Discovery of Novel Tyrosine Ammonia Lyases for the Enzymatic Synthesis of *p*-Coumaric Acid

Yannik Brack,[a] Chenghai Sun,[a] Dong Yi,*[a] and Uwe T. Bornscheuer*[a]

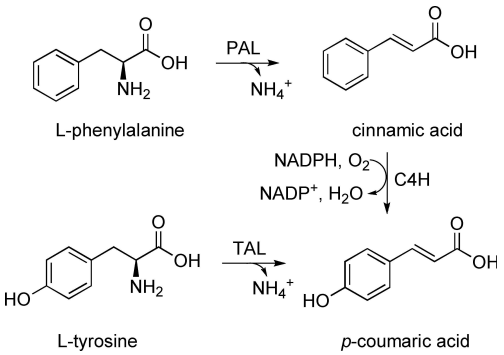
p-Coumaric acid (*p*-CA) is a key precursor for the biosynthesis of flavonoids. Tyrosine ammonia lyases (TALs) specifically catalyze the synthesis of *p*-CA from L-tyrosine, which is a convenient enzymatic pathway. To explore novel and highly active TALs, a phylogenetic tree-building approach was conducted including 875 putative TALs and 46 putative phenylalanine/tyrosine ammonia lyases (PTALs). Among them, 5 TALs and 3 PTALs were successfully characterized and found to exhibit the proposed enzymatic activity. The TAL from *Chryseobacterium luteum* sp. nov (TAL_{clu}) has the highest affinity ($K_M = 0.019$ mM) and conversion efficiency ($k_{cat}/K_M = 1631\ s^{-1}\cdot mM^{-1}$) towards L-tyrosine. The reaction conditions for two purified enzymes and their *E. coli* recombinant cells were optimized and *p*-CA yields of 2.03 g/L after 8 hours by TAL_{clu} and 2.35 g/L after 24 h by TAL *from Rivularia* sp. PCC 7116 (TAL_{rpc}) in whole cells were achieved. These TALs are thus candidates for the construction of whole-cell systems to produce the flavonoid precursor *p*-CA.

Introduction

Flavonoids are high value-added secondary metabolites in plants with anti-oxidative,[1] anti-inflammatory,[2] anti-mutagenic,[3] and anti-carcinogenic[2] properties for humans.[4] Since the extraction of flavonoids from plants is ineffective, the biotransformation by microbial cell factories became more promising.[5] In plants, the biosynthesis of flavonoids starts from the phenylalanine pathway, in which phenylalanine ammonia lyases (PALs, EC 4.3.1.24) catalyze the elimination of ammonia from L-phenylalanine to form the α,β-unsaturated cinnamic acid (Scheme 1).[6] This is subsequently hydroxylated by the P450 monooxygenase cinnamate-4-hydroxylase (C4H) to obtain *p*-coumaric acid (*p*-CA, Scheme 1).[7] Unfortunately, the plant-derived C4H is poorly expressed in prokaryotic host cells and requires a redox partner as electron supplier, thus the catalytic efficiency of this biosynthetic pathway is not satisfactory in microbial host cells.[8] Therefore, a more economical biosynthesis pathway for *p*-CA is required.

Tyrosine ammonia lyases (TALs, EC 4.3.1.23) and phenylalanine/tyrosine ammonia lyases (PTALs, EC 4.3.1.25) can directly catalyze the deamination of L-tyrosine to *p*-CA (Scheme 1).[9,10] Since the prices of L-tyrosine and L-phenylalanine are almost equal, a simplified one-step enzymatic reaction seems to be more economical for microbial cell factories than complex hydroxylation pathways. The TALs discovered so far are all derived from microorganisms, such as Proteobacteria,[10–12] Terrabacteria,[13–15] Spirochaetes,[12] and bacteria belonging to the Fibrobacteres, Chlorobi, and Bacteroidetes-group (FCB-group),[12,16] which is conducive to their expression in microbial host cells. However, their catalytic activity towards L-tyrosine is generally at a relative low level.[17,18] Besides, the PTALs found in Basidiomycota[12,15,19] and plants[9] favor L-tyrosine over L-phenylalanine. This low catalytic efficiency coupled with competitive inhibition of the product *p*-CA on TALs and PTALs[20,21] restricts the application of this enzymatic catalytic step. Therefore, identification of novel TALs with high activity is crucial.

Systematic bioinformatics analysis is an effective method to discover novel enzymes, by which we had already successfully explored the bacterial chalcone isomerase family.[22] The same concept has now been applied to search for novel TALs and PTALs.



Scheme 1. Biosynthetic pathway of *p*-CA formation. PAL: phenylalanine ammonia lyase, TAL: tyrosine ammonia lyase, and C4H: cinnamate-4-hydroxylase.

[a] Y. Brack, Dr. C. Sun, Prof. Dr. D. Yi, Prof. Dr. U. T. Bornscheuer
Dept. of Biotechnology & Enzyme Catalysis,
Institute of Biochemistry, University of Greifswald
Felix-Hausdorff-Straße 4, 17489 Greifswald (Germany)
E-mail: yid@sysbiosyn.com
uwe.bornscheuer@uni-greifswald.de

Supporting information for this article is available on the WWW under https://doi.org/10.1002/cbic.202200062

This article is part of a Special Collection dedicated to the Biotrans 2021 conference. Please see our homepage for more articles in the collection.

Results and Discussion

Systematic phylogenetic tree analysis of TALs and PTALs

In order to collect the TAL and PTAL sequences from protein databases for sequence alignment, we performed a BLAST search on the comprehensive and non-redundant UniParc database from UniProt (https://www.uniprot.org) with the protein sequences of TAL from *Rhodobacter sphaeroides* (TAL_{rsp}, UniProtKB accession no.: Q3IWB0) and PTALs from *Rhodotorula glutinis* ($PTAL_{rgl}$, UniProtKB accession no.: U5TV35) and *Petroselinum crispum* (PAL_{pcr}, UniProtKB accession no.: P24481) as search inquiries, and finally obtained 6,831 non-repetitive protein sequences. These sequences were used as the starting protein sequence set for a subsequent bioinformatic analysis.

TALs, PTALs and PALs belong to the enzyme class of aromatic ammonia lyases, which also includes histidine ammonia lyases (HALs, EC 4.3.1.3).[11] Although the enzymes in this family have high homology with each other and contain the specific cofactor 3,5-dihydro-5-methylidene-4*H*-imidazol-4-one (MIO) that is autocatalytically formed through cyclization and dehydration steps of the tripeptide Ala-Ser-Gly,[24] the characteristic amino acids in their active site pockets determine the specificity for substrate recognition (Figure 1). In TALs, the amino acid residues hydrogen-bonded to the *para*-hydroxy group on the benzene ring of L-tyrosine are H/L, Y/H, F/H, or S/H at the 89/90 positions related to TAL_{rsp}.[12] At the corresponding positions, H/Q, F/L and S/H are more common in PTALs, PALs and HALs, respectively.[25] Through the mutation of the residue H89 to Phe, the substrate specificity of the TAL_{rsp} could be switched from L-tyrosine to L-phenylalanine.[11] In addition, the amino acid residues that bind to the carboxylic acid group of L-tyrosine/ L-phenylalanine are also a characteristic binding site for the substrate specificity.[26] TALs, PALs and PTALs prefer a N/Q combination at the 435/436 positions (numbers related to TAL_{rsp}) instead of Q/E or T/E for HALs.[12] By analyzing the amino acid bias of these sites, this can help us to identify potential TALs and PTALs. Therefore, we screened all protein sequences containing the tripeptide Ala-Ser-Gly for the MIO-group, H/L, Y/H, F/H, S/H or H/Q at the 89/90 positions, and N/Q at the 435/436 positions. In total, 875 putative TALs and 46 putative PTALs were selected for a subsequent phylogenetic tree building.

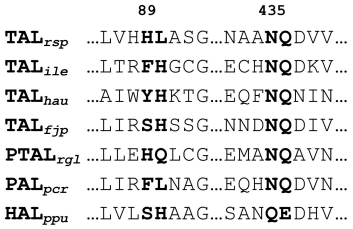


Figure 1. Substrate recognition sites of some well-known ammonia lyases. Alignment was performed with MUSCLE.[23] TAL_{rsp}: TAL from *Rhodobacter sphaeroides*; TAL_{ilo}: TAL from *Idiomarina loihiensis*; TAL_{hau}: TAL from *Herpetosiphon aurantiacus*; TAL_{fjo}: TAL from *Flavobacterium johnsoniae*; PAL_{pcr}: PAL from *Petroselinum crispum*; $PTAL_{rgl}$: PTAL from *Rhodotorula glutinis*; HAL_{ppu}: HAL from *Pseudomonas putida*.

The results of this phylogenetic tree building for TALs and PTALs are shown in Figure 2. These 875 putative TAL sequences are presented in 13 branches (Figure 2a). The proteins on the branches 1–4 contain F/H or Y/H at the 89/90 positions. Most of these sequences are from Proteobacteria, except those on branch 3, which are derived from Spirochaetes. The reported TAL from *Idiomarina loihiensis* (TAL_{ilo})[12] is located on a small sub-branch of branch 1. The proteins on the branches 5–8 contain S/H at the 89/90 positions, which originate from Terrabacteria, Proteobacteria, FCB-group, and unclassified bacteria, respectively. Unexpectedly, the tyrosine ammonia mutase from *Chondromyces crocatus* (TAM_{ccr})[27] is located on branch 5, which probably indicates that the proteins on this branch belong to ammonia mutases. This result indicates that TAMs have high homology to TALs without the difference at the key residues at the 89/90 and 435/436 positions, but probably contain a more flexible specific lid-like loop structure.[28] Branch

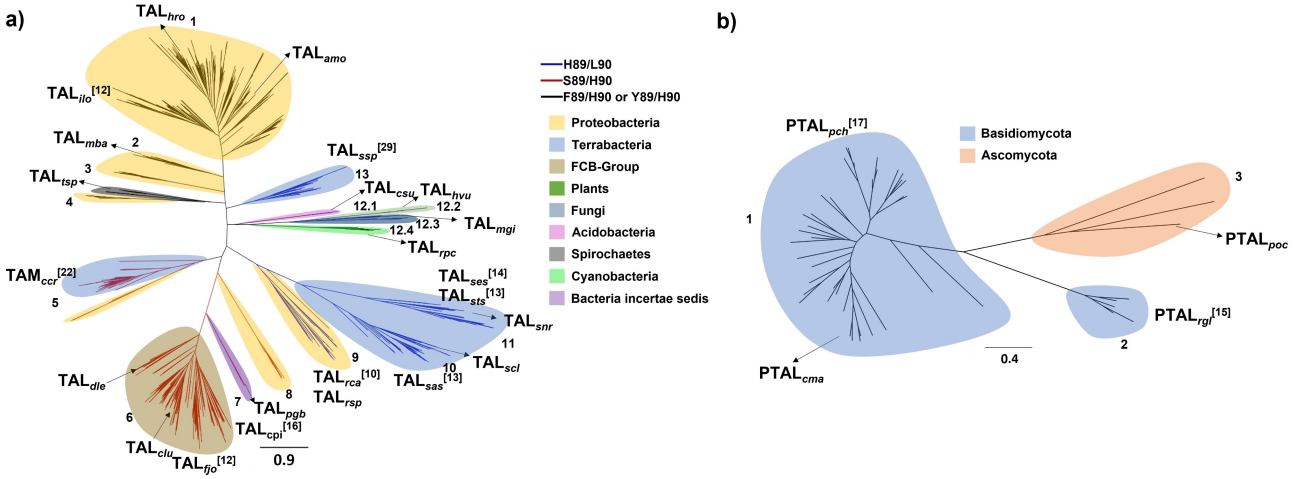


Figure 2. Phylogenetic tree analysis of putative TALs (a) and PTALs (b). The color of the branches symbolizes the residues at the substrate recognition positions. The background color indicates the origin of the sequences.

6 includes two known TALs from *Flavobacterium johnsoniae* (TAL_{fjo})[12] and *Chitinophaga pinensis* (TAL_{cpi}),[16] respectively. The proteins on the branches 9–13 contain H/L at the 89/90 positions. The reported TALs from *Rhodobacter capsulatus* (TAL_{rca}),[10] *Saccharothrix* sp. (TAL_{sas}),[13] *Streptomyces* sp. (TAL_{sts}),[13] *Saccharothrix espanaensis* (TAL_{ses}),[14] and *Streptomyces* sp. Tü 4128 (TAL_{ssp})[29] are presented on branches 9–11 and 13, respectively. Interestingly, the proteins on branch 12 are from diverse sources including bacterial, fungi, and even plants. The phylogenetic tree for PTALs contains three branches (Figure 2b), which includes 46 sequences derived from fungi. The PTAL from *Phanerochaete chrysosporium* ($PTAL_{pch}$)[19] and *Rhodotorula glutinis* ($PTAL_{rgl}$)[15] on branches 1 and 2, respectively, were reported to have high activity.[30]

After the phylogenetic tree analysis, we evaluated the differences in the active site pockets of the proteins on each branch. The protein structures for typical sequences from each branch were simulated and compared with TAL_{rsp}. Although their overall structures look consistent and 11 amino acid residues in all active site pockets are highly conserved, we could spot some differences in the specific lid-like loop structure (Figure 3, residues 67–69 related to TAL_{rsp}) besides the 89/90 positions. This loop probably influences the substrate specificity and the activity of the enzymes.[28] Therefore, it is relevant to characterize enzymes with different residues at these positions. To make the screening range as large as possible to cover the whole phylogenetic trees, 15 microbial-derived protein sequences (13 putative TALs and 2 putative PTALs) were chosen, which represent most branches and cover most active site pocket patterns (Table S1).

```
                60     67   89 149  297      333347  405
TAL_rsp     ...YGLTTGFGP...HL...ASG...QDAYSLR...N...GNF...MGAQV...
TAL_amo     ...YGVTTGYGD...FH...ASG...QDRYSLR...N...GHF...KAVCI...
TAL_hro     ...YGVTTGYGD...YH...ASG...QDRYSVR...N...GHF...KALQI...
TAL_mba     ...YGVTTGVGA...FH...ASG...QDRYSLR...N...RNF...KAMQI...
TAL_tsp     ...YGVTTGFGD...YH...ASG...QDRYSIR...N...GNF...KAAQI...
TAL_clu     ...YGVNTGFGP...SH...ASG...QEYYSLR...N...GNF...QGVQF...
TAL_dle     ...YGINTGFGP...SH...ASG...QEYYSIR...N...GNF...QGVQF...
TAL_pgb     ...YGINTGFGP...SH...ASG...QEVYSFR...N...GNF...QGLQF...
TAL_scl     ...YGATTGFGP...HL...ASG...QEPYSVR...N...GNF...QGVQL...
TAL_snr     ...YGLTSGFGP...HL...ASG...QEPYSLR...N...GNF...AGVQI...
TAL_csu     ...YGVNTGVGG...HL...ASG...QTCYSLR...N...GNF...KGMQL...
TAL_hvu     ...YGVTTGFGG...HL...ASG...QDRYALR...N...GNF...KGIEI...
TAL_mal     ...YGVNTGFGG...HL...ASG...QDRYSLR...N...GNF...KGLDV...
TAL_rpc     ...YGVTTGFGG...YH...ASG...QDRYSLR...N...GNF...KGLQI...
PTAL_poc    ...YGVTTGFGG...HQ...ASG...QDRYSLR...N...GNF...KGLDI...
PTAL_cma    ...YGLSTGFGG...HQ...ASG...QDRYPLR...N...GNF...KGVDI...
```

Figure 3. Alignment of amino acid residues in the active site pockets of putative TALs and PTALs. TAL_{rsp}: TAL from *Rhodobacter sphaeroides*; TAL_{amo}: TAL from *Aeromonas molluscorum* 848; TAL_{hro}: TAL from *Herbaspirillum robiniae*; TAL_{mba}: TAL from *Myxococcales bacterium*; TAL_{tsp}: TAL from *Treponema* sp.; TAL_{clu}: TAL from *Chryseobacterium luteum* sp. Nov; TAL_{dle}: TAL from *Dyadobacter* sp. Leaf189; TAL_{pgb}: TAL from *Patescibacteria* group bacterium; TAL_{scl}: TAL from *Streptomyces clavuligerus*; TAL_{snr}: TAL from *Saccharothrix* sp. NRRL B-1634; TAL_{csu}: TAL from *Candidatus sulfopaludibacter* sp.; TAL_{hvu}: TAL from *Hordeum vulgare* subsp. *Vulgare*; TAL_{mgi}: TAL from *Monosporascus* sp. GIB2; TAL_{rpc}: TAL from *Rivularia* sp. PCC 7116; $PTAL_{poc}$: PTAL from *Penicillium* sp. occitanis; $PTAL_{cma}$: PTAL from *Coprinopsis marcescibilis.*

Characterization of the TALs and PTALs

The genes for these 15 putative enzymes were codon-optimized for *E. coli* expression and transformed into BL21(DE3) host cells via a pET15b vector, of which 12 were expressed in soluble form (Figure S1). These 12 recombinant proteins were purified by affinity chromatography for further characterization using 2 mM L-tyrosine or 10 mM L-phenylalanine as substrates (Figure 4).

Among the ten characterized putative TALs, six of them presented activities towards L-tyrosine from 0.20 U/mg to 1.45 U/mg. The TAL from Bacteroides (*Chryseobacterium luteum* sp. Nov, TAL_{clu}) and Cyanobacteria (*Rivularia* sp. PCC 7116, TAL_{rpc}) showed the highest specific activity (0.8 U/mg and 1.45 U/mg, respectively). Regarding the active site residues, no residues which lead to the high activity could be identified. The substrate specificity of TALs from Proteobacteria (*Aeromonas molluscorum* 848, TAL_{amo}) and Cyanobacteria (TAL_{rpc}) is not very strict. They presented a slight activity towards L-phenylalanine. Unexpectedly, the TALs from Bacteria incertae sedis (*Patescibacteria* group bacterium, TAL_{pgb}) and plants (*Hordeum vulgare* subsp. *vulgare*, TAL_{hvu}) only showed activity towards L-phenylalanine, which indicated that these two enzymes are PALs instead of TALs. Unfortunately, the TALs from fungi (*Monosporascus* sp. GIB2, TAL_{mgi}) and Terrabacteria (*Streptomyces clavuligerus*, TAL_{scl}) had no activity at all. The two putative PTALs showed high activity towards both L-tyrosine and L-phenylalanine. To our knowledge, the PTAL from Ascomycota (*Penicillium* sp. *occitanis*, $PTAL_{poc}$) is the first characterized PTAL from the phylum Ascomycota.

Enzymatic properties of the TALs and PTALs

Furthermore, we determined the enzymatic properties of these novel TALs and PTALs, including pH and temperature optimum, thermostability, and kinetic constants. The deamination catalyzed by ammonia lyases requires alkaline conditions,[31] thus the activity of theses TALs and PTALs increased by increasing pH (Figure 5a). Most TALs and PTALs reached their highest

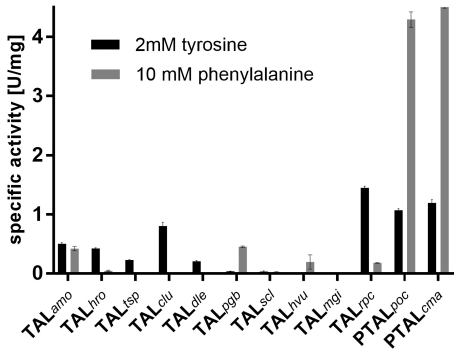


Figure 4. Specific activity of putative TALs/PTALs. Reaction system: 10 μL enzyme solution in 190 μL 20 mM PBS-Buffer (pH 7.5). Absorption measured at 290 nm for L-phenylalanine and 310 nm for L-tyrosine.

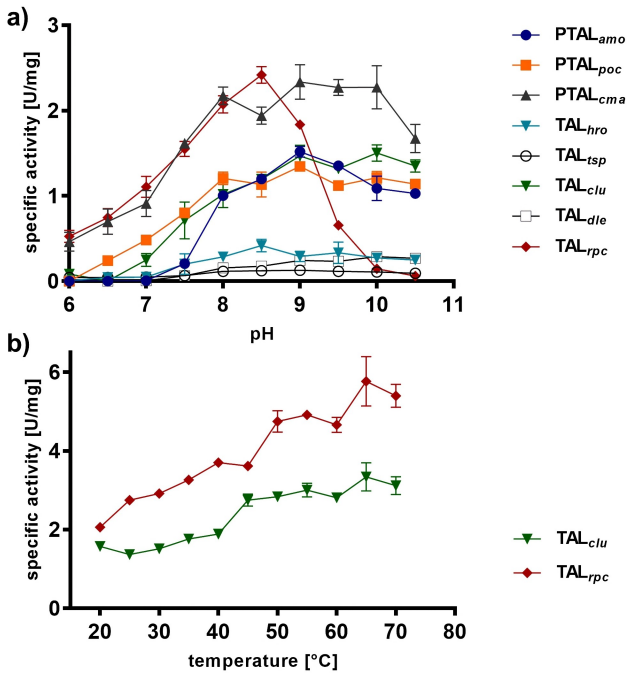


Figure 5. pH (a) and temperature (b) influence on the enzymatic activity of TALs and PTALs. Reaction system for pH analysis: 10 μL enzyme solution in 2 mM tyrosine in 50 mM PBS (pH 6.5–8) or 50 mM borate buffer (pH 8.5–10.5). The absorption was measured at 310 nm for 10 min. Reaction system for temperature analysis: 10 μL enzyme solution in 190 μL 2 mM L-tyrosine in 50 mM borate buffer (pH 8.5). Absorption was measured at 310 nm.

activity at pH 9.5–10, except for TAL$_{rpc}$, which showed its pH optimum at pH 8.5 and quickly lost activity at higher pH values. This distinctive feature shows that TAL$_{rpc}$ might have poor stability under strong alkaline conditions. Among the novel TALs and PTALs, TAL$_{rpc}$ presented the highest activity at pH 8.5 (2.42 U/mg), followed by PTAL$_{cma}$ (2.27 U/mg), TAL$_{clu}$ (1.50 U/mg), PTAL$_{poc}$ (1.21 U/mg), and PTAL$_{amo}$ (1.09 U/mg).

To determine the optimum temperature of these enzymes, we selected the most active specific TALs (TAL$_{clu}$ and TAL$_{rpc}$) and measured their activity in the range of 20 to 70 °C (Figure 5b). The activity of these enzymes was greatly promoted above 40 °C and kept increasing at higher temperatures. TAL$_{rpc}$ reached its maximum activity at 65 °C (5.8 U/mg), which is around 2.5 times higher than at 20 °C. TAL$_{clu}$ maintained high activity at 45–70 °C, with the highest activity also at 65 °C (3.2 U/mg). Both optimal temperatures are higher than that for TAL$_{sas}$ and TAL$_{sts}$ (55 °C and 50 °C, respectively).[13] These results indicate that TAL$_{clu}$ and TAL$_{rpc}$ are thermophilic enzymes.

Furthermore, we determined their thermostability via nanoDSF. Most enzymes melted between 70 and 80 °C, which indicates their excellent thermostability. TAL$_{scl}$ showed an extraordinary T_m value over 95 °C. TAL$_{pgb}$ and TAL$_{hro}$ also had high T_m values at 83.1 °C and 81.4 °C, respectively. TAL$_{amo}$, TAL$_{hvu}$, TAL$_{rpc}$ and PTAL$_{poc}$ presented lower T_m values below 70 °C (Table S2). The excellent thermophilicity of these enzymes contributes to their application in biocatalysis. We also determined the kinetic constants of these enzymes towards L-tyrosine (Table 1). Among these three novel enzymes, TAL$_{clu}$

Table 1. Kinetic parameters of TALs and PTAL towards L-tyrosine. Reaction system: 50 mM borate buffer (pH 9.5 for TAL$_{clu}$ and TAL$_{cma}$, pH 8.5 for TAL$_{rpc}$) with 2 mM, 1 mM, 0.5 mM, 0.25 mM, 0.125 mM, 0.075 mM, 0.0375 mM, or 0.01875 mM L-tyrosine. To determine The K_i values, the K_M values were measured at the *p*-CA concentrations of 0.03125 mM, 0.0625 mM and 0.125 mM and 0.250 mM.

Enzyme	K_M [mM]	k_{cat} [s^{-1}]	k_{cat}/K_M [$s^{-1}\cdot mM^{-1}$]	K_i [mM]
TALclu	0.019	31	1631	0.190
TALrpc	0.086	55	639	0.051
PTALcma	0.053	58	1094	0.068

exhibited the highest affinity towards L-tyrosine (K_M = 0.019 mM), which is similar to TAL$_{rca}$ (K_M = 0.016 mM)[10] but lower than TAL$_{rsp}$ (K_M = 0.008 mM).[12] However, the conversion efficiency of TAL$_{clu}$ reaches k_{cat}/K_M = 1631 $s^{-1}\cdot mM^{-1}$, which is higher than that for TAL$_{sts}$ (k_{cat}/K_M = 1305 $s^{-1}\cdot mM^{-1}$)[13] and around 4-fold of that for the mutated PTAL$_{rgl}$ (k_{cat}/K_M = 391 $s^{-1}\cdot mM^{-1}$).[32] In addition, we also determined the K_i values for the three novel enzymes (Table 1). TAL$_{rpc}$ and PTAL$_{cma}$ showed relative low K_i values and underlie therefore relative strong product inhibition, whereas the TAL$_{clu}$ is less inhibited by *p*-CA.

Biotransformation of L-tyrosine to *p*-CA by TALs and PTALs

To evaluate the synthetic capability of TAL$_{clu}$, TAL$_{rpc}$, and PTAL$_{cma}$, we carried out the biotransformation to *p*-CA from L-tyrosine on a 1.5 mL scale with 0.1 mg enzymes. The reactions took place under optimal reaction conditions according to pH and temperature as well as thermostability of these three enzymes. The results are shown in Figure 6a. All conversions reached their maximum within 17 hours. TAL$_{clu}$ exhibited the highest yield of 27% (1.38 g/L), while the others presented 22% (1.11 g/L) for TAL$_{rpc}$ and 18% (0.94 g/L) for PTAL$_{cma}$.

Furthermore, we evaluated the biotransformation by TAL$_{clu}$ and TAL$_{rpc}$ in *E. coli* whole cell systems. The conversion was carried out under several pH conditions in 15 mL falcons to avoid evaporation. Interestingly, there was no considerable difference in the product yield at different pH (Figures 6b and 6c). The highest transformation rate was even obtained at pH 7. This might be because the constant pH environment in the cell resists changes in external pH, resulting in the catalytic efficiency of the whole cell remaining constant. Meanwhile, the neutral pH was also conducive to cell stability and provided stable conditions for intracellular enzymes. The highest *p*-CA yield was 1.66 ± 0.020 g/L within 72 h by the cells expressing TAL$_{clu}$, whereas the cells expressing the TAL$_{rpc}$ reached the highest yield after 24 h with 1.48 ± 0.031 g/L.

Finally, we scaled-up the whole cell reaction system up to 40 mL at the optimal conditions (Figure 6d). The results showed that the highest yield of 2.35 g/L was reached in 24 h by TAL$_{rpc}$ in *E. coli* whole cells with a yield of 46% and a space-time yield of 0.097 ($g\cdot L^{-1}\cdot h^{-1}$). The biotransformation with TAL$_{clu}$ in whole cells led to a *p*-CA yield of 2.03 g/L within 8 h and a space-time yield of 0.25 ($g\cdot L^{-1}\cdot h^{-1}$). The application of TALs with strict

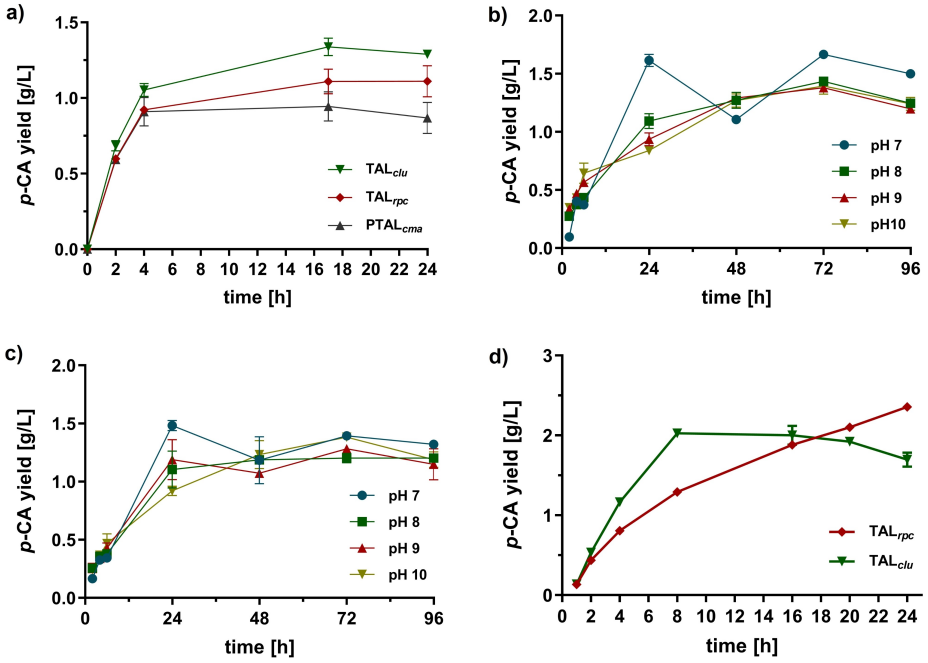


Figure 6. Biotransformation of L-tyrosine to *p*-CA by TAL$_{clu}$, TAL$_{rpc}$ and PTAL$_{cma}$. (a) Biotransformation to *p*-CA by purified enzymes. Reaction system: 10 mM glycine-NaOH (pH 9.5 for TAL$_{clu}$ and PTAL$_{cma}$, pH 8.5 for TAL$_{rpc}$) and 30 mM L-tyrosine at 65 °C for TAL$_{clu}$ and TAL$_{rpc}$, 40 °C for PTAL$_{cma}$. The reactions were stopped by adding 50 µL 3 M HCl. The amount of produced *p*-CA was measured by HPLC. (b) pH optimization of the biotransformation by TAL$_{clu}$ in the whole cell system. Reaction system: 30 mM L-tyrosine in 10 mM PBS for pH 7 and 8 or 30 mM L-tyrosine in 10 mM glycine-NaOH for pH 9 and 10 at 50 °C. (c) pH optimization of the biotransformation by TAL$_{rpc}$ in the whole cell system. (d) Biotransformation to *p*-CA under the optimized conditions in the whole cell system. Reaction system: 30 mM L-tyrosine in 10 mM PBS (pH 7) at 50 °C.

substrate specificity to L-tyrosine as biocatalysts in whole cell factories helps to produce *p*-CA with high purity and avoids the generation of cinnamic acid by-products caused by the side activity of PTALs.

Conclusion

We conducted a phylogenetic tree-building approach to get an overview of putative TALs and PTALs, and from these identified novel biocatalysts for the synthesis of *p*-CA. The obtained phylogenetic trees comprised 875 putative TAL and 46 putative PTAL enzymes, covering most putative TALs from a variety of bacteria and a small number of plants and fungi, as well as putative PTALs from Basidiomycota and Ascomycota. The activity screening identified 5 novel TALs and 3 novel PTALs. Among them, TAL$_{clu}$ and TAL$_{rpc}$ presented excellent thermostability with the highest activity of the novel TALs. By using novel TALs in *E. coli* whole cells, the maximum conversion from L-tyrosine to *p*-CA reached a yield of 2.03 g/L after 8 hours by TAL$_{clu}$ and 2.35 g/L after 24 hours by TAL$_{rpc}$. These results provide more valuable biocatalysts for the construction of cell factories to produce the flavonoid precursor *p*-CA.

Experimental Section

Building of phylogenetic trees for TALs and PTALs: The search of homologous protein sequences was carried out with a BLAST algorithm. The protein sequences of TAL$_{rsp}$ (UniProtKB accession no.: Q3IWB0), PTAL$_{rgl}$ (UniProtKB accession no.: U5TV35), and PAL$_{pcr}$ (UniProtKB accession no.: P24481) were used as search queries. The non-redundant protein sequences of UniParc were used as database for searching. Around 16,000 sequences were identified. The sequences that were too long (>800 amino acids) or too short (<400 amino acids) in comparison to TAL$_{rsp}$ were excluded from the sequence list. The sequences with a similarity of more than 90% were sorted out by CD-HIT[33] to remove multiple annotated sequences. The resulting sequences were aligned using MUSCLE.[23] The sequences with the key amino acid residues were chosen for the final alignment and carefully optimized to obtain an accurate alignment. The command line tool IQ-TREE (http://www.iqtree.org) was used to compute a phylogenetic tree from a multisequence alignment. The default parameters were used for the tree construction. The obtained phylogenetic tree was rooted to midpoint with the FigTree software (http://tree.bio.ed.ac.uk/software/figtree) to ensure a better comparability.

Protein structure analysis of novel TALs and PTALs: The protein sequences were submitted to the SWISS-MODEL[34] for protein structure simulation. The obtained protein structures were aligned by PyMOL software (Schrödinger, USA) to identify the overall structure similarity and active site pockets.

Expression and purification of putative TALs and PTALs: The synthetic genes were inserted into the vector pET15b with NdeI and BamHI restriction enzymes at the 5′- and 3′-terminii, respectively. The obtained plasmids were transformed into *E. coli* BL21(DE3) chemical competent cells by heat shock. The recombi-

nant *E. coli* cells were cultured in 1 L LB medium at 37 °C until OD reached 0.5 and induced by adding IPTG (0.5 mM) for protein expression at 20 °C for 20 h. The cells were harvested by centrifugation at 4,000×*g* for 30 min. The harvested cells were lysed by sonication. The crude extract was centrifuged at 10,000 × *g* for 30 min to remove the pellet, while the supernatant was used for further purification. Recombinant proteins were purified by affinity chromatography with a HisTrap HP 5 mL column (GE Healthcare). The target proteins were eluted by 200 mM imidazole with 20 mM PBS (pH 7.4) and 500 mM NaCl. Buffer exchange with 50 mM PBS (pH 7.5) was achieved by ultrafiltration with Amicon Ultra-15 (Merck). The concentration of proteins was measured using a NanoDrop device (ThermoFisher).

UV-assay for TALs and PTALs: The reactions were performed in UV-transparent-96-well plates at 25 °C (200 μL per well). The measurement was started by adding 10 μL enzyme solution (1 mg/mL) to 190 μL substrate solution which contained 20 mM PBS-buffer (pH 7.5) with either 2 mM L-tyrosine or 10 mM L-phenylalanine. For phenylalanine the increase of absorbance at 290 nm (production of cinnamic acid) was measured, while in the assays with L-tyrosine the absorbance at 310 nm (production of *p*-CA) was measured. 1 U of activity is defined as the amount of enzyme required to convert 1 μmol L-tyrosine or L-phenylalanine to cinnamic acid or *p*-CA under the assay conditions.

pH assay for TALs and PTALs: For the pH-assay, the activity of TALs and PTALs towards L-tyrosine was measured at pH 6–10.5. For this, 10 μL enzyme solution (1 mg/mL) were mixed with 190 μL substrate solution in 50 mM PBS buffer with 2 mM L-tyrosine (for the pH range of 6.5–8) or 50 mM borate buffer with 2 mM L-tyrosine (for the pH range from 8.5–10.5). The formation of *p*-CA was measured at 310 nm by the TECAN plate reader.

Temperature assay for TALs and PTALs: For the temperature assay, the activity of the enzymes with L-tyrosine was measured at the temperature range of 20–70 °C. For the temperatures between 20 and 70 °C, the incubation was performed in the BioTek Synergy H1 plate reader. The *p*-CA production was measured, in triplicate, by absorption (310 nm) for 10 min in a BioTek Synergy H1 plate reader.

Determination of thermostability for TALs and PTALs: The melting point of the enzymes was measured with the nanoDSF device. For the measurements, the enzymes were diluted in 50 mM PBS buffer (pH 7.4) at a concentration of 1 mg/mL and evaluated in the temperature range of 20–95 °C (increase of 1 °C/min).

Determination of kinetic contents: The kinetic contents were measured at various concentration of L-tyrosine (2, 1, 0.5, 0.25, 0.125, 0.075, 0.03750, and 0.01875 mM) with the UV-assay described above.

Determination of *p*-CA concentration via HPLC: Determination of the *p*-CA concentration was done by HPLC (Agilent) using EC-C18 Poroshell 120 (100 mm ×3 mm) and a diode array detector. The samples were eluted with 10% (v/v) acetonitrile-water, and the product was detected at 310 nm. The retention time of *p*-CA was 4.5 min. To determine the *p*-CA concentration, a standard curve was constructed with *p*-CA concentrations from 1 to 15 mm.

Biotransformation with purified enzymes: Four enzymes (TAL_{clu}, TAL_{rpc}, $PTAL_{poc}$, and $PTAL_{cma}$) were expressed and purified using the methods described above. Then 0.1 mg of the enzymes were added to 1.5 mL reaction solution (10 mM glycine-NaOH and 30 mM L-tyrosine). The reactions were performed at the optimal pH of the enzymes (pH 9.5 for TAL_{clu} and $PTAL_{cma}$, pH 8.5 for TAL_{rpc}, and pH 9 for $PTAL_{poc}$) and the two temperatures (TAL_{clu}: 65 °C; TAL_{rpc}: 65 °C; and $PTAL_{cma}$: 40 °C). The biotransformations were started by adding 0.1 mg of the freshly purified enzyme and then monitored for 72 h. For every time point, 200 μL samples were taken. 50 μL 3 M HCl were added to stop the reaction. The samples were then centrifuged (13,000 *g* ×30 min), and the supernatant was filtered (0.22 μm). The obtained samples were analyzed by the HPLC method.

Whole cell biotransformation to produce *p*-CA: The enzymes were expressed as described above. Subsequently, the *E. coli* cells were harvested and washed two times with distilled water. The biotransformation was started by adding the cells to the reaction system to a final OD_{600} of 10. The reaction system was composed of 10 mM glycine-NaOH buffer containing 30 mM L-tyrosine. For every time point, 200 μL samples were taken and 50 μL 3 M HCl were added to stop the reaction. Then the samples were centrifuged (13.000 *g*×30 min), and the supernatant was filtered (0.22 μm). The samples were analyzed by the HPLC method.

Acknowledgements

This work was supported by the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement no. 814650 for the project SynBio4Flavy. The authors are thankful to Ina Menyes for her support in analytics. Open Access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available in the supplementary material of this article.

Keywords: biocatalysis · flavonoids · *p*-coumaric acid · phenylalanine ammonia lyase · tyrosine ammonia lyase

[1] P. G. Pietta, *J. Nat. Prod.* **2000**, *63*, 1035–1042.
[2] S. J. Maleki, J. F. Crespo, B. Cabanillas, *Food Chem.* **2019**, *299*, 125124.
[3] P. W. Snijman, S. Swanevelder, E. Joubert, I. R. Green, W. C. A. Gelderblom, *Mutat. Res. Genet. Toxicol. Environ. Mutagen.* **2007**, *631*, 111–123.
[4] A. N. Panche, A. D. Diwan, S. R. Chandra, *J. Nutr. Sci.* **2016**, *5*, e47.
[5] A. Vargas-Tah, L. M. Martínez, G. Hernández-Chávez, M. Rocha, A. Martínez, F. Bolívar, G. Gosset, *Microb. Cell Fact.* **2015**, *14*, 6.
[6] A. M. Reyes Jara, M. E. Gómez Lobato, P. M. Civello, G. A. Martínez, *J. Food Biochem.* **2022**, *46*, e14054.
[7] Y. Zhang, Z. Hu, M. Zhu, Z. Zhu, Z. Wang, S. Tian, G. Chen, *J. Agric. Food Chem.* **2015**, *63*, 4160–4169.
[8] Y. Li, J. Li, B. Qian, L. Cheng, S. Xu, R. Wang, *Molecules* **2018**, *23*, 3185.
[9] J. Barros, R. A. Dixon, *Trends Plant Sci.* **2020**, *25*, 66–79.
[10] J. A. Kyndt, T. E. Meyer, M. A. Cusanovich, J. J. Van Beeumen, *FEBS Lett.* **2002**, *512*, 240–244.
[11] K. T. Watts, B. N. Mijts, P. C. Lee, A. J. Manning, C. Schmidt-Dannert, *Chem. Biol.* **2006**, *13*, 1317–1326.
[12] C. B. Jendresen, S. G. Stahlhut, M. Li, P. Gaspar, S. Siedler, J. Forster, J. Maury, I. Borodina, A. T. Nielsen, *Appl. Environ. Microbiol.* **2015**, *81*, 4458–4476.
[13] P. Cui, W. Zhong, Y. Qin, F. Tao, W. Wang, J. Zhan, *Bioprocess Biosyst. Eng.* **2020**, *43*, 1287–1298.

[14] M. Berner, D. Krug, C. Bihlmaier, A. Vente, R. Müller, A. Bechthold, *J. Bacteriol.* **2006**, *188*, 2666–2673.
[15] L. Zhu, W. Cui, Y. Fang, Y. Liu, X. Gao, Z. Zhou, *Biotechnol. Lett.* **2013**, *35*, 751–756.
[16] T. A. Schöner, S. W. Fuchs, C. Schönau, H. B. Bode, *Microb. Biotechnol.* **2014**, *7*, 232–241.
[17] L. Xiang, B. S. Moore, *J. Bacteriol.* **2005**, *187*, 4286–4289.
[18] J. Rosler, F. Krekel, N. Amrhein, J. Schmid, *Plant Physiol.* **1997**, *113*, 175–179.
[19] Z. Xue, M. McCluskey, K. Cantera, F. S. Sariaslani, L. Huang, *J. Ind. Microbiol. Biotechnol.* **2007**, *34*, 599–604.
[20] F. S. Sariaslani, *Annu. Rev. Microbiol.* **2007**, *61*, 51–69.
[21] C. W. Abell, R.-S. Shen, in *Methods Enzymology*, Academic Press, **1987**, pp. 242–248.
[22] H. Meinert, D. Yi, B. Zirpel, E. Schuiten, T. Geißler, E. Gross, S. I. Brückner, B. Hartmann, C. Röttger, J. P. Ley, U. T. Bornscheuer, *Angew. Chem. Int. Ed.* **2021**, *60*, 16874–16879; *Angew. Chem.* **2021**, *133*, 17011–17016.
[23] R. C. Edgar, *Nucleic Acids Res.* **2004**, *32*, 1792–1797.
[24] M. Baedeker, G. E. Schulz, *Structure* **2002**, *10*, 61–67.
[25] B. Wu, W. Szymański, H. J. Wijma, C. G. Crismaru, S. de Wildeman, G. J. Poelarends, B. L. Feringa, D. B. Janssen, *Chem. Commun.* **2010**, *46*, 8157–8159.
[26] S. Bartsch, U. T. Bornscheuer, *Protein Eng. Des. Sel.* **2010**, *23*, 929–933.
[27] D. Krug, R. Müller, *ChemBioChem* **2009**, *10*, 741–750.
[28] M. M. Heberling, M. F. Masman, S. Bartsch, G. G. Wybenga, B. W. Dijkstra, S. J. Marrink, D. B. Janssen, *ACS Chem. Biol.* **2015**, *10*, 989–997.
[29] Y. Zhu, S. Liao, J. Ye, H. Zhang, *Biotechnol. Lett.* **2012**, *34*, 269–274.
[30] Z. Xue, M. McCluskey, K. Cantera, A. Ben-Bassat, F. S. Sariaslani, L. Huang, *Enzyme Microb. Technol.* **2007**, *42*, 58–64.
[31] S. D. Tork, E. Z. A. Nagy, L. Cserepes, D. M. Bordea, B. Nagy, M. I. Toşa, C. Paizs, L. C. Bencze, *Sci. Rep.* **2019**, *9*, 20123.
[32] S. Zhou, P. Liu, J. Chen, G. Du, H. Li, J. Zhou, *Appl. Microbiol. Biotechnol.* **2016**, *100*, 10443–10452.
[33] Y. Huang, B. Niu, Y. Gao, L. Fu, W. Li, *Bioinformatics* **2010**, *26*, 680–682.
[34] T. Schwede, J. Kopp, N. Guex, M. C. Peitsch, *Nucleic Acids Res.* **2003**, *31*, 3381–3385.

Manuscript received: January 28, 2022
Revised manuscript received: March 29, 2022
Accepted manuscript online: March 30, 2022