Structure and evolution of Alanine/Serine Decarboxylases and the engineering of theanine production

Hao Wang[a,1], Biying Zhu[b,1], Siming Qiao[b], Chunxia Dong[b], Xiaochun Wan[b], Weimin Gong[a,*], Zhaoliang Zhang[b,*]

[a]Department of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China.

[b]State Key Laboratory of Tea Plant Biology and Utilization, Anhui Agricultural University, Hefei, Anhui 230036, China.

[1]These authors contributed equally to this work.

* To whom correspondence should be addressed. Email: zhlzhang@ahau.edu.cn or wgong@ustc.edu.cn.

Author contributions: W.G, Z.Z and X.W designed research; H.W, B.Z, S.Q and C.D performed research; H.W and B.Z analyzed data; H.W, B.Z, W.G, and Z.Z wrote the paper.

Competing interest statement: The authors declare no competing financial interests.

Classification: Biochemistry (major), Plant Sciences (minor)

Keywords: alanine decarboxylase, serine decarboxylase, substrate specificity, theanine, *Camellia sinensis*

Abstract

Ethylamine (EA), the precursor of theanine biosynthesis, is synthesized from alanine decarboxylation by alanine decarboxylase (AlaDC) in tea plants. AlaDC evolves from serine decarboxylase (SerDC) through neofunctionalization and has lower catalytic activity. However, lacking structure information hinders the understanding of the evolution of substrate specificity and catalytic activity. In this study, we solved the X-ray crystal structures of AlaDC from *Camellia sinensis* (CsAlaDC) and SerDC from *Arabidopsis thaliana* (AtSerDC). Tyr^{341} of AtSerDC or the corresponding Tyr^{336} of CsAlaDC is essential for their enzymatic activity. Tyr^{111} of AtSerDC and the corresponding Phe^{106} of CsAlaDC determine their substrate specificity. Both CsAlaDC and AtSerDC have a distinctive zinc finger and have not been identified in any other Group II PLP-dependent amino acid decarboxylases. Based on the structural comparisons, we conducted mutation screen of CsAlaDC. The results indicated that the mutation of L110F or P114A in the CsAlaDC dimerization interface significantly improved the catalytic activity by 110% and 59%, respectively. Combining a double mutant of $CsAlaDC^{L110F/P114A}$ with theanine synthetase increased theanine production 672% in an *in vitro* system. This study provides the structural basis for the substrate selectivity and catalytic activity of CsAlaDC and AtSerDC and provides a route to more efficient biosynthesis of theanine.

Introduction

Tea is one of the most popular natural non-alcoholic beverages consumed worldwide. About 2 billion cups of tea are consumed worldwide daily (1). Theanine (γ-glutamylethylamide) is a unique non-protein amino acid and the most abundant free amino acid in tea plants (*Camellia sinensis*). It accounts for more than 50% of the total free amino acids and approximately 1-2% of the dry weight of the new shoots of tea plants (2). Theanine is the secondary metabolite conferring the umami taste of tea infusion and also balances the astringency and bitterness of tea infusion caused by catechins and caffeine (3). It has also many health-promoting functions, including neuroprotective effects, enhancement of immune functions, and potential anti-obesity capabilities, among others (2, 4-9). Therefore, theanine content is highly correlated with green tea quality (10).

Theanine is synthesized from ethylamine (EA) and glutamate (Glu) by theanine synthetase (11, 12). Importantly, the large amount of theanine biosynthesis is determined by the high availability of EA in tea plants (13). Therefore, the evolution of EA biosynthesis in tea plants provided the basis for theanine biosynthesis and the formation of tea quality. EA is synthesized from alanine decarboxylation by alanine decarboxylase (CsAlaDC) (14) (Figure 1A). Indeed, CsAlaDC expression level and catalytic activity largely determine the theanine accumulation level in tea plants (15). As a novel gene, *CsAlaDC* had not been reported in other plant species before it was identified in tea plants. Previous studies indicated that *AlaDC* originated from the serine decarboxylase gene (*SerDC*) by gene duplication in tea plants (16). However, despite

61 sharing highly conserved amino acid sequences, CsAlaDC and CsSerDC specifically
62 catalyze the decarboxylation of alanine and serine, respectively (16) (Figure 1A and B).
63 In addition, CsAlaDC exhibited a much lower enzymatic activity compared with
64 CsSerDC (16). However, the structural basis and key amino acids underlying the
65 evolution of the substrate specificity and enzymatic activity of CsAlaDC are unknown.
66 SerDC belongs to the Group II pyridoxal-5-phosphate (PLP)-dependent
67 decarboxylase superfamily (17). These Group II amino acid decarboxylases produce
68 many bioactive secondary metabolites and signaling molecules (18-20). In plants,
69 SerDC catalyzes the biosynthesis of ethanolamine from serine, which was first
70 functionally characterized in *Arabidopsis thaliana* (21) (Figure 1A). Ethanolamine is
71 an important metabolite in the synthesis of phosphatidylethanolamine and
72 phosphatidylcholine, two main phospholipids involved in maintaining eukaryotic
73 membrane structures in eukaryotic cell membranes (22-24). Furthermore, the study has
74 shown that SerDC is essential in the embryogenesis of *Arabidopsis* (25). SerDC is of
75 vital importance for plant growth and development, but the mechanism of substrate
76 recognition and catalytic activity of SerDC also remains unclear.
77 In this work, we attempt to understand the mechanism of functional divergence of
78 plant AlaDC and SerDC by analyzing their structural characteristics. Here, we obtained
79 the X-ray crystal structure of CsAlaDC and AtSerDC. According to the crystal structure,
80 we found a distinctive zinc finger structure located at both CsAlaDC and AtSerDC,
81 which has not been identified in any other Group II PLP-dependent amino acid
82 decarboxylases that have been previously characterized. By comparing the substrate

83 binding pockets, we identified Phe^{106} of CsAlaDC and Tyr^{111} of AtSerDC as the crucial
84 sites for substrate specificity. By conducting mutation screening based on the protein
85 structures, we identified the amino acids repressing the catalytic activity and discovered
86 that $CsAlaDC^{L110F/P114A}$ exhibited a 2.3-fold increase in catalytic activity compared to
87 that of CsAlaDC and enhanced the engineering of theanine production *in vitro*.

88 **Results**

89 **Enzymatic properties of CsAlaDC, AtSerDC, and CsSerDC**

90 CsAlaDC originates from CsSerDC by gene duplication and neofunctionalization in tea
91 plants, but they catalyze different metabolic processes (Figure 1A). We performed a
92 multiple sequence alignment of the amino acid sequences of the 6 SerDCs from
93 kiwifruit (*Actinidia chinensis*), grape (*Vitis vinifera*), coffee (*Coffea eugenioides*),
94 cocoa (*Theobroma cacao*), Arabidopsis and tea plants and CsAlaDC (Figure 1B). The
95 results showed that the amino acid sequences of CsAlaDC and SerDCs were highly
96 conserved, but CsAlaDC has amino acid mutations in highly conserved regions
97 compared with SerDCs.

98 Next, to verify the substrate specificity and enzyme activity, we conducted enzyme
99 activity detection and enzyme kinetics analysis for CsAlaDC, CsSerDC and AtSerDC.
100 The 5′ truncated *CsAlaDC* and *AtSerDC* were inserted into the pET22b and pET28a
101 expression vectors to generate recombinant plasmids pET22b-*CsAlaDC* and pET28a-
102 *AtSerDC*, correspondingly. Additionally, the full-length protein of CsSerDC was
103 inserted into the pET28a expression vector to generate the recombinant plasmid
104 pET28a-*CsSerDC*. Subsequently, the recombinant proteins were expressed in

105 *Escherichia coli* and purified via nickel affinity chromatography. The purified proteins
106 were examined through SDS-PAGE analysis (Figure 2A), supplemented by size-
107 exclusion chromatography assessments (Supplementary figure 1A), it seems that the
108 protein manifests in an oligomeric configuration within the solution.

109 To verify the substrate specificity of CsAlaDC, AtSerDC, and CsSerDC, enzyme
110 activity assays were conducted using Ala and Ser as substrates, followed by product
111 identification via UPLC analysis. The results showed that CsAlaDC can effectively
112 catalyze the decarboxylation of alanine to generate EA, whereas its catalytic efficacy
113 towards serine was relatively inferior. Conversely, AtSerDC and CsSerDC selectively
114 catalyzed serine decarboxylation to yield ethanolamine but did not exhibit activity
115 towards alanine (Figure 2B). These findings confirmed that CsAlaDC and SerDCs do
116 not have promiscuous decarboxylase activity on Ala and Ser.

117 The kinetic properties of CsAlaDC, AtSerDC, and CsSerDC were determined
118 through the use of corresponding substrates (Figure 2C), and the results are presented
119 in Table 1. The kinetic parameter *Km* of CsAlaDC was determined to be 1.215 mM,
120 which is similar to that of AtSerDC and CsSerDC at 1.522 mM and 2.364 mM,
121 respectively. However, AtSerDC has a *Vmax* of 7.053 μmo L^{-1} s^{-1}, which is 4-fold that
122 of CsAlaDC's *Vmax* of 1.709 μmo L^{-1} s^{-1}; while CsSerDC has a *Vmax* of 9.031 μmo L^{-}
123 ^{1} s^{-1}, which is 5.3-fold that of CsAlaDC. The findings suggest that the enzymatic
124 catalytic efficiency of CsSerDC and AtSerDC is approximately triple compared to that
125 of CsAlaDC, and the *Vmax* of CsAlaDC is considerably lower than that of both
126 CsSerDC and AtSerDC. The observed disparity in *Vmax* suggests that CsAlaDC and

127 SerDCs may exhibit discrepancies in substrate-to-product conversion rates under
128 saturated substrate conditions. It is plausible that SerDCs possess a higher turnover rate,
129 facilitating the rapid conversion of substrate to product and subsequent release of the
130 enzyme. Another possibility is that SerDCs demonstrate enhanced stability, thereby
131 maintaining superior catalytic activity even at higher substrate concentrations,
132 consequently leading to a higher *Vmax* value. Due to the similarity in catalytic activity
133 between CsSerDC and AtSerDC, we chose the more representative AtSerDC for further
134 analysis.

135 **The overall structures of CsAlaDC and AtSerDC**

136 To enhance our comprehension of CsAlaDC and AtSerDC, we conducted structural
137 analyses of these two proteins. Through optimization of crystallization conditions, we
138 successfully determined the crystal structures of CsAlaDC, CsAlaDC-EA, and
139 AtSerDC at resolutions of 2.50 Å, 2.60 Å and 2.85 Å, respectively (Supplementary
140 table 1).

141 The overall structure of CsAlaDC or AtSerDC is homodimeric with two subunits
142 exhibiting an asymmetrical arrangement (Figure 3A and 3D). The monomer of
143 CsAlaDC or AtSerDC is divided into three distinct structural domains: an N-terminal
144 domain (N-terminal -104 aa in CsAlaDC, N-terminal -109 aa in AtSerDC), a large
145 domain (105-364 aa in CsAlaDC, 110-369 aa in AtSerDC), and a small C-terminal
146 domain (365 aa-C-terminus in CsAlaDC, 370-C-terminus as in AtSerDC), colored light
147 pink, khaki, and sky blue, respectively. Compared to the full-length protein, the N-
148 terminal structures of CsAlaDC and AtSerDC were truncated by 60 and 65 amino acid

149 residues, respectively. The N-terminal domain of the truncated protein contains a long
150 α-helix, which is connected to the large domain through a long loop. The helix of one
151 subunit is antiparallel to the corresponding helix in the neighboring subunit, forming a
152 clamp. Therefore, the N-terminal domain may not be an independent folding unit and
153 is likely only stable in dimers. The large domain contains a seven-stranded mixed β-
154 sheet surrounded by eight α-helices, while the C-terminal domain is composed of a
155 three-stranded antiparallel β-sheet and three α-helices (Supplementary figure 1). The
156 catalytic site of the enzyme is positioned within a superficial crevice at the junction of
157 two subunits forming a dimeric arrangement. The amino acid residues derived from
158 both subunits participate in the binding of the PLP, effectively anchoring it to the active
159 site. In the catalytic site, one monomer accommodates PLP, while the other monomer
160 also contributes to PLP binding and enzymatic function (Figure 3A and 3D).

161 Notably, our investigation has revealed the presence of a distinctive zinc finger
162 structure (as depicted in Figure 3A and 3D) located at the C-terminus of CsAlaDC and
163 AtSerDC. This structure is composed of a loop structure spanning 17 amino acid
164 residues, wherein coordination of Zn^{2+} is facilitated by three Cys residues and one His
165 residue. Importantly, this particular configuration is exclusive to the two proteins under
166 examination and has not been identified in any other Group II PLP-dependent amino
167 acid decarboxylases that have been previously characterized (26, 27).

168 The crystal structures of CsAlaDC-EA complex and AtSerDC were obtained and
169 further analyzed. In the former complex, PLP bound to Lys^{309} via a Schiff base linkage
170 (internal aldimine form), and the pyridine moiety of PLP is positioned between the

171 imidazole ring of His^{196} and the methyl group of Ala^{279} in a parallel orientation to the
172 imidazole ring. The carboxylic group of Asp^{277} stabilizes the N1 of PLP by a salt bridge
173 interaction, providing the latter with the strong electrophilicity necessary to stabilize
174 carbon ion intermediates during enzymatic catalysis. The O3 forms hydrogen bonds
175 with Thr^{247} and Lys^{309}. Additionally, His^{308}, Gly^{169}, Thr^{170}, Lys^{309}, and Ser^{347}*("*"
176 denotes the amino acids on adjacent subunits) establish a hydrogen bonding network
177 with the phosphate group of PLP, consequently fortifying its attachment to the protein
178 (Figure 3C). Similarly, in the latter structure, the pyridine ring of PLP was sandwiched
179 between the imidazole ring of His^{201} and the methyl group of Ala^{284}. The N1 of PLP
180 formed a salt bridge with Asp^{282}, while the O3 formed hydrogen bonds with Thr^{252}. The
181 phosphate group of PLP established hydrogen bonding interactions with Lys^{314}, His^{313},
182 Gly^{174}, Thr^{175}, and Ser^{352}* (Figure 3E). The aforementioned amino acid residues are
183 observed across various Group II PLP-dependent amino acid decarboxylases and
184 exhibit considerable stability (see further discussion below), suggesting that they could
185 share common features in terms of catalytic mechanisms.

186 **Factors affecting CsAlaDC and AtSerDC activity**

187 The Group II PLP-dependent amino acid decarboxylases are dimeric enzymes with
188 each monomer containing an active site located within a shallow cavity (28). A flexible
189 loop of amino acids originating from one monomer extends into the active site of the
190 other monomer and plays a crucial role in catalysis. The catalytic loop in AtSerDC,
191 consisting of amino acid residues 328-341, is well-structured and exhibits clear electron
192 density. However, the corresponding loop in CsAlaDC is disordered (Figure 3F). The

193 aforementioned loop harbors a conserved Tyr residue that is believed to function as a
194 proton donor of the carbanion in the catalytic process (29). In particular, CsAlaDC
195 Tyr^{336} and AtSerDC Tyr^{341} correspond to that Tyr residue (Supplementary figure 2).
196 Upon superimposing the active site residues of CsAlaDC apo and CsAlaDC-EA
197 complex, it was observed that the hydroxyl group of Tyr^{336}* in the CsAlaDC-EA
198 complex exhibits a substantial 60-degree deviation along the Cα-Cβ bond (Figure 3G).
199 To assess the function of the Tyr, we substituted the corresponding Tyr residues in
200 CsAlaDC and AtSerDC with Phe. The resulting $CsAlaDC^{Y336F}$ and $AtSerDC^{Y341F}$
201 mutants were then exposed to Ala and Ser, respectively. We measured the production
202 of EA or ethanolamine in the reaction mixtures. Our findings indicate that these mutants
203 catalyze abortive decarboxylation, as evidenced by the absence of detectable EA and
204 only a small amount of ethanolamine observed in the reaction mixture (Figure 3H). This
205 result suggested this Tyr is required for the catalytic activity of CsAlaDC and AtSerDC.

206 **Identification of key amino acids for the substrate specificity**

207 The superposition of amino acid residues in the substrate binding pocket of CsAlaDC
208 and AtSerDC revealed that all residues are identical, except for the substitution of Tyr
209 at position 111 in AtSerDC with Phe at position 106 in CsAlaDC (Figure 4A). This
210 observation suggests a potential role for this specific residue in determining the
211 selective binding of the appropriate substrate.

212 To gain further insights into the functional role of this residue (Tyr^{111} in AtSerDC
213 or Phe^{106} in CsAlaDC) in selective binding of the appropriate substrate, we identified
214 563 potential serine decarboxylases in Embryophyta using the amino acid sequences of

215 AtSerDC (Supplementary figure 3A). By comparing the amino acid sequences of these
216 serine decarboxylases, we observed that the corresponding residues to Tyr^{111} or Phe^{106}
217 are located within a conserved motif comprising 9 amino acids (Figure 4B). Within this
218 motif, the first two residues, Y (Tyrosine) and P (Proline) are completely conserved
219 across all 563 proteins. However, the third residue where Tyr^{111} in AtSerDC or Phe^{106}
220 in CsAlaDC is positioned exhibits variability among Y, T (Threonine), F
221 (Phenylalanine), V (Valine), A (Alanine), I (Isoleucine), L (Leucine), or other amino
222 acids (Figure 4B). Remarkably similar to Tyr^{111}, this residue is predominantly Y in 83.7%
223 of these homologs; conversely, F was found to be present in only 2.3% of these
224 homologs within plant species belonging to Poales, Asterales, Ranunculales, Ericales,
225 Caryophyllales and Nymphaeales orders (Figure 4C-4D).

226 To verify the role of Tyr^{111} in AtSerDC or Phe^{106} in CsAlaDC in the selective
227 binding of the appropriate substrate, we introduced mutations in Tyr^{111} to Phe
228 ($AtSerDC^{Y111F}$) and Phe^{106} to Tyr ($CsAlaDC^{F106Y}$), followed by *in vitro* enzymatic
229 activity assays. The results revealed that $AtSerDC^{Y111F}$ acquired alanine decarboxylase
230 activity, whereas $CsAlaDC^{F106Y}$ completely lost its alanine decarboxylase activity
231 (Figure 4E). Additionally, we generated other mutations of Tyr^{111} in AtSerDC, including
232 $AtSerDC^{Y111A}$, $AtSerDC^{Y111I}$, $AtSerDC^{Y111L}$, $AtSerDC^{Y111V}$ and $AtSerDC^{Y111W}$. Among
233 these mutants, only $AtSerDC^{Y111I}$ exhibited a marginal level of alanine decarboxylase
234 activity (Figure 4E). These findings indicate the essential role of Phe^{106} in the selective
235 binding of alanine for CsAlaDC.

236 On the other side, both CsAlaDC and $CsAlaDC^{F106Y}$ exhibit a very low level of

237 serine decarboxylase activity (Figure 4F). Moreover, compared with AtSerDC, the
238 serine decarboxylase activity of AtSerDC^{Y111F} was about 30% of AtSerDC;
239 AtSerDC^{Y111I} retained lower than 5% of serine decarboxylase activity of AtSerDC;
240 while other mutations, including AtSerDC^{Y111A}, AtSerDC^{Y111L}, AtSerDC^{Y111V} and
241 AtSerDC^{Y111W} abolished the serine decarboxylase activity (Figure 4F). These results
242 suggested the Tyr^{111} of AtSerDC is also important for the selective binding of serine for
243 AtSerDC.

244 To further verify that Phe^{106} of CsAlaDC and Tyr^{111} of AtSerDC were key amino
245 acid residues determining the selective binding of substrates *in planta*, we employed
246 the *Nicotiana benthamiana* transient expression system. To this end, *A. tumefaciens*
247 strain GV3101 (pSoup-p19), carrying recombinant plasmid, was infiltrated into leaves
248 of 5-week-old *N. benthamiana* plants, and the pCAMBIA1305 empty vector was used
249 as the control (EV). Relative mRNA levels were detected (Supplementary figure 4),
250 indicating that they had been successfully overexpressed in tobacco. Here, we found
251 that a high level of EA was produced in the CsAlaDC-expressing tobacco leaves; while
252 no EA product was detected in tobacco leaves infiltrated with the mutant CsAlaDC^{F106Y}.
253 In addition, we did not detect EA products in the AtSerDC-expressing tobacco leaves,
254 while a high level of EA was detected in tobacco leaves infiltrated with mutant
255 AtSerDC^{Y111F}. As anticipated, EA was not detected in tobacco leaves infiltrated with
256 EV (Figure 4G and H). These results further verified the critical role of Phe^{106} in the

257 selective binding of alanine for CsAlaDC.

258 **Key amino acids for the evolution of CsAlaDC enzymatic activity**

259 CsAlaDC and AtSerDC have a high sequence similarity of 74.5% and a nearly identical
260 structure with an RMSD (root mean square deviation) of only 0.77Å for monomer
261 structures. However, the two enzymes catalyze different amino acid decarboxylation
262 reactions, and AtSerDC exhibits significantly higher *Vmax* than CsAlaDC (Figure 2B
263 and C). In light of this observation, we postulated a hypothesis: EA, generated via Ala
264 decarboxylation by CsAlaDC, can be toxic and harmful to plants if accumulated
265 excessively. Thus, during the evolution of plant serine decarboxylase into alanine
266 decarboxylase, the enzyme has evolved not just to alter the substrate preference but also
267 to reduce catalytic activity to control EA production within a suitable range. Based on
268 this hypothesis, we suggest that mutating specific amino acids in CsAlaDC to those
269 corresponding amino acids in SerDCs could enhance its activity, and the results could
270 provide insights into the evolution of CsAlaDC enzymatic activity.

271 Dimerization is essential to AlaDC/SerDC activities because the active site is
272 composed of residues from two monomers. To identify the amino acids repressing the
273 enzymatic activity of CsAlaDC during evolution from SerDC, we analyzed the crystal
274 structures and the amino acids at the dimer interface between CsAlaDC and AtSerDC.
275 This analysis revealed that the amino acids at the dimer interface at positions 66, 97,
276 110, 114, 116, 117, 122, 315 and 345 are different in CsAlaDC and SerDCs (Figure 1B).
277 Therefore, we mutated these amino acids of CsAlaDC into those in the corresponding
278 positions of AtSerDC and CsSerDC, and performed enzyme activity assays (Figure 5A).

279 The results demonstrated that CsAlaDC^{L110F} and CsAlaDC^{P114A} exhibited significantly
280 enhanced enzyme activity compared to the wild-type CsAlaDC, with a 2.1-fold and
281 1.59-fold increase, respectively. Furthermore, the catalytic activity of the
282 CsAlaDC$^{L110F/P114A}$ double mutant exhibits a remarkable 2.3-fold increase compared to
283 that of the wild-type protein (Figure 5B). These findings suggested a critical role of
284 Leu^{110} and Pro^{114} of CsAlaDC in the evolution of enzymatic activity and provided a
285 basis to improve CsAlaDC activity. It is possible that these amino acid residues could
286 potentially augment the hydrophobic nature of the protein dimer interface.

287 ***In Vitro* synthesis of theanine**

288 The biosynthetic pathway of theanine in tea plants comprises two consecutive
289 enzymatic steps: alanine decarboxylase facilitates the decarboxylation of alanine to
290 generate EA, while theanine synthetase catalyzes the condensation reaction between
291 EA and Glu to synthesize theanine. By conducting mutation screens, we discovered a
292 CsAlaDC mutant protein (L110F/P114A) that exhibited a 2.3-fold higher catalytic
293 activity compared to the wild-type protein. Subsequently, we employed an *in vitro*
294 theanine synthesis system utilizing CsAlaDC and either glutamine synthetase (PsGS
295 [P*seudomonas syringae* pv. syringae]) or gamma-glutamate methylamine ligase
296 (MmGMAS [*Methylovorus mays*]) which have the ability to synthesize theanine from
297 EA and Glu (30, 31).

298 The results illustrated the successful synthesis of theanine using CsAlaDC in
299 conjunction with two theanine synthetases, employing Ala and Glu as substrates (Figure
300 5C). The theanine content generated by the combination of CsAlaDC^{L110F} and

301 MmGMAS, as well as CsAlaDC$^{L110F/P114A}$ and MmGMAS, was 4.57-fold and 6.72-fold
302 higher than the content produced in the combination of wild-type CsAlaDC and
303 MmGMAS, respectively. Similarly, when combined with the PsGS, comparable
304 outcomes are observed as well. The theanine content resulting from the combination of
305 CsAlaDC^{L110F} and PsGS, as well as CsAlaDC$^{L110F/P114A}$ and PsGS, exhibit
306 enhancements of 1.62-fold and 4.33-fold compared to the wild-type protein
307 combination, respectively (Figure 5C). Hence, the utilization of CsAlaDC$^{L110F/P114A}$
308 could effectively enhance the theanine production yield and thus, holds potential for
309 large-scale engineering production of theanine.

310 **Discussion**

311 **The common and distinctive structural characteristics of CsAlaDC and AtSerDC**
312 **compared with other amino acid decarboxylases**

313 CsAlaDC and AtSerDC are Group II PLP-dependent amino acid decarboxylases, which
314 exhibit numerous structural features common to other amino acid decarboxylases. The
315 Dali search (32) revealed that AtSerDC and CsAlaDC share structural similarities with
316 7CIG (MetDC [*Streptomyces sp*. 590]), 7ERV (HisDC1 [*Photobacterium*
317 *phosphoreum*]), 6KHO (TrpDC [*Oryza sativa Japonica* Group]), 4E1O (HisDC2
318 [*Homo sapiens*]), 6JY1 (AspDC [*Methanocaldococcus jannaschii*]), and 5GP4 (GluDC
319 [*Levilactobacillus brevis*]). The monomers of these amino acid decarboxylases all
320 consist of three characteristic domains: the C-terminal domain, the large domain and
321 the N-terminal domain (Supplementary figure 5). The amino acid residues that bind to
322 PLP cofactors are conserved across multiple enzymes (Supplementary figure 5 and 6).

323 Out of the eight proteins analyzed, a total of thirteen amino acid residues were found to
324 be conserved across all sequences (Glu^{143}, Glu^{171}, Lys^{201}, Gly^{245}, Thr^{247}, Asp^{253}, His^{275},
325 Asp^{277}, Ala^{279}, Pro^{286}, Ser^{302}, Lys^{309}, and Tyr^{336} in CsAlaDC). Some of them are situated
326 within the active center and play a role in stabilizing PLP or facilitating catalytic
327 reactions. Conversely, other residues reside outside the active center and their function
328 remains unclear (Supplementary figure 6).

329 Group II PLP-dependent amino acid decarboxylases are characterized by the
330 existence of a highly flexible loop, which is of great significance for the catalytic
331 mechanism of decarboxylase (33). The loops are located at the dimer interface and
332 extend to the active sites of other monomers in a closed conformation. Prior research
333 has established that the conserved amino acid residue Tyr within the loop plays a crucial
334 role in catalysis by donating protons to carbanions of quinone intermediates that arise
335 following decarboxylation (34, 35). Our experiments demonstrate that the substitution
336 of the corresponding Tyr with Phe in the loop renders the protein inactive. Through the
337 application of circular dichroism, we have corroborated that the stability of the protein
338 is not compromised by mutations (Supplementary figure 7A, B, D, E). Both the mutant
339 and wild-type proteins manifest absorption bands at the 420 nm wavelength, signifying
340 the formation of a Schiff base between PLP and the lysine residues found at the active
341 site. Given that PLP has been incorporated during the protein purification stage, the
342 absorbance observed for both the mutant and wild-type proteins at 420 nm is consistent
343 when compared at equivalent concentrations (Supplementary figure 8). This
344 corroboration implies that the protein mutation does not interfere with the binding to

345 PLP. Consequently, the inactivation observed in the mutant proteins is not a
346 consequence of instability incited by corresponding mutations, nor changes in the
347 binding affinity between the protein and PLP induced by these mutations.

348 We observed that the substitution of Phe^{106} for Tyr in CsAlaDC rendered CsAlaDC
349 inactive, while the substitution of Tyr^{111} for Phe in AtSerDC enabled alanine
350 decarboxylase activity, and the alteration in the activity of the mutant protein bears no
351 correlation to protein stability and affinity for PLP (Supplementary figure 7A, C, D, F
352 and Supplementary figure 8), suggesting that the amino acid in the position is critical
353 for determining substrate specificity. Integrative crystallographic structure of the two
354 enzymes, we speculate the hydrophilic nature of the Lys residue in AtSerDC
355 predisposes the active site to preferentially accommodate the hydrophilic amino acid
356 Ser as its substrate. In contrast, the equivalent position in CsAlaDC is occupied by Phe,
357 an amino acid lacking the hydroxyl group. This substitution enhances the hydrophobic
358 nature of the substrate-binding pocket. Consequently, CsAlaDC demonstrates a unique
359 predilection, selectively binding Ala (an amino acid with comparatively hydrophobic
360 properties) as its preferred substrate.

361 The monomeric configuration of CsAlaDC and AtSerDC is akin to that of other
362 Group II PLP-dependent amino acid decarboxylases, except for a conspicuous
363 dissimilarity. Unlike other Group II PLP-dependent amino acid decarboxylases, the C-
364 terminal of CsAlaDC and AtSerDC harbors an evident zinc finger structure. We propose
365 that this structure could potentially influence enzyme stability. To test this hypothesis,
366 we truncated the zinc finger structure from both proteins and expressed them. Our

367 findings indicate that, following the excision of the zinc finger structure, CsAlaDC
368 became insoluble, while AtSerDC displayed a similar trait to a certain extent
369 (Supplementary figure 9). Hence, we can establish that the zinc finger structure indeed
370 exerts an influence on the enzyme's stability. The potential for additional functions
371 necessitates further investigation.

372 **Evolution of AlaDC and SerDC**

373 The structure of CsAlaDC seems like with AtSerDC, however, there are differences in
374 substrate specificity. To clarify the relationship between CsAlaDC and other SerDCs,
375 we analyzed serine decarboxylase-like proteins in Embryophyta that are highly
376 homologous to CsAlaDC and constructed phylogenetic trees to gain insight into their
377 evolutionary relationships (Supplementary figure 3). Based on our experimental
378 findings and evolutionary evidence, we have identified a conserved YPX motif at the
379 substrate binding pocket. For most plants, the residue X in this motif is predominantly
380 Tyr, which corresponds to the serine decarboxylase protein. However, other residues
381 such as Thr, Phe, Val, Ala, and Ile can also occupy this position (Figure 4B and C).
382 Interestingly, serine decarboxylase-like proteins containing YPF motifs, which have the
383 potential to function as alanine decarboxylases, were found to be distributed throughout
384 the phylogenetic tree, including Asteroideae, Ericales, Chenopodiaceae, Poaceae, and
385 Ranunculales (Supplementary figure 3B). However, these proteins were absent in some
386 more recent species. This observation implies that the emergence of alanine
387 decarboxylase in these particular species could be attributed to convergent evolution.

388 Moreover, we have observed the enrichment of serine decarboxylase-like proteins

389 that possess a YPT motif within the Fabales (Supplementary figure 3B). Prior research
390 has demonstrated that XP_004496485-1, which possesses a YPT motif, displays
391 catalytic efficacy toward the processes of decarboxylation and oxidative deamination
392 of Phe, Met, Leu, and Trp (36). Given that the YPT motif is highly conserved and
393 widely distributed in Fabales, serine decarboxylase-like proteins bearing the YPT motif
394 may have developed a unique substrate specificity in Fabales, beyond their
395 conventional decarboxylation functions, as exemplified by XP_004496485-1 protein
396 mentioned above. We speculate that these proteins may be capable of catalyzing other
397 reactions, potentially involving non-protein amino acids or other substances as
398 substrates.

399 **Applied to improve the synthesis of theanine**

400 Theanine is an important indicator of green tea quality. Therefore, improving the
401 synthesis of theanine in tea plants is the focus of research. In this study, through crystal
402 structure analysis and mutation verification, we found that the catalytic activity of
403 $CsAlaDC^{L110F/P114A}$ is 2.3 times higher than that of the wild-type protein, resulting in a
404 more abundant synthesis of theanine *in vitro*. This gives us great inspiration to improve
405 the synthesis of EA in tea plants by gene editing, thus increasing the content of theanine
406 in tea plants (Figure 5).

407 Theanine is highly demanded, by the market, due to its health effects and medicinal
408 value, and as a constituent in food, cosmetics and other fields. To meet the market
409 demand, a variety of methods have been used to acquire theanine, with the main
410 methods including direct extraction, chemical synthesis, biotransformation (microbial

411 fermentation) and plant cell culture. Based on the findings of this study, site-directed
412 mutagenesis can be employed to modify enzymes involved in theanine synthesis. This
413 modification enhances the capacity of bacteria, yeast, model plants, and other
414 organisms to synthesize theanine, thereby facilitating its application in industrial
415 theanine production.

416 **Conclusions**

417 In conclusion, our structural and functional analyses have significantly advanced
418 understanding of the substrate-specific activities of alanine and serine decarboxylases,
419 typified by CsAlaDC and AtSerDC. Critical amino acid residues responsible for
420 substrate selection were identified-Tyr111 in AtSerDC and Phe106 in CsAlaDC-
421 highlighting pivotal roles in enzyme specificity. The engineered CsAlaDC mutant
422 (L110F/P114A) not only displayed enhanced catalytic efficiency but also substantially
423 improved L-theanine yield in a synthetic biosynthesis setup with PsGS or MmGMAS.
424 Our research expanded the repertoire of potential alanine decarboxylases through the
425 discovery of 13 homologous enzyme candidates across embryophytic species and
426 uncovered a special motif present in serine protease-like proteins within Fabale,
427 suggesting a potential divergence in substrate specificity and catalytic functions. These
428 insights lay the groundwork for the development of industrial biocatalytic processes,
429 promising to elevate the production of L-theanine and supporting innovation within the
430 tea industry.

431 **Materials and Methods**

432 **Plants materials.**

433 Tobacco (*Nicotiana benthamiana*) plants were grown in a controlled chamber, under a
434 16-h light and 8-h dark photoperiod at 25 °C. Leaves of 5-week-old tobacco plants were
435 used for transient transformation, mediated by *Agrobacterium tumefaciens* strain
436 GV3101.

437 **Gene cloning and protein expression**

438 The coding sequences of AtSerDC and CsAlaDC, amino acid residues 66-482 and 61-
439 478, respectively, were amplified using PCR and then ligated into the *Nde* I and *Xho* I
440 restriction sites of the pET-28a and pET-22b vector, respectively, providing the
441 recombinant vector pET-28a-*AtSerDC* and pET-22b-*CsAlaDC*. The cDNAs encoding
442 CsSerDC were amplified using PCR and then ligated into the *Nde* I and *Xho* I restriction
443 sites of the pET-28a vector. Gene-specific primers were listed in Supplementary table
444 2. The recombinant plasmid was transformed into *E. coli* BL21 (DE3) competent cells.
445 Positive transformants were grown in a 5 mL LB medium containing 30 μg/mL
446 kanamycin or 50 μg/mL ampicillin at 37 °C overnight and then subcultured into an 800
447 mL LB medium containing the corresponding antibiotic. Protein expression was
448 induced by the addition of 0.2 mM isopropyl-β-d-thiogalactoside (IPTG) for 20 h at
449 16 °C when the optical density (OD) at 600 nm reached 0.6-0.8, harvested by
450 centrifugation at 4 °C and 4,000 rpm for 30 min.

451 **Protein production and crystallization**

452 The cell pellet was suspended in 30 mL of lysis buffer (20 mM HEPES, pH 7.5, 200
453 mM NaCl, 0.1 mM PLP), disrupted by High-Pressure Homogenizer, and then

454 centrifuged at 16,000 rpm for 30 min at 4 °C to remove the cell debris. The supernatants
455 were purified with a Ni-Agarose resin column followed by size-exclusion
456 chromatography. Before crystallization, purified proteins were concentrated at 10
457 mg/mL. Crystallization conditions were screened by the sitting-drop vapor diffusion
458 method using the reservoir solutions supplied in commercially available screening kits
459 (Crystal Screen, Crystal Screen 2, PEGRx 1, 2, and SaltRx 1, 2). A droplet made by
460 mixing 1.0 μL of purified AtSerDC or CsAlaDC (10 mg/mL) with an equal volume of
461 a reservoir solution was equilibrated against 100 μL of the reservoir solution at 16°C.
462 The crystal of AtSerDC was obtained using buffer pH 8.0 containing 20% (w/v)
463 PEG400 as a precipitate and 0.2 M $CaCl_2$. The crystal of the CsAlaDC-EA complex
464 was obtained at pH 7.5 containing 2.6 M sodium acetate. The crystal of CsAlaDC was
465 obtained at pH 6.0 containing 3.5 M sodium formate.

466 **Data collection and processing**

467 Crystals were grown by sitting-drop vapor diffusion method at 16 °C. The volume of
468 the reservoir solution was 100 μL and the drop volume was 2 μL, containing 1 μL of
469 protein sample and 1 μL of reservoir solution. The reservoir solution of AtSerDC
470 contained 0.1 M Tris-HCl (pH 8.0), 0.2 M $CaCl_2$, and 20% PEG 400. The reservoir
471 solution of CsAlaDC contained 0.1 M HEPES (pH 7.5), 2.6 M sodium acetate or 0.1 M
472 Bis-Tris (pH 6.0), and 3.5 M sodium formate. Crystals grew in 3-5 days using a protein
473 concentration of 10 mg/mL. Diffraction data were collected at Shanghai Synchrotron
474 Radiation Facility (China). The collected data sets were indexed, integrated, and scaled
475 using the HKL3000 software package. The structure of AtSerDC was solved by

476 molecular replacement using the structure of HisDC (PDB code: 7ERV
477 [*Photobacterium phosphoreum*]) as the model, and utilizing AtSerDC as a molecular
478 displacement model for both CsAlaDC and the CsAlaDC-EA complex. The AtSerDC,
479 CsAlaDC, and CsAlaDC-EA complex exhibit resolutions of 2.85 Å, 2.50 Å, and 2.60
480 Å, respectively. The statistics for data collection and processing are summarized in
481 Supplementary table 1.

482 **Enzyme activity assays**

483 Decarboxylase activity was measured by detecting products (EA or ethanolamine) in
484 Waters Acquit ultraperformance liquid chromatography (UPLC) system (16, 37). The
485 100 μL reaction mixture, containing 20 mM substrate (Ala or Ser), 100 mM potassium
486 phosphate, 0.1 mM PLP, and 0.025 mM purified enzyme, was prepared and incubated
487 at standard conditions (45 °C and pH 8.0 for CsAlaDC, 40 °C and pH 8.0 for AtSerDC
488 for 10 min). Then, the reaction was stopped with 20 μL of 10% trichloroacetic acid. The
489 product was derivatized with 6-aminoquinolyl-N-hydroxy-succinimidyl carbamate
490 (AQC) and subjected to analysis by UPLC. All enzymatic assays were performed in
491 triplicate.

492 The detection methodology for theanine production remains consistent with the
493 aforementioned approach. The 100 μL reaction mixture, containing 20 mM Ala, 45 mM
494 Glu, 50 mM HEPES (pH 7.5), 0.1 mM PLP, 30 mM $MgCl_2$, 10 mM ATP, 0.03 mM
495 PsGS/MmGMAS, and 0.025 mM CsAlaDC/CsAlaDC ^{L110F}/CsAlaDC $^{L110F/P114A}$, was
496 prepared and incubated at standard conditions for 1 h. Subsequently, the reaction was
497 terminated via immersion of the reaction vessel in a metal bath at 96 °C for 3 min (31).

498 Theanine was derivatized with AQC and subjected to analysis by UPLC. All enzymatic
499 assays were performed in triplicate.

500 **Site-directed mutagenesis**

501 Site-directed mutagenesis experiment was conducted using a PCR method from the
502 wild-type construct pET-28a-*AtSerDC* and pET-22b-*CsAlaDC*, respectively. *Dpn* I
503 endonucleases were used to digest the parental DNA template. The reaction mixture
504 was used to transform *E. coli* DH5α competent cells and the plasmids from positive
505 strains were extracted to *E. coli* BL21 (DE3) for protein expression, purification, and
506 analysis of the enzymatic activity.

507 ***In vivo* enzyme activity assay in *N. benthamiana***

508 The amplified PCR products were fused to the plant expression vector, pCAMBIA1305.
509 Linearization was conducted by restriction digest with *Spe* I and *Bam*H I. The
510 recombinant colonies were selected for PCR validation on the appropriate antibiotics
511 plate. After validation, the plasmids were electroporated into *Agrobacterium*
512 *tumefaciens* strain GV3101(pSoup-p19). Empty pCAMBIA1305 vector, with an intron
513 containing the GFP gene, was treated as the control.

514 *Agrobacterium* transient expression assays were performed on 5-week-old *N.*
515 *benthamiana* plants. *Agrobacterium tumefaciens* strain GV3101 (pSoup-p19), carrying
516 the above-described vectors, were cultured in Luria Bertani (LB) medium, containing
517 appropriate antibiotics, at 28 °C. When the absorbance of bacteria colonies reached
518 OD_{600} = 0.6-0.8, bacterial cells were collected and resuspended in MMA solution (10
519 mM $MgCl_2$, 10 mM 2-(N-morpholino) ethane sulfonic acid (MES), pH 5.6). After the

520 OD_{600} of the resuspended bacterial solution was adjusted to approx. 1.0, acetosyringone
521 (AS) was added with a final concentration of 200 μM, and this solution was then
522 incubated, at room temperature, for at least 3 h in darkness. Next, cell suspensions were
523 infiltrated into *N. benthamiana* leaves with a needle-free syringe. These *N.*
524 *benthamiana* leaves were then collected 3 days post infiltration, frozen in liquid
525 nitrogen and stored at -80 °C. Internal EA in *N. benthamiana* leaves was extracted as
526 previously described (18), and then the solvent was subjected to gas chromatography-
527 mass spectrometry GS-MS system.

528 **Transcript level analysis in *N. benthamiana***

529 Total RNA was isolated from samples using the RNAprep Pure Plant Kit (Tiangen,
530 Beijing, China), according to the manufacturer's protocol. The cDNAs were
531 synthesized using TransScript One-Step gDNA Removal and cDNA Synthesis
532 SuperMix Kit (TransGen Biotech, Beijing, China). The qRT-PCR assays were
533 performed, as previously described (38). Primers used for qRT-PCR assays were listed
534 in Supplementary table 3 and the qRT-PCR was run on a Bio-Rad CFX96TM RT PCR
535 detection system and CFX Manager Software. Each reaction reagent (20 μL) contained
536 0.4 μL forward and reverse primers (10 μM), 2 μL cDNA (200±5 ng/μL), 10 μL SYBR
537 Green Supermix (Vazyme, Nanjing, China) and 6.2 μL double-distilled water. Reaction
538 was performed by a two-step method: 95 ℃ for 5 min; 40 cycles of 95 ℃ for 10 s; and
539 60 ℃ for 30 s. The glyceraldehyde-3-phosphate dehydrogenase (GAPDH) gene was
540 used for internal normalization in each RT-PCR, and the $2^{-\Delta\Delta CT}$ method (39) was used
541 to calculate the relative gene expression. All samples were performed with three

replicates.

Multiple sequence alignment

MUSCLE was used to generate the protein multiple sequence alignment of amino acid decarboxylases with default settings (40). ESPript 3.x was used to display the multiple sequence alignment (41).

Phylogenetic analysis

The AtSerDC was used as a query for searching its homologs in Embryophyta using BLASTP (42) (version 2.13.0). To fetch homologs of AtSerDC in distant species, we performed a two-round blast. We first search for the best hit in the given order under Embryophyta. Then we blast each species under a provided order using the order best hit as a query and as many as 10 hits were kept. The sequences with e values <= 0.001 were removed. We further dropped sequences with lengths less than 200 or larger than 800 amino acids. Then multiple sequence alignment was performed via clustalo (version 1.2.4) with default settings on these filtered sequences (43). The NJ algorithm implemented in clustalw (version 2.1) was adopted to build the phylogenetic tree (44). The R package ggtree (version 3.2.1) and universalmotif (version 1.12.4) were used for tree and motif visualization (45).

Acknowledgments

We thank the team of beamline BL18U1 in the Shanghai Synchrotron Radiation Facility for diffraction data collection.

Funding

This study is supported by grants from the Ministry of Science and Technology of China

(2019YFA0904100 to W.G, 2022YFF1003103 to X.W) and the Natural Science Foundation of China (T2221005 to W.G, 32072624 to Z.Z).

Data availability

The structures of AtSerDC, CsAlaDC-EA complex and CsAlaDC have been deposited in the Protein Data Bank (PDB) under accession numbers 8JG7, 8JIK and 8JIJ respectively.

References

1. Drew, L. (2019) The growth of tea. *Nature* **566**, S2-S4.
2. Juneja, L. R., Chu, D. C., Okubo, T., Nagato, Y., Yokogoshi, H. (1999) L-theanine - a unique amino acid of green tea and its relaxation effect in humans. *Trends Food Sci. Technol.* **10**, 425-425.
3. Lin, J. Z. *et al*. (2022) Effect of the presence of stem on quality of oolong tea. *Foods* **11**, 3439.
4. Haskell, C. F., Kennedy, D. O., Milne, A. L., Wesnes, K. A., Scholey, A. B. (2008) The effects of L-theanine, caffeine and their combination on cognition and mood. *Biol. Psychol.* **77**, 113-122.
5. Kakuda, T. *et al*. (2000) Protective effect of gamma-glutamylethylamide (theanine) on ischemic delayed neuronal death in gerbils. *Neuro. Sci. Lett.* **289**, 189-192.
6. Liu, Q., Duan, H., Luan, J., Yagasaki, K., Zhang, G. (2009) Effects of theanine on growth of human lung cancer and leukemia cells as well as migration and invasion of human lung cancer cells. *Cytotechnology* **59**, 211-217.
7. Lu, K. *et al*. (2004) The acute effects of L-theanine in comparison with alprazolam on anticipatory anxiety in humans. *Hum. Psychopharmacol.* **19**, 457-465.
8. Takagi, Y. *et al*. (2010) Combined administration of (L)-cystine and (L)-theanine enhances immune functions and protects against influenza virus infection in aged mice. *J. Vet. Med. Sci.* **72**, 157-165.
9. Zheng, G., Sayama, K., Okubo, T., Juneja, L. R., Oguni, I. (2004) Anti-obesity effects of

three major components of green tea, catechins, caffeine and theanine, in mice. *In Vivo* **18**, 55-62.

10. Lin, S. *et al*. (2023) Theanine metabolism and transport in tea plants (*Camellia sinensis* L.): advances and perspectives. *Crit. Rev. Biotechnol*. **43**, 327-341.
11. Deng, W., Ogita, S., Ashihara, H. (2008) Biosynthesis of theanine (γ-ethylamino-l-glutamic acid) in seedlings of *Camellia sinensis*. *Phytochem. Lett.* **1**, 115-119.
12. Deng, W.-W., Ogita, S., Ashihara, H. (2009) Ethylamine content and theanine biosynthesis in different organs of camellia sinensis seedlings. *Z. Naturforsch. C* **64**, 906-906.
13. Cheng, S. *et al*. (2019) Differential accumulation of specialized metabolite L-theanine in green and albino-induced yellow tea (*Camellia sinensis*) leaves. *Food Chem.* 276, 93-100.
14. Takeo, T. (1974) L-alanine as a precursor of ethylamine in *Camellia sinensis*. *Phytochem.* **13**, 1401-1406.
15. Zhu, B. *et al*. (2021) CsAlaDC and CsTSI work coordinately to determine theanine biosynthesis in tea plants (*Camellia sinensis* L.) and confer high levels of theanine accumulation in a non-tea plant. *Plant Biotechnol. J.* **19**, 2395-2397.
16. Bai, P. *et al*. (2021) Biochemical characterization of specific Alanine Decarboxylase (ADC) and its ancestral enzyme Serine Decarboxylase (SDC) in tea plants (*Camellia sinensis*). *BMC Biotechnol.* **21**,17.
17. Han, S. W., Shin, J. S. (2022) Aromatic L-amino acid decarboxylases: mechanistic features and microbial applications. *Appl. Microbiol. Biotechnol.* **106**, 4445-4458.
18. Berger, M., Gray, J. A., Roth, B. L. (2009) The expanded biology of serotonin. *Annu. Rev. Med.* **60**, 355-366.
19. Khan, M. Z., Nawaz, W. (2016) The emerging roles of human trace amines and human trace amine-associated receptors (hTAARs) in central nervous system. *Biomed. Pharmacother*. **83**, 439-449.
20. Ng, J., Papandreou, A., Heales, S. J., Kurian, M. A. (2015) Monoamine neurotransmitter disorders--clinical advances and future perspectives. *Nat. Rev. Neurol.* **11**, 567-584.
21. Rontein, D. *et al*. (2001) Plants synthesize ethanolamine by direct decarboxylation of serine using a pyridoxal phosphate enzyme. *J. Biol. Chem.* **276**, 35523-35529.

22. Gibellini, F., Smith, T. K. (2010) The Kennedy pathway-De novo synthesis of phosphatidylethanolamine and phosphatidylcholine. *IUBMB Life* **62**, 414-428.
23. Mudd, S. H., Datko, A. H. (1989) Synthesis of ethanolamine and its regulation in *Lemna paucicostata*. *Plant Physiol.* **91**, 587-597.
24. Zinser, E. *et al*. (1991) Phospholipid synthesis and lipid composition of subcellular membranes in the unicellular eukaryote *Saccharomyces cerevisiae*. *J. Bacteriol.* **173**, 2026-2034.
25. Yunus, I. S., Liu, Y.-c., Nakamura, Y. (2016) The importance of SERINE DECARBOXYLASE1 (SDC1) and ethanolamine biosynthesis during embryogenesis of *Arabidopsis thaliana*. *Plant J.* **88**, 559-569.
26. Komori, H. et al. (2012) Structural Study Reveals That Ser-354 Determines Substrate Specificity on Human Histidine Decarboxylase* *J Biol Chem.* **287**, 29175-83.
27. Huang, J. et al. (2018) Lactobacillus brevis CGMCC 1306 glutamate decarboxylase: Crystal structure and functional analysis. *Biochem Biophys Res Co*. **503**, 1703-1709
28. Paiardini, A., Giardina, G., Rossignoli, G., Voltattorni, C. B., Bertoldi, M. (2017) New insights emerging from recent investigations on human group II pyridoxal 5'-phosphate decarboxylases. *Curr. Med. Chem.* **24**, 226-244.
29. Torrens-Spence, M. P. *et al*. (2020) Structural basis for divergent and convergent evolution of catalytic machineries in plant aromatic amino acid decarboxylase proteins. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 10806-10817.
30. Hagihara, R., Ohno, S., Hayashi, M., Tabata, K., Endo, H. (2021) Production of L-theanine by *Escherichia coli* in the absence of supplemental ethylamine. *Appl. Environ. Microbiol.* **87**, e00031-21.
31. Yamamoto, S., Wakayama, M. & Tachiki, T. (2007) Characterization of theanine-forming enzyme from *Methylovorus mays* no. 9 in respect to utilization of theanine production. *Biosci. Biotechnol. Biochem.* **71**, 545-552.
32. Holm, L. (2022) Dali server: structural unification of protein families. *Nucleic. Acids Res.* **50**, W210-W215.
33. Okawa, A. *et al*. (2021) Structural basis for substrate specificity of L-methionine

decarboxylase. *Protein Sci.* **30**, 663-677.

34. Fenalti, G. *et al*. (2007) GABA production by glutamic acid decarboxylase is regulated by a dynamic catalytic loop. *Nat. Struct. Mol. Biol.* **14**, 280-286.
35. Gayathri, S. C., Manoj, N. (2019) Structural insights into the mechanism of internal aldimine formation and catalytic loop dynamics in an archaeal Group II decarboxylase. *J. Struct. Biol.* **208**, 137-151.
36. Torrens-Spence, M. P., von Guggenberg, R., Lazear, M., Ding, H., Li, J. (2014) Diverse functional evolution of serine decarboxylases: identification of two novel acetaldehyde synthases that uses hydrophobic amino acids as substrates. *BMC Plant Biol.* **14**, 247.
37. Qu, L., Gu, S., Zhang, J., Zhao, C., Deng, X. (2021) Determination of 18 amino acids in three different kinds of milk powder by ultra performance liquid chromatography coupled with pre-column derivatization. *Chin. J. Chromatogr.* **39**, 472-477.
38. Madeira, F. *et al*. (2022) Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* **50**, W276-W279.
39. Dong C. *et al*. (2020) Theanine transporters identified in tea plants (*Camellia sinensis* L.). *Plant J.* **101**, 57-70.
40. Schmittgen T. D. *et al*. (2000) Quantitative reverse transcription-polymerase chain reaction to study mRNA decay: comparison of endpoint and real-time methods. *Anal. Biochem.* **285**, 194–204.
41. Robert X., Gouet P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320-W324.
42. Camacho, C. *et al*. (2009) BLAST+: architecture and applications. *BMC Bioinf.* **10**, 421.
43. Sievers, F., Higgins, D. G. (2018) Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci.* **27**, 135-145.
44. Larkin, M. A. *et al*. (2007) Clustal W and clustal X version 2.0. *Bioinformatics.* **23**, 2947-2948.
45. Yu, G., Lam, T. T.-Y., Zhu, H., Guan, Y. (2018) Two methods for mapping and visualizing associated data on phylogeny using ggtree. *Mol. Biol. Evol.* **35**, 3041-3043.

Figures and table

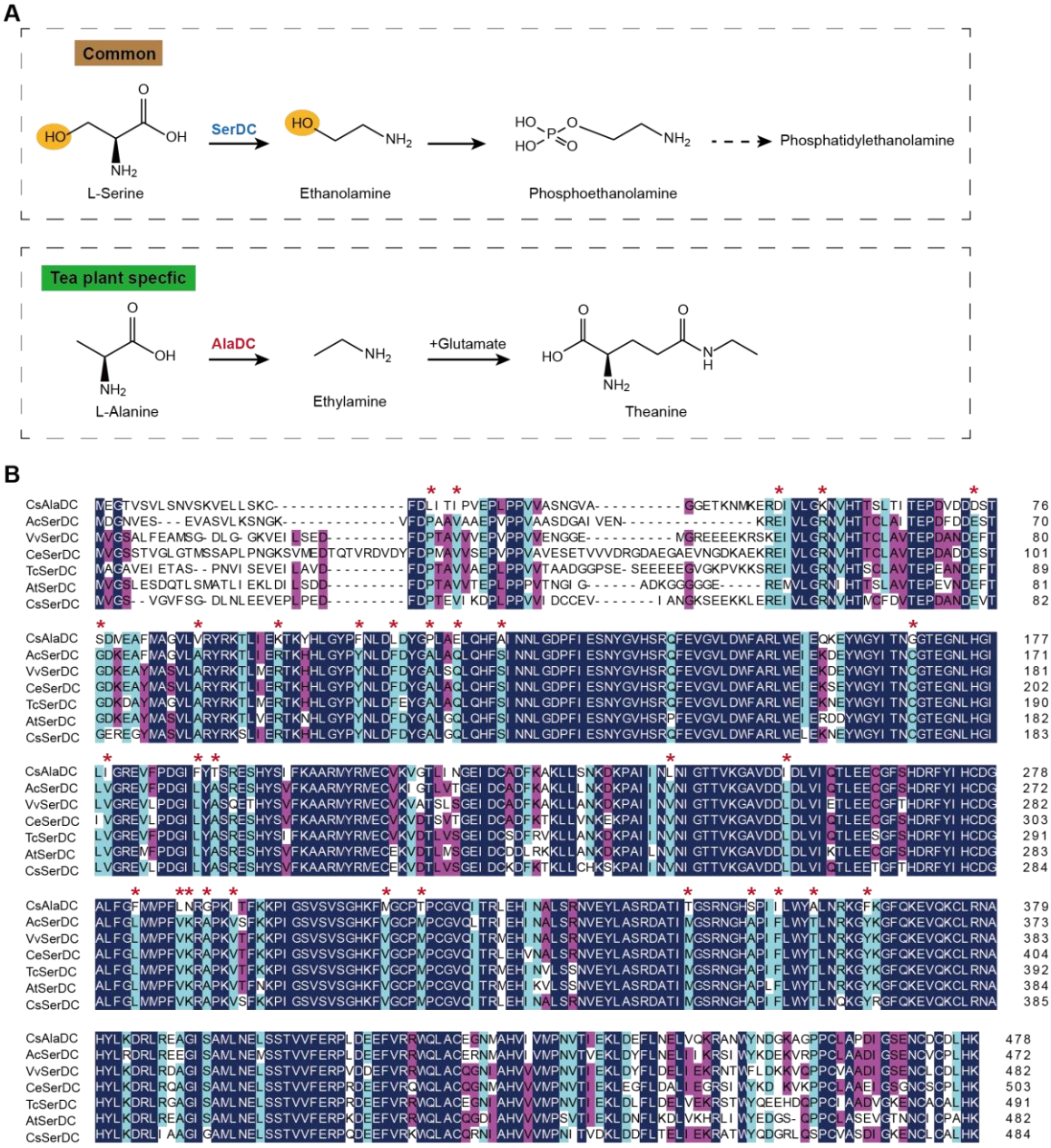


Figure 1. Metabolic pathways and sequence analysis of CsAlaDC and SerDCs in plants. (**A**) The decarboxylation of serine and alanine in plants. SerDC, serine decarboxylase; AlaDC, alanine decarboxylase. (**B**) Multiple alignment of the amino acid sequences of AlaDC and SerDCs. The amino acid sequences of the 6 SerDCs from kiwifruit (*Actinidia chinensis*), grape (*Vitis vinifera*), coffee (*Coffea eugenioides*), cocoa (*Theobroma cacao*) and Arabidopsis. Primary (100%), secondary (80%), and

687 tertiary (60%) conserved percent of similar amino acid residues were shaded in deep
688 blue, light blue and cheer red, respectively. “*” indicated amino acid residues mutated
689 only in CsAlaDC.
690

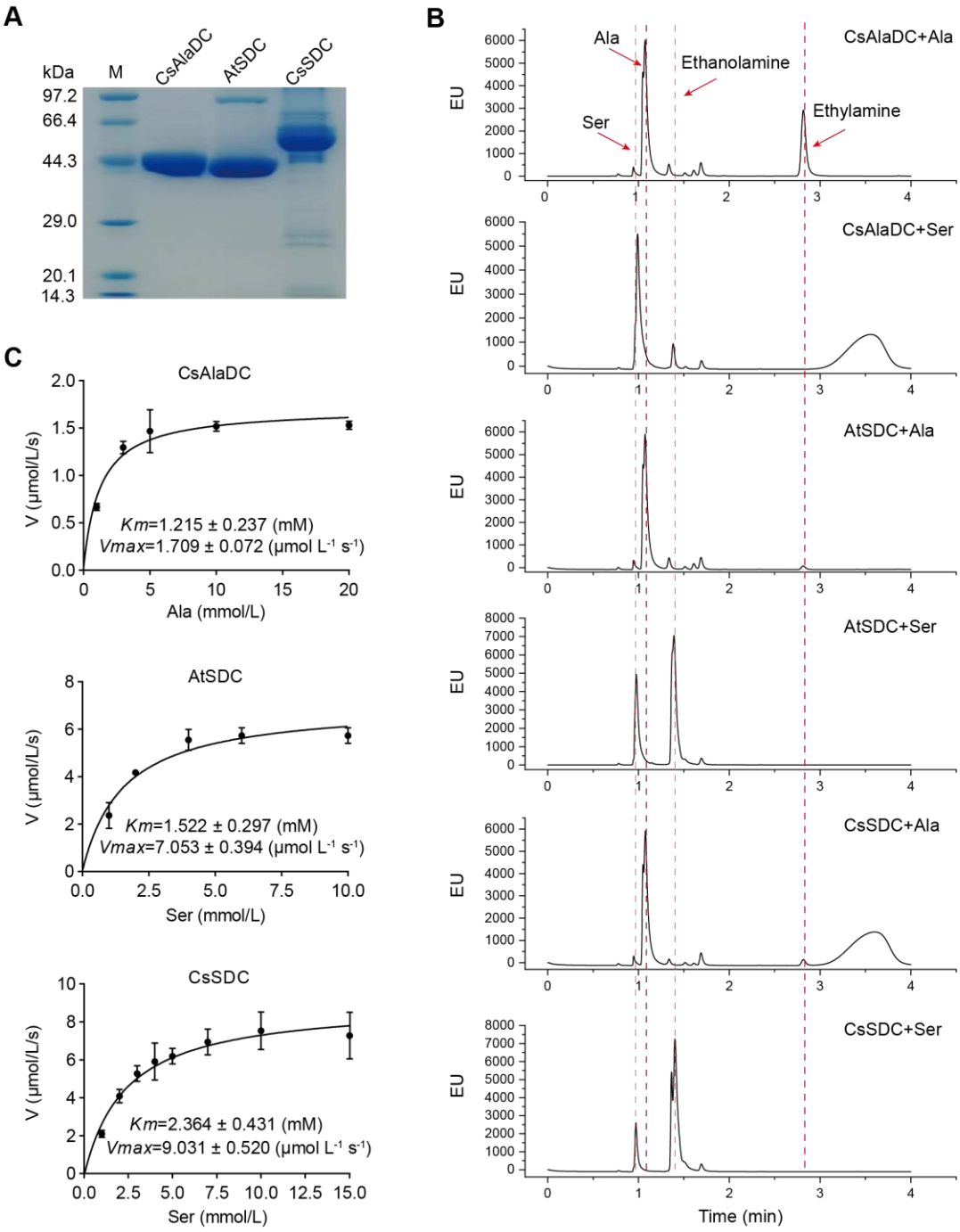


Figure 2. Purification and characterization of CsAlaDC, AtSerDC, and CsSerDC. (**A**) Identification of CsAlaDC, AtSerDC, and CsSerDC using SDS-PAGE. (**B**) Detection of enzyme activities of CsAlaDC, AtSerDC, and CsSerDC by UPLC. (**C**) Reaction rates of substrates with different concentrations catalyzed by CsAlaDC, AtSerDC, and CsSerDC.

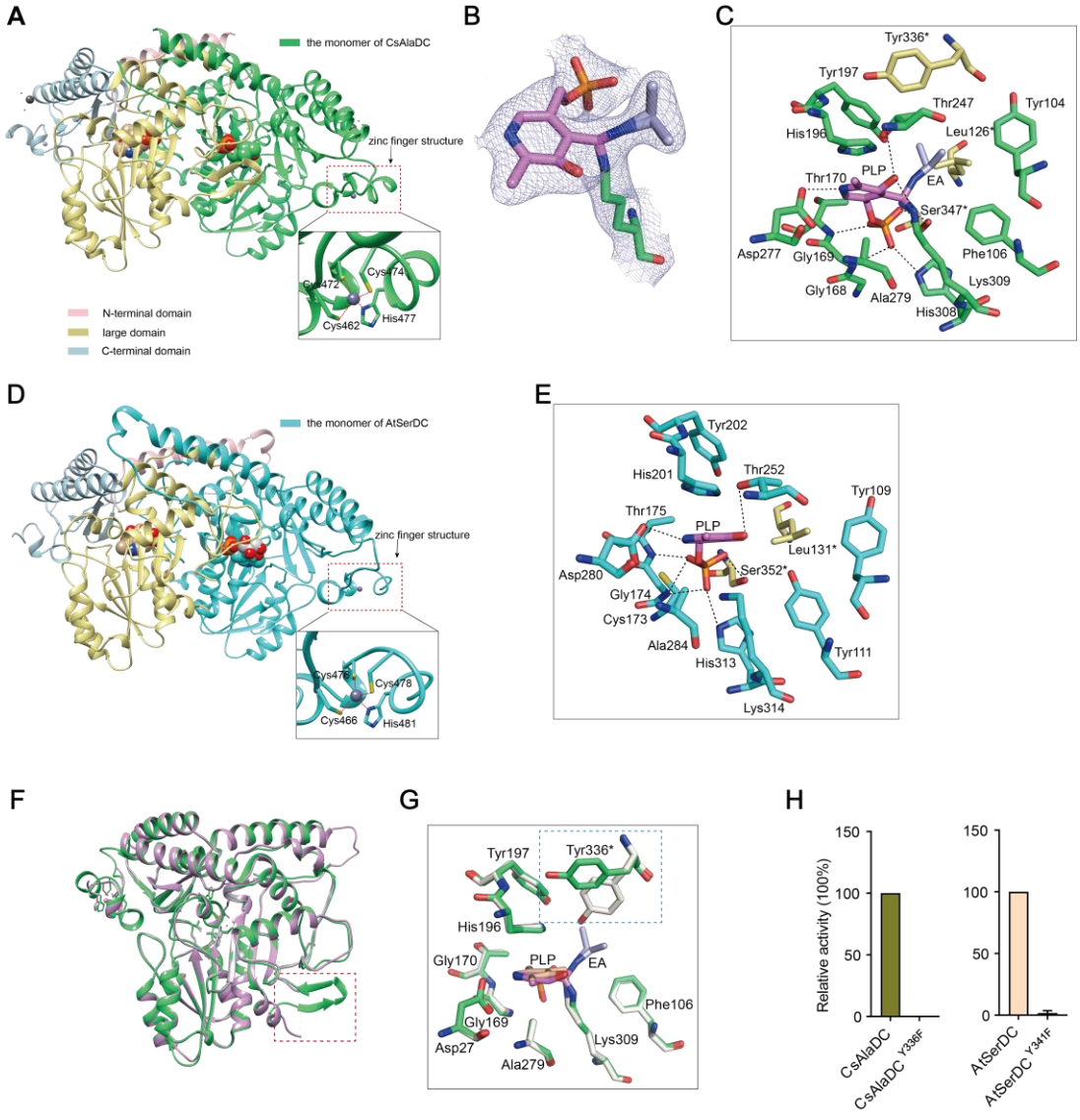


Figure 3. Crystal structures of CsAlaDC and AtSerDC. (**A**) Dimer structure of CsAlaDC. The color display of the N-terminal domain, large domain, and C-terminal domains of chain A is shown in light pink, khaki and sky blue, respectively. Chain B is shown in spring green. The PLP molecule is shown as a sphere model. The zinc finger structure at the C-terminus of CsAlaDC is indicated by the red box. The gray spheres represent zinc ions, while the red dotted line depicts the coordination bonds formed by zinc ions with cysteine and histidine. (**B**) The 2Fo-Fc electron density maps of K309-PLP-EA (contoured at 1σ level). The PLP is shown in violet, the K309 is shown in

706 spring green, and the EA is shown in lightblue. (**C**) Active center of the CsAlaDC-EA
707 complex, with hydrogen bonds denoted by black dotted lines. "*" denotes the amino
708 acids on adjacent subunits. (**D**) Dimer structure of AtSerDC. The color display of the
709 N-terminal domain, large domain, and C-terminal domains of chain A is shown in light
710 pink, khaki and sky blue, respectively. Chain B is shown in cyan. The PLP molecule is
711 shown as a sphere model. The zinc finger structure at the C-terminus of AtSerDC is
712 indicated by the red box. The gray spheres represent zinc ions, while the red dotted line
713 depicts the coordination bonds formed by zinc ions with cysteine and histidine. (**E**)
714 Active center of the AtSerDC, with hydrogen bonds denoted by black dotted lines. "*"
715 denotes the amino acids on adjacent subunits. (**F**) The monomers of CsAlaDC and
716 AtSerDC are superimposed. CsAlaDC is depicted in spring green, while AtSerDC is
717 shown in plum. The conserved amino acid catalytic loop is indicated by the red box.
718 (**G**) Amino acid residues of the active center in CsAlaDC apo and CsAlaDC-EA
719 complex are superimposed. CsAlaDC apo is shown in floral white, while CsAlaDC-EA
720 complex is shown in spring green. (**H**) The relative activity of wild-type CsAlaDC and
721 its Y336F mutant (left), as well as wild-type AtSerDC and its Y341F mutant (right) is
722 shown.
723

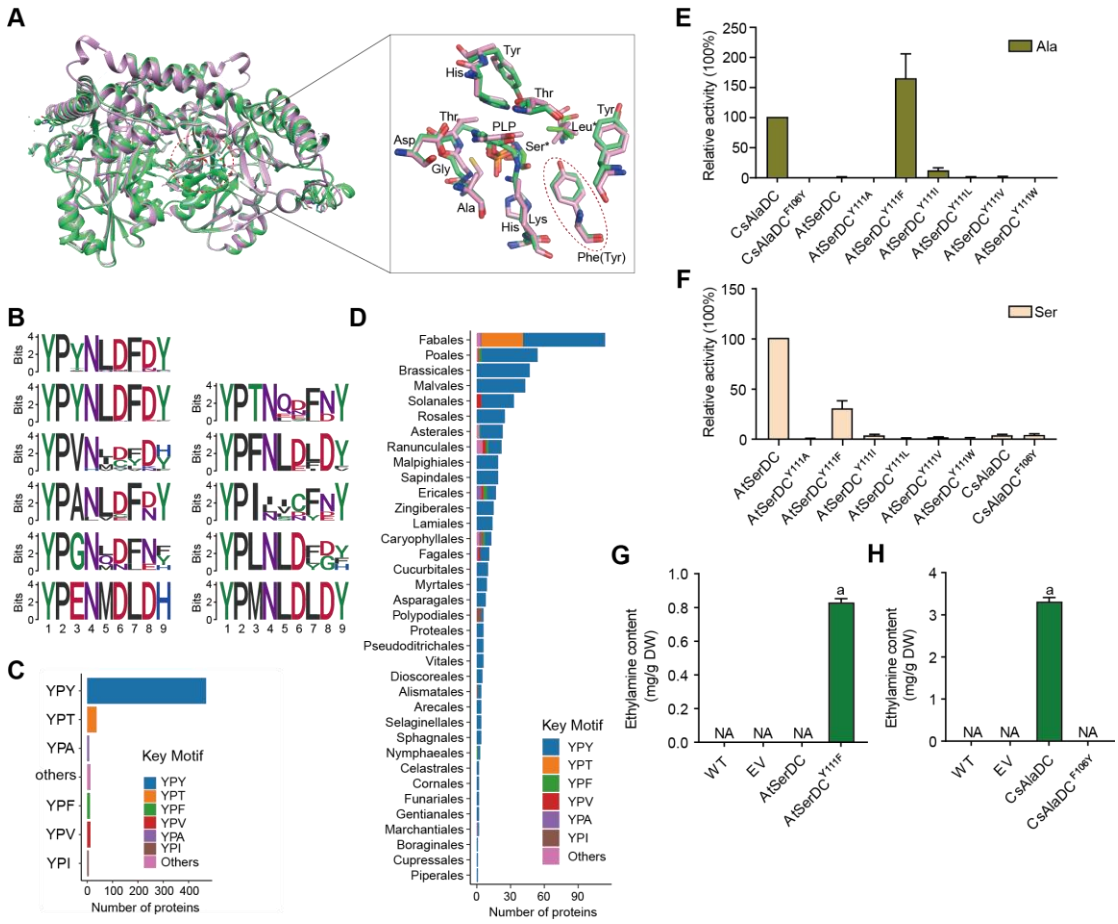


Figure 4. Key amino acid residues for substrate recognition. (**A**) Superposition of substrate binding pocket amino acid residues in CsAlaDC and AtSerDC. The amino acid residues of CsAlaDC are shown in spring green, the amino acid residues of AtSerDC are shown in plum, with the substrate specificity-related amino acid residue highlighted in a red ellipse. (**B**) Active-site-lining amino acid residues of SDC homologs from Embryophyta were identified. The height of each amino acid is scaled proportionally to the amount of information content (measured in bits). The first line depicts the conserved motif in all SerDC homologs from Embryophyta, whereas lines 2-5 represent the conserved motifs based on the variable third amino acid residue. (**C**) Histogram showing the distribution of the number of key motifs. (**D**) Histogram

735 showing the number of key motifs in different plant orders. (**E**) Relative enzyme
736 activities of wild-type CsAlaDC and mutant protein CsAlaDC^{F106Y} against Ala
737 substrate (columns 1 and 2), and enzyme activities of wild-type AtSerDC and various
738 AtSerDC mutant proteins against Ala substrate (columns 3-9) are presented. The
739 percentage graph shows the relative activity of each protein compared to wild-type
740 CsAlaDC activity (taken as a 100% benchmark). (**F**) Relative enzyme activities of wild-
741 type AtSerDC and AtSerDC mutant proteins (columns 1-7) against Ser substrates, and
742 enzyme activities of wild-type CsAlaDC and mutant protein CsAlaDC^{F106Y} against Ser
743 substrates(columns 8, 9) were measured. The percentage graph shows the relative
744 activity of each protein compared to the wild-type AtSerDC (taken as a 100%
745 benchmark). Three independent experiments were conducted. (**G**) The EA contents of
746 AtSerDC and its mutant AtSerDC^{Y111F} in *N. benthamiana*. (**H**) The EA contents of
747 CsAlaDC and its mutant CsAlaDC^{F106Y} in *N. benthamiana*. The significance of the
748 difference (P<0.05) was labeled with different letters according to Duncan's multiple
749 range test.

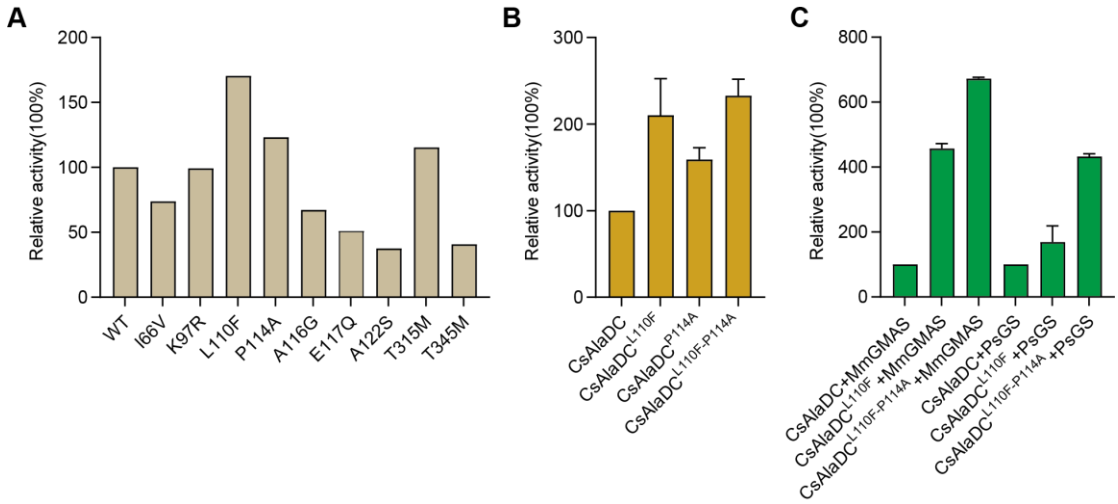


Figure 5. Mutations enhance CsAlaDC enzyme activity and theanine synthesis *in vitro*. (**A**) Relative enzyme activities of CsAlaDC mutant proteins against Ala substrate. (**B**) Relative enzyme activities of CsAlaDC^{L110F}, CsAlaDC^{P114A}, and CsAlaDC$^{L110F/P114A}$ against Ala substrate. (**C**) Histogram showing the relative content of theanine resulting from different combinations of alanine decarboxylase and theanine synthetase. Three independent experiments were conducted.

758 **Table 1.** Kinetic parameters of CsAlaDC, AtSerDC and CsSerDC

Enzyme	Substrate	*Km* (mM)	*Vmax* (μmol L^{-1} s^{-1})	*Kcat* (s^{-1})	*Kcat/Km* (s^{-1} M^{-1})
CsAlaDC	Alanine	1.215 ± 0.237	1.709 ± 0.072	0.068	55.7
AtSerDC	Serine	1.522 ± 0.297	7.053 ± 0.297	0.281	184.6
CsSerDC	Serine	2.364 ± 0.431	9.031 ± 0.520	0.361	152.7

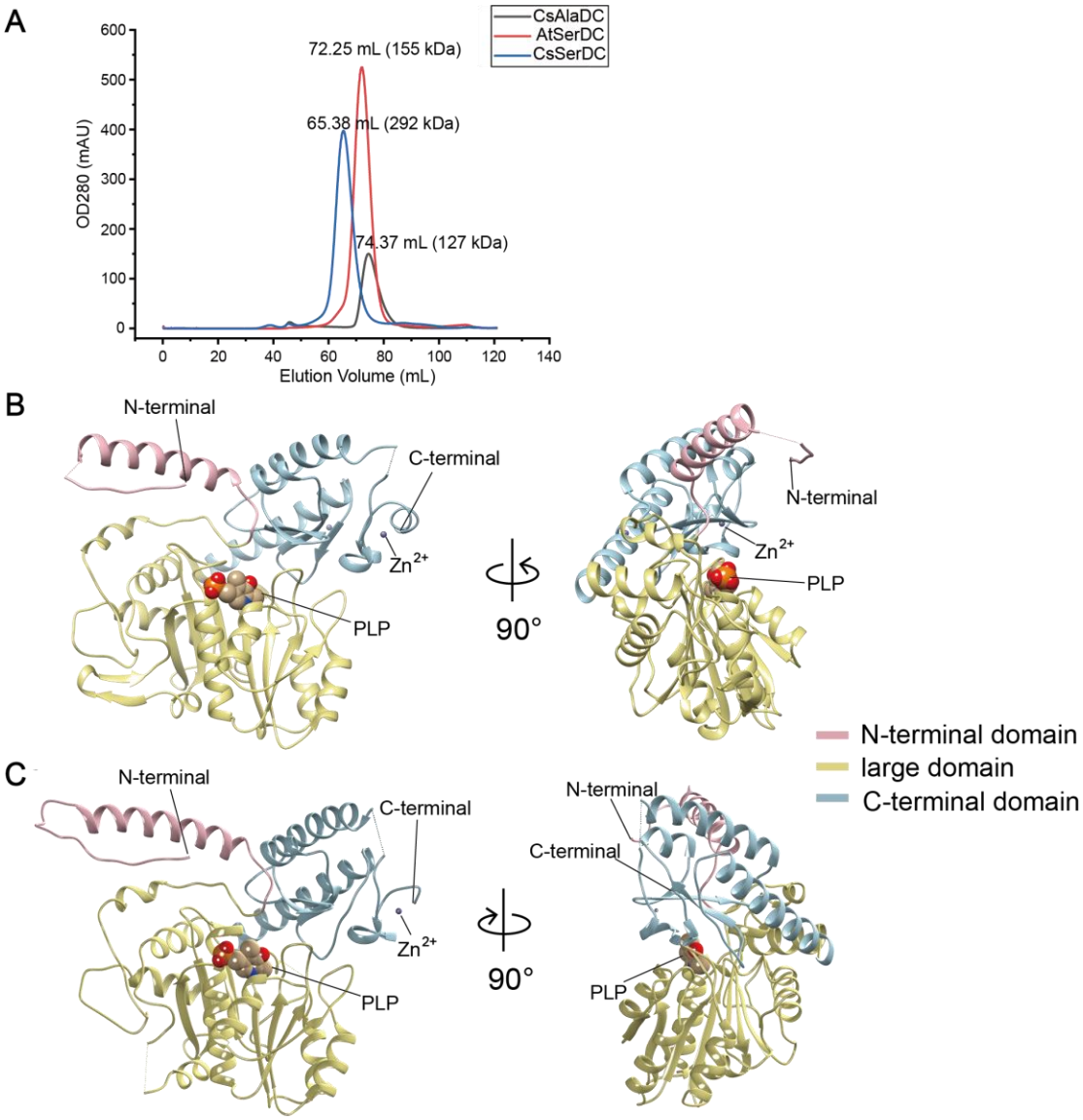
759

760

761

762 **Supplementary figures**

763



764

765 **Supplementary figure 1. Purification of CsAlaDC, AtSerDC, and CsSerDC and**

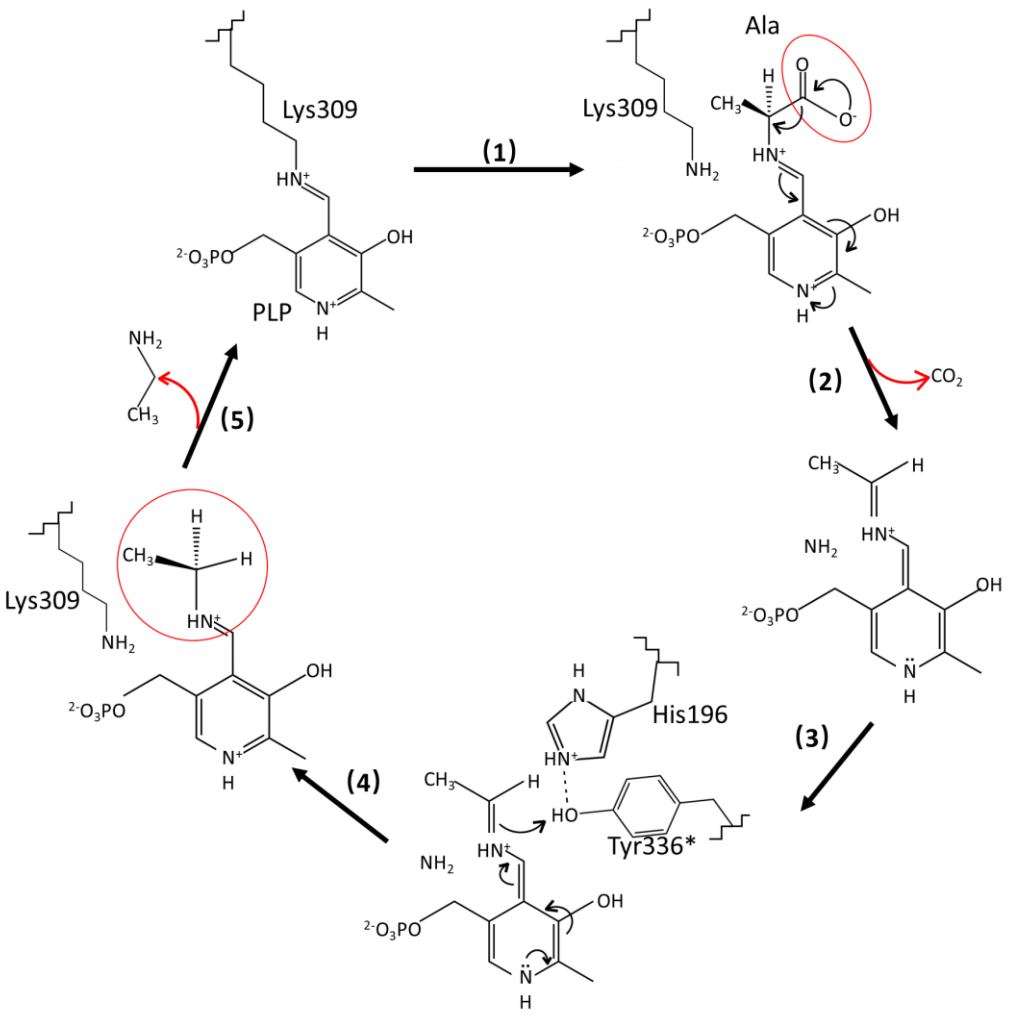
766 **Crystal structures of CsAlaDC and AtSerDC**. (**A**) Comparison of elution profiles of

767 CsAlaDC, AtSerDC and CsSerDC. (**B**) Monomer structure of CsAlaDC. (**C**) Monomer

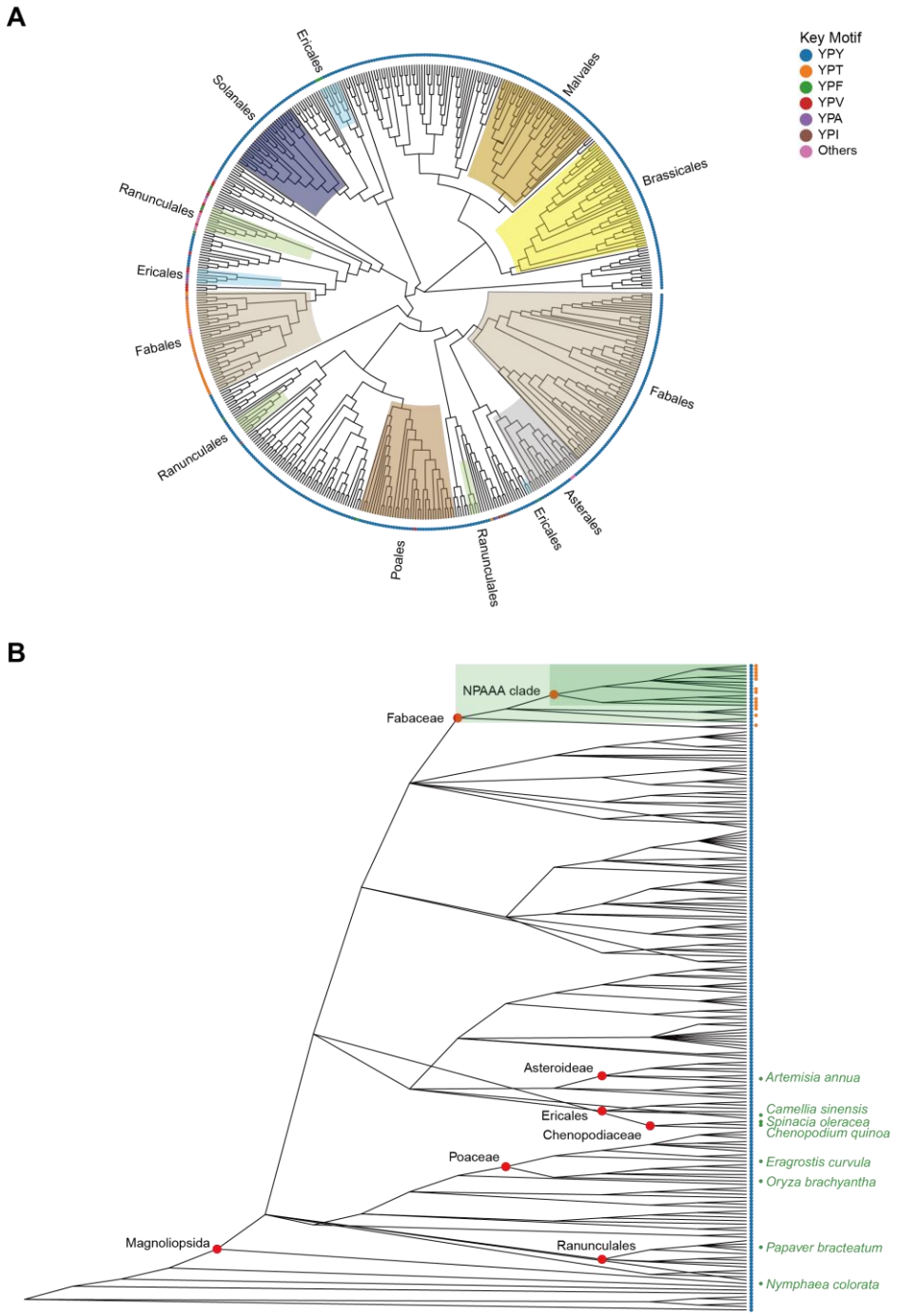
768 structure of AtSerDC. The color display of the N-terminal domain, large domain, and

769 C-terminal domains is shown in light pink, khaki and sky blue, respectively. The PLP

770 molecule is shown as a sphere model. The gray spheres represent zinc ions.

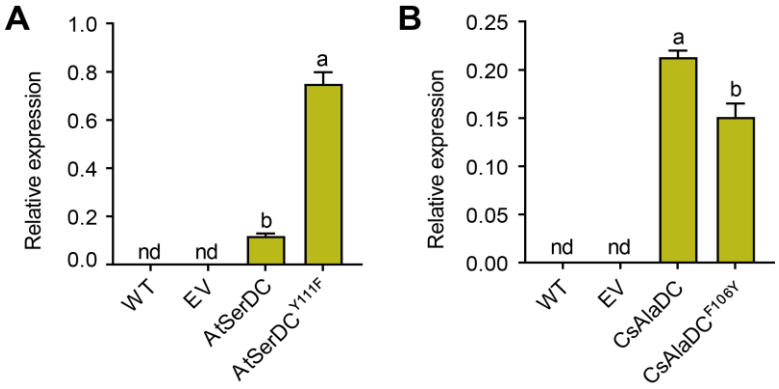


Supplementary figure 2. Catalytic mechanisms and conformational changes of CsAlaDC. After the transaldimation of the internal aldimine within CsAlaDC, resulting in the release of the active-site residue Lys309, the PLP amino acid external aldimine undergoes decarboxylation, leading to the removal of the α-carboxyl group as CO2. This process generates a quinonoid intermediate that is stabilized by the delocalization of paired electrons (1,2). The carbanion at Cα is subsequently protonated by the acidic p-hydroxyl group of Tyr336* located on the large loop, facilitated by its neighboring residue His196, which is situated on the small loop (3,4). Simultaneously, the internal aldimine LLP309 in CsAlaDC is restored, resulting in the release of the product (5).

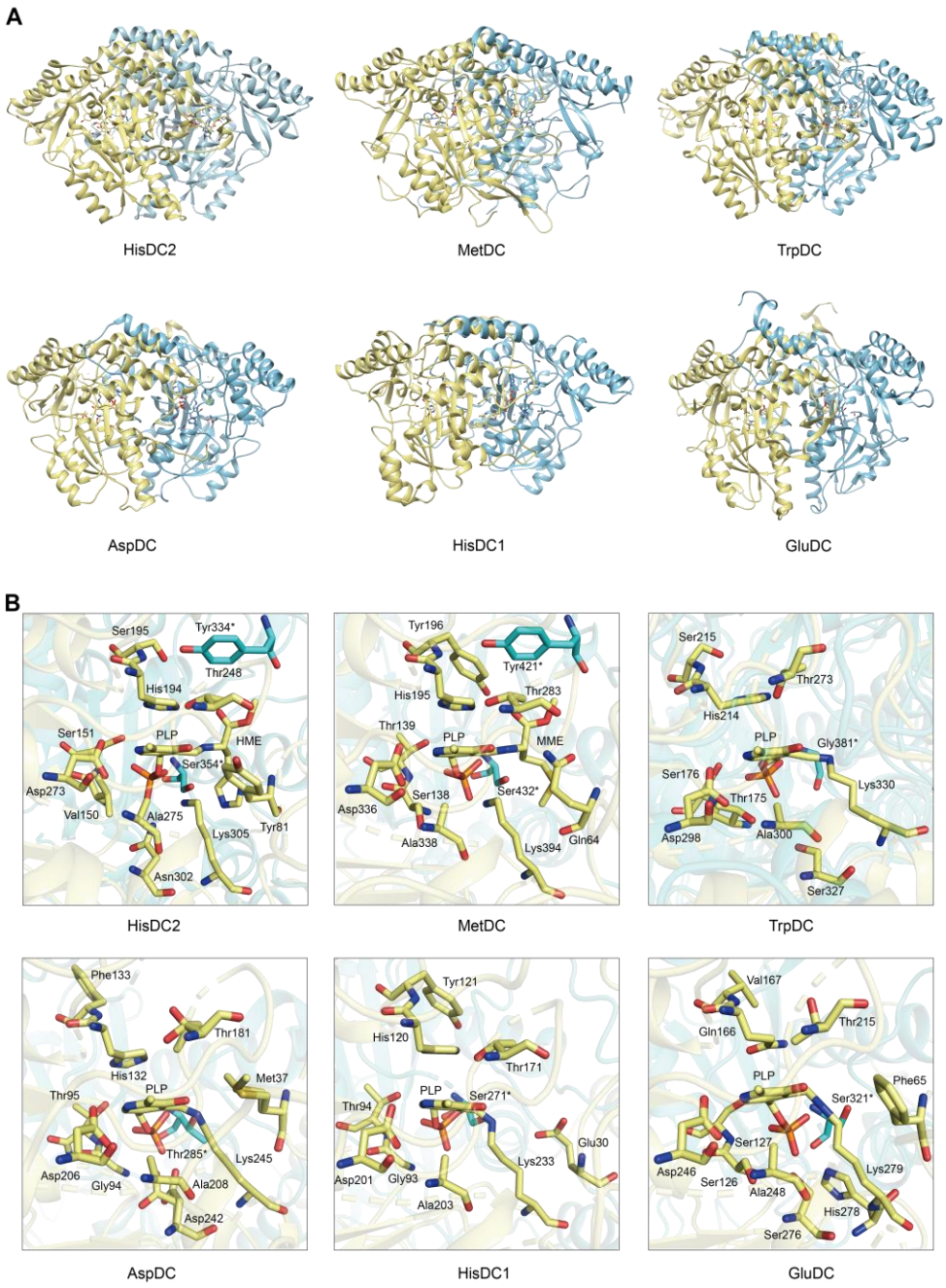


Supplementary figure 3. Evolutionary analysis of CsAlaDC in Embryophyta. (**A**) The presented diagram depicts an evolutionary tree of CsAlaDC, which is devoid of a root and solely portrays the topological structure of the tree without including distance information. The color of the inner ring corresponds to various orders, while the outer ring's leaf nodes are colored based on the motif types that the sequence exhibits. (**B**)

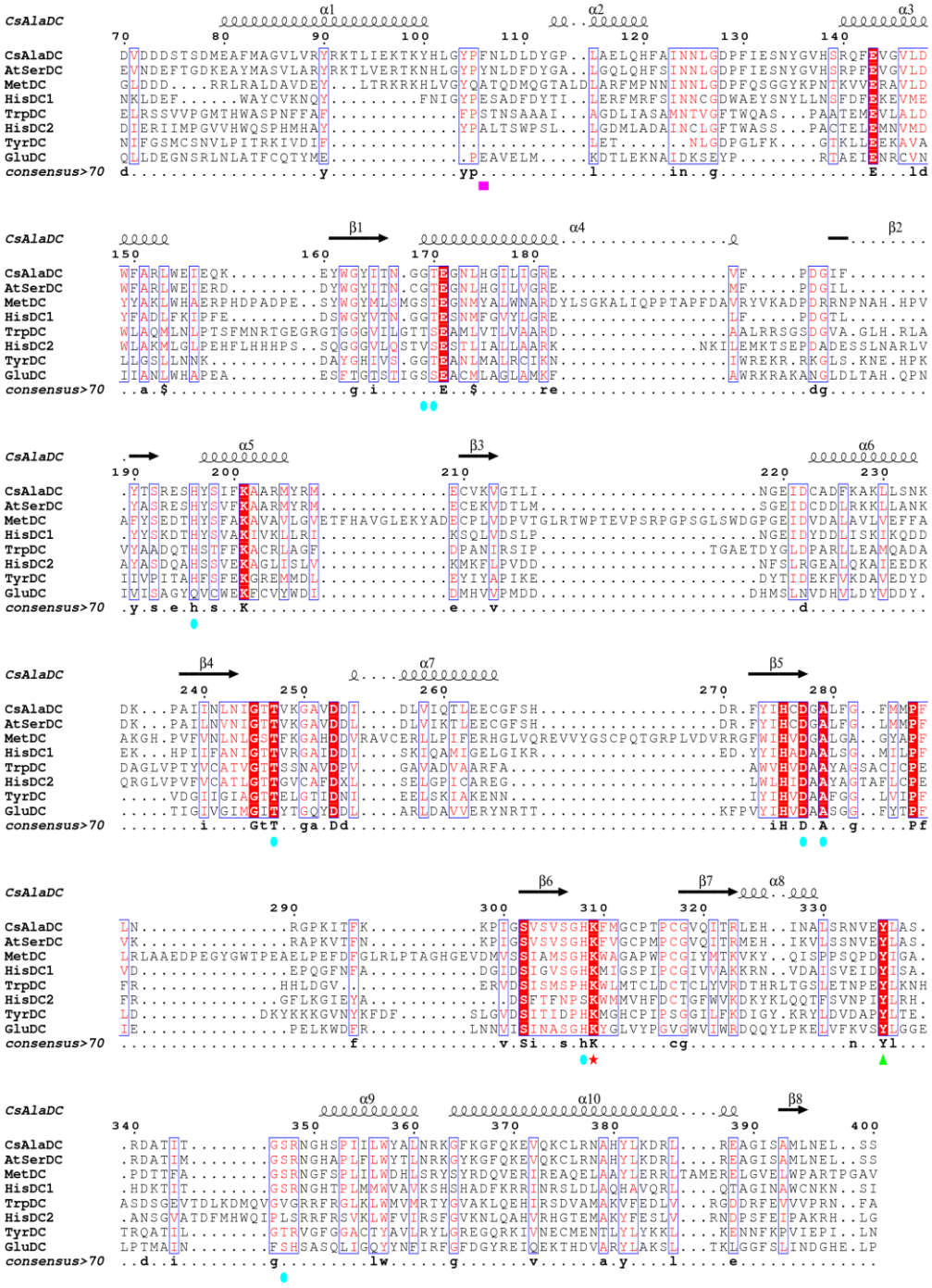
Diversity of serine decarboxylase-like proteins in Embryophyta (196 species). Colored scatter spots on the right side of leaf nodes correspond to the respective motifs shown in Figure A.



Supplementary figure 4. The relative mRNA levels of AtSerDC and its mutant Y111F (*A*), CsAlaDC and its mutant F106Y (*B*) in *N. benthamiana* leaves were measured by two primers. WT, wild type of *N. benthamiana*; EV, empty vector control; *NbGAPDH* was used as an internal control. Data represent mean ± SD (n=3). The significance of the difference (P<0.05) was labeled with different letters according to Duncan's multiple range test.



Supplementary figure 5. Structures of HisDC2, MetDC, TrpDC, AspDC, HisDC1 and GluDC. (**A**) The Overall Structures of HisDC2, MetDC, TryDC, AspDC, HisDC1 and GluDC. Chain A is shown in khaki, chain B is shown in cyan. (**B**) Amino acid residues in the substrate binding pocket of HisDC2, MetDC, TryDC, AspDC, HisDC1 and GluDC. The amino acid residues in chain A are shown in khaki, and the amino acid residues in chain B are shown in cyan.

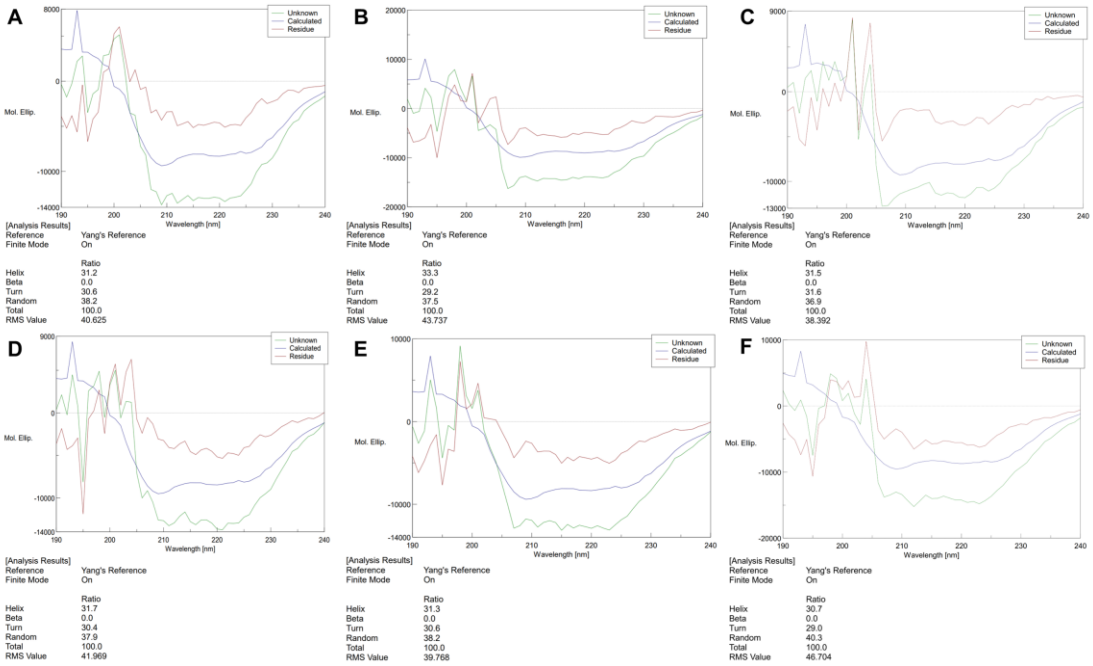


806

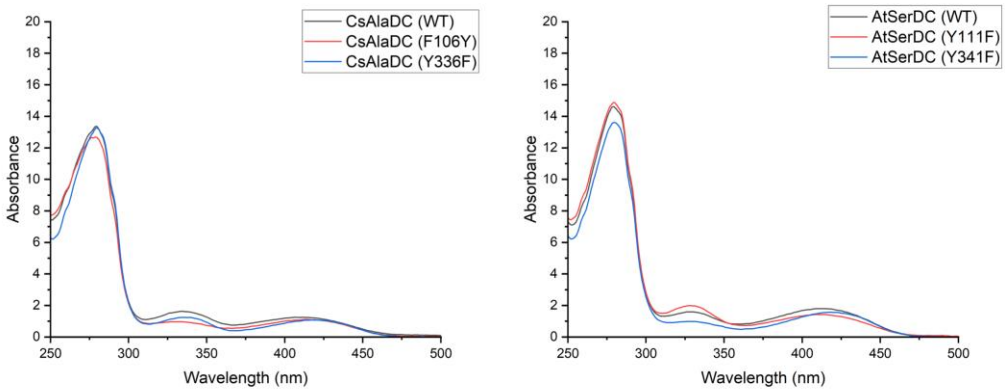
807 **Supplementary figure 6.** Multiple sequence alignment of CsAlaDC, AtSerDC, MetDC,

808 HisDC1, TrpDC, HisDC2, TyrDC, and GluDC were generated using MUSCLE and

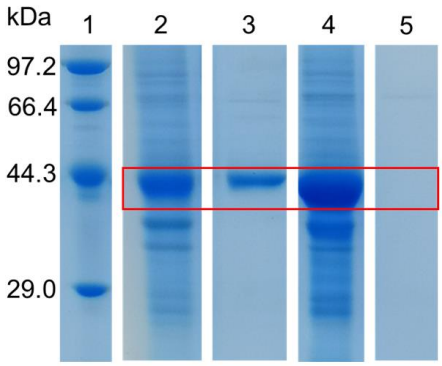
visualized with ESPript 3.x. Conserved amino acid residues in all eight proteins are highlighted with red backgrounds. The magenta box marks key amino acid residues involved in substrate recognition for CsAlaDC and AtSerDC. The Lys residue covalently bound to the PLP cofactor is denoted by a red star, while the green triangle indicates the Tyr residue associated with enzymatic activity. Amino acid residues involved in the CsAlaDC substrate binding pocket are marked with blue circles.



Supplementary figure 7. Circular Dichroism Spectra of proteins. (**A**) Circular Dichroism Spectra of CsAlaDC (WT). (**B**) Circular Dichroism Spectra of CsAlaDC^{Y336F}. (**C**) Circular Dichroism Spectra of CsAlaDC^{F106Y}. (**D**) Circular Dichroism Spectra of AtSerDC (WT). (**E**) Circular Dichroism Spectra of AtSerDC^{Y341F}. (**F**) Circular Dichroism Spectra of AtSerDC^{Y111F}.



Supplementary figure 8. Absorption Spectra of different proteins. (**A**) Absorption Spectra of CsAlaDC (WT), CsAlaDC^{Y336F} and CsAlaD^{F106Y}. (**B**) Absorption Spectra of AtSerDC (WT), AtSerDC^{Y341F} and AtSerDC^{Y111F}.



Supplementary figure 9. Purification of truncated CsAlaDC and AtSerDC. Lane 1 represents marker. The precipitated and eluted samples of AtSerDC with a truncated zinc finger structure are shown in lanes 2 and 3 respectively, the precipitated and eluted samples of CsAlaDC with a truncated zinc finger structure are illustrated in lanes 4 and 5 respectively.

Supplementary tables

Supplementary table 1. Data collection and refinement statistics

	CsAlaDC	CsAlaDC-EA complex	AtSerDC
Data Collection			
Space Group	*P3121*	*P3121*	*C2*
Cell Dimensions			
a, b, c (Å)	163.5,163.5, 50.2	164.0,164.0,50.2	115.4,149.5,139.1
α, β, γ (°)	α=β=90, γ=120	α=β=90, γ=120	α=γ=90, β=109.2
Wavelength (Å)	0.979	0.979	0.979
Resolution (Å)	50-2.50	50-2.60	50-2.85
Completeness (%)	99.3 (99.9)	99.6 (100.0)	99.6 (98.1)
Redundancy	5.4 (5.1)	13.0 (12.5)	5.9 (5.0)
$<I>/<\sigma I>$	9.3 (1.9)	18.9 (4.6)	10.6 (2)
R_{merge}	0.128 (0.603)	0.156 (0.562)	0.173 (0.551)
Refinement			
R_{work}/ R_{free}	0.173/0.222	0.172/0.223	0.214/0.262
No. reflections	25175	22813	49543
B-facator (Å^2)			
Protein	43.3	48.3	57.6
PLP	29.5	31.5	52.8
Zn^{2+}	58.7	43.8	71.6
Ca^{2+}	-	55.5	-
Ethylamine	-	27.5	-
Glycerol	-	-	51.7
Water	38.5	40.9	-
R.m.s. deviations			
Bond lengths (Å)	0.008	0.008	0.007
Bond angles (°)	1.531	1.601	1.410
No. of atoms			
Protein	3292	3265	3250
PLP	15	15	60
Zn^{2+}	2	1	4
Ca^{2+}	-	1	-
Ethylamine	-	5	-
Glycerol	-	-	24
Water	115	60	-
Ramachandran statistics (%)			
Favoured	94.01	93.58	91.52
Allowed	5.74	5.68	7.61
Outliers	0.25	0.74	0.87

R_{merge}=∑hl∑i|Ii(hkl)-<I(hkl)>|/∑hl∑i Ii(hkl), where<I(hkl)>is the main value of I(hkl).

R_{work}=∑||Fobs|-|Fcalc||/∑|Fobs|, where Fobs and Fcalc are observed and calculated structure factors.

The free R factor was calculated using 5% of reflections ommitted from the refinement.

Highest resolution shell is shown in parenthesis

Supplementary table 2. Primers used for gene cloning.

Name	Forward primers (5'-3')	Reverse primers (5'-3')
CsAlaDC	GGAATTCCATATGACAACAAGCCTCACCATCACGG	GATCTCGAGTCATTTATGAAGATCACAATCACAATTCTCACTTC
AtSerDC	GGAATTCCATATGACCACCAGCCTGGCCG	GATCTCGAGTTATTTATGGGCCGGACAAATGCAATTATTG
CsSerDC	ATGGTGGGAAGTGTTGGGGT	TCACTTGTGTAGTGCACAAAGACAGT

Supplementary table 3. Primers used for real-time PCR.

Name	Forward primers (5'-3')	Reverse primers (5'-3')
CsAlaDC	CACTGTGATGGGGCTCTGTT	TGTTATCTGGACACCGCACG
AtSerDC	TCACTCGCTGTAACGGAACC	AACTGACCAAGCGCACCATA
NbGAPDH	AGCTCAAGGGAATTCTCGATG	AACCTTAACCATGTCATCTCCC