Hydroxycinnamoyltransferase and CYP98 in phenolic metabolism in the rosmarinic acid-producing hornwort *Anthoceros agrestis*

Lucien Ernst¹ · Julia Wohl¹ · Elke Bauerbach¹ · Maike Petersen¹

Received: 4 November 2021 / Accepted: 12 February 2022 / Published online: 2 March 2022
 © The Author(s) 2022

Abstract

Main conclusion *Anthoceros agrestis* hydroxycinnamoyltransferase accepts shikimic and 3-hydroxyanthranilic acids while hydroxycinnamoyl ester/amide 3-hydroxylase (CYP98A147) preferred *p*-coumaroyl-(3-hydroxy)anthranilic acid compared to the shikimic acid derivative. Alternative pathways towards rosmarinic acid have to be considered.

Abstract Rosmarinic acid (RA) is a well-known ester of caffeic acid and 3,4-dihydroxyphenyllactic acid. In the search for enzymes involved in RA biosynthesis in the hornwort *Anthoceros agrestis*, the hydroxycinnamoyltransferase sequence with the highest similarity to rosmarinic acid synthase from Lamiaceae has been amplified and heterologously expressed in *Escherichia coli*. In parallel, the single cytochrome P450 sequence belonging to the CYP98 group in *Anthoceros agrestis* was isolated and expressed in *Saccharomyces cerevisiae* which did not result in protein formation. Codon optimization and co-expression with NADPH:cytochrome P450 reductase (CPR) from *Coleus blumei* resulted in the formation of active enzymes. Both, the hydroxycinnamoyltransferase and CYP98 were characterized with respect to their temperature and pH optimum as well as their substrate acceptance. The hydroxycinnamoyltransferase (AaHCT6) readily accepted *p*-coumaroyl- and caffeoyl-CoA with a slightly higher affinity towards *p*-coumaroyl-CoA. The best acceptor substrate was shikimic acid (K_m 25 μ M with *p*-coumaroyl-CoA) followed by 3-hydroxyanthranilic acid (K_m 153 μ M with *p*-coumaroyl-CoA). Another accepted substrate was 2,3-dihydroxybenzoic acid. Anthranilic acid and 4-hydroxyphenyllactic acid (as precursor for RA) were not used as substrates. *p*-Coumaroyl esters and -amides are substrates hydroxylated by CYP98 hydroxylases. The only CYP98 sequence from *Anthoceros agrestis* is CYP98A147. The best substrates for the NADPH-dependent hydroxylation were *p*-coumaroyl anthranilic and *p*-coumaroyl-3-hydroxyanthranilic acids while *p*-coumaroyl shikimic and *p*-coumaroyl-4-hydroxyphenyllactic acids were poor substrates. The biosynthetic pathway towards rosmarinic acid thus still remains open and other enzyme classes as well as an earlier introduction of the 3-hydroxyl group to afford the caffeic acid substitution pattern must be taken into consideration.

Keywords Anthocerotaceae · BAHD acyltransferases · Evolution · Hydroxycinnamic acids · Hydroxycinnamoyl ester/amide 3-hydroxylase · Model organism · Phenylpropanoid derivatives · Phylogeny

Abbreviations

BAHD Acyltransferase family named according to the first four identified enzymes

CPR	NADPH:cytochrome P450 reductase
HCT	Hydroxycinnamoyltransferase
HST	Hydroxycinnamoyl-CoA:shikimic acid hydroxycinnamoyltransferase
RA	Rosmarinic acid
RAS	Rosmarinic acid synthase = hydroxycinnamoyl-CoA:hydroxyphenyllactic acid hydroxycinnamoyltransferase

Communicated by Dorothea Bartels.

Lucien Ernst and Julia Wohl have contributed equally to this work.

✉ Maike Petersen
 petersen@staff.uni-marburg.de

¹ Institut für Pharmazeutische Biologie und Biotechnologie, Philipps-Universität Marburg, Robert-Koch-Str. 4, 35037 Marburg, Germany

Introduction

The model plant *Anthoceros agrestis* PATON (Anthocerotaceae) belonging to the hornworts is known for its accumulation of various hydroxycinnamic acid derivatives, such as rosmarinic acid (RA) and (hydroxy)megacerotonic acid or the alkaloid anthocerodiazonin (for structures see Suppl. Fig. S1) (Takeda et al. 1990a; Trennheuser et al. 1994) as well as numerous other hydroxycinnamic and benzoic acid derivatives (Trennheuser 1992). As such a wealth of phenolic compounds has rarely been described in mosses and liverworts, investigations into the biochemical machinery of *Anthoceros* and the encoding genes are interesting also with regard to its evolutionary development. RA is better known as a typical compound of members of the seed plant families Lamiaceae (subfamily Nepetoideae) and Boraginaceae, although its wide occurrence in other taxa of the seed plants and in ferns has been documented (Petersen et al. 2009). Its biosynthesis has first been established in a member of the Lamiaceae, namely *Coleus blumei* (syn. *Plectranthus scutellarioides*, *Solenostemon scutellarioides*) (Petersen et al. 1993), and verified in other plant species (Ma et al. 2015).

The biosynthesis of the above-mentioned hydroxycinnamic acid derivatives requires the presence of the general phenylpropanoid pathway to afford activated hydroxycinnamic acids. Phenylalanine ammonia-lyase (PAL), cinnamic acid 4-hydroxylase (C4H) as well as *p*-coumarate:CoA-ligases (4CL) have previously been detected and characterized from this hornwort species on molecular and biochemical levels (Petersen 2003; Pezeshki 2016; Wohl and Petersen 2020a, b) showing the conservation of the general phenylpropanoid pathway also in non-seed plants as already proposed previously (Labeeuw et al. 2015; de Vries et al. 2017; Renault et al. 2019). Furthermore, in all the above-mentioned compounds (except anthocerodiazonin), a hydroxyphenyllactic acid moiety is incorporated which is derived from L-tyrosine by tyrosine aminotransferase (TAT) and hydroxyphenylpyruvic acid reductase (HPPR), two enzymes also present in *Anthoceros agrestis* and currently under investigation in our laboratory (Busch and Petersen (2021) and own unpublished results). The further biosynthesis of RA (Fig. 1) (and putatively also of anthocerotonic and megacerotonic acids) requires a hydroxycinnamoyl-transferase to link the two precursors, 4-coumaroyl-CoA and 4-hydroxyphenyllactic acid, and a hydroxylase to finalize the substitution patterns at the aromatic rings as described for *Coleus blumei* (Lamiaceae) by Berger et al. (2006) and Eberle et al. (2009) or for *Phacelia campanularia* (Boraginaceae) by Levsh et al. (2019).

The first appearance of true phenolic compounds in the plant kingdom has been related to the formation of

cuticle-like structures and thus to the protection of the land-colonizing plants against desiccation. Therefore, the phenylpropanoid pathway has been regarded as essential for plant terrestrialization, not only to protect against drought but also to provide a screen against UV radiation and to protect against pathogens and herbivores as well as to enable vertical growth by providing lignin for cell wall reinforcement (Weng and Chapple 2010; de Vries et al. 2021). In *Physcomitrium patens*, the formation of caffeoylthreonic acid from *p*-coumaroylthreonic acid by CYP98A34 (the only member of the CYP98 group in this plant species) has been reported to be essential for the formation of a phenolic-containing cuticle while shikimic or quinic esters were not detected in first investigations (Renault et al. 2017). However, the CYP98A34 hydroxylating *p*-coumaroylthreonic acid was later found to prefer *p*-coumaroylanthranilic acid as substrate, while *p*-coumaroylshikimic acid was less accepted and *p*-coumaroyl-2-threonic and *p*-coumaroyl-4-threonic acids were only poorly converted. Similar results were shown for CYP98A38 from the lycopod *Selaginella moellendorffii* and CYP98A104 from the fern *Pteris vittata* (Alber et al. 2019). These observations made a pathway to caffeic acid derivatives leading through phenolamides probable, while in seed plants, generally *p*-coumaroylshikimic acid was shown to be the responsible intermediate in the monolignol biosynthetic pathway (Schoch et al. 2001). The missing ability of *Physcomitrium* CYP98A34 to complement a CYP98A3 knockout mutant of *Arabidopsis thaliana* furthermore made an in vivo formation of caffeoylshikimic acid by the moss cytochrome P450 improbable (Alber et al. 2019). Findings by Kriegshauser et al. (2021), however, showed that the main intermediates formed by *Physcomitrium patens* hydroxycinnamoyltransferase *Pp3c2_29140* are hydroxycinnamoyl shikimic acid esters and not amides and thus a pathway similar to the one in seed plants was suggested (see Fig. 1). In seed plants, the main pathway towards higher substituted hydroxycinnamic acid derivatives (e.g., caffeic, ferulic or sinapic acids) and monolignols leads via a *p*-coumaric acid ester with shikimic acid (Hoffmann et al. 2003) which is then hydroxylated by a member of the CYP98 family to caffeoylshikimic acid (Schoch et al. 2001) after which caffeic acid (Vanholme et al. 2013; Saleme et al. 2017) or caffeoyl-CoA is released (Hoffmann et al. 2003) for further biosyntheses. Silencing of caffeoylshikimic acid esterase has been shown to alter lignification in poplar and to lead to an accumulation of caffeoylshikimic acid (Vanholme et al. 2013; Saleme et al. 2017). Reinvestigation of the substrate specificities of *Physcomitrium patens* hydroxy-cinnamoyltransferase (encoded by *Pp3c2_29140*) and CYP98A34 finally led to the conclusion that also mosses synthesize their caffeic acid metabolites via caffeoylshikimic acid (Kriegshauser et al. 2021).

Fig. 1 Biosynthetic pathways leading to rosmarinic acid and caffeoylshikimic acid. Abbreviations of enzyme names are as follows: PAL L-phenylalanine ammonia-lyase, C4H cinnamic acid 4-hydroxylase, 4CL 4-coumaric acid coenzyme A ligase, TAT L-tyrosine aminotransferase, HPPR hydroxyphenylpyruvic acid reductase, RAS “rosmarinic acid synthase” = hydroxycinnamoyl-CoA:hydroxyphenyllactic acid

hydroxycinnamoyltransferase, 3-H, 3'-H 3- and 3'-hydroxylase(s), AaHCT6 4-coumaroyl-CoA:shikimic acid hydroxycinnamoyltransferase from *Anthoceros agrestis*, AaCYP98 4-coumaroylshikimic acid 3-hydroxylase from *Anthoceros agrestis*. Full arrows show reactions which have been characterized in *Anthoceros agrestis*, dashed arrows have not yet been characterized

Hydroxycinnamoyltransferases of the BAHD superfamily are known for their sometimes astonishing substrate promiscuity (Landmann et al. 2011; Sander and Petersen 2011) and some structural aspects important for substrate selection have been investigated (Chiang et al. 2018). It is assumed, that gene duplication and neo-functionalization events lead to the extended numbers of hydroxycinnamoyltransferases in angiosperms (Tuominen et al. 2011) with varying degrees of substrate specificity. While many hydroxycinnamoyl-CoA:shikimic/quinic acid hydroxycinnamoyltransferases accept shikimic as well as quinic acid (Schoch et al. 2001; Mahesh et al. 2007; Moglia et al. 2009), the hydroxycinnamoyl-CoA:shikimic acid hydroxycinnamoyltransferase (CbHST) from *Coleus blumei* accepts shikimic acid but not quinic and hydroxyphenyllactic acids, while hydroxycinnamoyl-CoA:hydroxyphenyllactic acid hydroxycinnamoyltransferase (better known as RA synthase (RAS); CbRAS)—involved in RA biosynthesis—from the same species accepts hydroxyphenyllactic acid derivatives and does not catalyze esterification of shikimic or quinic acids. However, also other acceptor substrates were accepted, e.g., 3-hydroxyanthranilic acid (but not anthranilic acid), 3-hydroxybenzoic, 2,3-dihydroxybenzoic and 3-aminobenzoic acids by CbHST and D-DOPA, D-tyrosine and D-phenylalanine by CbRAS (Sander and Petersen 2011). Comparable substrate promiscuities have been reported for RAS from *Lavandula angustifolia* by Landmann et al. (2011) and for three partially specific hydroxycinnamoyltransferases from *Phacelia campanularia* (Levsh et al. 2019) as well as the hydroxycinnamoyltransferase encoded by *Pp3c2_29140* from *Physcomitrium patens* (Kriegshauser et al. 2021). Notably, these hydroxycinnamoyltransferases catalyze ester as well as amide formation.

The finalization of the substitution pattern of *p*-coumaric acid esters and amides is catalyzed by cytochrome P450 monooxygenases of the CYP98 family. As reported by Alber et al. (2019), only one copy of CYP98 is present in bryophytes, lycophytes, monilophytes and gymnosperms, but multiple CYP98 family members evolved in angiosperms with varying substrate acceptance spectra. The single copy of CYP98 in the above-mentioned taxa may not in all cases be related to the biosynthesis of lignin since lignin is missing in bryophytes, and the lycophyte *Selaginella* uses a different cytochrome P450 (CYP788A1) for the formation of monolignols (Weng et al. 2010; Alber et al. 2019). It seems that CYP98 monooxygenases can accept a wide range of *p*-coumaric acid esters or amides, thus mainly showing specificity towards the hydroxycinnamoyl moiety, while, however, *p*-coumaric acid itself is generally no or only a very poor substrate (Alber et al. 2019). It must be noted, however, that a broad substrate range has only been tested for very few of these hydroxylases.

We here report on a hydroxycinnamoyltransferase from the hornwort *Anthoceros agrestis* forming esters between a hydroxycinnamic acid and shikimic acid or 2,3-dihydroxybenzoic acid as well as an amide between the hydroxycinnamic acid and 3-hydroxyanthranilic acid. The second enzyme reported here is the single CYP98 monooxygenase from the same organism introducing the 3-hydroxyl group into several *p*-coumaric acid esters and amides.

Materials and methods

Plant cell cultures

Anthoceros agrestis PATON cell suspension cultures originated from a callus culture from Binding and Mordhorst (1991) and were cultivated as described previously (Petersen 2003).

Preparation of cDNA and amplification of partial sequences of AaHCT6 and AaCYP98

Isolation of RNA was performed according to Chomczynski and Sacchi (1987). After checking RNA integrity, cDNA was prepared with the RevertAid First Strand cDNA Synthesis Kit (ThermoFisher). Internal PCR primers were designed based on *Anthoceros punctatus* scaffold 3426 (AaHCT6) and on *Anthoceros agrestis* scaffold 26091 (AaCYP98) after searching the *Anthoceros punctatus* and *Anthoceros agrestis* genome databases with the rosmarinic acid synthase (ENA: AM283092) and the CYP98A14 (ENA: AJ427452) sequences from *Coleus blumei*, respectively (Steven Kelly, University of Oxford, UK, personal communication; Peter Szövényi, Universität Zürich, Switzerland, personal communication; Szövényi et al. 2015). All primers were synthesized by Eurofins Genomics and are listed in Suppl. Table S1.

For the amplification of partial sequences encoding AaCYP98 and AaHCT6, PCR assays of 25 µl were conducted with up to 0.2 µg cDNA, 0.5 µl 10 mM dNTP mix, 0.5 µl of each primer (Ap3426_f + Ap3426_r 100 µM each; Aa26091_f + Aa26091_r 10 µM each), 3.0 µl 25 mM MgCl₂, 5.0 µl 5 × GoTaq buffer and 0.1 µl GoTaq polymerase (5 U/µl, Promega). The following programs were used for AaCYP98 (a) or AaHCT6 (b): 1 cycle 94 °C 120 s, 52 °C 60 s (a) / 60 °C 30 s (b), 70 °C 90 s; 38 cycles 94 °C 30 s, 52 °C 60 s (a) / 60 °C 30 s (b), 70 °C 90 s; 1 cycle 94 °C 120 s, 52 °C 60 s (a) / 60 °C 30 s (b), 70 °C 600 s. Amplification of the target sequences was checked by gel electrophoresis on a 0.7% agarose gel in TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA) using the GeneRuler DNA Ladder Mix (ThermoFisher) as marker. The PCR products were isolated with the NucleoSpin Gel and PCR Clean-up Kit (Macherey–Nagel). After ligation

into pDrive (Qiagen), transformation and multiplication in *E. coli* EZ (Qiagen), the sequences were determined (Microsynth SeqLab).

RACE-PCR and amplification of full-length AaHCT6 and AaCYP98 sequences

According to the sequences determined in the previous step, rapid amplification of cDNA ends (RACE)-PCR primers (Aa3426_5'R, Aa3426_3'R; Aa26091_5'R, Aa26091_3'R; Suppl. Table S1) was designed. RACE-PCR cDNA synthesis, PCR and InFusion®-cloning (only for AaCYP98) were performed using the SMARTer® RACE 5'/3' kit (Takara/Clontech). After purification of the PCR products (NucleoSpin Gel and PCR Clean-up Kit, Macherey–Nagel), insertion into pRACE (AaCYP98) or ligation into pDrive (for AaHCT6), *E. coli* EZ were transformed and grown overnight. The plasmid was isolated and the sequence determined (Microsynth SeqLab). For amplification of the full-length sequence of AaHCT6, primers with restriction sites for NdeI (Aa3426_VL-f) in the forward and BamHI (Aa3426_VL-r) in the reverse primer were designed for integration into the expression vector pET-15b (Novagen). For amplification and integration of the full-length sequence of AaCYP98 into the expression vector pESC-URA (Agilent Technologies), the sequence was equipped with restriction sites for EcoRI (5') and NotI (3') and the stop codon was removed. PCR reactions of 25 µl were performed as above, but using Pfu-Polymerase® (2 U/µl; Promega; AaCYP98) or Phusion polymerase (NEB) and the respective buffers using the same PCR program as above, but with annealing temperatures of 64 °C for AaHCT6 and 56 °C for AaCYP98. After ligation of the purified PCR product into pDrive, *E. coli* EZ cells were transformed, grown in LB medium with 100 µg/ml ampicillin overnight at 37 °C and the sequence was determined (Microsynth SeqLab) after plasmid isolation. The AaHCT6 (= Aa3426) full-length sequence (GenBank MW248389) was released from pDrive by restriction digest with NdeI and BamHI and ligated into the vector pET-15b into the same restriction sites. The AaCYP98 (= Aa26091) full-length sequence (GenBank MT119883) was cut from the cloning plasmid with EcoRI and NotI and integrated into the respective restriction sites of the multiple cloning site I (MSCI) of the vector pESC-URA. By integration, AaHCT6 was attached to the sequence encoding an *N*-terminal 6xHis-tag already present on the expression vector pET-15b and AaCYP98 was attached to the sequence encoding a *C*-terminal FLAG-tag of the expression vector pESC-URA. The sequences were again checked for correctness (Microsynth SeqLab) after transformation (*E. coli* EZ), growth in LB medium with 100 µg/ml ampicillin at 37 °C overnight and plasmid isolation.

Construction of phylogenetic trees

For phylogenetic analysis, the translated amino acid sequences of AaHCT6 or AaCYP98 were aligned with other amino acid sequences from different species using the Maximum Likelihood method of the MEGA7 software package. The robustness of the branch structure was evaluated by bootstrap analysis (1000 replicates). The sequences for the phylogenetic trees were accessed from the BRENDA enzyme database, NCBI searches and Alber et al. (2019) (see Suppl. Figs. S2 and S3 for accession numbers).

Protein expression and isolation of AaHCT6

For expression of AaHCT6, *E. coli* SoluBL21 (Amsbio) bacteria were transformed with the pET-15b vector carrying the full-length sequence of AaHCT6. The bacteria were cultivated overnight in 2 ml LB-medium containing 100 µg/ml ampicillin at 37 °C, 220 rpm and then transferred into a 250 ml Erlenmeyer flask containing 100 ml TB-medium with 100 µg/ml ampicillin (for media composition see Lesard 2013). The bacteria were cultivated at 220 rpm and 37 °C to an optical density at 600 nm (OD₆₀₀) of 0.4–0.6. Expression was then induced by addition of isopropyl-β-D-1-thiogalactopyranoside (final concentration 1 mM). The cultures were incubated for 24 h at 180 rpm and 25 °C. Bacteria were harvested by centrifugation for 15 min at 8000g, re-suspended in 4 ml per g fresh weight potassium phosphate buffer (KP_i 0.1 M, pH 8.0) and incubated on ice with 50 mg lysozyme for 30 min. The bacteria were then disrupted by ultra-sonication for 1 min at 0 °C and centrifuged for 15 min at 8000g. The supernatant was collected as crude protein extract. Purification of the AaHCT6 protein carrying an *N*-terminal 6xHis-tag was attempted by metal-chelate chromatography with Ni–NTA resin (Novagen) followed by desalting through PD-10 columns (GE Healthcare) but resulted in severe activity losses and was therefore not routinely applied.

Codon optimization, protein expression and isolation of AaCYP98

At first, *Saccharomyces cerevisiae* BY4742 (ATCC 201389) was transformed with the native AaCYP98 coding sequence in pESC-URA, but after expression, AaCYP98 protein was not detectable by Western blot analysis. For this reason, we decided to codon-optimize AaCYP98 (= coAaCYP98) for the expression in yeast; this was carried out commercially by GeneCust (Boynes, France). The sequence was equipped with restriction sites for EcoRI (5') and NotI (3') and the stop codon was removed. The codon-optimized sequence was delivered in the vector pUC57 which was used to transform *E. coli* EZ for

plasmid multiplication. The codon-optimized AaCYP98 sequence was released from the vector with EcoRI and NotI and integrated into the suitable restriction sites of the multiple cloning site I (MSCI) of the vector pESC-URA. Additionally, a sequence encoding a NADPH:cytochrome P450 reductase (CPR) from *Coleus blumei* (CbCPR) characterized by Eberle et al. (2009) was ligated into multiple cloning site II (MCSII) into the restriction sites for BamHI and SalI, alternatively MCSII was left empty. *Saccharomyces cerevisiae* BY4742 was transformed with the plasmids containing only coAaCYP98 or coAaCYP98 + CbCPR essentially as described by Gietz and Schiestl (2007) and cultivated on SCD_{-ura} plates (Dymond 2013). Plasmid uptake was checked by colony-PCR: single colonies were picked and suspended in 100 µl 200 mM lithium acetate, 1% SDS solution and incubated for 5 min at 70 °C. For DNA precipitation, 300 µl ethanol (96%) was added and mixed vigorously. The samples were centrifuged for 3 min at 15,000g and the resulting pellet was washed with 300 µl 70% ethanol. Afterwards, the tubes were centrifuged again for 3 min at 15,000g and the pellet was dissolved in 50 µl H₂O. 2 µl of the solution was used in a PCR reaction as described above with GoTaq polymerase and specific primers for the respective multiple cloning sites (MCSI_f + MCSI_r or MCSII_f + MCSII_r; Suppl. Table S1). Amplification was checked by agarose gel electrophoresis. For protein expression, a small amount of transformed yeast cells was transferred from a SCD_{-ura} plate to 100 ml liquid SCD_{-ura} in a 500-ml baffled flask and incubated for 48 h at 30 °C at 160 rpm. Then the cells were transferred to 100 ml SCG_{-ura} medium supplemented with 200 µM 5-aminolevulinic acid and 12.5 mg l⁻¹ FeSO₄ after cell collection by centrifugation for 5 min at 3000g at 4 °C. The cultures were incubated for another 24 h at 30 °C and 160 rpm and the cells then collected by centrifugation for 5 min at 3000 g and 4 °C. The cell pellet was re-dissolved in 10 ml buffer (0.1 M Tris-HCl pH 7.0, 1 mM dithiothreitol (DTT), 1 mM sodium diethyldithiocarbamate (DIECA)) and left on ice for 5 min. To remove the remains of the medium, the cells were again sedimented for 5 min at 3000g (4 °C) and the supernatant was discarded. The pellet was re-suspended in 1.5 ml buffer (as above) and transferred to a 7-ml tube with 2.5 g glass beads (~0.5 mm ø). Disintegration was done by vigorous shaking in a benchtop homogenizer (Minilys®) at 4000 rpm for 30 s, followed by cooling on ice for 30 s and 6–8 repetitions. The supernatant was removed after the glass beads had settled and glass beads were washed with 0.5 ml buffer. The crude protein extract was obtained after the supernatants were pooled and centrifuged for 20 min at 5000g (4 °C). The extract was used directly for the determination of enzyme activities. Samples for SDS-PAGE or Western blot analysis were stored at –20 °C.

SDS-PAGE and western blot

Protein extracts were subjected to SDS-PAGE essentially according to Laemmli (1970). After electrophoresis, the gel was either stained with Coomassie Brilliant Blue R250 or Western blotting was performed basically as specified by Mahmood and Yang (2012), but applying the Towbin et al. (1979) buffer system. The expressed proteins were detected with mouse anti-6xHis-tag (ThermoFisher, MA1-21315) or mouse anti-FLAG-tag (ThermoFisher, MA1-91878) monoclonal antibodies, followed by goat anti-mouse antibodies conjugated to alkaline phosphatase (Life Technologies, A16087) as secondary antibody. Staining was performed with nitro blue tetrazolium chloride (NBT)/5-bromo-4-chloro-3-indolylphosphate (BCIP) essentially following standard protocols (<https://www.sysy.com/protocols/westernblot-ap-detection>). Here the alkaline phosphatase coupled to the secondary antibody dephosphorylates BCIP followed by oxidation and indoxyl dimerization to dibromodichloroindigo. Additionally, reduction equivalents reduce NBT to its formazan. Both insoluble pigments can be detected by their purple coloration (Sambrook and Russell 2001).

Production of substrates for AaHCT6 and AaCYP98

Chemical synthesis of *p*-coumaroyl-CoA and caffeoyl-CoA

The synthesis of hydroxycinnamoyl-CoA thioesters followed essentially the protocol of Stöckigt and Zenk (1975). The CoA-esters were purified using Chromabond C18ec cartridges (Macherey–Nagel) with 4% aqueous ammonium acetate to rinse the column and water for elution (Beuerle and Pichersky 2002). The CoA-ester concentrations were determined photometrically using the absorption maxima and molecular extinction coefficients published by Stöckigt and Zenk (1975) and Zenk (1979).

Enzymatic synthesis of *p*-coumaroylshikimic, *p*-coumaroylquinic and *p*-coumaroyl-3-hydroxyanthranilic acids

p-Coumaroylshikimic, *p*-coumaroylquinic and *p*-coumaroyl-3-hydroxyanthranilic acids were produced enzymatically using heterologously synthesized hydroxycinnamoyltransferases from *Anthoceros agrestis* and *Sarcandra glabra* (from our laboratory collection) with *p*-coumaroyl-CoA and the corresponding acid as substrates. Assays were composed as follows: 100 µl buffer (0.1 M potassium phosphate buffer (KP_i) pH 7.0), 10 µl *p*-coumaroyl-CoA (2.5 mM), 5 µl ascorbic acid (12.5 mM), 5 µl substrate (10 mM) and 5 µl protein preparation. All enzyme assays were incubated up to 3 h at 30 °C and then heated to 95 °C for 5 min to inactivate the enzymes. Products were used directly for qualitative assays

without further purification. Alternatively, *p*-coumaroylshikimic acid was purified by HPLC using a semi-preparative Equisil ODS column (250×8 mm; pre-column: 20×8 mm; particle size 5 µm) and isocratic elution with 35% aqueous methanol containing 0.3% acetic acid at a flow rate of 4 ml/min and detection at 312 nm. Samples were stored at -20 °C until use.

Isolation of *p*-coumaroyl-4'-hydroxyphenyllactic acid and caffeoyl-4'-hydroxyphenyllactic acid from suspension cultures of *Melissa officinalis*

The precursors of RA, *p*-coumaroyl-4'-hydroxyphenyllactic acid and caffeoyl-4'-hydroxyphenyllactic acid, were isolated from *Melissa officinalis* suspension cultures treated with the cytochrome P450 inhibitor tetracycline (BASF). 100 µM tetracycline (in acetone, filter-sterilized) was added to a four-day-old suspension culture of *Melissa officinalis* (cultivated in CB2 medium (Petersen and Alfermann 1988) at 25 °C in the dark at 100 rpm) and the cells were incubated for another 7 days under the same conditions. The cells were harvested by vacuum filtration and stored at -20 °C until further use. 0.01% aqueous H₃PO₄ was heated to 100 °C, added to the cells (1 ml/g cell fresh weight) and the mixture was incubated at 80 °C for 20 min. The cells were filtered off and the filtrate was extracted three times with 25 ml ethyl acetate (EtOAc) each. The organic phases were collected and evaporated and the residue was re-suspended in 2 ml EtOAc. The cell extract was applied to a preparative silica gel TLC plate (silica gel 60 F₂₅₄) and developed in EtOAc:chloroform:formic acid (5:4:1, by vol.) as mobile phase. The expected bands were scraped off and eluted four times with 1 ml methanol each. The concentration was determined photometrically using the following absorption maxima and molecular extinction coefficients (*p*-coumaroyl-4'-hydroxyphenyllactic acid 312 nm, $\epsilon = 11.94 \text{ mM}^{-1} \text{ cm}^{-1}$; caffeoyl-4'-hydroxyphenyllactic acid 328 nm, $\epsilon = 9.15 \text{ mM}^{-1} \text{ cm}^{-1}$).

Chemical synthesis of *p*-coumaroylanthranilic acid, *p*-coumaroyl-3-hydroxyanthranilic acid and *p*-coumaroyltyramine

p-Coumaroylanthranilic acid and *p*-coumaroyl-3-hydroxyanthranilic acid were synthesized essentially according to Alber et al. (2019). For the synthesis of *p*-coumaroyltyramine, one equivalent (0.3 g) *p*-coumaric acid was dissolved in 10 ml acetonitrile, then triethylamine (0.642 ml, 2.5 equivalents) and 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide × HCl (0.420 g, 1.2 equivalents) were added and the solution was stirred at 0 °C for 30 min. After addition of 1 equivalent (0.250 g) tyramine, the mixture was stirred at room temperature overnight. The solvent was evaporated and the remaining solid was re-suspended in 50 ml water and

extracted three times with 50 ml EtOAc. After evaporation of the organic solvent, the residue was purified by column chromatography using silica gel as solid phase with isocratic elution using EtOAc:cyclohexane (3:1, v/v).

All products were analyzed by ¹H-NMR, *p*-coumaroyl-3-hydroxyanthranilic acid was additionally analyzed by ¹³C-NMR (Suppl. Table S2). Data unequivocally show that the product of the chemical synthesis is an amide of *p*-coumaric acid and 3-hydroxyanthranilic acid and not an ester.

p-Coumaroyl-2-threonic acid was kindly provided by Dr. Hugues Renault, Université de Strasbourg.

Standard assay for AaHCT6 and reaction kinetics

AaHCT6 standard enzyme assays with shikimic acid were performed in 1.5 ml reaction tubes containing 105 µl potassium phosphate buffer (KP_i) buffer pH 7.0, 5 µl 12.5 mM ascorbic acid, 5 µl 2.5 mM *p*-coumaroyl- or caffeoyl-CoA, 5 µl 10 mM shikimic acid and 5 µl crude protein extract (appr. 5 µg protein), assays with 3-hydroxyanthranilic acid were composed similarly but contained 105 µl 1 M KP_i buffer pH 9.0, 5 µl 12.5 mM ascorbic acid, 5 µl 2.5 mM *p*-coumaroyl- or caffeoyl-CoA, 5 µl 10 mM 3-hydroxyanthranilic acid (in 1 M HCl) and 5 µl crude protein extract (appr. 0.5 µg protein). The reactions were pre-warmed at 30 °C for 15 min and then started by addition of the protein. Incubation lasted for 5 min and the reaction was stopped by addition of 20 µl 6 M HCl for assays with shikimic acid and 60 µl 6 M HCl for assays with 3-hydroxyanthranilic acid. The reaction products were extracted twice with 500 µl EtOAc each. After evaporation of the solvent, the residues were re-dissolved in 100 µl of the respective HPLC mobile phase (see below). The reaction products were analyzed and quantified by HPLC using an Equisil ODS column (250×4 mm; pre-column: 20×4 mm; particle size 5 µm) and isocratic elution with 35% (products with shikimic acid) or 50% (products with 3-hydroxyanthranilic or 2,3-dihydroxybenzoic acids) aqueous methanol containing 0.01% H₃PO₄ at a flow rate of 1 ml/min and detection at 312 nm for *p*-coumaric acid derivatives and 333 nm for caffeic acid derivatives. For the determination of kinetic values, the substrates and substrate concentrations were varied (for details Suppl. Table S3). Data were analyzed with the GraphPad Prism 5 software using the Michaelis–Menten, Lineweaver–Burk (not shown) and Hanes–Woolf models.

Standard assay for AaCYP98 and reaction kinetics

Standard assays for AaCYP98 in 1.5 ml reaction vials contained 100 µl crude protein extract (0.2 mg protein), 7.5 µl 10 mM substrate (in 50% methanol) or HCT-enzyme assays (see above), 12.5 µl 50 mM NADPH and 5 µl buffer (0.1 M Tris–HCl pH 7.0, 1 mM DTT, 1 mM DIECA). Assays

were mixed vigorously and incubated for 120 min at 30 °C under shaking at 1200 rpm in an Eppendorf Thermomixer. The incubation time for kinetic assays was kept at 5 min (*p*-coumaroyl-3-hydroxyanthranilic acid) and 10 min (*p*-coumaroylanthranilic acid and *p*-coumaroyltyramine) or 30 min (*p*-coumaroylshikimic acid) after having ensured the linearity of product formation at the lowest used substrate concentration at this time point. The reaction was stopped by addition of 50 µl 6 M HCl. The assays were extracted twice with 500 µl EtOAc each and the EtOAc extracts combined and evaporated. The residues were re-dissolved in 50 µl methanol containing 0.01% H₃PO₄ (85%) and centrifuged after addition of 50 µl aqueous 0.01% H₃PO₄. Quantification and analysis of the reaction product were performed by HPLC (injection volume 50 µl) or LC–MS (injection volume 10 µl). Analysis by HPLC was performed using a Hypersil ODS column (250×4 mm; pre-column: 20×4 mm; particle size 5 µm) and isocratic elution with 45% or 50% aqueous methanol containing 0.01% H₃PO₄ at a flow rate of 1 ml/min and detection at 333 nm. LC was performed on an Agilent Technologies HPLC 1260 with a Multospher 120 RP18 column (250×2 mm; particle size 5 µm) using a solvent system of A=0.1% (v/v) aqueous formic acid, B=acetonitrile with 0.1% (v/v) formic acid at a flow rate of 0.25 ml/min and a temperature of 20 °C using the following gradient: 0–40 min 5% B→100% B; 40–45 min 100% B; 45–45.10 min 100% B→5% B; 45.10–55 min 5% B. Detection was performed with the mass spectrometer micrOTOF-Q III with ESI source (Bruker Daltonics) calibrated with 5 mM CHNaO₂ using the negative mode. Data from kinetic analysis are deduced from at least three independent protein isolations with three technical replicates for each substrate concentration and were analyzed with the GraphPad Prism 5 software using the Michaelis–Menten, Lineweaver–Burk (not shown) and Hanes–Woolf models.

Results

Isolation and analysis of a cDNA encoding a hydroxycinnamoyltransferase (AaHCT6) from *Anthoceros agrestis*

Searching the genome database of *Anthoceros punctatus* (Kelly, personal communication), a close relative of *Anthoceros agrestis*, with the protein sequence of *Plectranthus scutellarioides* RAS (UniProtKB: A0PDV5) resulted in the identification of several putative BAHD hydroxycinnamoyltransferase sequences. The sequence with the highest identity with the query sequence, scaffold 3426, was further used for this study. PCR with cDNA from *Anthoceros agrestis* suspension cultures and primers directed against this partial putative BAHD acyltransferase scaffold resulted in a

partial sequence (1016 bp) similar to known hydroxycinnamoyltransferase sequences. RACE-PCR completed the full-length sequence comprising 1305 bp encoding a protein of 434 amino acid residues, which we named AaHCT6. The AaHCT6 sequence was deposited in GenBank under the accession number MW248389. The translated amino acid sequence displayed identities of 55–59% and similarities of 78–80% to hydroxycinnamoyltransferases from di- and monocotyledonous plants (FASTA). The similarity/identity (EMBOSS Needle) to RA synthases from *Melissa officinalis* (UniProtKB: G0LD36) and *Plectranthus scutellarioides* (UniProtKB: A0PDV5) were at 62.6/46.2% and 62.6/44.8%, respectively. Due to the long loop found in *Physcomitrium patens* HCT *Pp3c2_29140* (Kriegshauser et al. 2021), the identity to this sequence is only at 47.4%. The translated protein had a molecular mass of 47.5 kDa and an isoelectric point of 6.22. Conserved sequence motifs for BAHD acyltransferases, such as HxxxDG and DFGWG (St. Pierre and De Luca 2000), are present as well as the “arginine handle” described by Chiang et al. (2018) to be essential for substrate recognition (see boxes in Suppl. Fig. S4). This motif is altered in RAS from *Melissa officinalis* (Weitzel and Petersen 2011) where the Arg residue is replaced by Gln. The conserved amino acid residues Thr and Trp located between the “arginine handle” and the DFGWG motif were reported to be involved in substrate binding and catalysis besides the catalytic His residue of the HxxxDG motif (Chiang et al. 2018). The conservation coloring in Suppl. Fig. S4 shows the rather high conservation within the two protein domains, but distinct differences in the lengths of the loops connecting these two domains as also described by Kriegshauser et al. (2021). This loop is much less extended in AaHCT6 than in hydroxycinnamoyltransferases from *Physcomitrium patens* or *Marchantia paleacea* and thus resembles more seed plant hydroxycinnamoyltransferases.

Isolation and analysis of a cDNA encoding CYP98 from *Anthoceros agrestis*

Based on PCR primers directed against a partial putative CYP98 sequence from *Anthoceros agrestis* (scaffold 26091; P. Szövényi, personal communication), an internal 1050 bp fragment was isolated and sequenced. This fragment showed high similarities to other plant CYP98 sequences. RACE-PCR was used to amplify the cDNA ends. The full open reading frame consisted of 1515 bp. This sequence can be translated to a 505 amino acid protein with an isoelectric point at pH 9.19. The CYP98 sequence from *Anthoceros agrestis* was deposited in GenBank under the accession number MT119883 and was classified as CYP98A147 by Dr. David Nelson (cytochrome P450 homepage: <https://drnelson.uthsc.edu/>). As depicted in Suppl. Fig. S5, the amino acid sequence of AaCYP98

showed elements generally found in the canonic P450 monooxygenases, such as the proline-rich region (Bak et al. 2011), the PERF motif, the heme-binding cysteine motif and the threonine-containing binding pocket motif (Schuler 1996; Mizutani et al. 1997; Chapple 1998). On amino acid level, AaCYP98 was analyzed by BLASTp and showed highest identity (71%) to CYP98 from *Physcomitrium patens*. The highest identity (65%; FASTA) with a seed plant CYP98 was to CYP98A3 from *Arabidopsis thaliana* (UniProtKB: O22203) which catalyzes the *meta*-hydroxylation of *p*-coumaroylshikimic or quinic acid esters (Schoch et al. 2001). The similarity/identity (EMBOSS Needle) on amino acid level to CYP98A14 from *Plectranthus scutellarioides* involved in RA biosynthesis was at 75.4/63.1%.

Phylogenetic analyses of AaHCT6 and AaCYP98 amino acid sequences

Phylogenetic trees were constructed with the MEGA7 software package for both, AaHCT6 and AaCYP98 amino acid sequences using the maximum likelihood algorithm (Suppl. Figs. S2 and S3).

In the phylogenetic analysis (Suppl. Fig. S2) based on amino acid sequences of BAHD hydroxycinnamoyltransferases, AaHCT6 turns up in the same sub-branch as hydroxy-cinnamoyltransferases from the moss *Physcomitrium patens* and two liverworts as well as two hydroxycinnamoyltransferases from the lycophyte *Selaginella moellendorffii*. The enzymes from this branch that were already tested for substrate acceptance convert shikimic, quinic and (hydroxy-)anthranilic acids. The phylogenetic tree shows two major branches. One branch contains two groups of dicotyledonous plant hydroxycinnamoyltransferases accepting shikimic and/or quinic acids. These two groups are separated by hydroxy-cinnamoyltransferases from grasses, gymnosperms, lycophytes and bryophytes. The phylogenetic analysis further shows that the relationships are not ordered according to the evolutionary level. The second major branch contains hydroxycinnamoyltransferases from seed plants as well as bryophytes with a very diverse substrate spectrum, among them those involved in RA biosynthesis. Anthranilic, malic and piscidic acids as well as spermine/spermidine, tetrahydroxyhexanedioic acid and fatty acids also occur as acceptor substrates for hydroxycinnamoyltransferases in this branch besides shikimic and quinic acids which are only rarely represented here. The observed scheme underlines the general hypothesis that substrate diversification has occurred several times during the evolution of hydroxycinnamoyltransferases and that hydroxycinnamoyltransferases with the same substrate range have evolved independently.

AaCYP98 groups closest with the single CYP98 sequences from mosses (*Physcomitrium patens*, *Sphagnum fallax*) and the liverwort *Marchantia polymorpha* followed

by the lycophyte *Selaginella moellendorffii* and the fern group represented by *Pteris vittata* and *Blechnum spicant* (Suppl. Fig. S3). These non-seed plant CYP98 members group on one branch which is neighbored by the branch of gymnosperm CYP98 members. The branch with monocotyledonous plants is divided into two sub-branches with the Poaceae CYP98s on one side-branch and *Phoenix dactylifera* (Arecaceae) on the other. CYP98s from dicotyledonous plants are found on a separate highly ramified main branch divided from all before-mentioned CYP98s. This can be explained by the already mentioned diversification of CYP98 in the angiosperms (Alber et al. 2019). Regarding substrate specificities, CYP98 enzymes metabolizing, e.g., shikimic acid or hydroxyphenyllactic acid esters appear on several separated branches indicating that substrate specificities have evolved several times independently as already postulated by Alber et al. (2019).

Expression of *Anthoceros agrestis* HCT6 in *Escherichia coli* and characterization of hydroxycinnamoyltransferase activity

The full-length coding sequence of AaHCT6 with an *N*-terminally attached His-tag was ligated into the expression vector pET-15b and introduced into *E. coli* SoluBL21 for expression. The crude protein extract contained enzymatically active hydroxycinnamoyltransferases which did not appear in protein extracts from *E. coli* carrying the empty pET-15b vector. Thus, hydroxycinnamoyltransferase activity was assigned to the transgene. Unfortunately, attempts to purify the His-tagged protein by metal-chelate chromatography resulted in almost complete loss of activity (an observation that was made for several different BAHD acyltransferases) and therefore the characterization experiments were performed with crude protein extracts. However, Western blot analyses clearly showed the presence of a His-tagged protein with the expected size missing in empty vector controls (Suppl. Fig. S6). First attempts aimed at identifying acceptor substrates using caffeoyl-CoA as hydroxycinnamoyl donor. AaHCT6 readily accepted shikimic acid together with *p*-coumaroyl-CoA and caffeoyl-CoA forming the respective esters. The identity of the formed *p*-coumaroylshikimic acid was verified by LC–MS analysis (Fig. 2a, b). Crude protein extracts from empty vector controls did not catalyze product formation (Suppl. Fig. S7). 3-Hydroxyanthranilic acid (but not anthranilic acid) was accepted as well; here, ester as well as amide formation would be possible. To verify the identity of the formed product, LC–MS analyses were performed which firstly indicated the correct mass at *m/z* 298 [M–H] and secondly displayed the identical retention time as the chemically synthesized *p*-coumaroyl-3-hydroxyanthranilic acid which was proven to be the amide (Fig. 2c, d). Ester formation was moreover observed with *p*-coumaroyl-CoA

Fig. 2 LC–MS analysis of *p*-coumaroylshikimic acid (**a**) UV-chromatogram at 312 nm. **b** Extracted ion chromatogram (EIC) at m/z 319 [M-H][−], *p*-coumaroyl-3-hydroxyanthranilic acid (**c**) UV-chromatogram at 296 nm. **d** Extracted ion chromatogram (EIC) at m/z 298 [M-H][−] and caffeoyl-2,3-dihydroxybenzoic acid (**e**) UV-chromatogram at 296 nm. **f** Extracted ion chromatogram (EIC) at m/z 315 [M-H][−] formed in enzyme assays by AaHCT6 (black line) and the respective standards (gray dashed line) or assay with reaction time 0 min for 2,3-dihydroxybenzoic acid (gray dashed line)

or caffeoyl-CoA and 2,3-dihydroxybenzoic acid (Fig. 2e, f). In total, we have tested 55 putative acceptor substrates (see Suppl. Table S4), among them 4-hydroxyphenyllactic acid and quinic acid, but only the above-mentioned three compounds resulted in product formation. The temperature and pH optima were determined with caffeoyl-CoA and shikimic acid to be at 35 °C and pH 7.3, respectively (Suppl. Fig. S8).

Kinetic data were determined for shikimic, 3-hydroxyanthranilic and 2,3-dihydroxybenzoic acids together with *p*-coumaroyl- and caffeoyl-CoA as hydroxycinnamoyl donors (Table 1 and Suppl. Fig. S9 and S10). The highest affinities were determined for *p*-coumaroyl-CoA as donor substrate (K_m 14.4 ± 1.4 μM with shikimic acid, K_m 29.3 ± 6.6 μM with 3-hydroxyanthranilic acid), caffeoyl-CoA displayed a slightly higher K_m value together with shikimic acid (19.2 ± 16.2 μM), whereas the K_m value

with 3-hydroxyanthranilic acid was considerably higher (314.4 ± 28.7 μM). Interestingly, caffeoyl-CoA had a higher affinity with 2,3-dihydroxybenzoic acid as acceptor substrate (K_m caffeoyl-CoA 171.9 ± 13.2 μM, *p*-coumaroyl-CoA 388.7 ± 44.1 μM). Shikimic acid was accepted with a remarkably higher affinity than 3-hydroxyanthranilic acid and 2,3-dihydroxybenzoic acid, especially together with *p*-coumaroyl-CoA as donor substrate (K_m 24.6 ± 10.4 μM versus 153.1 ± 33.8 μM and 1025.9 ± 59.7 μM, respectively). With caffeoyl-CoA as donor substrate, the K_m values were quite high but in the same range (K_m 587.0 ± 26.5 μM for shikimic acid, 761.0 ± 74.7 μM for 3-hydroxyanthranilic acid, 686.7 ± 38.3 μM for 2,3-dihydroxybenzoic acid) indicating unfavorable substrate combinations. Since crude protein extracts had to be used for enzyme assays, real turnover numbers could not be calculated.

Table 1 Kinetic values of AaHCT6 for various acceptor and donor substrates in combination

K_m value for	With	MM: K_m [μ M]	HW: K_m [μ M]
<i>p</i> -Coumaroyl-CoA	Shikimic acid	14.4 $\pm$ 1.4	5.1 $\pm$ 1.3
Caffeoyl-CoA	Shikimic acid	19.2 $\pm$ 16.2	19.2 $\pm$ 17.1
<i>p</i> -Coumaroyl-CoA	3-Hydroxyanthranilic acid	29.3 $\pm$ 6.6	34.5 $\pm$ 5.2
Caffeoyl-CoA	3-Hydroxyanthranilic acid	314.4 $\pm$ 28.7	343.3 $\pm$ 15.6
<i>p</i> -Coumaroyl-CoA	2,3-Dihydroxybenzoic acid	388.7 $\pm$ 44.1	323.0 $\pm$ 31.5
Caffeoyl-CoA	2,3-Dihydroxybenzoic acid	171.9 $\pm$ 13.2	162.4 $\pm$ 15.7
Shikimic acid	<i>p</i> -Coumaroyl-CoA	24.6 $\pm$ 10.4	57.6 $\pm$ 13.0
Shikimic acid	Caffeoyl-CoA	587.0 $\pm$ 26.5	592.9 $\pm$ 35.5
3-Hydroxyanthranilic acid	<i>p</i> -Coumaroyl-CoA	153.1 $\pm$ 33.8	202.7 $\pm$ 34.2
3-Hydroxyanthranilic acid	Caffeoyl-CoA	761.0 $\pm$ 74.7	780.6 $\pm$ 67.7
2,3-Dihydroxybenzoic acid	<i>p</i> -Coumaroyl-CoA	1025.9 $\pm$ 59.7	1033.7 $\pm$ 57.4
2,3-Dihydroxybenzoic acid	Caffeoyl-CoA	686.7 $\pm$ 38.2	747.5 $\pm$ 50.8

Values were calculated with the help of Michaelis–Menten (MM) and Hanes–Woolf plots (HW). Mean values of at least six determinations $\pm$ standard error

Expression of codon-optimized *Anthoceros agrestis* CYP98 in *Saccharomyces cerevisiae* and characterization of CYP98 activity

Since protein formation was not detectable after expression of the native AaCYP98 sequence in yeast (data not shown), AaCYP98 was codon-optimized (coAaCYP98). After transformation of *Saccharomyces cerevisiae* BY4742 with pESC-URA harboring either only coAaCYP98 (coAaCYP98) or coAaCYP98 together with the coding sequence of the *Coleus blumei* CPR (coAaCYP98 + CbCPR), plasmid uptake was verified by colony-PCR (Suppl. Fig. S11a). With an expected size of approximately 1800 bp, coAaCYP98 was detectable in all transformants. For transformants carrying also the CPR from *Coleus blumei*, colony-PCR was additionally performed using primers directed against the multiple cloning site II (MCSII) of the expression vector and a signal was detected at the expected size of around 2600 bp (Suppl. Fig. S11b). After expression, protein formation of coAaCYP98 was verified by Western blot analysis with anti-FLAG antibody, showing a band of the expected size (~60 kDa) (Suppl. Fig. S11a, b). Western blot analysis using anti-Myc antibody against CbCPR with a C-terminally Myc-tag was not attempted.

First activity assays were performed using crude protein extract from transformed yeast cells and *p*-coumaroyl-3-hydroxyanthranilic acid together with NADPH as substrates, since microsome preparation resulted in strongly reduced activities. HPLC analysis showed the formation of caffeoyl-3-hydroxyanthranilic acid for both coAaCYP98 and coAaCYP98 + CbCPR. However, co-expression of CPR resulted in 13-fold caffeoyl-3-hydroxyanthranilic acid formation (Fig. 3a, b). LC–MS analysis resulted in a peak at m/z 314.07 [M–H] corresponding to the exact mass of caffeoyl-3-hydroxyanthranilic acid (Fig. 3c). Protein

extracts from BY4742 transformed with the empty vector were used as a negative control and displayed no product formation. *p*-Coumaroylanthranilic acid, *p*-coumaroyl-tyramine, *p*-coumaroylshikimic acid and *p*-coumaroyl-4'-hydroxyphenyllactic acid were converted as well by crude protein extracts of coAaCYP98 + CbCPR (Suppl. Fig. S12), but not by the less active coAaCYP98. Due to unknown reasons, both, *p*-coumaroylanthranilic acid and caffeoyl-3-hydroxyanthranilic acid split into two peaks in HPLC and LC–MS analysis (Suppl. Fig. S12). *p*-Coumaroylquinic acid, *p*-coumaroyl-2-threonic acid and caffeoyl-4'-hydroxyphenyllactic acid were not converted by AaCYP98.

p-Coumaroylanthranilic acid was used as substrate for the determination of temperature and pH optima. Suppl. Fig. S13 shows the pH and temperature optima for coAaCYP98 + CPR. They point out a temperature optimum at 30–35 °C and a pH-optimum around pH 7.

Kinetic data for AaCYP98 (coAaCYP98 + CPR) were determined for *p*-coumaroylanthranilic acid, *p*-coumaroyltyramine, *p*-coumaroyl-3-hydroxyanthranilic acid and for the co-substrate NADPH (Table 2; Suppl. Fig. S14). For determination of the kinetic values of *p*-coumaroylanthranilic acid, the areas for both product peaks were summed up. *p*-Coumaroylanthranilic and *p*-coumaroyl-3-hydroxyanthranilic acids were the substrates accepted with the highest affinities (K_m 6.4 $\pm$ 0.8 μ M and 9.3 $\pm$ 0.7 μ M, respectively), followed by *p*-coumaroyl-tyramine (K_m 36.3 $\pm$ 3.6 μ M). For the co-substrate NADPH coAaCYP98 + CPR displayed a K_m value of 53.2 $\pm$ 4.0 μ M. In our hands, no saturation could be achieved up to 1 mM *p*-coumaroylshikimic acid and thus a K_m value for this substrate could not be determined. This observation indicates a rather high K_m value and thus a low affinity towards this substrate.

Fig. 3 Activity of crude protein extract (200 μ g protein) of coAaCYP98 or coAaCYP98+CbCPR with *p*-coumaroyl-3-hydroxyanthranilic acid (pC-3-OH-An) analyzed by HPLC (333 nm) and LC-MS. **a** Enzyme assay with *p*-coumaroyl-3-hydroxyanthranilic acid after 120 min at 25 °C with either coAaCYP98 (black) or coAaCYP98+CbCPR (gray). HPLC with isocratic elution with 50% methanol+0.01% H_3PO_4 . **b** *p*-Coumaroyl-3-hydroxyanthranilic acid incubated for 120 min at 25 °C with coAaCYP98 (solid line)

or crude protein extract from the empty vector control (dashed line). HPLC with isocratic elution with 45% methanol+0.01% H_3PO_4 . **c** LC-MS analysis of coAaCYP98 incubated with *p*-coumaroyl-3-hydroxyanthranilic acid for 0 min (gray) or 120 min (black). HPLC with acetonitrile/water with 0.1% formic acid gradient elution. Chromatograms show the extracted ion chromatogram (EIC) of caffeoyl-3-hydroxyanthranilic acid in the negative mode. The exact mass of caffeoyl-3-hydroxyanthranilic acid is m/z 314.07 [M-H]

Discussion

One of our main interests in looking into the biosynthesis of hydroxycinnamic acid derivatives, especially RA,

in *Anthoceros agrestis* was to investigate whether the biosynthesis of RA and the genes lying behind this pathway are evolutionary old traits that have been passed on to the seed plant families or whether this biosynthetic pathway has

Table 2 Kinetic values of AaCYP98 (in combination with CPR from *Coleus blumei*) for *p*-coumaroyl-3-hydroxyanthranilic acid, *p*-coumaroylanthranilic acid, *p*-coumaroyltyramine and NADPH

K_m value for	With	MM: K_m [μ M]	HW: K_m [μ M]
<i>p</i> -Coumaroyl-3-hydroxyanthranilic acid	NADPH	9.3 $\pm$ 0.7	7.2 $\pm$ 0.9
<i>p</i> -Coumaroylanthranilic acid	NADPH	6.4 $\pm$ 0.8	5.5 $\pm$ 1.6
<i>p</i> -Coumaroyltyramine	NADPH	36.3 $\pm$ 3.6	34.3 $\pm$ 0.7
NADPH	<i>p</i> -Coumaroylanthranilic acid	53.2 $\pm$ 4.0	50.6 $\pm$ 3.4

Values were calculated with the help of Michaelis–Menten (MM) and Hanes–Woolf plots (HW). Mean values of at least 9 determinations $\pm$ standard error

evolved several times independently. Recent investigations of members of the Lamiaceae and the Boraginaceae indicate that hydroxycinnamoyltransferases catalyzing the transesterification of a hydroxycinnamoyl-CoA and a hydroxyphenyllactic acid moiety have evolved independently in these two plant families (Levsh et al. 2019). The hydroxycinnamoyltransferase from *Phacelia campanularia* (Boraginaceae) proved to be more closely related to spermidine hydroxycinnamoyltransferases. Furthermore, two different CYP98 proteins (CYP98A112, CYP98A113) were found to be involved in finalizing the substitution patterns of the aromatic rings (Levsh et al. 2019) while only one (CYP98A14) is known to date from *Coleus blumei* as a member of the Lamiaceae (Eberle et al. 2009). Both enzyme families, BAHD acyltransferases and CYP98, have widely expanded at different time points during evolution (Tuominen et al. 2011; Alber et al. 2019). While only one candidate sequence for CYP98 has been identified in the genome database of *Anthoceros agrestis* and other early diverged plant groups as well as gymnosperms, already multiple gene candidates for acyltransferases of the BAHD superfamily are present in these plant taxa.

In search for the enzyme catalyzing the formation of RA or one of its less hydroxylated precursors in the hornwort *Anthoceros agrestis*, we searched the genome database of the closely related species *Anthoceros punctatus* (Kelly, personal communication) and came upon the partial sequence of scaffold 3426 with the highest identity to RAS from *Melissa officinalis* (Weitzel and Petersen 2011) among the BAHD-sequences in this hornwort. However, other less similar sequences of BAHD acyltransferases were found as well, none of which at the present time encoded a protein with RAS activity (own unpublished results). After publication of whole genome sequences of *Anthoceros agrestis* (Szövényi 2016; Carpenter et al. 2019; Leebens-Mack et al. 2019), we later used our full-length AaHCT6 sequence from *Anthoceros agrestis* suspension-cultured cells to search the 1kp database (<https://db.cngb.org/onekp/>) and identified scaffolds 2004822, 2004823 and 2050582 (with identical amino acid sequences) that displayed an identity of 48% to RAS from *Melissa officinalis* and 100% to our AaHCT6 amino acid sequence. The identities to other scaffolds ranged

between 30 and 40%. For this reason, we directed our interest primarily towards AaHCT6 (corresponding to scaffold 3426 from *A. punctatus*) as the gene most similar to RAS from Lamiaceae. The encoded protein, however, was not active with 4-hydroxyphenyllactic acid, the substrate involved in RA biosynthesis, but accepted shikimic acid as acceptor substrate and—to a lesser extent—3-hydroxyanthranilic acid and 2,3-dihydroxybenzoic acid, but not anthranilic acid. 3-Hydroxyanthranilic acid and 2,3-dihydroxybenzoic acid have previously been shown to be substrates for hydroxycinnamoyl-CoA:shikimic acid hydroxycinnamoyltransferase from *Coleus blumei* (CbHST) (Sander and Petersen 2011) and the former was also converted by RAS from *Lavandula angustifolia* (Landmann et al. 2011).

Kinetic characterization showed that *p*-coumaroyl-CoA was preferred as hydroxycinnamoyl donor over caffeoyl-CoA (except with 2,3-dihydroxybenzoic acid), while the enzyme displayed a higher affinity towards shikimic acid compared to 3-hydroxyanthranilic and 2,3-dihydroxybenzoic acids, the latter two probably do not serve as substrates *in planta*. Reactions with shikimic and 2,3-dihydroxybenzoic acids led to ester formation while the reaction with 3-hydroxyanthranilic acid has been shown to result in an amide (Suppl. Table S2). It must be pointed out here, that we did not yet determine which of the hydroxyl groups of 2,3-dihydroxybenzoic acid has been used for ester formation. Substrate promiscuity had already been shown for a number of other hydroxycinnamoyltransferases from plants forming esters as well as amides (Sander and Petersen 2011; Landmann et al. 2011; Eudes et al. 2016; Kriegshauser et al. 2021). A hydroxycinnamoyltransferase acting on anthranilic acid as substrate has been described from carnation (*Dianthus caryophyllus*; Caryophyllaceae), while hydroxycinnamoyltransferases from *Avena sativa* transfer the hydroxycinnamoyl moiety to 5-hydroxyanthranilic acid in avenanthramide biosynthesis (Yang et al. 1997, 2004). Hydroxycinnamoyltransferases from the moss *Physcomitrium patens* and the lycophyte *Selaginella moellendorffii* accepted 3-hydroxy- and 5-hydroxyanthranilic acids besides shikimic acid and a variety of other phenolic compounds (Eudes et al. 2016). Liverwort hydroxycinnamoyltransferases from *Marchantia paleacea* and *Plagiochasma appendiculatum* were shown

to be active with shikimic and quinic acids, however, other substrates have not been tested (Wu et al. 2018).

In the phylogenetic tree (Suppl. Fig. S2), the sequences encoding the above-mentioned enzymes from bryophytes and lycophytes group closely together, while the *Avena sativa* hydroxycinnamoyltransferases group together with other sequences from monocotyledonous angiosperms and the *Dianthus caryophyllus* enzyme is in a mixed group of hydroxycinnamoyltransferases far apart. This supports the hypothesis that ester as well as amide formation by BAHD hydroxycinnamoyltransferases has evolved independently several times and may be an intrinsic capability of this enzyme family. Overall, the substrate specificities of hydroxycinnamoyltransferases from early diverged plants do not seem to differ substantially from those of seed plants, since all accept shikimic acid for ester formation as well as various other substrates, some of them leading also to the synthesis of hydroxycinnamoyl amides. This has recently been shown by Kriegshauser et al. (2021) who demonstrated that bryophyte hydroxycinnamoyltransferases from *Marchantia polymorpha* and *Physcomitrium patens* could readily complement an *Arabidopsis thaliana* mutant defective in the hydroxycinnamoyltransferase involved in monolignol formation. The *Physcomitrium patens* hydroxycinnamoyltransferase analyzed by Kriegshauser et al. (2021) finally turned out to prefer shikimic acid as acceptor substrate although previously a threonic acid accepting enzyme had been expected deduced from data from previous studies (Renault et al. 2017). Interestingly, feruloyl-CoA was described as best donor substrate for the *Physcomitrium patens* hydroxycinnamoyltransferase, although feruloylshikimic acid (as well as caffeoylshikimic acid) has not been identified as metabolite in plant extracts from *P. patens* in contrast to *p*-coumaroylshikimic acid (Kriegshauser et al. 2021).

The *meta*-hydroxylation of *p*-coumaric acid esters and amides to the respective caffeic acid derivatives is generally assigned to members of the CYP98 family; free *p*-coumaric acid itself is no or only a very poor substrate. The substrate specificity of CYP98 proteins from different plant species can vary considerably. As stated by Alber et al. (2019), the number of CYP98 genes and isoforms is restricted to one in bryophytes, lycophytes, monilophytes and gymnosperms but duplication and neo-functionalization events in angiosperms have increased the number of CYP98 candidates. Among the CYP98 enzymes, substrate promiscuity is a common feature. However, CYP98s from early diverged plants may display different substrate preferences than those from gymno- and angiosperms. While the latter preferably use *p*-coumaroylshikimic acid as substrate but can also hydroxylate *p*-coumaroylquinic acid to varying extents, the single *Physcomitrium patens* CYP98 was reported to accept *p*-coumaroylthreonic acid and *p*-coumaroylanthranilic acid (Renault et al. 2017), but also *p*-coumaroylshikimic acid

(Alber et al. 2019). As reported by Alber et al. (2019), phenol amides (e.g., with anthranilic acid and spermidine) were preferred over phenol esters. However, this view is changing due to the recent publication of Kriegshauser et al. (2021) who postulated that *p*-coumaroylthreonic acid is not a native substrate of CYP98 from *Physcomitrium patens*. Since only one CYP98 sequence is present in *Anthoceros agrestis*, we assumed this to be directly involved in RA biosynthesis. After identifying the respective full-length sequence, first attempts with expression in the frequently used expression system *Saccharomyces cerevisiae* were not successful. We speculated that either the different GC-content of yeast (average 39% in yeast; Nakase and Komagata 1971) or *Anthoceros agrestis* CYP98 (62%) and/or the differing codon usage (codon usage database: www.kazusa.or.jp) might be the reason. Codon optimization of the hornwort gene was therefore the matter of choice and besides adapting the codons for *Saccharomyces cerevisiae*, the GC content was lowered to 36%. Active protein was synthesized by transformed yeast to a low extent using the electron transfer from yeast CPR. Combination of AaCYP98 with CPR from *Coleus blumei* (Eberle et al. 2009), however, strongly increased the hydroxylation activities and therefore this system was used for further characterization. Using a number of putative substrates, substrate promiscuity with preference for hydroxycinnamoyl amides was found for CYP98A147 from the hornwort *Anthoceros agrestis*. Similar affinities were found for *p*-coumaroylanthranilic acid (K_m 6.4 ± 0.8 μ M) and *p*-coumaroyl-3-hydroxyanthranilic acid (K_m 9.3 ± 0.7 μ M) and lower affinity for *p*-coumaroyltyramine (K_m 36.3 ± 3.6 μ M), all of them *p*-coumaroyl amides. The *p*-coumaroyl esters with shikimic and 4-hydroxyphenyllactic acids were accepted as well, but only resulted in rather low product formation (Table 2 and Suppl. Fig. S12). Even at 1 mM *p*-coumaroylshikimic acid, the hydroxylase was not saturated pointing at a rather low affinity towards this substrate. On the other hand, AaHCT6 catalyzed preferentially the formation of *p*-coumaroylshikimic acid, followed by *p*-coumaroyl-3-hydroxyanthranilic acid, while *p*-coumaroylanthranilic acid (one of the preferred substrates of CYP98A147) was not formed. Similarly, Eudes et al. (2016) reported that a *Physcomitrium patens* hydroxycinnamoyltransferase did not catalyze *p*-coumaroylanthranilic acid formation but only accepted 3- and 5-hydroxyanthranilic acid as acceptors. Threonic acid, esterified to *p*-coumaroylthreonic acids, which were postulated to play an essential role in the formation of phenol-enriched cuticles in *Physcomitrium patens* (Renault et al. 2017), was not a substrate for AaHCT6 and *p*-coumaroylthreonic acid was not converted by AaCYP98.

Interestingly, the hydroxylation of the RA precursor *p*-coumaroyl-4'-hydroxyphenyllactic acid by AaCYP98 was very low, although *Anthoceros agrestis* accumulates substantial amounts of RA (Vogelsang

et al. 2006). In *Coleus blumei*, CYP98A14 only accepts *p*-coumaroyl-4'-hydroxyphenyllactic acid, *p*-coumaroyl-3',4'-dihydroxyphenyllactic acid and caffeoyl-4'-hydroxyphenyllactic acid while *p*-coumaroylshikimic acid and *p*-coumaroylquinic acid were not converted (Eberle et al. 2009). The only identified CYP98 in *Anthoceros agrestis*, namely CYP98A147, might therefore not be directly involved in RA biosynthesis. An alternative way towards RA might follow a pathway including an early hydroxylation to the caffeic acid substitution pattern and the transfer of this caffeoyl moiety from an activated precursor to 4-hydroxyphenyl- or 3,4-dihydroxyphenyllactic acid. Recently, the 3-hydroxylation of 4-coumaric acid to caffeic acid by a bifunctional peroxidase that oxidizes both ascorbic acid and 4-coumaric acid and its impact on monolignol formation has been described from grasses, *Medicago truncatula* and *Arabidopsis thaliana* (Barros et al. 2019).

Recently, Kriegshauser et al. (2021) performed metabolic analyses of plant material of different bryophytes and detected *p*-coumaroyl-5-*O*-shikimic acid as well as caffeoyl-5-*O*-shikimic acid in extracts of *Marchantia polymorpha* and *Anthoceros agrestis*. Only the *p*-coumaroyl derivative was found in gametophore extracts of *Physcomitrium patens*, although CYP98 expression—presumably leading to caffeoylshikimic acid—had been shown in *P. patens* at this developmental stage (Renault et al. 2017). This shows the presence of hydroxycinnamoylshikimic acid esters in all bryophyte groups. At the same time, the respective threonic acid esters have only been detected in *Physcomitrium patens* (Kriegshauser et al. 2021). Earlier reports by Trennheuser (1992)—a very detailed phytochemical investigation—and Takeda et al. (1990b) do not mention the presence of *p*-coumaroyl- or caffeoylshikimic acids in *Anthoceros*. However, the presence of hydroxycinnamoylshikimic acid esters in *Anthoceros agrestis* is in line with our observation that shikimic acid is the best substrate for AaHCT6. To our knowledge, Renault et al. (2017) did not test PpCYP98 with *p*-coumaroylshikimic acid. The only mediocre acceptance of *p*-coumaroylshikimic acid as substrate for AaCYP98 may be due to the test conditions although our assays with *p*-coumaroyl(hydroxy)anthranilic acids performed under the same conditions have been successful.

The presented results leave a final conclusion towards the biosynthesis of caffeoyl esters, especially rosmarinic acid, in the hornwort *Anthoceros agrestis* open and strengthen the probability that the formation of RA might follow a different way in this hornwort. This is especially the case for hydroxycinnamoyltransferases, because different classes of enzymes might catalyze the formation of hydroxycinnamic acid esters (see review by Petersen 2016). Since only a single CYP98 member has been found in hornworts, we either did not meet the requirements of this enzyme in vitro or the hydroxylation may take place prior to esterification. All

in all, the formation of phenolic compounds in *Anthoceros agrestis* remains a field of research still to be conquered.

Author contribution statement LE, JW, EB and MP performed experiments and wrote the article. MP finalized the data and the text. All authors read and approved the manuscript.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00425-022-03856-9>.

Acknowledgements We are sincerely grateful to Dr. Peter Szövényi (Universität Zürich) for analyzing his preliminary genome data of *Anthoceros agrestis* for CYP98 sequences. Prof. Dr. Martin Schlitzer and his group and Dr. Regina Ortmann (Philipps-Universität Marburg) introduced us to chemical synthesis of *p*-coumaroyl amides and performed NMR analyses—thank you very much. We are grateful to Dr. Hugues Renault (University of Strasbourg) for samples of *p*-coumaroyl-2-threonic acid as substrate and caffeoylshikimic acid for HPLC quantification. LC-MS analyses were performed by Rixa Kraut and Lena-Christin Ludwig-Radtke (Philipps-Universität Marburg)—thank you for your important help. Some caffeic acid and *p*-coumaric acid esters and amides were enzymatically synthesized by Lucien Ernst and Paul Bömeke—many thanks for this support. The Bruker micrOTOF QIII mass spectrometer was financially supported in part by a grant from the Deutsche Forschungsgemeinschaft (INST 160/620-1) to Prof. Dr. Shu-Ming Li, Philipps-Universität Marburg. This work has been performed in connection with the Priority Programme “MAdLand—Molecular Adaptation to Land” (SPP2237) financially supported as PE360/37-1 by the Deutsche Forschungsgemeinschaft (DFG).

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability statement Sequence data were deposited in GenBank (<https://https.ncbi.nlm.nih.gov/genbank/>) under the accession numbers MW248389 and MT119883. The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alber AV, Renault H, Basilio-Lopes A, Bassard JE, Liu Z, Ullmann P, Lesot A, Bihele F, Schmitt M, Werck-Reichhart D, Ehrling J (2019) Evolution of coumaroyl conjugate 3-hydroxylases in land plants:

- lignin biosynthesis and defense. *Plant J* 99:924–936. <https://doi.org/10.1111/tj.14373>
- Bak S, Beisson F, Bishop G, Hamberger B, Höfer R, Paquette S, Werck-Reichhart D (2011) Cytochromes P450. *Arabidopsis Book/Am Soc Plant Biol* 9:e0144. <https://doi.org/10.1199/tab.0144>
- Barros J, Escamilla-Trevino L, Song L et al (2019) 4-Coumarate 3-hydroxylase in the lignin biosynthesis pathway is a cytosolic ascorbate peroxidase. *Nat Commun* 10:1994. <https://doi.org/10.1038/s41467-019-10082-7>
- Berger A, Meinhard J, Petersen M (2006) Rosmarinic acid synthase is a new member of the superfamily of BAHD acyltransferases. *Planta* 224:1503–1510. <https://doi.org/10.1007/s00425-006-0393-y>
- Beuerle T, Pichersky E (2002) Enzymatic synthesis and purification of aromatic coenzyme A esters. *Anal Biochem* 302:305–312. <https://doi.org/10.1006/abio.2001.5574>
- Binding H, Mordhorst G (1991) Gametophyte regeneration and apospory from archegoniate protoplasts under conditions devised for higher plants. *Bot Acta* 104:330–335. <https://doi.org/10.1111/j.1438-8677.1991.tb00238.x>
- Busch T, Petersen M (2021) Identification and biochemical characterisation of tyrosine aminotransferase from *Anthoceros agrestis* unveils the conceivable entry point into rosmarinic acid biosynthesis in hornworts. *Planta* 253:98. <https://doi.org/10.1007/s00425-021-03623-2>
- Carpenter EJ, Matasci N, Ayyampalayam S, Wu S, Sun J, Yu J, Jimenez-Vieira FR, Bowler C, Dorrell RG, Gitzendanner MA, Li L, Du W, Ullrich K, Wickett NJ, Barkmann TJ, Barker MS, Leebens-Mack JH, Wong GKS (2019) Access to RNA-sequencing data from 1173 plant species: The 1000 plant transcriptomes initiative (1KP). *GigaScience* 8. <https://doi.org/10.1093/gigascience/giz126>
- Chapple C (1998) Molecular-genetic analysis of plant cytochrome P450-dependent monooxygenases. *Annu Rev Plant Physiol Plant Mol Biol* 49:311–343. <https://doi.org/10.1146/annurev.arplant.49.1.311>
- Chiang YC, Levsh O, Lam CK, Weng JK, Wang Y (2018) Structural and dynamic basis of substrate permissiveness in hydroxycinnamoyltransferase (HCT). *PLOS Comput Biol* 14:e1006511. <https://doi.org/10.1371/journal.pcbi.1006511>
- Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 162:156–159. [https://doi.org/10.1016/0003-2697\(87\)90021-2](https://doi.org/10.1016/0003-2697(87)90021-2)
- de Vries J, de Vries S, Slamovits CH, Rose LE, Archibald JM (2017) How embryophytic is the biosynthesis of phenylpropanoids and their derivatives in streptophyte algae? *Plant Cell Physiol* 58:934–945. <https://doi.org/10.1093/pcp/pcx037>
- de Vries S, Fürst-Jansen JMR, Irisarri I, Dhabalia Ashok A, Ischebeck T, Feussner K, Abreu IN, Petersen M, Feussner I, de Vries J (2021) The evolution of the phenylpropanoid pathway entailed pronounced radiations and divergences of enzyme families. *Plant J* 107:975–1002. <https://doi.org/10.1111/tj.15387>
- Dymond JS (2013) *Saccharomyces cerevisiae* growth media. *Methods Enzymol* 533:191–204. <https://doi.org/10.1016/B978-0-12-420067-8.00012-X>
- Eberle D, Ullmann P, Werck-Reichhart D, Petersen M (2009) cDNA cloning and functional characterisation of CYP98A14 and NADPH:cytochrome P450 reductase from *Coleus blumei* involved in rosmarinic acid biosynthesis. *Plant Mol Biol* 69:239–253. <https://doi.org/10.1007/s11103-008-9420-7>
- Eudes A, Pereira JH, Yogiswara S, Wang G, Teixeira Benites V, Baidoo EEK, Lee TS, Adams PD, Keasling JD, Loqué D (2016) Exploiting the substrate promiscuity of hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyl transferase to reduce lignin. *Plant Cell Physiol* 57:568–579. <https://doi.org/10.1093/pcp/pcw016>
- Gietz RD, Schiestl RH (2007) High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc* 2:31–34. <https://doi.org/10.1038/nprot.2007.13>
- Hoffmann L, Maury S, Martz F, Geoffroy P, Legrand M (2003) Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism. *J Biol Chem* 278:95–102. <https://doi.org/10.1074/jbc.M209362200>
- Kriegshauser L, Knosp S, Grienemberger E, Tatsumi K, Gütle DD, Sørensen I, Herrgott L, Zumsteg J, Rose JKC, Reski R, Werck-Reichhart D, Renault H (2021) Function of the HYDROXY-CINNAMOYL-CoA:SHIKIMATE HYDROXYCINNAMOYL TRANSFERASE is evolutionarily conserved in embryophytes. *Plant Cell* 33:1472–1491. <https://doi.org/10.1093/plcell/koab044>
- Labeuw L, Martone PT, Boucher Y, Case RJ (2015) Ancient origin of the biosynthesis of lignin precursors. *Biol Direct* 10:23. <https://doi.org/10.1186/s13062-015-0052-y>
- Laemmli UK (1970) Cleavage of structural proteins during assembly of the head of bacteriophage T4. *Nature* 227:680–685. <https://doi.org/10.1038/227680a0>
- Landmann C, Hücherig S, Fink B, Hoffmann T, Dittlein D, Coirer HA, Schwab W (2011) Substrate promiscuity of a rosmarinic acid synthase from lavender (*Lavandula angustifolia* L.). *Planta* 234:305–320. <https://doi.org/10.1007/s00425-011-1400-5>
- Leebens-Mack JH, Barker MS, Carpenter EJ et al (2019) One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574:679–685. <https://doi.org/10.1038/s41586-019-1693-2>
- Lessard JC (2013) Growth media for *E. coli*. *Methods Enzymol* 533:181–189. <https://doi.org/10.1016/B978-0-12-420067-8.00011-8>
- Levsh O, Pluskal T, Carballo V, Mitchell AJ, Weng JK (2019) Independent evolution of rosmarinic acid biosynthesis in two sister families under the Lamiids clade of flowering plants. *J Biol Chem* 294:P15193–15205. <https://doi.org/10.1074/jbc.RA119.010454>
- Ma XH, Ma Y, Tang JF, He YL, Liu YC, Ma XJ, Shen Y, Cui GH, Lin HX, Rong QX, Guo J, Huang LQ (2015) The biosynthetic pathways of tanshinones and phenolic acids in *Salvia miltiorrhiza*. *Molecules* 20:16235–16254. <https://doi.org/10.3390/molecules200916235>
- Mahesh V, Million-Rousseau R, Ullmann P, Chabrilange N, Bustamante J, Mondolot L, Morant M, Noirot M, Hamon S, de Kochko A, Werck-Reichhart D, Campa C (2007) Functional characterization of two *p*-coumaroyl ester 3 β -hydroxylase genes from coffee tree: evidence of a candidate for chlorogenic acid biosynthesis. *Plant Mol Biol* 64:145–169. <https://doi.org/10.1007/s11103-007-9141-3>
- Mahmood T, Yang PC (2012) Western blot: technique, theory, and trouble shooting. *N Am J Med Sci* 4:429–434. <https://doi.org/10.4103/1947-2714.100998>
- Mizutani M, Ohta D, Sato R (1997) Isolation of a cDNA and a genomic clone encoding cinnamate 4-hydroxylase from *Arabidopsis* and its expression manner in planta. *Plant Physiol* 113:755–763. <https://doi.org/10.1104/pp.113.3.755>
- Moglia A, Comino C, Portis E, Acquadro A, De Vos RCH, Beekwilder J, Lanteri S (2009) Isolation and mapping of a C3'H gene (CYP98A49) from globe artichoke, and its expression upon UV-C stress. *Plant Cell Rep* 28:963–974. <https://doi.org/10.1007/s00299-009-0695-1>
- Nakase T, Komagata K (1971) Significance of DNA base composition in the classification of yeast genus *Saccharomyces*. *J Gen Appl Microbiol* 17:227–238. <https://doi.org/10.2323/jgam.17.227>
- Petersen M (2003) Cinnamic acid 4-hydroxylase from cell cultures of the hornwort *Anthoceros agrestis*. *Planta* 217:96–101. <https://doi.org/10.1007/s00425-002-0960-9>
- Petersen M (2016) Hydroxycinnamoyltransferases in plant metabolism. *Phytochem Rev* 15:699–727. <https://doi.org/10.1007/s11101-015-9417-1>
- Petersen M, Alfermann AW (1988) Two new enzymes of rosmarinic acid biosynthesis from cell cultures of *Coleus blumei*:

- hydroxyphenylpyruvate reductase and rosmarinic acid synthase. *Z Naturforsch* 43c:501–504
- Petersen M, Häusler E, Karwatzki B, Meinhard J (1993) Proposed biosynthetic pathway for rosmarinic acid in cell cultures of *Coleus blumei* Benth. *Planta* 189:10–14. <https://doi.org/10.1007/BF00201337>
- Petersen M, Abdullah Y, Benner J, Eberle D, Gehlen K, Hücherig S, Janiak V, Kim KH, Sander M, Weitzel C, Wolters S (2009) Evolution of rosmarinic acid biosynthesis. *Phytochemistry* 70:1663–1679. <https://doi.org/10.1016/j.phytochem.2009.05.010>
- Pezeshki S (2016) Biosynthese von Kaffeeäuremetaboliten im Ackerschornmoos *Anthoceros agrestis* und im Kleinen Blasenmützenmoos *Physcomitrella patens*. Dissertation, Philipps-Universität Marburg
- Renault H, Alber A, Horst NA, Basilio Lopes A, Fich EA, Kriegshauser L, Wiedemann G, Ullmann P, Herrgott L, Erhardt M, Pineau E, Ehlting J, Schmitt M, Rose JKC, Reski R, Werck-Reichhart D (2017) A phenol-enriched cuticle is ancestral to lignin evolution in land plants. *Nat Commun* 8:14713. <https://doi.org/10.1038/ncomms14713>
- Renault H, Werck-Reichhart D, Weng JK (2019) Harnessing lignin evolution for biotechnological applications. *Curr Opin Biotechnol* 56:105–111. <https://doi.org/10.1016/j.copbio.2018.10.011>
- Saleme MdLS, Cesarino I, Vargas L et al (2017) Silencing *CAFFEYOYL SHIKIMATE ESTERASE* affects lignification and improves saccharification in poplar. *Plant Physiol* 175:1040–1057. <https://doi.org/10.1104/pp.17.00920>
- Sambrook J, Russell D (2001) Molecular cloning: a laboratory manual, 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor
- Sander M, Petersen M (2011) Distinct substrate specificities and unusual substrate flexibilities of two hydroxycinnamoyltransferases, rosmarinic acid synthase and hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyltransferase, from *Coleus blumei* Benth. *Planta* 233:1157–1171. <https://doi.org/10.1007/s00425-011-1367-2>
- Schoch G, Goepfert S, Morant M, Hehn A, Meyer D, Ullmann P, Werck-Reichhart D (2001) CYP98A3 from *Arabidopsis thaliana* is a 3'-hydroxylase of phenolic esters, a missing link in the phenylpropanoid pathway. *J Biol Chem* 276:36566–36574. <https://doi.org/10.1074/jbc.M104047200>
- Schuler MA (1996) Plant cytochrome P450 monooxygenases. *Crit Rev Plant Sci* 15:235–284. <https://doi.org/10.1080/07352689609701942>
- St. Pierre B, De Luca V (2000) Evolution of acyltransferase genes: origin and diversification of the BAHD superfamily of acyltransferases involved in secondary metabolism. In: *Evolution of Metabolic Pathways*. Recent Adv Phytochem 34:285–315
- Stöckigt J, Zenk MH (1975) Chemical syntheses and properties of hydroxycinnamoyl-coenzyme A derivatives. *Z Naturforsch* 30c:352–358. <https://doi.org/10.1515/znc-1975-5-609>
- Szővényi P (2016) The genome of the model species *Anthoceros agrestis*. In: Rensing SA (ed) *Advances in botanical research*. Genomes and evolution of charophytes, bryophytes, lycophytes and ferns, vol 78. Academic Press, pp 189–211
- Szővényi P, Frangedakis E, Ricca M, Quandt D, Wicke S, Langdale JA (2015) Establishment of *Anthoceros agrestis* as a model species for studying the biology of hornworts. *BMC Plant Biol* 15:1–7. <https://doi.org/10.1186/s12870-015-0481-x>
- Takeda R, Hasegawa J, Shinozaki M (1990a) The first isolation of lignans, megacerotonic acid and anthocerotonic acid, from non-vascular plants, Anthocerotae (hornworts). *Tetrahedron Lett* 31:4159–4162. [https://doi.org/10.1016/S0040-4039\(00\)97569-5](https://doi.org/10.1016/S0040-4039(00)97569-5)
- Takeda R, Hasegawa J, Sinozaki K (1990b) Phenolic compounds from Anthocerotae. Bryophytes: Their chemistry and chemical taxonomy. In: *Proc Phytochemical Society of Europe*, vol 29. Oxford Science Publications, Oxford
- Towbin H, Staehelin T, Gordon J (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets—procedure and some applications. *Proc Natl Acad Sci USA* 76:4350–4354. <https://doi.org/10.1073/pnas.76.9.4350>
- Trennheuser F, Burkhard G, Becker H (1994) Anthocerodiazonin, an alkaloid from *Anthoceros agrestis*. *Phytochemistry* 37:899–903. [https://doi.org/10.1016/S0040-4039\(00\)97569-5](https://doi.org/10.1016/S0040-4039(00)97569-5)
- Trennheuser F (1992) Phytochemische Untersuchung und in vitro Kultur ausgewählter Vertreter der Anthocerotopsida. Dissertation, University of Saarbrücken
- Tuominen LK, Johnson VE, Tsai CJ (2011) Differential phylogenetic expansions in BAHD acyltransferases across five angiosperm taxa and evidence of divergent expression among *Populus* paralogues. *BMC Genomics* 12:236. <https://doi.org/10.1186/1471-2164-12-236>
- Vanholme R, Cesarino I, Rataj K et al (2013) Caffeoyl shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway in *Arabidopsis*. *Science* 341:1103–1106. <https://doi.org/10.1126/science.1241602>
- Vogelsang K, Schneider B, Petersen M (2006) Production of rosmarinic acid and a new rosmarinic acid 3'-O-β-D-glucoside in suspension cultures of the hornwort *Anthoceros agrestis* Paton. *Planta* 223:369–373. <https://doi.org/10.1007/s00425-005-0089-8>
- Weitzel C, Petersen M (2011) Cloning and characterisation of rosmarinic acid synthase from *Melissa officinalis* L. *Phytochemistry* 72:572–578. <https://doi.org/10.1016/j.phytochem.2011.01.039>
- Weng JK, Chapple C (2010) The origin and evolution of lignin biosynthesis. *New Phytol* 187:273–285. <https://doi.org/10.1111/j.1469-8137.2010.03327.x>
- Weng JK, Akiyama T, Bonawitz ND, Li X, Ralph J, Chapple C (2010) Convergent evolution of syringyl lignin biosynthesis via distinct pathways in the lycophyte *Selaginella* and flowering plants. *Plant Cell* 22:1033–1045. <https://doi.org/10.1105/tpc.109.073528>
- Wohl J, Petersen M (2020a) Functional expression and characterization of cinnamic acid 4-hydroxylase from the hornwort *Anthoceros agrestis* in *Physcomitrella patens*. *Plant Cell Rep* 39:597–607. <https://doi.org/10.1007/s00299-020-02517-z>
- Wohl J, Petersen M (2020b) Phenolic metabolism in the hornwort *Anthoceros agrestis*: 4-coumarate CoA ligase and 4-hydroxybenzoate CoA ligase. *Plant Cell Rep* 39:1129–1141. <https://doi.org/10.1007/s00299-020-02552-w>
- Wu YF, Zhao Y, Liu XY, Gao S, Cheng AX, Lou HX (2018) Isolation and functional characterization of hydroxycinnamoyltransferases from the liverworts *Plagiochasma appendiculatum* and *Marchantia paleacea*. *Plant Physiol Biochem* 129:400–410. <https://doi.org/10.1016/j.plaphy.2018.06.019>
- Yang Q, Reinhard K, Schiltz E, Matern U (1997) Characterization and heterologous expression of hydroxycinnamoyl/benzoyl-CoA:anthranilate N-hydroxycinnamoyl/benzoyltransferase from elicited cell cultures of carnation, *Dianthus caryophyllus* L. *Plant Mol Biol* 35:777–789. <https://doi.org/10.1023/A:1005878622437>
- Yang Q, Trinh HX, Imai S, Ishihara A, Zhang L, Nakayashiki H, Tosa Y, Mayama S (2004) Analysis of the involvement of hydroxyanthranilate hydroxycinnamoyltransferase and caffeoyl-CoA 3-O-methyltransferase in phytoalexin biosynthesis in oat. *Mol Plant-Microbe Interact* 17:81–89. <https://doi.org/10.1094/MPMI.2004.17.1.81>
- Zenk MH (1979) Recent work on cinnamoyl CoA derivatives. Recent advances in phytochemistry: biochemistry of plant phenolics, vol 12. Plenum Press, New York