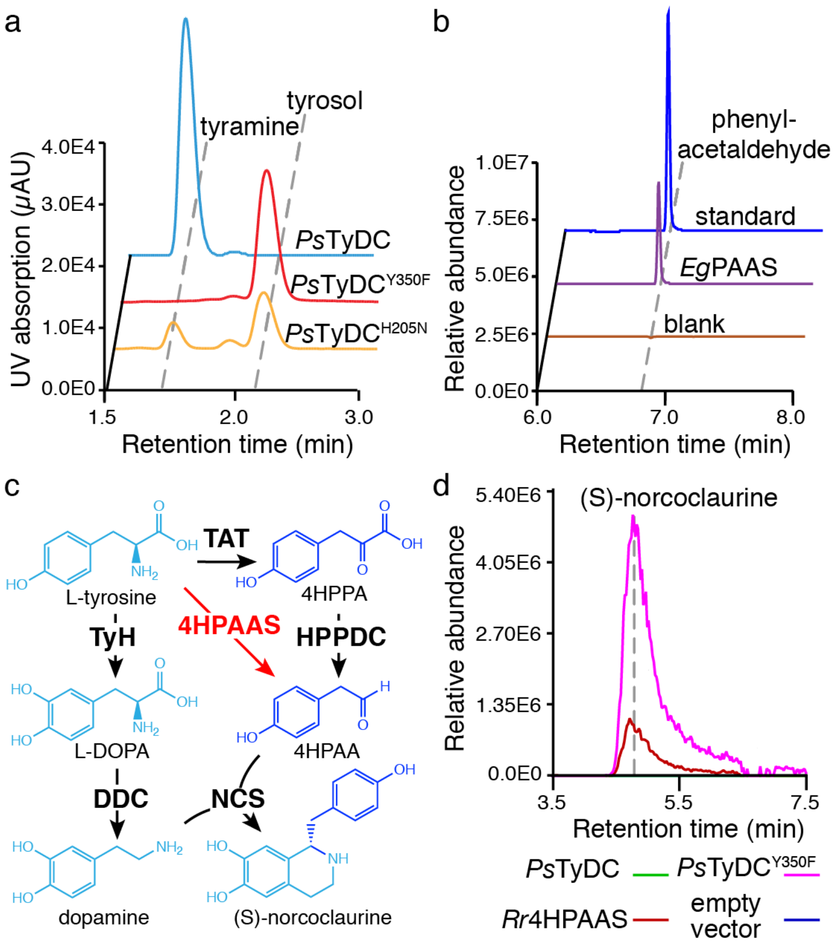


Figure 4 | Two alternative molecular strategies to arrive at aldehyde synthase chemistry from a canonical AAAD progenitor and the utility of 4HPAAS in metabolic engineering of (S)-norcoclaurine production in yeast. (**a**) Relative decarboxylation and aldehyde synthase products formed from L-tyrosine and recombinant *Ps*TyDC, *Ps*TyDC$^{\text{H204N}}$ and *Ps*TyDC$^{\text{Y350F}}$ reactions. After enzymatic reaction, the 4HPAA aldehyde product was chemically reduced by sodium borohydride to yield tyrosol prior to LC-UV detection. (**b**) Phenylacetaldehyde formation from the incubation of *Eg*PAAS with phenylalanine measured by LC-MS. (**c**) Canonically proposed (S)-norcoclaurine biosynthetic pathway (black) rerouted by the use of an 4HPAAS (red). 4HPPA, 4-hydroxyphenylacetaldehyde. (**d**) Engineering of (S)-norcoclaurine production in yeast using two AAAD proteins with the 4HPAAS activity. All (S)-norcoclaurine multigene vectors used to transform yeast contain the requisite *PpDDC*, *PsNCS2* and *BvTyH* in addition to variable *PsTyDC*, *PsTyDC*^{Y350F} or *Rr4HPAAS* genes.

References

1. Weng, J.-K., Philippe, R. N. & Noel, J. P. The rise of chemodiversity in plants. *Science* **336,** 1667–1670 (2012).
2. Chae, L., Kim, T., Nilo-Poyanco, R. & Rhee, S. Y. Genomic signatures of specialized metabolism in plants. *Science* **344,** 510–513 (2014).
3. Kries, H. *et al.* Structural determinants of reductive terpene cyclization in iridoid biosynthesis. *Nat. Chem. Biol.* **12,** 6–8 (2016).
4. Kaltenbach, M. *et al.* Evolution of chalcone isomerase from a noncatalytic ancestor. *Nat. Chem. Biol.* **14,** 548–555 (2018).
5. Burkhard, P., Dominici, P., Borri-Voltattorni, C., Jansonius, J. N. & Malashkevich, V. N. Structural insight into Parkinson's disease treatment from drug-inhibited DOPA decarboxylase. *Nat. Struct. Biol.* **8,** 963–967 (2001).
6. Facchini, P. J., Huber-Allanach, K. L. & Tari, L. W. Plant aromatic L-amino acid decarboxylases: evolution, biochemistry, regulation, and metabolic engineering applications. *Phytochemistry* **54,** 121–138 (2000).
7. Noé, W., Mollenschott, C. & Berlin, J. Tryptophan decarboxylase from Catharanthus roseus cell suspension cultures: purification, molecular and kinetic data of the homogenous protein. *Plant Mol. Biol.* **3,** 281–288 (1984).
8. Facchini, P. J. & De Luca, V. Differential and tissue-specific expression of a gene family for tyrosine/dopa decarboxylase in opium poppy. *J. Biol. Chem.* **269,** 26684–26690 (1994).
9. Gutensohn, M. *et al.* Role of aromatic aldehyde synthase in wounding/herbivory response and flower scent production in different Arabidopsis ecotypes. *Plant J.* **66,** 591–602 (2011).

10. Torrens-Spence, M. P., Pluskal, T., Li, F.-S., Carballo, V. & Weng, J.-K. Complete Pathway Elucidation and Heterologous Reconstitution of Rhodiola Salidroside Biosynthesis. *Mol. Plant* **11,** 205–217 (2018).
11. Alexander, F. W., Sandmeier, E., Mehta, P. K. & Christen, P. Evolutionary relationships among pyridoxal-5'-phosphate-dependent enzymes. Regio-specific alpha, beta and gamma families. *Eur. J. Biochem.* **219,** 953–960 (1994).
12. Jansonius, J. N. Structure, evolution and action of vitamin B6-dependent enzymes. *Curr. Opin. Struct. Biol.* **8,** 759–769 (1998).
13. Torrens-Spence, M. P., Lazear, M., von Guggenberg, R., Ding, H. & Li, J. Investigation of a substrate-specifying residue within Papaver somniferum and Catharanthus roseus aromatic amino acid decarboxylases. *Phytochemistry* **106,** 37–43 (2014).
14. Dunathan, H. C. Conformation and reaction specificity in pyridoxal phosphate enzymes. *Proceedings of the National Academy of Sciences* **55,** 712–716 (1966).
15. Ishii, S., Mizuguchi, H., Nishino, J., Hayashi, H. & Kagamiyama, H. Functionally important residues of aromatic L-amino acid decarboxylase probed by sequence alignment and site-directed mutagenesis. *J. Biochem.* **120,** 369–376 (1996).
16. Giardina, G. *et al.* Open conformation of human DOPA decarboxylase reveals the mechanism of PLP addition to Group II decarboxylases. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 20514–20519 (2011).
17. Han, Q., Ding, H., Robinson, H., Christensen, B. M. & Li, J. Crystal structure and substrate specificity of Drosophila 3,4-dihydroxyphenylalanine decarboxylase. *PLoS One* **5,** e8826 (2010).
18. Torrens-Spence, M. P. *et al.* Biochemical evaluation of the decarboxylation and decarboxylation-deamination activities of plant aromatic amino acid decarboxylases. *J. Biol. Chem.* **288,** 2376–2387 (2013).
19. Torrens-Spence, M. P., von Guggenberg, R., Lazear, M., Ding, H. & Li, J. Diverse functional evolution of serine decarboxylases: identification of two novel acetaldehyde synthases that uses hydrophobic amino acids as substrates. *BMC Plant Biol.* **14,** 247 (2014).

20. Kaminaga, Y. *et al.* Plant phenylacetaldehyde synthase is a bifunctional homotetrameric enzyme that catalyzes phenylalanine decarboxylation and oxidation. *J. Biol. Chem.* **281,** 23357–23366 (2006).
21. Vacca, R. A., Christen, P., Malashkevich, V. N., Jansonius, J. N. & Sandmeier, E. Substitution of apolar residues in the active site of aspartate aminotransferase by histidine. Effects on reaction and substrate specificity. *Eur. J. Biochem.* **227,** 481–487 (1995).
22. Mehta, P. K., Hale, T. I. & Christen, P. Aminotransferases: demonstration of homology and division into evolutionary subgroups. *Eur. J. Biochem.* **214,** 549–561 (1993).
23. Dominici, P., Tancini, B. & Borri Voltattorni, C. Chemical modification of pig kidney 3,4-dihydroxyphenylalanine decarboxylase with diethyl pyrocarbonate. Evidence for an essential histidyl residue. *J. Biol. Chem.* **260,** 10583–10589 (1985).
24. Hayashi, H., Mizuguchi, H. & Kagamiyama, H. Rat liver aromatic L-amino acid decarboxylase: spectroscopic and kinetic analysis of the coenzyme and reaction intermediates. *Biochemistry* **32,** 812–818 (1993).
25. Stavrinides, A. *et al.* Structural investigation of heteroyohimbine alkaloid synthesis reveals active site elements that control stereoselectivity. *Nat. Commun.* **7,** 12116 (2016).
26. Matasci, N. *et al.* Data access for the 1,000 Plants (1KP) project. *Gigascience* **3,** 17 (2014).
27. Pereira, M. do C. *et al.* Chemical Composition and Antimicrobial Activity of the Essential Oil fromMicrolicia crenulata. *Journal of Essential Oil Bearing Plants* **18,** 18–28 (2015).
28. Özek, T., Demirci, B. & H.C. Baser, K. Chemical Composition of Turkish Myrtle Oil. *J. Essent. Oil Res.* **12,** 541–544 (2000).
29. D'Arcy, B. R., Rintoul, G. B., Rowland, C. Y. & Blackman, A. J. Composition of Australian Honey Extractives. 1. Norisoprenoids, Monoterpenes, and Other Natural Volatiles from Blue Gum (Eucalyptusleucoxylon)and Yellow Box (Eucalyptus melliodora)Honeys. *J. Agric. Food Chem.* **45,** 1834–1843 (1997).
30. Tsurusaki, K.-I., Takeda, K. & Sakurai, N. Conversion of Indole-3-Acetaldehyde to Indole-3-Acetic Acid in Cell-Wall Fraction of Barley (Hordeum vulgare) Seedlings. *Plant Cell Physiol.* **38,** 268–273

(1997).

31. Hagel, J. M. & Facchini, P. J. Benzylisoquinoline Alkaloid Metabolism: A Century of Discovery and a Brave New World. *Plant Cell Physiol.* **54,** 647–672 (2013).
32. Weng, J.-K. & Noel, J. P. The remarkable pliability and promiscuity of specialized metabolism. *Cold Spring Harb. Symp. Quant. Biol.* **77,** 309–320 (2012).
33. Yoshikuni, Y., Ferrin, T. E. & Keasling, J. D. Designed divergent evolution of enzyme function. *Nature* **440,** 1078–1082 (2006).
34. Kopycki, J. G. *et al.* Biochemical and structural analysis of substrate promiscuity in plant Mg2+-dependent O-methyltransferases. *J. Mol. Biol.* **378,** 154–164 (2008).
35. Huang, R. *et al.* Enzyme functional evolution through improved catalysis of ancestrally nonpreferred substrates. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 2966–2971 (2012).
36. Austin, M. B., Bowman, M. E., Ferrer, J.-L., Schröder, J. & Noel, J. P. An aldol switch discovered in stilbene synthases mediates cyclization specificity of type III polyketide synthases. *Chem. Biol.* **11,** 1179–1194 (2004).
37. Liang, J., Han, Q., Ding, H. & Li, J. Biochemical identification of residues that discriminate between 3,4-dihydroxyphenylalanine decarboxylase and 3,4-dihydroxyphenylacetaldehyde synthase-mediated reactions. *Insect Biochem. Mol. Biol.* **91,** 34–43 (2017).
38. Thompson, J. D., Gibson, T. J. & Higgins, D. G. Multiple sequence alignment using ClustalW and ClustalX. *Curr. Protoc. Bioinformatics* **Chapter 2,** Unit 2.3 (2002).
39. Goodstein, D. M. *et al.* Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40,** D1178–86 (2012).
40. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **33,** 1870–1874 (2016).
41. Gouet, P., Robert, X. & Courcelle, E. ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.* **31,** 3320–3323 (2003).
42. Crooks, G. E. WebLogo: A Sequence Logo Generator. *Genome Res.* **14,** 1188–1190 (2004).

43. Torrens-Spence, M. P. *et al.* Biochemical evaluation of the decarboxylation and decarboxylation-deamination activities of plant aromatic amino acid decarboxylases. *J. Biol. Chem.* **288,** 2376–2387 (2013).

44. Torrens-Spence, M. P., Pluskal, T., Li, F.-S., Carballo, V. & Weng, J.-K. Complete Pathway Elucidation and Heterologous Reconstitution of Rhodiola Salidroside Biosynthesis. *Mol. Plant* **11,** 205–217 (2018).

45. Mumberg, D., Müller, R. & Funk, M. Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene* **156,** 119–122 (1995).

46. Lee, M. E., DeLoache, W. C., Cervantes, B. & Dueber, J. E. A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synth. Biol.* **4,** 975–986 (2015).

47. Han, Q., Ding, H., Robinson, H., Christensen, B. M. & Li, J. Crystal structure and substrate specificity of Drosophila 3,4-dihydroxyphenylalanine decarboxylase. *PLoS One* **5,** e8826 (2010).

48. Vagin, A. & Teplyakov, A. MOLREP: an Automated Program for Molecular Replacement. *J. Appl. Crystallogr.* **30,** 1022–1025 (1997).

49. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67,** 355–367 (2011).

50. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 486–501 (2010).

51. Pluskal, T., Castillo, S., Villar-Briones, A. & Oresic, M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11,** 395 (2010).

52. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2,** 19–25 (2015).

53. Best, R. B. *et al.* Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ(1) and χ(2) dihedral angles. *J. Chem. Theory Comput.* **8,** 3257–3273 (2012).

54. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox. *Gaussian 09*. (2016).
55. Mayne, C. G., Saam, J., Schulten, K., Tajkhorshid, E. & Gumbart, J. C. Rapid parameterization of small molecules using the Force Field Toolkit. *J. Comput. Chem.* **34,** 2757–2770 (2013).
56. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14,** 33–38 (1996).
57. Vanommeslaeghe, K. & MacKerell, A. D., Jr. Automation of the CHARMM General Force Field (CGenFF) I: bond perception and atom typing. *J. Chem. Inf. Model.* **52,** 3144–3154 (2012).
58. Vanommeslaeghe, K., Raman, E. P. & MacKerell, A. D., Jr. Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges. *J. Chem. Inf. Model.* **52,** 3155–3168 (2012).
59. Vanommeslaeghe, K. *et al.* CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* NA–NA (2009).
60. Yu, W., He, X., Vanommeslaeghe, K. & MacKerell, A. D. Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations. *J. Comput. Chem.* **33,** 2451–2468 (2012).

61. Søndergaard, C. R., Olsson, M. H. M., Rostkowski, M. & Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J. Chem. Theory Comput.* **7,** 2284–2295 (2011).

62. Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **7,** 525–537 (2011).

63. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98,** 10089–10092 (1993).

64. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126,** 014101 (2007).

65. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52,** 7182–7190 (1981).

66. Nosé, S. & Klein, M. L. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **50,** 1055–1076 (1983).

67. Hess, B., Bekker, H., Berendsen, H. J. C. & Johannes G E. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18,** 1463–1472 (1997).

68. Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **4,** 116–122 (2008).

69. Laio, A. & Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 12562–12566 (2002).

70. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **185,** 604–613 (2014).

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

Supplementary Information for

Structural basis for independent origins of new catalytic machineries in plant AAAD proteins

Michael P. Torrens-Spence[1], Ying-Chih Chiang[2], Tyler Smith[1,4], Maria A. Vicent[1,3], Yi Wang[2], and Jing-Ke Weng[1,4]*

[1]Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA.
[2]Department of Physics, the Chinese University of Hong Kong, Shatin, N.T., Hong Kong.
[3]Department of Biology, Williams College, Williamstown, Massachusetts 01267, USA.
[4]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

*Corresponding author: wengj@wi.mit.edu

Table of contents

Materials and Methods 3
Fig. S1-18 11
Supplementary Video Information 29
Note S1-5 30
Table S1-4 37
Supplementary References 42

Supplementary Figures

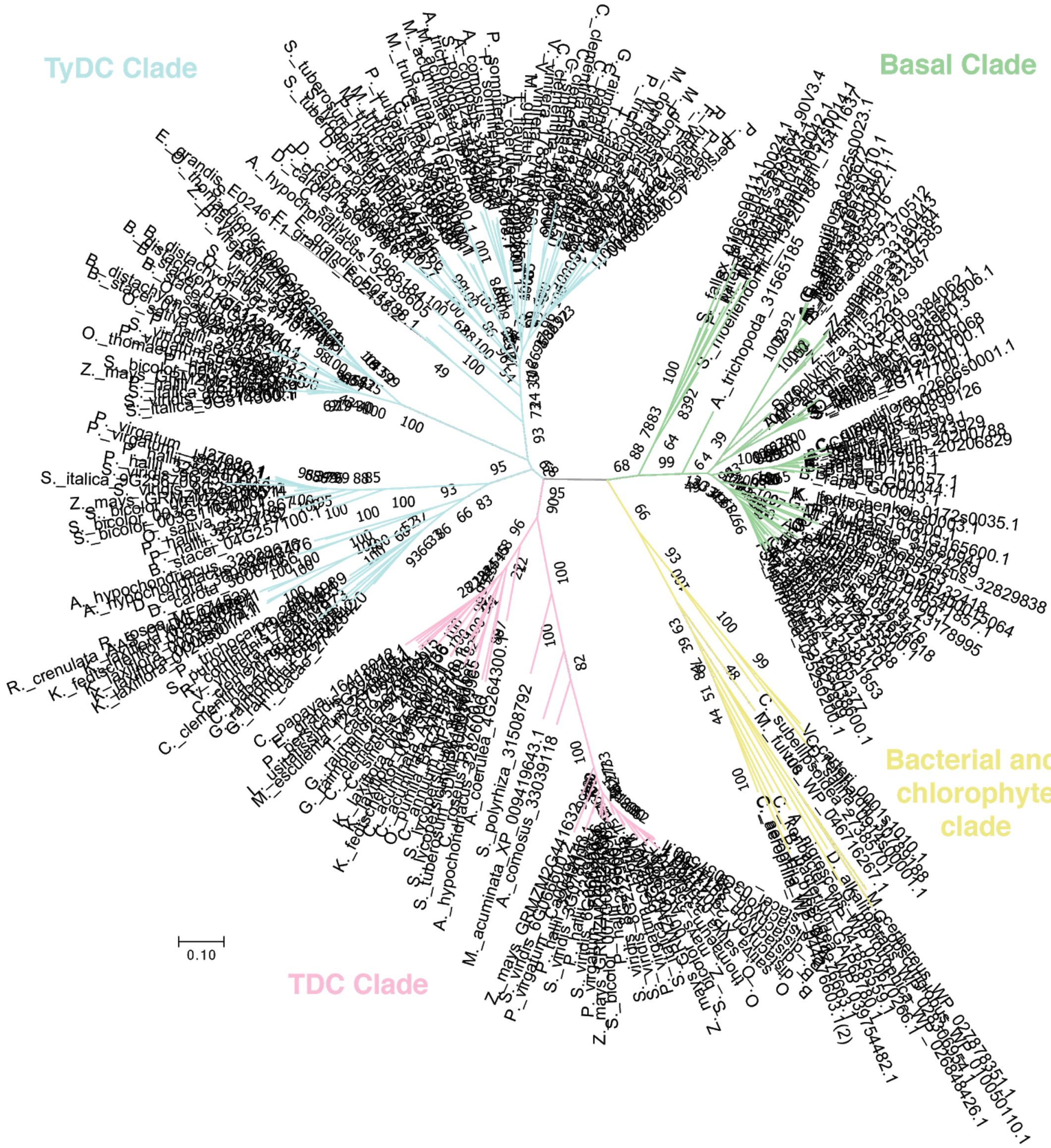


Fig. S1 | Phylogenetic tree of plant and eubacteria AAADs. This tree is populated with sequences from all Phytozome V12 species, all attainable characterized NCBI plant AAAD sequences and select eubacteria NCBI sequences. Green, pink and blue branches correspond to the plant AAAD basal, TDC and TyDC clades, respectively. The yellow branches correspond to the chlorophytes and bacterial AAAD sequences. The plant AAAD clades were annotated according to their relation to ancestral sequences (the green basal is most closely related to bacterial and chlorophytes AAADs) and their apparent substrate selectivity (the pink TDC clade contains a number of characterized enzymes with exclusive indolic substrate specificity while the blue TyDC clade is represented by characterized sequences with phenolic substrate selectivity).

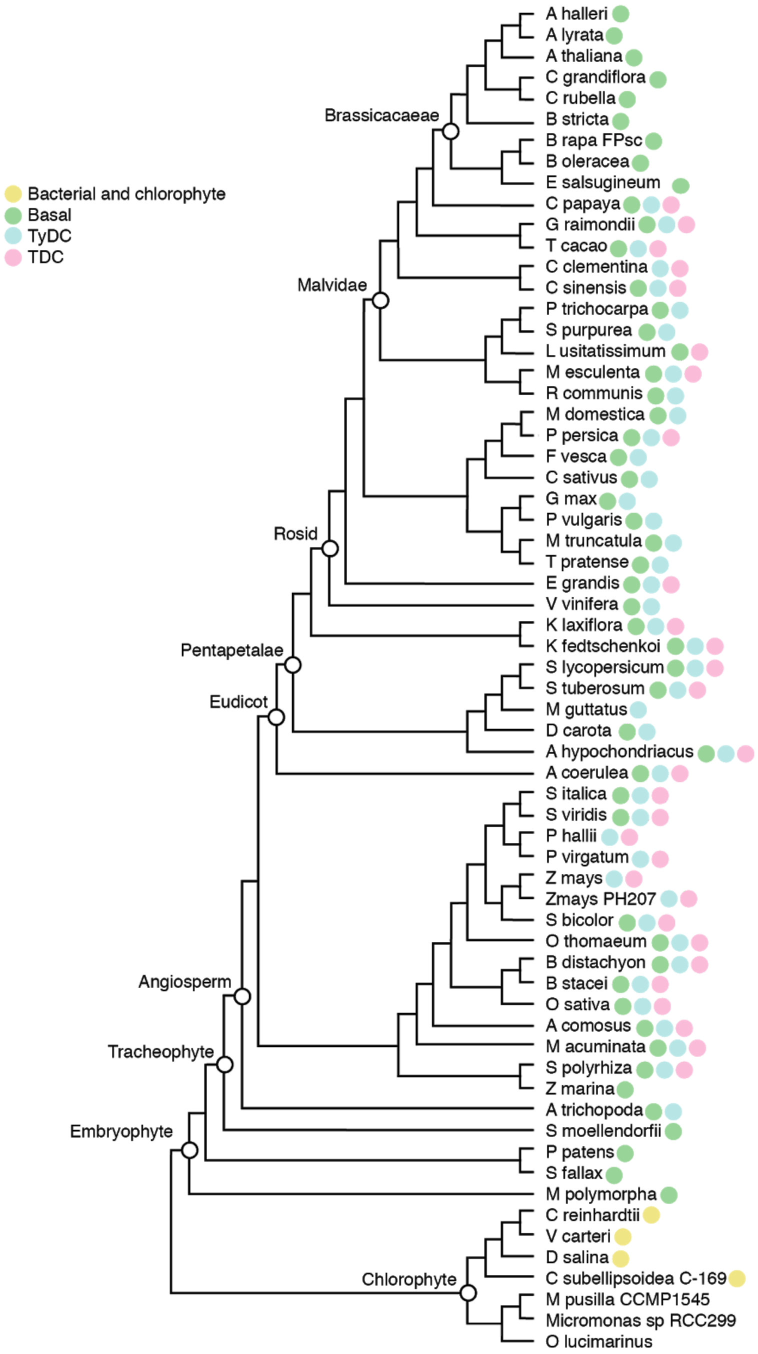
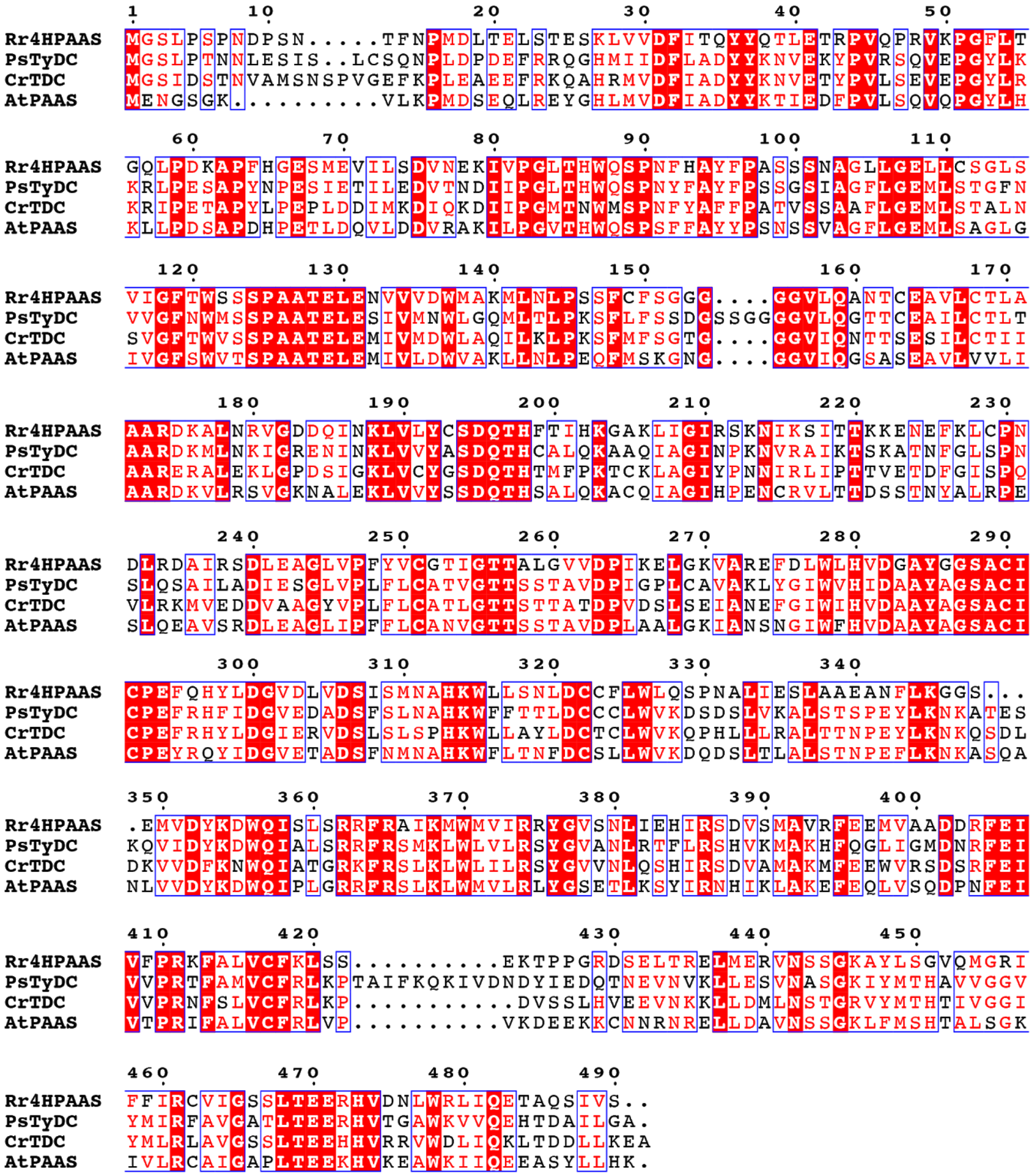


Fig. S2 | Taxonomic distribution of plant AAAD sequences. This evolutionary tree illustrates the inferred relationships of sequenced plant species from Phytozome V12. The presence of a yellow, green, blue or pink circle next to the species name indicates the presence of one or more AAAD sequences from the bacterial/chlorophyte, basal, TyDC or TDC clades, respectively.



```
Rr4HPAAS  MGSLPSPNDPSN.....TFNPMDLTELSTESKLVVDFITQYYQTLETRPVQPRVKPGFLT
PsTyDC    MGSLPTNNLESIS..LCSQNPLDPDEFRRQGHMIIDFLADYYKNVEKYPVRSQVEPGYLK
CrTDC     MGSIDSTNVAMSNSPVGEFKPLEAEEFRKQAHRMVDFIADYYKNVETYPVLSEVEPGYLR
AtPAAS    MENGSGK..........VLKPMDSEQLREYGHLMVDFIADYYKTIEDFPVLSQVQPGYLH

Rr4HPAAS  GQLPDKAPFHGESMEVILSDVNEKIVPGLTHWQSPNFHAYFPASSSNAGLLGELLCSGLS
PsTyDC    KRLPESAPYNPESIETILEDVTNDIIPGLTHWQSPNYFAYFPSSGSIAGFLGEMLSTGFN
CrTDC     KRIPETAPYLPEPLDDIMKDIQKDIIPGMTNWMSPNFYAFFPATVSSAAFLGEMLSTALN
AtPAAS    KLLPDSAPDHPETLDQVLDDVRAKILPGVTHWQSPSFFAYYPSNSSVAGFLGEMLSAGLG

Rr4HPAAS  VIGFTWSSSPAATELENVVVDWMAKMLNLPSSFCFSGGG....GGVLQANTCEAVLCTLA
PsTyDC    VVGFNWMSSPAATELESIVMNWLGQMLTLPKSFLFSSDGSSGGGGVLQGTTCEAILCTLT
CrTDC     SVGFTWVSSPAATELEMIVMDWLAQILKLPKSFMFSGTG....GGVIQNTTSESILCTII
AtPAAS    IVGFSWVTSPAATELEMIVLDWVAKLLNLPEQFMSKGNG....GGVIQGSASEAVLVVLI

Rr4HPAAS  AARDKALNRVGDDQINKLVLYCSDQTHFTIHKGAKLIGIRSKNIKSITTKKENEFKLCPN
PsTyDC    AARDKMLNKIGRENINKLVVYASDQTHCALQKAAQIAGINPKNVRAIKTSKATNFGLSPN
CrTDC     AARERALEKLGPDSIGKLVCYGSDQTHTMFPKTCKLAGIYPNNIRLIPTTVETDFGISPQ
AtPAAS    AARDKVLRSVGKNALEKLVVYSSDQTHSALQKACQIAGIHPENCRVLTTDSSTNYALRPE

Rr4HPAAS  DLRDAIRSDLEAGLVPFYVCGTIGTTALGVVDPIKELGKVAREFDLWLHVDGAYGGSACI
PsTyDC    SLQSAILADIESGLVPLFLCATVGTTSSTAVDPIGPLCAVAKLYGIWVHIDAAYAGSACI
CrTDC     VLRKMVEDDVAAGYVPLFLCATLGTTSTTATDPVDSLSEIANEFGIWIHVDAAYAGSACI
AtPAAS    SLQEAVSRDLEAGLIPFFLCANVGTTSSTAVDPLAALGKIANSNGIWFHVDAAYAGSACI

Rr4HPAAS  CPEFQHYLDGVDLVDSISMNAHKWLLSNLDCCFLWLQSPNALIESLAAEANFLKGGS...
PsTyDC    CPEFRHFIDGVEDADSFSLNAHKWFFTTLDCCLWVKDSDSLVKALSTSPEYLKNKATES
CrTDC     CPEFRHYLDGIERVDSLSLSPHKWLLAYLDCTCLWVKQPHLLLRALTTNPEYLKNKQSDL
AtPAAS    CPEYRQYIDGVETADSFNMNAHKWFLTNFDCSLLWVKDQDSLTLALSTNPEFLKNKASQA

Rr4HPAAS  .EMVDYKDWQISLSRRFRAIKMWMVIRRYGVSNLIEHIRSDVSMAVRFEEMVAADDRFEI
PsTyDC    KQVIDYKDWQIALSRRFRSMKLWLVLRSYGVANLRTFLRSHVKMAKHFQGLIGMDNRFEI
CrTDC     DKVVDFKNWQIATGRKFRSLKLWLILRSYGVVNLQSHIRSDVAMAKMFEEWVRSDSRFEI
AtPAAS    NLVVDYKDWQIPLGRRFRSLKLWMVLRLYGSETLKSYIRNHIKLAKEFEQLVSQDPNFEI

Rr4HPAAS  VFPRKFALVCFKLSS...........EKTPPGRDSELTRELMERVNSSGKAYLSGVQMGRI
PsTyDC    VVPRTFAMVCFRLKPTAIFKQKIVDNDYIEDQTNEVNVKLLESVNASGKIYMTHAVVGGV
CrTDC     VVPRNFSLVCFRLKP...........DVSSLHVEEVNKKLLDMLNSTGRVYMTHTIVGGI
AtPAAS    VTPRIFALVCFRLVP...........VKDEEKKCNNRNRELLDAVNSSGKLFMSHTALSGK

Rr4HPAAS  FFIRCVIGSSLTEERHVDNLWRLIQETAQSIVS..
PsTyDC    YMIRFAVGATLTEERHVTGAWKVVQEHTDAILGA.
CrTDC     YMLRLAVGSSLTEEHHVRRVWDLIQKLTDDLLKEA
AtPAAS    IVLRCAIGAPLTEEKHVKEAWKIIQEEASYLLHK.
```

Fig. S3 | Multiple sequence alignment of crystallized plant AAADs. The multiple sequence alignment was generated with ClustalW2[1] and displayed with ESPript 3.0[2].

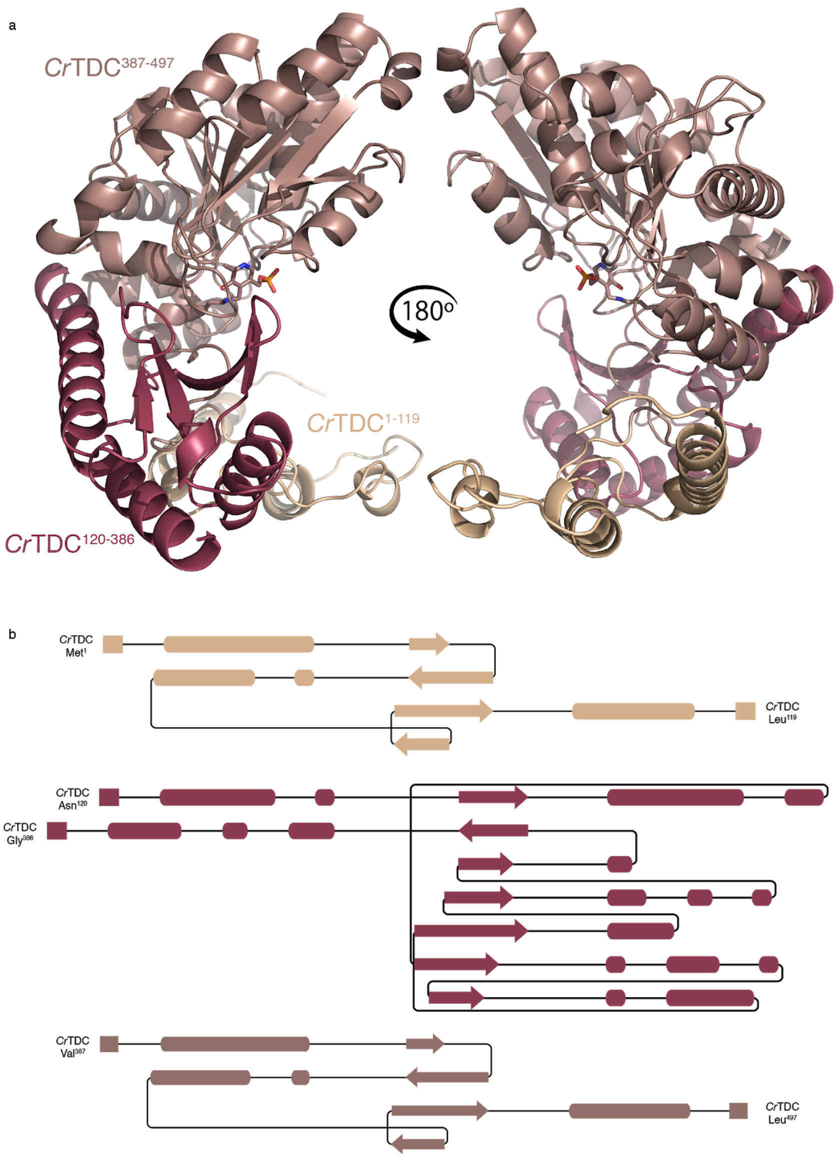


Fig. S4 | Three segment architecture of *Cr*TDC. (**a**) Plant AAAD segment structure as displayed by the *Cr*TDC B-chain. Each monomer is composed of the N-terminal *Cr*TDC$^{1-119}$(beige), middle *Cr*TDC$^{120-386}$(maroon) and C-terminal *Cr*TDC$^{387-497}$(brown) segments. (**b**) Topology diagram for each of the three *Cr*TDC. The segment diagrams were generated though Pro-origami using DSSP secondary structure program[3].

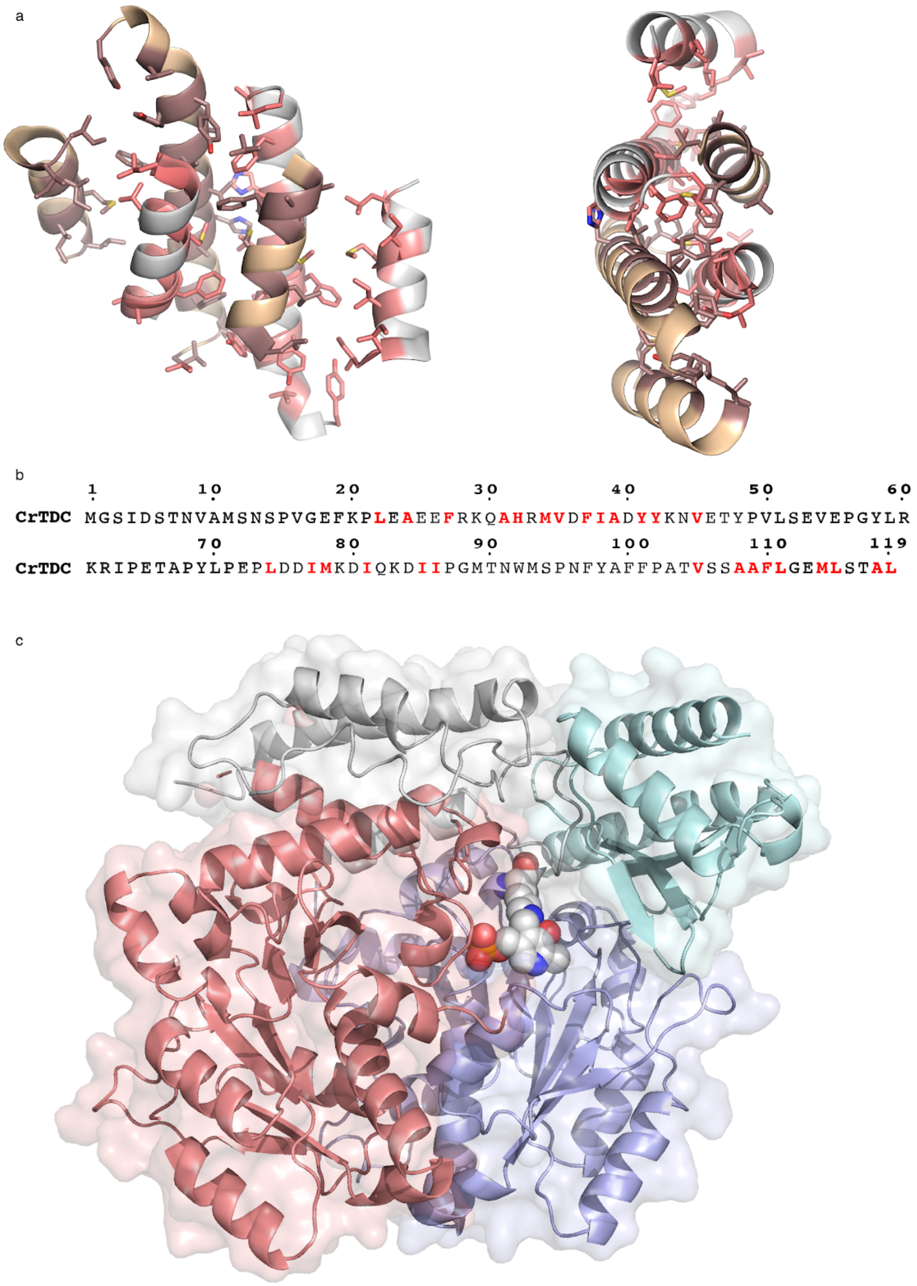


Fig. S5 | Intermonomer association various segments. (**a**) A side and top view of the aromatic and hydrophobic residues forming the intermolecular interaction of the *Cr*TDC homodimer. One monomer is orange with maroon hydrophobic or aromatic residues, whereas the second monomer is white with pink hydrophobic or aromatic residues. (**b**) Sequence of the *Cr*TDC N-terminal segment with hydrophobic or aromatic helices residues marked in red. (**c**) View of the segment structure composing a single *Cr*TDC active site pocket. The N-terminal, middle and C-terminal segments from Chain A are shown in grey, blue and cyan, respectively, while the middle segment of Chain B is shown in pink. The LLP of the active site is displayed as a ball model.

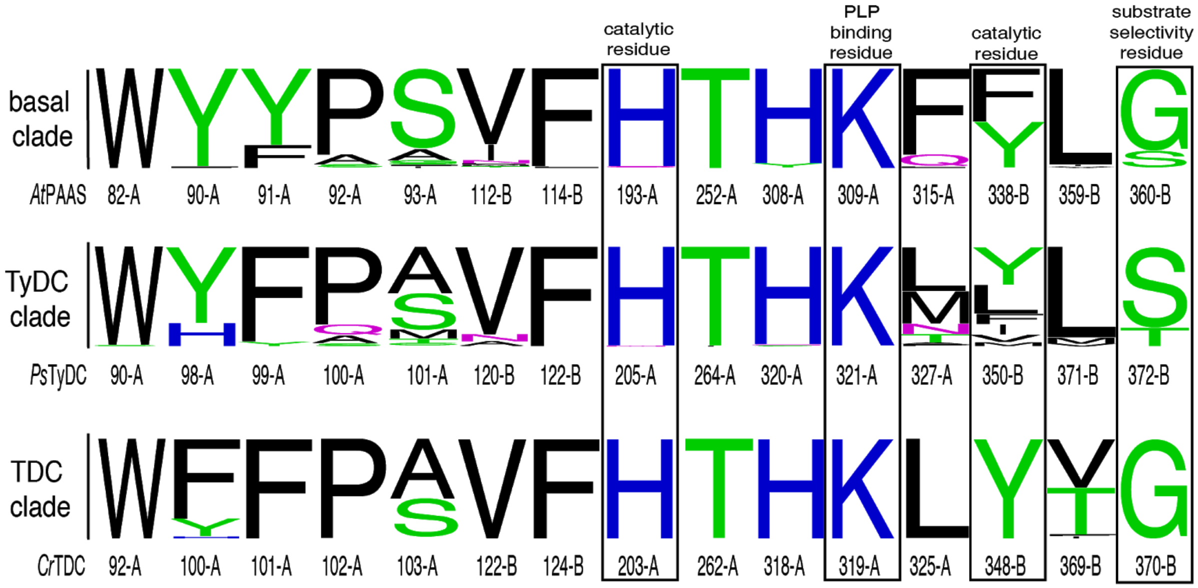


Fig. S6 | Active-site sequence conservation within the three plant AAAD clades. Active site residues from plant AAADs were identified through the evaluation of the four plant AAAD crystal structures. Multiple sequence alignments of the queried AAAD sequences within each phylogenetic clade were generated and then evaluated for active site conservation. Polar amino acids are green, basic amino acids are blue, acidic amino acids are red and hydrophobic amino acids are black. The amino acid composition of the active site is displayed for the basal (top), TyDC (middle) and TDC (bottom) clades. The polypeptide chain and position of each active site location is listed by one of the representative *At*PAAS, *Ps*TyDC and *Cr*TDC references sequences. The y-axis has been scaled to display the relative amino acid frequency of a given residue. Key residues are displayed in boxes with header titles indicating their function.

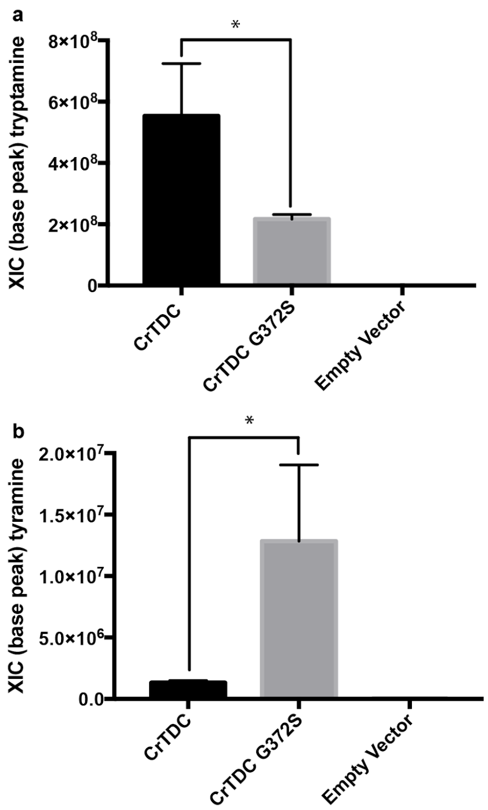


Fig. S7 | Metabolite profiling of tryptamine and tyramine in transgenic *S. cerevisiae.* (**a**) Transgenic yeast overexpressing *Cr*TDC$^{\text{G370S}}$ display reduced intracellular tryptamine in comparison to transgenic yeast overexpressing wild type *Cr*TDC. (**b**) *S. cerevisiae* overexpressing *Cr*TDC$^{\text{G370S}}$ demonstrates increased intracellular tyramine in comparison to transgenic yeast overexpressing wild type *Cr*TDC. Cultures were grown and metabolically profiled in triplicate. *Adjusted P Value < 0.02.

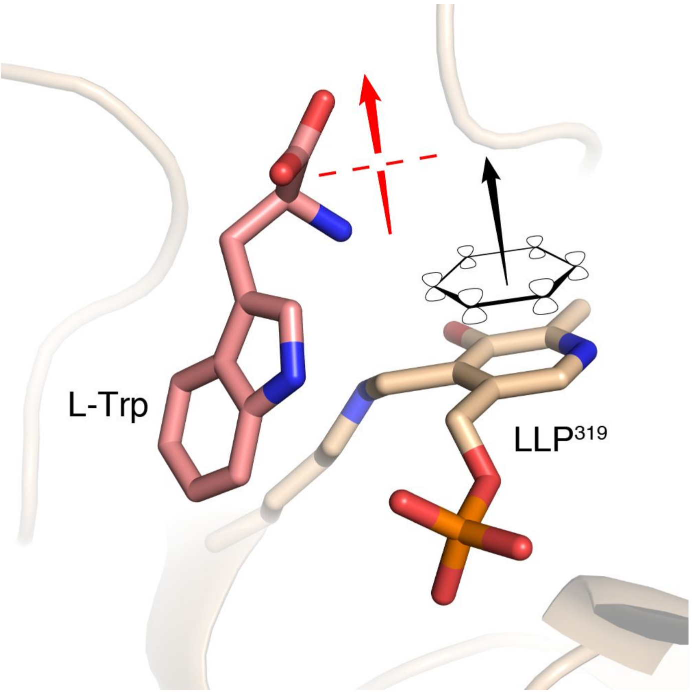


Fig. S8 | Orientation of the alpha carbon carbonyl bond relative to the plane of the pyridoxal imine system. As per the Dunathan hypothesis, PLP enzymes exhibit stereospecific cleavage of bonds orthogonal to the pyridine ring pi system electrons (shown as black ring and arrow)[4]. In the case of PLP decarboxylases, the alpha carbon carbonyl bond of the substrate is positioned perpendicular to the plane of the pyridine ring (shown as red arrow).

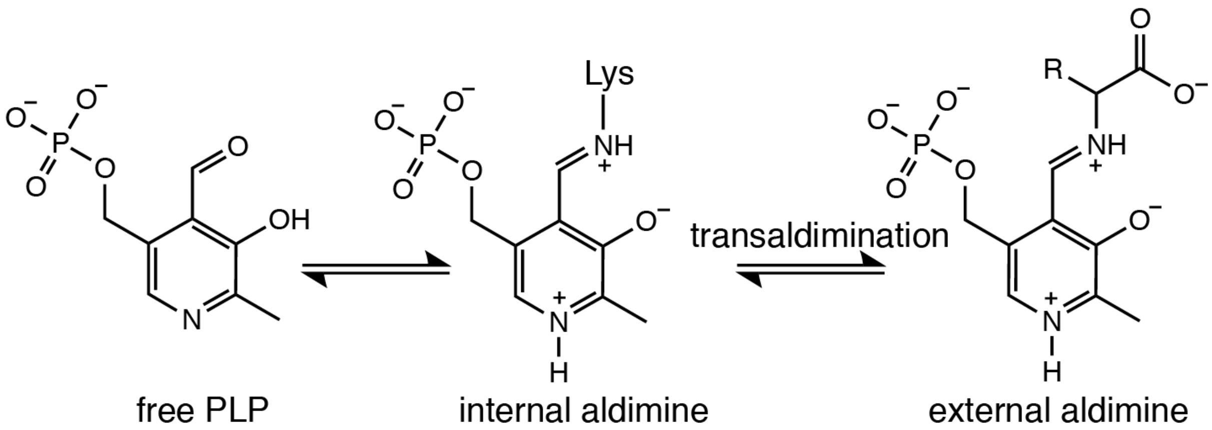


Fig. S9 | Schematic of the formation of PLP internal and external aldamines. First, the internal aldimine is formed when the aldehyde group of the PLP coenzyme forms an imine with the conserved active site lysine. Second, the external aldimine is formed upon the imine exchange between the ζ-amino group of the lysine and the α-carbon amine of the substrate.

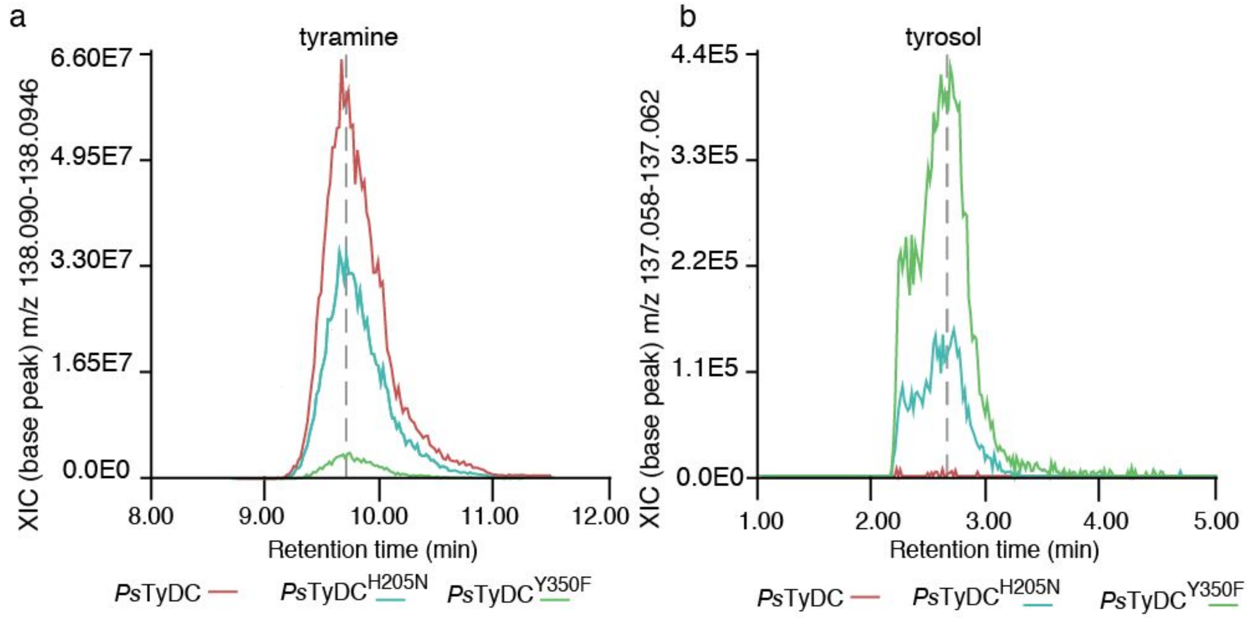


Fig. S10 | *S. cerevisiae* expressing various *Ps*TyDC mutants displaying altered ectopic tyramine and tyrosol levels (**a**) Relative in vivo tyrosine decarboxylase activities of the wild type *Ps*TyDC, *Ps*TyDC$^{\mathrm{H204N}}$ and *Ps*TyDC$^{\mathrm{Y350}}$ enzymes as observed though ectopic tyramine levels. (**b**) Relative in vivo tyrosine aldehyde synthase activities of the wild type *Ps*TyDC, *Ps*TyDC$^{\mathrm{H204N}}$ and *Ps*TyDC$^{\mathrm{Y350}}$ enzymes as observed though ectopic tyrosol levels.

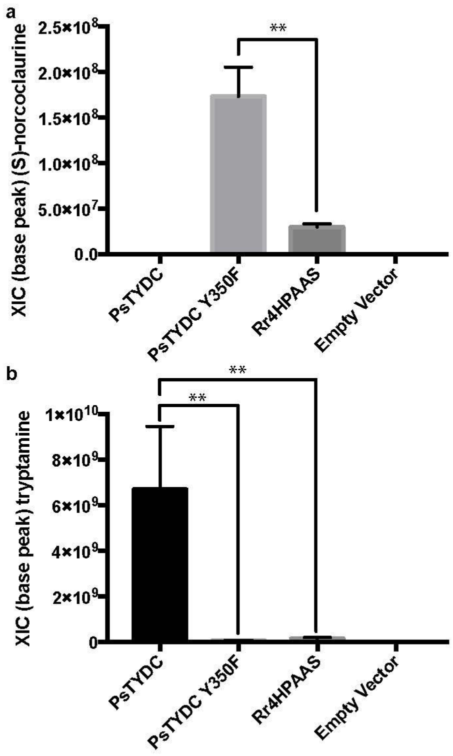


Fig. S11 | Metabolite profiling of tyramine and tyrosol in transgenic *S. cerevisiae* (**a**) Transgenic yeast overexpressing *Ps*TyDC H204N and *Ps*TyDC Y350F display reduces intracellular tyramine in comparison to transgenic yeast overexpressing wild type *Ps*TyDC. (**b**) Transgenic yeast overexpressing *Ps*TyDC H204N and *Ps*TyDC Y350F demonstrates increased intracellular tyrosol in comparison to transgenic yeast overexpressing wild type *Ps*TyDC. Cultures were grown and metabolically profiled in triplicate. * Adjusted P Value < 0.02, **Adjusted P Value < 0.002, ***Adjusted P Value < 0.0001

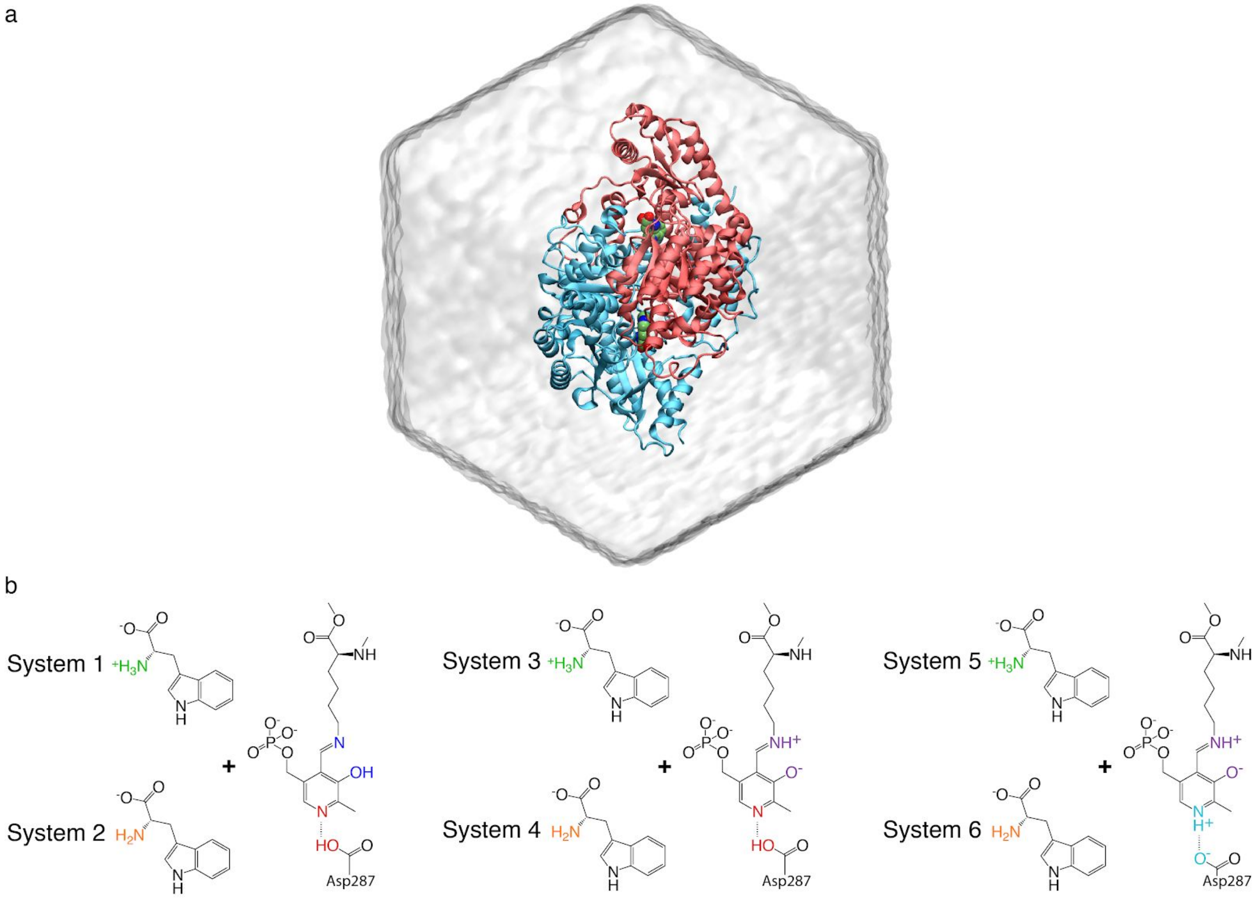


Fig. S12 | MD simulation systems of holo- *Cr*TDC with LLP and L-tryptophan in different protonation states. (**a**) The dodecahedron simulation box with the two monomers of *Cr*TDC colored in red and blue, respectively. Water molecules are shown as transparent surfaces. (**b**) Six holo- *Cr*TDC systems simulated in this work.

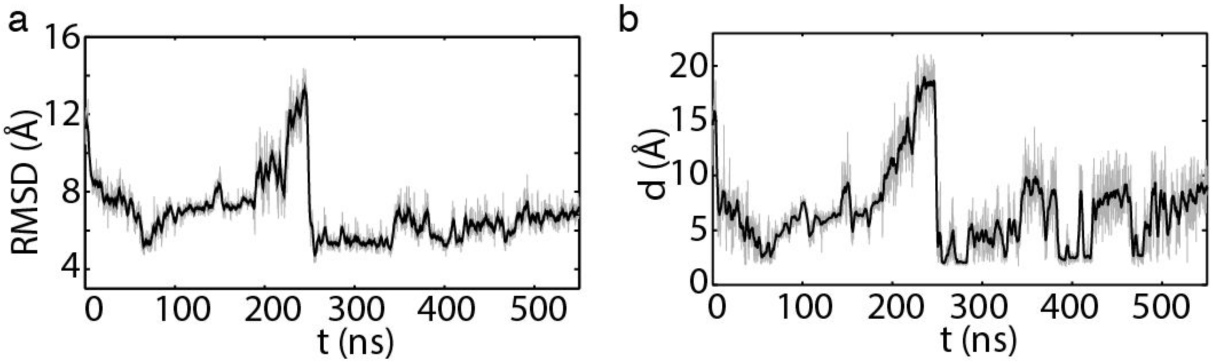


Fig. S13 | Large loop conformation as measured by RMSD and atomistic distances in the 550-ns simulation of *Cr*TDC system 1. (a) RMSD of large loop C_α atoms with respect to the modeled closed-state *Cr*TDC. (b) The minimal distance between $\text{Tyr}^{348\text{-B}}$ and $\text{His}^{203\text{-A}}$. Black curves represent running averages (window size: 101) performed on data colored in gray.

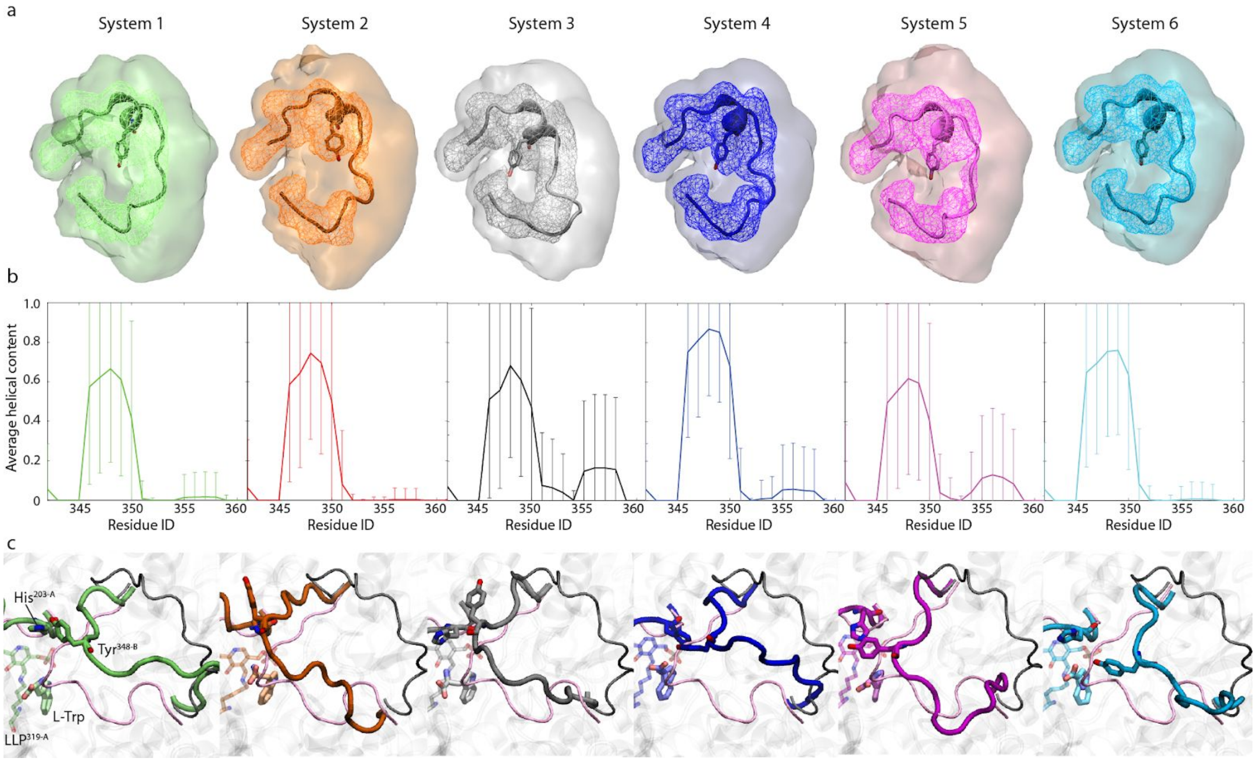


Fig. S14 | Large loop conformations revealed by MD simulations of holo- *Cr*TDC. (**a**) Clustering analysis and occupancy calculation results performed on the six replicas of 100-ns simulations of each *Cr*TDC system. Centroid structure of the largest cluster from clustering analysis is shown in Cartoon representation, where a short helix (residues 346-350) is seen across all systems. Isosurfaces of 30% and 1% occupancy are shown in wireframes and transparent surfaces, respectively. (**b**) Average helical content of the large loop in the simulations described in (**a**). Error bars indicate standard deviations. (**c**) Snapshots from selected 50-ns simulations of *Cr*TDC systems 1-6 initiated with the short helix in an unfolded state (Table S7). Structure of the large loop in this unfolded state is shown in black thin tube, with the closed conformations of the loops from crystal structure shown in pink thin tubes.

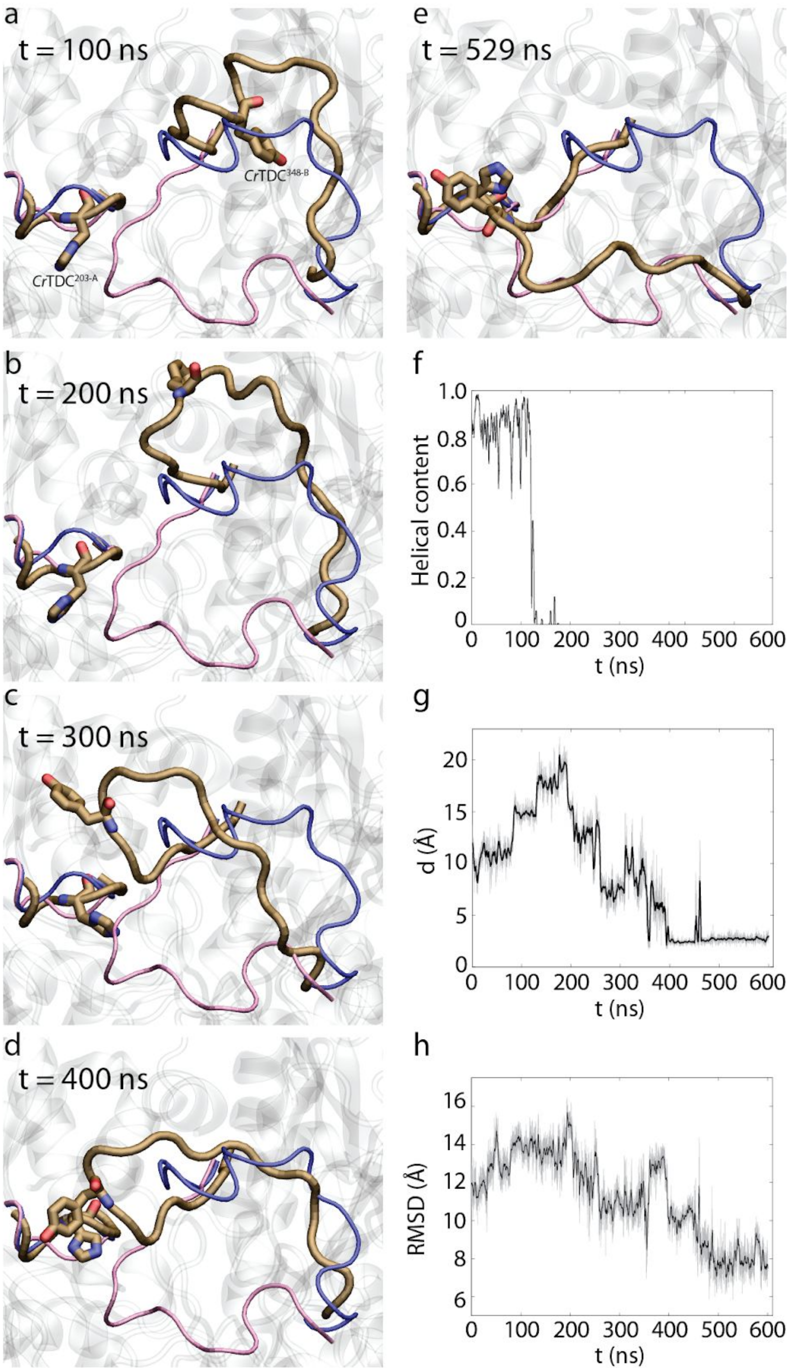


Fig. S15 | Large loop conformations revealed by a 600-ns MD simulation of apo- *Cr*TDC. (**a**-**e**) Simulation snapshots with residues Tyr$^{348\text{-}B}$ and His$^{203\text{-}A}$ highlighted. Loop conformations in the open and the modeled closed-state *Cr*TDC are colored in blue and red, respectively. (**f**) Helical content of the large loop during the 600-ns apo- simulation. Note that the loss of helical content precedes the large-scale loop closing motion shown in (**a**-**e**). (**g**) Minimal distance between His$^{203\text{-}A}$ and Tyr$^{348\text{-}B}$. (**h**) C$_\alpha$ RMSD of the large loop with respect to the modeled closed-state *Cr*TDC.

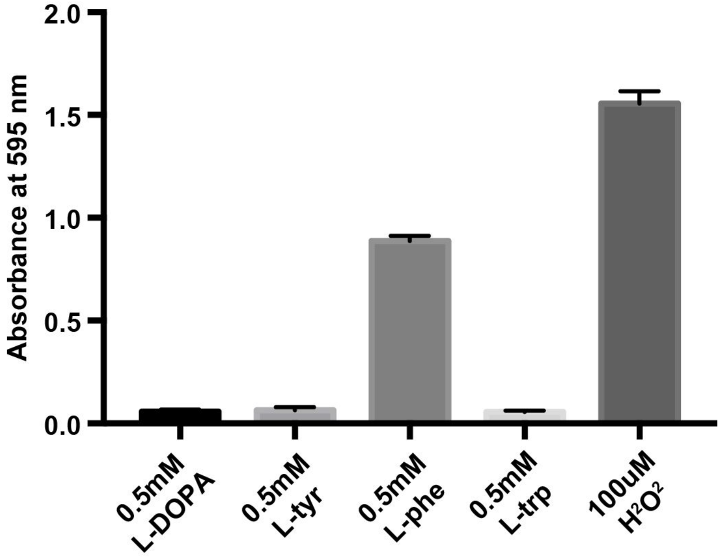


Fig. S16 | Relative selectivity of EgPAAS towards various aromatic amino acid substrates. Acetaldehyde synthesis activity of the *Eg*PAAS was measured against various aromatic amino acid substrates through the detection of the hydrogen peroxide co-product via Pierce Quantitative Peroxide Assay Kit (Pierce) against a hydrogen peroxide standard.

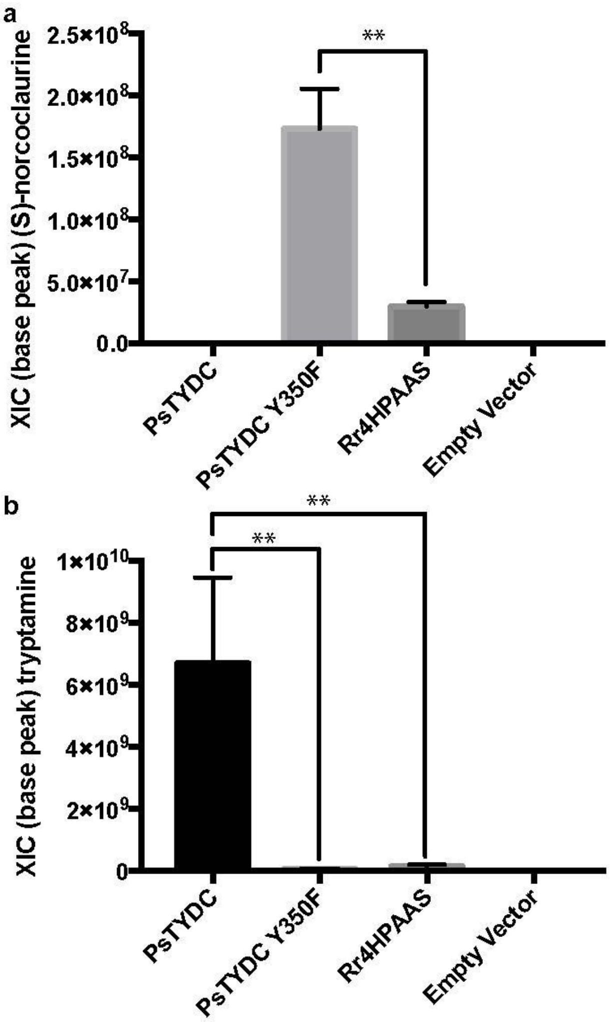


Fig. S17. Metabolite profiling of (S)-norcoclaurine and tyramine in transgenic *S. cerevisiae.* (**a**) Intracellular (S)-norcoclaurine in transgenic yeast containing a l-tyrosine overproduction multi gene plasmid (ARO4 K229L and ARO7 G141S) in addition to a (S)-norcoclaurine overproduction plasmid containing *Pp*DDC, *Ps*NCS2, *Bv*TyH and the listed plant AAAD. (**b**) Intracellular tyramine in transgenic yeast containing a l-tyrosine overproduction multi gene plasmid (ARO4 K229L and ARO7 G141S) in addition to a (S)-norcoclaurine overproduction plasmid containing *Pp*DDC, *Ps*NCS2, *Bv*TyH and the listed plant AAAD. Cultures were grown and metabolically profiled in triplicate. **Adjusted P Value < 0.002.

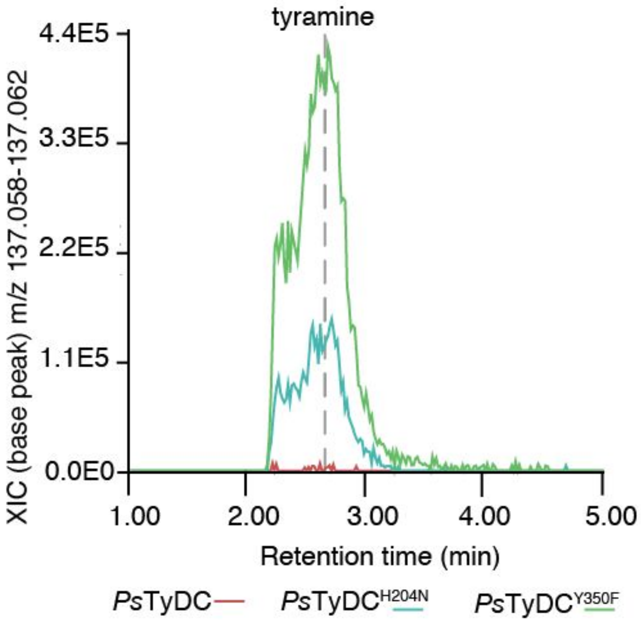


Fig. S18 | Metabolite profiling of tyramine in transgenic *S. cerevisiae* containing (S)-norcoclaurine multi gene vectors. Tyramine production in the media of *ARO4* and *ARO7*-overexpressing yeast transformed with (S)-norcoclaurine multigene vectors containing *PsTyDC*, *PsTyDC*^{H203N}, or *PsTyDC*^{Y350F}.

Supplementary Video | Trajectory of a 550-ns simulation of *Cr*TDC System 1. The large loop reached a semi-closed state during this simulation, with Tyr$^{348\text{-}B}$ and His$^{203\text{-}A}$ in frequent contact. For visualization clarity, water molecules and a large part of *Cr*TDC were not shown. Image smoothing was performed with a window size of 5 frames, which may have produced slight distortion of certain structures.

Supplementary Notes

Supplementary Note 1

To investigate their evolutionary history, AAAD sequences from chlorophytes, plants and select eubacteria were curated from Phytozome V12.1 and NCBI for phylogenetic analyses. Results illustrate a bacterial and chlorophytes clade in addition to three evolutionary distinct plant AAAD groups annotated as the basal, TyDC and TDC clades, respectively (Fig. 1a and Fig. S1). The basal clade contains AAAD sequences from the full taxonomic breadth of the surveyed plants including basal bryophytes and lycophytes lineages and is mostly closely related to the accessorial chlorophytes and bacterial clade (Fig. 1a and Fig. S1 and S2). Biochemically characterized enzymes from this clade demonstrates the full suite of AAAD biochemical functions including the functionally characterized *A. thaliana* PAAS (NP_849999)[5], *A. thaliana* TyDC (NP_001078461)[6], *Capsicum annuum* TDC (ACN62126)[7], and *Oryza sativa* (AK103253) TDC[8]. Upon the emergence of angiosperms (flowering plants), this progenitor clade speciated into two additional AAAD clades with seemingly more stringent biochemical functions (Fig. S2). Previously characterized enzymes from each neofunctionalized clade demonstrate exclusive indolic and phenolic amino acid substrate specificity, respectively. The indolic substrate selectivity of the TDC clade has been well documented through the biochemical characterization of selection of enzymes from taxonomically diverse species including *O. sativa* (XP_015648701)[8], *Camptotheca acuminate* (AAB39708, AAB39709)[9], *C. annuum* (NP_001312016)[7], *Ophiorrhiza pumila* (BAC41515)[10], and *C. roseus* (P17770)[11]. Likewise, the phenolic substrate specific TyDC clade is represented by characterized AAADs and AASs such as the *Petroselinum crispum* 4HPAAS (Q06086)[12], *Rosa hybrid* PAAS (ABB04522)[13], *Petunia* *hybrida P*AAS (ABB72475)[13], *Rhodiola rosea* 4HPAAS (MF674522)[14], *Thalictrum flavum* TyDC (AAG60665)[12], *Brachypodium distachyon* TyDC (XP_003569907)[15] and *P. somniferum* TyDC (AAC61842)[16]. The evolutionary history of plant AAADs suggests that biochemically flexible and evolutionary plastic progenitor AAADs present in lower plants speciated in angiosperms to yield a TDC clade with tryptophan selective decarboxylation activity and a TyDC clade with both decarboxylation and aldehyde synthase chemistry towards phenolic amino acid substrates.

Supplementary Note 2

All four enzymes purify and crystallize as homodimers, however, the *Rr*4HPAAS structure exists as a crystallographic monomer and the *Cr*TDC structure is represented by a crystallographic tetramer. Each monomer contains three distinct segments (Fig. S4a and S4b) previously described as domains[17]. This segments are unlikely to be stable as autonomously folding units, rather, each topologically associated segments supports the overall architecture of the α2-dimer. The principle hydrophobic association of the dimer is formed through the intermolecular antiparallel helices of the N-terminal segments (Fig. S5a and S5b). The middle and C-terminal segments mostly associate with each other and the other subunit through the LLP cofactor or water-mediated electrostatic interactions (Fig. S5c).

Supplementary Note 3

The residue range defined for the catalytic loops covers the area lacking significant secondary structure in the final *Cr*TDC model. The homologous sequence for the large loop in *Ps*TyDC, *At*PAAS and *Rr*4HPAAS corresponds to residues 344-363, 332-351 and 337-352, respectively. The small loop is represented in the *Ps*TyDC, *At*PAAS and *Rr*4HPAAS sequences by residues 202-207,190-195 and 195-200, respectively. In the CrTDC structure, the open conformation large loop lies on top of the upstream anchoring helix (*Cr*TDC residue range 335-341). This is particularly notable as the open conformation of this loop has not been observed in previously solved homologs. The large loop structure includes a single two turn helix containing the catalytic tyrosine. This loop helix interacts minimally with the preceding helix displaying tentative ionic interactions with Arg^{340}. A similar two turn helix is not observed in the large loop of the closed conformation *Ps*TyDC structure, possibly due to the missing electron density of *Ps*TyDC residues 354-359. The paralogous human HDC, however, displays a homologous two turn helix in the closed conformation suggest that the secondary structure of this loop may be important throughout the conformational change. The B-chain of the *At*PAAS structure displays a partially modeled catalytic loop in the open conformation. Residues 339-445 were not built into this catalytic loop as this sequence range displayed poor electron density. Likewise, the majority of the large loop was not modeled in the *Rr*4HPAAS structure as there was poor electron density support.

Supplementary Note 4

The lysine-conjugated coenzyme PLP is simulated in either the enolimine (systems 1-2) or the ketoenamine form (systems 3-6) shown in Fig. S12. In aqueous solution, PLP aldimine is known to undergo reversible proton tautomerism between these two forms[18–20]. Although when conjugated with an enzyme, the ketoenamine form is believed to dominate[18,19], we decided to simulate both forms for the sake of completeness. The electron density map of *Cr*TDC suggests an electron shared between the pyridine nitrogen of LLP and Asp287. While PROPKA calculation supports a protonated Asp287, a deprotonated Asp287 is known to stabilize LLP with its pyridine nitrogen protonated[21–23]. Therefore, while we modeled the enolimine form of LLP with Asp287 protonated (systems 1-2), both states of this residue were modeled in the ketoenamine form of the coenzyme (systems 3-6). The substrate L-tryptophan, which is a zwitterion at pH=7, is expected to lose the proton on its amine group prior to the formation of the external aldimine. Given that it is unclear when such deprotonation process occurs, we simulated L-tryptophan in both forms (Fig. S12). Taken together, six holo- *Cr*TDC systems were constructed (Table S4) and six replicas of 100-ns simulations were initially launched for each system. Analysis of these simulation trajectories reveals highly similar dominating conformation of the large loop, represented by the centroid structure from the largest cluster shown in Fig. S14a. The cartesian space visited by loop residues as enclosed by the occupancy isosurfaces (Fig. S14a) as well as the average helical content of the large loop measured by the program DSSP (Fig. S14b) are also similar across all six systems. These results suggest that in its transition from an open to a semi-closed state, conformational change of the large loop is not dictated by LLP and L-tryptophan protonation states. Indeed, in one of the three replicas of 600-ns apo *Cr*TDC simulations where neither PLP nor L-tryptophan was present, we observed a loop closing motion resembling that shown in Fig. 3g (Fig. S15). While interactions with LLP and L-tryptophan are expected to be relevant upon the large loop reaching its fully closed state and establishing canonical contacts with these molecules, our results shown above demonstrate that the initial loop closing motion is largely decoupled from the chemical details of the coenzyme and the substrate.

Supplementary Note 5

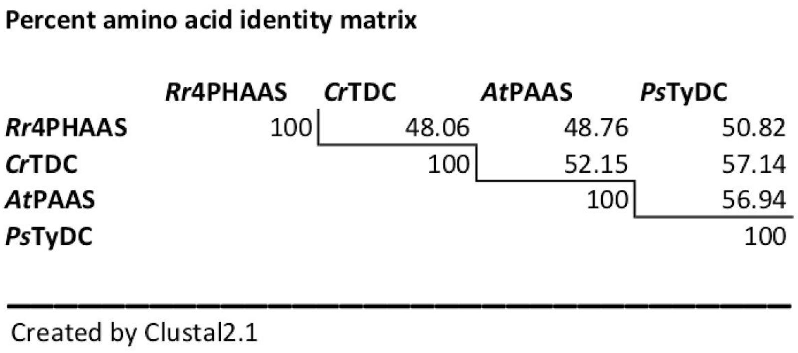
The model of the closed-state *Cr*TDC was constructed by superimposing the *Cr*TDC structure onto the closed conformation of *Ps*TyDC and subsequently threading the *Cr*TDC loops on the *Ps*TyDC structure. The thus generated *Cr*TDC model was used as a reference for MD results so that the difference between the structures can be measured via RMSD calculations.

Supplementary Table 1 | Data collection and structure refinement statistics.

Statistics for the highest-resolution shell are shown in parentheses.

	***Ps*TyDC**	***Cr*TyDC**	***AtPAAS***	***Rr*4HPAAS**
Wavelength	0.9793 Å	0.9793 Å	0.9793 Å	0.9793 Å
Resolution range	69.04 - 2.61 (2.703 - 2.61)	133.2 - 2.05 (2.123 - 2.05)	78.54 - 1.99 (2.061 - 1.99)	59.18 - 2.6 (2.693 - 2.6)
Space group	P 41 21 2	P 1 21 1	P 21 21 21	P 43 21 2
Unit cell	122.771 122.771 166.998 90 90 90	107.612 69.607 133.472 90 93.778 90	78.064 106.791 115.903 90 90 90	118.36 118.36 67.04 90 90 90
Total reflections	332655 (33114)	851987 (82988)	724886 (67838)	387791 (39193)
Unique reflections	39480 (3848)	123215 (9909)	67099 (5224)	15164 (1489)
Multiplicity	8.4 (8.6)	6.9 (6.8)	10.8 (10.2)	25.6 (26.3)
Completeness (%)	99.95 (99.74)	89.01 (80.54)	94.93 (78.61)	99.95 (99.66)
Mean I/sigma(I)	9.16 (1.59)	10.79 (0.58)	13.06 (0.77)	29.05 (2.97)
Wilson B-factor	34.24	42.66	41.64	60.13
R-merge	0.3504 (1.789)	0.1475 (2.547)	0.5424 (2.045)	0.3799 (1.006)
R-meas	0.373 (1.902)	0.1593 (2.763)	0.5699 (2.165)	0.3884 (1.025)
R-pim	0.1267 (0.644)	0.05974 (1.06)	0.173 (0.7018)	0.07923 (0.1962)
CC1/2	0.978 (0.558)	0.996 (0.362)	0.922 (0.267)	0.906 (0.909)
CC*	0.995 (0.846)	0.999 (0.729)	0.979 (0.65)	0.975 (0.976)
Reflections used in refinement	39464 (3847)	110319 (9907)	63725 (5216)	15158 (1484)
Reflections used for R-free	1979 (175)	1991 (175)	3228 (272)	1518 (149)
R-work	0.1872 (0.2909)	0.2201 (0.4525)	0.2045 (0.3889)	0.2108 (0.3298)
R-free	0.2345 (0.3360)	0.2557 (0.4525)	0.2398 (0.4182)	0.2686 (0.4258)
CC(work)	0.954 (0.769)	0.960 (0.635)	0.907 (0.738)	0.770 (0.637)
CC(free)	0.926 (0.680)	0.964 (0.627)	0.896 (0.732)	0.876 (0.537)
Number of non-hydrogen atoms	7985	15428	7655	3705
macromolecules	7670	15194	7443	3694
ligands	63	4	10	0
solvent	252	230	202	11
Protein residues	975	1882	939	467
RMS(bonds)	0.004	0.006	0.005	0.006
RMS(angles)	0.69	0.85	0.70	0.75
Ramachandran favored (%)	95.10	97.03	96.51	95.43
Ramachandran allowed (%)	4.38	2.75	3.17	3.91
Ramachandran outliers (%)	0.52	0.22	0.33	0.65
Rotamer outliers (%)	2.64	1.25	0.25	0.00
Clashscore	11.59	12.28	9.37	10.87
Average B-factor	37.54	63.45	61.76	67.80
macromolecules	37.31	63.63	61.94	67.84
ligands	83.57	76.38	131.64	0
solvent	32.91	51.43	51.57	54.73

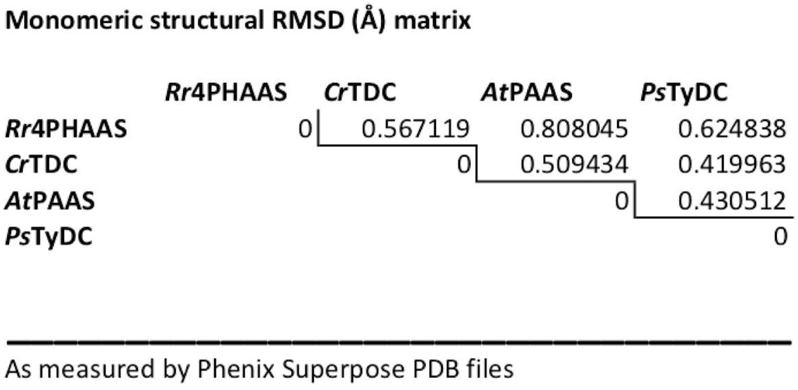
Supplementary Table 2 | Pairwise sequence identity between select AAAD proteins and the root-mean-square deviation (RMSD) between their monomeric structures.



Percent amino acid identity matrix

	***Rr*4PHAAS**	***Cr*TDC**	***At*PAAS**	***Ps*TyDC**
***Rr*4PHAAS**	100	48.06	48.76	50.82
***Cr*TDC**		100	52.15	57.14
***At*PAAS**			100	56.94
***Ps*TyDC**				100

Created by Clustal2.1



Monomeric structural RMSD (Å) matrix

	***Rr*4PHAAS**	***Cr*TDC**	***At*PAAS**	***Ps*TyDC**
***Rr*4PHAAS**	0	0.567119	0.808045	0.624838
***Cr*TDC**		0	0.509434	0.419963
***At*PAAS**			0	0.430512
***Ps*TyDC**				0

As measured by Phenix Superpose PDB files

Supplementary Table 3 | Cloning primers.

Gene	Vector/direction	Sequence
*At*PAAS	pHis8-4 Forward	GAAAACTTGTACTTCCAGGCCCATGGCATGGAAAATGGAAGCGGGAAGGTG
*At*PAAS	pHis8-4 Reverse	CTCGAATTCGGATCCGCCATGGTTACTTGTGAAGCAAGTAAGATGCTTCTTCCTG
*Ps*TyDC	pHis8-4 Forward	GAAAACTTGTACTTCCAGGCCCATGGCATGGGAAGCCTTCCGACTAATAACCTTG
*Ps*TyDC	pHis8-4 Reverse	CTCGAATTCGGATCCGCCATGGCTAGGCACCAAGTATGGCATCTGTATG
*Cr*TDC	pHis8-4 Forward	GAAAACTTGTACTTCCAGGCCCATGGCATGGGCAGCATTGATTCAACAAATGTAGC
*Cr*TDC	pHis8-4 Reverse	CTCGAATTCGGATCCGCCATGGTCAAGCTTCTTTGAGCAAATCATCGG
*Rr*4HPAAS	pHis8-4 Forward	GAAAACTTGTACTTCCAGGCCCATGGCATGGGCAGCTTGCCTTCTCCTAATG
*Rr*4HPAAS	pHis8-4 Reverse	CTCGAATTCGGATCCGCCATGGCTAAGACACGATGCTTTGAGCTGTTTCTTG
*Eg*PAAS	pTYB12 Forward	GTTGTTGTACAGAATGCTGGTCATATGACTAGTATGAACCCTCTCGATCCTGGAGAG
*Eg*PAAS	pTYB12 Reverse	CCGTCGACTCGCGAACTAGTTTATGCGGAATGTTGCTCACTGGC
*Ps*TyDC	H204N Forward	GTTTATGCTTCTAATCAAACCAACTGTGCACTTCAAAAAGCTG
*Ps*TyDC	H204N Reverse	CAGCTTTTTGAAGTGCACAGTTGGTTTGATTAGAAGCATAAAC
*Ps*TyDC	Y350F Forward	GCATTATCAACAAGTCCAGAATTCTTGAAGAACAAAGCAACGG
*Ps*TyDC	Y350F Reverse	CCGTTGCTTTGTTCTTCAAGAATTCTGGACTTGTTGATAATGC
*Cr*TDC	G370S Forward	CAAATCGCAACGAGCCGAAAATTTCGG
*Cr*TDC	G370S Reverse	CCGAAATTTTCGGCTCGTTGCGATTTG
*Ps*TyDC	p423TEF Forward	GCATAGCAATCTAATCTAAGTTTTCTAGAACTAGTATGGGAAGCCTTCCGACTAATAACC

*Ps*TyDC	p43TEF Reverse	CAGCCCGGGGGATCCACTAGTCTAGGCACCAAGTATGGCATCTGTATG
*Cr*TDC	p423TEF Forward	GCATAGCAATCTAATCTAAGTTTTCTAGAACTAGTATGGGCAGCATTGATTCAACAAATGTAGC
*Cr*TDC	p423TEF Reverse	CAGCCCGGGGGATCCACTAGTTCAAGCTTCTTTGAGCAAATCATCGG
*Ps*TyDC	pYTK001 Forward	GCATCGTCTCATCGGTCTCATATGGGAAGCCTTCCGACTAATAACC
*Ps*TyDC	pYTK001 Reverse	ATGCCGTCTCAGGTCTCAGGATCTAGGCACCAAGTATGGCATCTGTATG
*Rr*4HPAAS	pYTK001 Forward	GCATCGTCTCATCGGTCTCATATGGGCAGCTTGCCTTCTCCTAATG
*Rr*4HPAAS	pYTK001 Reverse	ATGCCGTCTCAGGTCTCAGGATCTAAGACACGATGCTTTGAGCTGTTTCTTG
*Pp*DDC	pYTK001 Forward	GCATCGTCTCATCGGTCTCATATGACGCCCGAGCAATTCAGACAG
*Pp*DDC	pYTK001 Reverse	ATGCCGTCTCAGGTCTCAGGATCTATCCCTTAATAACGTCCTGAAGTCTAGCCC
*Ps*NCS2	pYTK001 Forward	GCATCGTCTCATCGGTCTCATATGAGGAAAGTCATAAAATACGATATGGAGGTTGC
*Ps*NCS2	pYTK001 Reverse	ATGCCGTCTCAGGTCTCAGGATTTACAAAAGCCTGGGAATATCTGGGC
*Bv*TyH	pYTK001 Forward	GCATCGTCTCATCGGTCTCATATGGACAACACGACGTTAGCATTG
*Bv*TyH	pYTK001 Reverse	ATGCCGTCTCAGGTCTCAGGATTTACTTCCTTGGGACCGGGATTACC

Supplementary Table 4 | List of MD production runs performed in this work. Simulations marked with * were initiated from the end structure of a metadynamics run where the short helix (residues 346-350) on the large loop was forced to unfold (see Methods).

*Cr*TDC state	LLP and L-tryptophan protonation states	MD simulations performed
Holo	System 1	6 replicas of 100-ns runs with one run extended to 550 ns
		12 replicas of 50-ns runs*
	System 2	6 replicas of 100-ns runs
		12 replicas of 50-ns runs*
	System 3	6 replicas of 100-ns runs
		12 replicas of 50-ns runs*
	System 4	6 replicas of 100-ns runs
		12 replicas of 50-ns runs*
	System 5	6 replicas of 100-ns runs
		12 replicas of 50-ns runs*
	System 6	6 replicas of 100-ns runs
		12 replicas of 50-ns runs*
Apo	-	3 replicas of 600-ns runs
Aggregated simulation time		~9.5 μs

Supplementary References

1. Gouet, P., Robert, X. & Courcelle, E. ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.* **31,** 3320–3323 (2003).
2. Gouet, P., Robert, X. & Courcelle, E. ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.* **31,** 3320–3323 (2003).
3. Stivala, A., Wybrow, M., Wirth, A., Whisstock, J. C. & Stuckey, P. J. Automatic generation of protein structure cartoons with Pro-origami. *Bioinformatics* **27,** 3315–3316 (2011).
4. Dunathan, H. C. Conformation and reaction specificity in pyridoxal phosphate enzymes. *Proceedings of the National Academy of Sciences* **55,** 712–716 (1966).
5. Gutensohn, M. *et al.* Role of aromatic aldehyde synthase in wounding/herbivory response and flower scent production in different Arabidopsis ecotypes. *Plant J.* **66,** 591–602 (2011).
6. Lehmann, T. & Pollmann, S. Gene expression and characterization of a stress-induced tyrosine decarboxylase from Arabidopsis thaliana. *FEBS Lett.* **583,** 1895–1900 (2009).
7. Park, S. *et al.* Induction of serotonin biosynthesis is uncoupled from the coordinated induction of tryptophan biosynthesis in pepper fruits (Capsicum annuum) upon pathogen infection. *Planta* **230,** 1197–1206 (2009).
8. Kang, S., Kang, K., Lee, K. & Back, K. Characterization of rice tryptophan decarboxylases and their direct involvement in serotonin biosynthesis in transgenic rice. *Planta* **227,** 263–272 (2007).
9. Noe, W., Mollenschott, C. & Berlin, J. Tryptophan decarboxylase from Catharanthus roseus cell suspension cultures: purification, molecular and kinetic data of the homogenous protein. *Plant Mol. Biol.* **3,** 281–288 (1984).
10. Yamazaki, Y., Sudo, H., Yamazaki, M., Aimi, N. & Saito, K. Camptothecin biosynthetic genes in hairy roots of Ophiorrhiza pumila: cloning, characterization and differential expression in tissues and by stress compounds. *Plant Cell Physiol.* **44,** 395–403 (2003).

11. De Luca, V., Marineau, C. & Brisson, N. Molecular cloning and analysis of cDNA encoding a plant tryptophan decarboxylase: comparison with animal dopa decarboxylases. *Proc. Natl. Acad. Sci. U. S. A.* **86,** 2582–2586 (1989).
12. Torrens-Spence, M. P. *et al.* Biochemical evaluation of a parsley tyrosine decarboxylase results in a novel 4-hydroxyphenylacetaldehyde synthase enzyme. *Biochem. Biophys. Res. Commun.* **418,** 211–216 (2012).
13. Kaminaga, Y. *et al.* Plant phenylacetaldehyde synthase is a bifunctional homotetrameric enzyme that catalyzes phenylalanine decarboxylation and oxidation. *J. Biol. Chem.* **281,** 23357–23366 (2006).
14. Torrens-Spence, M. P., Pluskal, T., Li, F.-S., Carballo, V. & Weng, J.-K. Complete Pathway Elucidation and Heterologous Reconstitution of Rhodiola Salidroside Biosynthesis. *Mol. Plant* **11,** 205–217 (2018).
15. Noda, S. *et al.* Evaluation of Brachypodium distachyon L-Tyrosine Decarboxylase Using L-Tyrosine Over-Producing Saccharomyces cerevisiae. *PLoS One* **10,** e0125488 (2015).
16. Facchini, P. J. & De Luca, V. Differential and tissue-specific expression of a gene family for tyrosine/dopa decarboxylase in opium poppy. *J. Biol. Chem.* **269,** 26684–26690 (1994).
17. Burkhard, P., Dominici, P., Borri-Voltattorni, C., Jansonius, J. N. & Malashkevich, V. N. Structural insight into Parkinson's disease treatment from drug-inhibited DOPA decarboxylase. *Nat. Struct. Biol.* **8,** 963–967 (2001).
18. Chan-Huot, M. *et al.* NMR Studies of Protonation and Hydrogen Bond States of Internal Aldimines of Pyridoxal 5′-Phosphate Acid–Base in Alanine Racemase, Aspartate Aminotransferase, and Poly-l-lysine. *J. Am. Chem. Soc.* **135,** 18160–18175 (2013).
19. Caulkins, B. G. *et al.* Protonation states of the tryptophan synthase internal aldimine active site from solid-state NMR spectroscopy: direct observation of the protonated Schiff base linkage to pyridoxal-5'-phosphate. *J. Am. Chem. Soc.* **136,** 12824–12827 (2014).

20. Sharif, S. *et al.* 15N Nuclear Magnetic Resonance Studies of Acid−Base Properties of Pyridoxal-5‘-Phosphate Aldimines in Aqueous Solution. *J. Phys. Chem. B* **111,** 3869–3876 (2007).

21. Facchini, P. J., Huber-Allanach, K. L. & Tari, L. W. Plant aromatic L-amino acid decarboxylases: evolution, biochemistry, regulation, and metabolic engineering applications. *Phytochemistry* **54,** 121–138 (2000).

22. Eliot, A. C. & Kirsch, J. F. Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations. *Annu. Rev. Biochem.* **73,** 383–415 (2004).

23. Major, D. T. & Gao, J. A Combined Quantum Mechanical and Molecular Mechanical Study of the Reaction Mechanism and α-Amino Acidity in Alanine Racemase. *J. Am. Chem. Soc.* **128,** 16345–16357 (2006).