Supplementary Information for

Structural basis for divergent and convergent evolution of catalytic machineries in plant aromatic amino acid decarboxylase proteins

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

[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, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.
[4]Department of Biology, Williams College, Williamstown, Massachusetts 01267, USA.
†Present address: School of Chemistry, University of Southampton, Southampton, SO17 1BJ, UK.

*Corresponding author: wengj@wi.mit.edu

Table of contents

Supplementary Fig. S1-19	3-21
Supplementary Video Information	22
Supplementary Note S1-3	23-24
Supplementary Table S1-4	25-29
Supplementary References	30

Supplementary Figures

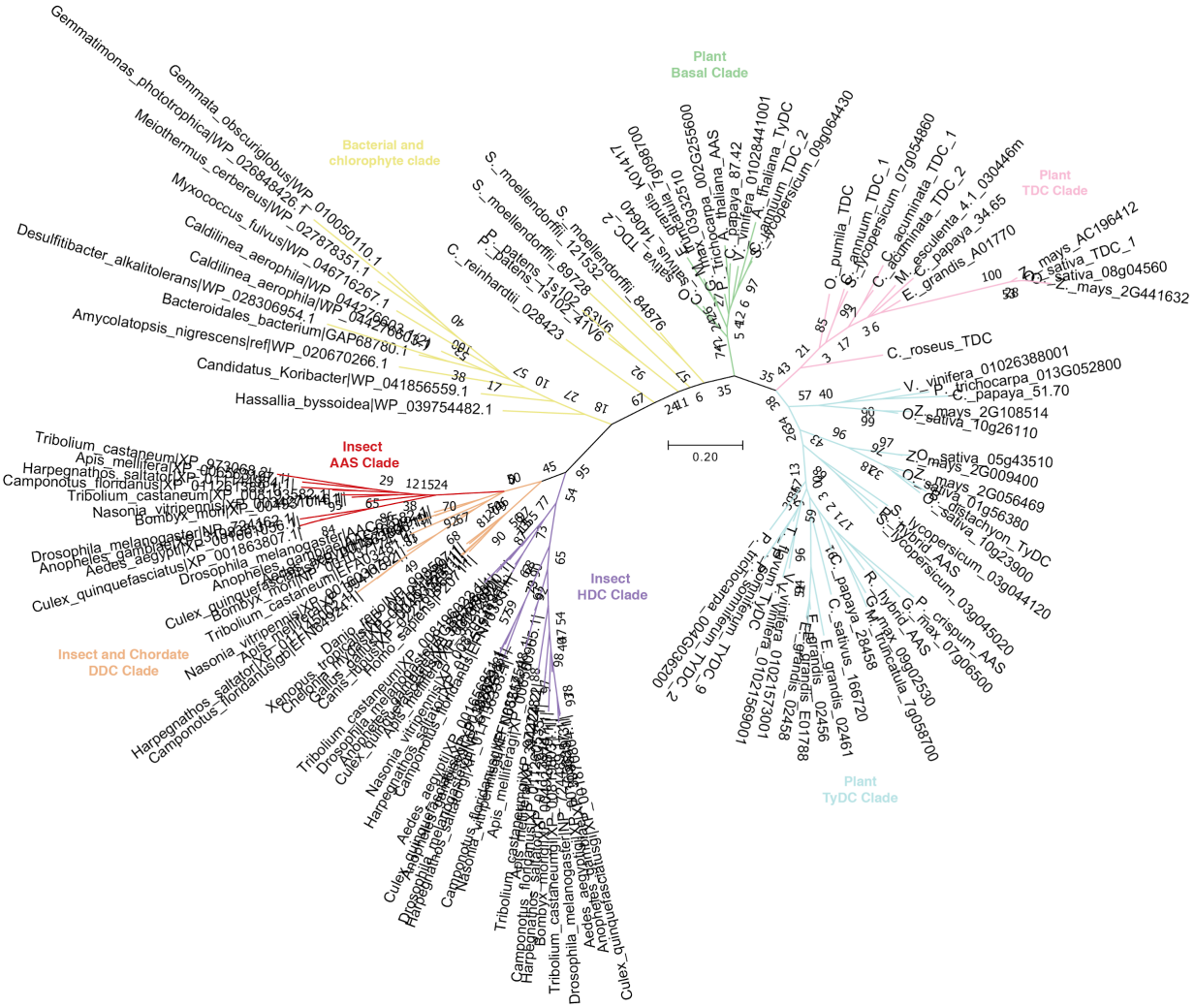

Fig. S1. A maximum-likelihood phylogenetic tree of select chordata, insect, bacterial, and plant AAADs. This tree is populated with sequences from all Phytozome V12 species, all attainable characterized NCBI plant AAAD sequences and select eubacteria, chordata, and insect NCBI sequences. Bootstrap values are indicated at the tree nodes. The bootstrap consensus unrooted trees were inferred from 500 replicates. The scale measures evolutionary distances in substitutions per amino acid. The green, pink and blue branches correspond to the basal, TDC and TyDC plant clades, respectively. The yellow branches correspond to the chlorophytes and bacterial AAAD sequences. The purple clade corresponds to insect histidine decarboxylase (HDC) sequences, while the orange clade represents insect and chordata DDC sequences. The red insect AAS clade emerged from the insect DDC clade and contains asparagine substitution in place of the typically considered active-site histidine. The evolutionary history of these enzymes indicates that animal and plant AAADs are monophyletic and have evolved independently from a universal common ancestor.

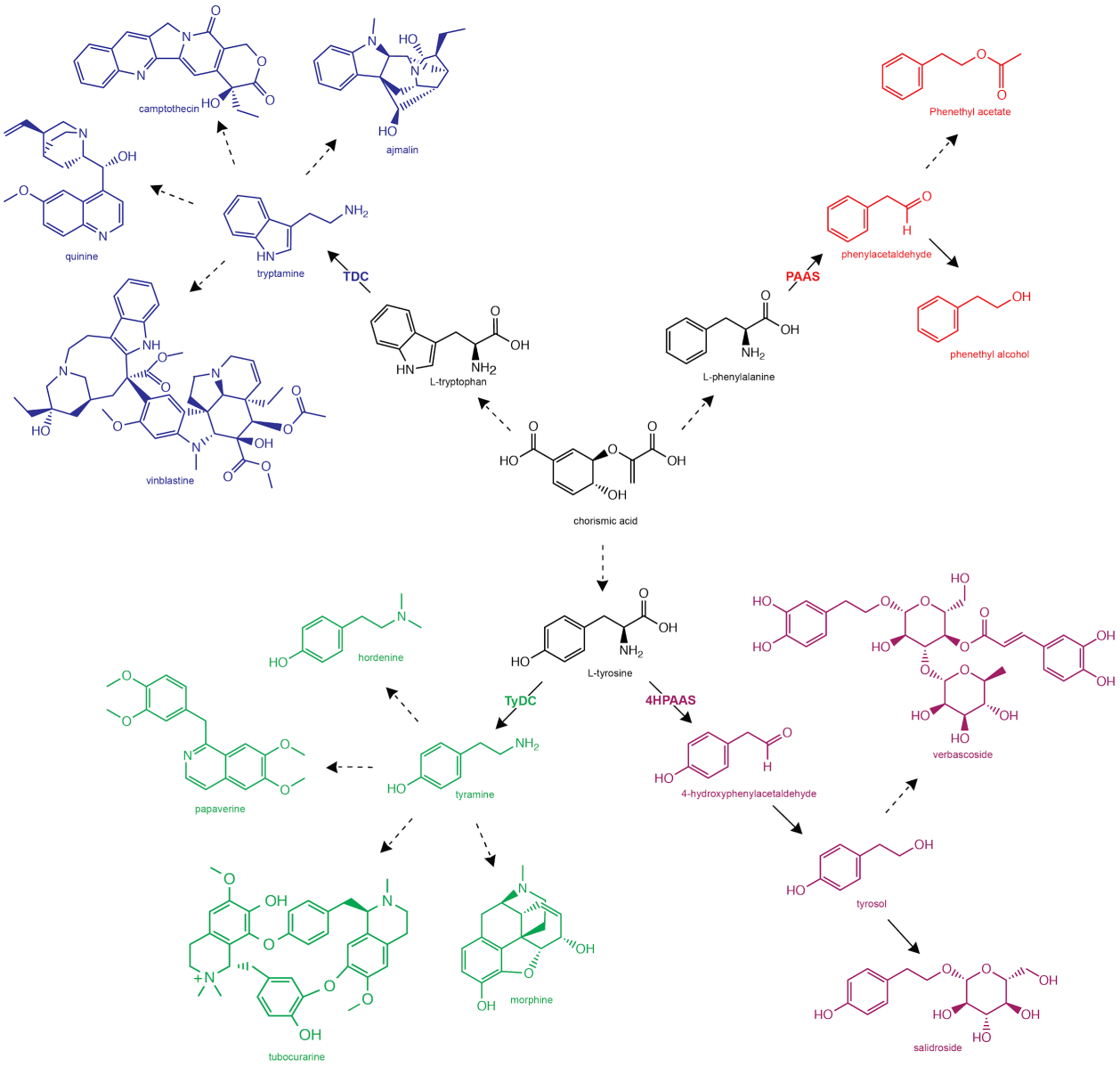

Fig. S2. Diverse specialized metabolic pathways downstream of various plant AAADs. In plants, the three proteinegic L-aromatic amino acids, L-tryptophan, L-tyrosine, and L-phenylalanine are all downstream of chorismate derived from the shikimate pathway. From these primary metabolites, plant AAADs catalyze the first biotransformation in divergent specialized metabolic pathways. The TDC enzyme and some downstream products are shown in blue, the PAAS and several downstream products are shown in red, the TyDC and select downstream products are shown in green, and the 4HPAAS and a few downstream products are shown in purple. Solid arrows represent single enzyme catalyzed reactions while the dotted arrows indicate multiple enzymatic steps.

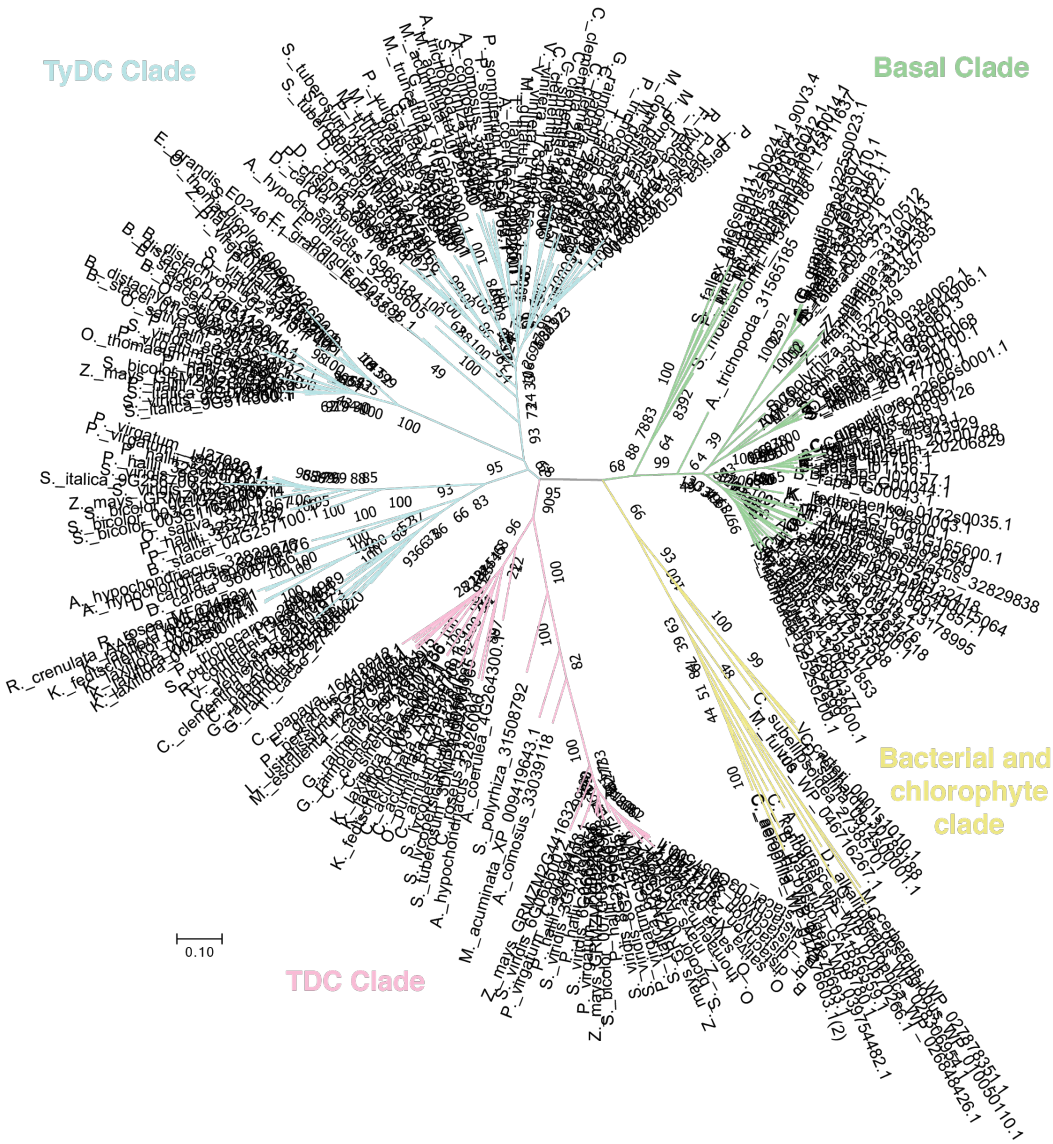

Fig. S3. A maximum-likelihood phylogenetic tree of AAADs. This tree is populated with sequences from all Phytozome V12 species, all attainable characterized plant AAAD sequences and select eubacteria from NCBI. Bootstrap values are indicated at the tree nodes. The bootstrap consensus unrooted trees were inferred from 500 replicates. The scale measures evolutionary distances in substitutions per amino acid. Green, pink and blue branches correspond to the 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 basal clade is most closely related to bacterial and chlorophyte AAADs) and their apparent substrate selectivity (the TDC clade contains a number of characterized enzymes with exclusive indolic substrate specificity, while the TyDC clade is represented by characterized enzymes with phenolic substrate selectivity). While the basal and TyDC clades contain substitutions impliciative of AAS chemistry, these mutations occur independently and sporadically through plant taxonomy.

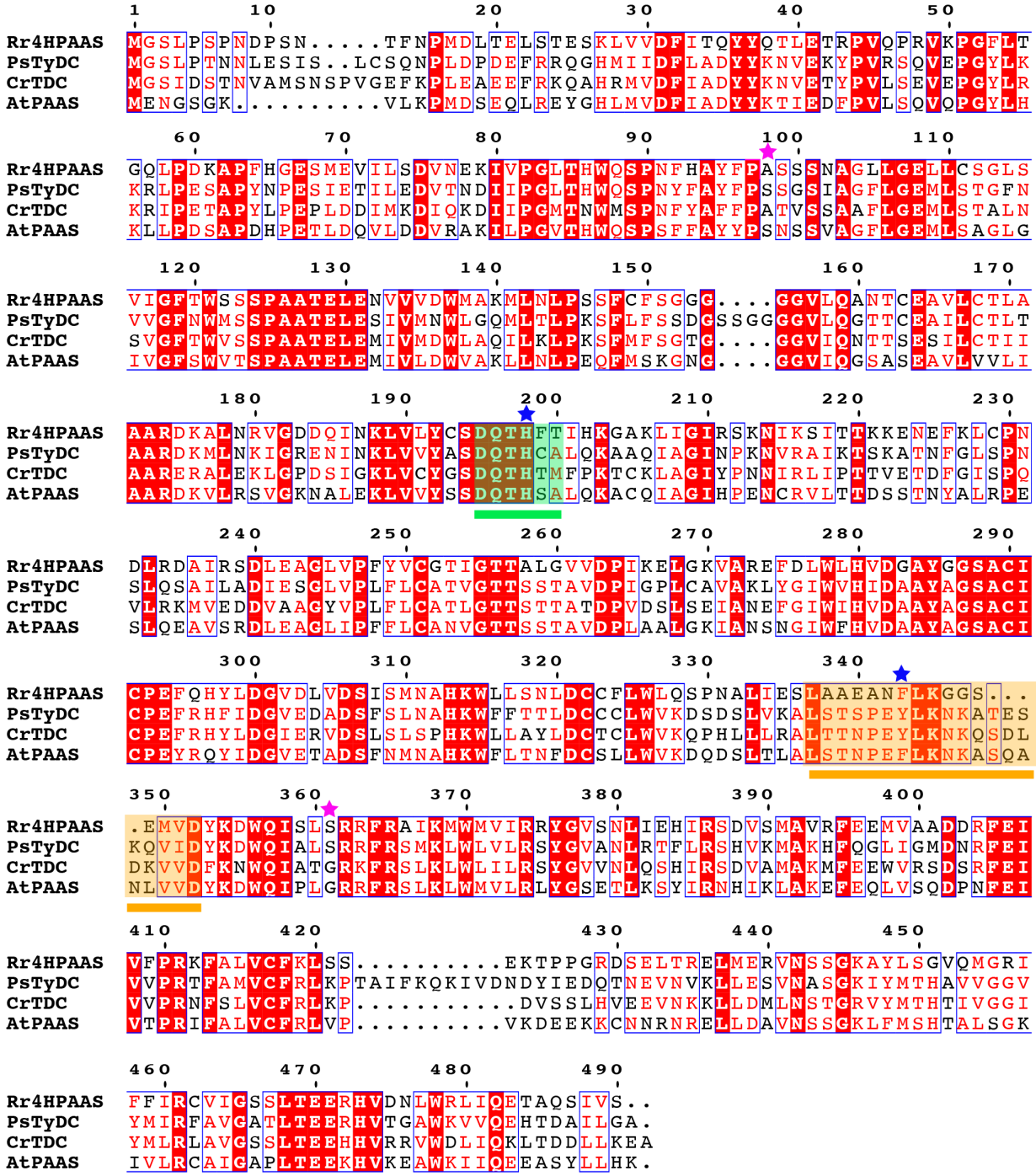

Fig. S4. Multiple sequence alignment of four crystallized plant AAADs. The short catalytic loop is highlighted and underlined in green and the large catalytic loop is highlighted and underlined in orange. The substrate selectivity residues are marked with pink stars and the catalytic mechanism dictating residues are marked with blue stars. The multiple sequence alignment was generated with ClustalW2 (1) and displayed with ESPript 3.0 (2).

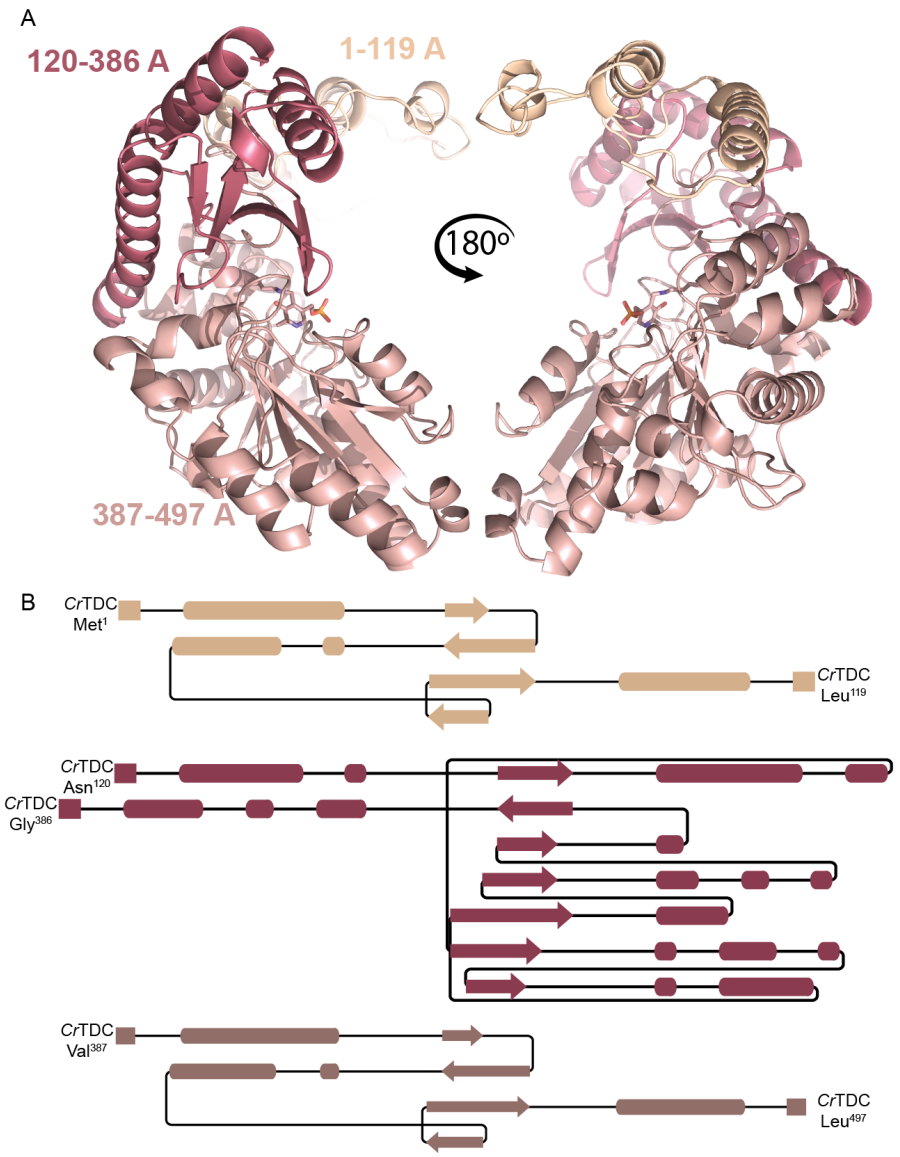

Fig. S5. Topology of plant AAADs as observed in *Cr*TDC. (*A*) Plant AAAD segments as displayed by the *Cr*TDC structure. Each monomer is composed of the N-terminal *Cr*TDC$^{1\text{-}119}$ (beige), middle *Cr*TDC$^{120\text{-}386}$ (maroon) and C-terminal *C*rTDC$^{387\text{-}497}$ (salmon) segments. (*B*) Topology diagram for each of the three *Cr*TDC segments. The segment diagrams were generated though Pro-origami using DSSP secondary structure program (3).

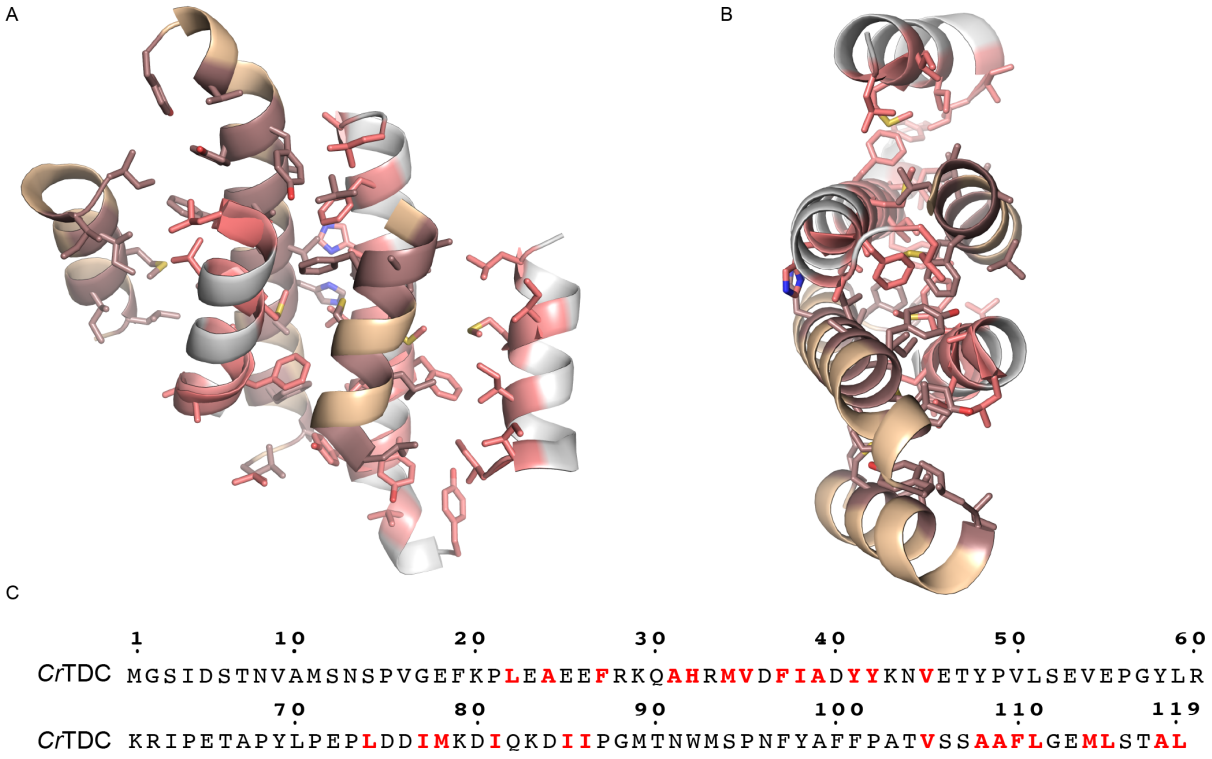

Fig. S6. Intermonomer association of N-terminal segments from two *Cr*TDC chains. Side (*A*) and top (*B*) views of the aromatic and hydrophobic residues forming the intermolecular interaction of the *Cr*TDC homodimer. One monomer is colored in orange with maroon hydrophobic or aromatic residues, whereas the second monomer is colored in white with pink hydrophobic or aromatic residues. (*C*) Sequence of the *Cr*TDC N-terminal segment with hydrophobic or aromatic residues involved in intermonomer association highlighted in red.

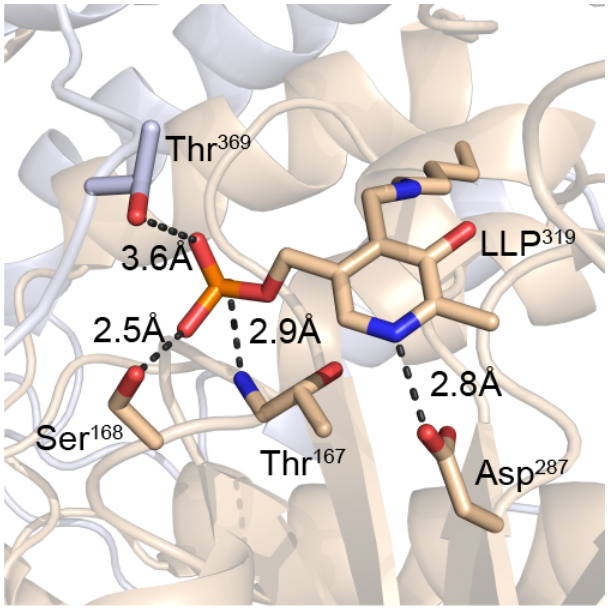

Fig. S7. Coordination of LLP by the *Cr*TDC active-site residues. Cartoon and stick representation of LLP coordination in *Cr*TDC. Chain A is colored in beige and chain B is colored in white.

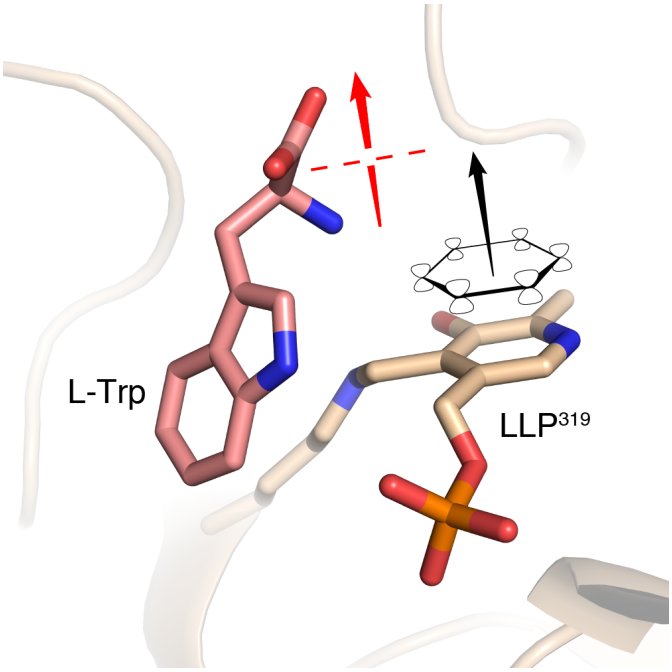

Fig. S8. Orientation of the alpha carbon carbonyl bond relative to the plane of the pyridoxal imine system. As per the Dunathan hypothesis, PLP enzymes exhibit stereospecific cleavage of bonds orthogonal to the pyridine ring pi-system electrons (shown as black ring and arrow) (4). In the case of PLP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

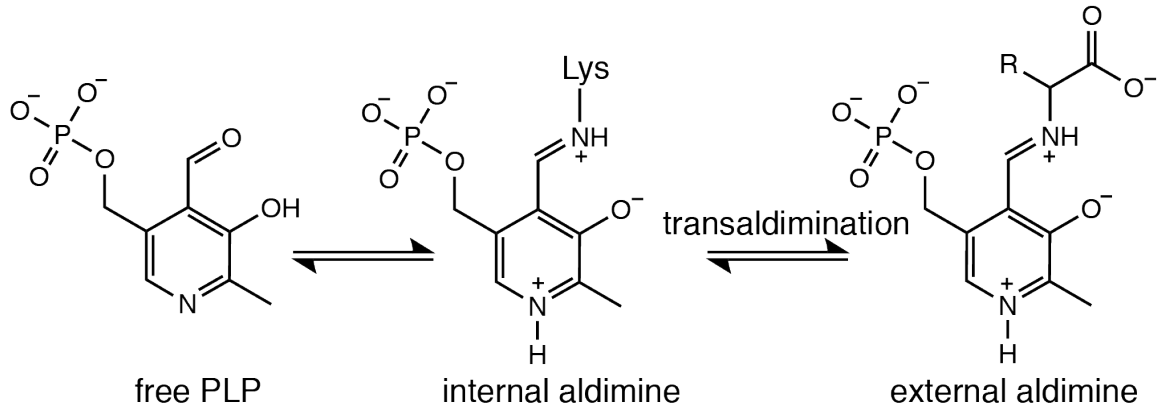

Fig. S9. Schematic of the formation of PLP internal and external aldamines. First, the internal aldimine is formed when the aldehyde group of the PLP coenzyme forms an imine with the conserved active site lysine. Second, the external aldimine is formed upon the imine exchange between the ζ-amino group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

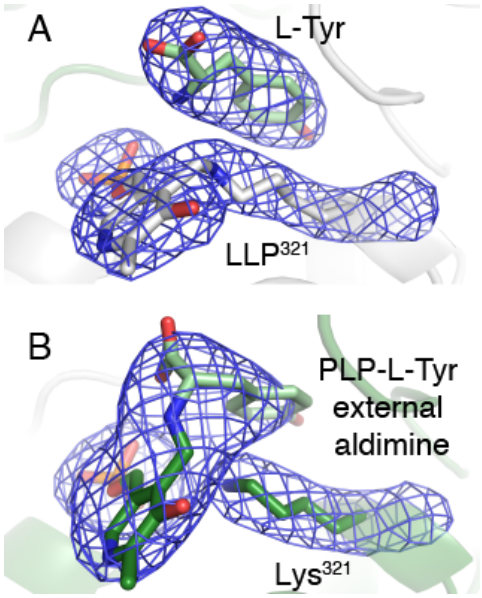

Fig. S10. Transaldimination as captured by two active sites of the *Ps*TyDC homodimer. (*A*) The PLP-Lys^{321} internal aldimine as modeled in one of the active sites of the *Ps*TyDC homodimer. (*B*) The PLP-L-tyrosine external aldimine as captured by the other active site of the *Ps*TyDC homodimer. 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.

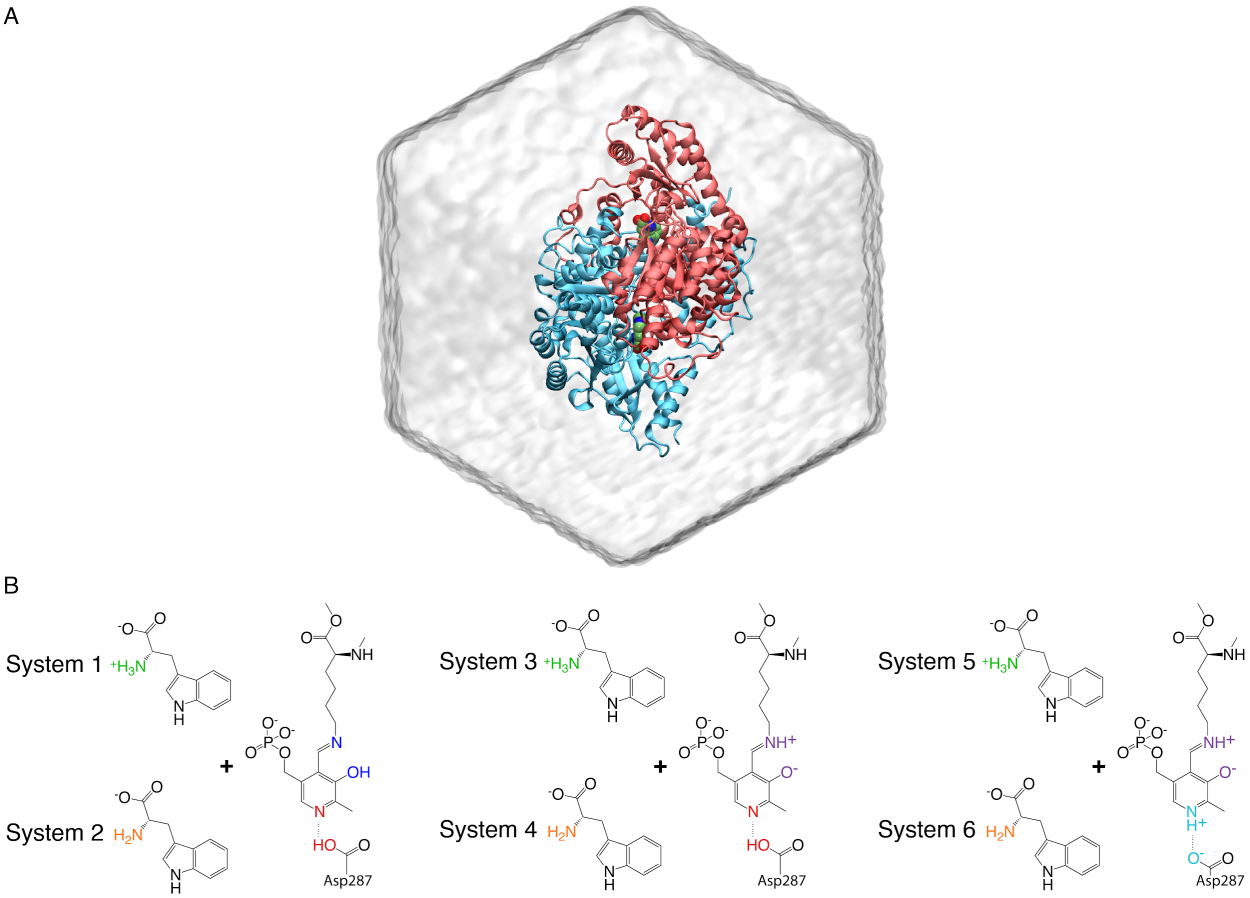

Fig. S11. MD simulation systems of holo- *Cr*TDC with LLP and L-tryptophan in different protonation states. (*A*) The dodecahedron simulation box with the two monomers of *Cr*TDC colored in red and blue, respectively. Water molecules are shown as transparent surfaces. (*B*) Six holo-*Cr*TDC systems simulated in this work.

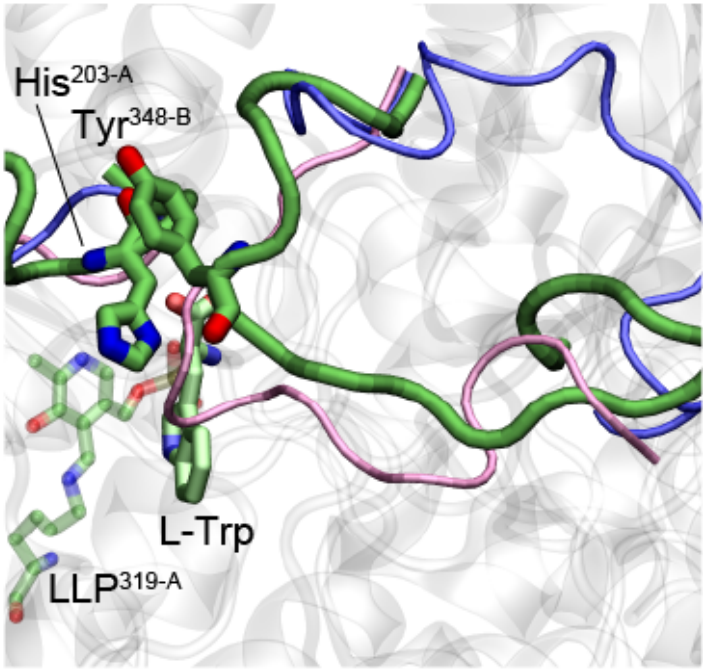

Fig. S12. *Cr*TDC MD semi-closed conformation. A snapshot from the MD simulation of *Cr*TDC System 1 at t=398 ns, exhibiting a semi-closed loop conformation. The open and closed conformations of the loops observed from the crystal structures are shown in blue and pink tubes, respectively.

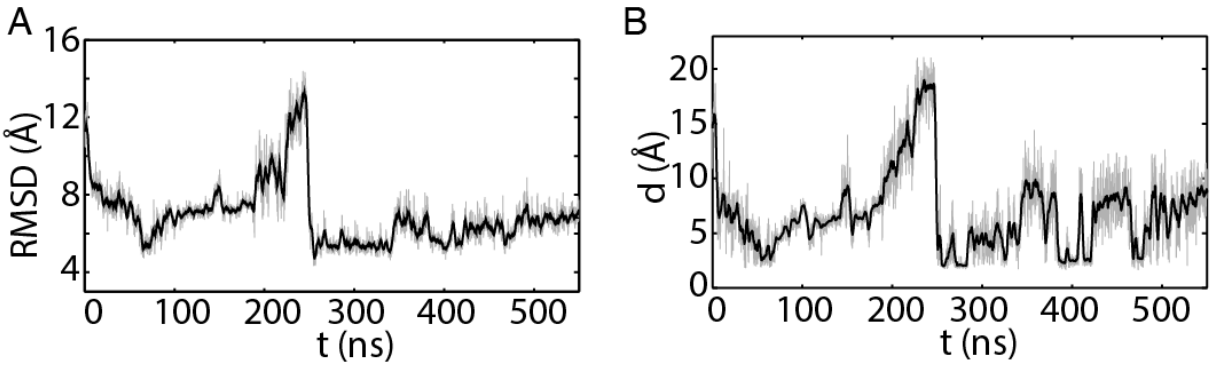

Fig. S13. Large loop conformation as measured by RMSD and atomistic distances in the 550-ns simulation of *Cr*TDC system 1. (*A*) RMSD of large loop C_α atoms with respect to the modeled closed-state *Cr*TDC. (*B*) The minimal distance between $Tyr^{348\text{-}B}$ and $His^{203\text{-}A}$. Black curves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

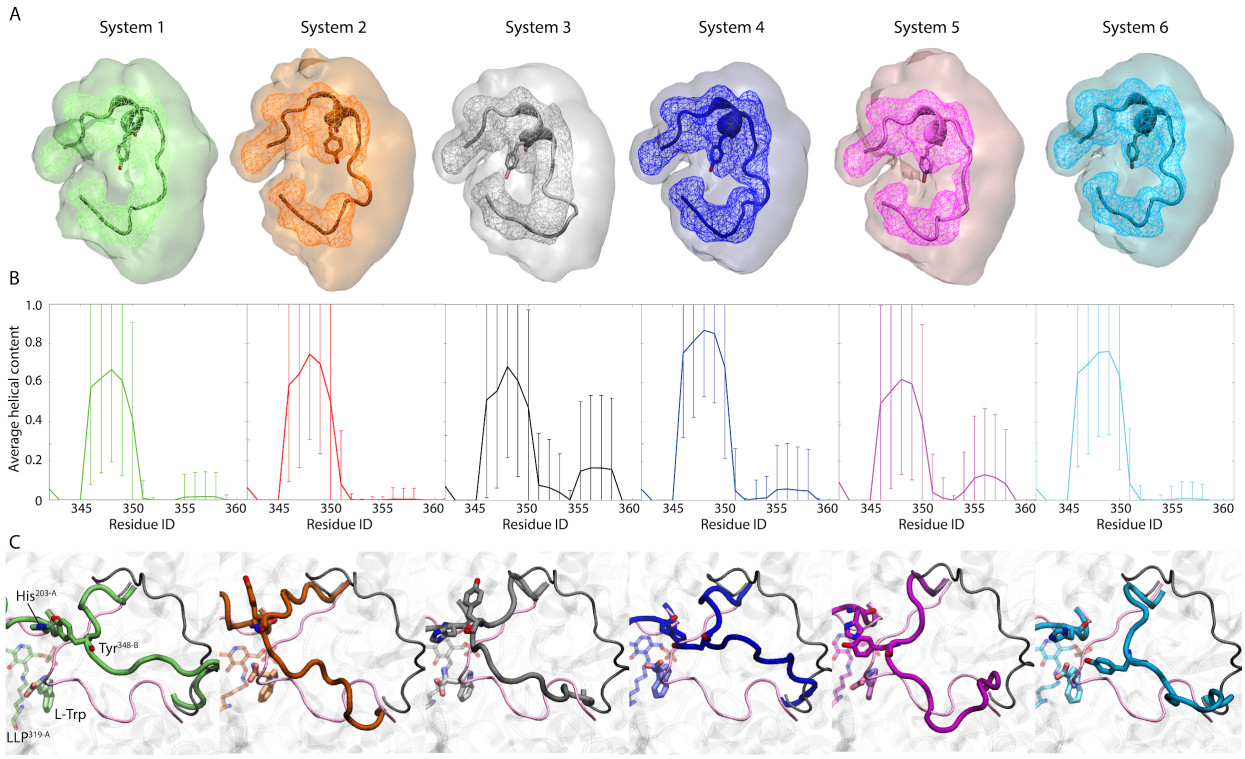

Fig. S14. Large loop conformations revealed by MD simulations of holo- *Cr*TDC. (*A*) Clustering analysis and occupancy calculation results performed on the six replicas of 100-ns simulations of each *Cr*TDC system. Centroid structure of the largest cluster from clustering analysis is shown in Cartoon 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 S4). Initial 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.

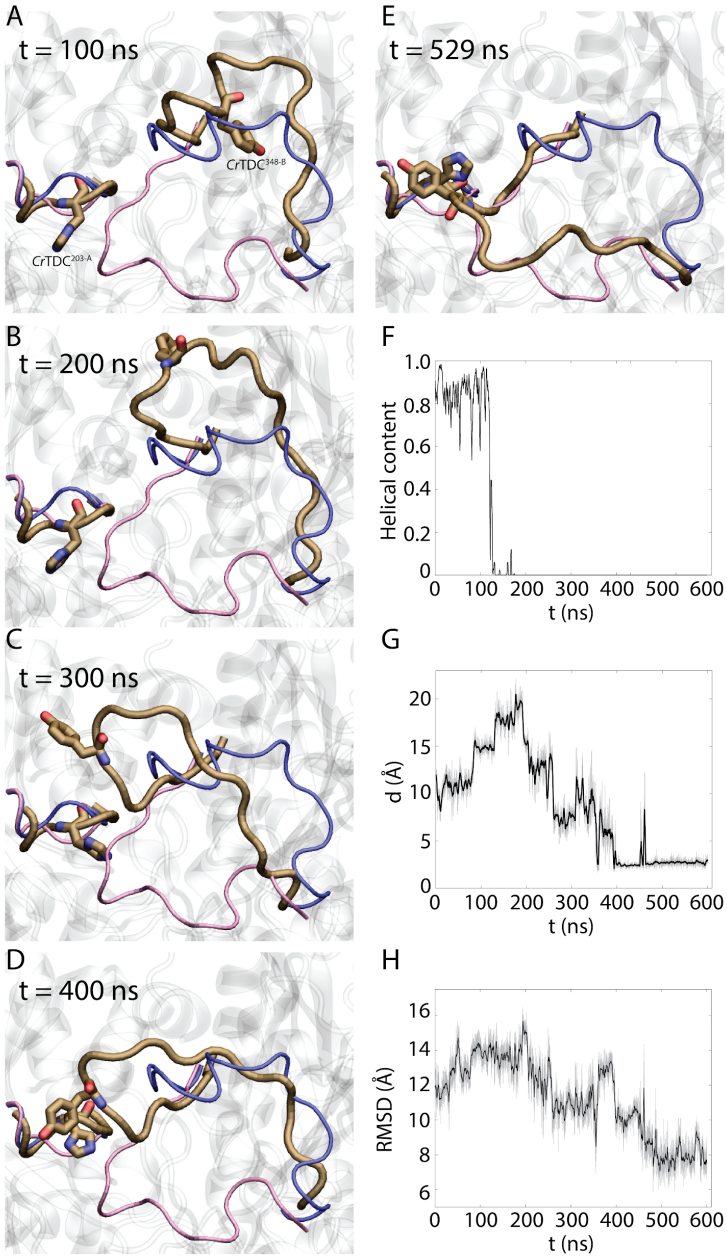

Fig. S15. Large loop conformations revealed by a 600-ns MD simulation of apo- *Cr*TDC. (*A-E*) Simulation snapshots with residues Tyr$^{348\text{-B}}$ and His$^{203\text{-A}}$ highlighted. Loop conformations in the open and the modeled closed-state *Cr*TDC are colored in blue and red, respectively. (*F*) 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_α RMSD of the large loop with respect to the modeled closed-state *Cr*TDC.

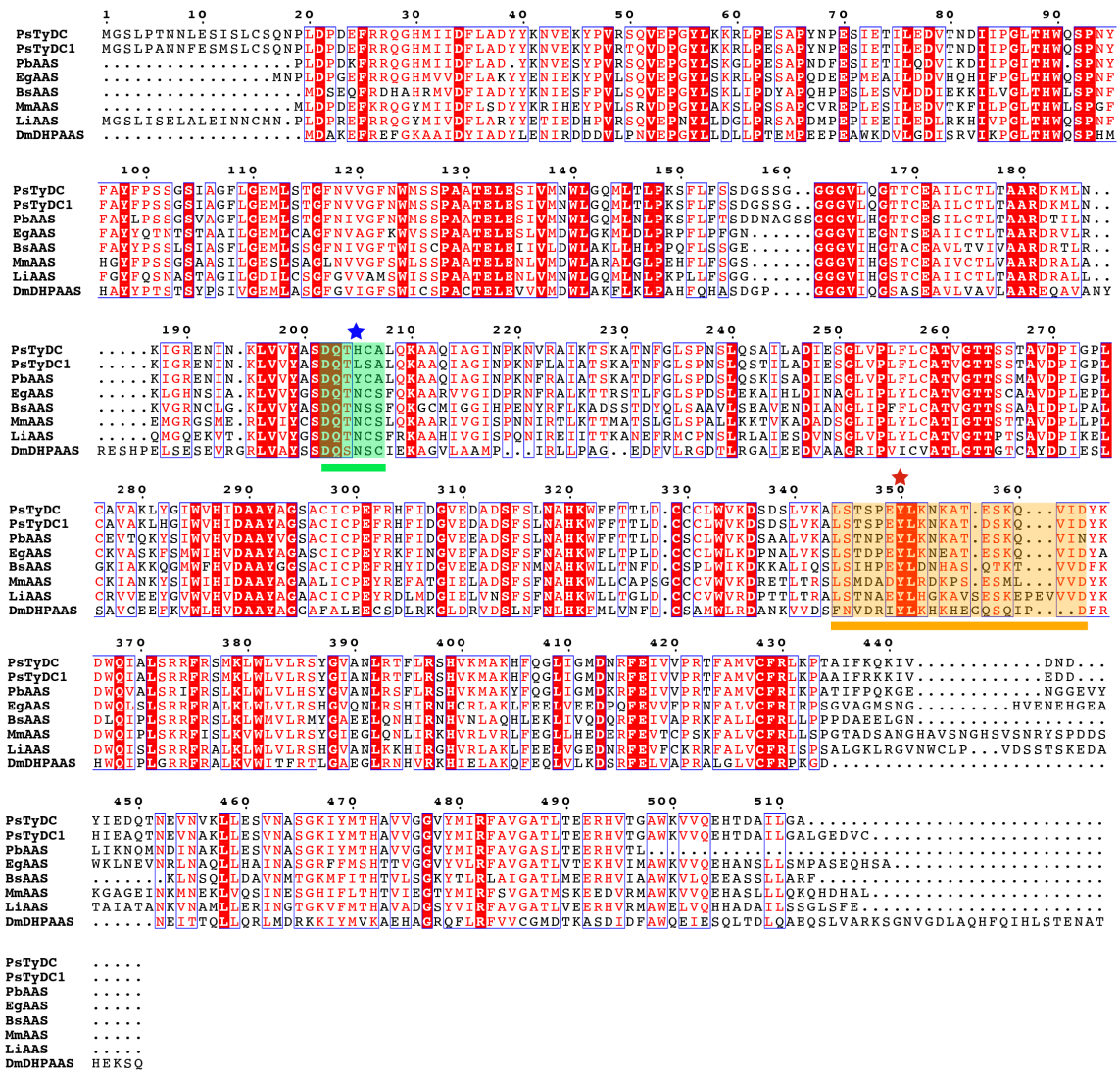

Fig. S16. Multiple sequence alignment of plant and insect AASs together with *Ps*TyDC highlighting naturally occurring substitutions at the small-loop histidine. The short catalytic loop is highlighted and underlined in green. The small-loop histidine conserved among canonical decarboxylases is labeled with a blue star. Variation in this residue may be implicative of alternative reaction mechanisms. The large loop is highlighted and underlined in orange. Although all the sequences display the red-stared catalytic tyrosine typically conserved in decarboxylation chemistry, *Eg*PAAS and *Drosophila melanogaster* 3,4-dihydroxyphenylacetaldehyde synthase (*Dm*DHPAAS) (5) display confirmed aldehyde synthase activity. *Ps*TyDC is crystallized in this study, while *Ps*TyDC1 (NCBI accession P54768) displays a small-loop histidine substitution but was previously annotated as a decarboxylase (6). *Papaver bracteatum* scaffold TMWO-2021544 (*Pb*AAS), B*egonia sp*. scaffold OCTM-2012326 (*Bs*AAS), *Medinilla magnifica* scaffold WWQZ-2007373 (*Mm*AAS), and *Lagerstroemia indica* scaffold RJNQ-2017655 (*Li*AAS) all contain small-loop His-to-Asn substitution characteristic of insect AASs. The multiple sequence alignment was generated using ClustalW2 (1) and displayed with ESPript 3.0 (2).

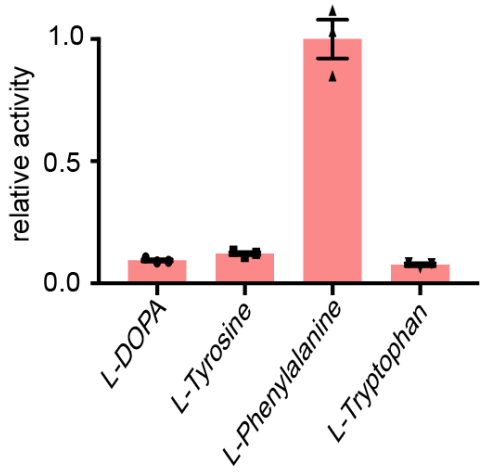

Fig. S17. Relative selectivity of *Eg*PAAS towards various aromatic L-amino acid substrates. AAS activity of *Eg*PAAS was measured against various L-aromatic amino acid substrates at 0.5 mM substrate concentration. The relative activity was quantified through detection of the hydrogen peroxide co-product using the Pierce Quantitative Peroxide Assay Kit (Pierce). Error bars indicate standard error of the mean (SEM) based on biological triplicates.

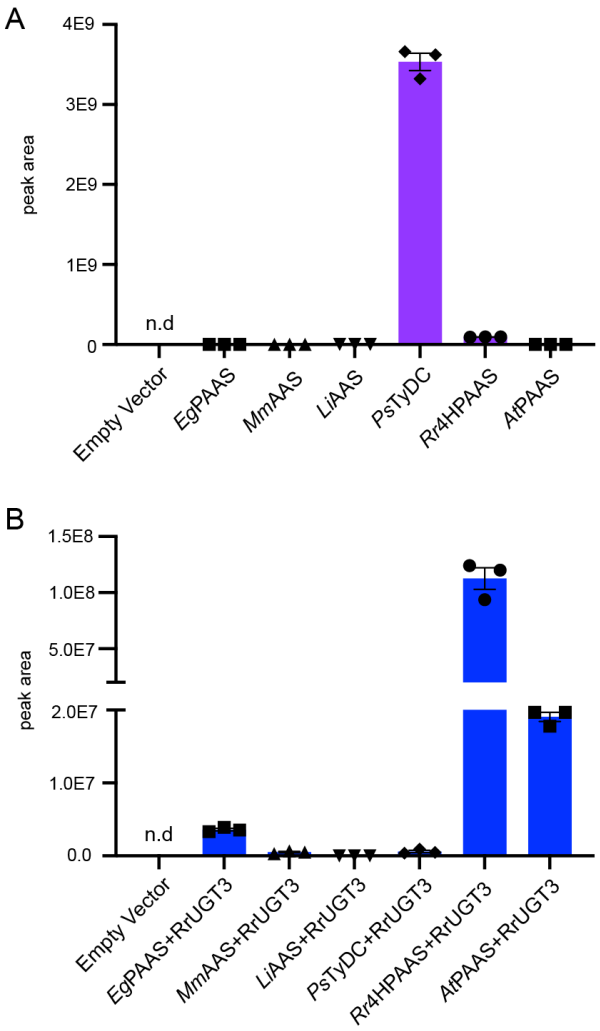

Fig. S18. Relative canonical AAAD activity and AAS activity of various plant AAADs as assessed by tyramine and icariside D2 production, respectively, in transgenic yeast. (*A*) Tyramine accumulation from *in vivo* decarboxylation of L-tyrosine by various plant AAAD proteins in transgenic yeast. *Eg*PAAS, *Mm*AAS (Phytozome: *M. magnifica* scaffold-WWQZ-2007373), and *Li*AAS (Phytozome: *L. indica* scaffold-RJNQ-2017655) contain the signature His-to-Asn substitution as observed in insect AASs. (*B*) Icariside D2 accumulation in yeast expressing various plant AAADs alongside the *R. rosea Rr*UGT3 required for the 4-*O*-glycosylation of tyrosol (7). The 4-hydroxyphenylacetaldehyde product of 4HPAAS is reduced endogenously in yeast to form tyrosol prior to 4-*O*-glycosylation (7). Note that phenylacetaldehyde and phenylethyl alcohol are highly volatile, therefore could not be retained in yeast culture to be measured by LC-MS. Therefore, the relative canonical AAAD activity and AAS activity of these enzymes were assessed against the substrate L-tyrosine. The addition of *Rr*UGT3 facilitates the more reliable LC-MS detection of the AAS activity. Error bars indicate standard error of the mean (SEM) based on biological triplicates.

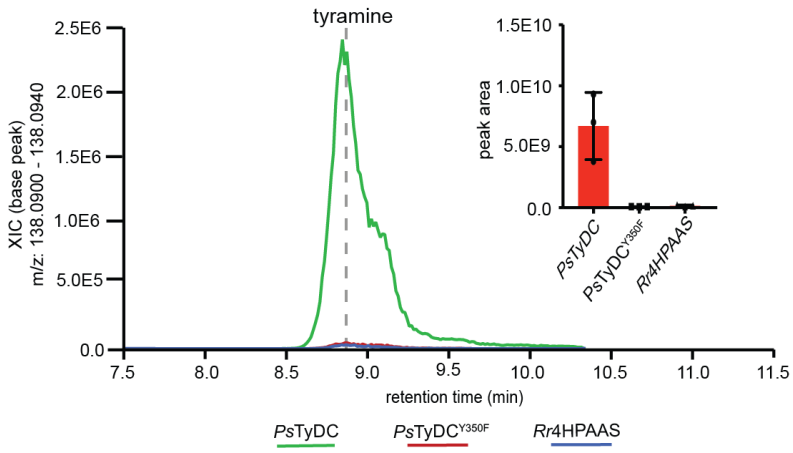

Fig. S19. Relative *in vivo* L-tyrosine decarboxylase activity of various AAAD proteins in transgenic yeast. Yeast transformed with the multi-gene vector containing the wild-type *PsTyDC* produces tyramine rather than (S)-norcoclaurine. The multi-gene vectors used to transform yeast contain the requisite *PpDDC*, *PsNCS* and *BvTyH* in addition to either *PsTyDC*, *PsTyDC*^{Y350F} or *Rr4HPAAS*. Error bars indicate standard error of the mean (SEM) based on biological triplicates.

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

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 *Cr*TDC, *Ps*TyDC, *At*PAAS and *Rr*4HPAAS corresponds to residues 342-361, 344-363, 332-351 and 337-352, respectively. The small loop is represented in the *Cr*TDC, *Ps*TyDC, *At*PAAS and *Rr*4HPAAS sequences by residues 200-205, 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 2

The lysine-conjugated coenzyme PLP is simulated in either the enolimine (systems 1-2) or the ketoenamine form (systems 3-6) shown in Fig. S11. In aqueous solution, PLP aldimine is known to undergo reversible proton tautomerism between these two forms (8–10). Although when conjugated with an enzyme, the ketoenamine form is believed to dominate (8, 9), 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 Asp^{287}. While PROPKA calculation supports a protonated Asp^{287}, a deprotonated Asp^{287} is known to stabilize LLP with its pyridine nitrogen protonated (11–13). Therefore, while we modeled the enolimine form of LLP with Asp^{287} 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. S11). 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. S14*A*. The cartesian space visited by loop residues as enclosed by the occupancy isosurfaces (Fig. S14*A*) as well as the average helical content of the large loop measured by the program DSSP (Fig. S14*B*) are also similar across all six systems. These results suggest that in its initial 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. S12. While interactions with LLP and L-tryptophan are certainly 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 3

The model of the closed-state *Cr*TDC was constructed by superimposing the open *Cr*TDC structure onto the closed conformation of *Ps*TyDC and subsequently threading the *Cr*TDC loops on the *Ps*TyDC structure. The differences between the MD results and the open *Cr*TDC model were 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	*At*PAAS	*Rr*4HPAAS
Wavelength (Å)	0.9793	0.9793	0.9793	0.9793
Resolution range (Å)	69.04 - 2.61 (2.70 - 2.61)	133.2 - 2.05 (2.12 - 2.05)	78.54 - 2.1 (2.18 - 2.1)	59.18 - 2.6 (2.69 - 2.6)
Space group	P 4_1 2_1 2	P 1 2_1 1	P 2_1 2_1 2_1	P 4_3 2_1 2
Unit cell a, b, c (Å)	122.77, 122.77, 167.00	107.61, 69.61, 133.47)	78.06, 106.79, 115.90	118.36, 118.36, 67.04
α, β, γ (°)	90, 90, 90	90, 93.778, 90	90, 90, 90	90, 90, 90
Total reflections	332655 (33114)	851987 (82988)	623771 (61394)	387791 (39193)
Unique reflections	39480 (3848)	123215 (9909)	57211 (5641)	15164 (1489)
Multiplicity	8.4 (8.6)	6.9 (6.8)	10.9 (10.9)	25.6 (26.3)
Completeness (%)	99.95 (99.74)	89.01 (80.54)	97.11 (99.47)	99.95 (99.66)
Mean I/sigma(I)	9.16 (1.59)	10.79 (0.58)	15.14 (2.26)	29.05 (2.97)
Wilson B-factor	34.24	42.66	41.06	60.13
R-merge	0.3504 (1.789)	0.1475 (2.547)	0.5252 (1.322)	0.3799 (1.006)
R-meas	0.373 (1.902)	0.1593 (2.763)	0.5515 (1.388)	0.3884 (1.025)
R-pim	0.1267 (0.644)	0.05974 (1.06)	0.1665 (0.4192)	0.07923 (0.1962)
CC1/2	0.978 (0.558)	0.996 (0.362)	0.93 (0.909)	0.906 (0.909)
CC*	0.995 (0.846)	0.999 (0.729)	0.982 (0.976)	0.975 (0.976)
Reflections used in refinement	39464 (3847)	110319 (9907)	55587 (5620)	15158 (1484)
Reflections used for R-free	1979 (175)	1991 (175)	1994 (207)	1518 (149)
R-work	0.1872 (0.2909)	0.2201 (0.4525)	0.2011 (0.2932)	0.2108 (0.3298)
R-free	0.2345 (0.3360)	0.2557 (0.4525)	0.2022 (0.3093)	0.2686 (0.4258)
CC(work)	0.954 (0.769)	0.960 (0.635)	0.916 (0.891)	0.770 (0.637)
CC(free)	0.926 (0.680)	0.964 (0.627)	0.934 (0.882)	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	GCATAGCAATCTAATCTAAGTTTTCTAGAACTAGTATGGGC

		AGCATTGATTCAACAAATGTAGC
*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. M. A. Larkin, *et al.*, Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).

2. P. Gouet, X. Robert, E. Courcelle, ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.* **31**, 3320–3323 (2003).

3. A. Stivala, M. Wybrow, A. Wirth, J. C. Whisstock, P. J. Stuckey, Automatic generation of protein structure cartoons with Pro-origami. *Bioinformatics* **27**, 3315–3316 (2011).

4. H. C. Dunathan, Conformation and reaction specificity in pyridoxal phosphate enzymes. *Proceedings of the National Academy of Sciences* **55**, 712–716 (1966).

5. J. Liang, Q. Han, H. Ding, J. Li, 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).

6. P. J. Facchini, V. De Luca, Differential and tissue-specific expression of a gene family for tyrosine/dopa decarboxylase in opium poppy. *J. Biol. Chem.* **269**, 26684–26690 (1994).

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

8. M. Chan-Huot, *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).

9. B. G. Caulkins, *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).

10. S. Sharif, *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).

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

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

13. D. T. Major, J. Gao, 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).