Molecular and biochemical investigations of genes and enzymes involved in the phenolic metabolism of the hornwort *Anthoceros agrestis*

Dissertation zur Erlangung des

Doktorgrades der Naturwissenschaften

(Dr. rer. nat.)

dem Fachbereich der Pharmazie

der Philipps-Universität Marburg

vorgelegt von

Julia Wohl

aus Lauingen

Marburg/Lahn 2020

Dem Fachbereich Pharmazie der Philipps-Universität Marburg als Dissertation

eingereicht am 18.03.2020

Erstgutachterin: Prof. Dr. Maike Petersen

Zweitgutachter: Prof. Dr. Lars Voll

Tag der mündlichen Prüfung: 28.05.2020

Hochschulkennziffer: 1180

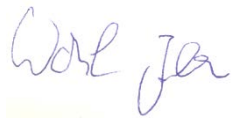
Eidesstattliche Erklärung

Ich versichere, dass ich meine Dissertation

Molecular and biochemical investigations of genes and enzymes involved in the phenolic metabolism of the hornwort *Anthoceros agrestis*

selbständig ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen bedient habe. Alle vollständig oder sinngemäß übernommenen Zitate sind als solche gekennzeichnet.

Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.



Marburg, den 17.03.2020

Acknowledgements

I hereby like to thank my PhD supervisor Prof. Dr. Maike Petersen for her support during this thesis. You gave me the freedom to choose the direction of my projects on my own, allowing me to learn a lot of new skills. I will miss the many pleasant conversations, your guidance, encouragement and advice throughout these exciting and joyful years.

I am very grateful to Prof. Dr. Lars Voll for being the co-examiner of my thesis and accompanying me throughout my entire scientific career.

Furthermore, I would like to thank the current and former colleagues of the Petersen working group for their support and help, especially my life partner Dr. Lennart Poppe. I will miss all the fun times together.

Although not directly involved in my projects, I would also like to express my gratitude to the employees of the Institute of Pharmaceutical Biology and the former and current colleagues of the Li working group for the good companionship and joyful atmosphere.

I would like to thank the working groups of Prof. Dr. Schlitzer and Prof. Dr. Rensing, especially Dr. Pierre-Francois Perroud and Rolf Emmerich, who welcomed me into their groups and gave me the opportunity to pick up new methods.

I am particularly thankful to my family, especially my parents, for always supporting me and investing a lot of time and effort in my education over the years. I am very grateful that I was able to pursue my scientific goals, because without you this would not have been possible.

Publications and Presentations

2020 **Publication** Plant Cell Reports

Phenolic metabolism in the hornwort *Anthoceros agrestis*: 4-Coumarate CoA Ligase and 4-hydroxybenzoate CoA ligase. Wohl J. & Petersen M. doi.org/10.1007/s00299-020-02552-w

2020 **Publication** Plant Cell Reports 39:597–607

Functional expression and characterization of cinnamic acid 4-hydroxylase from the hornwort *Anthoceros agrestis* in *Physcomitrella patens.* Wohl J. & Petersen M. doi.org/10.1007/s00299-020-02517-z

2019 **Scientific lecture** International Plant Science Conference, *Botanikertagung*, Rostock

Alternative expression systems for plant cytochrome P450s – cinnamic acid 4-hydroxylase from the hornwort *Anthoceros agrestis*

2018 **Scientific lecture** Section of plant natural products of the German Botanical Society, Warberg

Recombinant expression of enzymes potentially involved in rosmarinic acid biosynthesis from *Anthoceros agrestis*

2017 **Poster presentation** International Plant Science Conference, *Botanikertagung*, Kiel

Different expression approaches for cytochrome P450 enzymes from *Anthoceros agrestis*

2016 **Scientific lecture** Section of plant natural products of the German Botanical Society, Meisdorf

Hydroxylation reactions in the biosynthesis of rosmarinic acid in the hornwort *Anthoceros agrestis*

Table of content

I Introduction

1 The conquest of the land by plants

Without land plants (formally called embryophytes) there would be no agriculture, paper, plant fibers, crucial pharmaceuticals and other key industrial products. All land plants are derived from a common ancestor - green algae. To be more precise, freshwater living charophyte green algae were the ancestors of land plants. They had a haplontic life cycle with a multicellular haploid gametophyte and a unicellular diploid zygote (Wickett et al. 2014; Kenrick 2017). Over 450 million years ago these charophyte green algae were able to evolve, allowing them to remain hydrated and still reproduce while colonizing the land and eventually gain access to subsurface water (Gensel and Edwards 2001; Kenrick et al. 2012; Delwiche and Cooper 2015). Now they were able to use sunlight unfiltered by water and CO_2 during photosynthesis, which reduced the concentration of CO_2 in the atmosphere, forming the foundation of our terrestrial ecosystems (Lenton et al. 2012). And although they were probably not the first photosynthetic organisms to conquer terrestrial surroundings, they evolved quickly and occupied the harshest environments (Wellman et al. 2015).

Three of the lineages of charophytes, namely Charophyceae, Coleochaetophyceae and Zygnematophyceae were long discussed to be the ancestor of embryophytes. Charophyceae, Coleochaetophyceae and embryophytes share complex characteristics, for example plasmodesmata in the gametophyte, the ability of apical growth with branching, oogamous sexual reproduction and parental retention of the egg. Charophyceae, Coleochaetophyceae, embryophytes and some members of the Zygnematophyceae all possess a phragmoplast, a collection of actin microfilaments and microtubules which direct formation of the cell plates during cytokinesis (Fowke and Pickett-Heaps 1969; Pickett-Heaps and Wetherbee 1987; Galway and Hardham 1991). Nevertheless, most probably the Zygnematophyceae are the closest ancestors to land plants. This was based on phylogenomic analyses of plastid genomes, ribosomal protein genes and other nuclear genes (Becker 2013; Wickett et al. 2014; Delwiche and Cooper 2015; Rensing 2018). The question remains, whether all characteristics mentioned above

missing in Zygnematophyceae (for example apical growth and branching) evolved independently within Charophyceae, Coleochaetophyceae and embryophytes or were already present in a common ancestor and were subsequently lost in most lineages within the Zygnematophyceae. However, Zygnematophyceae had two decisive survival advantages which helped them conquer the terrestrial environment: Because of their structural simplicity and small size they were able to grow fast and even a thin film of water was sufficient for survival. Moreover, Zygnematophyceae experienced some molecular changes, for example the transfer of some plastid-encoded genes to the nucleus. This event helped establishing the embryoplast, as well as the functionalization into special plastid types (e.g. amyloplasts or chromoplasts) as they are known from higher plants (Delwiche and Cooper 2015; de Vries et al. 2016; Rensing 2018)

An essential process in the origin and evolution of land plants was the expansion of the diploid phase (sporophyte) from a single cell (as in charophytic algae) to a multicellular free-living organism (as in vascular plants) with different tissues and organs like leaves, roots, stems, stomata and a vascular system. This was accompanied by the reduction of the haploid phase (gametophyte) from bryophytes to seed plants (Qiu et al. 2012). The first group of organisms evolved through early events in the diversification of embryophytes were bryophytes comprising mosses, liverworts and hornworts. They all share distincitive characteristics: for example, the diploid sporophyte is nutrient-dependent on the dominant haploid gametophyte. Moreover, they are not able to develop flowers, instead they all are monosporangic. Although all bryophytes already possess the required enzymes for its biosynthesis, true lignin is absent. Instead some forms of ‘pre-lignin’ were found in *Physcomitrella patens* (Renault et al. 2017a; Alber et al. 2019; Renault et al. 2019).

The phylogenetic classification of Anthocerotophyta, Bryophyta and Marchantiophyta is often discussed, but yet it is still unresolved. Since dating via fossils was hardly possible (Troitsky et al. 2007), data were collected mostly by DNA sequence comparisons. Recent publications came to varying results regarding their phylogenetic placement under different points of view. The four most strongly supported possible phylogenetic trees are displayed in Figure 1.

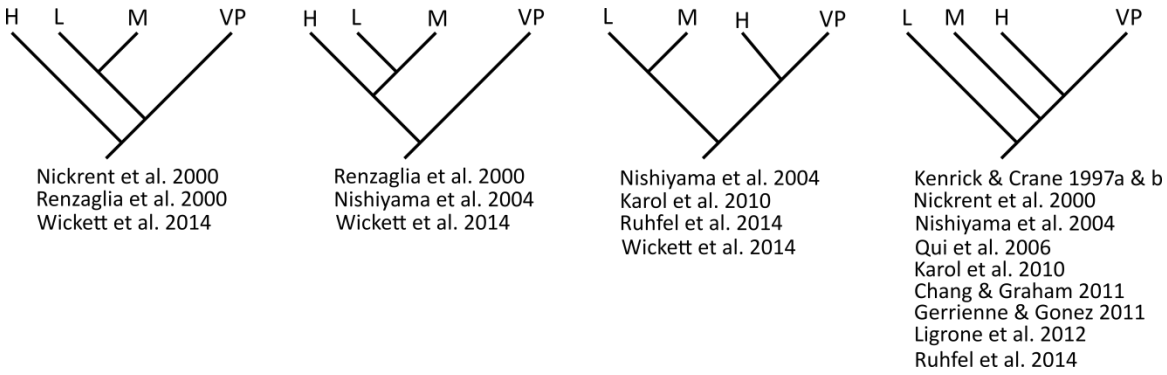


Figure 1 Hypothesized evolution of bryophytes. The abbreviations are H - hornworts, L - liverworts, M - mosses and VP - vascular plants.

On the one hand, Nickrent et al. (2000), Nishiyama et al. (2004), Qiu et al. (2006), Karol et al. (2010), Chang and Graham (2011), Ruhfel et al. (2014) and Wickett et al. (2014) built their suggestions by comparisons of DNA sequences. On the other, Kenrick and Crane (1997ab), Renzaglia et al. (2000), Gerrienne and Gonez (2011) and Ligrone et al. (2012a) investigated on phenotype level. Most authors share the opinion that liverworts and mosses form a clade. Puttick et al. (2018) rejected the hypothesis of the last tree. On the basis of maximum-likelihood analyses and Bayesian supertree estimation they support bryophyte monophyly, which was also proposed by Zhang et al. (2020).

In all proposed phylogenetic trees, the bryophytes are the sister group to vascular plants. This also can be shown with enzymes from mosses which perform a similar task in higher plants. For example MpBNB from *Marchantia polymorpha* is involved in the development of gametangia, while the ortholog gene pair AtBNB1/2 from *Arabidopsis thaliana* controls pollen generative cell formation (Yamaoka et al. 2018). The formation of rhizoids in bryophytes and the formation of root hairs in flowering plants are both under the control of the transcription factor bHLH (Tam et al. 2015; Proust et al. 2016).

During evolution the sporophyte gained more and more importance. While the sporophyte was only a small stalked capsule in bryophytes, the sporophytes of lycopods, ferns, gymnosperms, and flowering plants are free-living, large and leafy. Therefore bryophytes are an intermediate between their green algae ancestor and the vascular plants (Ligrone et al. 2012a; Qiu et al. 2012; Wickett et al. 2014). The Rhynie chert fossils

of early land plants were classified as the closest members to vascular plants. Like ferns, these plants had independent gametophytes and sporophytes and both life sections possessed tissues such as rooting structures, a vascular system and stomata. Moreover the gametophyte was smaller and axes ended in cup-shaped structures which contained the sexual organs. Some of them already had tracheids. In vascular plants the sporophyte became the dominant stage, whereas the gametophyte lost tissues like the vascular system and stomata (Kenrick 2017).

2 Phenolic metabolism and its occurrence in land plants

With the conquest of land, plants were confronted with new challenges (for example UV radiation, loss of water and pathogens). Some of those challenges were answered by the biosynthesis of specialized compounds. Among them were phenolic compounds derived from the shikimic acid and phenylpropanoid metabolism. The phenolic metabolism is supposed to be one of the oldest and most common pathways of specialized metabolites in the plant kingdom (Ferrer et al. 2008; Emiliani et al. 2009; Vogt 2010; Cheynier et al. 2013). In several ways phenolic compounds were essential:

- Phenolic compounds can protect the plant against UV radiation and UV-induced damage and mutations can be reduced (e.g. flavonoids, hydroxycinnamic esters).
- By the incorporation of phenolic substances into water-repellent compounds water loss by transpiration decreases (cutin, suberin).
- Moreover, phenolic compounds are required for cell wall reinforcing structures, so the plant can build water-conducting tracheids and grow geotropically (lignin and lignin-like compounds).
- Phenolic compounds act as pathogen defence and herbivore repellents because of their antibiotic properties (e.g. hydroxycinnamic acid derivatives).

2.1 Biosynthesis of phenylpropanoids and derived compounds

Many phenolic compounds are synthesized via the phenylpropanoid pathway (Figure 2). Starting with the aromatic amino acid L-phenylalanine, the enzyme L-phenylalanine ammonia-lyase (PAL) catalyzes the non-oxidative deamination of phenylalanine to *t*-

cinnamic acid. *t*-Cinnamic acid is then hydroxylated by the cytochrome P450 enzyme cinnamic acid 4-hydroxylase (C4H) and the resulting 4-coumaric acid is activated for further reactions by 4-coumarate CoA-ligase (4CL) as coenzyme A thioester. 4-Coumaric acid can alternatively be directly synthesized from L-tyrosine by L-tyrosine ammonia-lyase (TAL). 4-Coumaroyl-CoA is the key compound for the biosynthesis of secondary metabolites, e.g. coumarins, flavonoids, isoflavonoids, lignans and lignin, cutin and suberin as well as hydroxycinnamic acid esters and amides (e.g. chlorogenic acid and rosmarinic acid) (Vogt 2010). 4-Coumaric acid as well as other hydroxycinnamic acids can also be activated by the formation of a glucose ester. This reaction is catalyzed by a UDP-glucose:hydroxycinnamic acid glucosyltransferase (Petersen 2016).

In this work, the further transformation of 4-coumaroyl-CoA to 4-coumaroyl esters or amides is of particular interest. This is accomplished by a hydroxycinnamoyltransferase (HCT) belonging to the superfamiliy of BAHD acyltransferases. The introduction of a hydroxyl group in 3-position of the 4-coumaroyl ester/amide is catalyzed by a CYP98 hydroxylase (C3H) (Alber et al. 2019). By this way, for example caffeoylshikimate is formed, a precursor for monolignols. Also 4-coumaroylquinic acid is hydroxylated to chlorogenic acid, which is the best known UV-protecting compound from the phenolic metabolites (Clé et al. 2008; Comino et al. 2009). The coupling of 4-coumaric acid from 4-coumaroyl-CoA and 4-hydroxyphenyllactic acid and successive hydroxylation in positions 3 and 3´ results in rosmarinic acid. Rosmarinic acid is accumulated up to 5 % of the cell dry weight of *Anthoceros agrestis* suspension cultures (Petersen 2013). Hydroxylation of the amides 4-coumaroylthreonic acid and 4-coumaroyl-(3-hydroxy)anthranilic acid to caffeoylthreonic acid and caffeoyl-(3-hydroxy)anthranilic acid is supposed to be involved in synthesis of compounds forming the cuticule in *Physcomitrella patens* (Renault et al. 2017a; Alber et al. 2019).

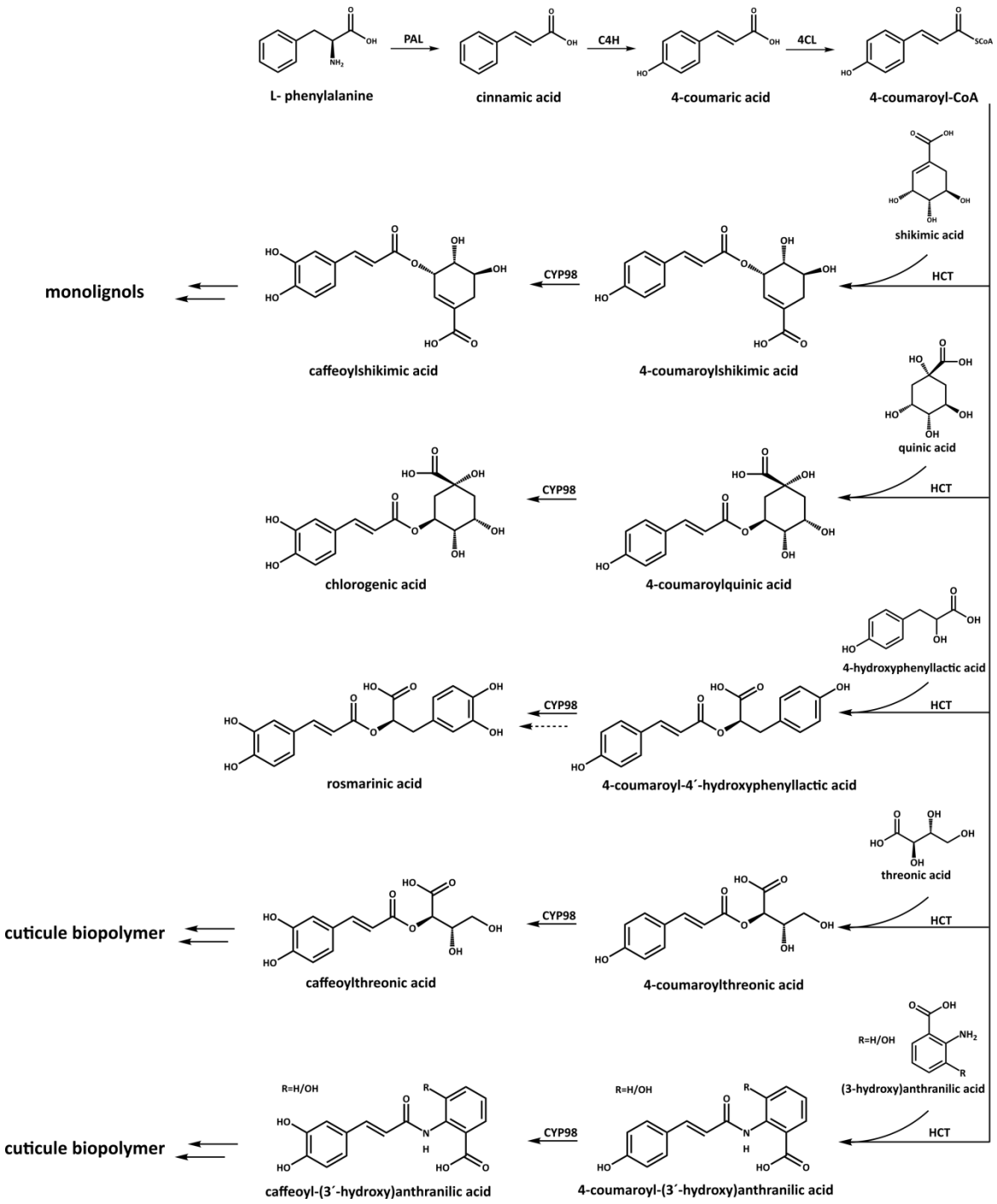


Figure 2 Scheme of phenylpropanoid metabolism and derived metabolic pathways. The abbreviations are PAL - phenylalanine ammonia-lyase, C4H - cinnamic acid 4-hydroxylase, 4CL - 4-coumarate CoA-ligase, HCT - hydroxycinnamoyltransferase, CYP98 - hydroxycinnamoyl ester/amide 3-hydroxylase.

2.2 Phenylpropanoid pathway genes and phenolic compounds found in lower plants

The ability to produce phenolic compounds is wide-spread in the plant kingdom. Niklas et al. (2017) postulated, that prokaryotes and fungi already had the capability to produce cinnamic acid, whereas 4-coumaroyl-CoA was first detected in algae. Orthologues of core phenylpropanoid pathway genes were investigated in Chlorophyta (*Ostreococcus* and *Chlamydomonas*), Streptophyta (*Mesostigma, Klebsormidium, Nitella, Coleochaete, Spirogyra, Chara*), Embryophyta (*Marchantia* and *Physcomitrella*), and Tracheophyta (*Selaginella, Salvinia, Picea* and *Arabidopsis*) (Figure 3). Hornworts were not included in the surveys. None of the genes encoding proteins from the core phenylpropanoid pathway were found in Chlorophyta. Orthologue genes for PAL were identified in Streptophyte algae and many land plants. While C4H was missing in Streptophyte algae, C4H was found in Embryophyta. The presence of 4CL was found in all land plants and *Mesostigma, Klebsormidium* and *Spirogyra.* Genes encoding proteins with the putative ability to produce 4-coumaroylesters or amides (HCT) were found in three of the six Streptophyte algae and all land plants. CYP98 was found in *Klebsormidium* and Embryophyta. Most of these results were based on sequence similarities and there is often no proof of the functionality and catalytic activity of these genes (Labeeuw et al. 2015; de Vries et al. 2017; Renault et al. 2019).

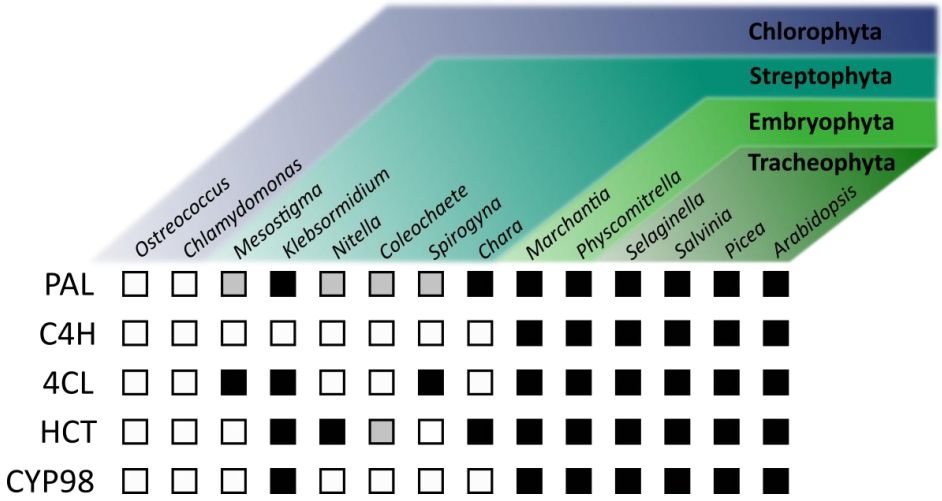


Figure 3 Presence of phenylpropanoid pathway genes in different plant lineages. Present genes are boxed black, no clear presence grey and absent genes are boxed white (de Vries et al. 2017; Renault et al. 2019). The abbreviations are PAL - phenylalanine ammonia-lyase, C4H - cinnamic acid 4-hydroxylase, 4CL - 4-coumarate CoA-ligase, HCT - hydroxycinnamoyltransferase, CYP98 - hydroxycinnamoyl ester 3-hydroxylase.

Surprisingly, H- G- and S-units of lignin were identified in the red alga *Calliarthron cheilosporioides* (Martone et al. 2009); this was seen as a convergent evolution to land plants. The units are named after the substitution pattern of the aromatic ring, namely those of 4-coumaryl (H), coniferyl (G), and sinapyl (S) alcohols. In green algae only weak signals were detected through antibody labeling for G- and GS- units in *Nitella flexilis* (Ligrone et al. 2008). Members of the Zygnematophyceae were able to accumulate phenolic compounds under UV-irradiation. One of these compounds was putatively a galloyl derivative (Aigner et al. 2013; Pichrtova et al. 2013). Flavonoids were putatively present in Cyanobacteria, Ochrophyta, Haptophyta, Rhodophyta and Chlorophyta (Goiris et al. 2014).

Liverworts are rich in phenolic compounds. H- G- and S-units of lignin were identified in *Marchanita polymorpha* (Espineira et al. 2011). Furthermore *Marchantia* and other liverworts contain flavonoids (e.g. luteolin and apigenin) and are rich in bis(bibenzyls) (Asakawa et al. 2000 & 2013). Tsugawa et al. (2019) detected anthocyanins and flavones. Simple phenolic acids and phenolic esters, for example chlorogenic acid, were

identified by Murugan and Krishnan (2013). Wounding increased PAL expression and the accumulation of luteolin and apigenin and the bisbibenzyl isoriccardine C (Yoshikawa et al. 2018). Moreover luteolin glycoside increased after UV irradiation in *Marchantia polymorpha*. Also 4-coumaric, ferulic acid as well as soluble apigenin and luteolin derivatives were accumulated in *Plagiochila asplenioides* (Soriano et al. 2019).

G- and GS-lignin like polymers were detected in *Dawsonia superba*, *Dendroligotrichum dendroides*, *Polytrichum formosum*, *Notoligotrichum minimum*, *Takakia ceratophylla* and mostly in *Sphagnum cuspidatum* (Ligrone et al. 2008). Flavonoids and flavonoid glycosides were also found in Bryophyta (Asakawa et al. 2013). In *Physcomitrella patens* the pre-lignins caffeoylthreonic acid and caffeoylanthranilic acid are involved in gametophyte development and cuticle formation (Renault et al. 2017a & 2019).

G- and GS-lignin like units were mainly found in the spores and the pseudoelaters of hornwort species (*Megaceros*; Ligrone et al. 2008). Until now, flavonoids have not been detected in hornworts (Yonekura-Sakakibara et al. 2019). In contrast to liverworts and mosses they contain lignans (Zinsmeister et al. 1991). Moreover other interesting phenolic compounds, like rosmarinic acid, (hydroxy-)megacerotonic acid, anthocerotonic acid, anthocerodiazonin and rosmarinic acid 3´-*O*-β-D-glucoside were detected (Takeda et al. 1990; Vogelsang et al. 2006). Other phenolic esters and amides were described in the PhD thesis of Trennheuser (1992).

3 Objects of study

3.1 *Anthoceros agrestis*

The hornwort *Anthoceros agrestis* belongs to the family Anthocerotaceae, class Anthocerotopsida, division Anthocerotophyta. It occurs worldwide, mostly on the northern hemisphere, on humid, damp and bright areas. The haploid gametophyte is dominant (as with all bryophytes) and formed as a flattened thallus, lacking specialized tissue. It is monoecious, meaning both male (antheridia) and female (archegonia) reproductive organs develop on the same thallus. After fertilization, a diploid horn-shaped sporophyt develops, which produces spores via meiosis. A unique feature of the hornworts is that the basal meristem of the sporophyte remains active throughout the

life of the sporophyte (Renzaglia et al. 2000). *Anthoceros* is usually associated with the cyanobacterium *Nostoc* in a symbiotic relationship (Renzaglia et al. 2008). In 2015, the genome sequence of *Anthoceros agrestis* was published (Szövényi et al. 2015). So far, the hornwort has the smallest genome of all bryophytes (around 83 Mbp) (Leitch and Bennett 2007). It is of interest, since mostly *Marchantia* (liverwort) (Bowman et al. 2017; Soriano et al. 2018) and *Physcomitrella* (moss) (Rensing et al. 2008; Lang et al. 2018) are investigated and hornworts are often left out in studies concerning land plant evolution. However, hornworts have some unique features among byrophytes, some closer related to algae and others to vascular plants. While the first zygotic division is longitudinal in hornworts (like in vascular plants), it is transverse in liverworts and mosses (Renzaglia 1978; Ligrone et al. 2012b). On the other hand, they are the only land plant having chloroplasts with pyrenoids like algae (Duckett and Renzaglia 1988; Villarreal and Renner 2012).

3.2 *Physcomitrella patens*

The moss *Physcomitrella patens* belongs to the family Funariaceae, class Bryopsida, division Bryophyta. It occurs mostly in Europe, North America and Japan and grows besides lakes, rivers, ditches and on damp open ground in fields. Germinating spores first produce a protonemal stage. This protonema initially consists of chloronemal cells, filled tightly with large chloroplasts. Apical cell division creates further branches of chloronemal cells and some of them evolve into caulonemal cells. These cells contain fewer and less developed chloroplasts. Most of the side branches from subapical caulonemal filaments develop into chloronemal filaments, however some side branches develop into either caulonemal filaments or leafy shoots (= gametophores) (Cove 2005). As *Anthoceros* also *Physcomitrella* is monoecious and self-fertilization is usual. After fertilization the diploid embryo develops into a sporophyte consisting a spore capsule containing around 4000 spores (Engel 1968; Schween et al. 2003; Cove 2005).

The plant is nowadays often used as a model organism for differentiation analysis and the investigation of gene function in molecular and cellular development, particularly because of its dominant haploid life cycle and the high and efficient rates of homologous recombination. In 2008, the genome of *Physcomitrella patens* was published (Rensing et

al. 2008). Thus targeted knockouts are possible in order to analyse gene functions (Cove 1992; Cove and Knight 1993; Schaefer and Zrÿd 1997; Cove et al. 2009d). Moreover, *Physcomitrella* is comparable to higher plants in terms of gene content, expression, and regulation (Reski 1999).

Since the homologous expression of proteins in the moss is very cheap, it is used in high volume bioreactors for the production of complex biopharmaceutical products (Reski et al. 2015; Reski et al. 2018). The advantage as an expression host is furthermore that it has the ability to perform posttranslational protein modifactions, such as the formation of disulfide bridges and complex glycosylations, similar to higher plants (Koprivova et al. 2003).

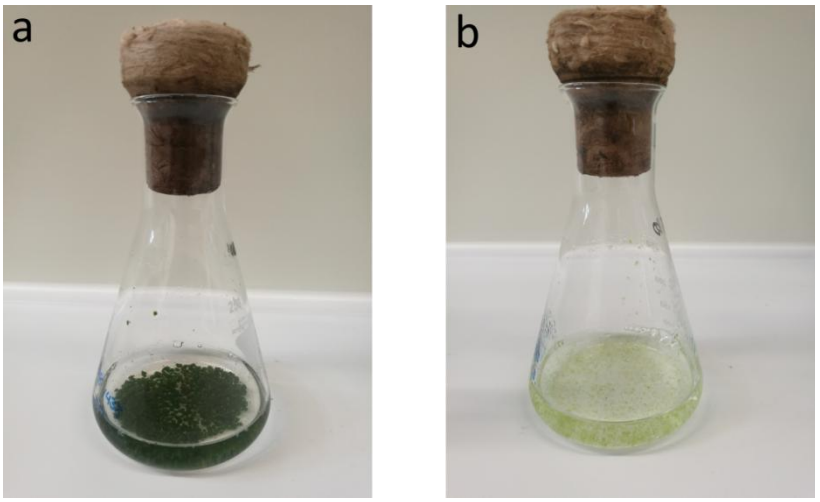


Figure 4 Plant cell cultures of *Anthoceros agrestis* and *Physcomitrella patens*. a Suspension culture of *Anthoceros* in CBM-medium; **b** Protonema culture of *Physcomitrella* in BCD-medium.

4 Relevant enzyme families

4.1 Cytochrome P450 monooxygenases

Cytochrome P450 enzymes are heme-containing enzyme that are found in almost all organisms. The name is derived from the absorption maximum at 450 nm in the CO-difference spectrum of the enzyme in its reduced state. CO binds to the active center with a higher affinity than molecular oxygen shifting the absorption maximum of the heme to 450 nm. This inhibits the reaction, but the bond can be released by exposure to

light (λ = 450 nm) (Omura and Sato 1964a; Omura and Sato 1964b). The nomenclature of P450s is based on similarity of the amino acid sequence, assigning proteins with more than 40% sequence identity into the same family and proteins with more than 55% sequence identity into the same subfamily (Nelson 2006a). Amino acid sequences of known P450 enzymes can vary substantially, but nevertheless the protein structure and folding is highly conserved (Werck-Reichhart and Feyereisen 2000). They are classified as heme-thiolate proteins with a protoporphyrin IX as a prosthetic group (Figure 5).

Usually, cytochrome P450s act as terminal monooxygenases using molecular oxygen, which transfer one oxygen atom to the substrate while the other one is reduced to water. The two required electrons are mostly provided by NADPH (or NADH) and transferred via a NADPH:cytochrome P450 reductase (CPR). This enables the P450s to carry out the following reactions: hydroxylations, N-, O- and S-dealkylations, C-C- or C-O-C- coupling, sulphoxidations, epoxidations, deaminations, desulphurations, dehalogenations, peroxidations, and N-oxide reductions (Sono et al. 1996; Guengerich 2001; Bernhardt 2006). In general, the following equation represents a hydroxylation reaction of monooxygenases:

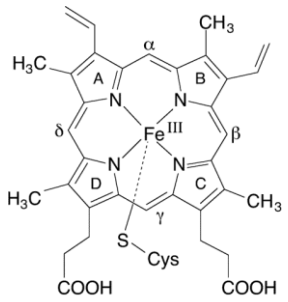


Figure 5 Structure of an iron-(III) protoporphyrin-IX linked with a proximal cysteine ligand (from Meunier et al. 2004)**.**

$$RH + O_2 + NADPH + H^+ \rightarrow ROH + H_2O + NADP^+$$

In some cases cytochrome P450 act as oxidases. Here the oxygen serves as an acceptor molecule and is reduced to water or hydrogen peroxide. The following equation can be formulated for oxidase reactions:

$$R + O_2 + 4\ H^+ \rightarrow R + 2\ H_2O$$

While most prokaryotic P450s are soluble, eukaryotic enzymes are mostly bound to membranes of the endoplasmic reticulum (ER), mitochondria or chloroplasts. Depending on the proteins involved in the electron transfer to the P450 enzymes, nine classes of P450s were identified. Most of the plant cytochrome P450 belong to the second class,

depending on a CPR, with both enzymes being anchored in the ER membrane (Hannemann et al. 2007; Gamenara et al. 2013).

In plants cytochrome P450 enzymes play an important role, since they are involved in the biosynthesis of e.g. phenylpropanoids, alkaloids, benzoxazinones, flavonoids, fatty acids, glucosinolates, sterols, terpenoids, plant hormones (e.g. gibberellins), pigments (e.g. anthocyanidins) and signaling molecules (e.g. jasmonic acid, salicylic acid). Furthermore they are needed for the detoxification of herbicides and other xenobiotics (Durst and O´Keefe 1995; Kahn and Durst 2000; Schuler and Werck-Reichhart 2003).

The amino acid sequences of these enzymes often share highly conserved regions. The membrane anchor is formed by the first 20-30 predominantly hydrophobic amino acids, followed by the proline rich region with the consensus sequence [(P/I)PGPX(G/P)XP]. This region is important for enzyme stability, because it provides flexibility between the membrane anchor and the globular domain. In addition, the protein requires an oxygen binding site [(A/G)GX(D/E)T(T/S)]. The ERR triad is involved in fixing the heme in its position and to ensure stability of the conserved core structures (Hasemann et al. 1995). The triad is defined by the amino acids Glu and Arg of the K-helix consensus [K**E**TL**R**] and the Arg in the [PE**R**F] consensus sequence. The most highly conserved region is the heme-binding domaine [FXXGXRX**C**XG] with the cysteine residue acting as a thiolate-ligand for the prosthetic heme-group (Halkier 1996; Werck-Reichhart et al. 2002).

The catalytic cycle of P450s is described as follows (Figure 6): In the first step, the water molecule bound to the heme group (**A**) is cleaved from the substrate and the iron (III) changes to a *high-spin* state (**B**). This is followed by the first electron transfer from NADPH via the prosthetic groups FAD and FMN of the reductase and the iron (III) is reduced to iron (II) (**C**). After binding molecular oxygen, a ferrous dioxygen complex (**D**) is formed, which is a good electron acceptor. The transfer of the second reducing equivalent [electron] of NADPH by the CPR to the monooxygenase forms the ferric-peroxo anion species (**E**). This is quickly converted into the ferric-hydroperoxide species (= compound 0) by protonation (**F**). The further protonation by the oxidized flavoprotein and the cleavage of a water molecule leads to the oxo iron complex, also called compound I (**G**). The exact mechanism of substrate hydroxylation is still discussed

controversely (Werck-Reichhart and Feyereisen 2000; Meunier et al. 2004; Shaik et al. 2005). Three possible principles are under consideration. The first is called oxygen rebound mechanism, where the cleavage of a hydrogen atom from the substrate leads to the formation of a radical. In a second step the Fe^{IV}-OH intermediate binds via the oxygen atom to the substrate and the substrate is hydroxylated, separating it from the iron complex. The concerted mechanism works without the formation of a radical intermediate and is based on the use of ultrafast radical clocks. The last mechanism is called two-state reactivity, a theoretical mechanism calculated with density functional theory. Hydrogen abstraction, alkyl rotation and alkyl rebound are the key steps for alkane hydroxylation and moreover a change between a doublet spin state (*low spin*) to a quartet spin state (*high spin*). After substrate release the oxyferryl cation returns to its initial condition with help of water (Werck-Reichhart and Feyereisen 2000; Meunier et al. 2004; Shaik et al. 2005).

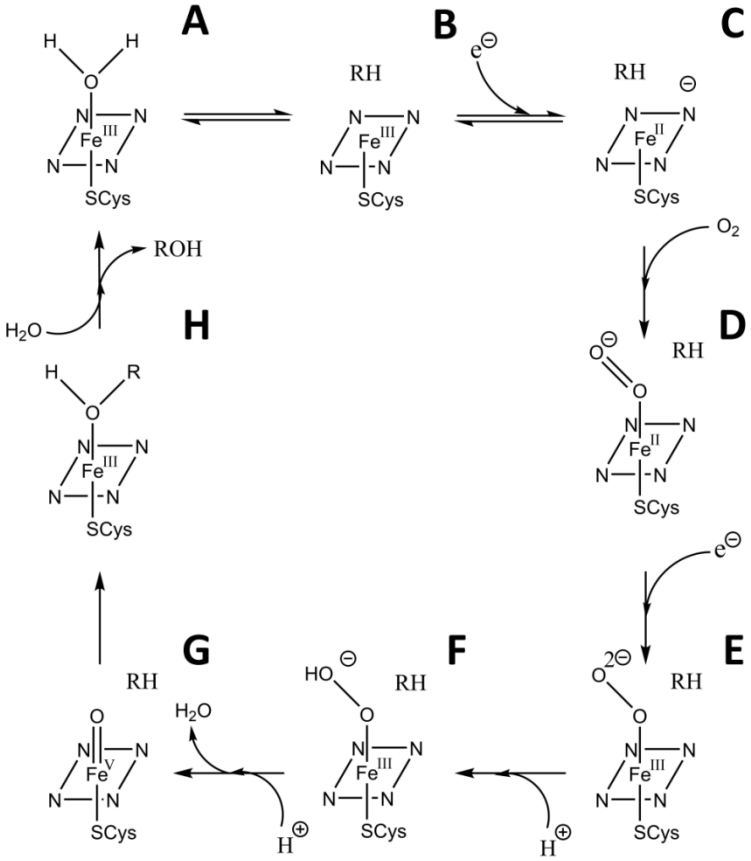


Figure 6 Schematic representation of the catalytic cycle of a cytochrome P450 (Werck-Reichhart and Feyereisen 2000; Meunier et al. 2004; Shaik et al. 2005).

4.1.1 Cinnamic acid 4-hydroxylase (C4H)

One of the best characterized cytochrome P450 hydroxylases from higher plants is cinnamic acid 4-hydroxylase (Werck-Reichhart 1995). The enzyme catalyzes the formation of 4-coumaric acid from *trans*-cinnamic acid. In the 1960s, the enzyme was first discovered in spinach and pea seedlings (Nair and Vining 1965; Russell and Conn 1967; Russell 1971) and is therefore one of the first described plant cytochrome P450 enzymes. Up to now, genes encoding a C4H have been isolated and characterized from liverworts (*Marchantia* and *Plagiochasma*) and several vascular plants (e.g. *Arabidopsis, Brassica, Gossypium, Petroselinum, Populus, Ruta* and *Scutellaria*) (Kawai et al. 1996; Urban et al. 1997; Koopmann et al. 1999; Gravot et al. 2004; Chen et al. 2007; Xu et al. 2010; Ni et al. 2014; Liu et al. 2017). Several other genes were annotated as C4H through genome projects or expression studies (Rensing et al. 2008; Kim et al. 2013; Shen et al. 2013 and many more). In the moss *Physcomitrella patens* six genes potentially encoding C4H were identified (Phytozome: Pp3c25_10190V3.1, Pp3c4_21680V3.1, Pp3v3_17840V3.1, Pp3c13_14870V3.1, Pp3c16_23740V3.1 and Pp3c12_6560V3.1) and activity was proven for Pp3c4_21680V3.1 after recombinant expression in yeast (Renault et al. 2017b). Recently, the genome of *Anthoceros angustus* was published and three genes putatively encoding a C4H were identified (Zhang et al. 2020). In 2003, Petersen characterized a C4H from cell cultures of the hornwort *Anthoceros agrestis* Paton. C4H could not be detected in streptophytes, therefore an alternative pathway to 4-coumaric acid is suggested (Labeeuw et al. 2015; de Vries et al. 2017; Renault et al. 2019).

4.1.2 Hydroxycinnamoyl ester/amide 3-hydroxylase

Homologs of CYP98 are involved in the hydroxylation of many different hydroxycinnamic esters or amides. Aspects of the evolution of CYP98 in different plants concerning their gene copy number and substrate specificity was published just recently (Alber et al. 2019). It is proposed, that all CYP98 are derived from a common ancestor. While there is only one gene copy number in bryophytes, lycophytes, monilophytes and gymnosperms, several copies can be found in angiosperms. This was explained by independent gene duplications and losses. Also the substrate specificity changed.

CYP98s from a moss (*Physcomitrella patens*), a lycopod (*Selaginella moellendorffii*) and a fern (*Pteris vittata*) preferred phenolamides instead of esters. All of them favoured 4-coumaroylanthranilate as substrate. CYP98A104 from *P. vittata* produced tri-caffeoylspermidine. Although *Selaginella* is able to produce lignin, CYP98A38 did not accept 4-coumaroylshikimate. Nevertheless, the lycopod is able to produce lignin independent from CYP98 by a distinct CYP788A1, which catalyzes the aromatic 3- and 5-hydroxylations of the free 4-coumaroyl aldehyde or 4-coumaryl alcohol (Weng et al. 2010). CYP98A34 from *P. patens* also accepted several other substrates like 4-coumaroylshikimate, isoprenyl-4-coumarate, 4-coumaroyltyramine and 4-coumaroylthreonate (Renault et al. 2017a; Alber et al. 2019).

Angiosperm CYP98 mostly favour 4-coumaroylshikimate (Schoch et al. 2001; Franke et al. 2002; Mahesh et al. 2007; Moglia et al. 2009; Sullivan and Zarnowski 2010; Pu et al. 2013) as a substrate, in order to mediate the production of G- and S-lignin units (Schoch et al. 2006; Adams et al. 2019). Some angiosperms moreover accumulate large amounts of chlorogenic acid (Schoch et al. 2001; Franke et al. 2002; Mahesh et al. 2007; Morant et al. 2007; Karamat et al. 2012). In *Arabidopsis thaliana* another homolog gained activity for tri-hydroxycinnamoylspermidines, which is involved in pollen coat and pollen wall biosynthesis (Matsuno et al. 2009; Xu et al. 2014; Liu et al. 2016). CYP98A14 from *Coleus blumei* produces rosmarinic acid by 3- and 3´-hydroxylation of 4-coumaroyl-4'-hydroxyphenyllactate (Eberle et al. 2009). This substrate was also accepted by enzymes from *Ocimum basilicum* (Gang et al. 2002) and CYP98A6 from *Lithospermum erythrorhizon* accepted coumaroyl-4'-hydroxyphenyllactic acid (Matsuno et al. 2002).

4.2 NADPH:cytochrome P450 reductase

Due to different environmental influences, plants must be able to produce a large number of different natural compounds. Many of them are synthesized with the involvement of heme-containing cytochrome P450 enzymes (Halkier and Møller 1991; Kahn and Durst 2000; Morant et al. 2003; Nielsen and Møller 2005). Depending on which partners are involved in the electron transfer cytochrome P450s are divided into nine classes (Hannemann et al. 2007). In plants and most other eukaryotes the P450s belong to the second class, which consists of two membrane bound proteins, the P450

and a NADPH:cytochrome P450 oxidoreductase (CPR). The reaction of the P450 requires two electrons, which are transferred from NADPH via the prosthetic groups FAD and FMN of the CPR. In 2007, Denisov et al. postulated a conformational change of the P450 after substrate binding towards the CPR to ensure efficient electron donation. The first isolated CPR from yeast was initially referred to as cytochrome *c* reductase, because of its ability to reduce cytochrome *c* (Haas et al. 1940). In the 1960s, it was already clear that P450s are their redox partners and that they are anchored in the endoplasmic reticulum (ER) (Williams and Kamin 1962; Lu and Coon 1968; Lu et al. 1969). The first plant CPRs found to be localized in the ER were from *A. thaliana* and hybrid poplar (Urban et al. 1997; Ro et al. 2002).

The *N*-terminal region of the CPR protein is particularly hydrophobic and serves as a membrane anchor. Based on their conservation and length of the *N*-terminal amino acid sequence, the CPRs were divided into two classes (Ro et al. 2002). While class I CPRs have short *N*-terminal sequences, class II CPRs have longer *N*-terminal sequences. The FMN- (*N*-terminal) and the NADPH/FAD- (*C*-terminal) binding domains are supposed to result from the gene fusion of an ancestral FMN-containing bacterial flavodoxin (Fld) and a FAD-containing ferredoxin-$NADP^+$ reductase (FNR) (Porter and Kasper 1986). This was also supported by the functional expression of individual domains of CPR (Smith et al. 1994; Hodgson and Strobel 1996). First crystal structures from rat liver, spinach FNR and Fld from *Desulfovibrio vulgaris* also revealed a high structural similarity within the organisms (Watenpaugh et al. 1973; Karplus and Daniels 1991; Wang et al. 1997). The crystal structure of rat liver CPR confirmed the presence of the four indiviual domains, a FMN-binding domain, a linker domain, the FAD-binding domain and the NADPH-binding domain (situated between the two conserved areas of the FAD-binding domain). A *C*-terminally located aromatic tryptophan residue in the NADPH-binding domain regulates the binding and release of NADPH (Hubbard et al. 2001). After binding of NADPH the entire conformation of the protein changes, allowing the electrons to be transported from FAD to FMN and later to the P450. Change of the conformation is accomplished by the linker domain and as a result the FMN region gains flexibility (Narayanasami et al. 1995; Vermilion and Coon 1978). CPR and P450s interact electrostatically. While the

region near FMN is charged negatively, the area near the heme of the P450 is charged positively (Ravichandran et al. 1993; Zhao et al. 1999).

In vascular plants the number of CPR homologs varies from one to three (Benveniste et al. 1991; Meijer et al. 1993; Shet et al. 1993; Koopmann and Hahlbrock 1997; Rosco et al. 1997; Urban et al. 1997; Mizutani and Ohta 1998; Ro et al. 2002; Eberle et al. 2009). Only in the moss *Physcomitrella patens* four different homologs were found (Rensing et al. 2008; Jensen and Møller 2010). Surprisingly, in many plants only one homolog was induced after elicitor exposure, although more than one homolog of CPR exist (Meijer et al. 1993; Koopmann and Hahlbrock 1997; Ro et al. 2002; Ohta and Mizutani 2004). This also implies the possibility, that some plant P450s only interact with specific CPRs (Jensen and Møller 2010).

4.3 4-Coumaric acid CoA-ligase

4-Coumaric acid CoA-ligase (4CL), involved in the activation of hydroxycinnamic acids, is a member of acyl-activating enzymes (AAEs). AAEs are associated with the biosynthesis or degradation of various metabolic products such as lipids, amino acids, sugars and natural compounds. They are responsible for the activation of carboxylic acids. They activate carboxylic acids through a two-stage reaction (Shockey et al. 2003). In the first step, ATP is cleaved and the substrate is activated by binding to AMP. The resulting unstable adenylate intermediate remains bound in the active center. In the second step, free electrons of the sulphur group of the acyl acceptor cause a nucleophilic attack on the carbon atom of the carboxyl group, thereby AMP is released and a thioester is formed. In general, the reaction can be described by the following equations:

(1) $\text{Acid} + \text{ATP} \rightarrow \text{acyl-AMP} + PP_i$

(2) $\text{Acyl-AMP} + \text{CoA} \rightarrow \text{acyl-CoA} + \text{AMP}$

There are several AAEs that allow the activation of various substrates like acetate and fatty acids of different length (Watkins 1997), benzoic acid derivatives (Chang et al. 1997), cinnamic acid derivatives (Ehlting et al. 1999) or citrate, malate and malonate (An et al. 1999).

Mostly coenzyme A is the acceptor of the acyl moiety, but others are also described. For example some bacterial enzymes use amino acids for the biosynthesis of antibiotics. These enzymes form an amino acid-adenylate and then transfer the amino acid residue to the thio-group of an enzyme-bound 4-phosphopantheteine moiety (Conti et al. 1997). Another acceptor is molecular oxygen (Staswick and Tiryaki 2004). An example is the firefly luciferase producing the chemiluminescent oxyluciferin by oxidative decarboxylation of luciferin-AMP (Conti et al. 1996).

In plants, most of these enzymes are soluble, but some are also membrane-bound, for example in peroxisomes (Schneider et al. 2005). They all have two highly conserved regions. The AMP-binding domain, also called box I, is represented by the consensus sequence [STG][STAG]G[ST][STEI][SG]X[PASLIVM][KR]. The consensus sequence of box II is [GEICIRG] (Shockey et al. 2003). The two conserved motifs correspond to the core motifs A3 and A6 of the phenylalanine-activating subunit (PheA) of gramicidin synthetase 1 (Conti et al. 1997; Stuible and Kombrink 2001).

The production of 4-coumaroyl-CoA is the last step in the core phenylpropanoid pathway and 4-coumaroyl-CoA is the precursor for many secondary metabolites. The reaction is catalyzed by a 4-coumaric acid CoA-ligase (4CL). In addition caffeic acid, cinnamic acid, ferulic acid, isoferulic acid and sinapic acid are often accepted. Nevertheless, the enzymes differ substantially in their substrate specificities and preferences (Lindermayr at al. 2002). In 2003, Schneider et al. found 12 amino acids proposed to function as the 4CL substrate specificity code. Moreover, two of these 12 amino acids were essential for acceptance of sinapic acid. In plants the protein is found soluble in the cytosol.

Usually several 4CL paralogs can be found in plants and they mostly have different expression patterns (Renault et al. 2019). A phylogenetic analysis of almost 200 4CL sequences revealed a duplication event occuring in seed plants before the split of gymnosperms and angiosperms (Li et al. 2015). Homologs of 4CL are already present in the genomes of green algae, red algae, glaucophytes, diatoms, dinoflagellates, haptophytes, cryptophytes and oomycetes (Labeeuw et al. 2015). Four genes from *Physcomitrella patens* encoding a putative 4CL were heterologously expressed and

converted preferably 4-coumaric acid, but also accepted cinnamic acid, caffeic acid and ferulic acid (Silber et al. 2008). Further characterized 4CLs can be found, for example, in *Arabidopsis thaliana* (Ehlting et al. 1999), *Nicotiana tabacum* (Lee and Douglas 1996), *Petroselinum hortense* (Knobloch and Hahlbrock 1977), as well as *Populus trichocarpa x deltoides* and *Populus tremuloides* (Hu et al. 1998). One of the four *Arabidopsis* paralogs gained a distinct role in flavonoid and sinapoyl malate biosynthesis (Li and Nair 2015).

5 Aim of this work

So far, *Anthoceros agrestis* has mostly not been considered in studies of the phenylpropanoid pathway. Aim of this work was to get a closer insight into the first steps of polyphenol biosynthesis, especially for the two cytochrome P450 enzymes C4H and CYP98 as well as the CoA-Ligase 4CL.

C4H is one of the best known plant cytochrome P450 enzymes, but until now, hornworts were the only clade among the bryophytes, in which no gene encoding a C4H had been identified. So far, only genes from the liverworts *Marchantia* and *Plagiochasma* have been isolated and characterized (Liu et al. 2017). Six potential candidate genes are present in *Physcomitrella*, but only one of them has been isolated and actively expressed (Renault et al. 2017b).

Suspension cultures of *Anthoceros* can accumulate up to 5 % of rosmarinic acid in the dry weight (Vogelsang et al. 2006). Therefore it was interesting to know, whether a CYP98 is present and able to convert 4-coumaroyl-4’-hydroxyphenyllactate to rosmarinic acid or caffeoyl-4’-hydroxyphenyllactate. On the other hand, CYP98 from lower plants like *Physcomitrella*, *Selaginella* and *Pteris* favoured phenolamides like 4-coumaroylanthranilate (Alber et al. 2019). If 4-coumaroyl-4’-hydroxyphenyllactate is not accepted, production of rosmarinic acid would probably differ from biosynthesis in Lamiaceae, since persumably there is only one copy of CYP98 in lower plants.

For the heterologous expression of CYP enzymes coexpression with a plant NADPH:cytochrome P450 oxidoreductase (CPR) was shown to be benefical. Therefore a CPR from *Anthoceros agrestis* had to be discovered.

Activation of hydroxycinnamic acids by a 4CL displays the last step of the core phenylpropanoid pathway. 4CL homologs can be found throughout the plant kingdom. Often 4CLs differ in substrate specificity/preference and there is mostly more than one paralog in plants. In *Physcomitrella,* four genes were heterologously expressed (Silber et al. 2008).

With the help of biochemical and molecular methods enzymes encoding a C4H, CYP98, CPR and 4CL had to be identified. Potential genes had to be introduced into appropriate expression vectors. Baker's yeast was initially intended to be used as the expression organism for the membrane-bound enzyme system consisting of cytochrome P450 and NADPH:cytochrome P450 reductase (CPR). *E. coli* was chosen as the expression host for the recombinant CoA-ligase. Subsequently, the expressed proteins had to be characterized biochemically.

II Material and Methods

1 General methods

1.1 Maintenance of plant cell cultures

Suspension cultures of *Anthoceros agrestis* Paton were cultivated as described in Petersen (2003). The suspension cultures were transferred into fresh CB-M medium (50 ml) with help of a perforated spoon once every week, using 5.5 g cell wet weight. The cultures were kept at 25 °C under continuous light on a gyratory shaker at 100 rpm.

Physcomitrella patens (provided by Dr. Stefan Martens, Fondazione Edmund Mach, Italy) was either cultivated as gametophores on solid BCD medium or as protonemata in liquid BCD medium (Cove et al. 2009a). Gametophores were cultivated on solid BCD medium in Petri dishes and subcultured every 3 months. Protonema tissue was kept in 50 ml BCD in 250 ml Erlenmeyer flasks. Every 7 days the tissue was disrupted with a sterilised tissue blender (Omni International) for 30 s and 4 ml of the old suspension was transferred to 50 ml fresh BCD medium and then kept on a gyratory shaker (100 rpm). All cultures were kept at 25 °C under continuous light.

1.2 Culture characterisation

1.2.1 Fresh weight and pH

Transformed and untransformed *Physcomitrella patens* cultures were cultivated for 21 days and different parameters (fresh weight and medium pH) were checked every second or third day. For this 2 ml of a 7-day old disrupted protonemata culture were transferred to 50 ml fresh BCD medium in 250 ml Erlenmeyer flasks and then kept at 25 °C under continuous light on a gyratory shaker (100 rpm). For the complete characterisation over 21 days around 60 flasks were incubated. Since the fresh weight in one flask was too low for determination in the first days of the characterisation, two or more flasks were combined. The measured fresh weight was then divided by the number of flasks used. Around 10 flasks were taken directly after addition of the old culture. This corresponded to day 0. Plant tissue and medium were separated with the

help of a suction filter and a vacuum pump. After determination of the fresh weight the tissue was stored at -80 °C for further use (Chapter II.2.1). In addition to the fresh weight, the pH of the medium was measured after calibration with adequate buffers solutions at room temperature with a pH-electrode.

1.2.2 Content of phenolic compounds

The total content of phenolic compounds was analyzed in untransformed and transformed *P. patens* cultures. Over 21 days *P. patens* was cultivated in 50 ml BCD medium at 25 °C under continuous light on a gyratory shaker (100 rpm). Every seven days tissue samples were collected and stored at -80 °C. Samples were incubated twice for 10 min at 80 °C in an ultrasonic bath after addition of 100 µl 70 % EtOH per 20 mg fresh weight. After centrifugation for 10 min at 13000 g, 25 µl of the supernatant was mixed with 475 µl water. 250 µl Folin-Ciocalteu reagent (Merck) was added and incubated for 15 min at room temperature. 2.5 ml alkaline reagent (0.1 N NaOH, 2 % Na_2CO_3) was added and the solution was once more incubated for 15 min at room temperature. The absorbance was measured photometrically at 760 nm (Jennings 1981). Different caffeic acid solutions (0, 0.25, 0.375, 0.5, 0.75 and 1 mg/ml in 70 % ethanol) were used to obtain a calibration curve using 25 µl caffeic acid solution instead of plant extract.

1.3 Phylogenetic analysis

For phylogenetic analysis, the translated amino acid sequence of the respective gene was aligned with amino acid sequences from different species using the Maximum Likelihood method of the MEGA X software package. The robustness of the branch structure was evaluated with a bootstrap analysis (1000 replicates). Amino acid sequences for the phylogenetic tree were accessed from the BRENDA enzyme database and suitable publications.

2 Molecular biology

2.1 RNA extraction

Total RNA extracts of *Anthoceros agrestis* and *Physcomitrella patens* were prepared using a phenol-chloroform extraction essentially according to Chomczynski and Sacchi (1987). Before extraction all solutions and materials were autoclaved twice or heated at 200 °C for 2 h. Furthermore, all steps were performed with gloves. Plant material was frozen and pulverized in liquid nitrogen in a mortar. Approximately 50 mg of this powder were mixed with 500 µl (1 equivalent) `Solution D´ (4 M guanidinium thiocyanate solution in 25 mM citrate buffer pH 7.0 and 0.5 % laurylsarcosine) and incubated at room temperature. Then 50 µl (0.1 equivalent) 2 M sodium acetate pH 4 was added and the solution was mixed gently. After addition of 500 µl (1 equivalent) phenol saturated with citrate buffer pH 2 the tube was again gently mixed and 100 µl ice-cold chloroform was added. The solutions were mixed for 10 sec, incubated on ice for 15 min and centrifuged at 12000 g at 4 °C for 15 min. 400 µl of the upper phase (aqueous supernatant) was transferred to a new reaction tube and 1 equivalent (around 400 µl) of ice-cold 2-propanol was added. The RNA was precipitated for 15 min at -20 °C with occasional inversion. The tube was centrifuged for 10 min at 12 000 g and 4 °C and the supernatant was removed. The pellet was washed with 500 µl 70 % ethanol and the tube was centrifuged for 5 min at 7500 g and 4 °C. After removal of the supernatant the pellet was additionally washed with 500 µl 96 % ethanol and again centrifuged at 7500 g for 5 min and 4 °C. After removal of the supernatant the remaining pellet was dried at 37 °C. The RNA was dissolved in 20 µl demineralized H_2O at 50 °C. The integrity of the RNA was controlled on an agarose gel (Chapter II.2.5) and the amount and purity was determined photometrically. For this 2 µl RNA solution were mixed with 98 µl water and the absorption at 260 and 280 nm was measured. Samples with high amounts and purity ($A_{260/280}$~ 2.0) were used for further experiments. The samples were stored at -20 °C.

2.2 Digestion with DNase

RNA samples from *Physcomitrella patens* later used for RT qPCR were digested with DNase (Thermo Scientific) to remove DNA contaminants. 5 µg RNA was mixed with 5 µl

10x buffer and 5 µl DNase and the samples were adjusted to a total volume of 50 µl with doubly autoclaved H_2O. The samples were incubated at 37 °C for 30 min and then 50 µl water were added. The reaction was stopped and the DNAse was removed from the RNA sample by phenol-chloroform extraction as already described in chapter II.2.1. Samples were stored at -20 °C.

2.3 cDNA synthesis

2.3.1 cDNA synthesis for PCR

cDNA for PCR amplification of partial and full-length sequences was synthesized with the RevertAid First Strand cDNA Synthesis Kit (Fermentas).

2 µg RNA were mixed with 1 µl oligo$(dT)_{18}$ primer and the samples were adjusted to a total volume of 12 µl with sterile H_2O. The tubes were mixed gently, centrifuged for a few seconds and incubated for 5 min at 65 °C. After cooling the samples on ice 4 µl reaction buffer, 1 µl RiboLock RNAse inhibitor and 2 µl 10 mM dNTP mix were added. After gentle mixing and short centrifugation, the tubes were incubated at 37 °C for 5 min. Finally, 1 µl RevertAid reverse transcriptase was added and the tube was incubated for 1 h at 42 °C. The reaction was stopped by heating for 5 min at 70 °C. Samples were stored at -20 °C.

2.3.2 cDNA synthesis for RACE-PCR

For RACE-PCR, cDNA was prepared with the SMARTer®RACE Kit (Clontech). In contrast to the `normal´ cDNA the reverse transcriptase attaches a poly(C) tail to the 3'-end of the cDNA (corresponding to the 5'-end of the template RNA) for the preparation of 5´-RACE cDNA. These nucleotides serve as a template for a G-rich second primer with an adapter sequence. For 3´-RACE cDNA the same adapter sequence is attached by a primer binding to the poly(A)-tail of each mRNA. Thus, the RACE-ready cDNA contains the same adapter region, either at the 5'- or 3'-end.

At first a buffer mix was prepared containing 4 µl 5x First-Strand buffer, 0.5 µl DTT (100 mM) and 1 µl dNTP mixture (20 mM). Then the following reagents were combined:

1 µg RNA was mixed with either 5´-CDS Primer A or 3´-CDS Primer A and the volume was adjusted to 11 µl (5´) or 12 µl (3´) with water. The tubes were mixed and centrifuged shortly. Then the RNA was denatured by incubation at 72 °C for 3 min and afterwards cooled down on ice and centrifuged for 10 sec at 14000 g. To the 5´-RACE cDNA synthesis reaction 1 µl of the SMARTer II A oligonucleotide was added. This was not necessary for 3´-RACE cDNA synthesis. A master mix was prepared at room temperature with 5.5 µl buffer mix (prepared at the beginning), 0.5 µl RNase inhibitor (40 U/µl) and 2 µl SMARTScribe reverse transcriptase (100 U). This master mix was added to the denatured RNA. The samples were gently mixed by pipetting, spun down and incubated for 90 min at 42 °C. After the reaction was stopped by heating for 10 min at 70 °C, the samples were diluted with 240 µl Tricine-EDTA buffer. The cDNA was stored at -20 °C.

2.3.3 cDNA synthesis for RT qPCR

0.5 µg DNase-digested RNA was transcribed using the qScript™ cDNA SuperMix kit (Quanta). 4 µl qScript cDNA Supermix (5x) was mixed with 0.5 µg RNA and water in a total volume of 20 µl. The samples were incubated for 5 min at 25 °C, 30 min at 42 °C and finally for 5 min at 85 °C. After cooling on ice, 80 µl water was added. The samples were stored at -20 °C.

2.4 Polymerase chain reaction

2.4.1 Amplification of partial and full-length sequences

Partial (missing the unknown 5´- and/or 3´-end of the gene) and full-length sequences were amplified by standard PCR. In contrast to partial sequences, full-length sequences were amplified using a polymerase with high-fidelity and proofreading function.

For the amplification of partial DNA sequences the following standard PCR protocol was used with GoTaq polymerase (Promega) in a final volume of 25 µl:

1 µl	cDNA (template)
5 µl	GoTaq buffer (5x)
0.5 µl	dNTP mix (10 mM)
3 µl	$MgCl_2$ (25 mM)
0.5 µl	primer 1 (10 or 100 mM)
0.5 µl	primer 2 (10 or 100 mM)
0.1 µl	GoTaq polymerase (5 U/µl)
14.4 µl	H_20

Full-length sequences were either amplified with Phusion® High-Fidelity DNA Polymerase (2 U/µl; NEB) or Pfu DNA Polymerase (Promega). The suitable buffer, $MgCl_2$ concentration and total volume were adjusted accordingly. Usually, the following protocols were used:

Phusion® High-Fidelity DNA Polymerase:

1 µl	cDNA (template)
10 µl	HF buffer with $MgCl_2$ (10x)
1 µl	dNTPs (10 mM)
1 µl	primer f (10 mM)
1 µl	primer r (10 mM)
2 µl	Phusion® Polymerase
ad 50 µl H_2O	

Pfu DNA Polymerase:

2 µl	cDNA (template)
10 µl	Pfu buffer (10x)
1 µl	dNTPs (10 mM)
6 µl	$MgCl_2$ (25 mM)
1 µl	primer f (10 mM)
1 µl	primer r (10 mM)
0.2 µl	Pfu Polymerase
ad 50 µl H_2O	

The PCR program was the following:

	Denaturation	Annealing	Extension
1st cycle:	95 °C, 120 sec		
2nd-40th cycle:	95 °C, 15 sec	GSP T_m -5 °C, 45 sec	70 °C, 1 min
End:	cooling at 6 °C		

If the expected amplicon was longer than 1500 bp, the extension time was increased by 1 min/1000 bp. T_m is the melting temperature of the primer. The melting point was calculated with a tool on the Eurofins website (https://www.eurofinsgenomics.eu). In case of additional method-specific overhangs (e.g. restriction sites or codons for a *C*-terminal 6xHis-tag), these additional nucleotides were not taken into account. Usually an annealing temperature of T_m – 5 °C was chosen as starting point with a gradient of 5 °C above and beyond with steps of 1-2 °C.

The PCR product was analyzed by agarose gel electrophoresis (Chapter II.2.5). Fragments of the expected size were cut out and purified by gel extraction (Chapter II.2.6).

2.4.2 RACE-PCR

RACE-PCR was performed using the SMARTer®RACE Kit (Clontech). After synthesis of RACE-ready cDNA a long universal primer (in the Universal primer mix (UPM)) can bind to the attached adapter sequence in the 5´ or 3´ end. This long universal primer contains further nucleotides, which enable a nested-PCR with a short universal primer. Additionally a gene specific primer (GSP) is used. The gene specific primer had preferably a T_m > 70 °C which enabled the use of touchdown PCR. Moreover, some primers had a 15 bp overhang for successful In-Fusion cloning into the vector pRACE.

Each RACE-PCR reaction had a total volume of 50 µl and was composed as follows:

25 µl	SeqAmp PCR Buffer
1 µl	SeqAmp DNA Polymerase
2.5 µl	5´/3´-RACE-ready cDNA
5 µl	Universal Primer Mix (10x) (UPM)
1 µl	GSP (10 µM)
15.5 µl	H_2O

The PCR programs were the following:

Touchdown PCR (if GSP T_m > 70 °C):

	Denaturation	Annealing	Extension
1st - 5th cycle:	94 °C, 30 sec	72°C, 3min	
6th - 10th cycle:	94 °C, 30 sec	70 °C, 30 sec	72 °C, 3 min
11th - 40th cycle:	94 °C, 30 sec	68 °C, 30 sec	73 °C, 3 min
End:	cooling at 6 °C		

Alternative PCR program (if GSP T_m < 70 °C)

	Denaturation	Annealing	Extension
1st – 40th cycle:	94 °C, 30 sec	68 °C, 30 sec	72 °C, 3 min
End:	cooling at 6 °C		

For both PCR programs the extension time was increased by 1 min/1000 bp, if the expected fragment was longer than 3000 bp. The amplicons were checked by agarose gel electrophoresis. If instead of a clear band a smear was seen, nested PCR was performed. For this purpose the PCR product was diluted 1:50 with water and 5 µl of this solution was used for nested PCR. Moreover, the included Universal Primer Short (UPM short) was used instead of the Universal primer mix (UPM). If available a second gene specific primer, binding further inside in the sequence, was applied. Otherwise the same GSP was used again. For amplification the alternative PCR program was chosen.

2.4.3 Colony-PCR

Colony-PCR was used to determine the presence or absence of insert DNA in plasmid constructs introduced into *S. cerevisiae*. A colony was picked after transformation (Chapter II.2.14), suspended in 100 µl 200 mM LiAc, 1 % SDS solution and incubated for 5 min at 70 °C. This step causes the release of DNA from the cells. For DNA precipitation 300 µl EtOH (96 %) was added and mixed vigorously. The samples were centrifuged for 3 min at 15000 g and the resulting pellet was washed with 300 µl 70 % EtOH. Afterwards the tubes were centrifuged again for 3 min at 15000 g and the pellet was dissolved in 50 µl H_2O. 2 µl of the solution was used in a standard PCR reaction with GoTaq

polymerase and specific primers for the respective DNA sequence. Amplification was checked by agarose gel electrophoresis.

If there was no PCR product, plasmid DNA was isolated from *S. cerevisiae* and introduced into *E. coli*. Then the plasmid was multiplied, isolated and checked by restriction digest.

2.4.4 RT qPCR

In this work, expression studies were performed with DNase-digested cDNA obtained from RNA samples isolated from AaC4H-transformed *P. patens*. The plant material was collected in the process of culture characterisation (Chapter II.1.2).

Primers were designed based on AaC4H and the two highly expressed putative *P. patens* C4Hs (PpC4H_1 = Pp3c25_10190V3.1 and PpC4H_2 = Pp3c4_21680V3.1) to obtain fragments with sizes of 212 to 219 bp. As a reference gene serine threonine protein phosphatase 2a regulatory subunit (St-P 2a) was used. St-P 2a is involved in the regulation of signalling processes. All amplified fragments were at first checked by agarose gel electrophoresis and sequenced after insertion into *E. coli*. Other tested reference genes (actin 1, actin 5, ubiquitin-conjugating enzyme E2 and elongation factor 1α) did either show an unspecific fragment after agarose gel electrophoresis, or the melting curve revealed more than one amplified product. Most primers for the reference genes were designed after Le Bail et al. (2013).

RT qPCR was performed in a 96 well thermocycler (PikoReal96 from Thermo Scientific) with the PerfeCTa SYBR Green SuperMix (Quanta). Each reaction consisted of 5 µl cDNA, 6.5 µl 2x PerfeCTa SYBR Green SuperMix and primers with a concentration of 192 nM (PpC4H1 and PpC4H2) or 385 nM (AaC4H and St-P 2a) in a volume of 13 µl. The following PCR program was used:

	Denaturation	Annealing	Extension
1st cycle:	95 °C, 2 min		
2nd – 50th cycle:	95 °C, 15 sec	52 °C, 45 sec	68 °C, 60 sec
melting curve:	50 °C to 95 °C		
End:	cooling at 6 °C		

For the determination of the amplification efficiency 20 µl of all isolated cDNAs were pooled and serial dilutions ranging from non-diluted to 1:256 were made. These diluted cDNAs were used in RT qPCR reactions. Each dilution and primer combination was used three times to account for technical varations. The resulting Cq values were plotted on the y-axis against the decadic logarithm of the dilutions. Due to the resulting slope of the linear regression, the amplification efficiency was calculated by the formula: $10^{(-1/slope)}$. Since in a PCR reaction the concentration of the desired fragment should theoretically be doubled with each cycle, the ideal amplification efficiency is 2 (≙ 100 %). Efficiencies between 1.9 -2.1 (≙ 90 – 110 %) were intended.

To account for biological variation, RNA was prepared twice for every time point and cDNA was prepared twice from the DNase-digested RNA samples with higher integrity (A_{260}/A_{280}) and once from the samples with lower integrity. Replica with cDNA of every isolated time point was measured twice on a 96 well plate and all measurements were repeated twice to account for technical variation. Water was used instead of cDNA as a negative control. Moreover, data for the reference gene were acquired simultaneously in every run.

The resulting Cq values displayed the level of gene expression of each candidate gene for all samples. Average ΔCq values were generated by comparison with the reference gene St-P 2a. To calculate fold-change of expression for each gene with respect to the average values of day 0, the method described by Pfaffl (2001) was used:

$$\frac{E_{C4H}^{\Delta CP_{C4H}\,(\text{day } 0 - \text{day X})}}{E_{\text{ST-P 2a}}^{\Delta CP_{\text{ST-P 2a}}\,(\text{day 0-day X})}}$$

Furthermore, the relative expression in comparison to the gene with the lowest overall expression (PpC4H_1) was calculated. This ratio was calculated according to the following formula:

$$E_{\text{St-P 2a}}^{\text{Cq day X}} / E_{C4H}^{\text{Cq day X}}$$

For better comparability PpC4H_1 was set to 1.

2.5 Agarose gel electrophoresis

DNA and RNA samples were separated by agarose gel electrophoresis. 1.4 g agarose were dissolved in 200 ml 1x TAE buffer (20 mM acetic acid, 1 mM Na_2-EDTA, 40 mM Tris) by heating in a microwave. Approximately 50 ml of the still hot solution was used to cast a gel, 1 µl of a 1 % ethidium bromide solution was added and mixed well. The electrophoresis chamber contained 1x TAE-buffer. The applied samples were mixed with a 6x loading dye (0.03 % bromophenol blue, 0.03 % xylene cyanol, 60 mM EDTA in 60 % glycerol) before loading. As a marker, 3 µl of a DNA ladder mix (GeneRuler™ or O´GeneRuler™) was used in one slot. The DNA or RNA was separated at 110 V for about 30-45 min until the lower bromophenol blue band from the loading dye had reached two thirds of the gel. The gel was checked under a blue-green LED light using an amber filter.

2.6 Purification of DNA fragments from agarose gels

Once the amplification by PCR or the digestion with restriction enzymes (Chapter II.2.17) led to a DNA fragment of the expected size, the respective bands were cut out of the gel. To isolate the DNA from the gel a NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel) was used.

At first, the cut out gel matrix was dissolved, using binding buffer NT in a ratio of 200 µl buffer/100 mg gel matrix. The samples were incubated at 50 °C until the gel was liquefied (5-10 min). Then the solution was applied on a NucleoSpin silica filter and spun down for 1 min at 11000 g. The flow-through was discarded. 600 µl wash buffer (NT3) was added and again the flow-through was discarded after centrifugation at 11000 g for 1 min. To dry the silica membrane, the columns were additionally centrifuged for 2 min at 11000 g. At last, 20-50 µl H_2O was added into the center of the silica filter. The flow-through was centrifuged into a fresh reaction tube after an incubation time of 5 min. The silica filter was cleaned with H_2O for further use.

2.7 Ligation

2.7.1 UA-ligation into pDrive

The principle of the UA-ligation is the attachment of an adenosine nucleotide at the end of the PCR amplicon. This adenosine can interact with an overhanging uridine nucleotide of the linear plasmid, thus facilitating the incorporation. The disadvantage of this method is that half of the constructed vectors result in a reverse incorporation of the insert, since both ends of the insert and the plasmid are identical.

UA-ligation into the vector pDrive was performed with the PCR cloning kit from Qiagen. 2 µl DNA fragment-containing solution was gently mixed with 0.5 µl pDrive cloning vector and 2.5 µl 2x Ligation Master Mix in a total volume of 5 µl. The samples were incubated at 4-10 °C overnight in a refrigerator or at 15 °C for at least 1 h. Finally, the reaction was stopped by denaturation of the ligase for 5-10 min at 65 °C. The ligated plasmid was then introduced into *E. coli* (Chapter II.2.13).

2.7.2 Ligation into restriction sites

DNA fragments previously cut with restriction endonucleases (Chapter II.2.17) were introduced into a plasmid by T4 DNA ligase. To create sticky ends, both plasmid and DNA fragment were cut with the same restriction enzymes. After digestion both, DNA fragment and plasmid were purified, isolated (Chapter II.2.5 and II.2.6) and the concentration was determined photometrically. For ligation, insert and plasmid were used in a ratio of around 5:1 in favor of the insert. The ligation mix had a total volume of 10 µl. First the insert was re-dissolved in the required volume of water.

1 µl	Plasmid
1 µl	T4 Ligase Buffer (10x)
1 µl	T4 DNA Ligase
7 µl	DNA fragment solution/water

The mixture was incubated for at least 20 h at 4 °C and the reaction was stopped at 65 °C for 10 min. For verification and multiplication the plasmid was then introduced into *E. coli* (Chapter II.2.13).

2.8 Fragment insertion by recombination

2.8.1 In-Fusion cloning

Fragments amplified by RACE-PCR were inserted into the vector pRACE by In-Fusion HD cloning, included in the SMARTer®RACE Kit (Clontech). The designed primers used for RACE-PCR had 15 bp overlaps with vector sequences at their 5' ends, necessary for successful In-Fusion cloning. The manufacturer's instructions were followed, using half of the indicated volumes:

0.5 µl	Linearised pRACE vector
3.5 µl	Gel-purified RACE product
1 µl	In-Fusion HD Master Mix

Samples were incubated for 15 min at 50 °C and cooled on ice until transformation of *E. coli* EZ cells (Chapter II.2.13). Later, the RACE-primers were designed without overhangs and the resulting amplicon from RACE-PCR was ligated directly into pDrive (Chapter II.2.7.1).

2.8.2 LR clonase reaction

The LR recombination reaction was performed with the Gateway® Technology kit from Invitrogen. The Gateway® technology is based on the bacteriophage lambda site-specific recombination between an *att*L-containing entry clone (pENTR1A) and an *att*R-containing destination vector to generate an expression clone. This results in an *att*B-containing expression clone and an *att*P-containing by-product after recombination. Using a different set of enzymes from the bacteriophage this reaction is also reversible (BP recombination reaction). This provides a rapid and highly efficient way to transfer DNA sequences into multiple vector systems for protein expression. The protocol was performed according to the manufacturer.

Following components were combined and mixed in a 1.5 ml reaction tube at room temperature:

0.5-5 µl	Entry clone (50-150 ng)
1 µl	Destination vector (150 ng μl^{-1})
2 µl	LR Clonase™ reaction buffer (5x)
ad 8 µl	TE Buffer, pH 8.0

After briefly mixing the LR Clonase™ enzyme mix, 2 µl were added to the components above and mixed well. The reaction was incubated for 1 h at 25 °C. Then 1 µl Proteinase K solution (2 µg μl^{-1}) was added, followed by incubation for 10 min at 37 °C. 3 µl of this solution were used to transform *E. coli* DH5α. The transformation mix was plated on solid LB medium in Petri dishes containing 100 µg ml^{-1} ampicillin, thus only bacteria carrying the expression clone could reproduce, because the entry clone carried only a kanamycin resistance and the destination vector was lethal to the cells because of the ccdB cassette.

2.9 Construction of a chimeric *Coleus blumei* and *Anthoceros agrestis* CPR

Heterologous expression of the native CPR failed in yeast (Chapter III.2.2). To overcome this issue a chimeric CPR was constructed, based on the results of Batard et al. (2000). The group managed the functional expression of monocot CYPs after codon optimization of the 5´-end using a megaprimer. They hypothesized that the high GC-content of monocot genes might be the issue and especially the 5´-region is of great importance. Here, the GC-content of the 5´-end of AaCPR was lowered by exchange of the first 277 bp with the first 382 bp from the CPR from *Coleus blumei* (Eberle et al. 2009). Both sequences were already ligated into the expression vector pYES2 using the same restriction sites. At first, both reductase sequences were cut out of the vector with BamHI and XbaI (Chapter II.2.17). Since AaCPR and CbCPR shared high identities, both had a restriction site for EcoRV between the highly conserved FMN and P450-binding regions (Fig. 7). CbCPR had an additional EcoRV restriction site at the 3'-end of the sequence. After digestion with EcoRV (Chapter II.2.17), AaCPR was split into two fragments with a size of 277 bp and 1772 bp. Digested CbCPR fragments had a size of

382 bp, 1582 bp and 169 bp. The 5'-fragment from CbCPR (382 bp) and the 3'-fragment from AaCPR (1772 bp) were isolated, purified (Chapter II.2.5 and II.2.6) and together ligated back into the expression vector pYES2 (Chapter II.2.7.2).

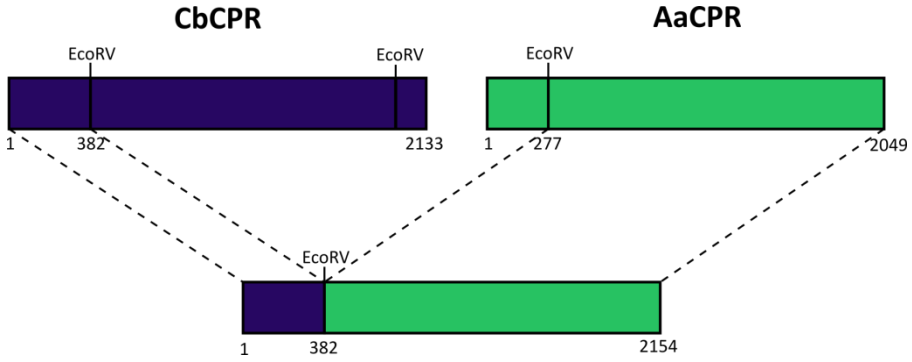


Figure 7 Construct of chimeric CPR from *Coleus blumei* and *Anthoceros agrestis*. EcoRV = restriction site for EcoRV

2.10 Construction of AaC4H-AaCPR fusion proteins for expression in *Escherichia coli*

For the heterologous expression in *E. coli*, AaC4H and AaCPR were fused following already published protocols (Hotze et al. 1995; Leonard et al. 2005; Quinlan et al. 2007; Leonard et al. 2007). AaCPR was modified: the membrane anchor was cut off (prediction with Phyre[2]) and the sequence (AaCPR [38-682]) was equipped with a *C*-terminal 6xHis-tag in front of the stop codon and restriction sites for BcuI (5´) and BamHI (3´). By adding the same BcuI restriction site at the 3'-end, AaC4H nucleotide sequences encoding amino acids 1 to 524 (*fus1[1-524]*) or 7 to 524 (*fus2[7-524]*) were amplified. *Fus2* was additionally fused to a sequence encoding seven amino acid residues of the mammalian peptide ε (*fus2*[ε:7-524]). The AaC4H gene sequence encoding amino acids 39 to 524 was also fused to the sequence encoding the mammalian peptide ε (*fus3*[ε:39-524]). All fus-constructs were linked to AaCPR [38-682] and ligated into pET15b. XbaI, used in the 5´-region of AaC4H fusion proteins, cuts off the ribosome binding site (rbs) in pET15b. This remained unnoticed during first experiments. After this issue had been noticed, new 5´-primers were designed, adding the same rbs as in pET15b (Fig. 8). All sequences were amplified by PCR with the full-length cDNA in pDrive as a template and ligation was carried out with T4 DNA ligase after a restriction digest and purification of the expected DNA fragments.

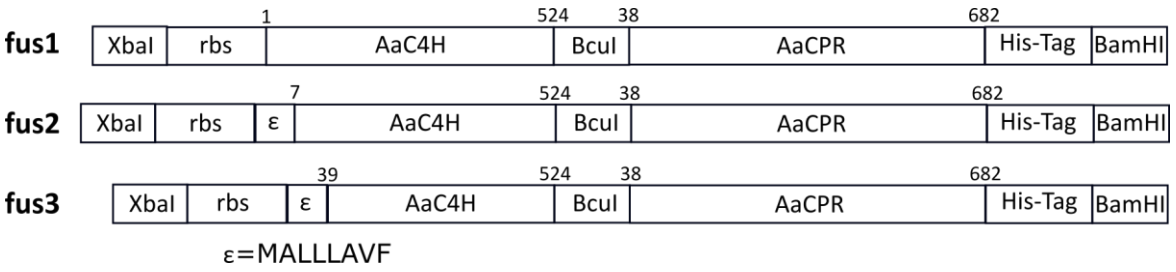


Figure 8 Fusion constructs for expression of AaC4H fused to AaCPR in *E. coli.* XbaI = restriction site for XbaI, rbs = ribosome binding site, ε = 8 amino acid residues (MALLLAVF) of mammalian peptide ε, BcuI = restriction site for BcuI, His-Tag = coding sequence for 6x His, BamHI = restriction site for BamHI.

2.11 Construction and plasmid preparation for transformation of *Physcomitrella patens*

For the heterologous expression of proteins in *P. patens* the respective genes were equipped with a sequence encoding six histidines in front of the stop codon. Usually the full-length sequence in pDrive was used for amplification. The full-length PCR product was integrated into the entry vector pENTR™1A into different restriction sites. By LR recombination (Chapter II.2.8.2), the sequence was transferred to the expression vector pTHUBiGate (kindly provided by Prof. Dr. S. Rensing, Philipps-Universität Marburg). After transformation and plasmid replication in *E. coli* DH5α, the sequence was checked (Seqlab) and the plasmid afterwards linearized with SwaI (20 µg plasmid DNA in each sample). After gel purification, contaminants were removed by ethanol precipitation (Crouse and Amorese 1987). For this purpose the samples were adjusted with 10 % sodium acetate solution (3 M). Then 1 equivalent 2-propanol was added and the samples were mixed. After incubation for at least 15 min at -20 °C the tubes were centrifuged for 20 min and 16000 g. The supernatant was removed and the pellets were washed twice with 70 % EtOH. The ethanol was removed and the pellet was air dried. Finally, the pellet was dissolved in 30 µl TE-buffer (10 mM Tris-HCl, 1 mM Na_2EDTA pH 8.0) at 37 °C. Samples were stored at -20 °C until transformation of *P. patens*.

2.12 Production of chemically competent *Escherichia coli* cells

Different safety strains of *E. coli* were used for plasmid multiplication and verification (EZ, DH5α) while others were used for expression studies (BL21 RIPL, SoluBL21, C41(DE3)). In order to get competent bacteria, cells were initially multiplied. With a sterile toothpick cells of a glycerol stock (stored at -80 °C) were picked and inoculated in a test tube containing 2 ml liquid LB medium and 25 µg tetracycline (EZ), 100 µg chloramphenicol (BL21 RIPL) or no antibiotic (DH5α, SoluBL21, C41(DE3)). The test tubes were closed with an aluminum cap and shaken for 14 to 18 h at 220 rpm and 37 °C (overnight culture). 100 ml liquid LB medium with the equivalent amount of antibiotics were prepared in a 250 ml Erlenmeyer flask and 1 ml was taken as a reference for the determination of the optical density at 600 nm (OD_{600}). 2 ml overnight culture were added to the remaining medium in the flask and cultivated at 220 rpm and 37 °C until an OD_{600} of 0.4-0.6 was reached. The cells were then centrifuged in two sterile 50 ml tubes at 3000 g and 4 °C for 10 min and the supernatant was discarded. The two pellets were combined after resuspension in 10 ml cold 100 mM $CaCl_2$ and centrifuged at 2500 g and 4 °C for 12 min. Again, the supernatant was discarded and the pellet resuspended in 10 ml cold 100 mM $CaCl_2$, followed by an incubation for 20 min at 4 °C and centrifugation under the same conditions as before. Finally, the supernatant was discarded and the cells were resuspended in 2 ml 100 mM $CaCl_2$ supplemented with 15 % glycerol. The bacteria were distributed in aliquots of 150 µl in sterile 1.5 ml reaction tubes, frozen in liquid nitrogen and stored at -80 °C.

2.13 Transformation of *Escherichia coli*

E. coli was transformed using a heat shock transformation protocol. Tubes with 150 µl frozen competent *E. coli* cells (stored at -80 °C until use) were thawed slowly on ice. To 70 µl of these cells, 5 µl ligation mix or 2 µl of an already constructed plasmid was added. The reaction tube was mixed carefully and put back on ice for 30 min. The tubes were then incubated in a water bath at 42 °C for 90 sec and cooled down for 2 min on ice. 150 µl SOC medium was added and incubated at 37 °C for 30 min with occasional gentle shaking. The mixture was pipetted onto a pre-warmed LB agar plate (containing the appropriate amount of antibiotic) in a Petri dish. Clones transformed with the vector

pDrive were additionally selected for integrated DNA fragments via a blue-white screening. For this purpose, X-Gal (80 µg ml^{-1}) and IPTG (12 µg ml^{-1}) was added to the LB agar plates. After drying of the liquid film, the plates were sealed and incubated overnight (~18 h) at 37 °C.

2.14 Transformation of *Saccharomyces cerevisiae*

Transformation of *S. cerevisiae* strains was done with the lithium acetate method according to Gietz and Schiestl (2007). All steps were performed with autoclaved buffers and toothpicks. First, 20 mg carrier DNA (fish sperm) was dissolved in 10 ml TE buffer (10 mM Tris-HCl, 1 mM Na_2EDTA pH 8.0) and stirred for 1-2 h at 4 °C. The solution was distributed in aliquots of 500 µl in sterile 1.5 ml reaction tubes and stored at -20 °C until further use. For transformation, the carrier DNA was thawed, heated to 99 °C for 5 min and then kept on ice. Before adding the carrier DNA to the transformation mixture, the tubes were mixed carefully. The plasmid DNA was diluted with H_2O to a concentration of 50 ng $µl^{-1}$. Yeast cells about as large as a pinhead were picked with a toothpick from a fresh YPD plate and resuspended in 1 ml H_2O. The suspension was centrifuged at 8000 g for 30 sec and the supernatant removed. The following components were pipetted to the cell pellet in the given order and mixed afterwards.

240 µl	PEG 4000 (50 % w/v)
36 µl	Lithium acetate (1 M)
50 µl	Carrier DNA (2 mg ml^{-1})
34 µl	Plasmid DNA (50 ng $µl^{-1}$)

The samples were incubated for at least 1 h at 42 °C, spun down at 8000 g and the supernatant was removed. Yeast cells were resuspended in 200 µl H_2O, poured on SCD_{-ura} agar plates, sealed and incubated for 3-4 days at 30 °C.

2.15 Overnight cultures and glycerol stocks

After successful transformation of *E. coli* and colony formation on solid media plates, colonies were picked with help of sterile toothpicks and placed in a tube containing 4 ml LB medium supplemented with either ampicillin (100 µg ml^{-1}) or kanamycin (50 µg ml^{-1}).

The tubes were closed and incubated for 18 h at 37 °C and 220 rpm. The next day, the cells were either used for plasmid isolation (Chapter II.2.16.1) or for protein expression (Chapter II.3.1.1). Glycerol cultures for long-term storage were established from these cultures as required. For this purpose, 425 µl *E. coli* overnight culture were mixed with 75 µl sterile glycerol. The samples were frozen in liquid nitrogen and stored at -80 °C.

2.16 Plasmid preparation

Plasmid preparation is a simple method to extract plasmids. The plasmids were then either sequenced for verification, or the insert was cut by restriction enzymes and ligated into another vector. Moreover, multiplied empty vectors could be obtained in this way.

2.16.1 Plasmid isolation from *Escherichia coli*

1.5 ml of a cell suspension from overnight cultures were pipetted into 1.5 ml tubes followed by centrifugation at 3000 g for 4 min. The supernatant was discarded and the procedure was repeated with another 1.5 ml cell suspension. After complete removal of the supernatant, the protocol from the QIAprep® Spin Miniprep Kit by Qiagen was conducted. At first, the pellets were resuspended in 250 µl buffer P1 (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 100 µg ml^{-1} RNase A). 250 µl buffer P2 (200 mM NaOH, 1 % SDS) were added and the tubes were mixed by inversion. 350 µl buffer N3 (4.2 M guanidinium-HCl, 0.9 M potassium acetate, pH 4.8) were added and the tubes were again mixed by inversion. After centrifugation for 10 min at 16000 g the supernatant was carefully transferred to QIAprep® spin columns and centrifuged for 1 min at 16000 g. The flow-through was removed and the columns were washed with 500 µl buffer PB (5 M guanidinium-HCl, 30 % 2-propanol) by centrifugation at 16000 g for 1 min. After discarding the flow-through the columns were washed again with 750 µl buffer PE (10 mM Tris-HCl pH 7.5, 80 % EtOH) by centrifugation at the same conditions. After removal of the flow-through the columns were dried by centrifugation for an additional minute. The columns were placed on fresh 1.5 ml reaction tubes and 50 µl sterile H_2O was applied to the columns. After incubation at room temperature for 5 min, the eluted plasmid was spun down for 1 min at 11000 g. Samples were stored at -20 °C.

2.16.2 Plasmid isolation from *Saccharomyces cerevisiae*

Plasmid preparation from yeast cells was basically very similar to the procedure described for *E. coli* (Chapter II.2.16.1). Initially yeast colonies were picked from a SCD_{-ura} plate and 10 ml liquid SCD_{-ura} medium were inoculated in 25 ml-Erlenmeyer flasks. The cells were incubated for 24 h at 30 °C and 160 rpm and afterwards spun down in 2 ml reaction tubes for 3 min at 3000 g. The supernatant was discarded and the centrifugation was repeated, until the complete cell culture was used. The pellets were resuspended in 250 µl P1 (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 100 µg ml^{-1} RNase A) and a blade tip of glass beads (~ 0.5 mm ø)(corresponding to a volume of approximately 100 µl) were added. The samples were vigorously mixed for 10 min. After the glass beads had settled, the supernatant was transferred into a fresh 1.5 ml reaction tube. Subsequently, the same procedure was followed as for the plasmid preparation from *E. coli*.

2.17 Restriction enzyme digestion

Digestion of DNA with restriction endonucleases was used either for cloning or for verification of insertion of a gene into a vector after multiplication in *E. coli*. Most of the restriction endonucleases used in this work were purchased from Fermentas. Since they often have different requirements regarding buffer and temperature, the optimal specifications of the manufacturer were used. If a double digest was performed, conditions were carried out as recommended by the supplier (DoubleDigest Calculator-Thermo Scientific). The standard digestion assays were performed in a total volume of 10 µl for the verification of ligation. If a fragment was inserted in pDrive, verification was usually carried out using EcoRI. In this case assays were composed as follows:

5 µl	Plasmid
3.5 µl	H_2O
1 µl	EcoRI buffer (10x)
0.5 µl	EcoRI

For digestion of high amounts of DNA (for example used for cloning or linearization of constructs used for transformation of *P. patens*), the final volume was extended up to

40 µl. In addition, buffers as well as restriction enzymes were upscaled accordingly and only plasmid DNA was used instead of further dilution with water. Samples were incubated overnight (~16 h) or at least for 3 h at 37 °C or 25 °C depending on the restriction enzyme. If only information about the actual presence of an insert had to be gathered quickly, short heating in the microwave at low radiation levels (~30 sec) was enough to detect bands by agarose gel electrophoresis (Chapter II.2.5).

2.18 Sequence verification

After subjecting the created plasmids to a restriction digest and examination of the products by agarose gel electrophoresis, plasmids with the expected fragment pattern were sent to Seqlab for sequencing. The concentration of the plasmid was determined and 60-120 ng μl^{-1} in a total volume of 12 µl were transferred to a fresh reaction tube.

2.19 Transformation of *Physcomitrella patens*

All steps for the transformation of *P. patens* were performed according to Cove et al. (2009abc). The transformation of *P. patens* took place in three steps: first, protoplasts were generated from protonema tissue, then the actual transformation was performed using PEG-mediated DNA uptake and finally protoplasts were regenerated over weeks on non-selective and selective media. All solutions (Chapter II.5) and materials were autoclaved or filter-sterilized.

2.19.1 Protoplastation

Before protoplastation a 2 % Driselase (Sigma-Aldrich) solution had to be prepared. The Driselase was dissolved in 8.5 % D-mannitol (w/v) and stirred at room temperature for 1 h. The solution was transferred to a 50 ml tube and centrifuged for 20 min at 4500 g. The supernatant was filter-sterilized, portioned in aliquots of 5 ml and stored at -20 °C. For transformation, tubes were slowly thawed 1 h before use.

7-Day old protonema tissue of the *P. patens* wild type in liquid BCD medium was initially concentrated by carefully decanting the medium after the moss had settled down. The remaining sample was disrupted with a sterilised tissue blender (Omni International) for 60 s. 3 ml were poured on Petri dishes with BCDA medium covered with cellophane. The

plates were sealed with Micropore™ tape (3M) and incubated under continuous light at 25 °C for six days. The protonema tissue of six plates was harvested, transferred to 10 ml 8.5 % D-mannitol (w/v) and 5 ml 2 % Driselase solution was added. The protoplastation reaction was incubated for 3 h at room temperature with occasional gentle swirling. The protoplasts were filtered through a 100 µm sieve and incubated at room temperature for another 15 min. Afterwards protoplasts were filtered through a 50 µM sieve and the flow-through was transferred to a 15 ml centrifugation tube. All centrifugation steps were carried out at 100 g in a swinging-bucket rotor. Protoplasts were sedimented by centrifugation for 5 min and the supernatant was discarded. The pellet was washed in 5 ml 8.5 % D-mannitol (w/v) supplemented with 10 mM $CaCl_2$ and spun down again for 5 min. Subsequently the pellet was resuspended in 5 ml 8.5 % D-mannitol (w/v) supplemented with 10 mM $CaCl_2$ and the number of protoplasts was determined microscopically using a Fuchs-Rosenthal counting chamber. For each transformation appr. 5-10 x 10^6 protoplasts were used in a volume of 300 µl MMM-buffer (1 % (w/v) 2-(*N*-morpholino)ethanesulfonic acid (MES), 8.5 % D-mannitol (w/v), 15 mM $MgCl_2$, pH 5.6 adjusted with KOH). After calculation of the protoplast number, they were centrifuged again for 5 min and dissolved in the corresponding volume of MMM-buffer.

2.19.2 PEG-mediated transformation

Transformation reactions were performed in 15 ml centrifuge tubes. 300 µl protoplasts in MMM-buffer were combined with 30 µl linearized plasmid DNA solution (Chapter II.2.17) and 300 µl filter-sterilized PEG transformation solution (0.38 M D-mannitol, 0.1 M calcium nitrate, 40 % (w/v) PEG 4000 and 1 ml l^{-1} 1 M Tris-HCl pH 8.0) and the tubes were mixed gently, producing a visually uniform solution. Reaction tubes were cooled down to room temperature for 10 min after incubation for 5 min at 45 °C in a water bath. Because the PEG transformation solution is toxic to the protoplasts, suspensions were diluted, at one minute intervals, five times with 300 µl followed by five times with 1 ml 8.5 % D-mannitol (w/v) supplemented with 10 mM $CaCl_2$. Tubes were mixed gently by swirling after each dilution step. Before the protoplasts could be transferred to regeneration medium they were left at room temperature for 30 min and concentrated by centrifugation for 5 min. The pellets were dissolved in 2 ml 8.5 % D-mannitol (w/v) supplemented with 10 mM $CaCl_2$ and 3 ml PRMT (previously melted in

the microwave and cooled down to 45 °C in a water bath). 2.5 ml of this mixture was quickly poured onto two PRMB plates each overlaid with sterile cellophane. The plates were covered and sealed with Micropore™ tape (3M) and left for 1 h at room temperature at a place with dim light. Afterwards Petri dishes were transferred to 25 °C with continuous light for regeneration and selection.

2.19.3 Plant regeneration and selection

Regeneration and selection of stable transformants took 5 weeks and was carried out at 25 °C under continuous light. All plates were sealed with Micropore™ tape (3M). During the first week protoplasts could regenerate on PRMB medium after transformation. After one week the cellophane, carrying the regenerating protoplasts, was transferred to BCDA medium containing 25 mg/l hygromycin B. A week later, the cellophane disks were transferred to antibiotic-free BCDA medium and incubated for 2 weeks. Since most transformants are still transient transformants at this stage, these two weeks of selection release led to a loss of resistance in transient transformants. At last the cellophane was again placed on a hygromycin B-containing BCDA medium for a week. Stable transformants were kept on solid BCD medium in Petri dishes at 25 °C under continuous light and sealed with Nescofilm™. Liquid protonema cultures for expression were established from stable transformants by transferring plants from solid BCD into 50 ml liquid BCD in a 250 ml-Erlenmeyer flask. The tissue was disrupted with a sterilised tissue blender (Omni International) for 30 s. Flasks were kept at 25 °C under continuous light on a gyratory shaker (100 rpm) and were maintained as described in chapter II.1.1.

3 Enzymology

3.1 Expression of recombinant proteins and isolation of crude protein extracts

3.1.1 *Escherichia coli*

For the heterologous expression of proteins in *E. coli* different strains were transformed with the created expression vectors. After insertion and correctness of the plasmid were verified, expression of the recombinant protein was the next step. First, an overnight culture was generated from a glycerol stock (Chapter II.2.15) and the cells were

incubated overnight (~ 16 h) at 37 °C and 220 rpm. The next morning 2 ml of the suspension were transferred to 100 ml LB or TB media supplemented with 100 µg ml^{-1} ampicillin in a 500 ml-baffled flask. For expression of the P450-CPR fusion constructs (Chapter II.2.10) 12.5 mg l^{-1} $FeSO_4$ and 500 µM 5-aminolevulinic acid were additionally added. The bacteria were incubated at 37 °C at 180 rpm until they reached an OD_{600} of 0.4-0.6. With addition of 1 mM isopropyl-β-D-galactopyranoside (IPTG) protein expression was induced. While the bacterial strains SoluBL21 and C41(DE3) were incubated for 16 h at 25 °C and 180 rpm, BL21 RIPL was incubated for 5 h at 37 °C and 180 rpm. After cultivation, cells were collected by centrifugation for 5 min at 3000 g and 4 °C in 50 ml-tubes. The supernatant was discarded and the cells were resuspended in 4 ml Tris-HCl buffer (0.1 M, pH 7.0 or 7.5) supplemented with 1 mM dithiothreitol (DTT), 1 mM sodium diethyldithiocarbamate (DIECA) (P450-CPR) or in 4 ml potassium phosphate buffer (100 mM, pH 7)(CoA-ligases) per 1 g of cells. If Ni-NTA-chromatography (Chapter II.3.4) was subsequently intended, the pellet was dissolved in 4 ml 0.1 M potassium phosphate buffer (pH 8). The cells were disrupted by ultrasonication on ice (5x30 s, 100 %, 0.3 cycles) after 30 min incubation with ~50 mg lysozyme. The crude protein extract was obtained by centrifugation (5000 g, 4 °C, 20 min).

3.1.2 *Saccharomyces cerevisiae*

At first, *S. cerevisiae* transformants were subcultivated on SCD_{-ura} agar plates. Colonies were picked from the transformation plate with a sterile toothpick and distributed on a fresh plate. The plates were incubated for 48 h at 30 °C and then stored at 4 °C until further use. Transformation was repeated every four weeks to ensure a persistent behaviour of the yeast cells. With a sterile toothpick a cell aliquot as large as a pinhead were transferred to 100 ml SCD_{-ura} in a 500 ml-baffled flask. The cultures were incubated for 48 h at 30 °C (180 rpm) to achieve a high cell density. Afterwards the cells were collected by centrifugation for 5 min at 3000 g and 4 °C in sterile 50 ml-tubes. The pellet was resuspended in 100 ml SCG_{-ura} in a 500 ml-baffled flask. In later experiments (expression of the codon optimized AaCYP98) SCG_{-ura} was supplemented with 12.5 mg l^{-1} $FeSO_4$ and 200 µM 5-aminolevulinic acid. The cultures were incubated for another 24 h at 30 °C (180 rpm) and then collected by centrifugation for 5 min at 3000 g and 4 °C in

50 ml-tubes. The pellet was redissolved in 10 ml buffer (0.1 M Tris-HCl pH 7.5, 1 mM dithiothreitol (DTT), 1 mM sodium diethyldithiocarbamate (DIECA)) and left for 5 min on ice. To remove the remains of the medium the cells were again spun down for 5 min at 3000 g (4 °C) and the supernatant was discarded. The pellet was resuspended 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 homogeniser (Minilys®) at 4.000 rpm for 30 sec, followed by cooling on ice for 30 sec and 6-8 repetitions. The supernatant was removed after the glass beads had settled and glass beads were washed with 0.5 ml buffer. Crude protein extracts were obtained after the supernatants were pooled and spun down for 20 min at 5000 g (4 °C) in a 50 ml-tube. The extracts were used directly for determination of enzyme activity or preparation of membranes. Samples for future SDS or Western blot analysis were stored at -20 °C or -80 °C.

3.1.3 *Physcomitrella patens*

7-day old suspension cultures in liquid BCD medium were homogenized with a tissue blender and 10 ml of these suspensions were transferred into 200 ml BCD in 1 l-Erlenmeyer flasks. The cells were cultivated for 12 d at 25 °C under continuous light on a gyratory shaker (100 rpm). *P. patens* was harvested by filtration and homogenized in a pre-cooled mortar together with 20 % (w/w) of the fresh weight (FW) Polyclar 10 and 6 ml per g FW buffer (0.1 M Tris-HCl pH 7.0, 1 mM DTT, 1 mM DIECA). Crude protein extracts were obtained after centrifugation of the homogenate for 20 min at 5000 g 4 °C.

3.2 Determination of protein concentration

Protein concentrations were determined according to Bradford (1976). The Bradford reagent was composed as follows: 100 mg l^{-1} Coomassie Brilliant Blue G250, 50 ml l^{-1} 96 % ethanol and 100 ml l^{-1} 85 % o-phosphoric acid in H_2O. The solution was filtered twice and stored at 4 °C until use. 10 µl protein sample was incubated with 1 ml Bradford reagent in disposable microcuvettes for 15 min. Absorbance was measured at 595 nm against Bradford reagent with 10 µl buffer as reference value. 10 µl bovine serum albumin (1 mg/ml) incubated with Bradford reagent served as a standard. In the case of very high amounts of protein (e.g. crude protein extracts from *E. coli*), the test

had to be slightly modified. Here only 5 µl of the protein sample, buffer or standard were used.

3.3 Preparation of membrane fractions

3.3.1 Preparation with $MgCl_2$ or PEG and NaCl

Precipitation of membranes with $MgCl_2$ or PEG 4000 and NaCl was performed to decrease the centrifugation speed needed to precipitate the membranes. This is usually the method of choice, if no ultracentrifuge is available (Pompon et al. 1996; Petersen 2003). The crude protein extract was slowly adjusted to 50 mM $MgCl_2$ (from a 1 M stock solution) or 0.1 g ml^{-1} PEG 4000 and 150 mM NaCl (from a 1 M stock solution) and the solution was stirred for 20 min on ice. Afterwards the aggregated membranes were spun down for 20 min at 60000 g (4 °C). The supernatant was removed and collected in a separate tube. The pellet containing aggregated membranes was resuspended in Tris-HCl buffer (100 mM, pH 7, supplemented with 1 mM DTT and 1 mM DIECA) with the help of a Potter-Elvehjem glass homogenizer. Resuspended membranes were kept on ice and used directly for the determination of enzyme activity.

3.3.2 Preparation by ultracentrifugation

This method was originally developed for large amounts of mammalian tissue, but it is also the most common procedure in studies involving plant membrane proteins (Sullivan and Zarnowski 2010; Liu et al. 2017; Renault et al. 2017a; Alber et al. 2019). The crude protein extract was centrifuged at 100000 g for 1 h at 4 °C. The supernatant was removed and collected in a separate tube. The pellet was resuspended in Tris-HCl buffer (as above) with the help of a Potter-Elvehjem glass homogenizer.

3.3.3 Preparation according to the protocol of Abas and Luschnig (2010)

In 2010, Abas and Luschnig published a method for isolating microsomal-type membranes from small amounts of plant material without requiring ultracentrifugation. The protocol was performed with some modifications. Thus, the crude protein extract was isolated on a different way as described in chapter II.3.1.3. At first an extraction buffer (EB) was prepared: 100 mM Tris-HCl pH 7.5 supplemented with 0.81 M sucrose, 5 % glycerol (v/v), 10 mM EDTA and 1 mM DTT. If not mentioned otherwise, all steps

were performed on ice. 2 µl EB per mg fresh weight were added to the plant cells and and the plant material was homogenized in a pre-cooled mortar. Then Polyclar 10 was added (0.05 g/mg cells) and the homogenate was incubated for 5 min at room temperature. The homogenate was centrifuged at 600 g for 3 min at 4 °C and the supernatant was collected in a fresh reaction tube. The pellet was again resuspended in half the volume EB and again spun down at 600 g for 3 min at 4 °C and the supernatant was collected. At last the pellet was centrifuged once more at 2000 g for 30 sec at 4 °C and the remaining supernatant was pooled with the previously collected supernatants. This crude protein extract was diluted with the same volume of water to decrease the sugar concentration. Samples were distributed in aliquots of 200 µl in 1.5 ml reaction tubes, followed by a centrifugation at 21000 g for 1.5 h at 4 °C. The pellets were resuspended in 100 µl Tris-HCl buffer (100 mM, pH 7) per reaction tube, pooled and carefully homogenized with the help of a Potter-Elvehjem glass homogenizer.

3.4 Purification of recombinant proteins by metal chelate chromatography

At first, the crude protein extract (in 0.1 M potassium phosphate buffer pH 8.0) was adjusted to 10 mM imidazole as well as 300 mM NaCl and then added to 1 ml pre-equilibrated Ni-NTA resin in a disposable column. The column was closed and incubated on ice for 1 hour on a rocking platform. After the solution had dripped out of the column (flow-through) the resin was washed three times with 2 ml wash buffer (50 mM K_2HPO_4/KH_2PO_4 pH 8.0, 15 mM imidazole, 300 mM NaCl) (wash fraction 1-3). Then the protein was eluted with 2.5 ml elution buffer (50 mM K_2HPO_4/KH_2PO_4 pH 8.0, 250 mM imidazole, 300 mM NaCl) (elution fraction). To remove the imidazole the elution fractions were desalted by gel filtration through PD-10 columns.

3.5 Desalting via PD-10 columns

All protein solutions, purified by metal chelate chromatography, were desalted with PD-10 columns. This was done based on the manufacturer's instructions (GE Healthcare). At first, the PD-10 columns were washed with 20 ml water and equilibrated with 20 ml elution buffer (0.1 M K_2HPO_4/KH_2PO_4 pH 7.0 for CoA-ligases and 0.1 M Tris-HCl pH 7.0 for AaCPR). Then 2.5 ml protein solution was added and the flow through was discarded. In the last step the PD-10 columns were eluted with 3.5 ml buffer, while collecting the

eluting protein fraction. Afterwards the columns were washed with 20 ml H_2O and stored at 4 °C until further use. The collected protein fractions were aliquoted, frozen with liquid nitrogen and stored at -80 °C after determination of the protein concentration (Chapter II.3.2).

3.6 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Protein extracts were subjected to SDS-PAGE, separating them according to their molecular weight. This is based on their differential rates of migration through a sieving matrix (a gel) under the influence of an applied electrical field. To mask the intrinsic charges of the protein the samples are heated with Laemmli buffer, a mixture of SDS, glycerol, β-mercaptoethanol and bromophenolblue (see below). The reducing agent β-mercaptoethanol disrupts the tertiary structure of proteins and SDS (sodium dodecyl sulphate) coats the protein with a uniform negative charge. Since SDS binds uniformly to the linear proteins (around 1.4 g SDS per 1 g protein), the charge of the protein is now almost proportional to its molecular weight. Laemmli buffer (4x) was composed as follows:

2.4 ml	1 M Tris-HCl pH 6.8
0.8 g	SDS
4 ml	glycerol
1 mg	bromophenol blue
2.5 ml	H_2O
1 ml	β-mercaptoethanol

Electrophoresis was performed based on a discontinuous system (Laemmli 1970), with a lower separating gel and an upper stacking gel layer. The gel was formed by polymerizing acrylamide with bisacrylamide using TEMED and APS.

First, the glass plates were placed into the gadget with an appropriate spacer in between, after they had been thoroughly cleaned with 70 % EtOH. Then the separating gel was mixed, quickly poured between the glass plates and a thin layer of water was overlaid to straighten the line between separating and stacking gel. The water was removed after polymerization of the separating gel. Then the stacking gel was mixed and poured on top of the separating gel layer and a comb for ten pockets was added. After

complete polymerization the gel was placed in the electrophoresis chamber and the buffer reservoir filled with electrophoresis buffer (192 mM glycine, 25 mM Tris and 0.1 % SDS, pH 5.3). The comb was removed carefully and the pockets repeatedly washed with running buffer to remove gel residues or air bubbles.

The separating and stacking gel were composed as follows:

Separating gel

1.25 ml/gel	1.5 M Tris-HCl pH 8.8
1.45 ml/gel	H_2O
2.05 ml/gel	Rotiphorese® Gel 30
200 µl/gel	SDS (10 % w/v)
8 µl/gel	TEMED
35 µl/gel	APS (10 % w/v)

Stacking gel

625 µl/gel	0.5 M Tris-HCl pH 6.8
1.4 ml/gel	H_2O
375 µl/gel	Rotiphorese® Gel 30
100 µl/gel	SDS (10 % w/v)
5 µl/gel	TEMED
20 µl/gel	APS (10 % w/v)

Protein samples (15 µl) were mixed with 5 µl Laemmli buffer and heated to 95 °C for 10 min before they were carefully pipetted into the gel pockets. Pellets from yeast expression cultures were dissolved in 100 µl 0.2 M NaOH and incubated for at least 30 min at room temperature. After centrifugation at 8.000 g for two min pellets were re-dissolved in 20 µl Laemmli buffer and denatured as described above. 5 µl Roti®-Mark TRICOLOR was used as a size marker. Running times for the electrophoresis were around 90 min at 150 V and 100 mA. Afterwards the stacking gel was removed and the separating gel was either further used for a Western blot analysis (Chapter II.3.7) or the gel was stained with Coomassie-Brilliant Blue. For staining, the separating gel was placed

in a solution of 0.3 mM Coomassie-Brilliant Blue R-250 in MeOH:AcOH:H_2O (4.5:1:4.5) and incubated for 0.5-2 h on a rocking platform. Short heating in the microwave at 800 W until the staining solution started to boil accelerated this process. The gel was destained in the same solvent mixture (without Coomassie reagent) for 1.5-2 h, with intermittent exchange of the destaining solution, until the optimal result was achieved.

3.7 Western blot

Western blotting, also known as protein immunoblot, is a technique to detect the presence of a specific protein in a complex mixture extracted from cells. At first, proteins are separated by size using SDS-PAGE and then they are transferred to the surface of a membrane followed by a detection of the target protein by specific antibodies. Finally, the target protein can be visualized as band with the help of color or chemoluminescence detection systems.

After successful SDS-PAGE the separating gel was equilibrated in transfer buffer (Towbin buffer: 25 mM Tris, 192 mM glycine, 20 % MeOH) for 10-30 min. If the expected proteins had a size over 100 kDa 0.1 % (w/v) SDS was added to the transfer buffer. Pre-cut filter paper and sponge material for the blotting chamber were briefly soaked in transfer buffer. The membrane (PVDF 2.0, 0.2 µm pore size, Roth) was rinsed in MeOH for about 15 sec, washed in water for 2 min and equilibrated in transfer buffer for at least 5 min. The single layers were assembled in the correct order from anode to cathode: sponge material, filter paper, membrane, separating gel, filter paper, sponge material. Air enclosures were carefully removed and the “sandwich” was installed in a tank blotting system filled with transfer buffer. After blotting for 1.5 h at 100 V the membrane was carefully removed and analyzed by immunodetection. In other experiments, a semi-dry blotting system was used. For this, blotting was done with a constant current, depending on the membrane size (length [cm] · width [cm] · 4 mA), for 1.5 h. The used SDS-gel was dyed as described in chapter II.3.6, to check for transfer efficiency.

The membrane was washed three times in TBS buffer (10 mM Tris-HCl, 0.9 % NaCl, pH 7.4) for 5 min. Blocking of the membrane was performed in TBS-T buffer (10 mM Tris-HCl, 0.9 % NaCl, 0.05 % Tween 20, pH 7.4) with 5 % milk powder for 1.5 h, followed by two washing steps in TBS-T for 5 min. After that, the membrane was incubated with the

first antibody (anti-6xHis or anti-FLAG antibody, diluted 1:10000 in TBS-T supplemented with 1 % milk powder) for at least 1 h at room temperature or overnight at 4 °C. The membrane was washed eight times for 5 min each in TBS-T and then incubated with the second antibody (anti-mouse IgG alkaline phosphatase, diluted 1:20000 in TBS-T supplemented with 1 % bovine serum albumin (w/v)) for 1 h at room temperature or overnight at 4 °C. Afterwards the membrane was washed five times in TBS-T for 5 min each and was finally dyed with freshly prepared NBT-BCIP solution. The membrane was equilibrated in substrate buffer (100 mM Tris-HCl, 100 mM NaCl, 5 mM $MgCl_2$, pH 9.5) for 5 min and then stained in 10 ml substrate buffer mixed with 80 µl BCIP (20 mg/ml in 100 % dimethylformamide) and 60 µl NBT (50 mg/ml in 70 % dimethylformamide) for 15-30 min and finally stopped by washing in H_2O for three times. This detection method leads to an intense purple coloration when it reacts with alkaline phosphatase coupled to the secondary antibodies. The alkaline phosphatase catalyses the cleavage of a phosphate group from BCIP to produce an indigo dye. The indoxyl group reduces a tetrazolium salt (NBT) to produce an insoluble formazan that couples with the indigo dye to form a colored precipitate (Altman 1976).

Afterwards the membrane was dyed with amido black staining solution (0.1 % in MeOH:AcOH:H_2O, 4:1:5) for 10 sec to check the protein transfer with subsequent destaining in MeOH:AcOH:H_2O (4.5:1:4.5) with intermittent exchange of the solution, until the optimal result was achieved.

3.8 Determination of enzyme activities

3.8.1 Standard assay for P450 monooxygenases

Standard assays of CYP enzymes contained 100 µl crude protein extract (0.2 mg protein), 7.5 µl substrate (10 mM, dissolved in 50 % methanol), 12.5 µl NADPH (50 mM) and 5 µl buffer (0.1 M Tris-HCl pH 7) (in 1.5 ml reaction tubes). Assays were mixed vigorously and incubated for at 25 °C under shaking at 1200 rpm in an Eppendorf Thermomixer. The reaction was stopped by addition of 50 µl 6 N HCl. The assays were extracted twice with 500 µl EtOAc each and the combined ethyl acetate extracts were evaporated. The residues were resuspended in 50 µl MeOH with 0.01 % H_3PO_4 and centrifuged for 10 min

at 16000 g after addition of 50 µl aqueous 0.01 % H_3PO_4. The supernatant was either analyzed by HPLC or LC-MS.

For the determination of the pH-optimum of AaC4H and AaCYP98 50 µl crude protein extract were mixed with 50 µl Tris-HCl buffer (500 mM, pH 5.5-9.0). The exact pH of the assays was determined before the addition of NADPH. Assays for the temperature optimum were usually performed in 100 mM Tris-HCl buffer pH 7.0.

The kinetic values for cinnamic acid and NADPH were recorded for AaC4H. The specific activity was calculated on the basis of a calibration curve of different 4-coumaric acid concentrations. Assays with cinnamic acid in concentrations up to 240 µM with 5 mM NADPH were carried out with 5 min reaction time at 25 °C to guarantee the determination of initial reaction velocities. K_m-values for NADPH were determined with enzyme assays using up to 960 µM NADPH with 400 µM cinnamic acid and an incubation time of 5 minutes at 25 °C.

The kinetic values of AaCYP98 were determined for the substrates 4-coumaroylanthranilate (pC-An), 4-coumaroyltyramine (pC-Ty) and the cosubstrate NADPH. Assays with pC-An or pC-Ty in concentrations up to 500 µM were incubated with 0.2 mg protein and 770 µM NADPH at 30 °C for 10 min (pC-An) or 15 min (pC-Ty) and 1200 rpm to guarantee the determination of initial reaction velocities. All assays contained the same amount of MeOH. Since no standard solutions of caffeoylanthranilate (Caf-An) or caffeoyltyramine (Caf-Ty) were available, substrate concentrations were plotted against the peak area of the resulting product. To determine the K_m-value for NADPH assays with 0.2 mg protein and 200 µM pC-An were incubated with up to 720 µM NADPH. The assays were incubated for 10 min at 30 °C and 1200 rpm.

3.8.2 NADPH:cytochrome P450 reductase (CPR)

The activity of CPR was determined spectrophotometrically at 550 nm and 40 °C using cytochrome c (horse heart, Fluka) as an artifical substrate. A standard assay contained 100 mM Tris-HCl buffer pH 7.5, 200 µg crude protein or 4 µg purified protein, 0.5 mM potassium cyanide, 0.01 mM FAD, 0.01 mM FMN and 0.05 mM cytochrome c. Before starting the assay, the absoption was recorded for 300 s without the substrate NADPH,

then 0.15 mM NADPH was added and the absorption monitored for additional 420 s. For the calculation of the specific activity, the slope was determined after the addition of the substrate minus the slope prior to the addition of the substrate. The activity was calculated with the extinction coefficient 21 mM^{-1} cm^{-1} for cytochrome c.

3.8.3 Determination of CoA-ligase activity

3.8.3.1 Cinnamic acid derivatives and 4-hydroxybenzoic acid as substrates

Formation of CoA-esters from (hydroxy)cinnamic acid derivatives and 4-hydroxybenzoic acid was determined photometrically at the absorption maxima published by Stöckigt and Zenk (1975), Zenk (1979) and Biegert et al. (1993) (Table 1).

A standard assay for (hydroxy)cinnamic acids consisted of 100 mM potassium phosphate buffer pH 7.0, 2 µg (Aa4CL) or 32 µg (Aa4HBCL) purified protein, 1 mM dithiothreitol (DTT), 2.5 mM ATP, 2.5 mM $MgCl_2$ and 500 µM (hydroxy)cinnamic acid. When testing caffeic acid as substrate, DTT was omitted. When Aa4HBCL was tested with 4-hydroxybenzoic acid (4HBA), only 15 µg purified protein was used. The assay was incubated at 45 °C for 1 min and then the reaction was started by addition of 500 µM CoA and incubated for another 8 min. Absorption was recorded throughout the incubation time in acryl or fused quartz cuvettes. The specific activity was calculated by using the measured slope of the absorbance and the respective extinction coefficient (Table 1).

Table 1 Wavelengths and extinction coefficients of the enzymatically synthesized hydroxycinnamoyl-CoA esters (Stöckigt and Zenk 1975; Zenk 1979; Biegert et al. 1993).

Product	Absorption maximum	Extinction coefficient
cinnamoyl-CoA	311 nm	22 cm^2/µmol
4-coumaroyl-CoA	333 nm	21 cm^2/µmol
caffeoyl-CoA	346 nm	18 cm^2/µmol
feruloyl-CoA	345 nm	19 cm^2/µmol
isoferuloyl-CoA	345 nm	18 cm^2/µmol
sinapoyl-CoA	352 nm	20 cm^2/µmol
4-hydroxybenzoyl-CoA	300 nm	13 cm^2/µmol

For the determination of temperature and pH-optimum, Aa4CL assays consisted of 500 µM 4-coumaric acid or ferulic acid, 2.5 mM ATP, 2.5 mM $MgCl_2$ and 100 µM CoA. Aa4HBCL assays contained 500 µM 4HBA, 2.5 mM ATP, 2.5 mM $MgCl_2$ and 100 µM CoA. 500 mM KPi buffer was used to determine the pH-optimum for both enzymes.

Kinetic values for Aa4CL were determined for the substrates cinnamic acid, 4-coumaric acid, caffeic acid, ferulic acid and isoferulic acid as well as the two cosubstrates ATP and CoA. For determination of the K_m value for cinnamic acid derivatives assays contained 2.5 mM ATP, 2.5 mM $MgCl_2$ and 100 µM CoA. 4-Coumaric, caffeic, ferulic and isoferulic acid concentrations ranged from 0 to 500 µM and cinnamic acid was added up to 2 mM. All assays contained the same amount of MeOH. The K_m-value for ATP was determined in enzyme assays using up to 3 mM ATP and the same concentration of $MgCl_2$, 0.5 mM ferulic acid and 100 µM CoA. For determination of the kinetic value for CoA, CoA was used in concentrations between 0 and 150 µM. Moreover, 0.5 mM ferulic acid as well as 2.5 mM ATP and 2.5 mM $MgCl_2$ were added.

Kinetic values for Aa4HBCL were determined for the substrate 4HBA and the cosubstrates ATP and CoA. 4HBA concentrations ranged from 0 to 3 mM, using 4 mM ATP and $MgCl_2$ and 500 µM CoA. All assays contained the same amount of MeOH. The K_m-value for ATP was determined in enzyme assays using up to 7.5 mM ATP and the same concentration of $MgCl_2$, 2.5 mM 4HBA and 750 µM CoA. For determination of the kinetic value for CoA, CoA was used in concentrations between 0 and 1 mM. Moreover, 2.5 mM 4HBA as well as 5 mM ATP and 5 mM $MgCl_2$ were added.

3.8.3.2 Indirect assay with hexokinase and glucose 6-phosphate dehydrogenase

For the determination of CoA-ligase activities with all other (hydroxy)benzoic acid derivatives and fatty acids an indirect spectrophotometric assay, using hexokinase and glucose-6-phosphate dehydrogenase (G6PDH), was established to determine the remaining ATP concentration by measuring the conversion of NADP to NADPH at 340 nm in a two step assay (Fig. 9).

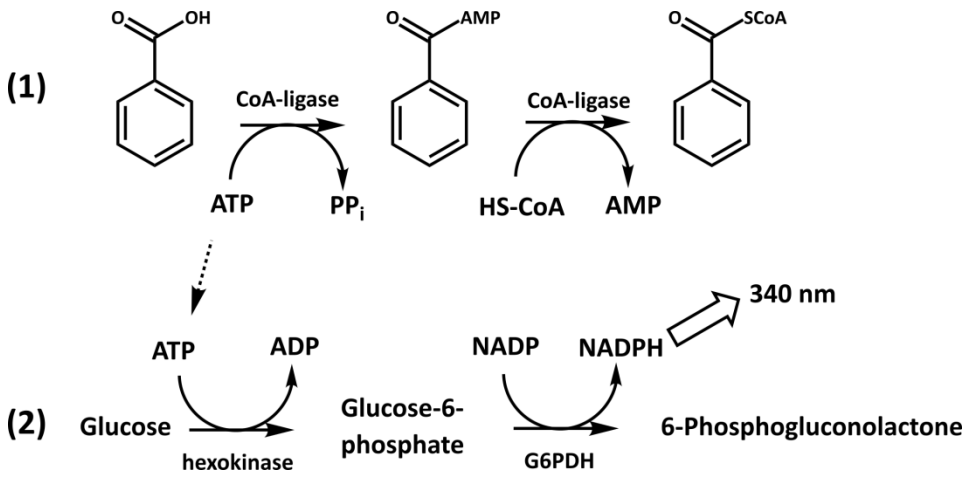


Figure 9 Reaction mechanism of the indirect assay to measure CoA-ligase activity. G6PDH = glucose-6-phosphate dehydrogenase.

The CoA-ligase assay **(1)** had a reaction volume of 250 µl and consisted of 100 mM Tris-HCl pH 7.5, 14 µg purified Aa4HBCL, 1 mM DTT, 500 µM ATP, 500 µM $MgCl_2$, 500 µM (hydroxy)benzoic acid derivative and 500 µM CoA. For measurement of fatty acids the CoA-ligase assay consisted of 30 µg purified Aa20832, 200 µM substrate, 200 µM ATP, 200 µM $MgCl_2$, 200 µM CoA and 1 mM DTT. After incubation for 1 h at 40 °C (benzoic acid derivatives) or 4 h at 25 °C (fatty acids), the reaction was stopped by heating for 5 min at 95 °C and cooling on ice. 750 µl reaction mixture consisting of 100 mM Tris-HCl pH 7.5, 200 µM NADP, 200 µM glucose, 1 U hexokinase and 1 U G6PDH were added and the absorption at 340 nm was recorded at 25 °C until until the absorption no longer changed **(2)**.

3.8.3.3 Assays for LC-MS analysis

Conversion of (hydroxy)benzoic acids as well as 2-coumaric and 3-coumaric acid was additionally analysed by LC-MS. Assays were performed in a 100 µl reaction volume consisting of 100 mM K_2HPO_4/KH_2PO_4 pH 7.0, 7.5 µg (Aa4CL), 28 µg (Aa4HBCL) purified enzyme or 100 µg crude protein extract of Aa20832, 1.25 mM ATP, 1.25 mM $MgCl_2$, 500 µM (hydroxy)benzoic acid or coumaric acid derivative and 1 mM CoA. Assays were incubated for 30 min at 40 °C (Aa4HBCL, Aa4CL with 2-coumaric and 3-coumaric acid) or 60 min at 35 °C (Aa4CL, Aa20832 with hydroxybenzoic acids). An assay with heat-denatured protein (10 min at 95 °C) served as negative control. Assays were stopped on

ice by addition of 100 µl methanol. After centrifugation for 10 min at 15 000 g, 15 µl of the supernatant was analysed by LC-MS (Chapter II.4.2).

4 Analytical and chemical methods

4.1 High-performance liquid chromatography (HPLC)

HPLC measurements were all performed under isocratic conditions. Thus, composition of the mobile phase remained stable throughout the time. A mixture of 20 % to 50 % MeOH supplemented with 0.01 % H_3PO_4 at a flow rate of 1 ml min^{-1} served as the mobile phase. If substances were purified by HPLC, acetic acid (3 ml l^{-1}) was used instead of H_3PO_4 to acidify the mobile phase. The stationary phase was an Equisil ODS-column (4 mm ø, 250 mm length with 20 mm pre-column). During the first years of this work an older HPLC system (Chapter II.6.7) without an autosampler was used. In this case, exactly 20 µl sample was injected onto the column by a sample loop. In the later course, analysis was carried out on a new HPLC system (Chapter II.6.7), including an autosampler and a column oven. Advantages were automated measurement and better reproducibility. Mostly 50 µl sample was analyzed in the new HPLC system. Elution was recorded at 309 nm (C4H assays) and 333 nm (CYP98 assays). Substrates or products in concentrations of 25 to 100 µM dissolved in eluent or MeOH were used as standards for identification and quantification.

4.2 Liquid chromatography/mass spectrometry (LC-MS)

LC was performed on a HPLC 1260 (Agilent Technologies) 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 with either a short (20 min) or long (55 min) method with different solvent gradients, flow rates and temperatures. 10 µl or 15 µl sample were injected onto the column.

The short method had the following gradient with a flow rate of 0.5 ml min^{-1} and a temperature of 25 °C: 0–10 min 5 % B → 100 % B; 10–15 min 100 % B; 15–15.10 min 100 % B → 5 % B; 15.10-20 min 5 % B.

The long method had the following gradient with a flow rate of 0.25 ml min^{-1} and a temperature of 20 °C: 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 by a mass-spectrometer micrOTOF-Q III with ESI source (Bruker Daltonics) calibrated with 5 mM sodium formate using the negative mode. Usually the mass-spectrometer was calibrated for low masses (m/z 100-400). In the later course of this work, mass-spectrometer was recalibrated for higher masses (m/z 400-1000) if samples from CoA-ligase assays were analyzed.

4.3 Production of substrates for CYP98

4.3.1 *p*-Coumaroyl-4´-hydroxyphenyllactic acid (pC-PHPL) and caffeoyl-4´-hydroxyphenyllactic acid (Caf-PHPL)

The precursors of rosmarinic acid, pC-4´-hydroxyphenyllactic acid (pC-PHPL) and Caf-4´-hydroxyphenyllactic (Caf-PHPL) were isolated from *Melissa officinalis* suspension cultures treated with the cytochrome P450 inhibitor tetcyclacis. Tetcyclacis (27.35 mg ml^{-1}) was dissolved in acetone and filter-sterilized. 100 µM tetcyclacis was added to a 4 day old suspension culture of *Melissa officinalis* (incubated in CB-2 medium at 25 °C in the dark at 100 rpm) and the cells were incubated for another 7 days under the same conditions. In this time the cells turned form a yellow-brownish colour to salmon-pink. Afterwards the cells were filtered and stored at -20 °C until further use. Acidified water (0.01 % H_3PO_4) was heated to 100 °C, added to the cells (1 ml/mg cell fresh weight) and the mixture was incubated at 80 °C for 20 min. After 10 min the cell suspension was mixed vigorously. Mushed cells were filtered off from the fluid and the supernatant was additionally centrifuged for 10 min at 300 g to remove the remaining cell residues. The resulting supernatant was extracted three times with 25 ml EtOAc. The organic phase was dried by evaporation and the remaining solid was resuspended in 2 ml EtOAc. The cell extract was applied onto a preparative silica gel TLC plate (DC Kieselgel 60 F_{254}) and developed in EtOAc:chloroform:formic acid (5:4:1) as mobile phase. Solutions of RA, Caf-PHPL and pC-PHPL (5 µl of a 10 mM stock solution in 50 % MeOH) were applied as a standard. The expected bands were scraped off and eluted four times with 1 ml MeOH. The concentration of the expected products was determined photometrically measuring

the absorption from 200-600 nm. Values obtained at 312 nm (pC-PHPL) and 328 nm (Caf-PHPL) were used with the extinction coefficients at this wavelength with 11.94 mM^{-1} cm^{-1} for pC-PHPL and 9.15 mM^{-1} cm^{-1} for Caf-PHPL to determine and calculate the concentration. The purity of the isolated products was checked by HPLC using an isocratic elution with 35 % MeOH with 0.01 % H_3PO_4.

4.3.2 *p*-Coumaroylshikimic acid (pC-Shik), *p*-coumaroylquinic acid (pC-Quin) and *p*-coumaroyl-3´-hydroxyanthranilic acid (pC-3OH-An)

4.3.2.1 Chemical Synthesis of pC-CoA

For the enzymatic synthesis of pC-Shik, pC-Quin and pC-3OH-An, pC-CoA had to be chemically synthesized and purified. First, 25.2 mg $NaHCO_3$ was dissolved in 6 ml H_2O and the solution was flushed with nitrogen for 15 min to remove the remaining oxygen. Then 23 mg CoA was added and the mixture was flushed for another 15 min. 56.4 mg 4-coumaroylsuccinimide (already synthesized in our lab previously) was dissolved in 200 µl acetone and added to the aqueous solution. After the filling level was marked, 4-6 ml acetone was slowly added until the yellow precipitate disappeared. The solution was incubated overnight in the dark at room temperature. On the next day, the acetone was evaporated with a stream of N_2. After total removal of the acetone, the reaction was centrifuged for 15 min at 15000 g. The supernatant was concentrated under vacuum. The remaining solvent was applied to a preparative TLC plate (Cellulose F, Merck) and a mixture of 1-butanol:acetic acid:H_2O (5:2:3) was used as a mobile phase. The bands corresponding to the standard, visualized under UV light at 254 nm, were scraped off and transferred to fresh reaction tubes. The cellulose was eluted with 2 ml water and centrifuged for 10 min at 5000 g. The supernatant was checked photometrically (diluted 1:50). Pure pC-CoA should display the highest absorption at 333 nm. Elution was repeated until no more product was detected. The aqueous solution was concentrated by evaporation. Finally the concentration was determined photometrically (diluted 1:200) using the extinction coefficient at 333 nm of 21 mM^{-1} cm^{-1}.

4.3.2.2 Enzymatic synthesis of pC-Shik, pC-Quin and pC-3OH-An

pC-Shik, pC-Quin and pC-3OH-An were produced enzymatically by HCTs from *Anthoceros agrestis* and *Sarcandra glabra* using pC-CoA and the corresponding acid. The sequences encoding these enzymes were identified by Maike Petersen and the enzymes are currently characterized by Lucien Ernst and Paul Bömeke. Enzyme assays to produce pC-Shik and pC-Quin were performed by Maike Petersen or Paul Bömeke. Assays to produce pC-3OH-An were composed as follows:

100 µl	potassium phosphate buffer (0.1 M, pH 7)
10 µl	pC-CoA (2.5 mM)
5 µl	ascorbic acid (12.5 mM)
5 µl	3-hydroxyanthranilic acid (10 mM)
5 µl	AaHCT6

All enzyme assays were incubated up to 3 h at 30 °C and then heated to 5 min at 95 °C. Except for pC-Shik, products were used directly without further purification. Samples were stored at -20 °C until use. pC-Shik was isolated by Olga Haag by HPLC, resulting in a small amount (~ 1 mg) of purified pC-Shik.

4.3.3 *p*-Coumaroylanthranilic acid (pC-An) and *p*-coumaroyltyramine (pC-Ty)

pC-Ty and pC-An were synthesized in the group of Prof. Dr. Martin Schlitzer (Philipps-Universität Marburg) under the supervision of Rolf Erik Emmerich and with the help of Jan Walter. All steps of the reactions or column chromatographic purification were checked by TLC (DC Kieselgel 60 F_{254} as solid phase and EtOAc/cyclohexane in different ratios as mobile phase).

For the synthesis of pC-Ty one equivalent (0.3 g) 4-coumaric acid was dissolved in 10 ml acetonitrile, then NEt_3 (0.642 ml, 2.5 equivalents) and EDC*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 over night (Fig. 10). The solvent was evaporated and the remaining solid was resuspended in 50 ml water and extracted three times with 50 ml EtOAc. After evaporation of the organic solvent the solid was purified by column chromatography using silica gel as solid phase

with isocratic elution using EtOAc/cyclohexane (3:1). Fractions containing pC-Ty were pooled and evaporated until dryness. The product was checked by ^{1}H-NMR.

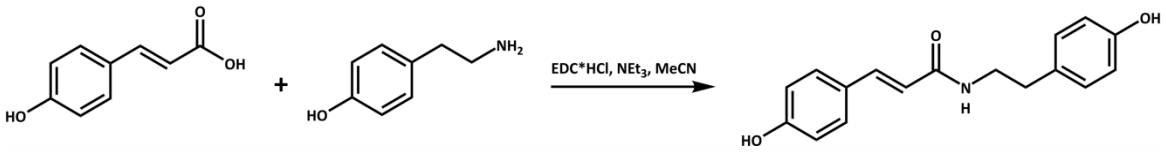


Figure 10 Chemical synthesis of pC-Ty. Abbreviations are EDC*HCl – 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride; NEt_3 – triethylamine; MeCN – acetonitrile.

While pC-Ty was synthesized in a one step reaction, this reaction failed for pC-An. Therefore, synthesis of pC-Aa was done according to Alber et al. (2019) with some modifications (Fig. 11). At first, 4-coumaric acid (5 g, 1 equivalent) was dissolved in 8 ml pyridine and 7.2 ml acetic anhydride (2.5 equivalents) was added (**1**). The solution was stirred for 72 h at room temperature and then for 0.5 h on ice. The mixture was acidified with 6 N HCl (5 ml) and the precipitated solid was filtered. The solid was washed with H_2O (20 ml) and diethyl ether (20 ml) and dried by evaporation. The product was checked by ^{1}H-NMR and then used directly in the next reaction. 1 equivalent of the (*E*)-3-(4-acetoxyphenyl)acrylic acid (0.5 g) was dissolved in 6 ml dichloromethane and stirred on ice after addition of oxalylchloride (0.42 ml, 2 equivalents). Then 4 drops dimethylformamide were added and the solution was incubated for 2 h at room temperature (**2**). After removal of the solvent by evaporation, the remains were co-evaporated with cyclohexane. The resulting 4-(acetyloxy)-(*E*)-cinnamoyl chloride was used in the next step without any further purification. It was resuspended in 8 ml pyridine and anthranilic acid was added (0.185 g, 0.55 equivalents) to obtain 4-(acetyloxy)-cinnamoylanthranilic acid (**3**). The solution was stirred at room temperature for 16 h. To check the reaction by TLC, 200 µl were transferred to a fresh reaction tube, acidified with 1 N HCl (1-2 drops) and extracted with EtOAc (400 µl). The upper phase was again mixed with 1 N HCl (200 µl) to remove the remaining pyridine. The remaining upper phase was checked by TLC with EtOAc/cyclohexane (1:1, acidified with 1-2 drops of formic acid) as mobile phase. The product displayed a R_f value of 0.625. Afterwards the remaining solution was acidified with 1 N HCl (4 ml) and extracted three times with 30 ml EtOAc. To remove the remaining pyridine, the organic phase was again extracted

with 1 N HCl (approximately 20 ml). The EtOAc was removed by evaporation and the solid was purified by column chromatography using silca gel as solid phase and a mobile phase of EtOAc/cyclohexane (1:4, acidified with 0.1 % formic acid). Finally, the acyl group of 4-(acetyloxy)-cinnamoylanthranilic acid was removed by incubation with sodium hydroxide. For this, 1 equivalent of 4-(acetyloxy)-cinnamoylanthranilic acid (60 mg) was dissolved in 5 ml H_2O and 9 ml THF, mixed with NaOH (approximately 20 mg, 3 equivalents) and stirred at room temperature for 15 min (**4**). The product was checked by TLC every 5 minutes by transferring 200 µl into a fresh reaction tube. The solution was acidified with 1 N HCl (1-2 drops) and extracted with EtOAc (400 µl) before the organic phase was applied to the plate. The separation was achieved using a mobile phase with EtOAc/cyclohexane (1:1, acidified with 1-2 drops of formic acid). pC-An displayed a R_f value of 0.31. The remaining solution was acidified with 1 N HCl (pH 1) and extracted four times with 20 ml EtOAc. The organic solvent was removed by evaporation and the remains were co-evaporated with cyclohexane to remove the remaining acetic acid. Since the substrate was completely transformed to pC-An after 15 min no further purification was necessary. The product was checked by ^{1}H-NMR.

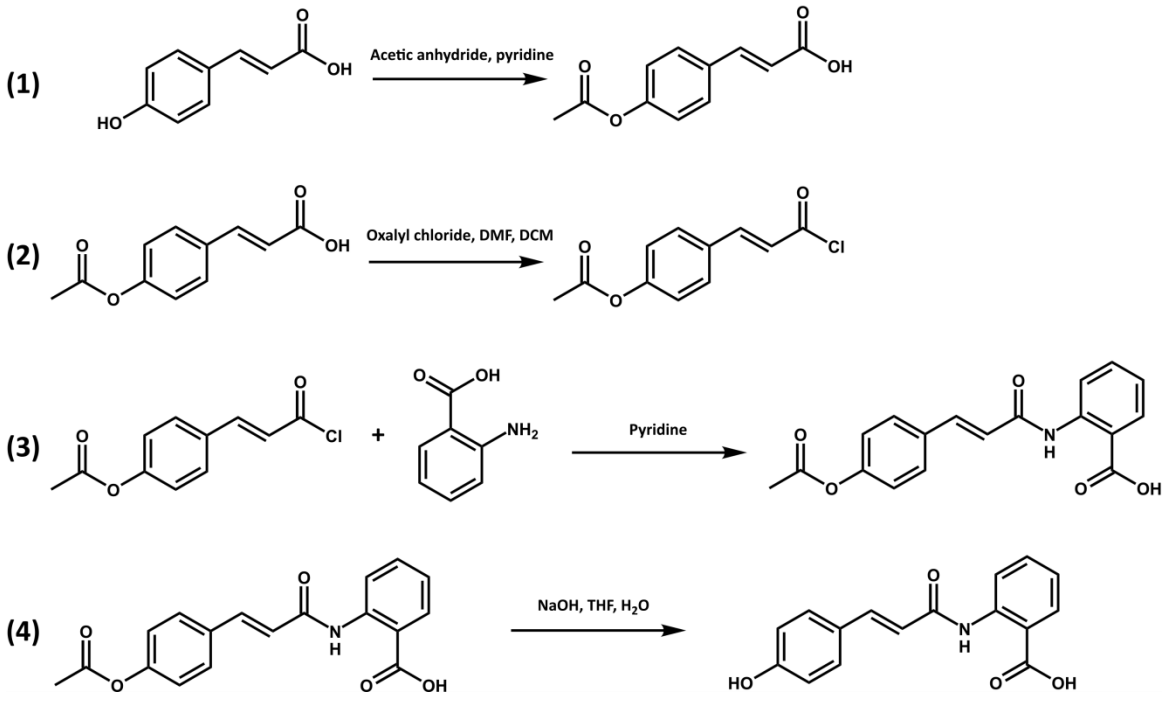


Figure 11 Chemical synthesis of pC-An. Abbreviations are DMF – dimethylformamide; DCM – dichloromethane; THF – tetrahydrofuran.

5 Culture Media

5.1 Cultivation media for *Escherichia coli* and *Saccharomyces cerevisiae*

Lysogeny Broth (LB)

The standard medium for the cultivation of *Escherichia coli* is LB (Bertani 1951). It is composed as follows:

Component	Concentration
Tryptone / peptone	10 g l^{-1}
Yeast extract	5 g l^{-1}
NaCl	10 g l^{-1}
± Agar	10 g l^{-1}

The medium was autoclaved after adjusting the pH to 7.0 with NaOH. Solid medium for plates additionally contained 1 % agar. After transformation of *E. coli* cells were plated on solid LB. For selection using antibiotic resistance, the medium was adapted with either ampicillin (100 mg l^{-1}) or other suitable antibiotics after cooling to ~ 55 °C. If a lacZ cassette was present on the plasmid, 50 µl l^{-1} IPTG (100 mM, filter-sterilized) and 160 µl l^{-1} X-Gal (50 mg l^{-1} in DMF) were additionally added.

Terrific Broth (TB)

TB medium (Hobbs and Tartoff 1987) was used for expression of recombinant proteins in *E. coli*. The medium is compatible with high cell densities, maintaining the organism in the logarithmic growth phase over a longer time period than LB. Thus, this often provides a higher yield of recombinant proteins.

Component	Concentration
Tryptone / peptone	12 g l^{-1}
Yeast extract	24 g l^{-1}
Glycerol	4 ml l^{-1}

All components were dissolved in 900 ml H_2O and autoclaved. After the solution cooled to ~ 60 °C, 100 ml buffer (0.17 M KH_2PO_4 and 0.72 M K_2HPO_4) was added.

Super Optimal broth with Catabolite repression (SOC)

SOC medium (Hanahan 1983) is a nutritious bacterial growth medium modified from the LB medium. It was used to regenerate bacteria after heat-shock transformation, because it is reported to provide higher transformation efficiency. It is composed as follows:

Component	Concentration
Tryptone / peptone	20 g l^{-1}
Yeast extract	5 g l^{-1}
NaCl	0.5 g l^{-1}

The solution was autoclaved, after adjustment to pH 7.0 with 1 M NaOH. After cooling, the solution was supplemented with filter-sterilized KCl (final concentration 2.5 mM), $MgSO_4$ (final concentration 10 mM) and glucose (final concentration 20 mM).

Yeast extract Peptone Dextrose medium (YPD)

YPD (Dymond 2013) is a specially adapted medium used for *S. cerevisiae* strains for simple cultivation purposes. The main focus was on the accumulation of cell mass of untransformed auxotrophic yeast strains. It is composed as follows:

Component	Concentration
Tryptone / peptone	20 g l^{-1}
Yeast extract	10 g l^{-1}
Glucose	20 g l^{-1}
L-adenine-sulphate	55 mg l^{-1}
Agar	15 g l^{-1}

The solution was autoclaved and the medium was poured into Petri dishes after cooling to ~ 55 °C. Plates were stored at 4 °C in the dark.

Synthetic Complete minimal defined medium (SCD/SCG) without uracil

SC medium (Dymond 2013) (supplemented with either glucose or galactose) was used if selective growth was required. The medium lacks the nucleobase uracil. Thus, it was used for expression of recombinant proteins in *S. cerevisiae*. Medium with glucose was used primarily for biomass accumulation of pre-cultures and the medium containing galactose for expression.

Component	Concentration
Yeast nitrogen base	6.7 g l^{-1}
Glucose/galactose	20 g l^{-1}
L-adenine-sulphate	65 mg l^{-1}
Phenylalanine	50 mg l^{-1}
Isoleucine, leucine, serine, threonine, tyrosine, valine	30 mg l^{-1} (each)
Tryptophan, methionine, lysine, histidine, arginine	20 mg l^{-1} (each)

Amino acids were prepared as a mixed stock solution (10x), divided into individual portions of 50 ml (stored at -20 °C) and added prior to autoclaving. The pH value was adjusted to pH 5.6 with 0.5 M HCl. For solid SC medium 15 g l^{-1} agar was added and plates were stored at 4 °C.

5.2 Media for plant cell cultures

CB-M

As a standard medium for the cultivation of suspension cells of *Anthoceros agrestis*, CB-M (Petersen and Alfermann 1988; Gertlowski and Petersen 1993) was used. CB-M is a modified version of CB-2 without hormones and NZ-amines and with only 1 % sugar. It is composed as follows:

	Component	Concentration
Macroelements	KNO_3	2500.0 mg l^{-1}
	$MgSO_4 \cdot 7\ H_2O$	250.0 mg l^{-1}
	$NaH_2PO_4 \cdot H_2O$	172 mg l^{-1}
	$CaCl_2 \cdot 2\ H_2O$	150.0 mg l^{-1}
	$(NH_4)_2SO_4$	134.0 mg l^{-1}
	$FeSO_4 \cdot 7\ H_2O$	25.6 mg l^{-1}
	Na_2-EDTA	34.3 mg l^{-1}
Microelements	H_3BO_3	3.00 mg l^{-1}
	$ZnSO_4 \cdot 7\ H_2O$	3.00 mg l^{-1}
	$MnSO_4 \cdot H_2O$	1.00 mg l^{-1}
	KI	0.75 mg l^{-1}
	$Na_2MoO_4 \cdot 2\ H_2O$	0.25 mg l^{-1}
	$CuSO_4 \cdot 5\ H_2O$	0.25 mg l^{-1}
	$CoCl_2 \cdot 6\ H_2O$	0.25 mg l^{-1}
Vitamins	Thiamine-HCl	10 mg l^{-1}
	Pyridoxine-HCl	1 mg l^{-1}
	Nicotinic acid	1 mg l^{-1}
Further components	Myo-inositol	100 mg l^{-1}
	Sucrose	10 g l^{-1}

After all components were dissolved, the pH was adjusted to pH 5.5 with 0.5 M KOH. At last, 50 ml medium were filled into 250 ml Erlenmeyer flasks each, closed with cellulose stoppers and autoclaved.

CB-2

CB-2 (Petersen and Alfermann 1988) was used for the cultivation of *Melissa officinalis* suspension cells. The medium CB-2 differs from CB-M only in terms of additionally added hormones, NZ-amines and a higher sugar concentration (2 %). Macroelements,

microelementes and vitamins were added in the same concentrations as mentioned for CB-M. Hormones and further components were added as follows:

	Component	Concentration
Hormones	2,4-D	2 mg l^{-1}
	NAA	0.5 mg l^{-1}
	Kinetin	0.2 mg l^{-1}
	IAA	0.5 mg l^{-1}
Further components	Myo-inositol	100 mg l^{-1}
	NZ-Amines	2 g l^{-1}
	Sucrose	20 g l^{-1}

After all components were dissolved, the pH was adjusted to pH 5.5 with 0.5 M HCl. At last, 50 ml medium were filled into 250 ml Erlenmeyer flasks each, closed with cellulose stoppers and autoclaved.

BCD

As a standard medium for the cultivation of *Physcomitrella patens*, BCD was used (Cove et al. 2009a). At first, stock solutions were prepared and the pH of solution C was adjusted to pH 6.5 with concentrated potassium hydroxide:

Stock solutions	Component	Concentration
Solution B	$MgSO_4 \cdot 7H_2O$	25 g l^{-1}
Solution C	KH_2PO_4	25 g l^{-1}
Solution D	KNO_3	101 g l^{-1}
TES (Hoaglands trace elements)	H_3BO_3	614 mg l^{-1}
	$Al_2(SO_4)_3 \cdot K_2SO_4 \cdot 24\ H_2O$	55 mg l^{-1}
	$CuSO_4 \cdot 5\ H_20$	55 mg l^{-1}
	KBr	28 mg l^{-1}
	LiCl	28 mg l^{-1}
	$CoCl_2 \cdot 4\ H_2O$	55 mg l^{-1}
	$ZnSO_4 \cdot 7\ H_2O$	55 mg l^{-1}
	KI	28 mg l^{-1}
	$SnCl_2$	28 mg l^{-1}
	$MnCl_2 \cdot 4\ H_2O$	389 mg l^{-1}

For BCD medium, the stock solutions were mixed and supplemented with the following components:

Component	Concentration
Solution B	10 ml l^{-1}
Solution C	10 ml l^{-1}
Solution D	10 ml l^{-1}
TES	1 ml l^{-1}
$CaCl_2$	219 mg l^{-1}
$FeSO_4 \cdot 7\ H_20$	12.8 mg l^{-1}
Na_2-EDTA	17.15 mg l^{-1}

For solid BCD medium 7 g l^{-1} agar was added. If cells were cultivated for protoplastation, the medium was supplemented with 920 mg l^{-1} (5 mM) di-ammonium tartrate (≙ **BCDA** medium). After all components were dissolved, 50 ml medium were filled into 250 ml Erlenmeyer flasks each, closed with cellulose stoppers and autoclaved. Solid BCDA was supplemented with 25 mg l^{-1} hygromycin B (filter-sterilized stock solution of 25 mg ml^{-1}) after the medium had cooled down to 55 °C and poured into Petri dishes.

Protoplast regeneration medium, bottom layer (PRMB)

PRMB (Cove et al. 2009b) is a modified version of BCD with an additional carbohydrate source (D-mannitol). All stock solutions prepared for the BCD medium were also used for PRMB. After transformation of *Physcomitrella patens* protoplasts were transferred to solid PRMB in Petri dishes covered with cellophane. The stock solutions described for BCD were mixed and supplemented with the following components:

Component	Concentration
Solution B	10 ml l^{-1}
Solution C	10 ml l^{-1}
Solution D	10 ml l^{-1}
TES	1 ml l^{-1}
$CaCl_2$	219 mg l^{-1}
$FeSO_4 \cdot 7\ H_20$	12.8 mg l^{-1}
Na_2-EDTA	17.15 mg l^{-1}
di-ammonium tartrate	920 mg l^{-1}
D-mannitol	60 g l^{-1}
Agar	7 g l^{-1}

After all components were dissolved the medium was autoclaved. After the solution cooled to ~ 55 °C it was poured into Petri dishes. After the medium had solidified it was covered with cellophane (pre-autoclaved between layers of filter paper).

Protoplast regeneration medium, top layer (PRMT)

After transformation of *P. patens* the protoplasts were mixed with PRMT (previously melted in the microwave and cooled to 45 °C) (Cove et al. 2009b) and quickly poured on PRMB plates overlaid with sterile cellophane. Again, the stock solutions described for BCD were mixed and the following components added:

Component	Concentration
Solution B	10 ml l^{-1}
Solution C	10 ml l^{-1}
Solution D	10 ml l^{-1}
TES	1 ml l^{-1}
$CaCl_2$	219 mg l^{-1}
$FeSO_4 \cdot 7\ H_20$	12.8 mg l^{-1}
Na_2-EDTA	17.15 mg l^{-1}
di-ammonium tartrate	920 mg l^{-1}
D-mannitol	60 g l^{-1}
Agar	4 g l^{-1}

After all components were dissolved, 30 ml medium were filled into 100 ml Erlenmeyer flasks each, closed with cellulose stoppers and autoclaved.

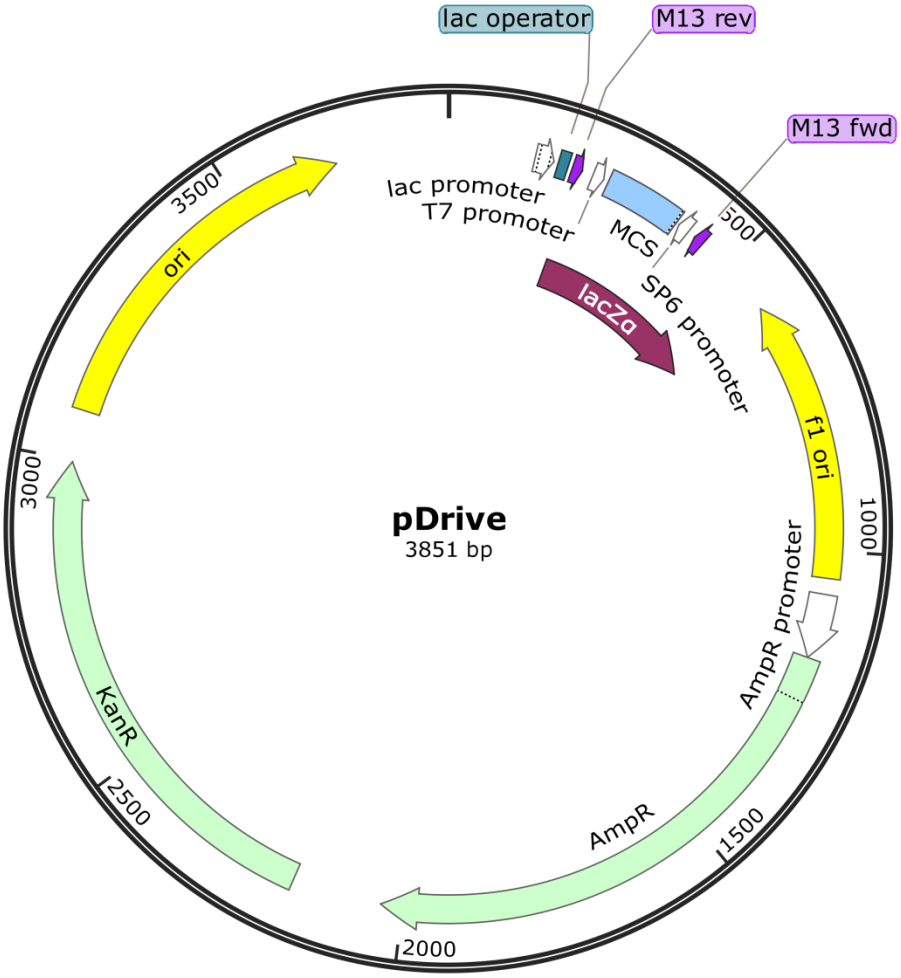
6 Materials and instruments

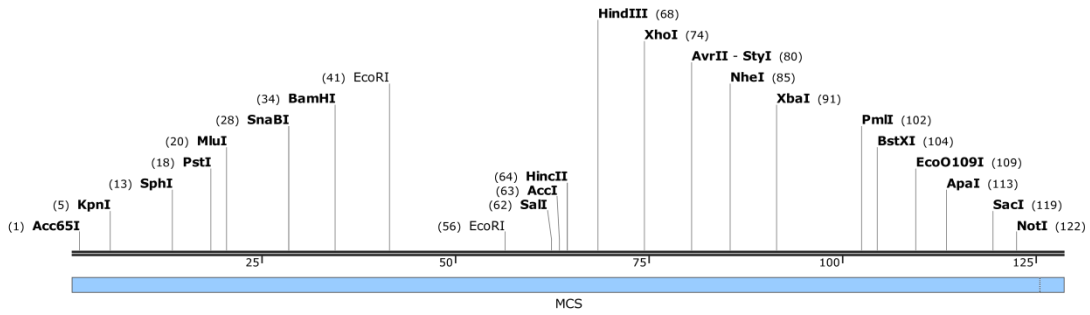
6.1 Genotypes of laboratory strains (*E. coli* and *S. cerevisiae*)

Strain	Genotype
***E. coli* EZ** (Qiagen)	[F'::Tn10(Tc^r) $proA^+B^+lacI^qZ\Delta M15$] recA1 end A1 hsdR17($r_{K12}^-m_{K12}^+$) lac glnV44 thi-1 gyrA96 relA1
***E. coli* BL21-CodonPlus (DE3)-RIPL** (Agilent Genomics, provided by Prof. Dr. Klebe, Marburg)	B F^-ompT hsdS(rB^-mB^-) dcm^+Tetrgal λ(DE3) endA Hte [argU proL Camr] [argU ileY leuW Strep/Specr]
***E. coli* SoluBL21** (amsbio)	F^-ompT hsdSB (rB^- mB^-) gal dcm (DE3)†
***E. coli* C41(DE3)** (Lucigen, provided by Prof. Dr. Dietrich Ober, Kiel)	F^- *ompT hsdSB* (rB^- *mB^-) gal dcm* (DE3)
***E. coli* DH5α** (Thermo Fisher)	F^-, endA1, hsdR17 (rk^-, mk^-), supE44, thi-1, recA1,gyrA96, relA1, Φ80d, lacZ[Δ]M15
***S. cerevisiae* BY4741** (provided by Dr. Backhaus, Marburg)	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0
***S. cerevisiae* BY4742** (provided by Dr. Backhaus, Marburg)	MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0
***S. cerevisiae* CB018** (provided by Dr. Backhaus, Marburg)	Matα his3-11,15 leu2-3,112 trp1-1 ura3-1 ade2-1ochrecan1-100 pep4Δ::HIS3 prb1Δ::hisG prc1Δ::hisG
***S. cerevisiae* InvSc1** (Invitrogen)	MATa his3Δ1 leu2 trp1-289 ura3-52 MAT his3D1 leu2 trp1-289 ura3-52

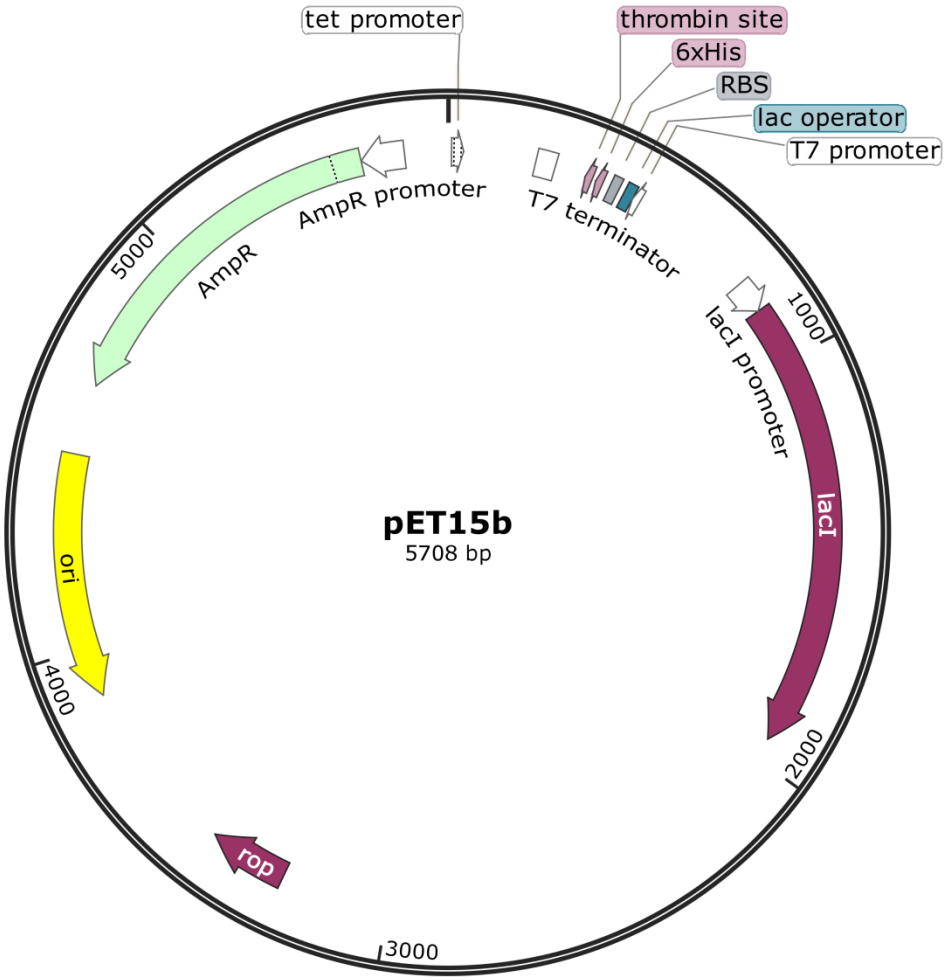
6.2 Vector maps

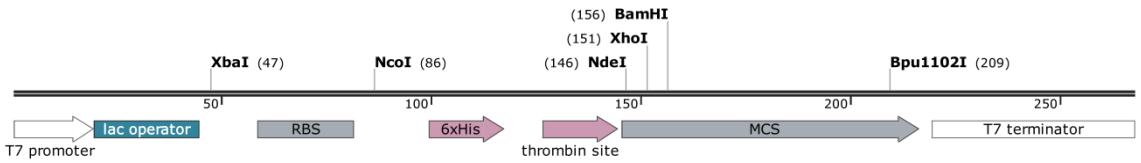
pDrive (Qiagen)



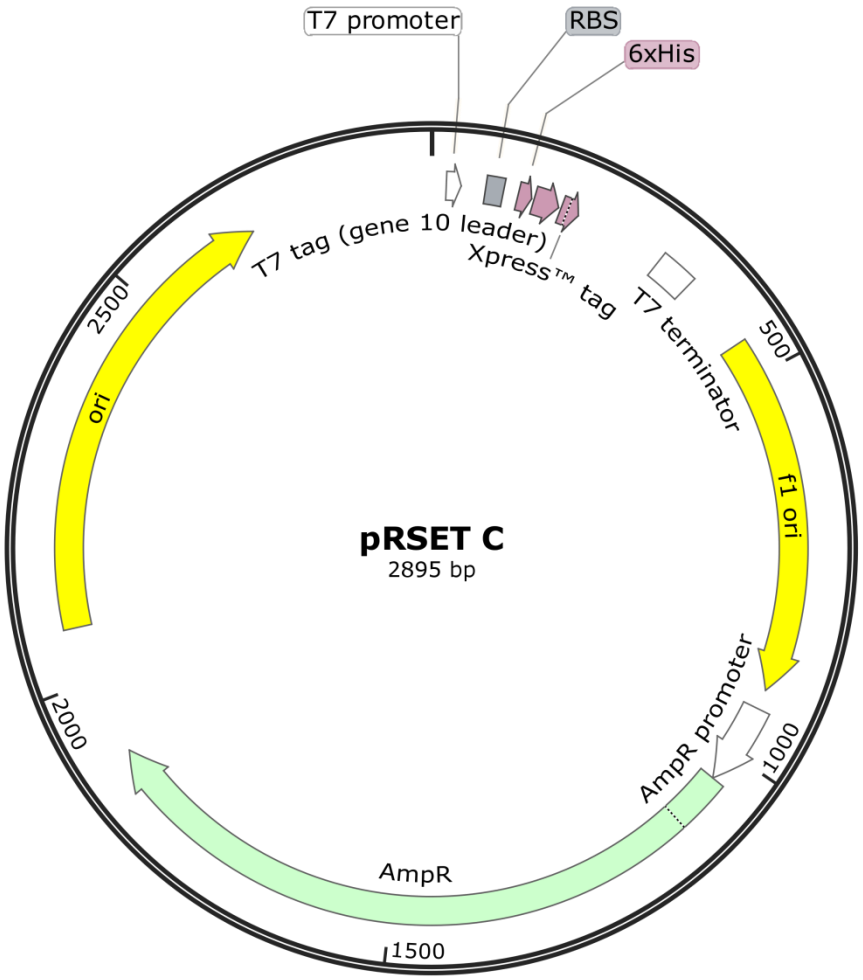


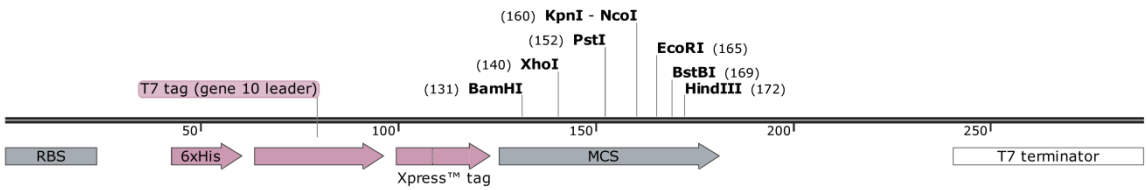
pET15b (Novagen)



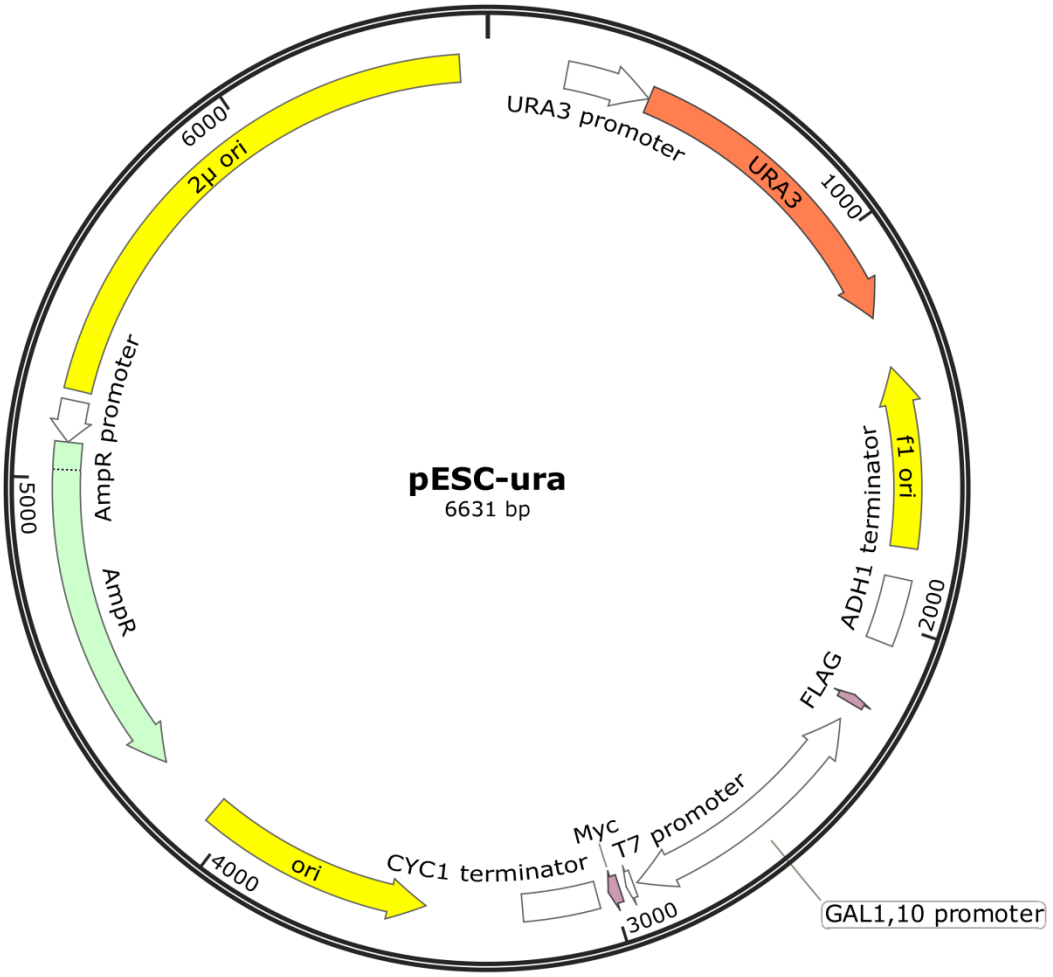


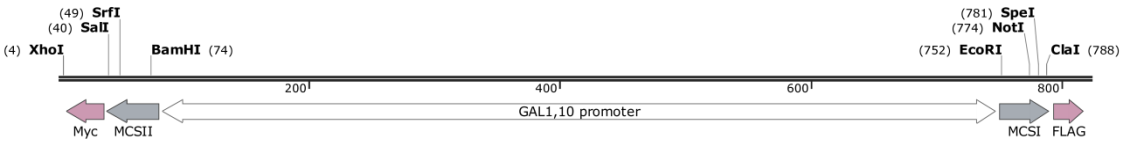
pRSET C (Invitrogen)



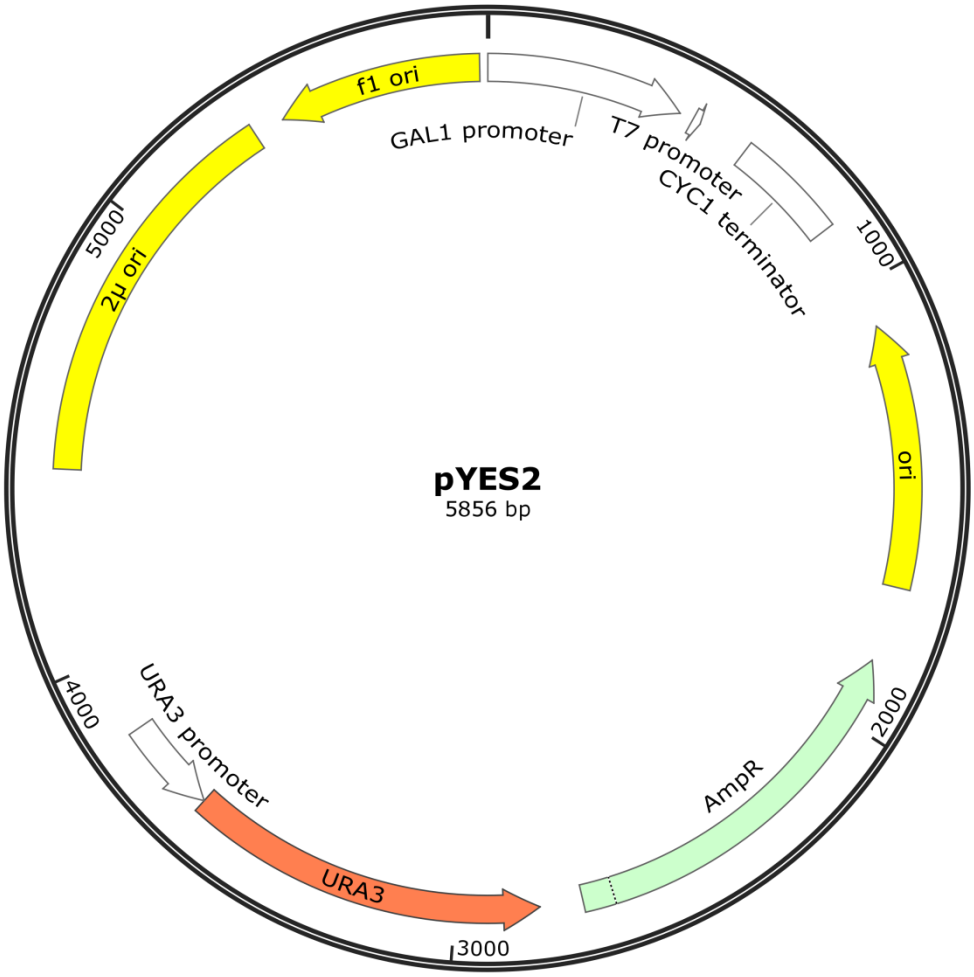


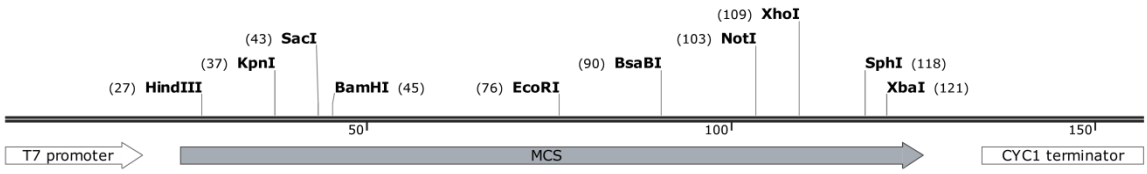
pESC$_{-ura}$ (Agilent Technologies)



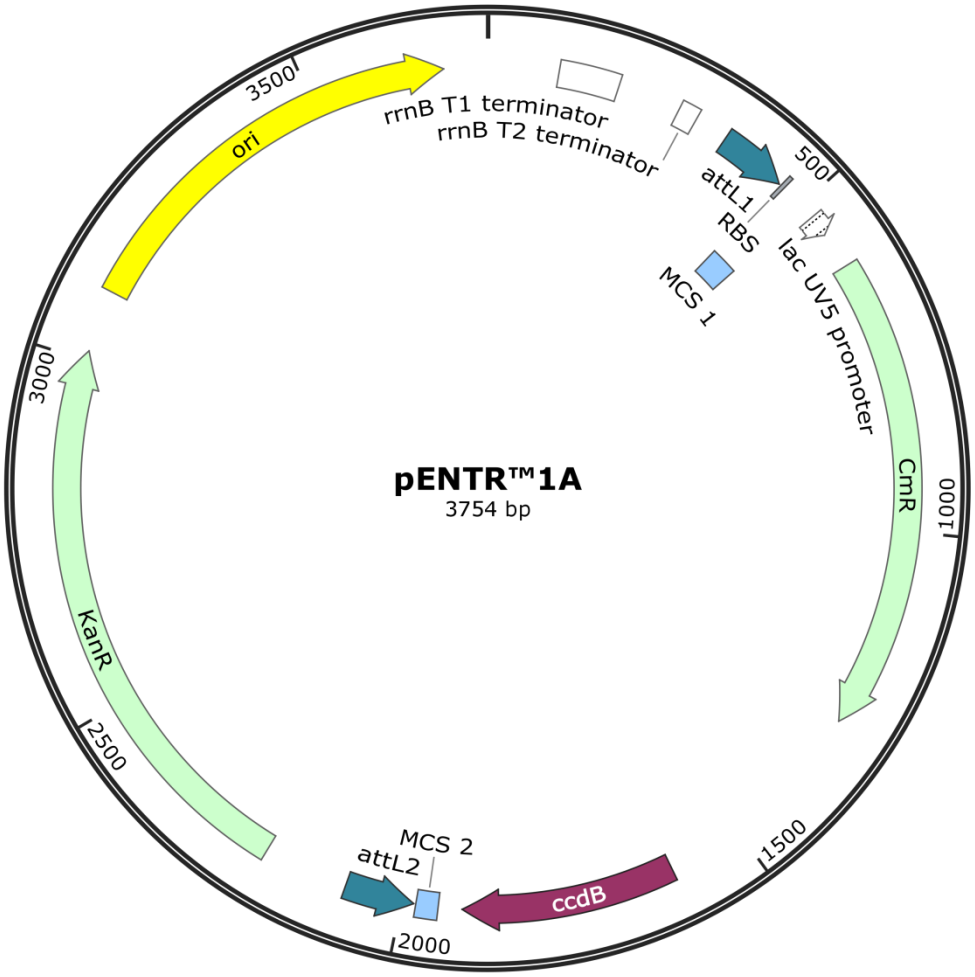


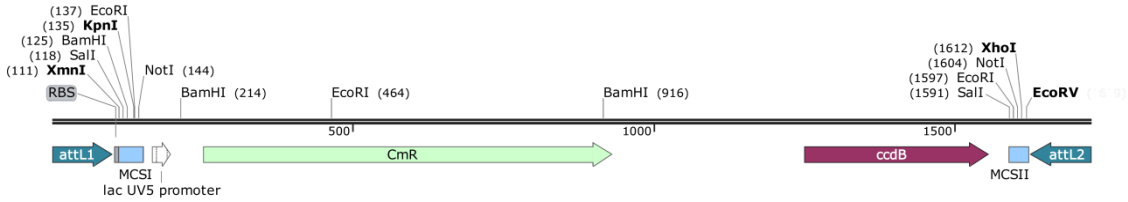
pYES2 (Invitrogen)



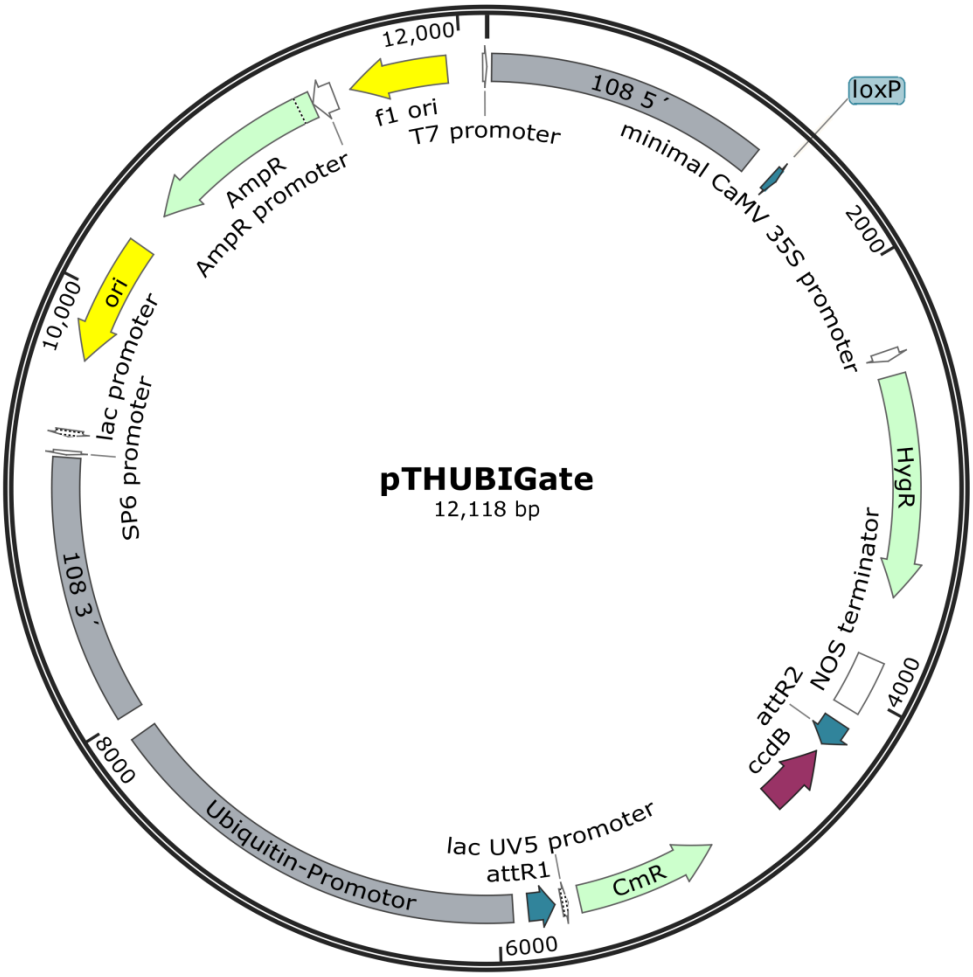


pENTR™1A (Invitrogen)





pTHUbiGate (kindly provided by Prof. Dr. S. Rensing, Philipps-Universität Marburg)



6.3 Primer list

Binding sequence underlined, **restriction sites written in bold letters**, ribosomal binding site marked red, His-Tag marked green, sequence encoding the mammalian peptide ε is marked blue

Partial length primers

Name	Sequence (5´-3´)	Restriction site	Comment
Aa11181_f	ATGGCTTCCGGTGAAACCAC	-	
Aa11181_r	GACAAATAGGGGATCGTCCTCG	-	
Aa10969_f	ATGGCGGTCGCTGCC	-	
Aa10969_r	CCACACATCCCTCAAGTAG	-	
AaAp6378_f	TTCCGCTCCAACTTGCCCG	-	
AaAp6378_r	ATCCTCCAGCTCCTTCCCC	-	
Aa35279_f	TCTTCGGGGACAACAGGACTG	-	
Aa35279_r	AACCTGGAACCCTTTGTACTTGATG	-	
Aa20832_f	TCCTCCGGTAACACGGGTTTG	-	
Aa20832_r	CTTACGCAATATTTTTCCTGAAGCCAAC	-	
Aa19917_f1	TCCTCTGGCACCACGGG	-	
Aa19917_r1	GCGTAGTATCTTTCCTGAAGGC	-	
Aa19917_f2	GGTAATCCCGTGAATGGAACC	-	
Aa19917_f3	GACTTCCGCCCTGGGAGCAACCTTTCCTG	-	
Aa19917_r2	GGTATCTCTCCAGCGATATCG	-	
Aa26091_f	ATGGTGCACAAGATCCACC	-	
Aa26091_r	GACGTTTCTGGCATACAAGG	-	
AaAp626_f	CTTGGACAGGATCTCACTGGT	-	
AaAp626_r	CTACAGATGGGTCTTGTACACG	-	

RACE primers

Name	Sequence (5´-3´)	Restriction site	Comment
Aa11181_5´R	GATTACGCCAAGCTTGGTGAAGAAGGGCACAGTCATGATCC	-	5´-overhang for In-Fusion cloning in pRACE
Aa11181_3´R	GATTACGCCAAGCTTGCAACCTGTCGGACCTTGCGAAG	-	5´-overhang for In-Fusion cloning in pRACE
Aa10969_5´R	GCTCACCATCCCCATAGGTGGCAATCAT	-	
AaAp6378_5´R	CTCGACACCACCTCCACCTCGCTGTAG	-	
AaAp6378_3´R	GTGATGCTGTGCGGGCTGCGTGCT	-	
Aa35279_5´R	CCGTTGTGCCGTAAATGTGGAAGAAGGGC	-	
Aa35279_3´R	GGAAAGTCCCTGCCTGCAAATCGTG	-	
Aa20832_5´R	CCGTATATGTGAAACTGAGGGACCACGC	-	
Aa20832_3´R	GCTGCCTATGGCTTTCGTTGTTCGCTTAC	-	
Aa19917_5´R	GGTTTGCCAGTGTGAGCACGACGAGGCC	-	
Aa19917_3´R	GACTTCCGCCCTGGGAGCAACCTTTCCTG	-	
Aa26091_5´R	GATTACGCCAAGCTTCCTTCCACACAGCAGGATCCCTGGC	-	5´-overhang for In-Fusion cloning in pRACE
Aa26091_3´R	GATTACGCCAAGCTTCGTTGACGCTCTGCTAGGGCTGC	-	5´-overhang for In-Fusion cloning in pRACE
AaAp626_5´R	GCCTCTATATAGGGCAGCTTCTGGAGGTCAG	-	

Full-length primers for expression in *S. cerevisiae*

Name	Sequence (5´-3´)	Restriction site	Comment
AaC4H_pESC_f	GCAT**GAATTC**ATGGCTTCCGGTGAAACCAC	EcoRI	
AaC4H_pESC_r	GCAT**GCGGCCGC**CATATCAGGGCGTGGCTTGA	NotI	
AaCPR_pYES_f	AT**GGATCC**AACACAATGGCGGTCGCTGCCGCG TTG	BamHI	
AaCPR_pYES_r	ATA**TCTAGA**TCAATGATGATGATGATGATGCC ACACATCCCTCAAGTAGCGACC	XbaI	
AaCYP98_pESC_f	GCAT**GAATTC**ATGGCGGAGCTGGTGGTGA	EcoRI	
AaCYP98_pESC_r	GCAT**GCGGCCGC**GACGTTTCTGGCATACAAGG CC	NotI	
CbCPRr	TAT**GTCGAC**CCACACGTCACGCAGTACCG	SalI	Restriction site was changed from XbaI to SalI

Full-length primers for expression in *E. coli*

Name	Sequence (5´-3´)	Restriction site	Comment
AaC4H_fus1_f	TAT**TCTAGA**ATGGCTTCCGGTGAAACCACTCT TGGCAGCACAAACACGATGTTCC	XbaI	
AaC4H_fus2_f	TAT**TCTAGA**ATGGCTCTGTTATTAGCAGTTTT TACTCTTGGCAGCACAAACACGATGTTCCTCG ATGCGCT	XbaI	
AaC4H_fus3_f	TAT**TCTAGA**ATGGCTCTGTTATTAGCAGTTTT TACGAGCAGCAAATTGAAGCTCCCACCAGGA CCCCTGG	XbaI	
AaC4H_fus_r	TAT**ACTAGT**CATATCAGGGCGTGGCTTGACA ACGATGTTAGAGTGAGCGGC	BcuI	
AaC4H_fus1_rbs_f	TAT**TCTAGA**AATAATTTTGTTTAACTTTAAGA AGGAGATATACCATGGCTTCCGGTGAAACCA CTCTTGGCAGCAC	XbaI	Primer with additional rbs of pET15b
AaC4H_fus2_rbs_f	TAT**TCTAGA**AATAATTTTGTTTAACTTTAAGA AGGAGATATACCATGGCTCTGTTATTAGCAG TTTTTACTCTTGGCAGCACAAACACG	XbaI	Primer with additional rbs of pET15b
AaC4H_fus3_rbs_f	TAT**TCTAGA**AATAATTTTGTTTAACTTTAAGA AGGAGATATACCATGGCTCTGTTATTAGCAG TTTTTACGAGCAGCAAATTGAAGCTCC	XbaI	Primer with additional rbs of pET15b
AaCPR_fus_f	TAT**ACTAGT**ACAGGGGGGAAGGTTGTGGAG GAGACCCATGTGTACAAGG	BcuI	

AaCPR_fus_r	TAT**GGATCC**TCAATGATGATGATGATGATGCCACACATCCCTCAAGTAGCGACCGTCC	BamHI
Aa4CL_1_pRSET_f	TAT**CTGCAG**ATGGCGCCGATTCTTGCACCTCC	PstI
Aa4CL_2_pRSET_f	TAT**CTGCAG**ATGCCTGCGGAGATGGAGGCC	PstI
Aa4CL_pRSET_r	TAT**AAGCTT**CTACACCCTGCTCCTCAGCTCC	HindIII
Aa4HBCL_pRSET_f	TAT**CTCGAG**ATATGGCCTCTCTCTCCGAGCC	XhoI
Aa4HBCL_pRSET_r	TAT**AAGCTT**TCAAATTAGTTGGTTCTCAGTCTTCTTGAG	HindIII
Aa20832_pRSET_f	TAT**GGATCC**ATATGTCGGATCGCTATGGTGAGAACT	BamHI
Aa20832_pRSET_r	TAT**GAATTC**TTAAAGCTTGGCTGATGTGAGACCG	EcoRI

Full-length primers for expression in *P. patens*

Name	Sequence (5´-3´)	Restriction site	Comment
AaC4H_pTHUbi_f	TTA**GTCGAC**AACCATGGCTTCCGGTGAAACCACTCTTGGCAGC	SalI	
AaC4H_pTHUbi_r	TAT**GAATTC**TCAATGATGATGATGATGATGCATATCAGGGCGTGGC	EcoRI	
AaCPR_pTHUbi_f	TAT**GGATCC**AACCATGGCGGTCGCTGCCGCGTTGGG	BamHI	
AaCPR_pTHUbi_r	TAT**GCGGCCGC**TCAATGATGATGATGATGATGTGCCACATCCCTCAAGTAGC	NotI	
AaCYP98_pTHUbi_f	ATT**GTCGAC**AACCATGGCGGAGCTGGTGGTGAGGGCG	SalI	
AaCYP98_pTHUbi_r	TAT**GAATTC**TCAATGATGATGATGATGATGGACGTTTCTGGCATACAAGGC	EcoRI	
AaAp626_pTHUbi_f	TAT**GCGGCCGC**ATGGCGTTGCTAGACATTGGCAGG	NotI	
AaAp626_pTHUbi_r	TAT**CTCGAG**TCAATGATGATGATGATGATGCAGATGGGTCTTGTACACGTTGACAG	XhoI	

Primer for qRT-PCR

Name	Sequence (5´-3´)	Restriction site	Comment
AaC4H_qPCR_f	CTTCTGGACACAAAGCGCA	-	
AaC4H_qPCR_r	CTCCCTCACTTTCTGCTG	-	
PpC4H_1_qPCR_f	GTTGTCAACCTTGGGACC	-	
PpC4H_1_qPCR_r	CTTCACGTATGCGGGTCT	-	
PpC4H_2_qPCR_f	GTTGAATAGCGTGAATCCTCC	-	
PpC4H_2_qPCR_r	CTCATTTCGAATCCTGGTTTGAAT	-	
St-P 2a_qPCR_f	GTCTAGTTAGTCCTTTGGTCCT	-	
St-P 2a_qPCR_r	GCCTATTTCTATAATGACTCCGT	-	
Act1_qPCR_f	TCCTGACTCTGAAGTACCCC	-	
Act1_qPCR_r	TGATCTGCGTCATCTTCTCC	-	
Act5_qPCR_f	ACCGAGTCCAACATTCTACC	-	
Act5_qPCR_r	GTCCACATTAGATTCTCGCA	-	
EF1α_qPCR_f	AATCATACATTTCACCTCGCC	-	
EF1α_qPCR_r	GATCAGTGGGTAGAAGTGAC	-	
EF1α_qPCR_f2	CAAGATTGGAGGCATTGGAAC	-	
EF1α_qPCR_r2	CCTCTCTTCAGGTCCTTCAC	-	
E2_qPCR_f	TACGGACCCTAATCCAGATGAC	-	
E2_qPCR_r	CAACCCATTGCATACTTCTGAG	-	

Vector specific sequencing or colony-PCR Primers

Name	Sequence (5´-3´)	Restriction site	Comment
pESC_MSCI_f	CTGAAAGTTCCAAAGAGAAG	-	
pESC_MSCI_r	TTTCTGGCAAGGTAGACAAG	-	
pESC_MSCII_f	GCCTTATTTCTGGGGTAATTAATCAGCG	-	
pESC_MSCII_r	GTCCCAAAACCTTCTCAAGCAAGG	-	
pTHUbiGate_f	CCTGTTTAGTTAACACAAGCAACAAA	-	
pTHUbiGate_r	ATGCATTTAAATTTGATTAGAGTTTTATCTAACA	-	

6.4 List of chemicals

Chemical	Company
1-butanol	Roth
1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC · HCl)	Merck
1-naphthaleneacetic acid (NAA)	Duchefa
2,4-dichlorophenoxyacetic acid (2,4-D)	Duchefa
2-aminobenzoic acid, anthranilic acid	Merck
2-coumaric acid	Sigma-Aldrich
2,3-dihydroxybenzoic acid	Sigma-Aldrich
2,4-dihydroxybenzoic acid	Merck
2-amino-3-hydroxybenzoic acid	Sigma-Aldrich
2-hydroxy-3-aminobenzoic acid	Sigma-Aldrich
2-(*N*-morpholino)ethanesulfonic acid (MES)	Roth
3-aminobenzoic acid	Sigma-Aldrich
3-coumaric acid	Sigma-Aldrich
3-hydroxyanthranilic acid	Sigma-Aldrich
3-hydroxybenzoic acid	Sigma-Aldrich
3,4-dihydroxybenzoic acid, protocatechuic acid	Roth
4-aminobenzoic acid	Applichem/ Pan Reac
4-coumaric acid	Sigma-Aldrich
4-hydroxybenzoic acid	Sigma-Aldrich
5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal)	Roth
5-bromo-4-chloro-3-indolyl phosphate (BCIP)	Roth
5-aminolevulinic acid	Roth
acetic anhydride	Merck

acetic acid, glacial	Roth
acetonitrile	Fisher Scientific
acetone	Roth
acrylamide/bisacrylamide (30%, 37.5:1)	Roth
agar-agar	Cero
agarose (SeaKem® LE)	Biozym
aluminium sulfate 16-hydrate	Roth
amidoblack	Alfa-Aesar
ammonium tartrate dibasic	Honeywell
ammonium nitrate	Roth
ammonium persulphate (APS)	Sigma
ammonium sulphate	Roth
ampicillin	Roth
ascorbic acid	Fluka
bovine serum albumin (BSA)	Roth
boric acid	Roth
bromophenol blue	Merck
benzoic acid	Applichem/ Pan Reac
caffeic acid	Alfa-Aesar
calcium chloride dihydrate	Roth
cellophane	Folia® Paper Bringmann
chloroform	Roth
cobalt(II) chloride	Merck
t-cinnamic acid	Merck
coenzyme A-trilithium salt dehydrate, CoA	AppliChem
Coomassie Brilliant Blue G250	Fluka
Coomassie Brilliant Blue R250	Fluka
copper(II) sulphate pentahydrate	Fluka
cyclohexane	Fisher Scientific
cytochrome c from horse heart	Fluka
D-(+)-galactose	Acros Organics / Roth
D-(+)-glucose	Roth
dichloromethane	Fisher Scientific
diethyl ether	Fisher Scientific
dimethylformamide (DMF)	Merck
dipotassium hydrogen phosphate	Roth
disodium hydrogen phosphate dihydrate	Roth
disodium ethylenediaminetetraacetate dihydrate (EDTA-Na_2)	Roth
dithiothreitol (DTT)	Roth
diethyldithiocarbamate (DIECA)	Merck

dNTPs (dATP, dCTP, dGTP, dTTP)	Fermentas
ethylenediaminetetraacetic acid (EDTA)	Merck
ethanol	Roth / Sigma-Aldrich
ethidium bromide	AppliChem
ethyl acetate	Roth
ethylenediaminetetraacetic acid ferric, sodium salt [EDTA-Fe(III)-Na]	Sigma-Aldrich
ferulic acid	Roth
fish sperm (carrier DNA)	Serva
flavin adenine dinucleotide, FAD	Fluka
flavin mononucleotide, FMN	Fluka
Folin-Ciocalteu´s phenol reagent	Merck
formic acid (98%)	Roth
galangin	Alfa Aesar
glycerol	Roth
glycine	Merck
glycolic acid	TCI
guanidine-hydrochloride	Roth
guanidine thiocyanate	Roth
hydrochloric acid (37 %)	Roth
p-hydroxyphenyllactic acid	TCI/ Santa Cruz Biotechnology
hygromycin B	Sigma-Aldrich/Alfa-Aesar
imidazole	Roth
indole-3-acetic acid (IAA)	Duchefa
iron(II) sulphate heptahydrate	Fluka
isoferulic acid	Roth
isopropanol	Roth
isopropyl-β-D-thiogalactopyranoside (IPTG)	Roth
kaempferol	Roth
kanamycin sulphate	Merck
kinetin	Duchefa
L-adenine	Roth
L-arginine	Roth
L-aspartic acid	Roth
L-cysteine	Roth
L-histidine	Roth
liquid nitrogen	Linde
L-isoleucine	Roth
lithium acetate, LiAc	Sigma-Aldrich
lithium chloride	Roth

L-leucine	Roth
L-lysine	Serva
L-methionine	Roth
L-phenylalanine	Roth
L-proline	Roth
L-serine	Roth
L-threonine	Roth
L-tryptophan	Roth
L-tyrosine	Fluka, Merck
L-valine	Roth
lysozyme	Fluka
magnesium chloride hexahydrate	Roth
magnesium sulphate heptahydrate	Merck
manganese(II) chloride tetrahydrate	Roth
manganese(II) sulphate pentahydrate	Duchefa
β-mercaptoethanol	Merck
methanol	Fisher Scientific
myo-inositol	Sigma-Aldrich
NADH	Roth
NADP	Roth
NADPH, sodium salt	Roth
naringenin	Sigma-Aldrich
nickel(II) sulphate hexahydrate	Roth
nicotinic acid	Duchefa
N-lauroylsarcosine sodium salt	Sigma-Aldrich
nitro-blue tetrazolium chloride (NBT)	Roth
N-Z-Amine®, casein hydrolysate	Sigma-Aldrich
oxalyl chloride	Sigma-Aldrich
palmitic acid	Roth
phenol (citrate buffer saturated)	Sigma-Aldrich
phenol/chloroform (1:1)	Roth
phenylmethylsulfonyl fluoride (PMSF)	Roth
phosphoric acid (85%)	Roth
Polyclar™10	Ashland Inc.
polyethylene glycol 4000 (PEG)	Roth
potassium chloride	Roth
potassium cyanide	Merck
potassium dihydrogen phosphate	Roth
potassium ferricyanide	Sigma-Aldrich
potassium fluoride	Roth

potassium hydroxide	Merck
potassium iodide	Merck
potassium nitrate	Roth
pyridine	Arcos organics
pyridoxine hydrochloride	Duchefa
quinic acid	Roth
Rotiphorese® Gel 30	Roth
salicylic acid	Sigma-Aldrich
shikimic acid	Roth
sinapic acid	Roth
sodium acetate trihydrate	Merck
sodium bicarbonate	Merck
sodium carbonate	Fluka
sodium chloride	Roth
sodium citrate	Roth
sodium dihydrogen phosphate monohydrate	Merck
sodium dodecyl sulphate (SDS)	Roth
sodium hydrogen carbonate	Roth
sodium hydroxide	Merck
sodium molybdate dihydrate	Fluka
stearic acid	Roth
sucrose	supermarket
tetcyclacis	BASF
tetracycline	Sigma
tetrahydrofuran	Sigma-Aldrich
tetramethylethylenediamine (TEMED)	Roth
thiamine hydrochloride	Roth
tin(II) chloride	Roth
titriplex III	Roth
triethylamine	Sigma-Aldrich
tris(hydroxymethyl)-aminomethane (Tris)	Roth
tryptone/peptone	Roth
Tween 20	Sigma
tyramine	Acros Organics
vanillic acid	Merck
xylene cyanol	Fluka
yeast extract	Roth
yeast nitrogen base	Conda
zinc chloride	Merck
zinc(II) sulphate heptahydrate	Merck

6.5 Reagents and kits

Product	Company
DNase I, RNase-free (1 U μl^{-1})	Thermo Fisher
Driselase	Sigma-Aldrich
Gateway® Technology	Invitrogen
GeneRuler™1 kb DNA Ladder	Fisher Scientific
GeneRuler™DNA Ladder Mix	Fisher Scientific
goat anti-Mouse IgG Fc	Fisher Scientific
Glucose-6-phosphate dehydrogenase (357 U $mg^{-1)}$	Fluka
GoTaq®Flexi DNA Polymerase Kit (5 U μl^{-1})	Promega
Hexokinase from *S. cerevisiae*	Sigma
mouse DYKDDDDK (FLAG) Tag Monoclonal Antibody	Thermo Fisher
Ni-NTA His-Bind®Superflow™	Novagen
Pfu DNA Polymerase	Promega
NucleoSpin®-Extract II Kit	Macherey-Nagel
PCR Cloning kit	Qiagen
PD-10 Columns Sephadex G-25M	GE Healthcare
PerfeCTa SYBR Green SuperMix	Quanta
Phusion®Polymerase (2 U μl^{-1})	NEB
Pierce™6x-His Epitope Tag Antibody (HIS.H8)	Fisher Scientific
Qiaprep® Spin Miniprep Kit	Qiagen
qScript™ cDNA SuperMix	Quanta
restriction enzymes: BamHI, BcuI, EcoRI, EcoRV, HindIII, NotI, PstI, SalI, XbaI, XhoI	Fermentas
restriction enzyme SwaI	NEB
Revert Aid First Strand cDNA Synthesis Kit	Fisher Scientific

RNase A	Fermentas
Roti®-Mark TRICOLOR Protein marker, prestained	Roth
SMARTer®RACE 5‘/3‘ Kit	Clontech
T4 DNA Ligase (5 U μl^{-1})	Fermentas

6.6 Consumables

Reaction tubes, tips, cuvettes and Petri dishes were all purchased from Sarstedt.

6.7 Instruments and analytical materials

Instrument	Product	Manufacturer/Distributer
-80 °C freezer	C585 Innova	New Brunswick Scientific
autoclaves	Systec VX-150	Systec GmbH
	Systec VX-95	
	AL02-02-100	Advantage–Lab
benchtop homogeniser	Minilys®	Bertin Instr.
blotting chamber	Semi Dry Blotter PROfessional	Roth
Bunsen burner	Flammy S	Schütt
cell culture shakers	Certomat SII	B. Braun Biotech.
	RS-306	Infors AG
	TR-150	
centrifuges	Biofuge 17RS	Heraeus Sepatech
	Sorvall MTX150	Thermo Scientific
	Fresco 17	
	Pico 17	
	Centrifuge 5415D	Eppendorf
	3-30KS	Sigma
freeze dryer	Christ L1	B. Braun Biotech
gel documentation system	FAS-Digi	Nippon Genetics
homogenizer	TH220	OMNI International
horizontal electrophoresis chambers	multiSUB Midi	Cleaver Scientific
	Agagel Mini Biometra	Biomed. Analytik GmbH

HPLC columns	Equisil ODS	Dr. Maisch HPLC GmbH
HPLC systems	Chromaster 5160 Pump + Organizer	VWR/Hitachi
	Chromaster 5280 Auto Sampler	
	Chromaster 5310 Column Oven	
	Chromaster 5430 Diode Array Detector	
	L-6200A Intelligent Pump	Merck/Hitachi
	L-4000 UV Detector	
	D-2500 Chromato-Integrator	
ice machine	AF 80	Scotsman
	RF-0244A	Manitowoc
laminar flow bench	Gelaire Laminar Air Flow Class 100	Gelman Instrument
	Laminar Flow Workstation	Microflow
LC-MS column	Multospher 120 RP 18 – 5 µM	CS-Chromatographie Service GmbH
LC-MS	Agilent 1260 series	Agilent
	micrOTOF-Q III with ESI source	Bruker Daltonics
magnetic stirrer	MR 3001	Heidolph Instr.
mixer	Vortex-Genie 2	Scientific Industries
NMR	ECA-500 MHz spectrometer	JEOL
oven	U40	Memmert
PCR thermocycler	Eppendorf Mastercycler gradient	Eppendorf
	MyCycler	Bio-Rad
PD-10 columns	PD-10 Entsalzungssäulen	GE Healthcare
pH-electrode	Accumed Basic	Fisher Scientific
	HI1330	HANNA instruments
photometer	BioPhotometer	Eppendorf
	Specord 200 plus	Analytik Jena
pipettes	P10ml, P1000, P200, P20, P10N, P2N	Gilson
protein electrophoresis chamber	Mini Protean® 3 Cell	Bio-Rad
rocking platform	Duomax 1030	Heidolph Instr.
rotary evaporator	Rotavapor RE120	Büchi

RT-qPCR Thermocycler	PikoReal96	Thermo Scientific
scales	EG 300-3M	Kern
	440-35A	
	440-47	
	Explorer EX2250	OHAUS
	H64	Mettler
sieves	40, 70, 100 µM	Sarstedt
	51, 100 µM	Bückmann GmbH & Co. KG
shaking incubator	Ecotron	Infors HT
	10X 400	Gallenkamp
thermomixer	Comfort	Eppendorf
	uniTHERMIX 2	LLG Labware
TLC plates	DC Kieselgel 60 F_{254}	Merck
	DC Cellulose F	
transfer membrane	Immobilon-P IPVH00010	Millipore
	Roti®-PVDF 2.0 – 0.2 µm	Roth
ultrapure water	OmniaPure	Stakpure GmbH
ultrasonic bath	Sonorex Super RK 510 H	Bandelin
ultrasonic processor	UP 200S	Dr. Hielscher
UV-hand lamp	HL-6-KM	Bachofer GmbH
vacuum centrifuge	RVC 2-18 CDplus	Christ
vacuum pump	MZ 2C NT	Vacuubrand
	Drehschieberpumpe P4Z	Ilmvac GmbH
voltage controller	E835	Consort
	E143	
	EV2310	
	EV3020	
water bath	SW	Julabo
	Thermomix ME	B. Braun Biotech.
	Thermomix 7P	

6.8 Implemented software

Software	Developer/Distributer
BioEdit	Tom Hall
BLASTp	NCBI
Bruker Data Analysis	Bruker Corporation
ChemDraw Professional 16.0	PerkinElmer Informatics
CLC Sequence Viewer 6	CLC bio
Chromaster System Manager D-7000	Hitachi
DeepLoc	Center for biological sequence analysis, Technical University of Denmark DTU
DoubleDigest Calculator	Thermo Scientific
GraphPasPrism 5	GraphPad
Inkscape 0.92	Inkscape
MEGA X	MEGA, Pennsylvania State University
Microsoft® Office 2007	Microsoft Corporation
Phyre2	Structural Bioinformatics Group, Imperial College, London
Plant-mPLoc	Kuo-Chen Chou and Hong-Bin Shen
PredictProtein Open	Guy Yachdav and Burkhard Rost, Technical University of Munich
Serial Cloner	Serialbasics
SnapGene	SnapGene®

III Results

1 Cinnamic acid 4-hydroxylase (C4H)

1.1 RNA extraction

RNA was extracted from *Anthoceros agrestis* according to Chomczynski and Sacchi (1987). After extraction, integrity and purity of RNA was checked by agarose gel electrophoresis (Fig. 12) and photometric measurement at 260 and 280 nm (Table 2). Integrity of RNA was proven, since 5S rRNA, 18S rRNA and 28S rRNA were still intact. Purity and concentration of RNA were measured by the determination of the A_{260}/A_{280} ratio. A ration of 2.0 is generally accepted as pure RNA. RNA_2, RNA_6 and RNA_8 had the best ratios and were therefore used for preparation of cDNA and RACE-ready cDNA. This cDNA was used for the amplification of all fragments in the following chapters.

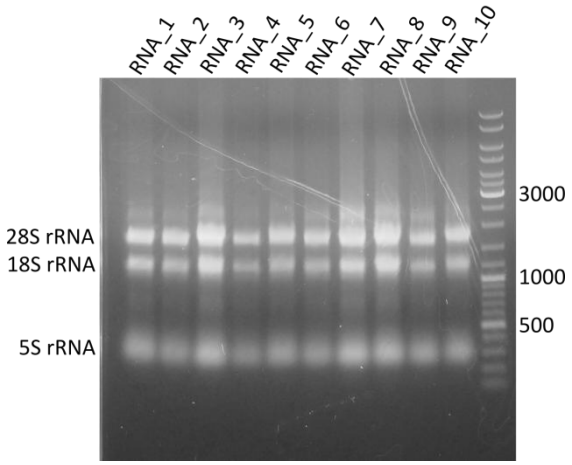


Figure 12 RNA extracted from a 7 day-old *Anthoceros agrestis* suspension culture.

Table 2 Concentration and purity of *Anthoceros agrestis* RNA.

RNA sample	Concentration [ng/µl]	A260/A280	RNA sample	Concentration [ng/µl]	A260/A280
1	1939.0	1.39	**6**	620.0	1.77
2	571.5	1.69	**7**	3100.2	1.59
3	1396.3	1.64	**8**	1449.5	1.78
4	1119.7	1.59	**9**	1267.9	1.56
5	1616.1	1.59	**10**	1357.2	1.29

1.2 Identification of a cDNA encoding C4H from *Anthoceros agrestis*

Based on the scaffold sequence 11181 (Szövenyi, personal communication), PCR primers were designed and used for amplification of a fragment of 657 bp (Fig. 13a). After 3´- and 5´-RACE PCR (Fig. 13b and c) the full open reading frame consisted of 1578 bp encoding an amino acid sequence of 525 amino acid residues. The calculated molecular mass was 59.13 kDa (531 aa/59.95 kDa with 6xHis). The *AaC4H* sequence was deposited in GenBank under the accession number MK778366. The similarities of the AaC4H amino acid sequence to other known sequences (BLASTp) are shown in Table 3.

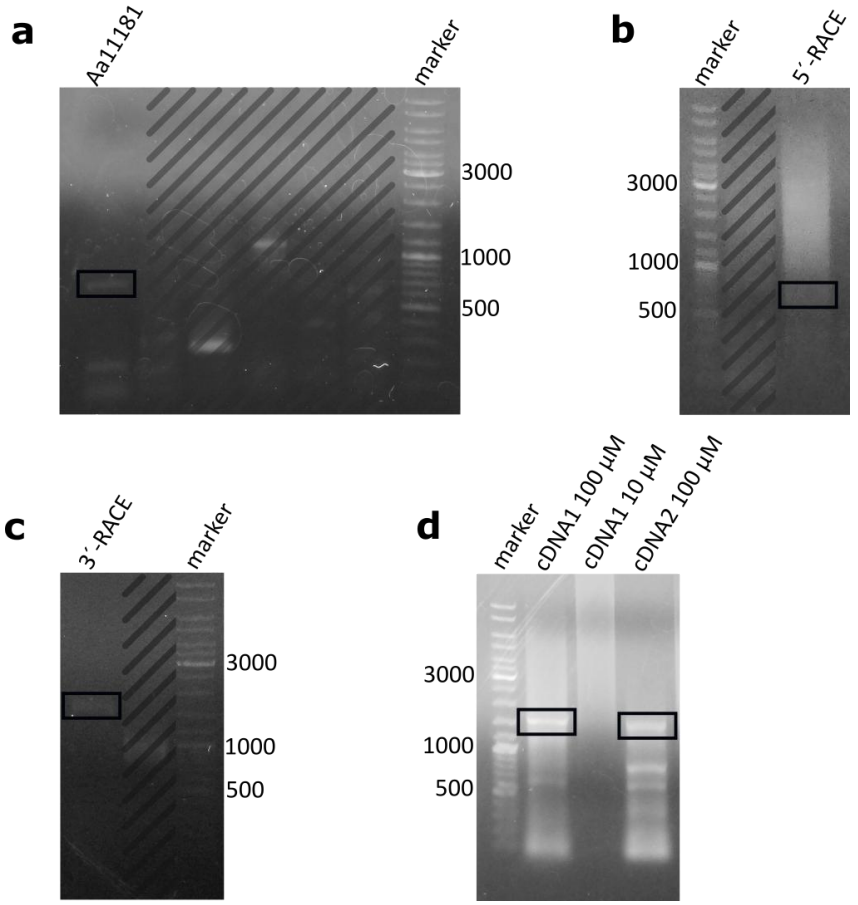


Figure 13 Amplification of *AaC4H* by PCR. a Partial sequence; **b** 5´-RACE; **c** 3´-RACE; **d** AaC4H full-length sequence with different cDNAs as template and two different primer concentrations (100 µM or 10 µM). Hatched areas belong to other experiments.

Table 3 Protein BLAST result of AaC4H. https://blast.ncbi.nlm.nih.gov/Blast.cgi 02.12.2019

Description	Organism	Identity	E value	Accession
***trans*-cinnamate 4-monooxygenase-like**	*Gossypium australe*	75 %	0.0	KAA3481243.1
***trans*-cinnamate 4-monooxygenase**	*Herrania umbratica*	75 %	0.0	XP_021297241.1
PREDICTED: *trans*-cinnamate 4-monooxygenase	*Theobroma cacao*	73 %	0.0	XP_007011365.2
PREDICTED: *trans*-cinnamate 4-monooxygenase-like	*Gossypium raimondii*	74 %	0.0	XP_012454096.1
Cinnamate-4-hydroxylase	*Theobroma cacao*	75 %	0.0	EOY20175.1
***trans*-cinnamate 4-monooxygenase**	*Jatropha curcas*	75 %	0.0	XP_012078176.1

In general, all known sequence motifs of canonical cytochrome P450 enzymes could be detected in the translated amino acid sequence (Fig. 14), such as the proline-rich region (Werck-Reichhart et al. 2002), the PERF motif, the heme-binding cysteine motif and the threonine-containing binding pocket motif (Schuler 1996; Mizutani et al. 1997; Chapple 1998).

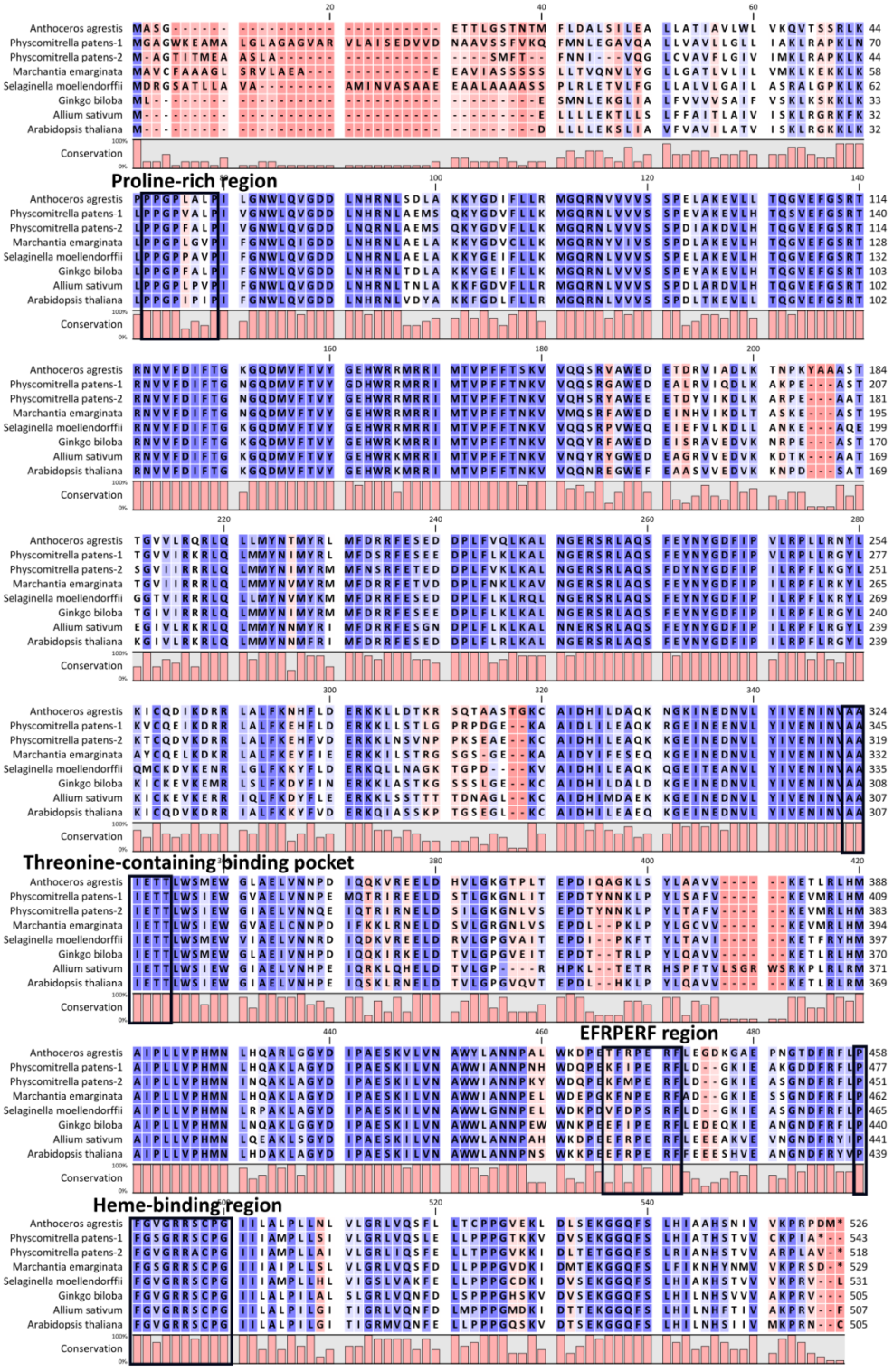


Figure 14 Alignment of C4H amino acid sequences. *Anthoceros agrestis* (MK778366), *Physcomitrella patens*-1 (Pp3c25_10190V3.1), *Physcomitrella patens-2* (Pp3c4_21680V3.1), *Marchantia emarginata* (A0A1Z2R7H6), *Selaginella moellendorffii* (D8S0Y1), *Ginkgo biloba* (AAW70021), *Allium sativum* (ADO24190) and *Arabidopsis thaliana* (P92994). Highly conserved amino acids are marked blue, low conservation is marked red. Typical conserved sequence motifs are marked by boxes (Wohl and Petersen 2020).

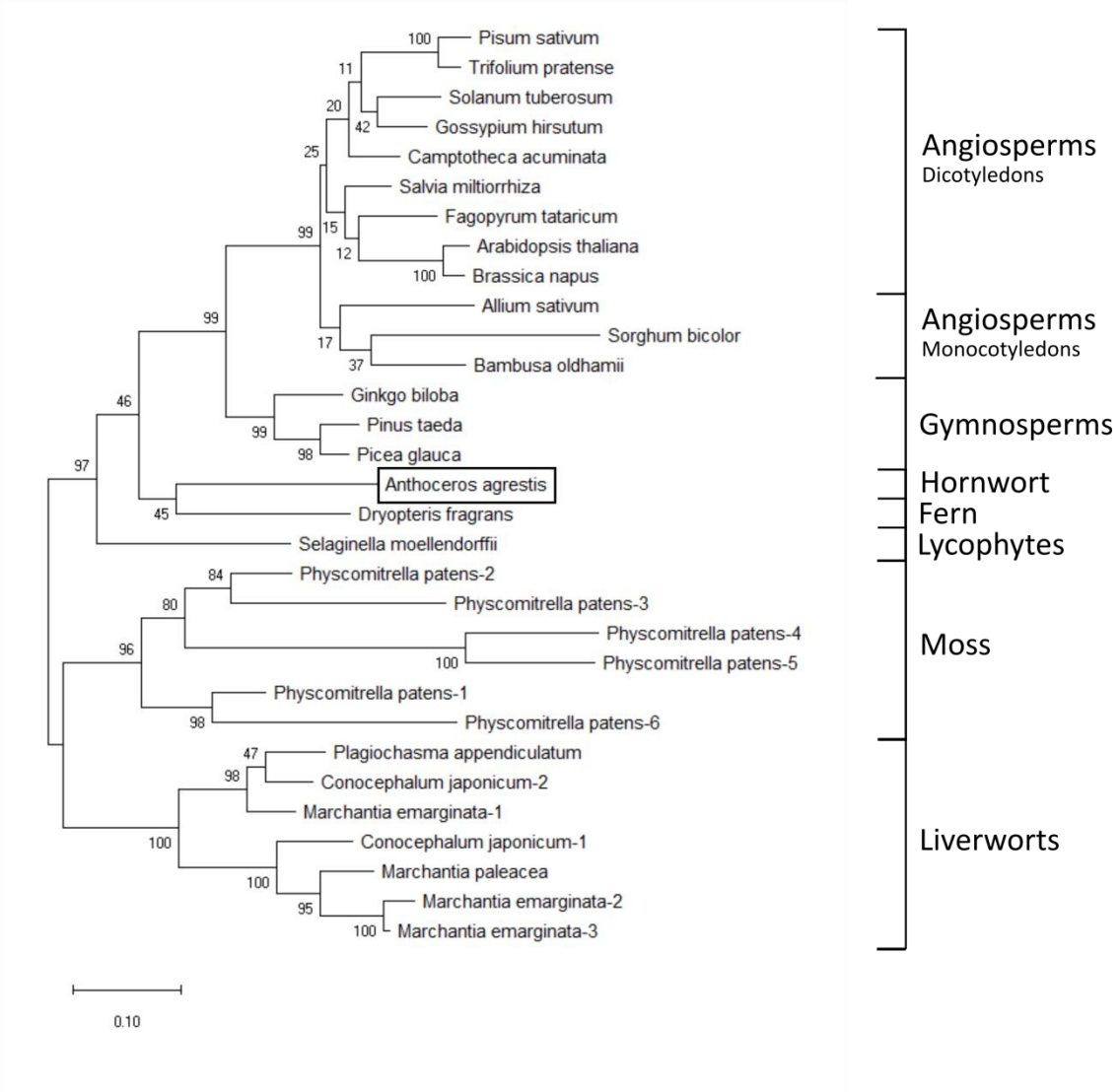


Figure 15 Phylogenetic analysis of C4H amino acid sequences. The maximum likelihood tree was constructed using the MEGA X software. The robustness of the branch structure was evaluated with a bootstrap analysis (1000 replicates). Evolutionary distance is represented with the branch lengths. *Pisum sativum* (Q43067), *Trifolium pratense* (B2LSD9), *Solanum tuberosum* (Q2LAD8), *Gossypium hirsutum* (E4W6M8), *Camptotheca acuminata* (AAT39513), *Salvia miltiorrhiza* (A3FIN3), *Fagopyrum tataricum* (A0A2R2ZJB8), *Arabidopsis thaliana* (P92994), *Brassica napus* (A5GZU5), *Allium sativum* (ADO24190), *Sorghum bicolor* (Q94IP1), *Bambusa oldhamii* (D2K8K9), *Ginkgo biloba* (AAW70021), *Pinus taeda* (AAD23378*), Picea glauca (A0A0G7ZP02), Anthoceros agrestis* (MK778366), *Dryopteris fragrans* (AHI17493), *Selaginella moellendorffii (D8S0Y1), Physcomitrella patens*-2 (Pp3c4_21680V3.1), *Physcomitrella patens*-3 (Pp3c12_6560V3.1), *Physcomitrella patens*-4 (Pp3c13_14870V3.1), *Physcomitrella patens*-5 (Pp3c3_17840V3.1), *Physcomitrella patens*-1 (Pp3c25_10190V3.1), *Physcomitrella patens*-6 (Pp3c16_23740V3.1), *Plagiochasma appendiculatum* (A0A1Z2R7G8), *Conocephalum japonicum*-2 (A0A1Z2R7G6), *Marchantia emarginata*-1 (A0A1Z2R7H6), *Conocephalum japonicum*-1 (A0A1Z2R7H4), *Marchantia paleacea* (A0A1Z2R7G9), *Marchantia emarginata*-2 (A0A1Z2R7H5) and *Marchantia emarginata*-3 (A0A1Z2R7I0) (Wohl and Petersen 2020).

The amino acid sequence was analyzed phylogenetically with the maximum likelihood algorithm of the MEGA X software package (Fig. 15). The tree indicates two branches, the first with liverworts and mosses and the second with lycophytes, ferns, hornworts, gymnosperms and angiosperms. According to this analysis, AaC4H has higher similarity to C4Hs from ferns and lycophytes, than to the corresponding enzymes from mosses and liverworts. It confirms the hypothesis of Qiu et al. (2006), Ligrone et al. (2012a) and Ruhfel et al. (2014) that hornworts are the youngest member of the bryophytes.

1.3 Expression of AaC4H in *Saccharomyces cerevisiae*

Saccharomyces cerevisiae was initially selected, since it is the most commonly used organism for the heterologous expression of cytochrome P450 enzymes. Therefore primers were designed, which equipped the full-length sequence of *AaC4H* with the restriction sites for EcoRI (5´) and NotI (3´). Furthermore the stop codon of *AaC4H* was deleted to ensure the attachment of a FLAG-tag in the *C*-terminus, already present on the expression vector. The sequence was amplified by PCR (Fig. 13d). The amplicon was purified, ligated into pDrive and the sequence verified after multiplication in *E. coli* EZ. After digestion with EcoRI and NotI the sequence was inserted into the multiple cloning site I (MCSI) of the $pESC_{-ura}$ vector (Agilent).

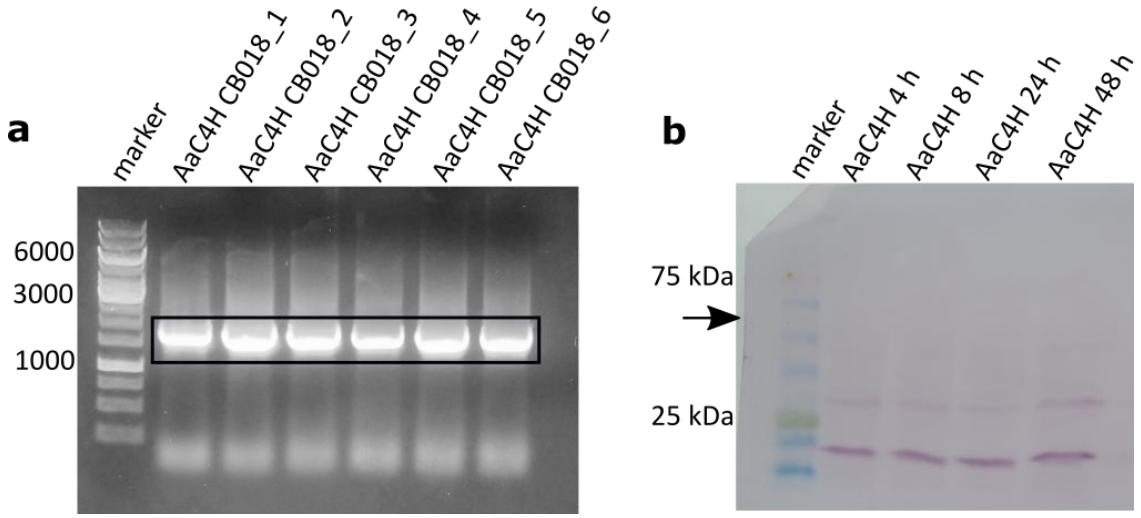


Figure 16 Expression of AaC4H in *S. cerevisiae* strain CB018. a Colony-PCR of 6 different transformants; **b** Western blot analysis with anti-FLAG-antibody. Samples were taken after 4, 8, 24 and 48 h of incubation in SCG_{-ura}-medium. Collected samples were adjusted to an OD_{600} of 4.

The *S. cerevisiae* strains BY4741, BY4742 and CB018 were transformed and the uptake of the plasmid was checked by colony PCR (Fig. 16a). Protein expression was performed as described in chapter II.3.1.2. The formation of the desired protein, however, could not be detected in any of the used yeast strains. Western blot analysis with anti-FLAG antibody showed no gene-specific bands in the expected size range (around 60 kDa; Fig. 16b for strain CB018 as an example). Likewise, enzyme activity tests with microsomal fractions were negative (Fig. 17). The expected enzyme product 4-coumaric acid had a retention time of approximately 3.7 min and could neither be detected in tests with microsomes from the transformed, nor from the untransformed cultures after 45 min incubation. After adding 4-coumaric acid to an AaC4H-assay, an additional peak appeared.

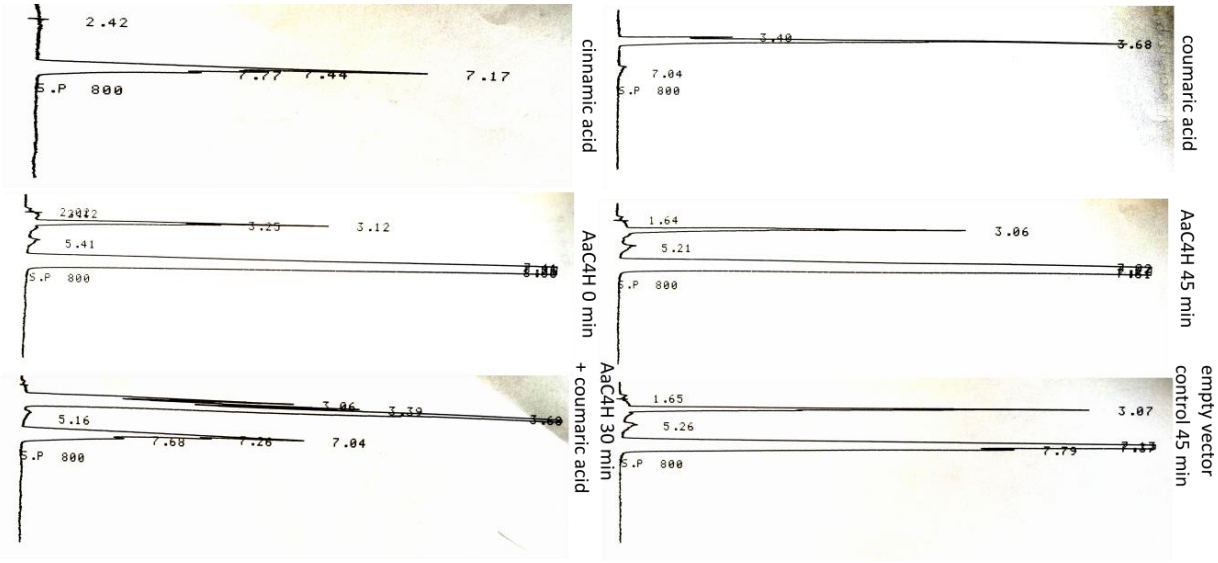


Figure 17 Reaction of *S. cerevisiae* BY4742 microsomes with *t*-cinnamic acid after 0 and 45 min. Assays were analysed by HPLC at 309 nm (isocratic elution with 50 % methanol/0.01 % H_3PO_4). Assays with microsomes from *S. cerevisiae* BY4742 transformed with the empty vector $pESC_{-ura}$ served as negative control. The cinnamic acid standard was detected after 7.2 min, the 4-coumaric acid standard after 3.7 min.

1.4 Expression of AaC4H in *Escherichia coli*

For the heterologous expression in *E. coli*, *AaC4H* and *AaCPR* were fused according to the constructs described in chapter II.2.10. All experiments described in this chapter were carried out with the constructs including the ribosomal binding site (rbs) of

pET15b, since constructs without showed no C4H activity. Results for the constructs without the rbs will be illustrated in chapter III.2.4. The fus-constructs were inserted into the expression vector pET15b and the *E. coli* strains BL21RIPL, C41(DE3) and SoluBL21 were transformed with these plasmids. After checking plasmid uptake, the proteins were expressed as described in chapter II.3.1.1. In this section only the activity of AaC4H is discussed, since activity of AaCPR will be described in chapter III.2.4.

Neither the expression in BL21RIPL nor in SoluBL21 resulted in any detectable fusion protein at around 130 kDa. In addition, the activity tests in these strains were negative. Expression in *E. coli* C41(DE3) resulted in faint bands with the correct size in fus1 C41(DE3)_1, fus2 C41(DE3)_2 and in both fus3 C41(DE3) transformants (Fig. 18).

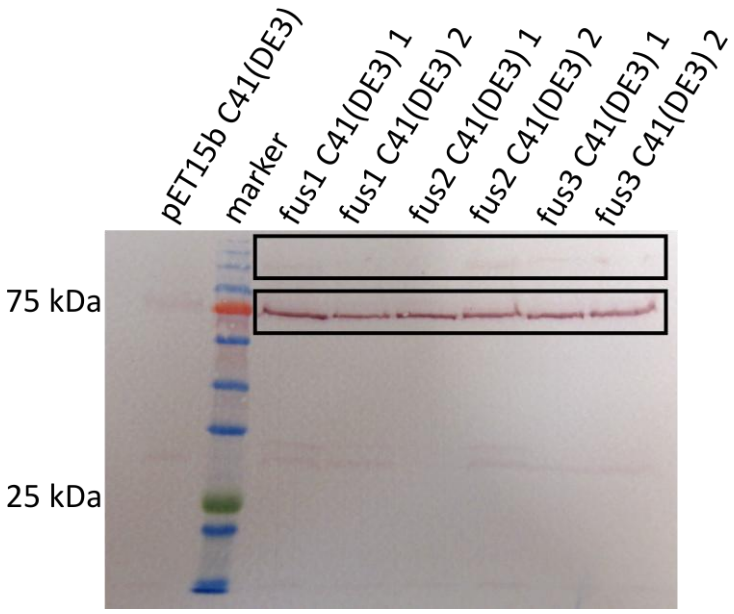


Figure 18 Western blot analysis of fusion AaC4H-AaCPR constructs expressed in *E. coli* C41(DE3) with anti-His-antibody. Fusion constructs expressed in C41(DE3) for 20 h at 25 °C after induction with 1 mM IPTG. Crude extract (4 mg/ml) was prepared from two transformants for each fusion construct. *E. coli* strains transformed with the empty vector pET15b served as a negative control. The upper box displays the detected fusionproteins and the lower bix displays the detected soluble AaCPR.

While the empty vector control and fus2 showed no C4H activity, 4-coumaric acid formation was catalyzed by fus3 and a small peak was also detected with the microsomal fraction of fus1. The highest 4-coumaric acid peaks could be measured in enzyme assays with microsomal fractions (Fig. 20). Western blot analysis of these

protein fractions did not show any fragments of the correct size, presumably because the concentration of fusion proteins was too low for detection (Fig. 19). The overall activity of the fusion proteins, however, was comparatively low and optimization would have been necessary in order to achieve higher expression levels and hydroxylation activities.

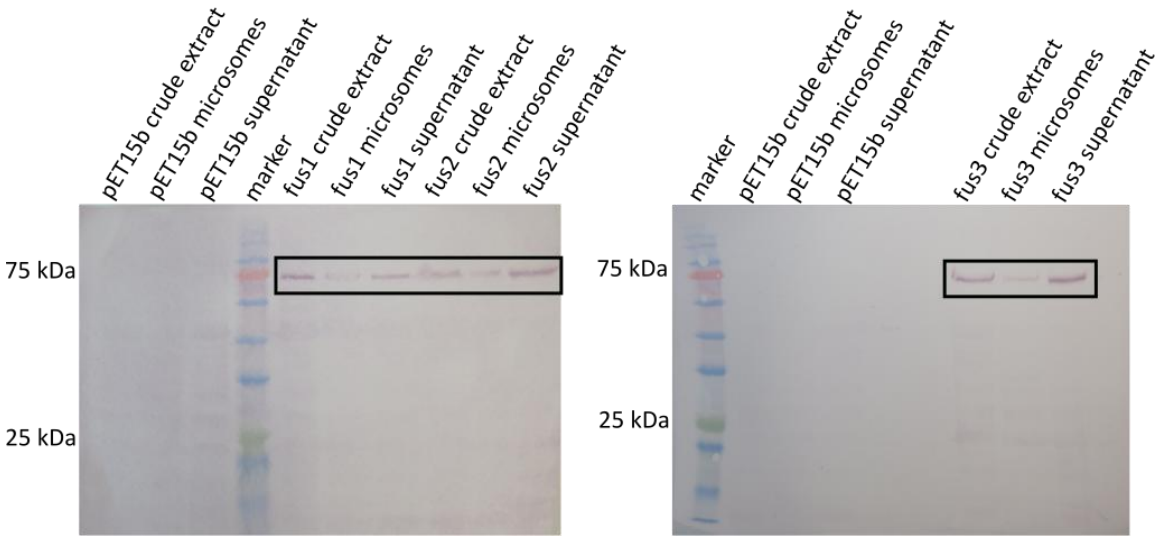


Figure 19 Western blot analysis of crude protein extract, microsomal fraction and supernatant of the empty vector control and the fusion constructs expressed in C41(DE3) with anti-His-antibody. The crude extract had a protein concentration of 3 mg/ml, 4.75 ml were used for preparation of the membrane fraction. Microsomes were resuspended in 550 µl. Boxes display the detection of the soluble AaCPR.

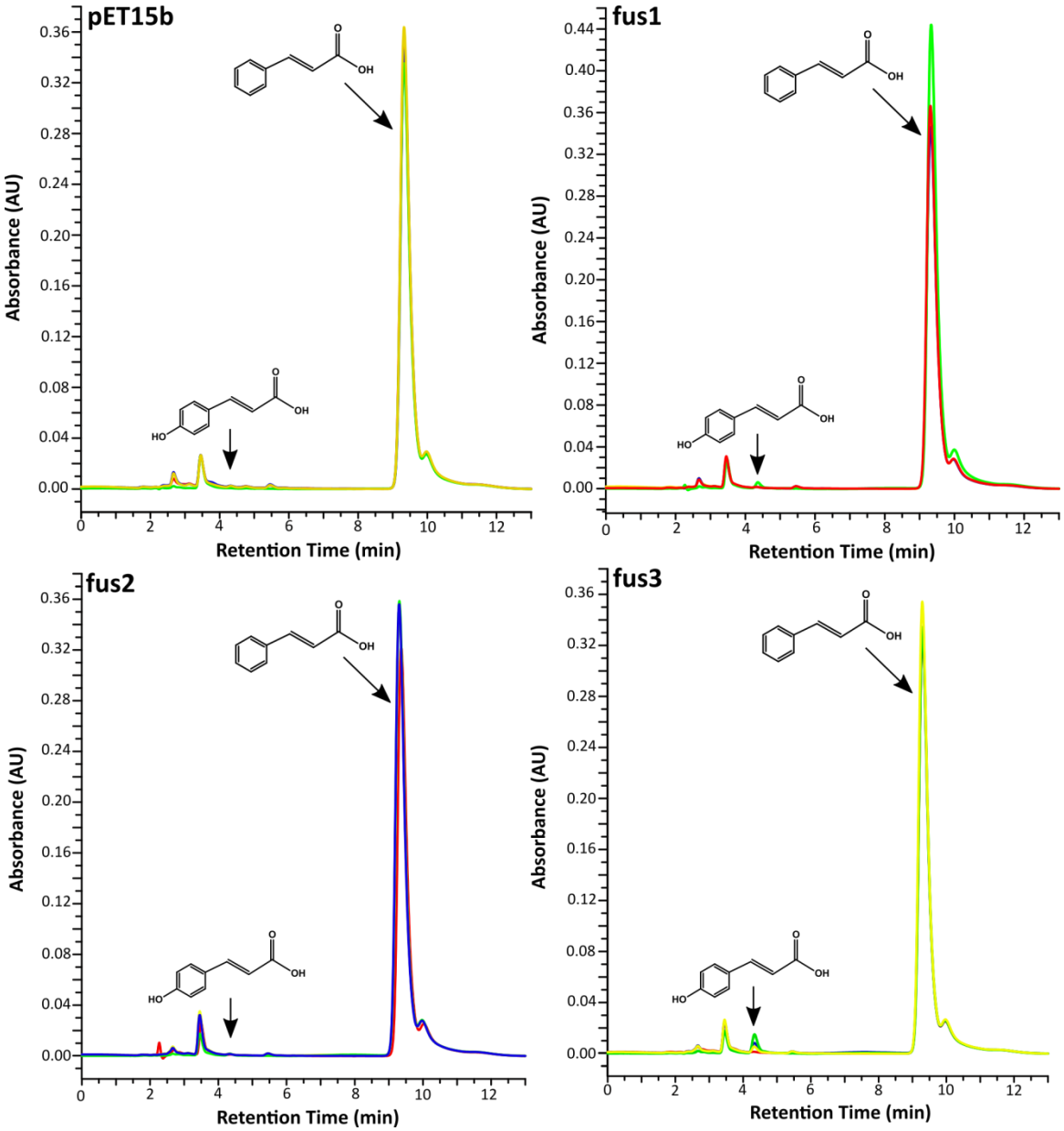


Figure 20 Reaction with *t*-cinnamic acid after 0 and 60 min with crude extracts, microsomes and supernatants isolated from fusion constructs and empty vector control expressed in C41(DE3). Assays with crude protein extract are marked red (0 min) and blue (60 min). The assay with microsome preparations after 60 min is marked green and the assays with supernatant fraction (60 min) is marked yellow.

1.5 Expression of AaC4H in *Physcomitrella patens*

In a next step, protein production in a plant expression system was tested. *Physcomitrella patens* was chosen, since it is closely related to *Anthoceros agrestis* and it is already established as a heterologous expression system. At first, the full-length sequence with *C*-terminal 6xHis codons was inserted into the expression vector pTHUbiGate as described in chapter II.2.8.2. Then *P. patens* protoplasts were transformed and stable transformants were selected (Chapter II.2.19). After selection three stable transformants were obtained but only one of them (PpAaC4H_1) showed a band of the correct size (~ 60 kDa) after Western Blotting. Figure 21 shows that AaC4H is localized in the microsomal fraction. There was no corresponding protein band in the control wild type culture. Since *Physcomitrella* presumably has own cinnamic acid 4-hydroxylases (e.g. the two most highly expressed putative C4Hs according to Phytozome: PpC4H_1: Pp3c25_10190V3.1 and PpC4H_2: Pp3c4_21680V3.1), characterization was carried out in parallel with the untransformed *Physcomitrella* wild type (PpWT).

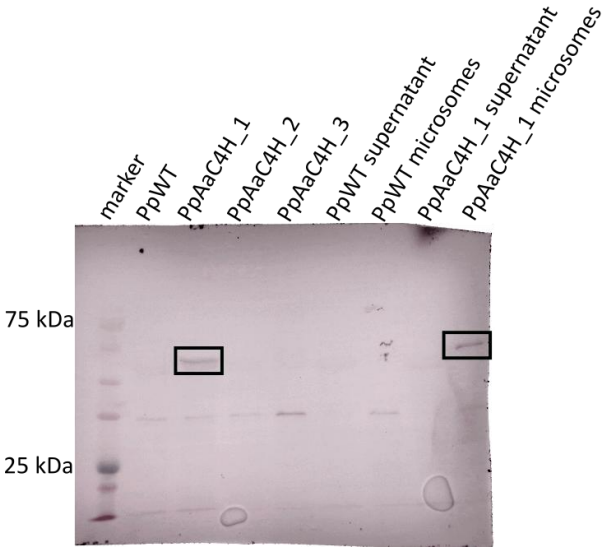


Figure 21 Western blot analysis of AaC4H expressed in *P. patens* with anti-His-antibody. 133 µg were applied from crude protein extract of the three different transformants and PpWT, 90 µg of the supernatant fractions and 30 µg of the microsomal fractions. *P. patens* wildtype served as negative control (Wohl and Petersen 2020).

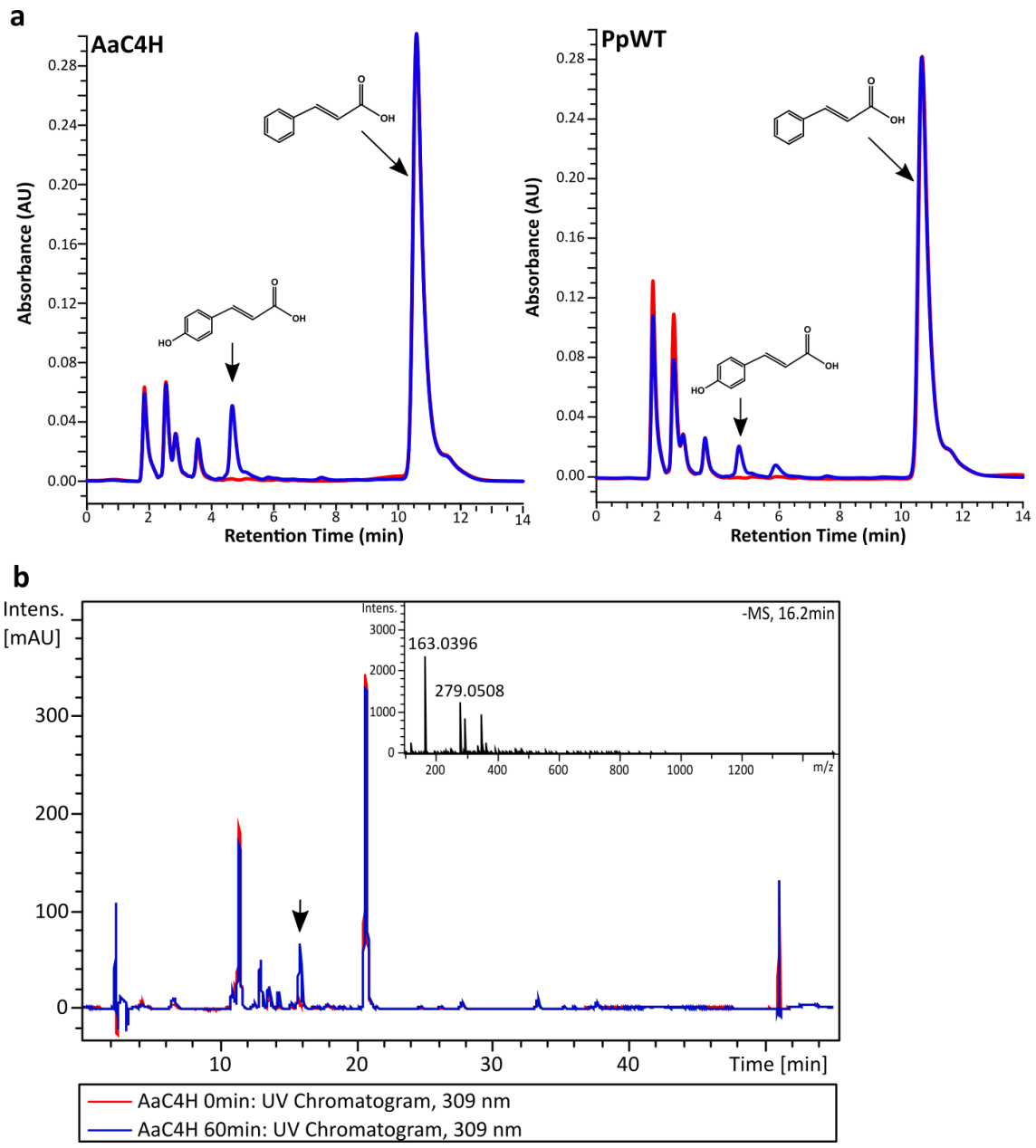


Figure 22 Enzyme assays with *t*-cinnamic acid and crude protein extracts (2 mg/ml) from AaC4H expressed in *P. patens* compared to the *P. patens* wild type. a Reaction with *t*-cinnamic acid after 0 (red) and 30 min (blue) measured by HPLC at 309 nm; **b** UV-chromatogram at 309 nm of LC-MS analysis from Pp_AaC4H assays after 0 (red) and 60 min (blue). The insert shows the mass of product peak in negative mode [M-H]. The exact mass of 4-coumaric acid is m/z 163.04 [M-H].

Analysis of C4H activity assays of the transformed and untransformed cultures revealed at least double to triple the amount of 4-coumaric acid in assays conducted with protein extracts from the transformed *AaC4H* culture (Fig. 22a). The formation of 4-coumaric acid was also confirmed by LC-MS analysis (Fig. 22b). Expression for 8, 12 and 15 days revealed that the protein was not detectable after 8 days, but that there were faint signals after 12 and 15 days which did not differ in intensity. Enzyme assays using crude extract with a protein concentration of 2 mg/ml also displayed the highest activity after expression for 12 d. For all further experiments the expression was carried out for 12 days. Moreover the storage stability of the protein was checked at 4 °C and -80 °C with 20 % glycerol. After storage at -80 °C the activity was still the same after 24 h and 1 week. At 4 °C the activity strongly decreased after 24 h and was completely lost after one week.

Using NADH as a cosubstrate resulted in a conversion rate of about 20 % compared to the presumed natural cosubstrate NADPH (Fig. 23). Addition of FAD and FMN, the prosthetic groups of the NADPH:cytochrome P450 reductase (CPR), had no effect on product formation. Enzyme assays with benzoic acid as alternative putative substrate showed no conversion to 4-hydroxybenzoic acid. Additionally, 4-coumaric acid and 3-coumaric acid were tested as potential substrates. The conversion rate was determined with the help of a caffeic acid standard since caffeic acid would be expected as reaction product. The area under the curve was the same for PpWT and Pp_AaC4H with approximately 0.06 nmol/assay after 60 min of incubation with 3-coumaric acid. Therefore it was assumed, that this conversion was catalyzed by a *Physcomitrella* enzyme and not by the additional AaC4H. Only approximately 2.5 % caffeic acid was produced in comparison to 4-coumaric acid from cinnamic acid as substrate. Production of caffeic acid from 4-coumaric acid was only detected in a very low amount (approximately 0.7 % compared to the substrate cinnamic acid) using the wild type crude protein extract. This conversion could not be detected in the transformed culture.

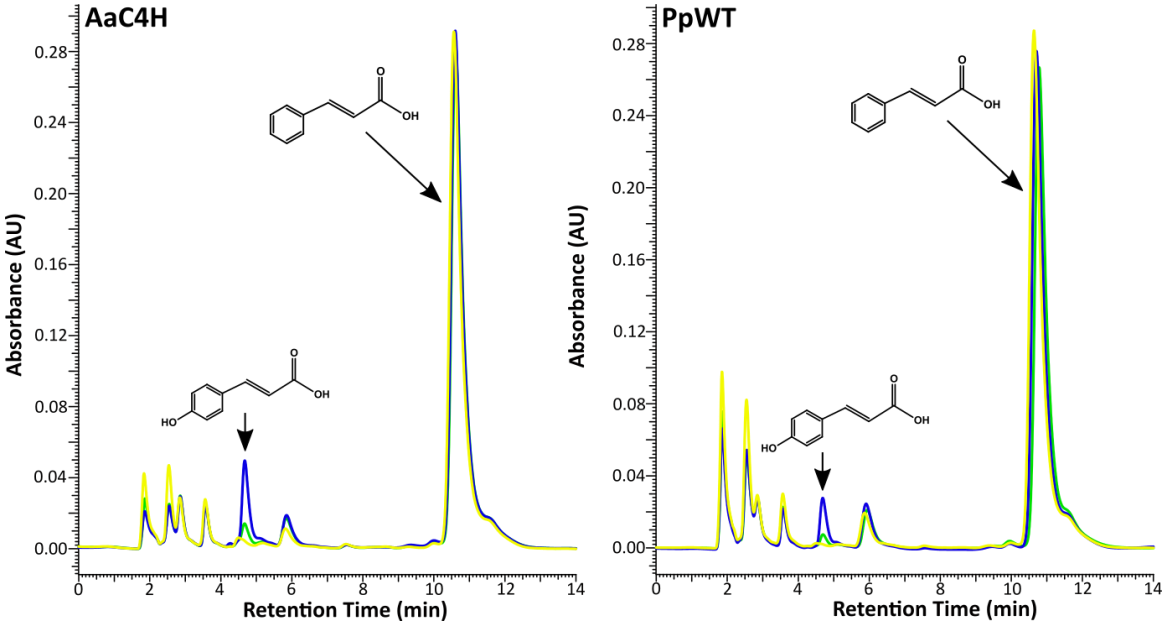


Figure 23 Cosubstrate-specificity of Pp_AaC4H and PpWT. Enyzme assays with crude protein extract (2 mg/ml) and *t*-cinnamic acid after 30 min using 500 µM NADPH (blue), NADH (green) or no electron donor (yellow).

1.5.1 Reduction of enzyme activity after preparation of microsomes

All attempts to isolate active membrane fractions resulted in the loss of enzyme activity although different methods were used. First attempts for isolations were performed with either $MgCl_2$ or polyethylene glycol with NaCl to aggregate the microsomes followed by a centrifugation for 20 min at 60000 g (Pompon et al. 1996; Petersen 2003). Furthermore, microsomes were prepared by ultracentrifugation of the crude protein extract for 1 h at 100000 g (Benveniste et al. 1986; Abas and Luschnig 2010). At last membrane fraction was isolated at 21000 g according to the low speed centrifugation protocol of Abas and Luschnig (2010). When microsomes were resuspended in the volume of the protein extract used for membrane precipitation, the activity was only about 25% compared to the crude protein extract (Fig. 24a). This effect could be observed for all methods used.

To restore enzyme activity different experiments were performed. First, AaC4H and AaCPR were expressed in parallel in different *P. patens* transformants and crude extract and microsomes were isolated. Assays with different protein solutions were mixed in the same ratio and incubated for 30 min (Table 4). Nevertheless the activity could not be restored.

Table 4 Assay combinations of protein solutions from Pp_AaC4H and Pp_AaCPR. Microsomes of Pp_AaC4H were mixed with the second protein extract in the same ratio (50 µl each).

Protein extract	In combination with
Pp_AaC4H microsomes	Pp_AaC4H crude protein extract
	Pp_AaC4H supernatant
	Pp_CPR microsomes
	Pp_CPR supernatant

Centrifugation with varied centrifugal forces (5000 g, 10000 g, 15000 g, 25000 g, 40000 g and 60000 g) after addition of 50 mM $MgCl_2$, resulted in a decreasing activity in the supernatant with higher centrifugation speeds while the activity of the microsomal fraction increased slightly after 40000 g. It was noticeable that the microsomal suspension became greener with higher centrifugation speeds, which was attributed to the increasing precipitation of thylakoid membranes (Fig. 24b).

Resuspension of the microsomes with Potter-Elvehjem homogeniser, by ultrasonication or by repeated pipetting led to no difference in activity. Purification of the crude extract through a PD10-column led to loss of activity.

Since the activity of the microsomal fraction could not be restored, all further experiments were carried out with crude protein extract.

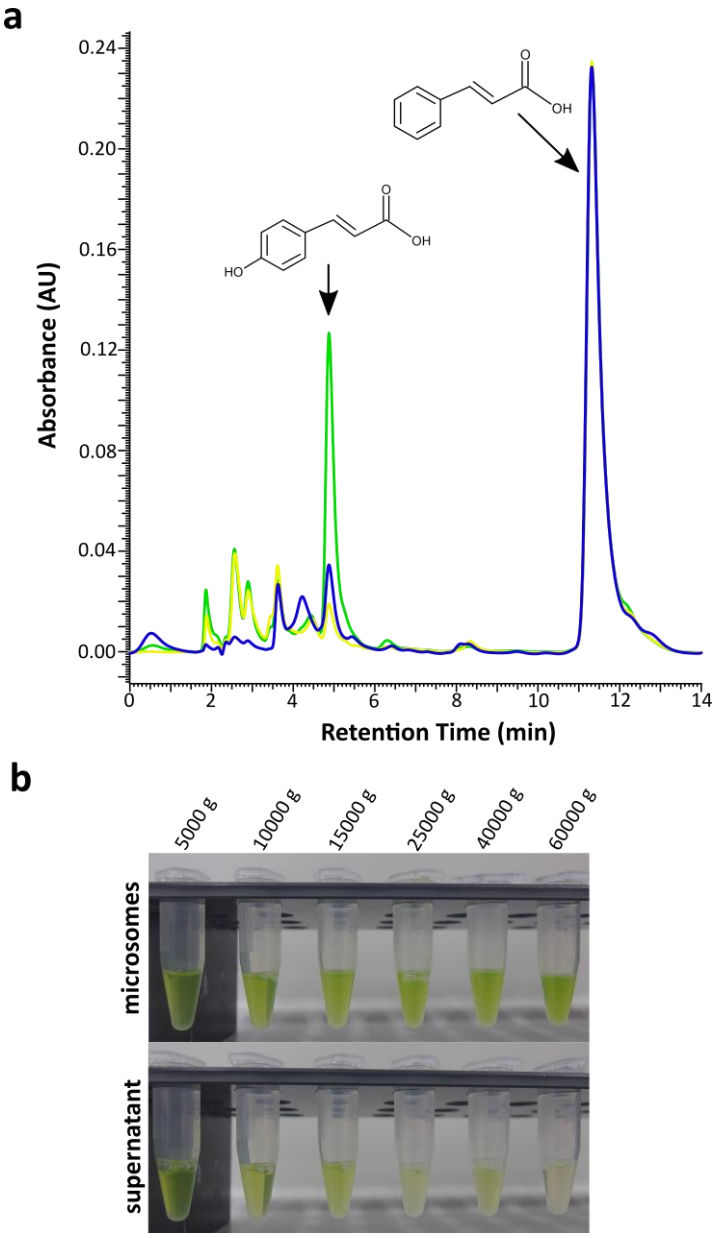


Figure 24 Extraction of microsomal fractions from *P. patens*. **a** Assays with crude protein extract (green), microsomes (blue) and supernatant (yellow) from AaC4H expressed in *P. patens* after 30 min. The crude extract had a protein concentration of 2 mg/ml. 1 ml was used for centrifugation and microsomes were resuspended in the same volume; **b** Extraction of microsomes with different centrifugation speeds after addition of 50 mM $MgCl_2$.

1.5.2 Temperature and pH-optimum of AaC4H

C4H assays using protein extract from the transformed Pp_AaC4H culture as well as PpWT showed a temperature optimum around 25 °C. The pH-optimum was around pH 7 for both (Fig. 25).

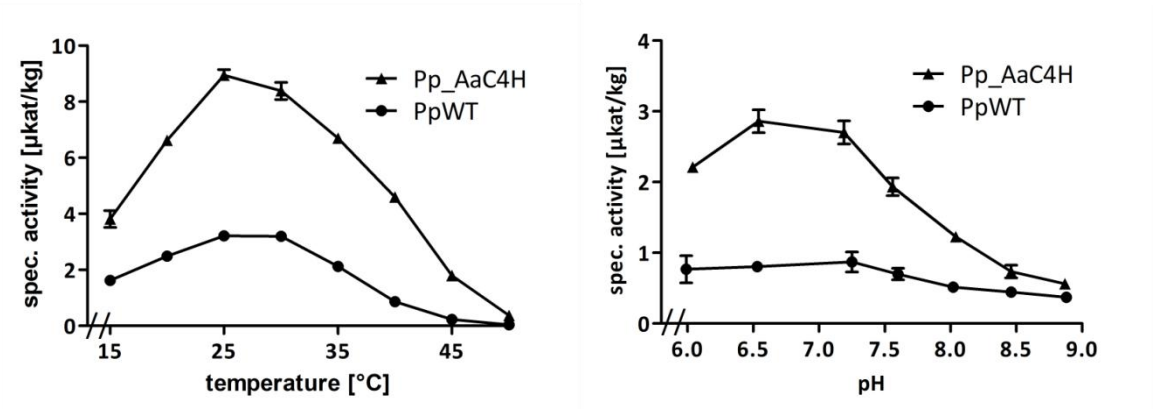


Figure 25 Temperature (left) and pH-optimum (right) of C4H assays performed with *P. patens*_AaC4H and *P. patens*_WT (Wohl and Petersen 2020)**.** Error bars represent the standard deviation (n=3).

1.5.3 AaC4H: reaction kinetics

For the determination of kinetic data for AaC4H (heterologously expressed in *P. patens*) and the C4Hs of *P. patens* protein was isolated at least four times to ensure biological variation. For all substrate concentrations assays were repeated three times to account for technical variation. The resulting substrate saturation curves of cinnamic acid led to an apparent K_m-value of 17.3 ± 2.5 µM for Pp_AaC4H and 25.1 ± 4.0 µM for PpWT (Fig. 26a; Table 5). The K_m-values for NADPH were approximately the same for the AaC4H transformant (88.0 ± 9.8 µM) and the PpWT C4H reaction (92.3 ± 5.4 µM) (Fig. 26b; Table 5), since both reactions are supplied by the same PpCPRs.

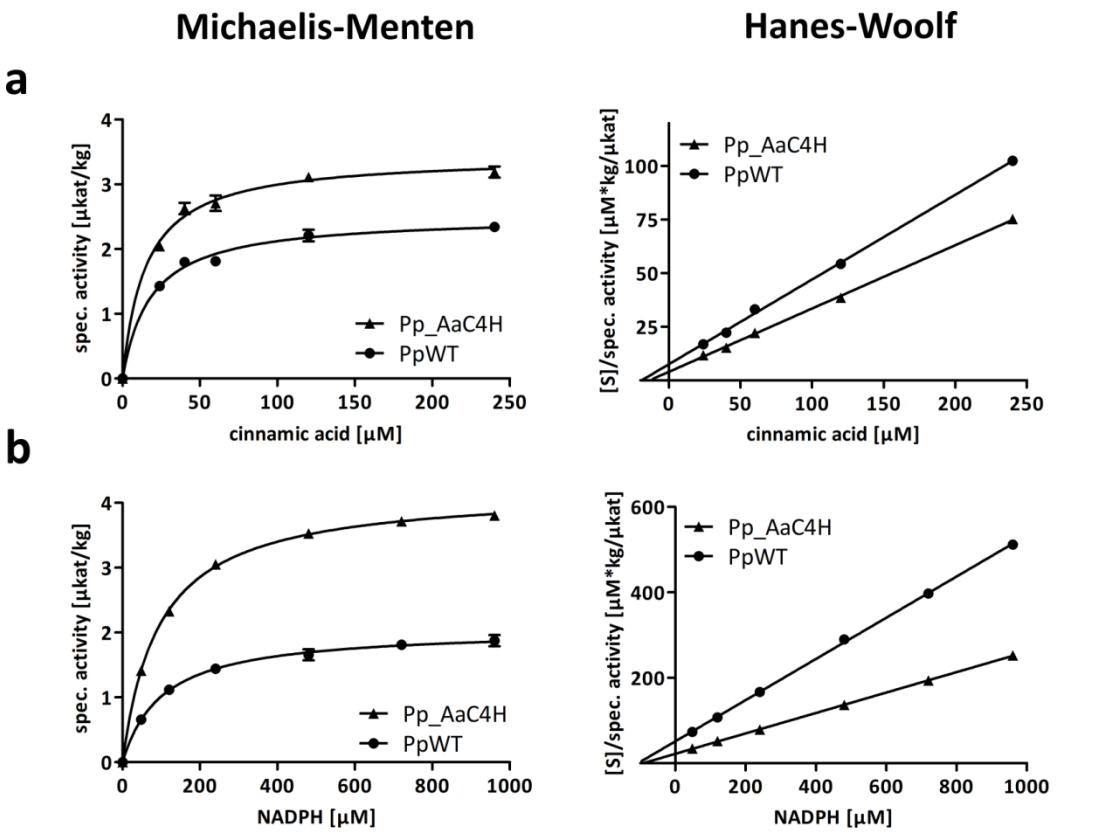


Figure 26 Dependence of C4H on cinnamic acid and NADPH (Wohl and Petersen 2020). Graphs represent the mean values of one biological replicate with three technical replicates. Michaelis-Menten diagrams are displayed on the left and Hanes-Woolf diagrams on the right. **a** *t*-cinnamic acid; **b** NADPH. Error bars represent the standard deviation.

Table 5 Kinetic values for Pp_AaC4H and PpWT. K_m-values were calculated with help of Michaelis-Menten plots (MM), Lineweaver-Burk plots (LB) and Hanes-Woolf plots (HW) (mean ± SE). Mean values represent at least four biological replicates with three technical replicates.

		Pp_AaC4H K_m [µM]	PpWT K_m [µM]
cinnamic acid	MM	17.3 ± 2.5	25.1 ± 4.0
	LB	16.8 ± 2.2	23.1 ± 4.1
	HW	15.3 ± 3.2	26.1 ± 2.4
NADPH	MM	88.0 ± 9.8	92.3 ± 5.4
	LB	85.4 ± 9.0	92.4 ± 6.9
	HW	82.9 ± 8.7	90.6 ± 7.4

1.5.3 Expression of AaC4H in *P. patens* during a cultivation period

The aim of the characterization of the transformed *P. patens* suspension culture was to determine the relation between *AaC4H* and the two most highly expressed *PpC4H* isoforms during a three week cultivation period.

1.5.3.1 Culture characterization during a cultivation period

During the culture characterization the fresh weight of the cells and the pH of the culture medium were determined every two days over a period of 21 days. The course of the fresh weight showed that the cells double their weight almost every two days (Fig. 27a). Even after 21 days, no weight reduction could be detected, which would indicate cell death. There was no significant difference in growth between the transformed and the wild-type culture.

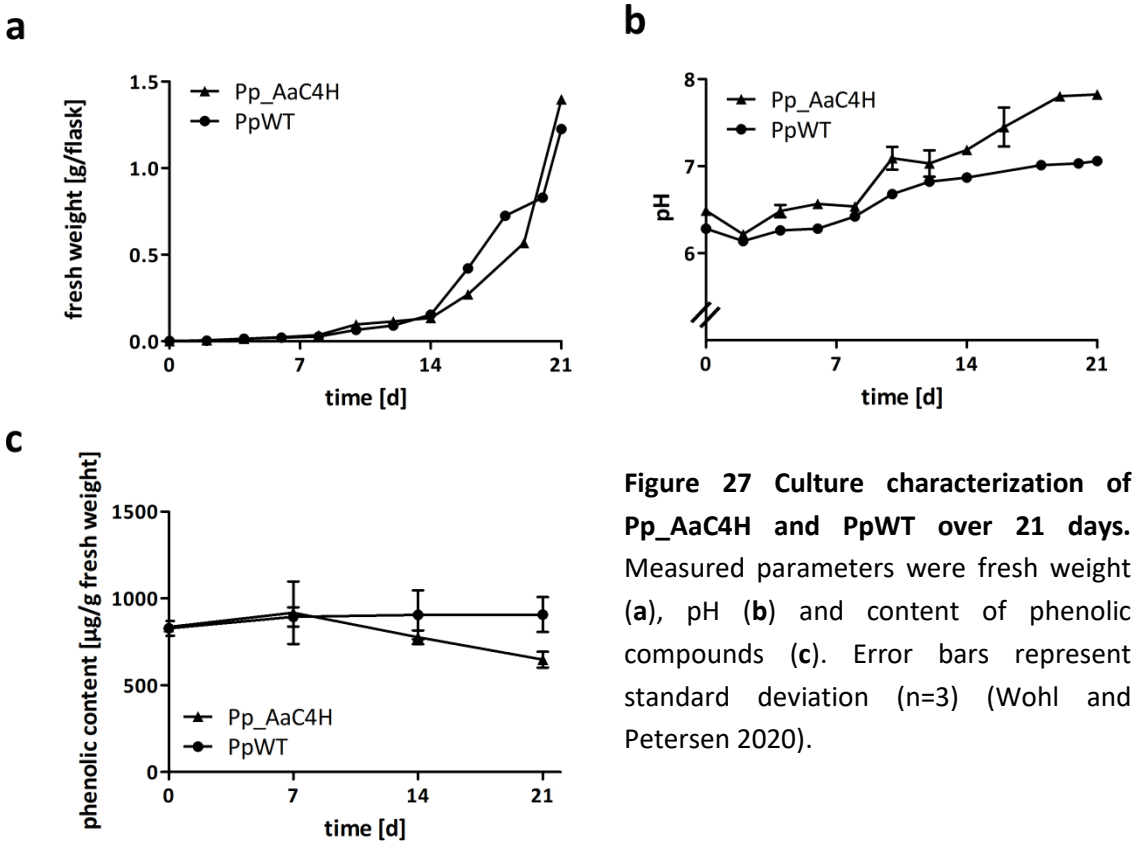


Figure 27 Culture characterization of Pp_AaC4H and PpWT over 21 days. Measured parameters were fresh weight (**a**), pH (**b**) and content of phenolic compounds (**c**). Error bars represent standard deviation (n=3) (Wohl and Petersen 2020).

The pH-value of the medium was around pH 6.5 at the beginning of the characterization. After two days the pH slightly dropped, before it started to rise again over the remaining 19 days. After three weeks the pH remained around 7.8 (Fig. 27b).

Determination of the phenolic content after 0, 7, 14 and 21 days showed, that the amount barely changed and was always around 650-950 µg/g fresh weight for both *P. patens*_AaC4H and *P. patens*_WT (Fig. 27c).

1.5.3.2 Expression analysis

To determine the relative amount of mRNA of AaC4H and the two most highly expressed C4Hs from *P. patens* (PpC4H_1: Pp3c25_10190V3.1 and PpC4H_2: Pp3c4_21680V3.1; Phytozome), RT-qPCR experiments were performed with mRNA from the transformed AaC4H culture. RNA was isolated of samples taken during the culture characterization (Fig. 28; Table 6). After digestion of 5 µg RNA with DNase (Table 7), the mRNA was reverse-transcribed into cDNA. To account for biological variation, RNA was isolated twice and both samples were digested with DNase. Moreover, cDNA was prepared twice for the RNA with a better A_{260}/A_{280} ratio and once for the other one.

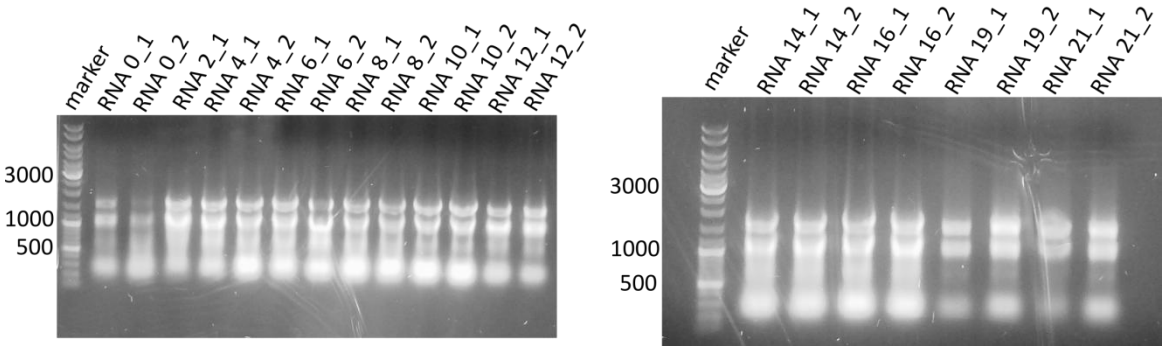


Figure 28 RNA extraction of transformed *P. patens* with AaC4H tissue samples for RT-qPCR experiments.

Table 6 Concentration and purity of isolated *P. patens* RNA

RNA sample	Concentration [ng/µl]	A_{260}/A_{280}	RNA sample	Concentration [ng/µl]	A_{260}/A_{280}
0_1	254.8	1.47	**12_1**	1997.8	1.56
0_2	2578.2	1.23	**12_2**	626.2	1.89
2	1375.4	1.81	**14_1**	1739.5	1.86
4_1	1877.4	1.90	**14_2**	1261.9	1.83
4_2	3794.9	1.81	**16_1**	1443.4	1.94

6_1	2469.9	1.84	**16_2**	2244.2	1.52
6_2	2606.0	1.91	**19_1**	331.0	1.69
8_1	3470.7	1.79	**19_2**	352.9	1.61
8_2	4167.3	1.48	**21_1**	978.0	1.25
10_1	4450.2	1.81	**21_2**	346.0	1.70
10_2	3621.4	1.99			

Table 7 Concentration and purity of isolated *P. patens* RNA after DNase digestion. RNA 1 was obtained by digestion of the RNA with the better A_{260}/A_{280} ratio, RNA 2 originates from the RNA with the poorer ratio. For both digestions 5 µg RNA were used.

DNase digested RNA 1			DNase digested RNA 2		
RNA sample	**Concentration [ng/µl]**	**A_{260}/A_{280}**	**RNA sample**	**Concentration [ng/µl]**	**A_{260}/A_{280}**
0	256.3	1.36	**0**	49.8	1.51
2	180.5	1.76	**2**	164.9	1.61
4	179.8	1.66	**4**	231.8	1.56
6	193.9	1.69	**6**	246.4	1.56
8	226.0	1.56	**8**	251.6	1.55
10	171.8	1.78	**10**	341.8	1.59
12	60.3	1.42	**12**	394.6	1.21
14	173.3	1.53	**14**	1451.8	1.05
16	167.4	1.86	**16**	170.4	1.56
19	166.7	1.73	**19**	2009.2	1.04
21	207.0	1.76	**21**	73.3	1.37

Before performing the RT-qPCR experiments, all primer pairs were used in normal PCR experiments in order to ensure the absence of unspecific PCR products. Moreover,

amplified fragments were sequenced to really ensure amplification of the very similar *C4H*s. *Serine threonine protein phosphatase 2a regulatory subunit [ST-P 2a]* (involved in regulation of signalling processes) was used as a reference gene. Other tested reference genes (*actin 1, actin 5, ubiquitin-conjugating enzyme E2 and elongation factor 1α*) showed either unspecific PCR products on an agarose gel or in the melting curves. Primers for *actin 5*, *ubiquitin-conjugating enzyme E2*, *elongation factor 1α* and *ST-P 2a* were ordered based on Le Bail et al. (2013).

The determination of the primer efficiency (E) showed that the PCR reaction of *ST-P 2a* and *AaC4H* required a doubled primer concentration (385 µM) compared to *PpC4H_1* and *PpC4H_2* (Fig. 29). Melting curves are shown in Appendix VIII.1.

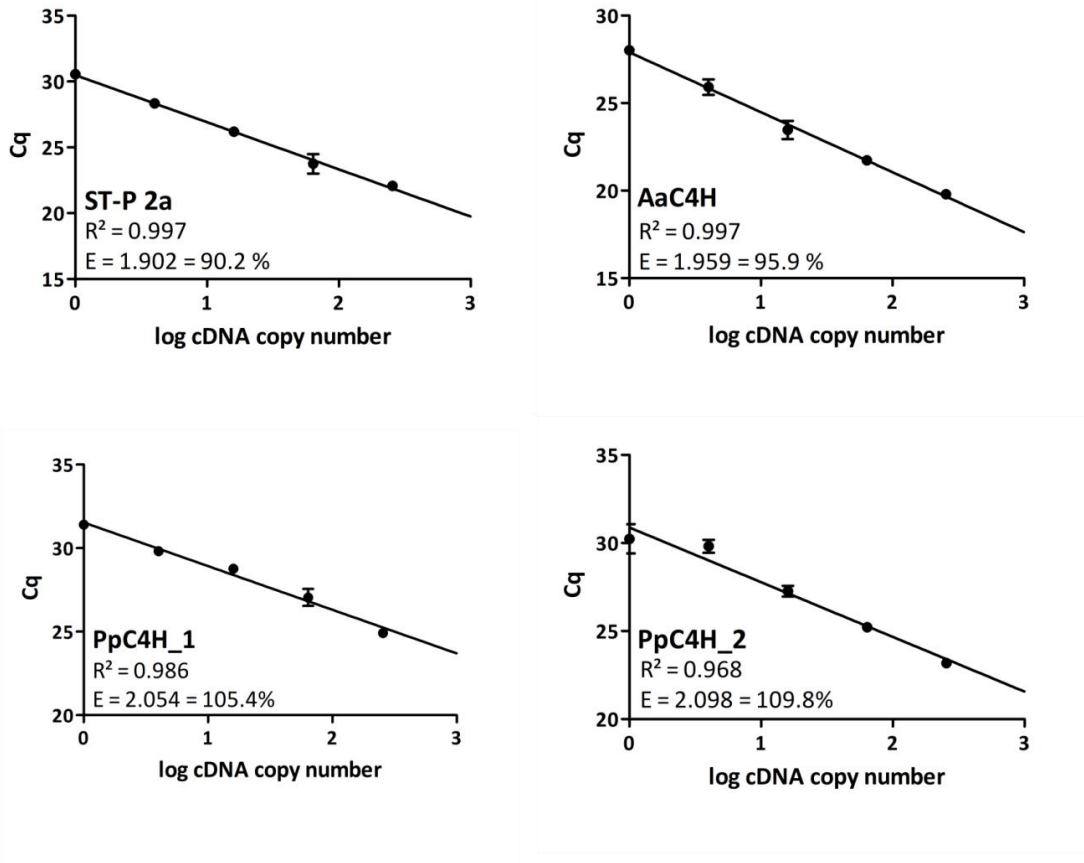


Figure 29 Primer efficiency (E) of *St-P 2a*, *AaC4H*, *PpC4H1* and *PpC4H2*. cDNA of all samples was mixed and serially diluted to concentrations ranging from non-diluted to 1:256 as a quantification standard to test amplification efficiency. Primers were added in concentrations of 385 µM for *ST-P 2a* and *AaC4H* and 192.5 µM for *PpC4H1* and *PpC4H2*. Error bars represent the standard deviation (n=3).

The mean Cq values of each measured duplicate can be found in the Appendix (Chapter VIII.2).

Relative expression in relation to day 0 was calculated according to Pfaffl (2001):

$$\text{Relative Expression (time)} = \frac{E_{C4H}^{\Delta CP_{C4H}\,(\text{day } 0 - \text{day X})}}{E_{\text{ST-P 2a}}^{\Delta CP_{\text{ST-P 2a}}\,(\text{day 0-day X})}}$$

While the mRNA concentration of *PpC4H_1* remained more or less steady, the expression of *PpC4H_2* and *AaC4H* behaved differently (Fig. 30a). Expression of *AaC4H* reached its first maximum after 4 days, dropped and then increased again after 12 days. On day 19 and 21 there was another increase in expression. *PpC4H_2* expression increased until day 8, afterwards the expression decreases and reached the same level as day 0 after 14 days.

ST-P 2a was used as a reference to determine the relative amount of mRNA. The following formula was used for this purpose:

$$\text{Relative expression (ST-P 2a)} = \frac{E_{\text{ST-P 2a}}^{\text{day X}}}{E_{C4H}^{\text{day X}}}$$

For a better visualization *PpC4H_1* was set as 1 and *AaC4H* and *PpC4H_2* were set in relation to *PpC4H_1*. In dependence to *PpC4H_1* it was demonstrated that *PpC4H_2* was present up to 7 times more than *PpC4H_1*. In comparison, the mRNA content of *AaC4H* was much higher, 270 to over 4000 fold (Fig. 30b). It was revealed that the template amount of *PpC4H_2* was the highest at day 6. *AaC4H* displayed two maxima, one between day 4 and 6 and the other one on day 19.

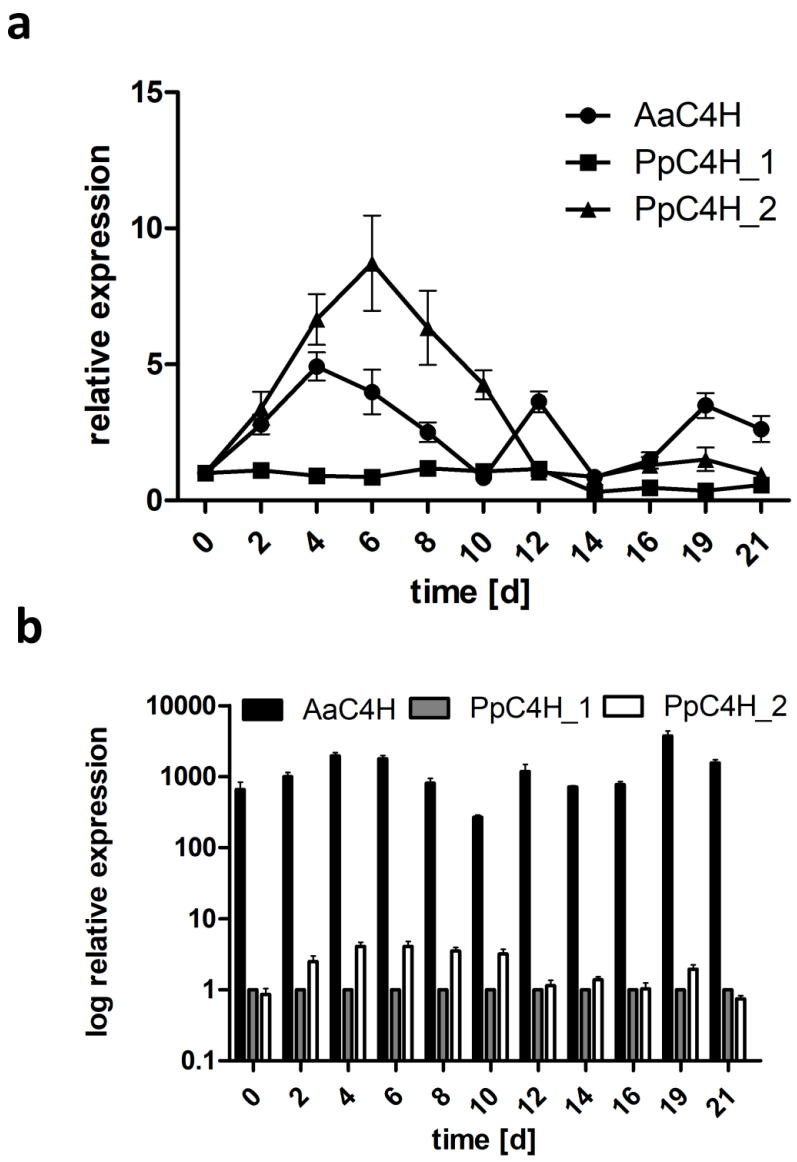


Figure 30 Quantitative real-time PCR analysis of Pp_AaC4H cinnamic acid 4-hydroxylases *AaC4H*, *PpC4H_1* and *PpC4H_2* (Wohl and Petersen 2020)**.** *ST-P 2a* was used as the reference gene. Each data point represents the mean average of at least six measured duplicates from two different RNAs, the error bars represent the standard error. **a** Time-dependent expression analysis calculated with the method described by Pfaffl (2001); **b** Expression in relation to *PpC4H_1*. Relative expression was calculated using the formula: ratio=$E_{ST\text{-}P\ 2a}^{Cq\ day\ X}$/ $E_{C4H}^{Cq\ day\ X}$ and *PpC4H_1* was set to 1. y-axis is displayed in a logarithmic scale.

2 NADPH:cytochrome P450 reductase (CPR)

2.1 Identification of a cDNA encoding CPR from *Anthoceros agrestis*

The putative nucleotide sequence of AaCPR was identified in a transcriptome project of *Anthoceros agrestis* (scaffold 10969; Szövenyi, personal communication; Szövenyi et al. 2015). PCR primers directed against this sequence led to a 2049 bp fragment (Fig. 31a). Successful amplification by 5´-RACE-PCR showed, that the sequence was already complete (Fig. 31b). The full open reading frame consisting 2052 bp (Fig. 31c) encoded a protein of 683 amino acids with a calculated molecular weight of 75.26 kDa. The sequence of *AaCPR* was deposited under the GenBank accession number MK778367. The highest identity on amino acid level was found to a CPR from *Pseudotsuga menziesii* with 67 % (Table 8).

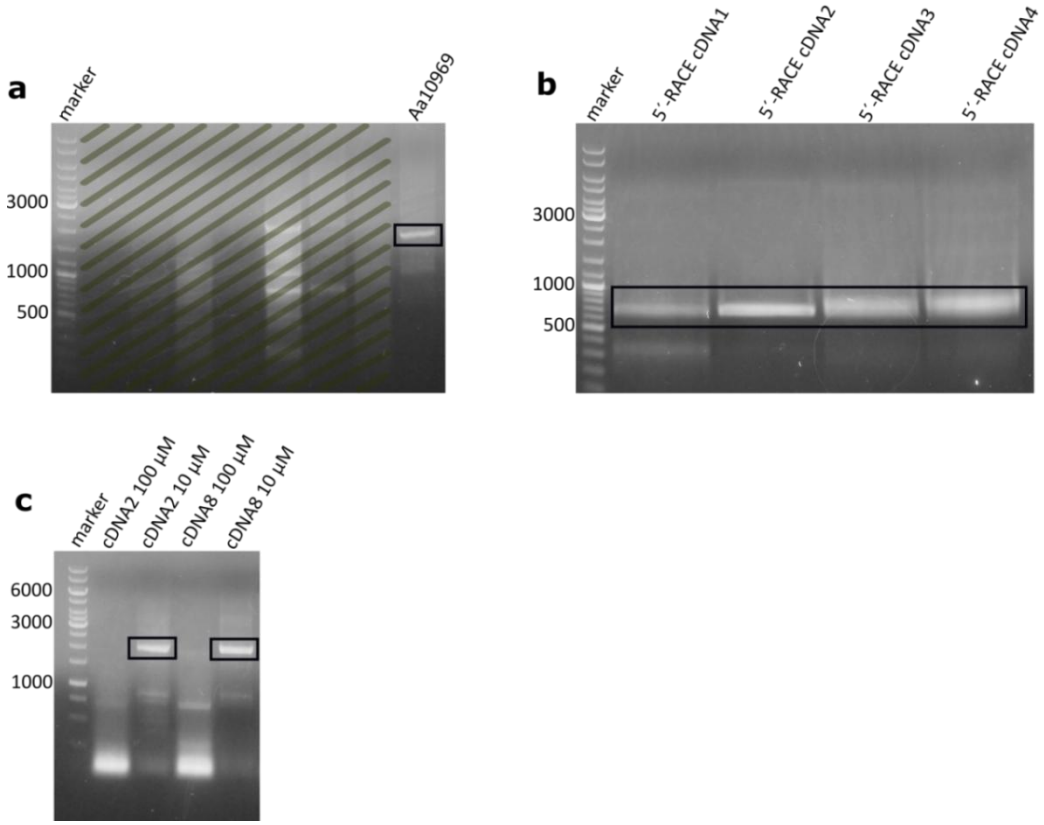


Figure 31 Amplification of *AaCPR* by PCR. a Partial sequence of *A. agrestis* scaffold 10969. Hatched areas belong to other experiments; **b** 5´-RACE-PCR with different cDNAs as template; **c** full-length sequence amplified with two different cDNAs (cDNA2 and cDNA8) and two different primer concentrations (10 µM or 100 µM).

Table 8 Protein BLAST result of AaCPR. https://blast.ncbi.nlm.nih.gov/Blast.cgi 02.12.2019

Description	Organism	Identity	E value	Accession
NADPH-cytochrome P450 reductase	*Pseudotsuga menziesii*	67 %	0.0	CAA89837.3
cytochrome P450 reductase	*Pinus contorta*	65 %	0.0	AIL29328.1
NADPH:cytochrome P450 reductase	*Taxus cuspidata*	66 %	0.0	AAT76449.1
hypothetical protein COLO4_35252	*Corchorus olitorius*	64 %	0.0	OMO57587.1
cytochrome P450 reductase	*Taxus chinensis*	66 %	0.0	AAX59902.1
flavodoxin	*Corchorus capsularis*	63 %	0.0	OMO50775.1

As shown in Figure 32, the amino acid sequence of AaCPR displayed elements generally found in plant CPRs like FMN-, FAD-, cytochrome P450-, cytochrome c- and NADPH-binding sites. In 2002, Ro et al. divided plant CPRs into two groups, based on their conservation and length of the *N*-terminal membrane anchor. AaCPR could not be associated with either class because the *N*-terminal end was even shorter than the short *N*-terminal class I CPRs. Phylogenetic analysis, using the neighbor-joining algorithm of the MEGA X software package, revealed that the hornwort can be evolutionary placed between liverworts and lycophytes (Fig. 33).

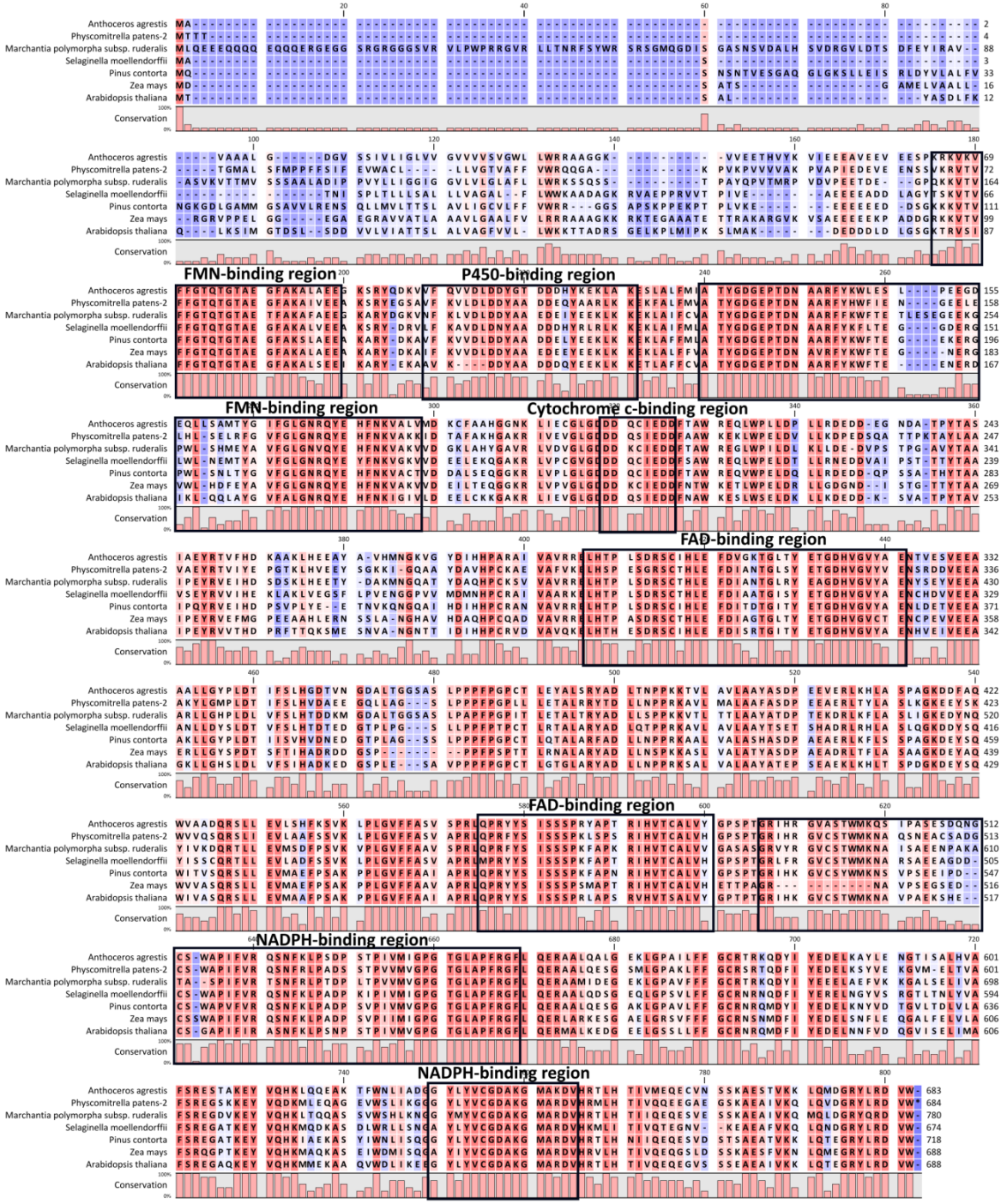


Figure 32 Alignment of amino acid sequences of CPR from *Anthoceros agrestis* (this publication, MK778367), *Physcomitrella patens*-2 (Pp3c24_17560V3.1), *Marchantia polymorpha* ssp. *ruderalis* (A0A176VZK1), *Selaginella moellendorffii* (D8SS42), *Pinus contorta* (A0A077D6G6), *Zea mays* (K7TVP4), *Arabidopsis thaliana* (Q9SB48). Highly conserved amino acids are marked red, low conservation is marked blue. The putative conserved regions are shown in boxes.

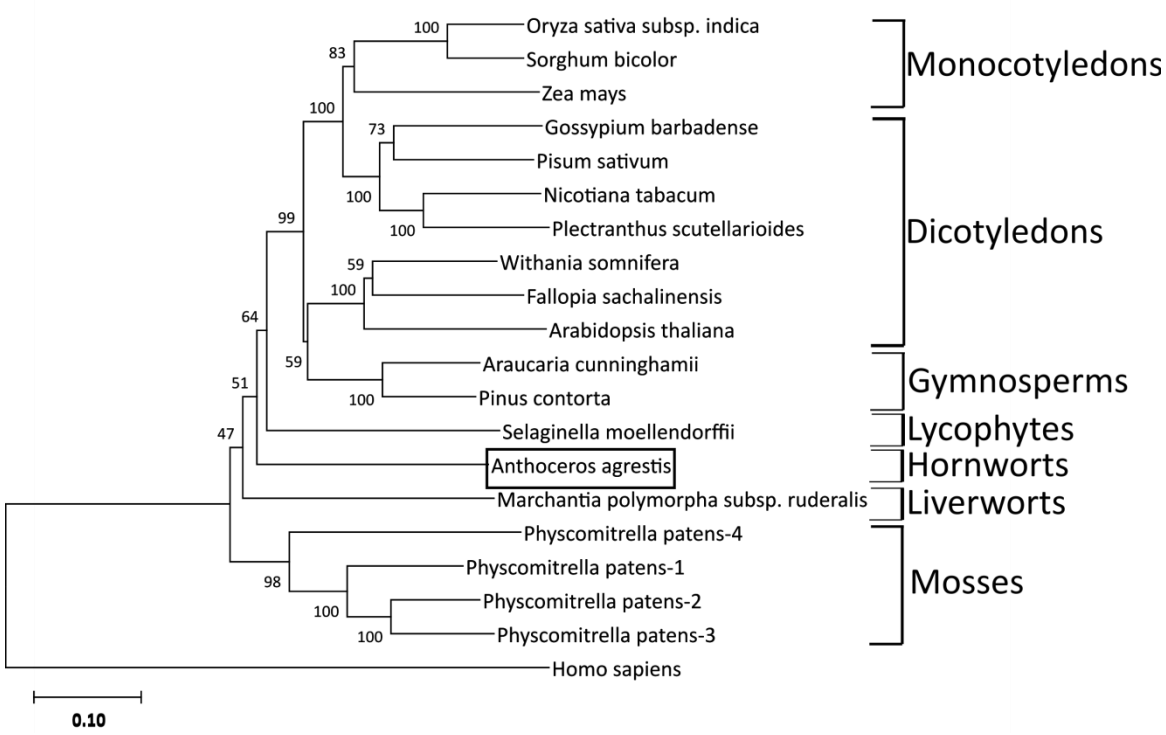


Figure 33 Phylogenetic analysis of CPR amino acid sequences. *Oryza sativa* ssp. *indica* (A2Z404), *Sorghum bicolor* (C5Y9G6), *Zea mays* (K7TVP4), *Gossypium barbadense* (A0A2P5WDR2), *Pisum sativum* (O04434), *Nicotiana tabacum* (A0A1S3X9I1), *Plectranthus scutellarioides* (CAQ37789.1), *Withania somnifera* (D6P3J1), *Fallopia sachalinensis* (A0A140JTI7), *Arabidopsis thaliana* (A0A178USJ1), *Araucaria cunninghamii* (A0A0D6QSM7), *Pinus contorta* (A0A077D6G6), *Selaginella moellendorffii* (D8SS42), *Anthoceros agrestis* (this publication, MK778367), *Marchantia polymorpha* ssp. *ruderalis* (A0A176VZK1), *Physcomitrella patens*-4 (Pp3c14_22890V3.1), *Physcomitrella patens*-1 (Pp3c20_9680V3.1), *Physcomitrella patens*-2 (Pp3c24_17560V3.1), *Physcomitrella patens*-3 (Pp3c8_19940V3.1) and *Homo sapiens* (B2R6E5; used as outgroup). The neighbor-joining tree was constructed using the MEGA X software. The robustness of the branch structure was evaluated with a bootstrap analysis (1000 replicates). Evolutionary distance is represented with the bars.

2.2 Expression of AaCPR in *Saccharomyces cerevisiae*

At first, expression in yeast was attempted. For this reason the full-length sequence of *AaCPR* was equipped with restriction sites for BamHI (5´) and XbaI (3´). Furthermore codons for a *C*-terminal 6xHis-tag in front of the stop codon were added. After ligation into pYES2 the *S. cerevisiae* strain CB018 was transformed. Plasmid uptake was checked by first isolating the plasmid and retransforming *E. coli*, since colony-PCR did not work. The plasmid was multiplied, isolated again and digestion with BamHI and XbaI was performed. For each yeast transformant, four *E. coli* colonies were picked and checked

for the plasmid (Fig. 34a). Yeast transformant 2 led to only one bacterial colony, which did not show any signs of plasmid uptake. For all other yeast transformants the presence of the plasmid could be verified. Based on this result, work continued with transformant 4. Expression in SCG_{-ura} medium was performed for up to 48 h and samples with an OD_{600}=4 were taken after 4, 8, 24 and 48 h. As a control CB018 transformed with the empty vector pYES2, was sampled accordingly. Detection with anti-His-antibody revealed no detectable protein (Fig. 34b). Furthermore there was no measurable difference in CPR activity between the crude protein extract of AaCPR and the empty vector control (Fig. 34c).

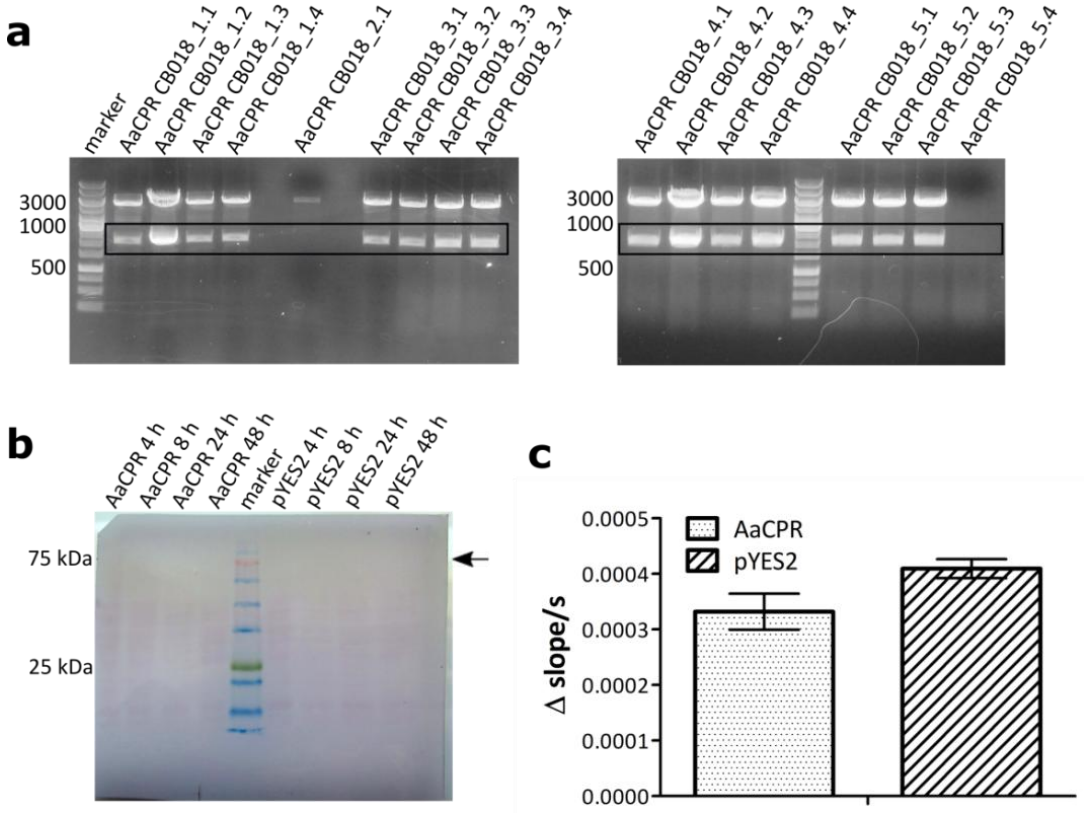


Figure 34 AaCPR expressed in *Saccharomyces cerevisiae* strain CB018. a Verification of the expression plasmid with *AaCPR* in CB018. *E. coli* was retransformed with the plasmid isolated from CB018 transformants. Four random *E. coli* transformants were chosen and the plasmid was isolated and digested with BamHI and XbaI; **b** Western blot analysis of AaCPR expressed in CB018 with anti-His-antibody. Samples with an OD_{600}=4 were taken after 4, 8, 24 and 48 h of expression in SCG_{-ura}. CB018 transformed with the empty vector (pYES2) served as a negative control; **c** Activity of heterologously expressed AaCPR in *S. cerevisiae* CB018. CB018 transformed with the empty vector (pYES2) served as a negative control. Error bars represent the standard deviation (n=3).

2.3 Expression of chimeric CPR from *Coleus blumei* (syn. *Plectranthus scutellarioides)* and *Anthoceros agrestis* in *Saccharomyces cerevisiae*

Because expression of the native *Anthoceros agrestis* CPR in *S. cerevisiae* did not work, the protein was changed in the *N*-terminal region, based on the results of Batard et al. (2000) (Chapter II.2.9). The publication explains numerous difficulties with the heterologous expression of P450s from monocots in yeast, for example a high GC content at the beginning of the sequence. With the help of a megaprimer they codon-optimized the first 18, 39 or 111 bp and were finally able to functionally express CYP73A17. Therefore, the first 277 bp of *AaCPR* were exchanged with 382 bp from *CPR* of *Coleus blumei* (CbCPR). CbCPR had already been successfully expressed in yeast by Eberle et al. (2009). Both sequences shared a restriction site for EcoRV between the highly conserved FMN and P450-binding regions. As a result, the GC content of the first 250 base pairs was reduced from the original 64% to 52%.

The sequence (*Cb+AaCPR*) was again ligated into pYES2 and the yeast strain CB018 was transformed. Plasmid uptake was checked by colony-PCR with two different primer pairs (Fig. 35a). The size of the two expected fragments were around 500 bp (A) (primers: *CbCPR* fwd new and *AaCPR* 5´R new) and 2150 bp (B) (primers: *CbCPR* fwd new and *AaCPR* rev). Both transformants showed these fragments. Expression was performed with tranformant 2. Again samples adjusted to an OD_{600}=4 were taken after 4, 8, 24 and 48 h hours and checked by Western blot analysis. As a control a sample of yeast transformed with the empty vector after 48 h expression was applied. Again no protein was detectable and there was no difference in CPR activity measurable (Fig. 35bc).

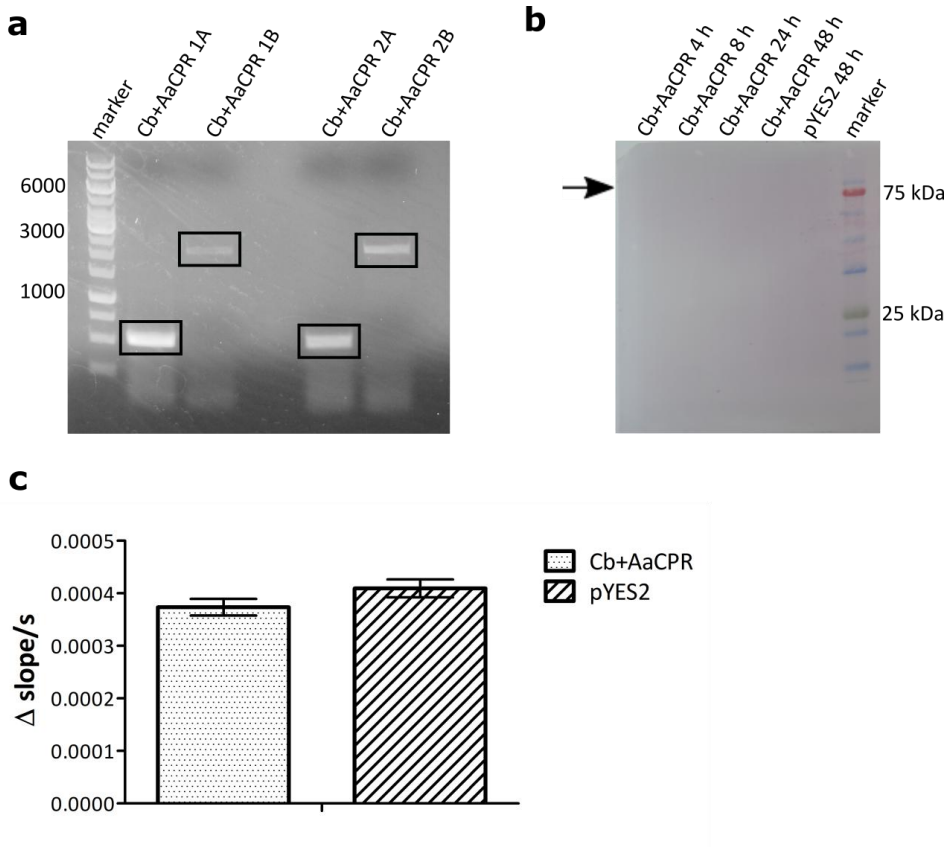


Figure 35 Expression of chimeric CPR from *Coleus blumeii* and *Anthoceros agrestis* (Cb+AaCPR) in *S. cerevisiae* CB018. a Colony PCR of *Cb+AaCPR* expressed in *S. cerevisiae* CB018. Two transformants were checked by amplification using two different primer combinations. Fragment A had a expected size of around 500 bp and fragment B of 2150 bp; **b** Western blot analysis of Cb+AaCPR with anti-His antibody. Samples were taken after 4, 8, 24 and 48 h of expression in SCG_{-ura} and adjusted to an OD_{600}=4. CB018, transformed with the empty vector (pYES2), expressed for 48 h in SCG_{-ura} served as a negative control; **c** Activity of heterologously expressed Cb+AaCPR in *S. cerevisiae* CB018. CB018 transformed with the empty vector (pYES2) served as a negative control. Error bars represent the standard deviation (n=3).

2.4 Expression of AaCPR in *Escherichia coli*

In the last years, expression of cytochrome P450 enzymes and P450-reductase has often been carried out in *E. coli* (Hotze et al. 1995; Leonard et al. 2005; Quinlan et al. 2007; Leonard et al. 2007). On the basis of these publications, *AaC4H* and *AaCPR* were fused with different modifications (Chapter II.2.10). Then the constructs were inserted into pET15b (Novagen). The first experiments were made with constructs lacking the ribosomal binding site (rbs), since it was cut off by the restriction digest with XbaI. At this time, only fus1 and fus3 were obtained. Although transformed *E. coli* strains BL21

RIPL and SoluBL21 showed no cinnamic acid 4-hydroxylase activity, both had a higher reductase activity compared to the empty vector control (Fig. 36b). For BL21 RIPL there was no significant difference between fus1 and fus3. Crude protein extracts of SoluBL21 showed a higher CPR activity for fus3. Overall, extracts from BL21 RIPL had higher activities. After separation of the membrane fractions from the supernatant, activity was predominantly detected in the supernatant. Moreover, there were two proteins detectable on Western blots for BL21 RIPL (Fig. 36a).

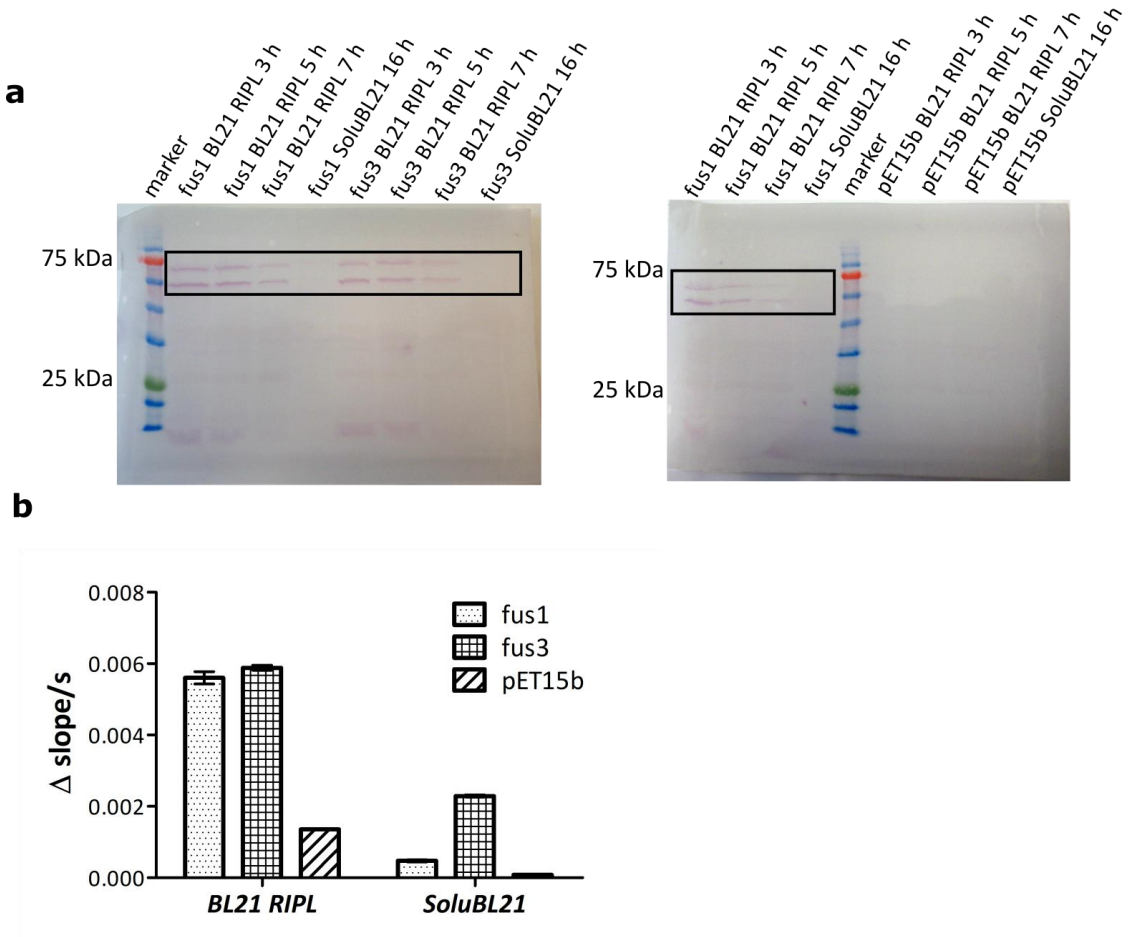


Figure 36 AaC4H/AaCPR fusion constructs fus1 and fus3 without ribosomal binding site expressed in *E. coli* BL21 RIPL and SoluBL21. a Western blot analysis with anti-His-antibody of fus1, fus3 or the empty vector control pET15b expressed in BL21 RIPL or SoluBL21. Samples with an OD_{600}=4 were taken after 3, 5 or 7 h (BL21 RIPL) or 16 h (SoluBL21) after induction with 1 mM IPTG; **b** Activity of heterologously expressed fus1 and fus3 in BL21 RIPL and SoluBL21. Crude protein extract of fus1 and fus3 (100 µg protein) was compared to a crude protein extract (100 µg protein) of the transformed empty vector control (pET15b). Error bars represent the standard deviation (n=3).

After discovering the lack of a rbs, the sequence was checked for the same rbs as occurring in pET15b. Four ribosomal binding sites were found in the fusion construct, but only two of them led to proteins in the correct reading frame with a molecular weight of 88 and 62 kDa. The two bands on the Western blot were correlated to these two proteins. It is unknown, whether both or only one of them was responsible for CPR activity. For the other two possible ribosome binding sites a stop codon followed shortly after the first possible start codon.

After insertion of the rbs into the constructs and transformation of the *E. coli* strain BL21 RIPL (only fus2 and fus3 were tested) there was no significant measurable difference in CPR activity between the fusion proteins and the empty vector control (Fig. 37a). Only transformation of *E. coli* C41(DE3) led to the expression of the fusion constructs and again a soluble AaCPR could be detected (Fig. 37b). Verification of C4H activity is described in chapter III.1.4. Higher CPR activity than the empty vector control could be detected for all fusion proteins (Fig. 37c). All fusion proteins showed almost the same activity, using crude protein extract. The membrane fractions had a comparably low reductase activity and only fus3 displayed a small difference to the empty vector control. The activity could again be traced back to a soluble AaCPR, since the supernatant displayed the same effect as the crude protein extract. For characterization, soluble AaCPR from fus3 was chosen, since this transformant showed also the highest C4H activity. After purification over a Ni-NTA column the empty vector control had only an insignificant activity left, while fus3 was still working (Fig. 37d). Electrophoresis on a SDS-gel showed a nearly pure protein of the expected size in the elution fraction (Fig. 37e).

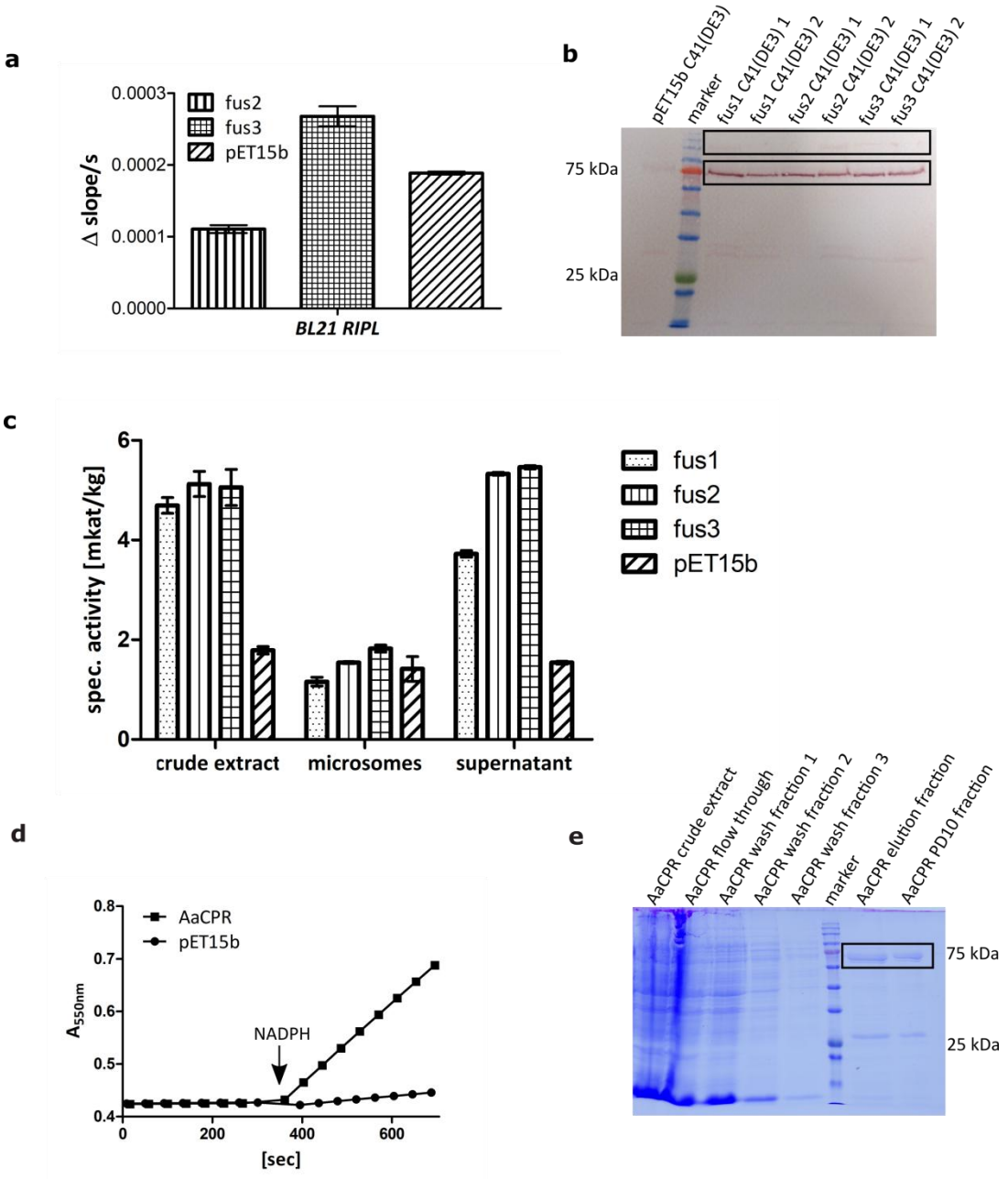


Figure 37 AaC4H/AaCPR fusion constructs expressed in *E. coli* BL21 RIPL and C41(DE3) with ribosomal binding site. a Activity of fus2 and fus3 compared to the empty vector control (pET15b) expressed in BL21 RIPL. Error bars represent the standard deviation (n=3); **b** Western blot analysis with anti-His-antibody of fusion constructs and the empty vector control expressed in C41(DE3). For each fusion construct two transformants were checked. The calculated molecular masses for the desired proteins were 131.85 kDa (fus1), 132.23 kDa (fus2) and 128.79 kDa (fus3); **c** CPR activity of the fusion constructs measured with crude protein extracts, microsomes and supernatant isolated from *E. coli* C41(DE3). *E. coli* transformed with empty vector (pET15b) served as a negative control. Error bars represent the standard deviation (n=3); **d** Assay of AaCPR and the empty vector control pET15b after metal chelate purification. Addition of NADPH after 350 sec is marked by an arrow; **e** SDS-PAGE of fractions obtained by metal chelate purification of AaCPR.

2.4.1 Temperature and pH-optimum of AaCPR

The pH-optimum of AaCPR was shown to be at pH 7.3 using a Tris-HCl buffer and the temperature optimum was at 40 °C (Fig. 38). These data were obtained from three technical replicates. The exact pH of the assay was measured before addition of NADPH. pH-optmimum was measured at 40 °C and temperature optimum was measured with a 100 mM Tris-HCl pH 7.5 buffer.

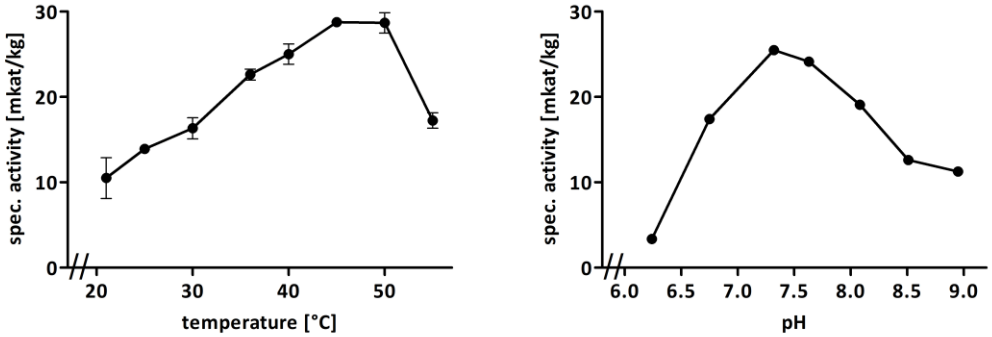


Figure 38 Temperature (left) and pH-optimum (right) of AaCPR (n=3, SD).

2.4.2 AaCPR: reaction kinetics

Kinetic parameters were determined, using up to 600 µM NADPH. The activity was calculated with the extinction coefficient 21 mM^{-1} cm^{-1} for cytochrome c. Data for kinetic analyses were obtained from four independent protein isolations with four technical replicates for each substrate concentration. The resulting substrate saturation curves led to an apparent K_m-value for NADPH of 9.2 ± 0.3 µM and a V_{max} of 49.2 ± 3.9 mkat/kg (Fig. 39; Table 9). Since the exact molecular mass of the soluble AaCPR was unknown, k_{cat} was calculated based on the first start codon before the AaCPR part of fus3. The resulting 72.8 kDa (including the His-tag) matched to the signal from the Western blot analysis. Nevertheless, k_{cat} is only an estimated value.

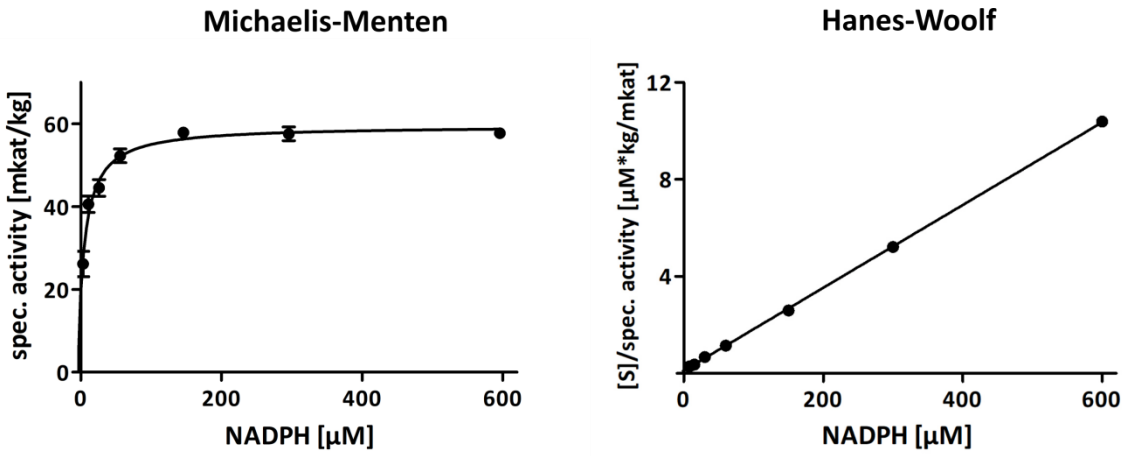


Figure 39 Dependence of AaCPR on NADPH displayed in Michaelis-Menten (left) and Hanes-Woolf (right) diagrams. Graphs represent the mean values of one biological replicate with four technical replicates. Error bars represent the standard deviation.

Table 9 Kinetic values of AaCPR for NADPH. K_m-values were calculated with help of Michaelis-Menten (MM), Lineweaver-Burk (LB) and Hanes-Woolf plots (HW) (mean ± SE). Mean values represent four biological replicates with four technical replicates for each substrate concentration. For the calculation of k_{cat} a predicted molecular mass of 72.82 kDa was used.

		AaCPR			
		K_m [µM]	V_{max} [mkat/kg]	k_{cat} [1/s]	k_{cat}/K_m [1/s*mM]
NADPH	MM	9.2 ± 0.3	49.2 ± 3.9	3.6 ± 0.4	390.4
	LB	9.8 ± 0.2	49.7 ± 4.1	3.6 ± 0.4	369.7
	HW	8.4 ± 0.6	49.1 ± 3.8	3.6 ± 0.4	426.6

2.5 Expression of AaCPR in *Physcomitrella patens*

For the expression in *Physcomitrella patens* the restriction site in the 3´-region was changed to NotI. The sequence was then ligated into the Gateway cloning entry vector pENTR™1A. The LR recombination reaction was performed and the sequence was introduced into the expression vector pTHUbiGate. After transformation of *P. patens*, five stable transformants were obtained and checked for protein production by Western blotting (Fig. 40a). Only transformants PpAaCPR_1 and PpAaCPR_3 clearly showed a

protein with the correct size and a faint band could be observed in PpAaCPR_2. CPR activity was measured with the crude protein extracts from AaCPR_Pp_3.

There was no detectable difference in activity between the untransformed wild type culture and PpAaCPR_1 (Fig. 40b), probably because Pp itself possesses at least four potential CPRs (Phytozome: PpCPR_1 = Pp3c20_9680V3.1, PpCPR_2 = Pp3c24_17560V3.1, PpCPR_3 = Pp3c8_19940V3.1 and PpCPR_4 = Pp3c14_22890V3.1). Nevertheless CPR activity was also checked in the microsomes and the supernatant. Again there was no difference to the wild type (Fig. 40b). However, it was conspicuous that the microsomes showed no activity at all, although the protein was located in the membrane fraction (Fig. 40c). A possible explanation could be that the photometric detection at 550 nm is disturbed by chlorophyll in the microsomal suspension. Loss of activity after preparation of microsomes was also observed for AaC4H (Chapter III.1.5.1) and AaCYP98, although 25 % activity still remained compared to the crude protein extract. Since expression in *P. patens* led to no detectable AaCPR activity, these cultures and heterologously expressed enzymes were not further characterized.

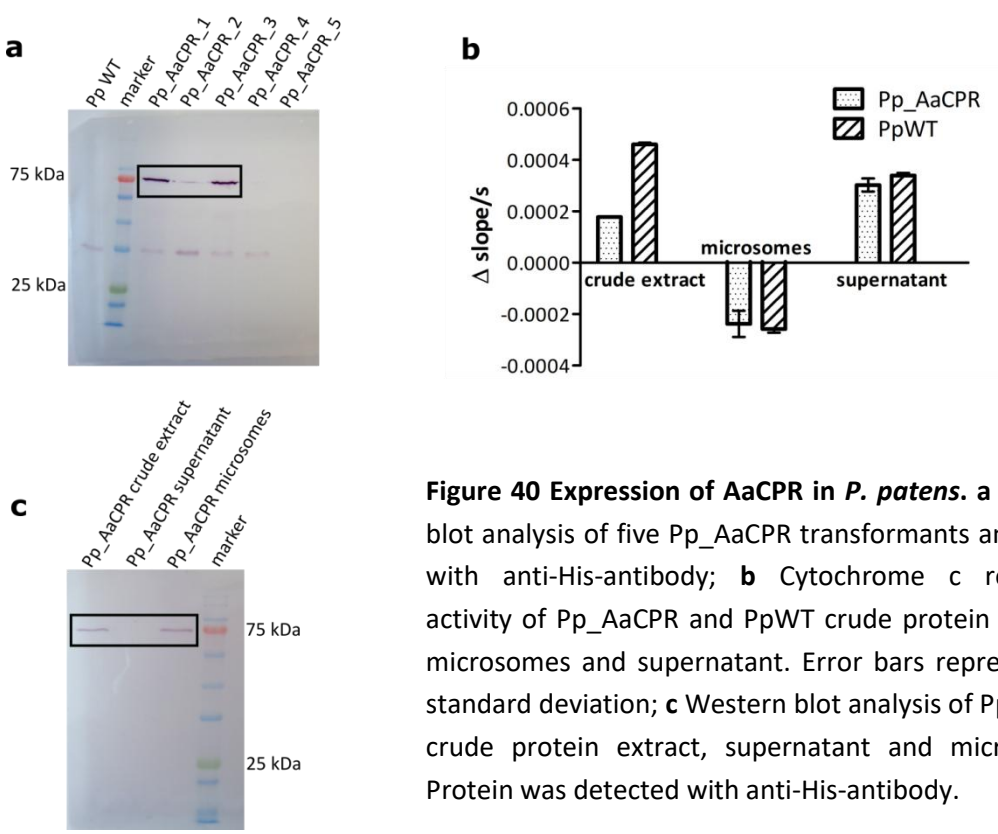


Figure 40 Expression of AaCPR in *P. patens*. a Western blot analysis of five Pp_AaCPR transformants and PpWT with anti-His-antibody; **b** Cytochrome c reductase activity of Pp_AaCPR and PpWT crude protein extracts, microsomes and supernatant. Error bars represent the standard deviation; **c** Western blot analysis of Pp_AaCPR crude protein extract, supernatant and microsomes. Protein was detected with anti-His-antibody.

3 4-Coumaric acid CoA-ligase (4CL)

3.1 Identification of a cDNA encoding 4-coumarate CoA-ligase (4CL) from *Anthoceros agrestis*

The next step of the phenylpropanoid pathway is the activation of 4-coumaric acid by a 4-coumarate CoA-ligase (4CL). Scaffold 6378 from the *Anthoceros punctatus* database was identified as a putative *4CL*. After amplification of a partial fragment of 909 bp from cDNA of *Anthoceros agrestis* (Fig. 41a), 3´- and 5´-RACE PCR was performed (Fig. 41b). 5´-RACE PCR revealed two possible start codons, therefore two forward full-length primers were designed. The full open reading frame of the two sequences consisted of 1647 bp (*4CL_1*) and 1599 bp (*4CL_2*) (Fig. 41c), encoding proteins of 548 aa (58.42 kDa) and 532 aa (56.86 kDa). The *Aa4CL_2* sequence was deposited in GenBank under the accession number MN922305. The highest identities on protein level for both possible amino acid sequences is displayed in Table 10.

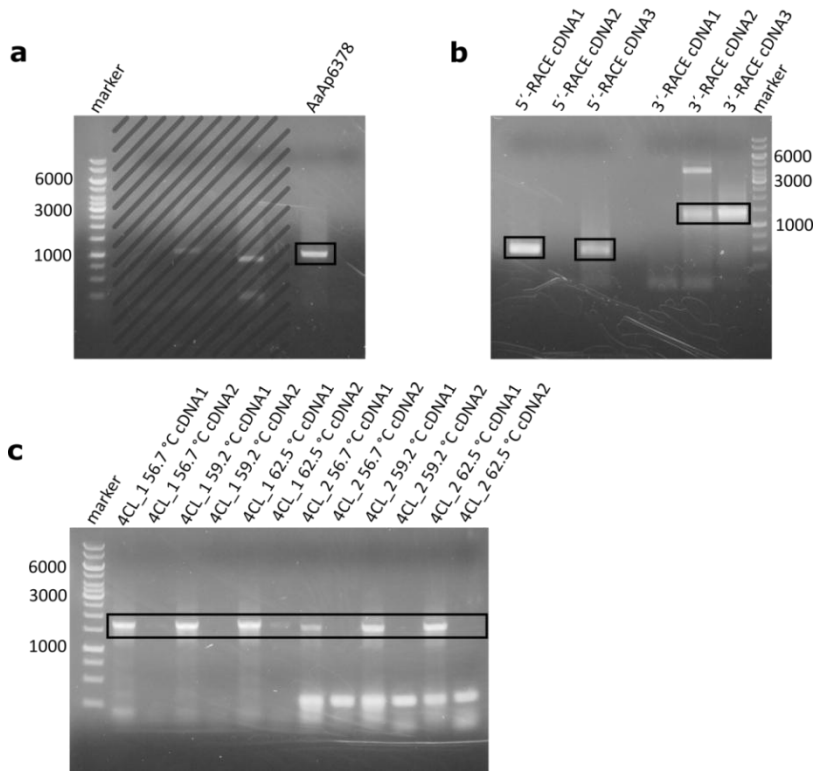


Figure 41 Amplification of *Aa4CL* by PCR. a Partial sequence; **b** 5´-RACE and 3´-RACE using three different cDNAs; **c** *Aa4CL_1* and *Aa4CL_2* full-length sequences with different cDNAs as template and three different annealing temperatures. Hatched areas belong to other experiments.

Table 10 Protein BLAST result of Aa4CL. https://blast.ncbi.nlm.nih.gov/Blast.cgi 02.12.2019

Description	Organism	Identity	E value	Accession
hypothetical protein AXG93_1953s1040	*Marchantia polymorpha subsp. ruderalis*	66 %	0.0	OAE23145.1
4-coumarate: coenzyme A ligase	*Cryptomeria japonica var. sinensis*	64 %	0.0	QDC33553.1
4-coumarate-CoA ligase 2-like	*Hevea brasiliensis*	63 %	0.0	XP_021655378.1
4-coumarate-CoA ligase	*Ginkgo biloba*	63 %	0.0	AMN10098.1
4-coumarate-CoA ligase enzyme	*Pinus taeda*	65 %	0.0	AAA92669.1
4-coumarate:coenzyme A ligase 4	*Physcomitrella readeri*	64 %	0.0	ABY21311.1

Figure 42 shows an alignment of different 4CL amino acid sequences. Outlined are the conserved peptide motifs in box I and box II. The twelve amino acid residues proposed to function as the 4CL substrate specificity code are marked in green (Schneider et al. 2003). Mostly all of them were conserved throughout the 4CLs in the plant kingdom. The amino acid sequence of Aa4CL_1 was moreover phylogenetically analyzed (Fig. 43). The resulting tree supports the thesis of a gene duplication before the split of gymnosperms and angiosperms (Li et al. 2015). While some angiosperm 4CLs can be found in the same branch as moss, lycophyte, hornwort, fern and gymnosperm 4CLs, most angiosperm 4CLs are located in a separate branch. In this case Aa4CL_1 is located between the lycophyte 4CLs from *Selaginella moellendorfii* and the gymnosperm 4CLs (*Pinus* species).

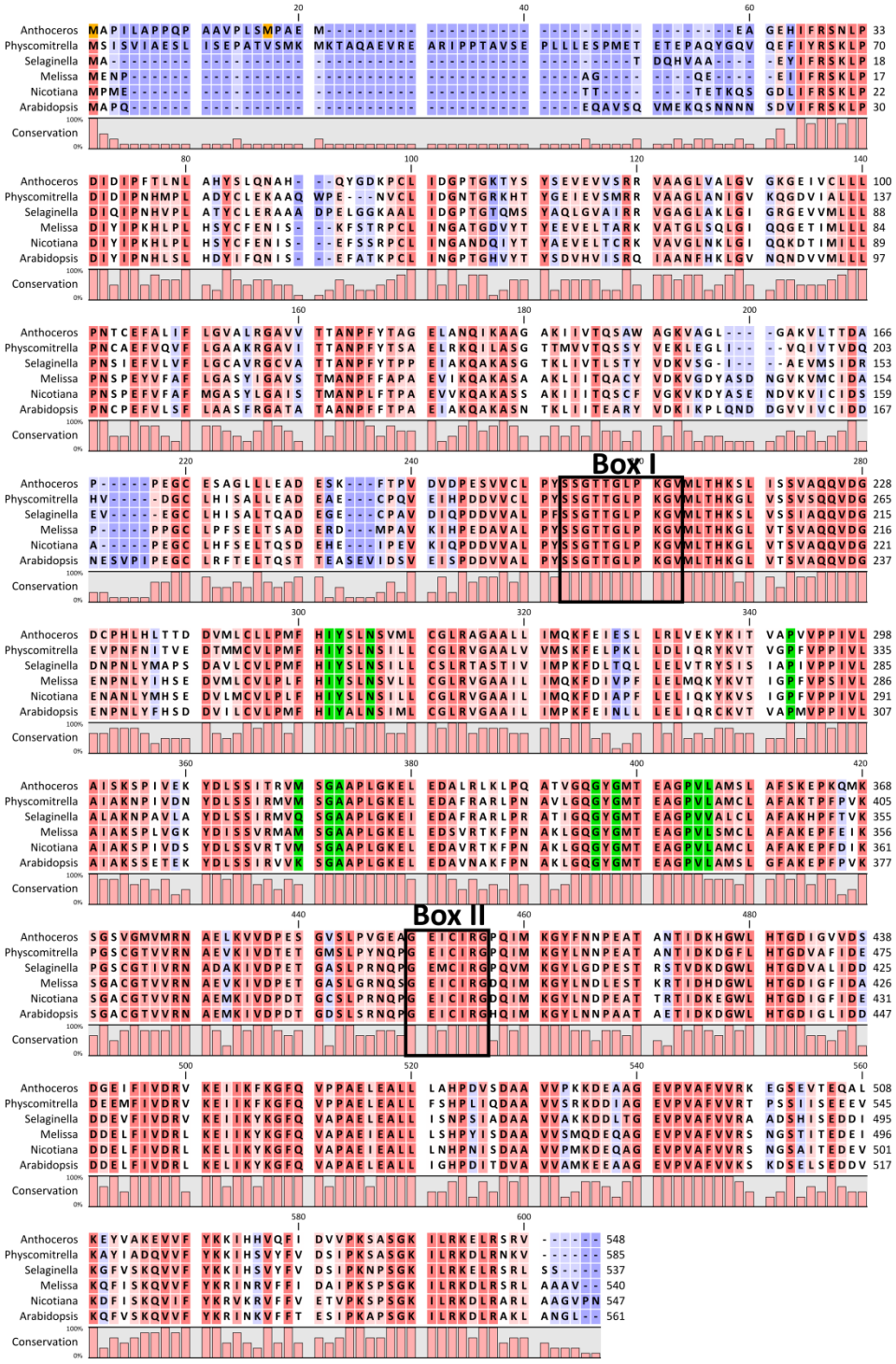


Figure 42 Alignment of 4CL amino acid sequences. *Anthoceros agrestis* (this work, MN922305*), Physcomitrella patens* (B7SBA0), *Selaginella moellendorffii* (D8RGA2), *Melissa officinalis* (E1UYU5), *Nicotiana tabacum* (O24145) and *Arabidopsis thaliana* (Q42524). Highly conserved amino acids are marked red, low conservation is marked blue. Typical conserved sequence motifs are marked by boxes. Amino acids from the 4CL substrate specificity code are highlighted green (Schneider et al. 2003) and the two possible start methionine residueas from Aa4CL are highlighted orange.

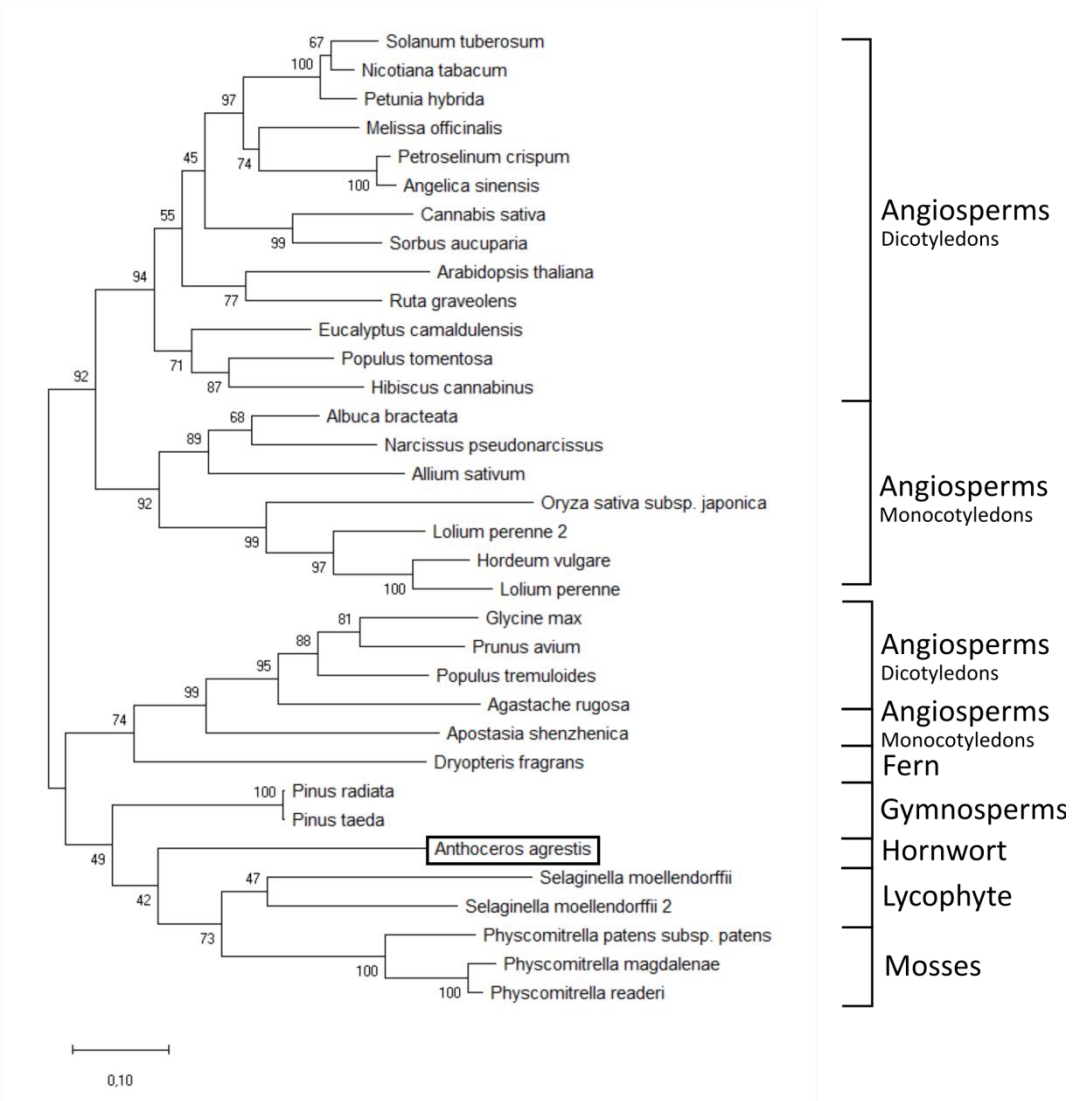


Figure 43 Phylogenetic analysis of 4CL amino acid sequences. The maximum likelihood tree was constructed using the MEGA X software. The robustness of the branch structure was evaluated with a bootstrap analysis (1000 replicates). Evolutionary distance is represented with the bars. Sequences were obtained from BRENDA database. *Solanum tuberosum* (P31684), *Nicotiana tabacum* (O24145), *Petunia hybrida* (I3PB37), *Melissa officinalis* (E1UYU5), *Petroselinum crispum* (P14913), *Angelica sinensis* (A0A140IPI1), *Cannabis sativa* (V5KXG5), *Sorbus aucuparia* (D5LLN7), *Arabidopsis thaliana* (Q42524), *Ruta graveolens* (B3TJI5), *Eucalyptus camaldulensis* (D0EY64), *Populus tomentosa* (H9AZ23), *Hibiscus cannabinus* (T1RK64), *Albuca bracteata* (U3N732), *Narcissus pseudonarcissus* (A0A2H5AIX5), *Allium sativum* (G3CU71), *Oryza sativa* subsp. *japonica* (P17814), *Lolium perenne* 2 (Q9M7S1), *Hordeum vulgare* (U6A2G2), *Lolium perenne* (Q9M7S2), *Glycine max* (P31687), *Prunus avium* (F2VR44), *Populus tremuloides* (O81140), *Agastache rugosa* (Q6PQ01), *Apostasia shenzhenica* (A0A2I0AJN1), *Dryopteris fragrans* (A0A0B4L7M5), *Pinus radiata* (B8QN50), *Pinus taeda* (P41636), *Anthoceros agrestis* (this work, MN922305), *Selaginella moellendorffii* (D8S6A2), *Selaginella moellendorffii* 2 (D8RGA2), *Physcomitrella patens* subsp. *patens* (B7SBA0), *Physcomitrella magdalenae* (B7SBB0) and *Physcomitrella readeri* (B7SBA6).

3.2 Expression of Aa4CL in *Escherichia coli*

For expression in *E. coli* restriction sites for PstI (5´) and HindIII (3´) were added to the two full-length sequences of *Aa4CL*. After ligation into the expression vector pRSET C (Invitrogen), the plasmids were introduced into *E. coli* SoluBL21. A polyhistidine-tag for purification, already present on the plasmid, was attached to the *N*-terminus. The expression procedure was performed as described in chapter II.3.1.1.

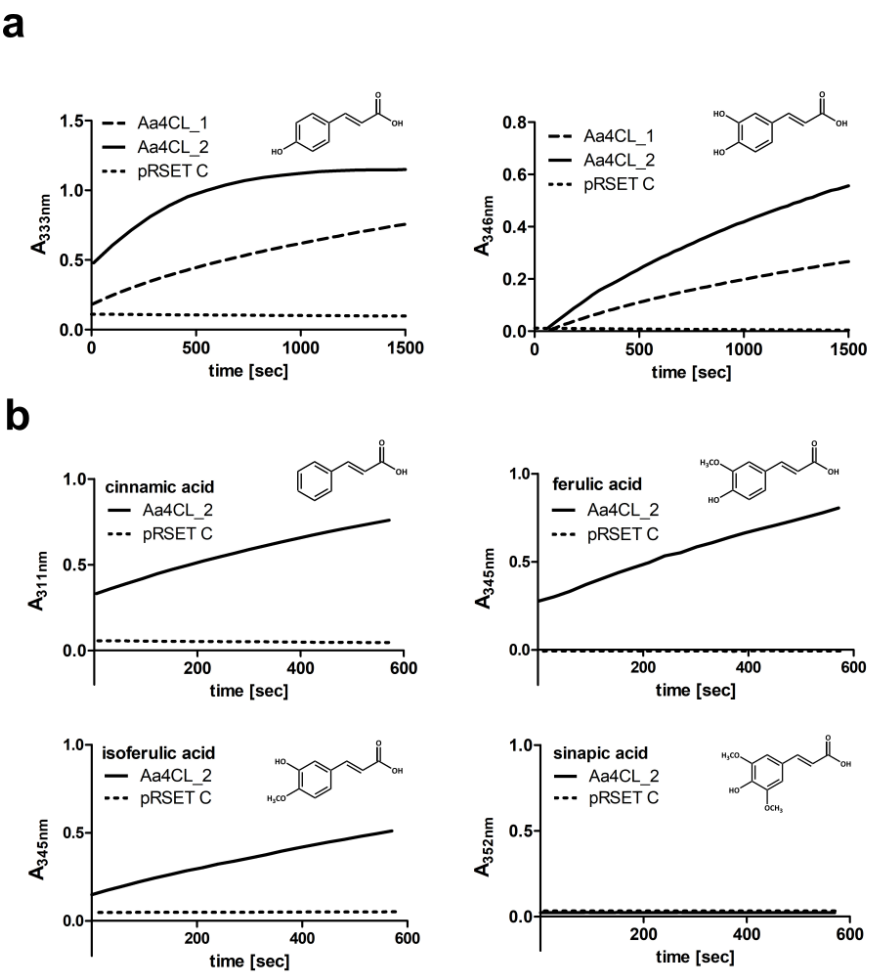


Figure 44 Activity of Aa4CL with different cinnamic acid derivatives. Crude protein extract (60 µg protein) from transformed *E. coli* with the empty vector pRSET C served as negative control. **a** Assays with crude protein extract of Aa4CL_1 and Aa4CL_2 (60 µg protein) with 500 µM 4-coumaric acid (left) or caffeic acid (right); **b** Assays with purified Aa4CL_2 (3 µg) with 500 µM cinnamic acid, ferulic acid, isoferulic acid or sinapic acid.

Enzyme activities were measured at first with crude protein extract of Aa4CL_1/ Aa4CL_2 or a crude protein extract from *E. coli* transformed with the empty vector using

4-coumaric acid and caffeic acid (Fig. 44a). Both enzymes showed an increase in absorption at the absorption maximum of the corresponding CoA ester compared to the empty vector control. Aa4CL_2 displayed a higher conversion rate than Aa4CL_1. Afterwards the proteins were purified by metal chelate chromatography to afford nearly pure protein, identified by SDS-PAGE and Western blotting (Fig. 45a, b). Although the protein concentration for the eluted 4CL_1 was too low for detection, there was a clear evidence of Aa4CL_2 in the crude extract and elution fractions. There was no protein traceable in the crude protein extract of the empty vector control. In all expression attempts the protein formation of the shorter Aa4CL_2 exceeded protein formation of Aa4CL_1.

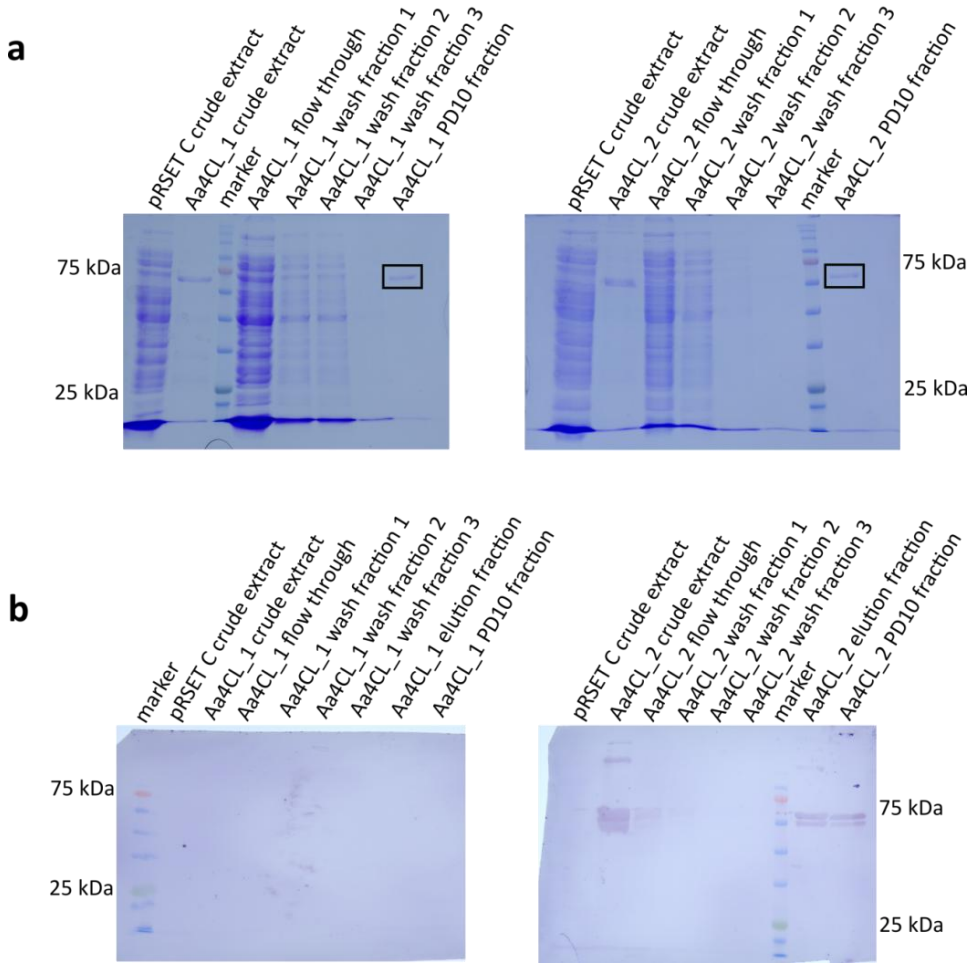


Figure 45 SDS-PAGE and Western blot of Aa4CL_1 and Aa4CL_2 expressed in *E. coli* after metal chelate purification. Crude protein extract (100 µg protein) from transformed *E. coli* with the empty vector pRSET C served as a negative control. **a** SDS-PAGE; **b** Western blot with anti-6x His-antibody.

Purified Aa4CL_2 was then tested with the substrates cinnamic acid, ferulic acid, isoferulic acid and sinapic acid. Again a crude protein extract of the empty vector served as negative control (Fig. 44b). Aa4CL_2 additionally activated cinnamic acid, ferulic acid and isoferulic acid but lacked affinity for sinapic acid. The empty vector control did not show any activity. Assays with 2-coumaric acid and 3-coumaric acid were analyzed by LC-MS. Both subtrates were accepted by Aa4CL_2 (Fig. 46). In this case, heat-denatured protein (10 min, 95 °C) served as negative control.

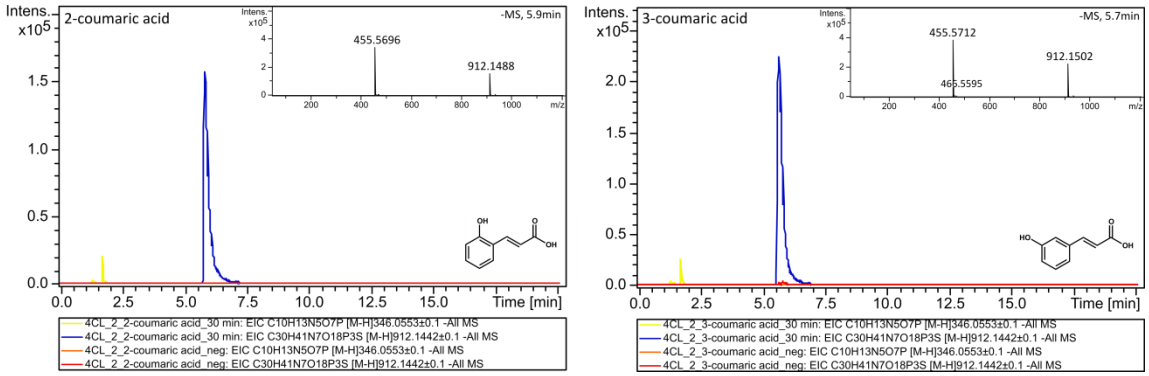


Figure 46 LC-MS result of purified Aa4CL_2 (8 µg protein) incubated with 500 µM 2-coumaric or 3-coumaric acid, 1.25 mM ATP and 1 mM CoA for 30 min at 40 °C. Heat-denatured protein (10 min 95 °C) was used as a negative control. The chromatograms show the EIC of the expected products AMP (m/z 346.0553 ± 0.1) in yellow for Aa4CL_2 and orange for the negative control. The corresponding CoA-ester is displayed in blue for Aa4CL_2 and red for the negative control. The exact mass of the resulting CoA-ester is shown in each chromatogram. For all produced CoA-esters the [M-H] pseudo molecular ion (m/z 912.15) as well as the doubly charged molecular ion (m/z 455.57) [M/2]-H was observed.

The relative substrate preference of Aa4CL_1 and Aa4CL_2 was compared to ferulic acid (Fig. 47). For both Aa4CL_1 and Aa4CL_2, 4-coumaric acid and ferulic acid were the substrates leading to the highest specific activity. Followed by caffeic acid with more than 75 % and isoferulic acid with 71 %/63 %. Cinnamic acid was the least accepted substrate for both enzymes with less than 55 % compared to ferulic acid.

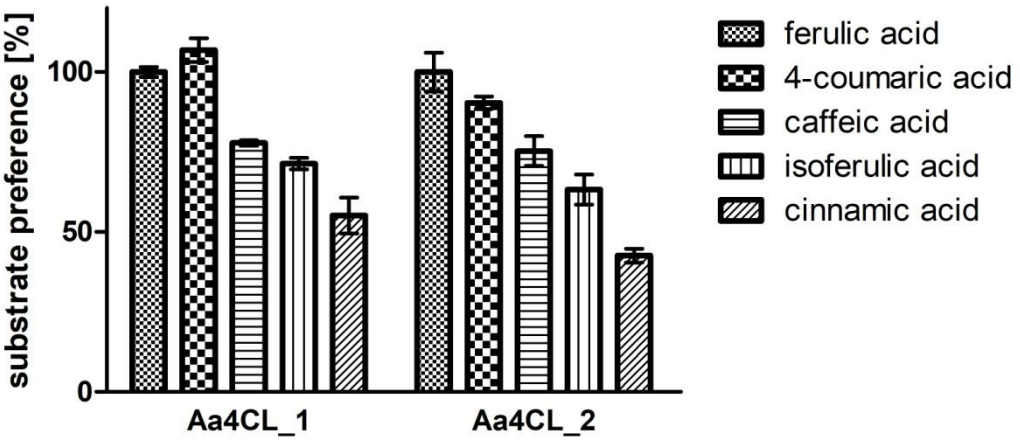


Figure 47 Relative substrate preference of Aa4CL_1 and Aa4CL_2. Specific activity measured with ferulic acid was set to 100 %. Assays were incubated at 40 °C with 500 µM substrate and 2.7 µg purified Aa4CL_1 or 2 µg purified Aa4CL_2. Bars represent the mean value of 6 technical replicates and the error bars represent the standard deviation.

Activation of benzoic acid derivatives by Aa4CL_2 was measured by LC-MS analysis. In this case, activation of caffeic acid was used as positive control. Chromatograms shown in Figure 48 display the EIC of the expected products. The corresponding CoA-ester is shown as the doubly charged molecular ion [M/2]-H. Furthermore, the formation of AMP was detected. Except for the positive control (caffeic acid), all tested benzoic acid derivatives were not activated. Only AMP was found, thus it is likely, that Aa4CL still has residual ATPase activity when a suitable substrate is lacking.

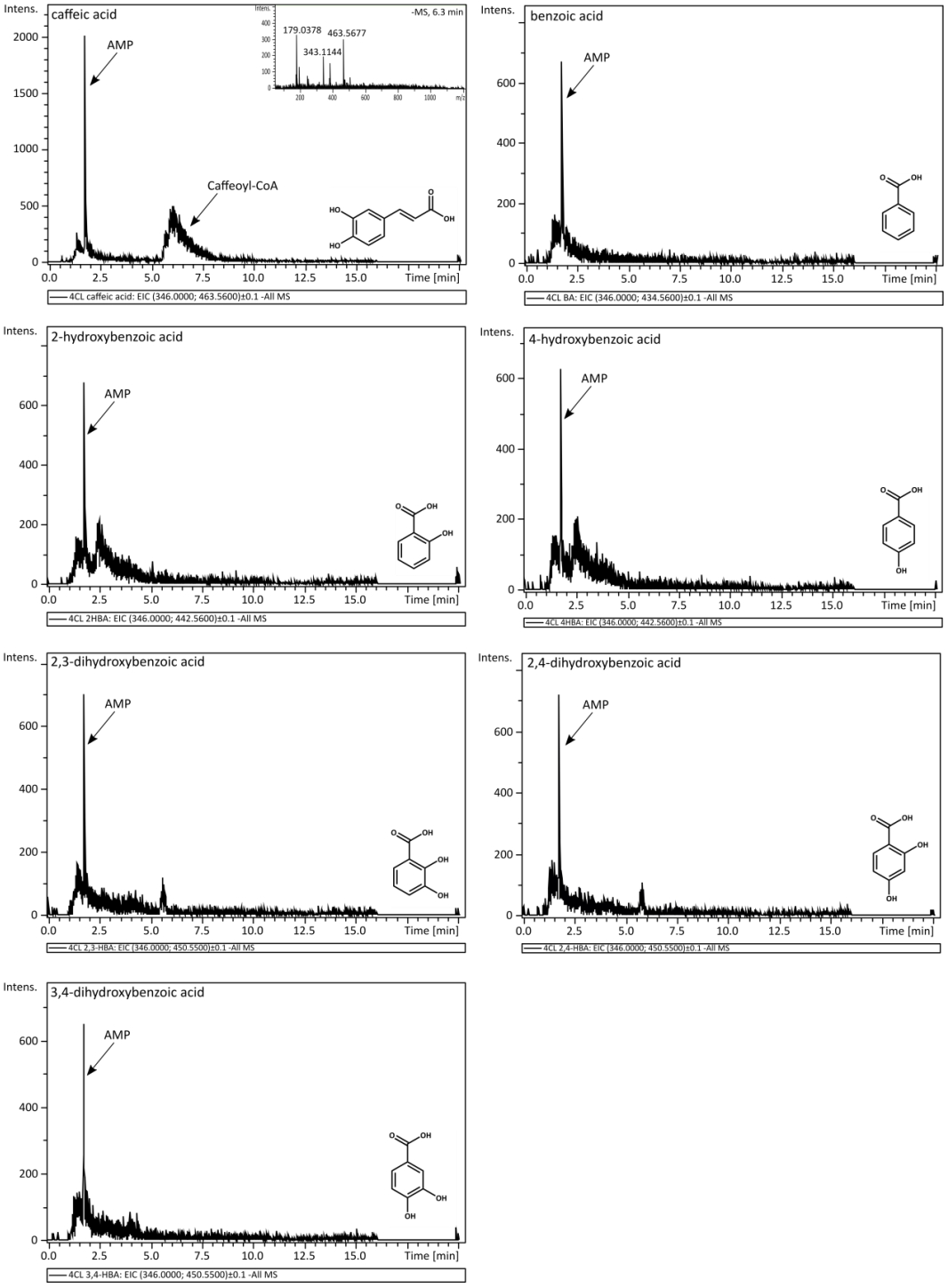


Figure 48 LC-MS result of assays with purified Aa4CL_2 with different benzoic acid derivatives or caffeic acid. 7.5 µg Aa4CL_2 was incubated with 500 µM caffeic acid, benzoic acid, 2-hydroxybenzoic acid, 4-hydroxybenzoic acid, 2,3-dihydroxybenzoic acid, 2,4-dihydroxybenzoic acid or 3,4-dihydroxybenzoic acid for 1 h at 35 °C. The chromatograms show the EIC of the expected products AMP (m/z 346 ± 0.1) and the corresponding CoA-ester (doubly charged molecular ion [M/2]-H) of caffeic acid (m/z 463.56 ± 0.1), benzoic acid (m/z 434.56 ± 0.1), monohydroxylated benzoic acid (m/z 442.56 ± 0.1) or dihydroxylated benzoic acid (m/z 450.55 ± 0.1).

3.2.1 Temperature and pH-optimum of Aa4CL

The temperature optimum was determined with the substrate 4-coumaric acid. The pH-optimum was measured with ferulic acid as substrate. Data were collected from two independent measurements with three or four technical replicates for each data point. Both Aa4CL protein variants showed a high temperature optimum at around 40 °C (Fig. 49a left). 4CL_1 even indicated a plateau between 40 and 50 °C. The pH-optimum was around pH 7.0-7.5 (Fig. 49a right). The use of buffers with higher pH led to a color change (Fig. 49b) and measurement at pH higher than 8.5 was not possible, because the data fluctuated too much over time. K_m-values were determined at 40 °C using buffer with pH 7.5.

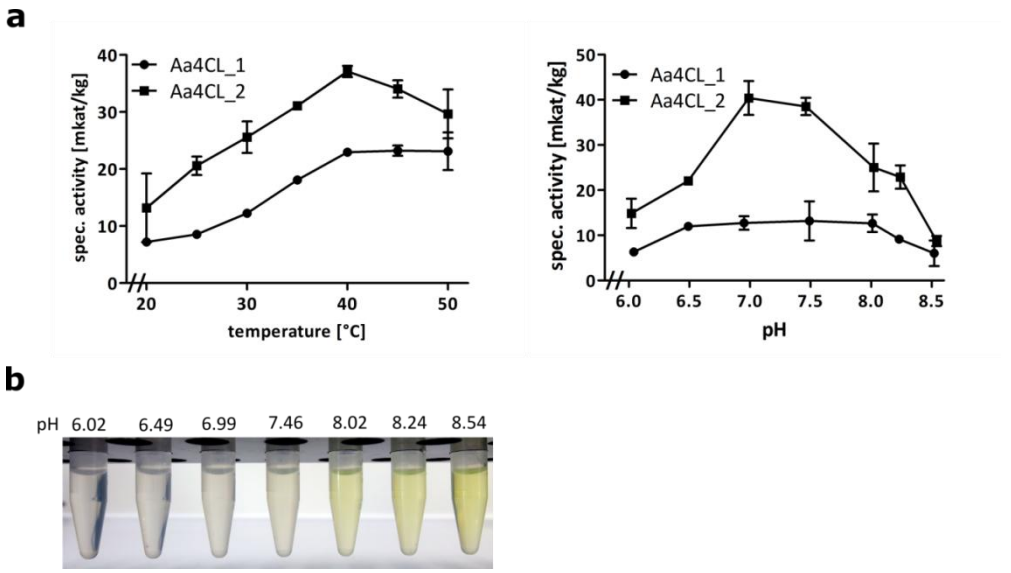
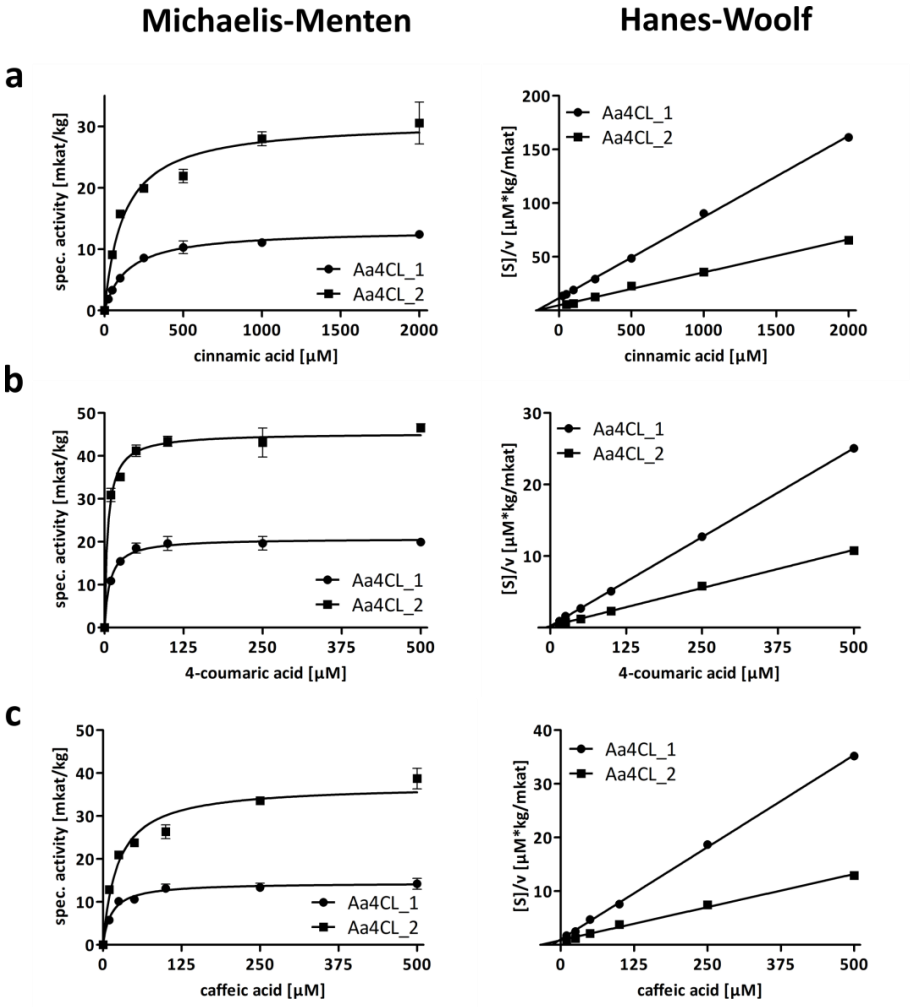


Figure 49 a Temperature (left) and pH-optimum (right) of Aa4CL_1 and Aa4CL_2**.** Mean values were calculated from two biological replicates with three or four technical replicates. Error bars represent the standard deviation; **b** Assay color after incubation of Aa4CL with ferulic acid and KPi-buffers with different pH values.

3.2.2 Aa4CL: reaction kinetics

Kinetic values were obtained from three independent protein isolations with three or four technical replicates for each substrate concentration. Parameters were measured for cinnamic acid, 4-coumaric acid, caffeic acid, ferulic acid and isoferulic acid, ATP and CoA. The K_m-values for ATP and CoA were determined using ferulic acid. All enzyme kinetic data are summarized in Table 11. The enzyme kinetic data were analyzed by

graphical displays of Michaelis-Menten, Lineweaver-Burk (data not shown) and Hanes-Woolf plots (Fig. 50). Both, 4CL_1 and 4CL_2 showed the same substrate affinity. 4-Coumaric acid and isoferulic acid were the substrates with the lowest K_m-values (Table 11). Significant differences could be observed for V_{max} between 4CL_1 and 4CL_2. 4CL_2 was significantly faster than 4CL_1. Since substrate affinities clearly differed from the specific activity, k_{cat} and k_{cat}/K_m were calculated based on the values obtained by the Michaelis-Menten plots (Table 12). The molecular mass including the additional *N*-terminal amino acids from the expression vector was calculated to be 63.28 kDa (Aa4CL_1) and 61.72 kDa (Aa4CL_2). k_{cat} also displayed the higher activity of Aa4CL_2. 4-Coumaric acid and isoferulic acid were shown as the best substrates by calculation of k_{cat}/K_m. It seems reasonable that Aa4CL displayed a much lower affinity for ATP compared to CoA, since CoA is exchanged with the bound AMP in the second step of the reaction.



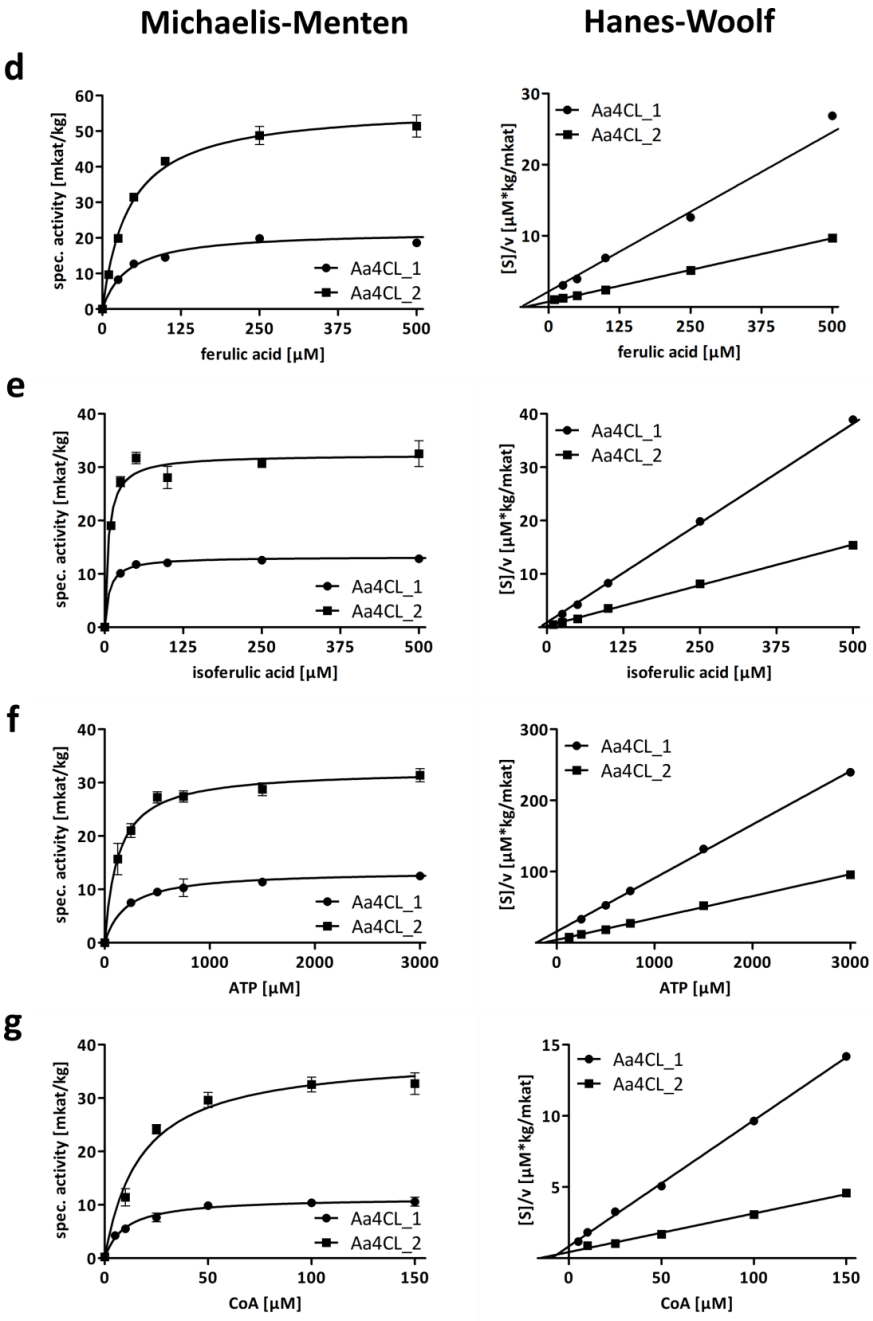


Figure 50 Kinetic data for Aa4CL_1 and Aa4CL_2 displayed by Michaelis-Menten and Hanes-Woolf diagrams. a Cinnamic acid; **b** 4-coumaric acid; **c** caffeic acid; **d** ferulic acid; **e** isoferulic acid; **f** ATP; **g** CoA. Graphs represent the mean values of one biological replicate with four technical replicates. Error bars represent the standard deviation.

Table 11 Kinetic values of Aa4CL_1 and Aa4CL_2. Data were analyzed with the help of Michaelis-Menten (MM), Lineweaver-Burk (LB) and Hanes-Woolf plots (HW); values represent the average of three biological replicates with three or four technical replicates (mean ± standard error).

		Aa4CL_1		Aa4CL_2	
		K_m [µM]	V_{max} [mkat/kg]	K_m [µM]	V_{max} [mkat/kg]
cinnamic acid	MM	218.2 ± 74.7	16.0 ± 2.9	216.7 ± 86.1	28.5 ± 2.9
	LB	169.3 ± 17.0	13.9 ± 0.7	133.4 ± 20.9	24.7 ± 5.5
	HW	225.5 ± 62.5	17.1 ± 2.3	247.0 ± 74.8	30.0 ± 3.6
4-coumaric acid	MM	12.8 ± 4.0	24.0 ± 3.1	11.8 ± 4.4	41.2 ± 4.8
	LB	16.3 ± 6.8	25.1 ± 3.9	16.7 ± 9.2	42.8 ± 3.3
	HW	12.0 ± 5.3	23.9 ± 3.5	15.0 ± 5.0	42.3 ± 4.5
caffeic acid	MM	15.5 ± 1.9	16.8 ± 2.2	20.5 ± 4.6	32.3 ± 5.8
	LB	14.7 ± 0.2	16.6 ± 2.2	15.1 ± 1.9	30.0 ± 4.7
	HW	19.8 ± 2.8	17.3 ± 2.5	28.9 ± 8.2	34.8 ± 6.8
ferulic acid	MM	61.4 ± 21.9	25.3 ± 3.9	65.3 ± 21.9	49.4 ± 8.1
	LB	71.1 ± 30.3	26.3 ± 4.7	62.8 ± 10.0	50.9 ± 11.0
	HW	68.0 ± 23.4	25.7 ± 3.9	69.4 ± 25.7	49.5 ± 7.7
isoferulic acid	MM	9.9 ± 1.8	13.1 ± 0.3	7.5 ± 1.8	28.0 ± 3.9
	LB	9.3 ± 1.3	13.1 ± 0.4	10.6 ± 4.2	29.0 ± 3.2
	HW	12.3 ± 0.4	13.3 ± 0.4	9.9 ± 0.9	28.6 ± 4.0
ATP	MM	227.1 ± 31.8	18.1 ± 4.8	151.1 ± 25.3	31.2 ± 1.0
	LB	247.1 ± 61.9	18.7 ± 5.5	171.4 ± 43.2	32.1 ± 0.3
	HW	244.0 ± 5.9	18.0 ± 4.3	174.0 ± 36.6	32.1 ± 0.4
CoA	MM	10.2 ± 0.2	15.9 ± 4.6	14.6 ± 3.2	35.9 ± 2.2
	LB	9.6 ± 0.8	15.7 ± 4.6	22.3 ± 7.5	40.3 ± 4.8
	HW	10.2 ± 0.5	15.9 ± 4.7	12.3 ± 4.1	34.7 ± 2.2

Table 12 Turnover number and catalytic efficiency of Aa4CL_1 and Aa4CL_2. Catalytic efficiency was calculated with the K_m values obtained from Michaelis-Menten diagrams; values represent the average of three biological replicates with three or four technical replicates. Average ± standard error.

	Aa4CL_1		Aa4CL_2	
	k_{cat} [1/s]	k_{cat}/K_m [1/s*mM]	k_{cat} [1/s]	k_{cat}/K_m [1/s*mM]
cinnamic acid	1.0 ± 0.2	4.6	1.8 ± 0.2	8.1
4-coumaric acid	1.5 ± 0.2	118.5	2.5 ± 0.3	215.1
caffeic acid	1.1 ± 0.1	68.3	2.0 ± 0.4	97.4
ferulic acid	1.6 ± 0.2	26.1	3.1 ± 0.5	46.7
isoferulic acid	0.8 ± 0.02	84.1	1.7 ± 0.2	231.6
ATP	1.1 ± 0.3	5.0	1.9 ± 0.1	12.7
CoA	1.0 ± 0.3	98.5	2.2 ± 0.1	151.6

4 4-Hydroxybenzoic acid CoA-ligase (4HBCL)

4.1 Identification of a cDNA encoding 4HBCL from *Anthoceros agrestis*

Looking for further *4CL* candidate genes, *Anthoceros agrestis* scaffold 35279 (Szövenyi, personal communication) was chosen. After PCR with primers directed against a partial sequence of 798 bp (Fig. 51a), 5´- and 3´-RACE-PCR was performed (Fig. 51b, c). The full open reading frame of *Aa35279* consisted of 1704 bp and could be translated into a 567 aa protein with a molecular weight of 61.84 kDa. The *Aa4HBCL* sequence was deposited in GenBank under the accession number MN922306. The BLASTp-algorithm still revealed the highest identity to 4-coumarate-CoA ligase-like 1 from several lower and higher plant species with 60 % on protein level (Table 13).

An alignment of different 4CLs and benzoate:CoA-ligases (BCL) showed a high conservation of amino acids in box I and box II (Fig. 52). The 12 amino acid residues proposed to function as the 4CL substrate specificity code (marked green) (Schneider et al. 2003) differed in some places (marked yellow) in Aa4HBCL and the other BCL

enzymes compared to 4CLs. The changes N288T, G380A and V390I or V390A were particularly noticeable.

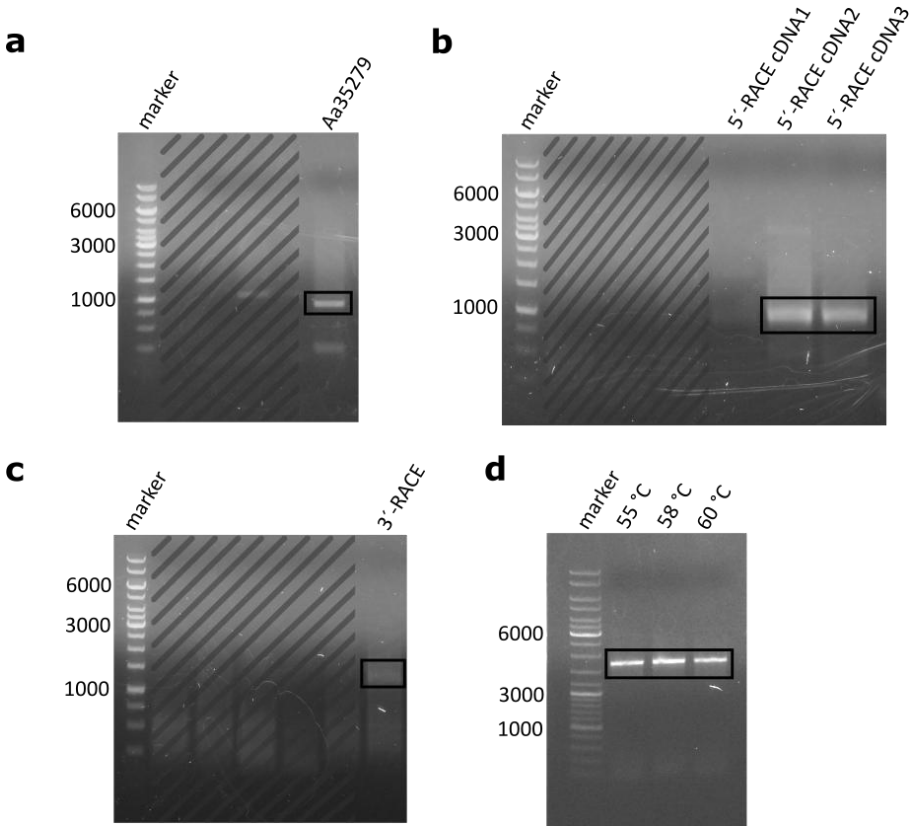


Figure 51 Amplification of *Aa4HBCL* by PCR. a Partial sequence; **b** 5´-RACE with different cDNAs as template; **c** 3´-RACE; **d** *Aa4HBCL* full-length sequence amplified at different annealing temperatures. Hatched areas belong to other experiments.

Table 13 Protein BLAST result of Aa4HBCL. https://blast.ncbi.nlm.nih.gov/Blast.cgi 17.10.2019

Description	Organism	Identity	E value	Accession
4-coumarate-CoA ligase-like 1	*Selaginella moellendorffii*	60 %	0.0	XP_002981856.1
4-coumarate-CoA ligase-like 1	*Physcomitrella patens*	55 %	0.0	XP_024370015.1
PREDICTED: 4-coumarate-CoA ligase-like 1 isoform X1	*Beta vulgaris subsp. vulgaris*	54 %	0.0	XP_010682749.1
PREDICTED: 4-coumarate-CoA ligase-like 1	*Cucumis sativus*	53 %	0.0	XP_004135516.2
4-coumarate-CoA ligase-like 1	*Brassica napus*	55 %	0.0	XP_013664375.1
4-coumarate-CoA ligase-like 1	*Rosa chinensis*	54 %	0.0	XP_024196337.1

Figure 52 Alignment of Aa4HBCL with different BCL and 4CL amino acid sequences. *Anthoceros*_4HBCL (this work, MN922306), *Arabidopsis*_BCL (Q9SS01), *Clarkia*_BCL (AEO52695.1), *Anthoceros*_4CL (this work, MN922305), *Arabidopsis*_4CL (Q42524), *Nicotiana*_4CL (O24145). Highly conserved amino acids are marked red, low conservation is marked blue. Typical conserved sequence motifs are marked by boxes. Amino acids from the 4CL substrate specificity code are highlighted green, and amino acids differing from this code are highlighted yellow (Schneider et al. 2003).

4.2 Expression of Aa4HBCL in *Escherichia coli*

As already Aa4CL also Aa4HBCL was expressed in *E. coli*. The full-length sequence was equipped with restriction sites for XhoI in the 5´-region and HindIII in the 3´-region. After amplification with PCR using four different cDNAs and different annealing temperatures (Fig. 51d), the fragment was first ligated into pDrive. After checking the sequence for mutations, it was inserted into pRSET C. On the vector were already 6xHis codons, which were attached *N*-terminally to the protein. After transformation of the *E. coli* strains BL21RIPL and SoluBL21, plasmid uptake and sequence correctness were checked. Then the protein was expressed for either 4.5 h at 37 °C (BL21RIPL) or 16 h at 25 °C (SoluBL21) in TB or LB media after induction with 1 mM IPTG. A crude protein extract of two transformants each was isolated and adjusted to a protein concentration of 2 mg/ml. 20 µl of these extracts were analysed by SDS-PAGE and immunoblotting with anti-His-antibodies. Unfortunately, the used glass panel had a crack, which caused the proteins to run crooked. The result was nonetheless evaluable. While there was a clear signal in all SoluBL21 samples, no protein could be detected in BL21RIPL protein extracts (Fig. 53a). Furthermore, the TB-media samples showed higher protein concentrations.

After protein purification (Fig. 53b) possible substrates for 4CL (4-coumaric, caffeic, cinnamic, ferulic and isoferulic acid) were tested first. Incubation with the substrates cinnamic acid, 4-coumaric acid, caffeic acid and isoferulic acid led to an increase of absorption compared to the empty vector control. Ferulic acid and sinapic acid led to no product formation. The production of 4-hydroxybenzoyl-CoA could also be directly measured photometrically. This substrate led to a significantly stronger increase of absorption compared to the cinnamic acid derivatives (Fig. 54). The overall specific activity using 500 µM substrate (5 mM ATP and $MgCl_2$ and 750 µM CoA) was 9.5 ± 1.2 mkat/kg for 4-hydroxybenzoic acid, 2.8 ± 0.3 mkat/kg for isoferulic acid, 1.9 ± 0.04 mkat/kg for 4-coumaric acid, 1.6 ± 0.03 mkat/kg for cinnamic acid and 0.4 ± 0.03 mkat/kg for caffeic acid.

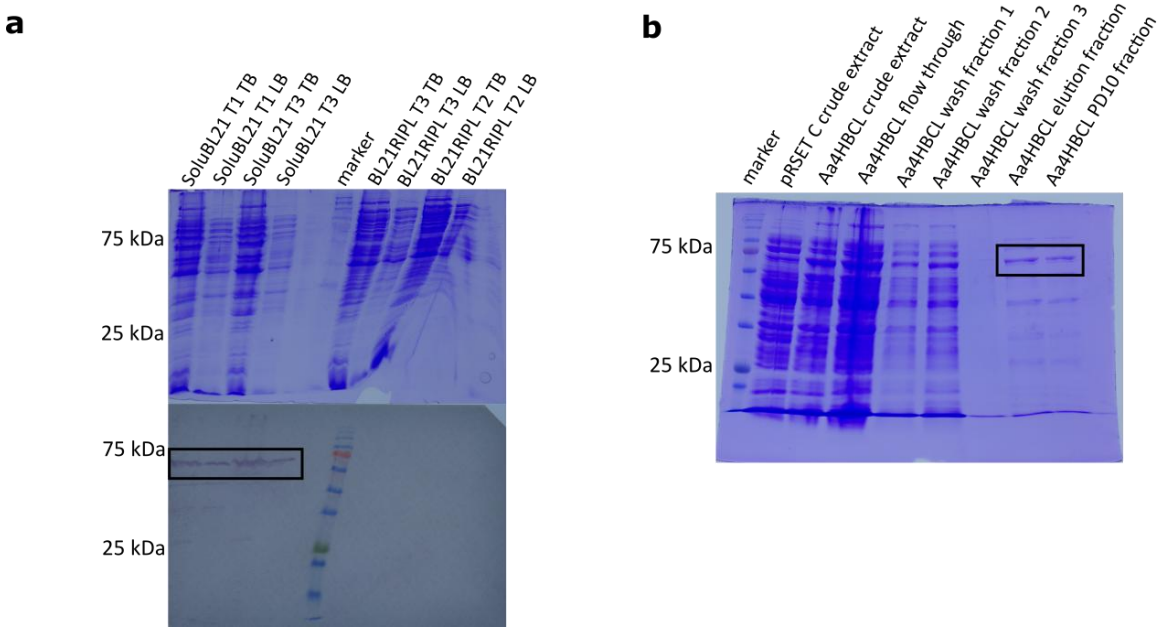
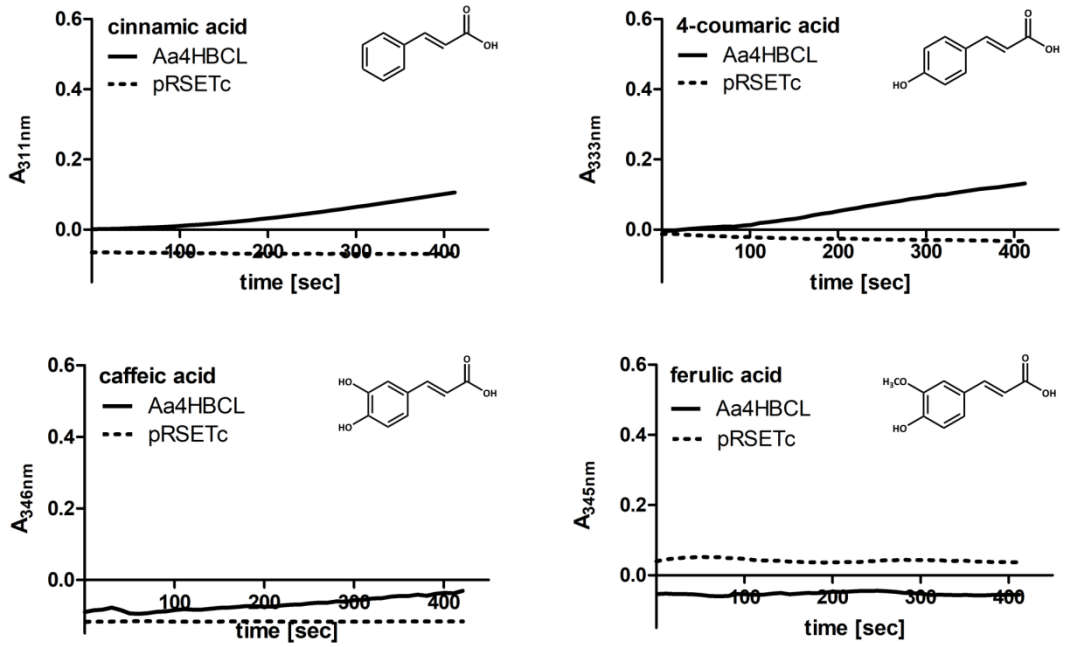


Figure 53 Expression of Aa4HBCL in *E. coli*. a SDS-PAGE and Western blot analysis of two different SoluBL21 or BL21 RIPL transformants of Aa4HBCL expressed in either TB or LB media induced with 1 mM IPTG. 40 µg crude protein was used for each sample; **b** SDS-PAGE of fractions obtained by metal chelate purification of Aa4HBCL together with a crude protein extract (100 µg protein) of the expressed empty vector pRSET C.



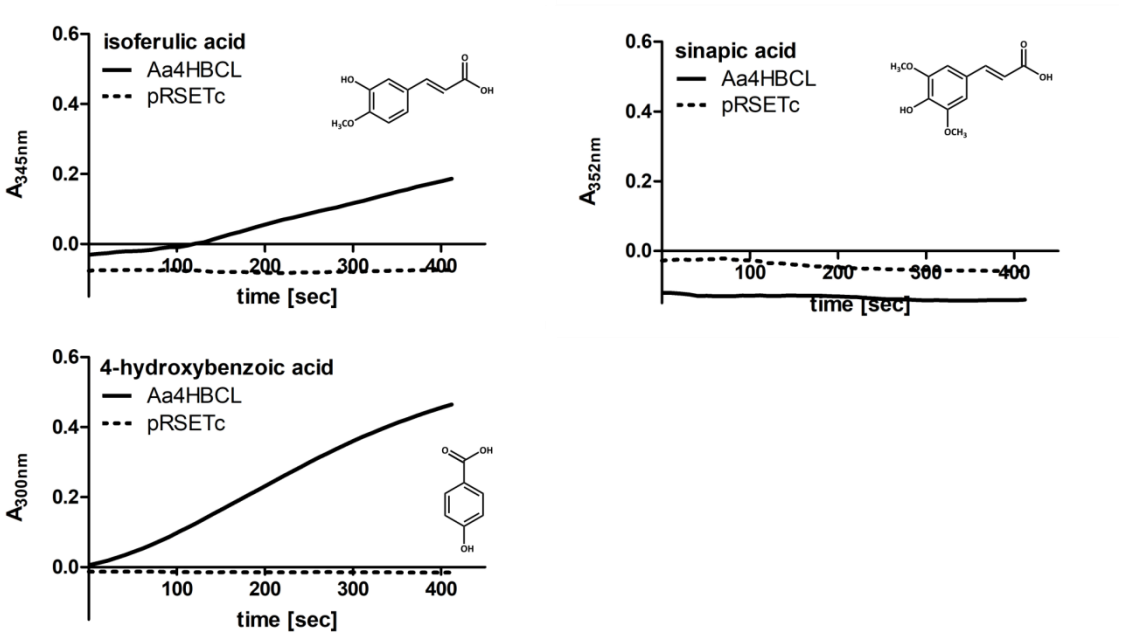


Figure 54 Activity of purified Aa4HBCL with different cinnamic acid derivatives and 4-hydroxybenzoic acid. Assays of Aa4HBCL (32 µg protein) were compared to assays with a crude protein extract of the expressed empty vector control pRSET C (100 µg protein). All assays contained 500 µM substrate, 5 mM ATP, 5 mM $MgCl_2$ and 750 µM CoA incubated at 40 °C. Aa4HBCL assays are displayed as a solid line and the empty vector control is shown as a dashed line.

Other benzoic acid derivatives were tested by an indirect measurement based on the conversion of ATP to AMP by the CoA-ligase. The residual ATP-concentration was determined by addition of the enzymes hexokinase and glucose-6-phosphate dehydrogenase (G6PDH) as well as glucose and NADP. After incubation of the CoA-ligase with the substrate the assay was stopped by heating. After cooling, hexokinase, G6PDH, glucose and NADP were added. The remaining ATP was then used by the hexokinase to produce glucose-6-phosphate. Glucose-6-phosphate was then oxidized by G6PDH and at the same time NADP is converted to NADPH. NADPH is then measured photometrically at 340 nm. Only the incubation with benzoic acid (BA) and 3-hydroxybenzoic acid (3HBA) led to a detectable reduction of the ATP concentration. Salicylic acid (2HBA) and all dihydroxylated substrates showed no difference to the negative control (methanol). It is unknown why 2,3-dihydroxybenzoic acid had a higher absorption compared to the negative control.

A decreasing ATP concentration over time was measured to compare the three accepted substrates (Fig. 55a). The percentual turnover of benzoic acid and 4-hydroxybenzoic acid was almost the same with 68 % and 67 % remaining ATP after 30 min incubation. 86 % of ATP was still present after 30 min using 3-hydroxybenzoic acid.

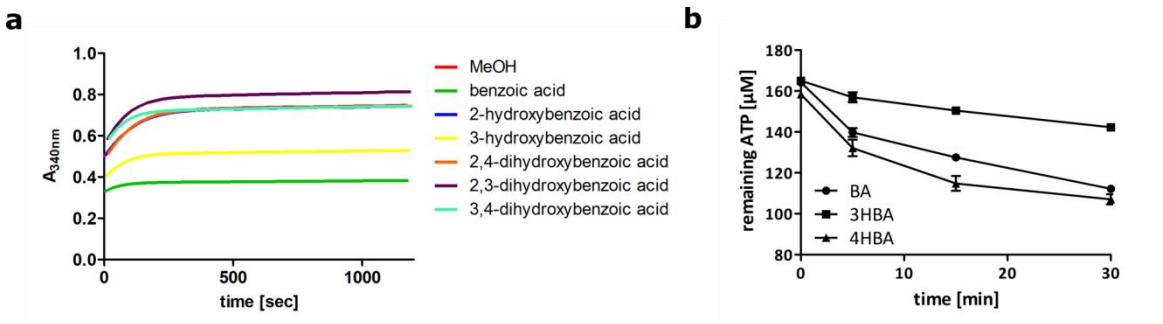


Figure 55 Activity of purified Aa4HBCL with different benzoic acid derivatives measured with the indirect method. a 14 µg Aa4HBCL was incubated with MeOH (negative control) or 500 µM substrate (dissolved in MeOH), 500 µM ATP, 500 µM $MgCl_2$ and 500 µM CoA for 1 h at 40 °C; **b** Determination of remaining ATP concentration after incubation of Aa4HBCL with benzoic acid (BA), 3-hydroxybenzoic acid (3HBA) and 4-hydroxybenzoic acid (4HBA) for 0, 5, 15 and 30 min at 45 °C (n= 3, SD).

To really ensure the substrate turnover of Aa4HBCL, the different benzoic acid derivatives (already tested by the indirect method) as well as 4-coumaric acid, caffeic acid and glycolic acid were analyzed again by LC-MS. At first, the LC-MS was calibrated for small masses (m/z 100-400). First assays were performed using 500 µM substrate and 70 µg purified protein (1.25 mM ATP and $MgCl_2$, 1 mM CoA) at 35 °C for 45 min. Heat-denatured protein (5 min, 95 °C) was used as negative control. The UV chromatogram at 254 nm displays that the substrate was completely converted after 45 min and only the doubly charged molecular ion [(M/2)-H] could be detected (Fig. 56).

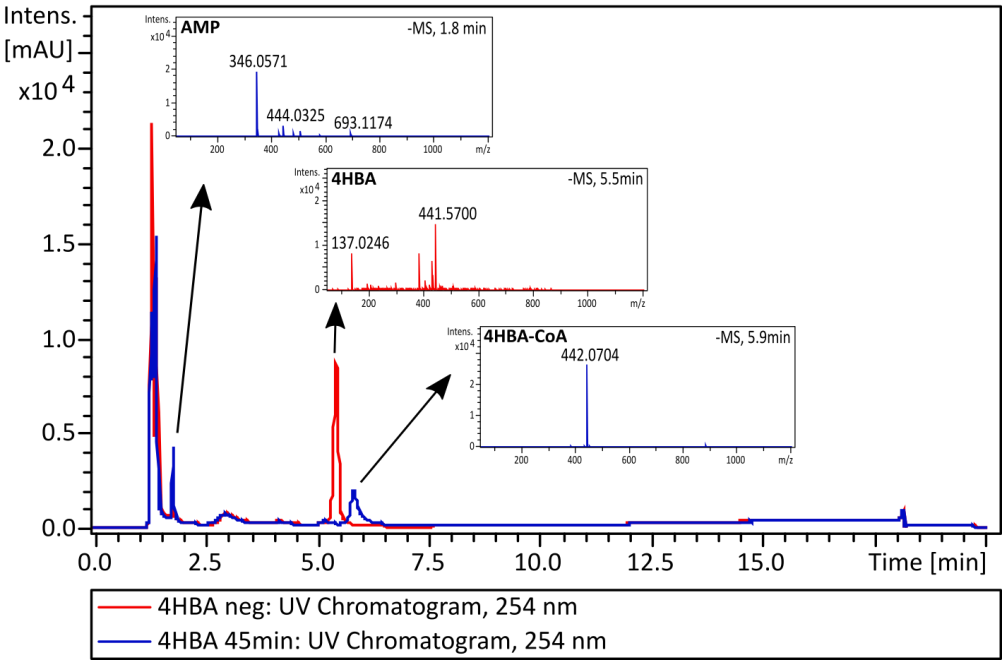
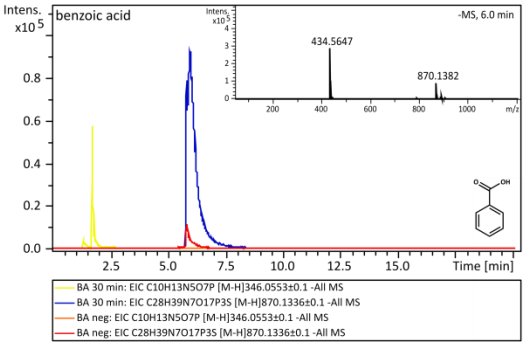
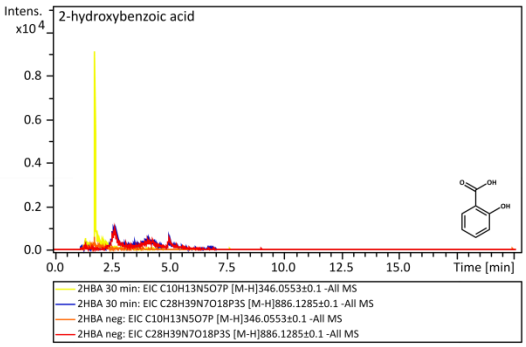
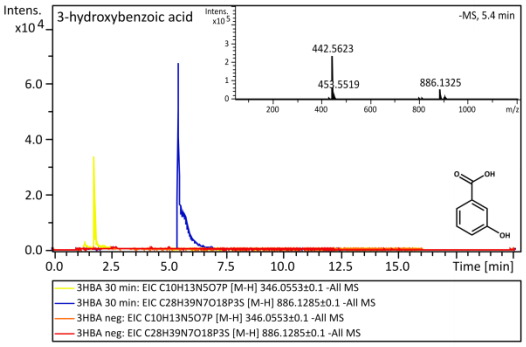


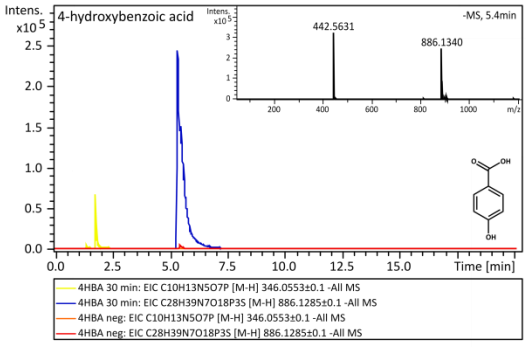
Figure 56 Reaction of Aa4HBCL with 4-hydroxybenzoic acid (4HBA) analyzed by LC-MS at 254 nm. The Aa4HBCL assay incubated for 45 min with 4HBA is displayed blue, the assays incubated with heat-denatured Aa4HBCL (5 min, 95 °C) is displayed red. The masses of the substrate 4HBA and the resulting products AMP and 4-hydroxybenzoyl-CoA (4HBA-CoA) are shown in the insets. LC-MS was calibrated for masses from m/z 100-400. Mass of 4HBA-CoA could only be detected as the doubly charged molecular ion [(M/2)-H].

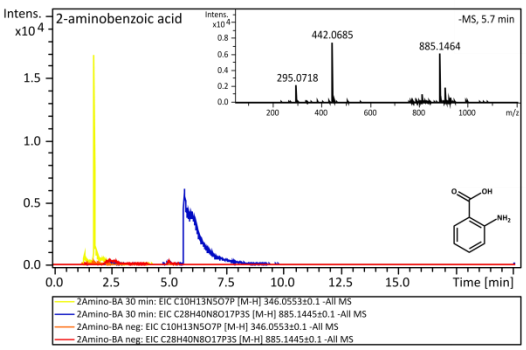
For easier interpretation of the analyses of the CoA-esters, the LC-MS was recalibrated by Rixa Kraut for higher masses between m/z 300-1000. Afterwards assays were repeated using 500 µM substrate and 28 µg purified Aa4HBCL incubated for 30 min at 40 °C. The same assays were performed with heat-denatured (10 min, 95 °C) as negative control. For all formed CoA-esters the [M-H] pseudo molecular ion as well as the doubly charged molecular ion [(M/2)-H] were observed (Fig. 57) (Beuerle and Pichersky 2002) and all of them had the expected mass. By using higher amounts of protein in a smaller volume, compared to the indirect assay, substrate conversion in combination with the formation of AMP was detectable for BA, 3HBA, 4HBA, 2-aminobenzoic acid, 3-aminobenzoic acid, 2,3-dihydroxybenzoic acid, 2,4-dihydroxybenzoic acid, 3,4-dihydroxybenzoic acid and 2-amino-3-hydroxybenzoic acid, 2-coumaric acid, 3-coumaric acid, 4-coumaric acid and caffeic acid. Only salicylic acid (2HBA), 3-aminosalicylic acid (2-hydroxy-3-aminobenzoic acid), vanillic acid (4-hydroxy-3-methoxybenzoic acid) and glycolic acid were not accepted. Instead only the formation of AMP was observed.

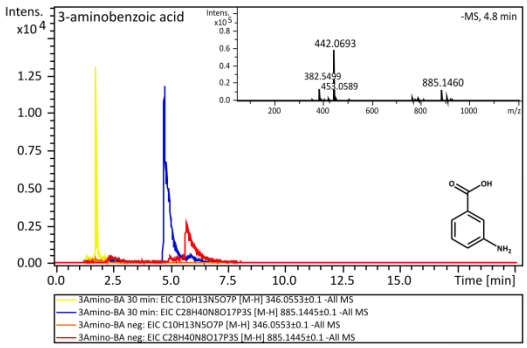

benzoic acid
-MS, 6.0 min
434.5647
870.1382
Intens.
Time [min]
BA 30 min: EIC C10H13N5O7P [M-H]346.0553±0.1 -All MS
BA 30 min: EIC C28H39N7O17P3S [M-H]870.1336±0.1 -All MS
BA neg: EIC C10H13N5O7P [M-H]346.0553±0.1 -All MS
BA neg: EIC C28H39N7O17P3S [M-H]870.1336±0.1 -All MS

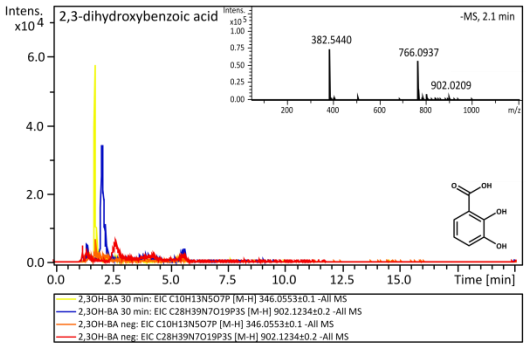

2-hydroxybenzoic acid
Intens.
Time [min]
2HBA 30 min: EIC C10H13N5O7P [M-H]346.0553±0.1 -All MS
2HBA 30 min: EIC C28H39N7O18P3S [M-H]886.1285±0.1 -All MS
2HBA neg: EIC C10H13N5O7P [M-H]346.0553±0.1 -All MS
2HBA neg: EIC C28H39N7O18P3S [M-H]886.1285±0.1 -All MS

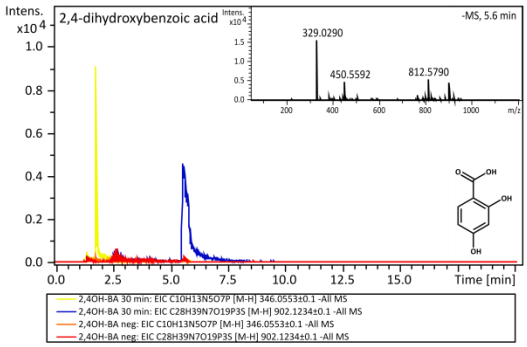

3-hydroxybenzoic acid
-MS, 5.4 min
442.5623
453.5519
886.1325
Intens.
Time [min]
3HBA 30 min: EIC C10H13N5O7P [M-H] 346.0553±0.1 -All MS
3HBA 30 min: EIC C28H39N7O18P3S [M-H] 886.1285±0.1 -All MS
3HBA neg: EIC C10H13N5O7P [M-H] 346.0553±0.1 -All MS
3HBA neg: EIC C28H39N7O18P3S [M-H] 886.1285±0.1 -All MS

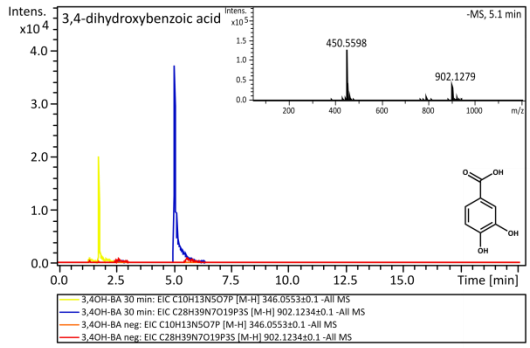

4-hydroxybenzoic acid
-MS, 5.4min
442.5631
886.1340
Intens.
Time [min]
4HBA 30 min: EIC C10H13N5O7P [M-H] 346.0553±0.1 -All MS
4HBA 30 min: EIC C28H39N7O18P3S [M-H] 886.1285±0.1 -All MS
4HBA neg: EIC C10H13N5O7P [M-H] 346.0553±0.1 -All MS
4HBA neg: EIC C28H39N7O18P3S [M-H] 886.1285±0.1 -All MS

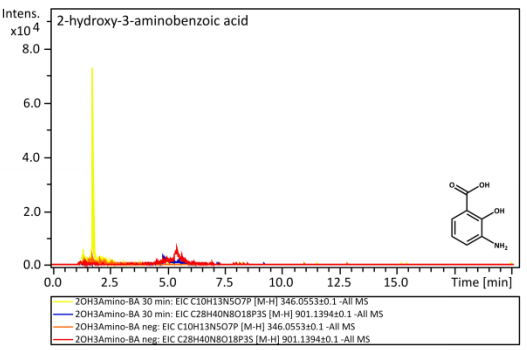

2-aminobenzoic acid
-MS, 5.7 min
442.0685
885.1464
295.0718
Intens.
Time [min]
2Amino-BA 30 min: EIC C10H13N5O7P [M-H] 346.0553±0.1 -All MS
2Amino-BA 30 min: EIC C28H40N8O17P3S [M-H] 885.1445±0.1 -All MS
2Amino-BA neg: EIC C10H13N5O7P [M-H] 346.0553±0.1 -All MS
2Amino-BA neg: EIC C28H40N8O17P3S [M-H] 885.1445±0.1 -All MS


3-aminobenzoic acid
-MS, 4.8 min
442.0693
382.5499
453.0589
885.1460
Intens.
Time [min]
3Amino-BA 30 min: EIC C10H13N5O7P [M-H] 346.0553±0.1 -All MS
3Amino-BA 30 min: EIC C28H40N8O17P3S [M-H] 885.1445±0.1 -All MS
3Amino-BA neg: EIC C10H13N5O7P [M-H] 346.0553±0.1 -All MS
3Amino-BA neg: EIC C28H40N8O17P3S [M-H] 885.1445±0.1 -All MS


2,3-dihydroxybenzoic acid
-MS, 2.1 min
382.5440
766.0937
902.0209
Intens.
Time [min]
2,3OH-BA 30 min: EIC C10H13N5O7P [M-H] 346.0553±0.1 -All MS
2,3OH-BA 30 min: EIC C28H39N7O19P3S [M-H] 902.1234±0.2 -All MS
2,3OH-BA neg: EIC C10H13N5O7P [M-H] 346.0553±0.1 -All MS
2,3OH-BA neg: EIC C28H39N7O19P3S [M-H] 902.1234±0.2 -All MS


2,4-dihydroxybenzoic acid
-MS, 5.6 min
329.0290
450.5592
812.5790
Intens.
Time [min]
2,4OH-BA 30 min: EIC C10H13N5O7P [M-H] 346.0553±0.1 -All MS
2,4OH-BA 30 min: EIC C28H39N7O19P3S [M-H] 902.1234±0.1 -All MS
2,4OH-BA neg: EIC C10H13N5O7P [M-H] 346.0553±0.1 -All MS
2,4OH-BA neg: EIC C28H39N7O19P3S [M-H] 902.1234±0.1 -All MS


3,4-dihydroxybenzoic acid
-MS, 5.1 min
450.5598
902.1279
Intens.
Time [min]
3,4OH-BA 30 min: EIC C10H13N5O7P [M-H] 346.0553±0.1 -All MS
3,4OH-BA 30 min: EIC C28H39N7O19P3S [M-H] 902.1234±0.1 -All MS
3,4OH-BA neg: EIC C10H13N5O7P [M-H] 346.0553±0.1 -All MS
3,4OH-BA neg: EIC C28H39N7O19P3S [M-H] 902.1234±0.1 -All MS


2-hydroxy-3-aminobenzoic acid
Intens.
Time [min]
2OH3Amino-BA 30 min: EIC C10H13N5O7P [M-H] 346.0553±0.1 -All MS
2OH3Amino-BA 30 min: EIC C28H40N8O18P3S [M-H] 901.1394±0.1 -All MS
2OH3Amino-BA neg: EIC C10H13N5O7P [M-H] 346.0553±0.1 -All MS
2OH3Amino-BA neg: EIC C28H40N8O18P3S [M-H] 901.1394±0.1 -All MS

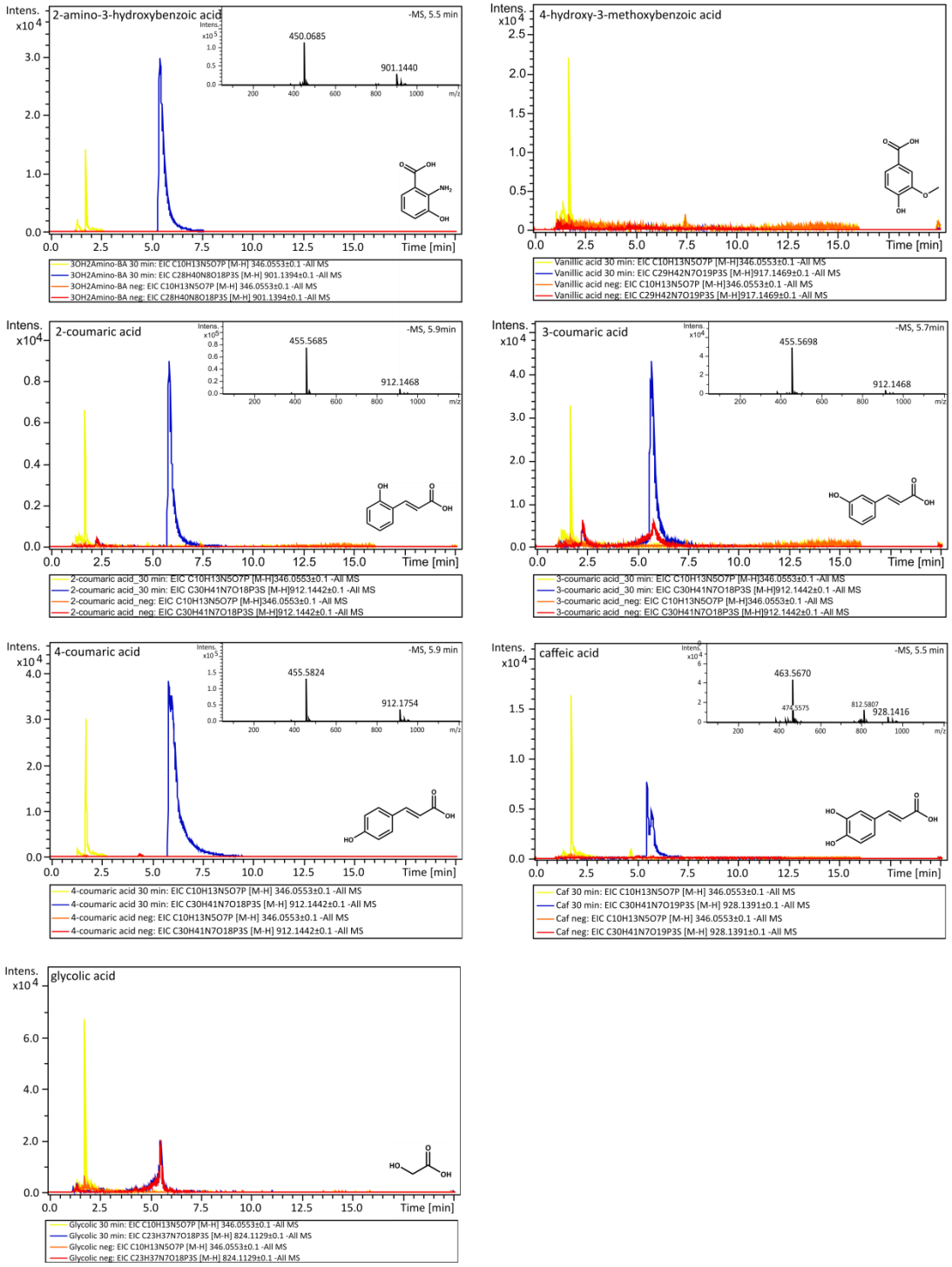


Figure 57 LC-MS results of enzyme assays of purified Aa4HBCL with different substrates. 28 µg Aa4HBCL was incubated with 500 µM substrate, 1.25 mM ATP and 1 mM CoA for 30 min at 40 °C. Heat-denatured protein (10 min 95 °C) was used as a negative control. The chromatograms show the EIC of the expected products AMP (m/z 346.0553 ± 0.1) in yellow for Aa4HBCL and orange for the negative control. The corresponding CoA-ester is displayed blue for Aa4HBCL or red for the negative control. The exact mass of the resulting CoA-ester is illustrated in each chromatogram. For all produced CoA-esters the [M-H] pseudo molecular ion as well as the doubly charged molecular ion [M/2]-H was observed.

The main substrates, measured photometrically in direct or indirect assays, were all compared to 4HBA to determine the relative substrate preference (Fig. 58). With 90 ± 8 % BA was accepted almost as good as 4HBA. 3HBA and isoferulic acid were converted with a relative activity of 35 % ± 7 % and 30 % ± 3 %. With 20 % or less 4-coumaric acid (20 % ± 0.4 %), cinnamic acid (17 % ± 0.3 %) and caffeic acid (5 % ± 0.3 %) were activated.

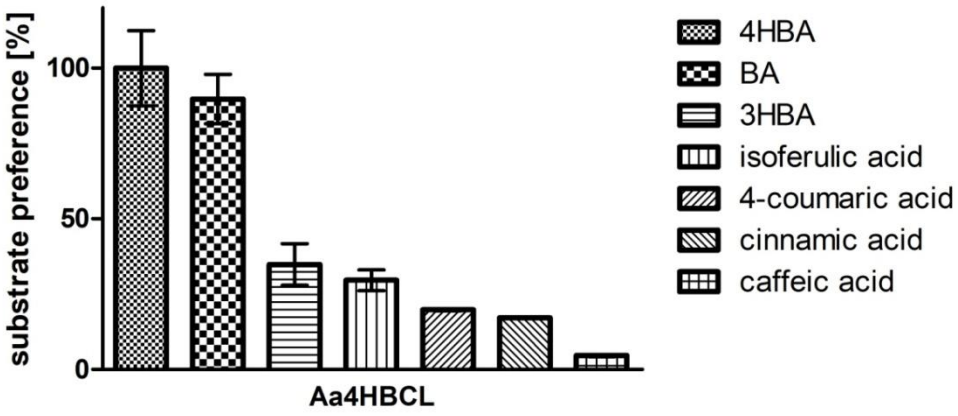


Figure 58 Relative substrate preference of Aa4HBCL compared to 4-hydroxybenzoic acid. 4-Hydroxybenzoic acid (4HBA), benzoic acid (BA) und 3-hydroxybenzoic acid (3HBA) were measured by the indirect method. Specific activity of 4HBA, isoferulic acid, 4-coumaric acid, cinnamic acid and caffeic acid were determined by direct photometric measurement. Each bar represents the mean average of three technical replicates and the error bars represent the standard deviation.

4.2.1 Temperature and pH-optimum of Aa4HBCL

Production of 4-hydroxybenzoyl-CoA was measured at 300 nm with 4HBA as substrate. Assays with increasing temperatures showed a temperature optimum at around 50 °C (Fig. 59). Higher temperatures than 55 °C were not determined, because the water bath connected to the photometer could not be heated to higher temperatures. 500 mM KPi buffer was used to determine the pH-optimum. The pH-optimum was at around pH 7.3 using a buffer with pH 7.5 (Fig. 59). Kinetic parameters were measured at 45 °C in 100 mM KPi-buffer pH 7.5.

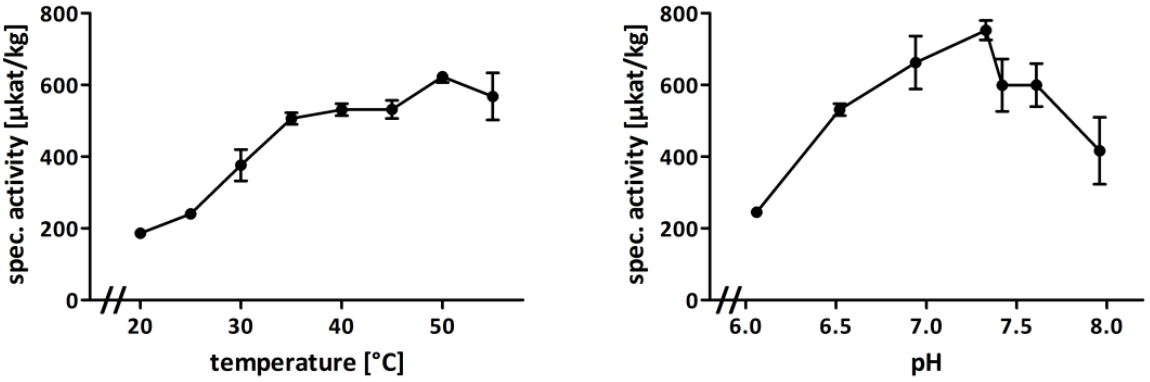


Figure 59 Temperature (left) and pH-optimum (right) of Aa4HBCL (n=3, SD).

4.2.2 Kinetic parameters of Aa4HBCL

Kinetic data were evaluated using Michaelis-Menten, Lineweaver-Burk (not shown) and Hanes-Woolf plots. The resulting substrate saturation curve of 4HBA led to an apparent K_m-value of 664.2 ± 1.5 µM and a V_{max} of 16.5 ± 0.7 mkat/kg. The apparent K_m-value for ATP was 1.2 ± 0.1 mM and V_{max} was 23.4 ± 4.7 mkat/kg. Different concentrations of CoA led to a substrate saturation curve with an apparent K_m-value of 247.8 ± 19.9 µM and a V_{max} of 32.5 ± 0.8 mkat/kg (Fig. 60; Table 14).

Aa4HBCL displayed a turnover number of 1-2 per second for all substrates and a catalytic efficiency of approximately 1.5 1/(s*mM) for 4HBA and ATP and 8.7 for CoA (Table 14). The molecular mass including the additional *N*-terminal amino acids from the expression vector of Aa4HBCL was calculated to be 66.46 kDa.

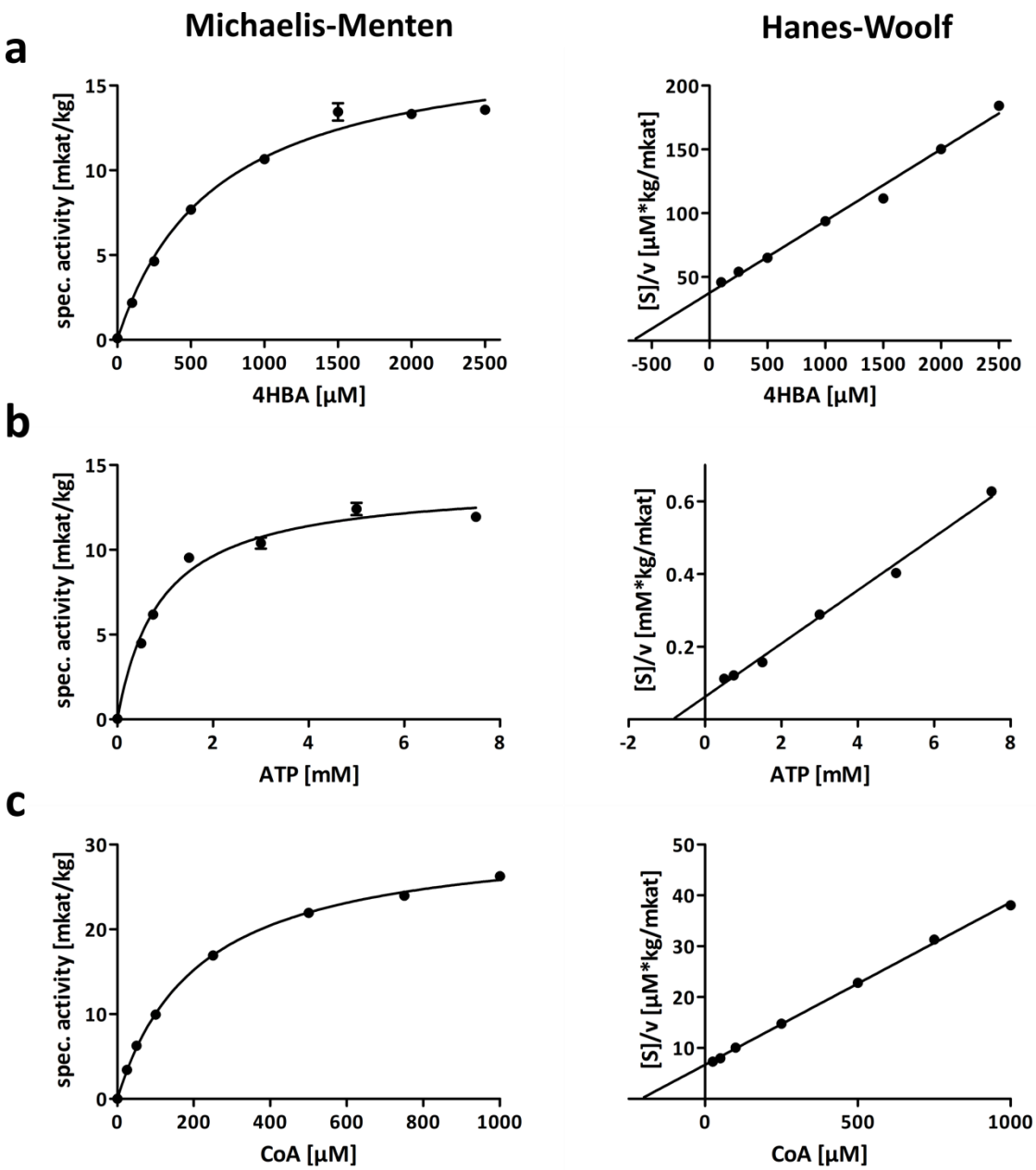


Figure 60 Kinetic data for Aa4HBCL displayed by Michaelis-Menten and Hanes-Woolf diagrams. **a** 4-Hydroxybenzoic acid (4HBA); **b** ATP; **c** CoA. Graphs represent the mean values of one biological replicate with three technical replicates. Error bars represent the standard deviation.

Table 14 Kinetic values of Aa4HBCL for the main substrate 4-hydroxybenzoic acid (4HBA) and the two cosubstrates ATP and CoA. Values were calculated with the help of Michaelis-Menten (MM), Lineweaver-Burk (LB) and Hanes-Woolf plots (HW); Measurements were reproduced three times with protein from three independent isolations. For each substrate or cosubstrate concentration three technical replicates were analyzed (mean ± SE).

		Aa4HBCL			
		K_m [µM]	V_{max} [mkat/kg]	k_{cat} [1/s]	k_{cat}/K_m [1/s*mM]
4HBA	MM	664.2 ± 1.5	16.5 ± 0.7	1.1 ± 0.1	1.7
	LB	765.8 ± 7.0	17.3 ± 0.8	1.2 ± 0.1	1.5
	HW	657.1 ± 5.5	16.4 ± 0.7	1.1 ± 0.1	1.7
ATP	MM	1175.3 ± 135.2	23.4 ± 4.7	1.6 ± 0.3	1.3
	LB	1392.8 ± 132.4	24.7 ± 4.9	1.6 ± 0.3	1.2
	HW	1098.0 ± 127.1	22.9 ± 4.6	1.5 ± 0.3	1.4
CoA	MM	247.8 ± 19.9	32.5 ± 0.8	2.2 ± 0.1	8.7
	LB	247.5 ± 25.3	30.4 ± 1.5	2.0 ± 0.1	8.2
	HW	248.0 ± 20.6	32.5 ± 0.5	2.2 ± 0.1	8.7

The attempt to repeat the K_m-value for 4HBA with the indirect method, using hexokinase and glucose-6-phosphate dehydrogenase, failed, since ATP was saturated and the concentration did not decrease perceptibly to see a difference between the added 4HBA concentrations. For this reason the K_m-values of benzoic acid and 3-hydroxybenzoic acid could not be determined.

5 Aa20832

5.1 Identification and isolation of Aa20832

Another potential 4CL candidate was *Aa20832*. Therefore a partial sequence of 1020 bp was amplified (Fig. 61a) and 5´- and 3´-RACE-PCR was performed (Fig. 61b, c). The full open reading frame revealed two possible start codons and the sequences consisted of 1914 bp and 1779 bp. The encoded protein had a size of 637 aa (69.07 kDa)/ 592 aa (64.29 kDa). Protein BLAST revealed the highest identities (around 40 %) mostly to

uncharacterized 4CLs from the moss *Physcomitrella patens* (XP_024366617.1) or the lycophyte *Selaginella moellendorfii* (XP_024514739.1). Furthermore, the sequence had a high identity (42 %) to the OPC-8:0 CoA-ligase 1 (involved in jasmonic acid biosynthesis) from *Dorcoceras hygrometricum* (KZV20099.1) (Table 15).

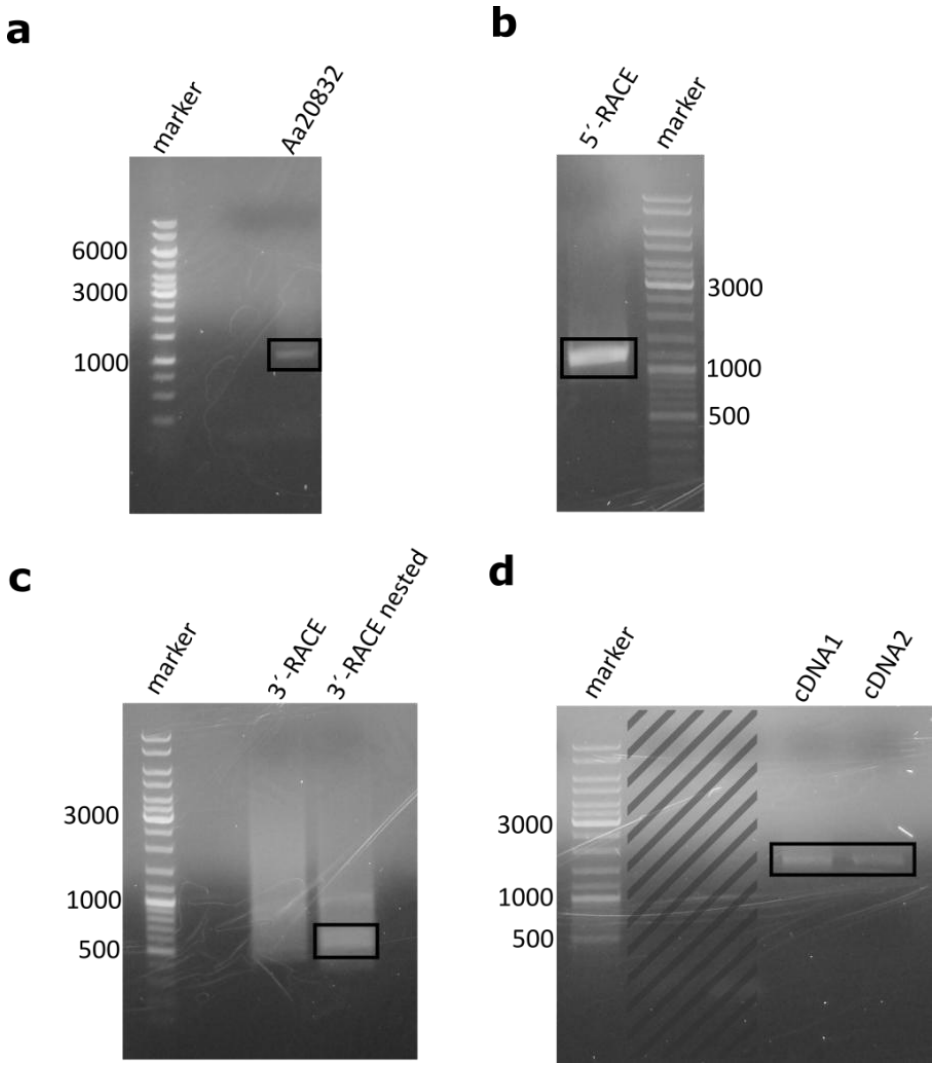


Figure 61 Amplification of *Aa20832* by PCR. a Partial sequence; **b** 5´-RACE; **c** 3´-RACE and nested 3´-RACE; **d** *Aa20832* full-length sequence with two different cDNAs as template. Hatched areas belong to other experiments

Table 15 Protein BLAST result of Aa20832. https://blast.ncbi.nlm.nih.gov/Blast.cgi 17.10.2019

Description	Organism	Identity	E value	Accession
4-coumarate-CoA ligase-like 7	*Physcomitrella patens*	42 %	1e-140	XP_024366617.1
4-coumarate-CoA ligase-like 7	*Selaginella moellendorffii*	43 %	2e-138	XP_024514739.1
4-coumarate-CoA ligase-like 7	*Vigna unguiculata*	41 %	9e-138	XP_027908923.1
Acyl-CoA synthetase	*Handroanthus impetiginosus*	40 %	2e-135	PIN09597.1
4-coumarate CoA ligase	*Boehmeria nivea*	40 %	9e-135	AWJ58443.1
OPC-8:0 CoA ligase1 isoform 1	*Dorcoceras hygrometricum*	42 %	5e-134	KZV20099.1

An alignment of the three potential amino acid sequences of 4CLs from *Anthoceros* (Aa4CL, Aa4HBCL and Aa20832) together with the two sequences from *Physcomitrella* and *Selaginella* from the BLASTp search, showed a high conservation of box I. In box II Aa20832 had a higher similarity with the sequences of *Physcomitrella* and *Selaginella* and differed from the other two *Anthoceros* CoA-ligases (Fig. 62). Also some of the twelve amino acid residues proposed to function as the 4CL substrate specificity code (marked in green) differed (marked in yellow) compared to Aa4CL (Schneider et al. 2003). Since the *N*-terminus of Aa20832 was much longer than in all other aligned CoA-ligases, it was assumed that the shorter sequence was the expressed one. Analysis of the subcellular localization (Plant-mPLoc and predictprotein.org) predicted the protein to be located in peroxisomes. This was also supported by the presence of a peroxisomal signal sequence (PTS1) (marked in orange) at the *C*-terminus (Reumann et al. 2016).

Figure 62 Alignment of the Aa20832 amino acid sequence with Aa4CL, Aa4HBCL (all this work) and two 4CL-like sequences from *Selaginella moellendorffii* (XP_024514739.1) and *Physcomitrella patens* (XP_024366617.1). Highly conserved amino acids are marked red, low conservation is marked blue. Typical conserved sequence motifs are marked by boxes. Amino acids from the 4CL substrate specificity code are highlighted green, and amino acids differing from this code are highlighted yellow (Schneider et al. 2003). The two possible start codons are highlighted pink. The peroxisomal signal sequence in the *C*-terminus is highlighted orange (Reumann et al. 2016).

5.2 Expression of Aa20832 in *Escherichia coli*

The two full-length sequences were equipped with restriction sites for BamHI (5´) and EcoRI (3´). After PCR-amplification (Fig. 61d), the sequence was first ligated into pDrive and checked by sequencing. Unfortunately, only the shorter sequence could be

amplified. Then both *Aa20832* and pRSET C were digested with BamHI and EcoRI and the sequence was integrated into pRSET C. A polyhistidine-tag for purification, already present on the plasmid, was attached to the *N*-terminus. The *E. coli* strains SoluBL21 and BL21 RIPL were transformed with the plasmid and plasmid uptake and sequence correctness were checked. First expression attempts were performed using two transformants of every strain and two different media (LB and TB). Only SoluBL21 showed formation of protein at the expected size after Western blot analysis with anti-His-antibody (Fig. 63a). At protein level there was no difference between LB and TB media, but the cells grew much better in TB (around 2.6 g) than in LB (around 0.7 g). Therefore, SoluBL21 and TB medium were used for further experiments. First activity assays with hydroxycinnamic acid derivatives and hydroxybenzoic acid derivatives were performed with crude protein extract. None of them showed any conversion in photometric assays and LC-MS analysis (Fig. 64 and 65). The protein was then purified by Ni-chelate chromatography (Fig. 63b) and the tests were repeated. Nonetheless, no activity could be observed.

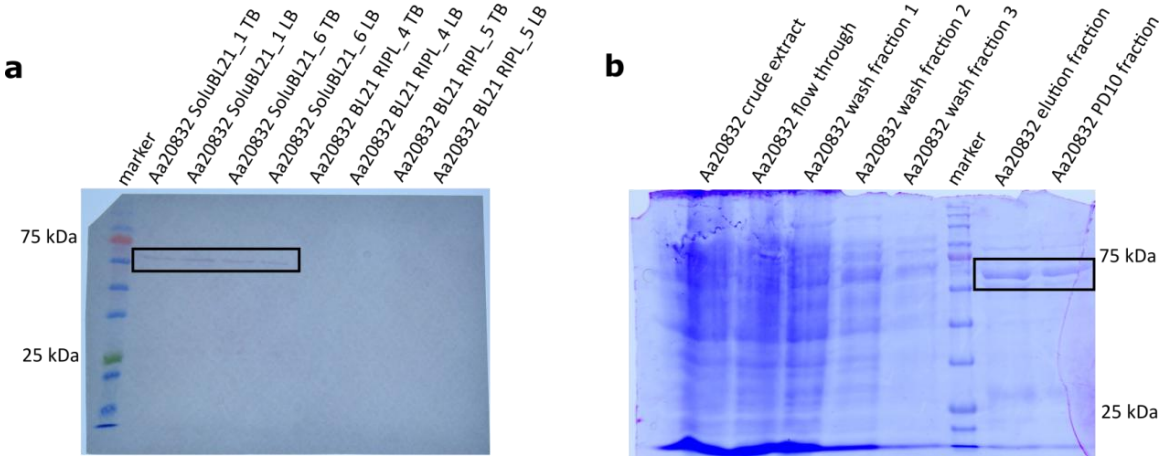


Figure 63 Expression of Aa20832 in *E. coli*. a Western blot analysis of two different SoluBL21 or BL21 RIPL transformants of Aa20832 expressed in either TB or LB media induced with 1 mM IPTG. 40 µg of crude protein was used for each sample; **b** SDS-PAGE of fractions obtained by metal chelate purification of Aa20832.

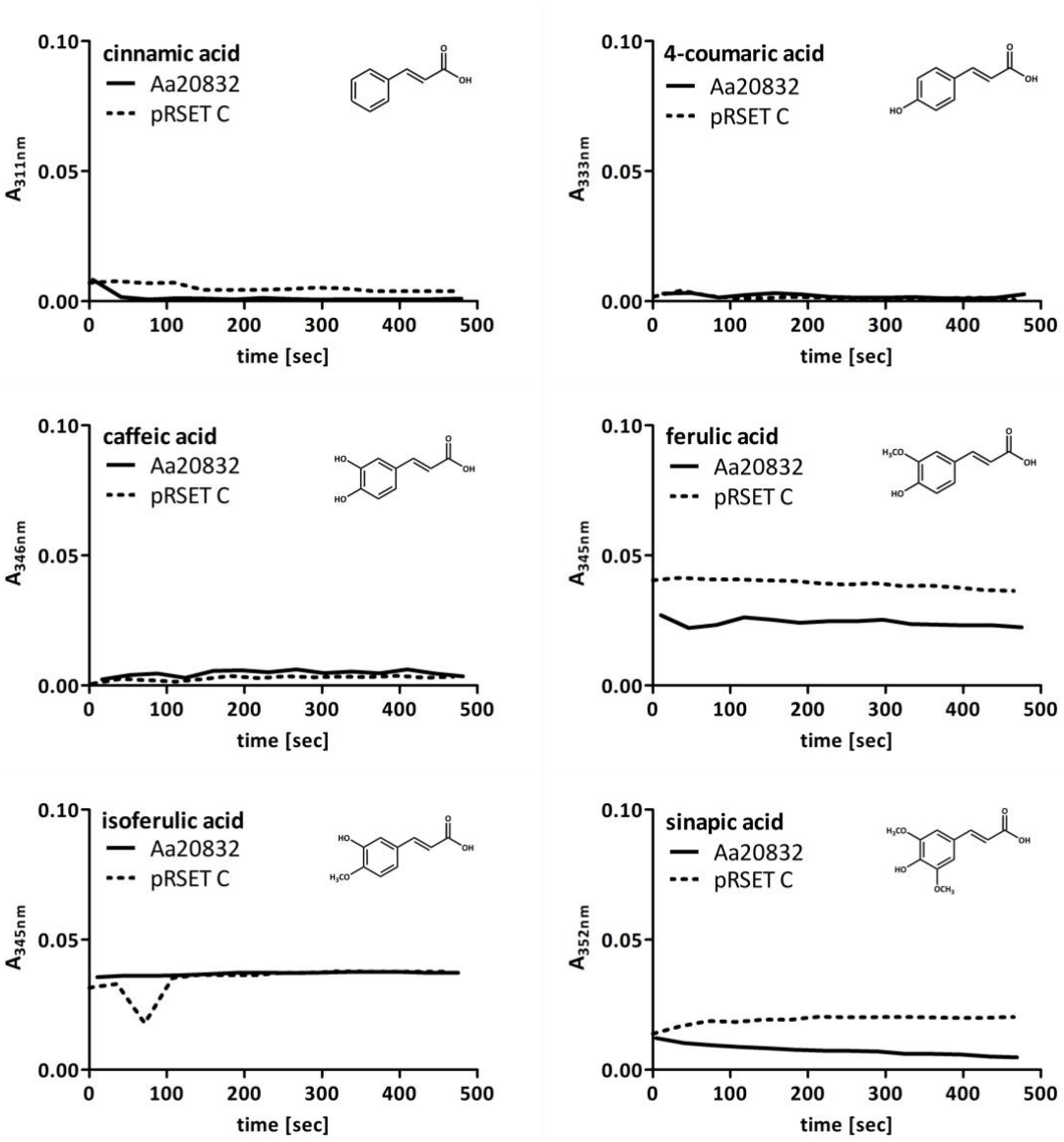


Figure 64 Activity of Aa20832 with different cinnamic acid derivatives (500 µM). Assays of Aa20832 crude protein extract (40 µg protein) were compared to assays with a crude protein extract of the expressed empty vector control pRSET C (100 µg protein).

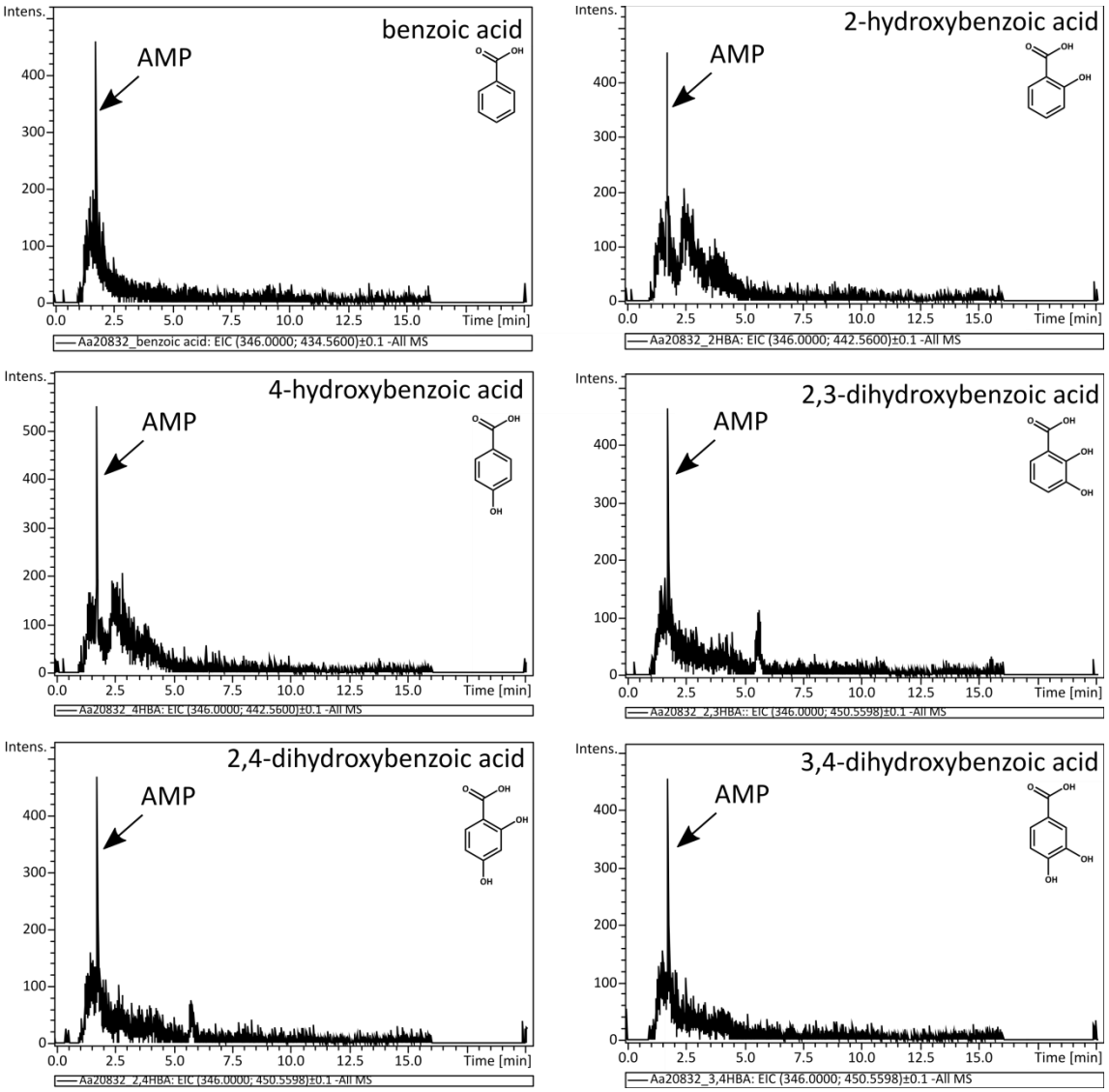


Figure 65 LC-MS analysis of enzyme assays of Aa20832 with different benzoic acid derivatives. 100 µg crude protein extract of Aa20832 was incubated with 500 µM benzoic acid, 2-hydroxybenzoic acid, 4-hydroxybenzoic acid, 2,3-dihydroxybenzoic acid, 2,4-dihydroxybenzoic or 3,4-dihydroxybenzoic acid for 1 h at 35 °C. The chromatograms show the EIC of the expected products AMP (m/z 346 ± 0.1 [M-H]) and the corresponding CoA-ester (doubly charged molecular ion [M/2]-H) of benzoic acid (m/z 434.56 ± 0.1), monohydroxylated benzoic acid (m/z 442.56 ± 0.1) or dihydroxylated benzoic acid (m/z 450.55 ± 0.1).

In a BLASTp search, the protein alignment also showed a high identity to an OPC-8:0 CoA-ligase. As OPC derivatives could not be procured, other substrates (stearic acid and palmitic acid) of *Arabidopsis thaliana* OPC-8:CoA-ligase (Kienow et al. 2008) were tested with the purified enzyme. This was done measuring the ATP reduction in a spectrophotometric assay with hexokinase and glucose-6-phosphate dehydrogenase

(Chapter II.3.8.3.2). MeOH, sinapic acid and salicylic acid were used as negative controls. All used substrates showed the same absorption pattern (Fig. 66). Although ATP reduction was highest with palmitic acid (52.9 % ATP left after 240 min) and stearic acid (63.3 % ATP left after 240 min), also the ATP concentration of assays with salicylic acid (64.9 % left after 240 min), sinapic acid (78.6 % left after 240 min) and MeOH (71 .3 % left after 240 min) decreased.

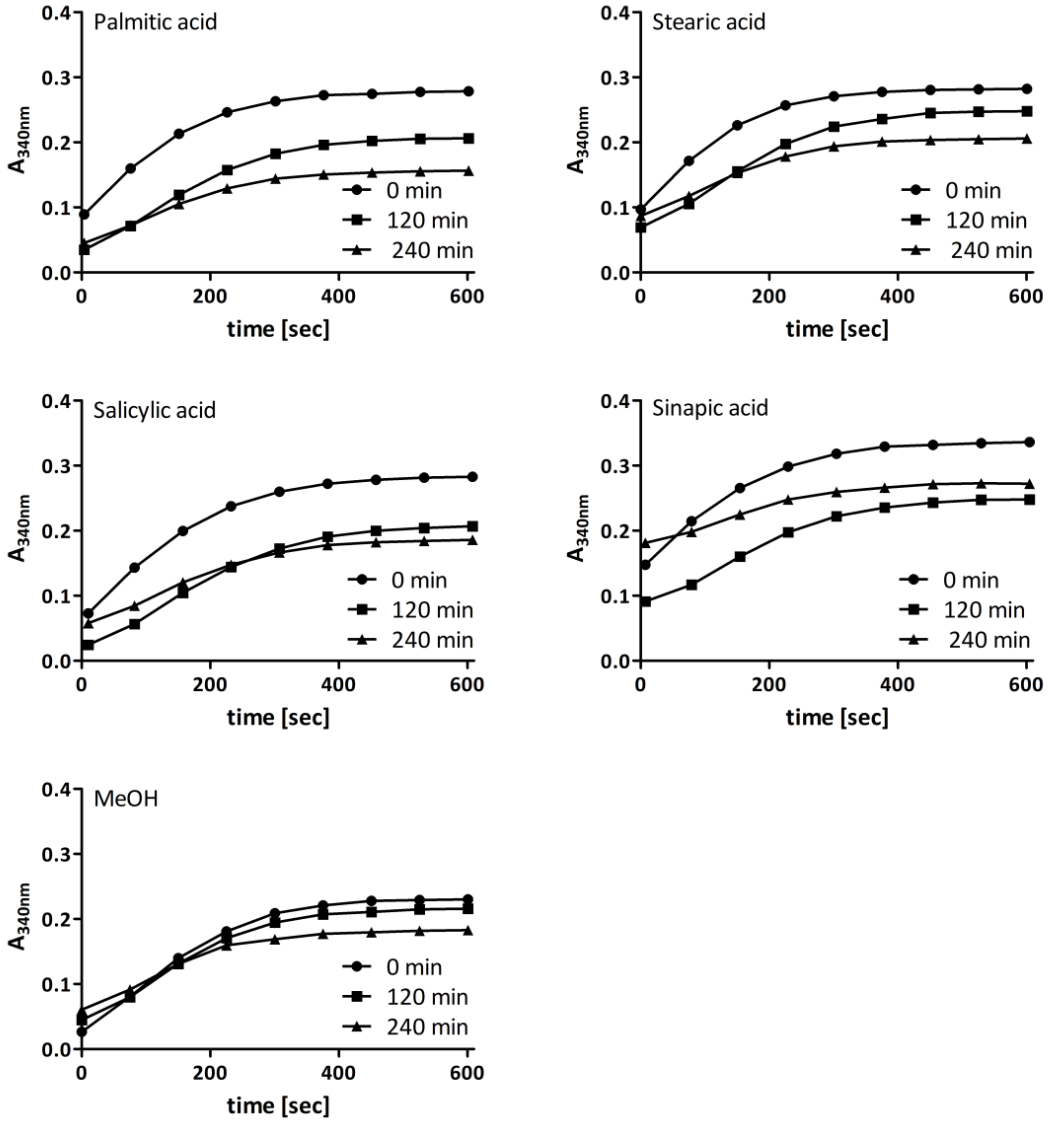


Figure 66 Activity of purified Aa20832 with different substrates measured with the indirect method. Assays were incubated at 25 °C for 0, 120 and 240 min and contained 30 µg purified protein, 200 µM substrate, 200 µM ATP, 200 µM $MgCl_2$, 200 µM CoA and 1 mM DTT. Methanol in the same volume as the added substrate was used as a negative control.

Given that this was no clear result, assays were performed with omitting different assay components using MeOH as a substrate (Fig. 67). This experiment revealed that the purified enzyme in combination with ATP and $MgCl_2$ was responsible for ATP reduction. Leaving out these components did not change absorption after 120 min, while it could still be observed without CoA and MeOH.

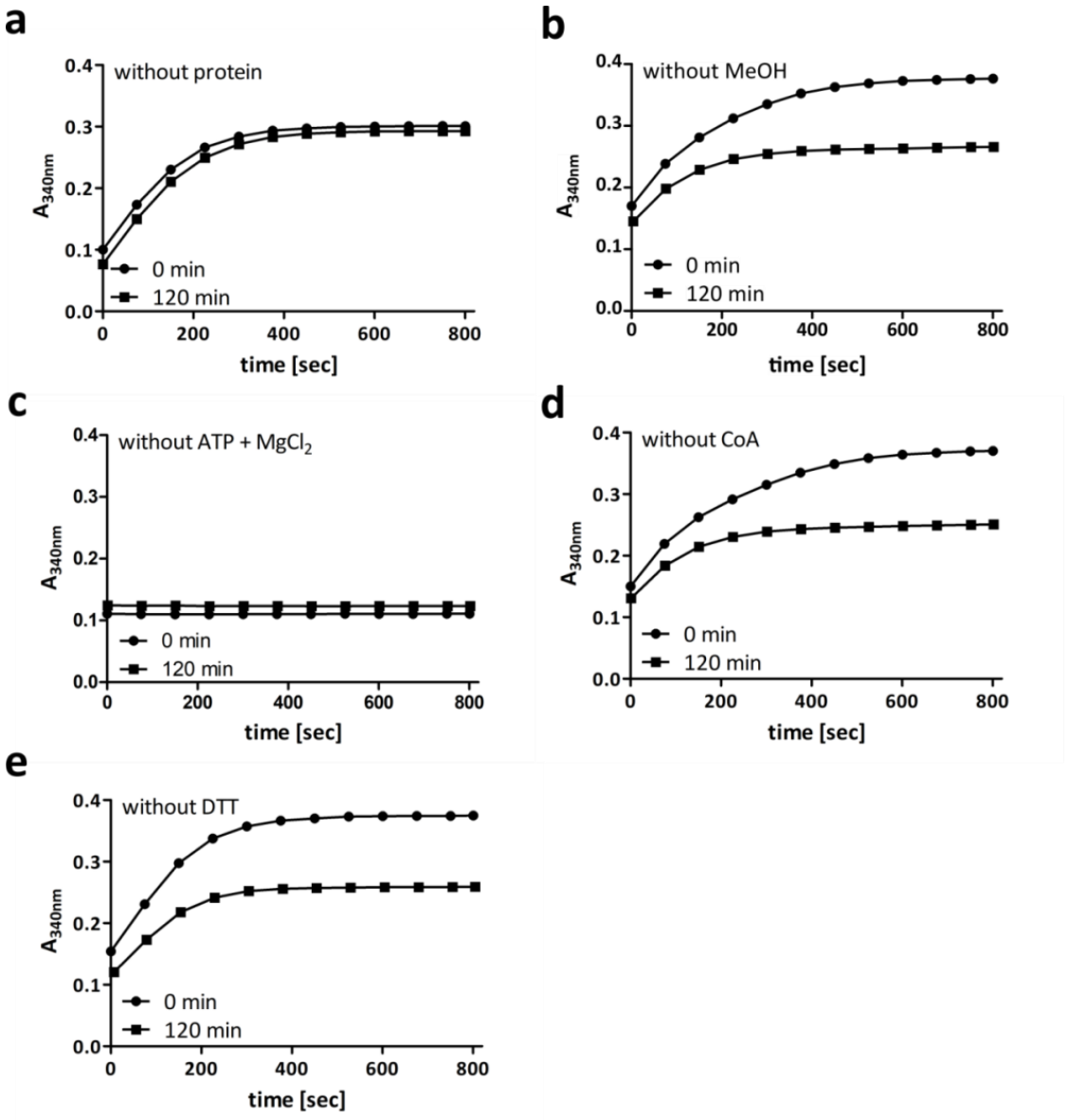


Figure 67 Omission tests of Aa20832 measured with the indirect method. In each test one component of the standard composition was omitted. MeOH was used instead of a substrate. In general, assays consisted of 27 µg purified Aa20832, 200 µM ATP, 200 µM $MgCl_2$, 200 µM CoA and 1 mM DTT. After incubation at 45 °C for 0 or 120 min, assays were heat-inactivated. The missing component was replaced by buffer. **a** without protein; **b** without methanol; **c** without ATP and $MgCl_2$; **d** without CoA; **e** without DTT.

The result gave rise to the suspicion that Aa20832 might still have ATPase function without any substrate activation. To confirm this, the protein was incubated with a high ATP concentration for 0 and 120 min. Buffer incubated with the same ATP concentration served as a negative control. Figure 68 shows the LC-MS analysis of this experiment. Incubation of the protein with ATP produced 4.5-fold AMP compared to the buffer control. This could also be observed for the formation of ADP (twofold). Based on these results measurement of ATP reduction for the determination of enzyme activity was unsuitable for this enzyme. The same effect in a lower degree was also observed for Aa4HBCL and Aa4CL.

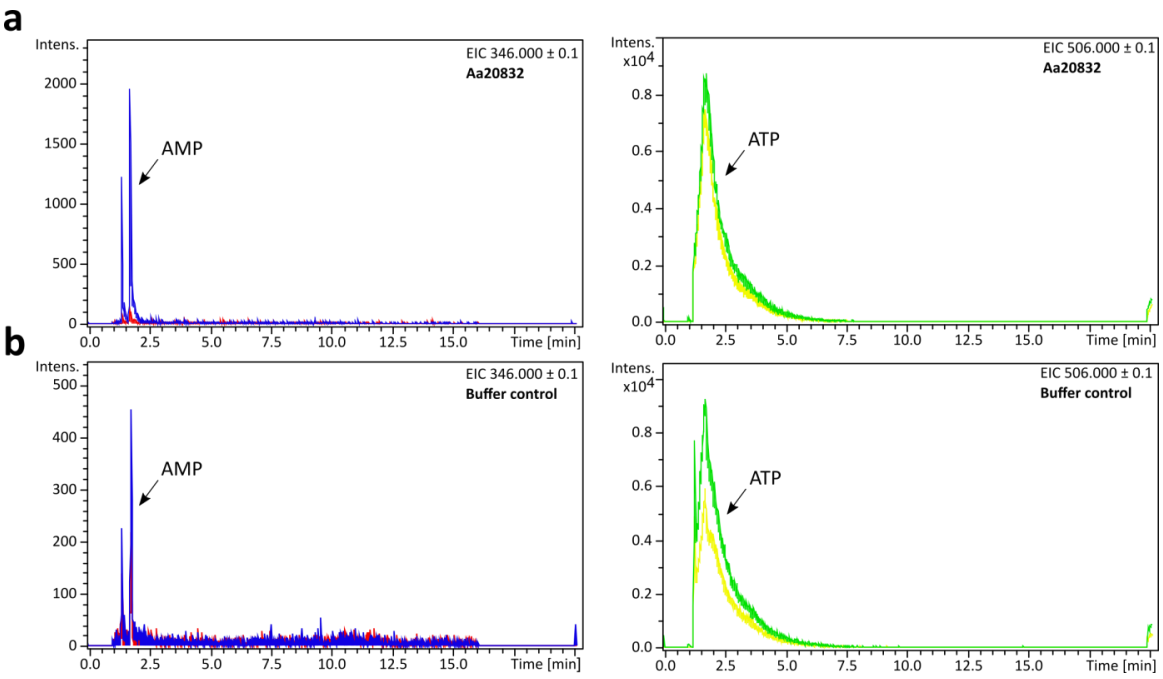


Figure 68 LC-MS analysis of Aa20832 incubated with ATP and $MgCl_2$. Either protein (27 µg) or buffer were incubated with 2.5 mM ATP and $MgCl_2$ at 25 °C for 0 or 120 min. Chromatograms show the EIC of AMP (m/z 346.000 ± 0.1 [M-H]) on the left side or ATP (506.000 ± 0.1 [M-H]) on the right side. Assays incubated for 0 min are displayed in red (AMP-chromatogram) or yellow (ATP-chromatogram). Assays incubated for 120 min are displayed in blue (AMP-chromatogram) or green (ATP-chromatogram). **a** Aa20832; **b** buffer control.

5.3 Potential membrane anchor of Aa20832

On the search for possible explanations for the strong ATPase activity, a secondary structure prediction for Aa20832 was performed (Phyre[2] and predictprotein.org). Both

predicted at least one transmembrane helix (TM-helix) for the protein sequence. While aa 129-137 and aa 282-297 were predicted as TM-helices by Phyre[2] (Fig. 69), predictprotein.org only predicted one TM-helix between aa 284 and 298.

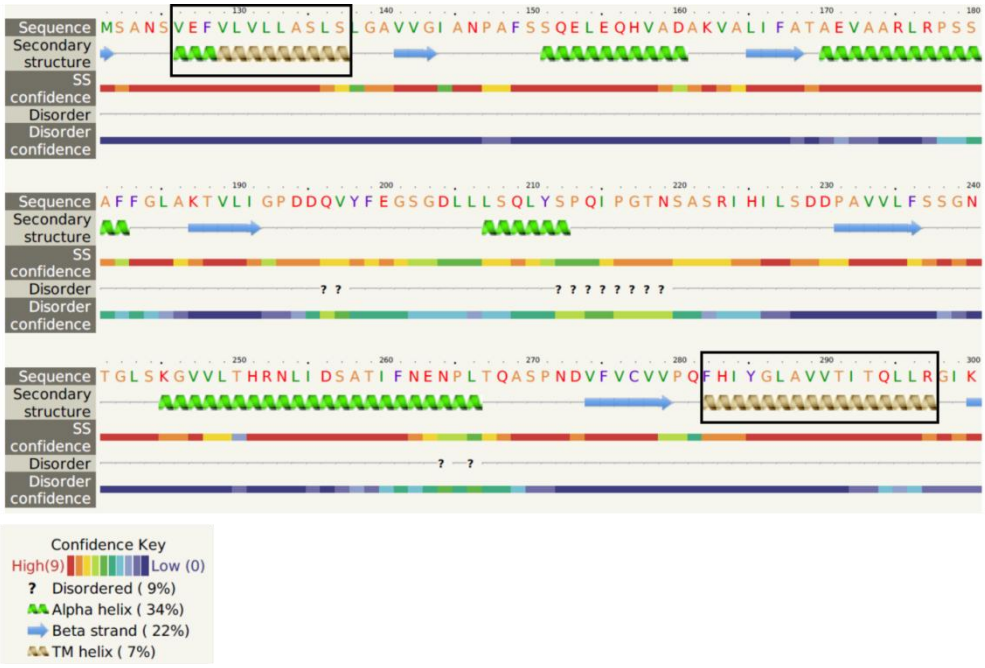


Figure 69 Secondary structure prediction of Aa20832 provided by Phyre[2]. Potential transmembrane helices are shown in boxes.

For this reason it was considered, that Aa20832 was partly denatured by purification with Ni-NTA and thus only ATPase activity was left. Since an ATPase function without the correct substrate was also observed in Aa4HBCL and Aa4CL, it could also be possible, that the suitable substrate has not yet been found.

For time reasons, experiments with this enzyme were not continued. It would be necessary to find a different method to measure CoA-ligase activity and test other substrates. A membrane fraction of *E. coli* should to be prepared and checked on Western blots.

6 Aa19917

Another potential 4CL candidate sequence was *Anthoceros agrestis* scaffold 19917 (Szövenyi, personal communication). Since PCR with the initially designed primers and

cDNA did not work, two additional forward primers and one additional reverse primer were configured. Nevertheless none of the primer combinations led to an amplification of the desired fragment (Fig. 70a). Changing the annealing temperature, isolation of new mRNA and using a different polymerase did not show amplification. The approximate sizes of the expected fragments are listed in Table 16. Amplification could only be observed after using gDNA, kindly provided by Tobias Busch (Fig. 70b). However, all fragments were longer (around 500 bp) than expected. This led to the conclusion, that there might be some introns. Furthermore, this gene might not be expressed and therefore no sequence amplification is possible using cDNA. Since RACE-PCR is only working with mRNA, this project was not carried on further.

Table 16 Primer combinations and expected fragment size after amplification of partial sequence of *Aa19917*.

Primer combination	Expected fragment size [bp]
Aa19917_fwd1 + Aa19917_rev1	1000
Aa19917_fwd2 + Aa19917_rev1	1200
Aa19917_fwd3 + Aa19917_rev1	900
Aa19917_fwd1 + Aa19917_rev2	900
Aa19917_fwd2 + Aa19917_rev2	1000
Aa19917_fwd3 + Aa19917_rev2	750

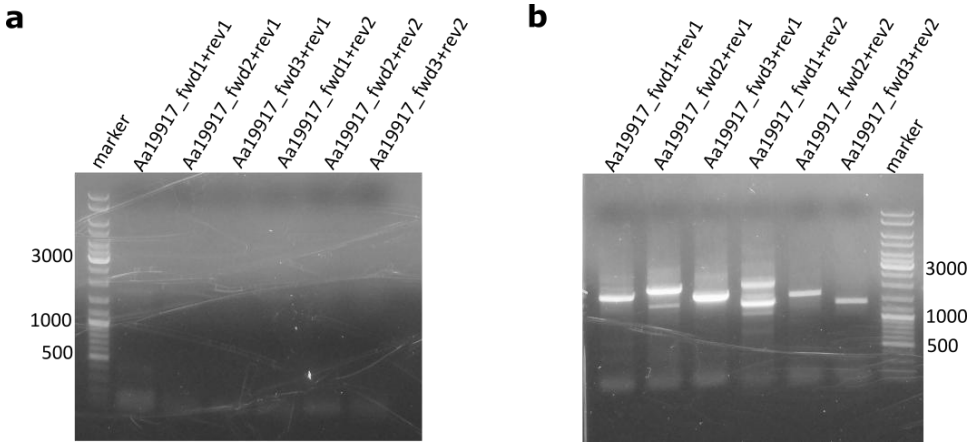


Figure 70 Amplification of a partial sequence of *Aa19917* by PCR using different primer combinations. a PCR with cDNA as template; **b** PCR with gDNA as template.

7 CYP98

7.1 Identification of a cDNA encoding CYP98 from *Anthoceros agrestis*

Scaffold sequence 26091 (Szövenyi et al. 2015) was identified as a potential nucleotide sequence of *AaCYP98*. A partial sequence of 1050 bp was obtained after PCR amplification. 5´- and 3´-RACE-PCR completed the full-length sequence (Fig. 71). The nucleotide sequence consisted of 1518 bp. This sequence was translated to a 56.58 kDa protein with 507 amino acid residues. The *AaCYP98* sequence was deposited in GenBank under the accession number MT119883. On amino acid level AaCYP98 was analysed by BLASTp (Table 17) and had the highest identity (71 %) to a cytochrome P450 98A3-like from *Physcomitrella patens*.

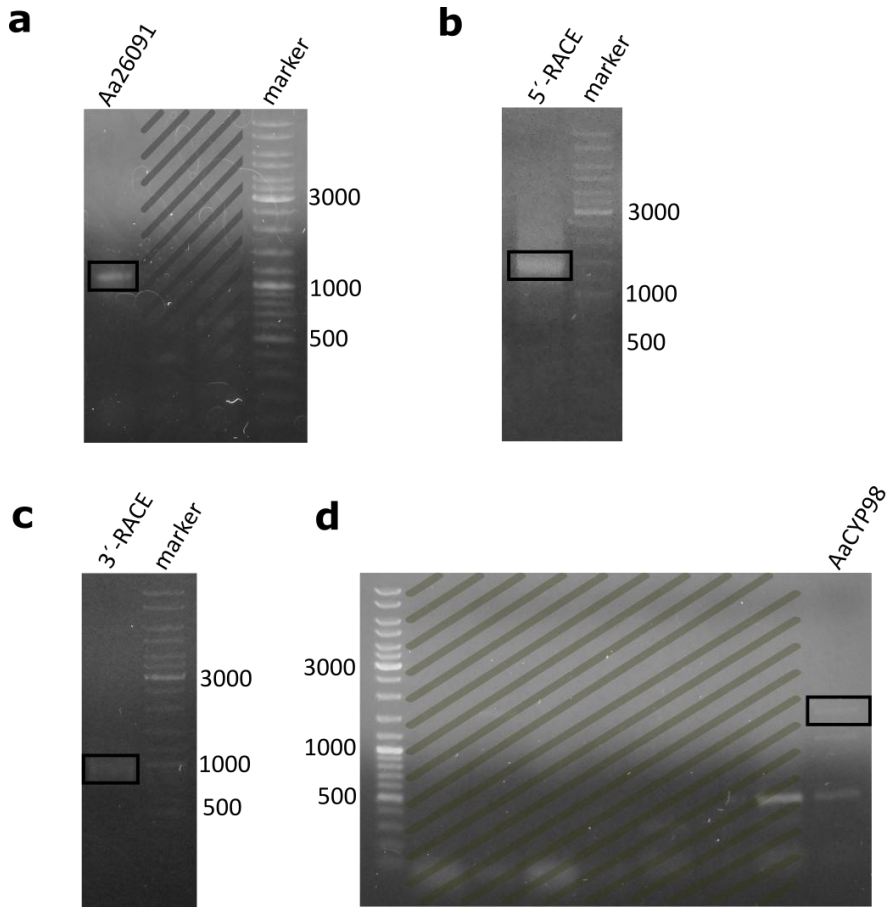


Figure 71 Amplification of *AaCYP98* by PCR. a Partial sequence; **b** 5´-RACE-PCR; **c** 3´-RACE-PCR; **d** *AaCYP98* full-length sequence. Hatched areas belong to other experiments.

Table 17 Protein BLAST result of AaCYP98. https://blast.ncbi.nlm.nih.gov/Blast.cgi 02.12.2019

Description	Organism	Identity	E value	Accession
cytochrome P450 98A3-like	*Physcomitrella patens*	71 %	0.0	XP_024360823.1
unknown	*Picea sitchensis*	67 %	0.0	ABR18076.1
***p*-coumarate 3-hydroxylase**	*Pinus taeda*	67 %	0.0	AAL47685.1
cytochrome P450 98A2	*Morella rubra*	67 %	0.0	KAB1222998.1
***p*-coumarate 3-hydroxylase**	*Cunninghamia lanceolata*	69 %	0.0	AFX98060.1
cytochrome P450 98A2	*Prosopis alba*	67 %	0.0	XP_028766773.1

Figure 72 displays an alignment of different CYP98 amino acid sequences. Characteristic motifs of canonical cytochrome P450 enzymes like the proline-rich region (Werck-Reichhart et al. 2002), the O_2-binding region, the PERF motif and the heme-binding cysteine motif (Schuler 1996; Mizutani et al. 1997; Chapple 1998) have been identified. Phylogenetic analysis (Fig. 73) also supports the BLASTp result of AaCYP98 being closely related to the CYP98 from *Physcomitrella patens*. As expected, the two *Arabidopsis* CYP98s, the only ones producing tricoumaroyl or triferuloyl spermidines, were located on a separate branch. Interestingly, all CYP98 genes of the monocots *Triticum aestivum* and *Panicum virgatum* are located separately, although it has been proven biochemically, that they accept coumaroylshikimate and coumaroylquinate (Morant et al. 2007; Escamilla-Trevino et al. 2014).

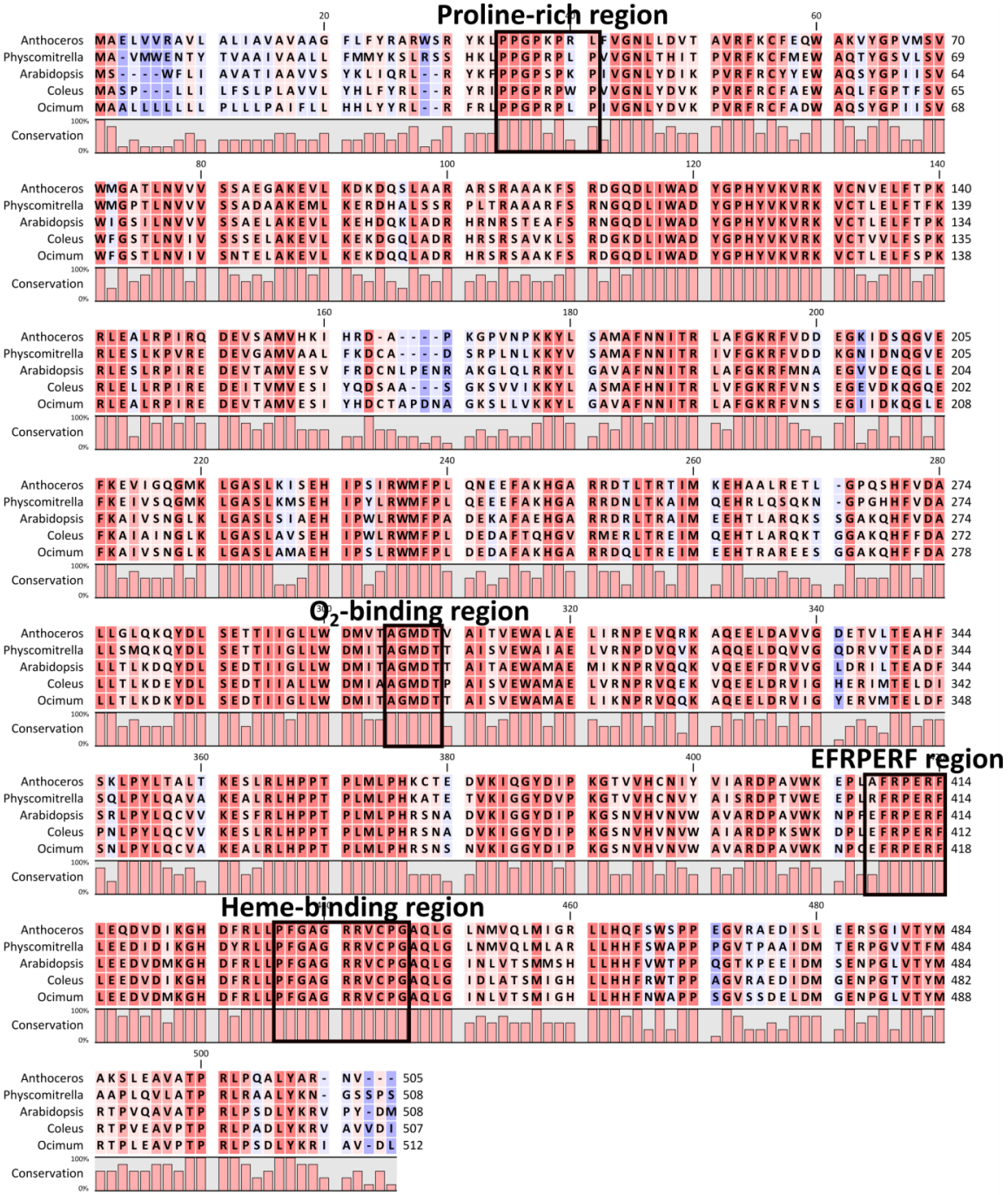


Figure 72 Alignment of CYP98 amino acid sequences. *Anthoceros agrestis* (this work), *Physcomitrella patens* (Pp3c22_19010), *Arabidopsis thaliana* (P850337), *Coleus blumei* (CAD20576) and *Ocimum basilicum* (AY082611). Highly conserved amino acids are marked red, low conservation is marked blue. Typical conserved sequence motifs are marked by boxes.

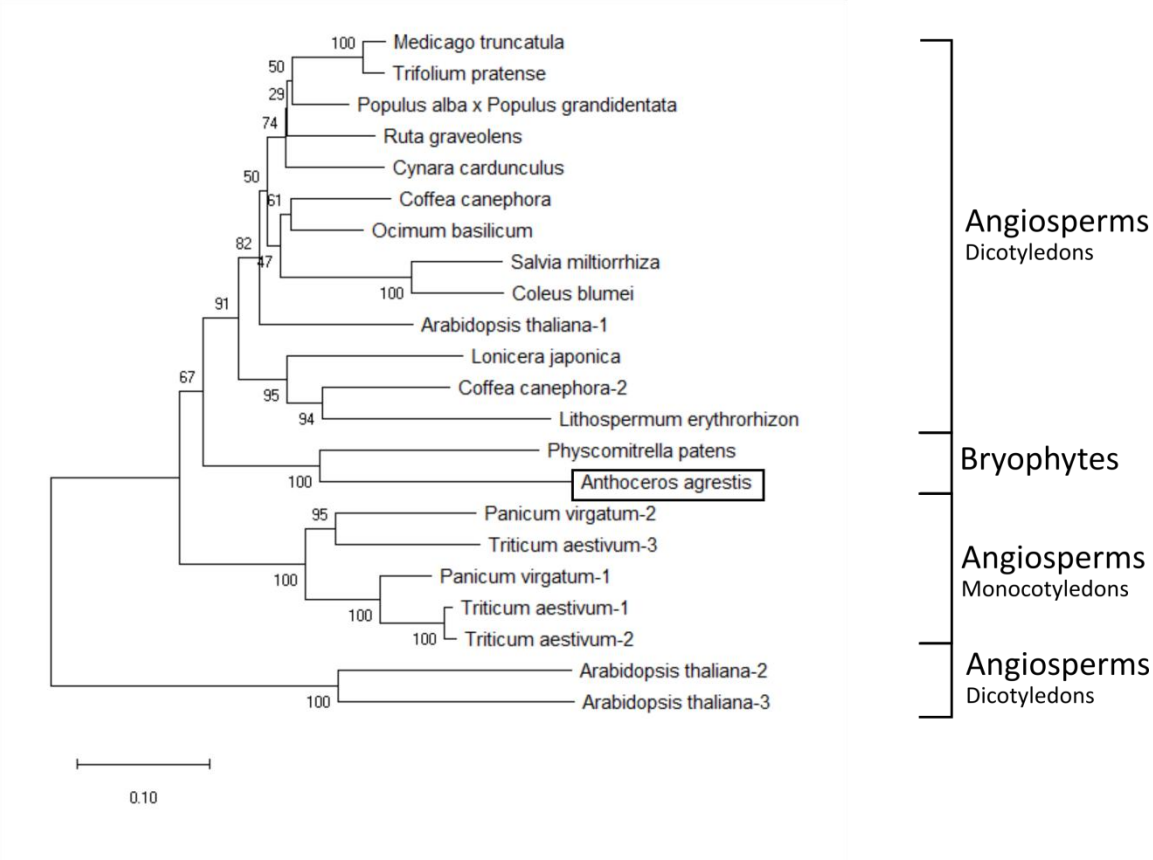


Figure 73 Phylogenetic analysis of CYP98 amino acid sequences. The maximum likelihood tree was constructed using the MEGA X software. The robustness of the branch structure was evaluated with a bootstrap analysis (1000 replicates). Evolutionary distance is represented with the bars. *Medicago truncatula* (ABC59086.1), *Trifolium pratense* (ACV91106.1), *Populus alba x Populus grandidentata* (EU391631), *Ruta graveolens* (JF799117), *Cynara cardunculus* (FJ225121), *Coffea canephora* (DQ269126), *Ocimum basilicum* (AY082611), *Salvia miltiorrhiza* (HQ316179.1), *Coleus blumei* (CAD20576), *Arabidopsis thaliana-1* (P850337), *Lonicera japonica* (KC765076), *Coffea canephora-2* (DQ269127), *Lithospermum erythrorhizon* (BAC44836), *Physcomitrella patens* (Pp3c22_19010), *Anthoceros agrestis* (this work), *Panicum virgatum*-2 (AB723824), *Triticum aestivum*-3 (CAE47491), *Panicum virgatum*-1 (AB723823), *Triticum aestivum*-1 (CAE47489), *Triticum aestivum*-2 (CAE47490), *Arabidopsis thaliana*-2 (AG52369) and *Arabidopsis thaliana*-3 (AM67314).

7.2 Substrates for AaCYP98

Most substrates used for the analyses of CYP98 were not commercially available. 4-Coumaroyl (pC) and caffeoyl (Caf) esters or amides (Fig. 74) were procured in different ways.

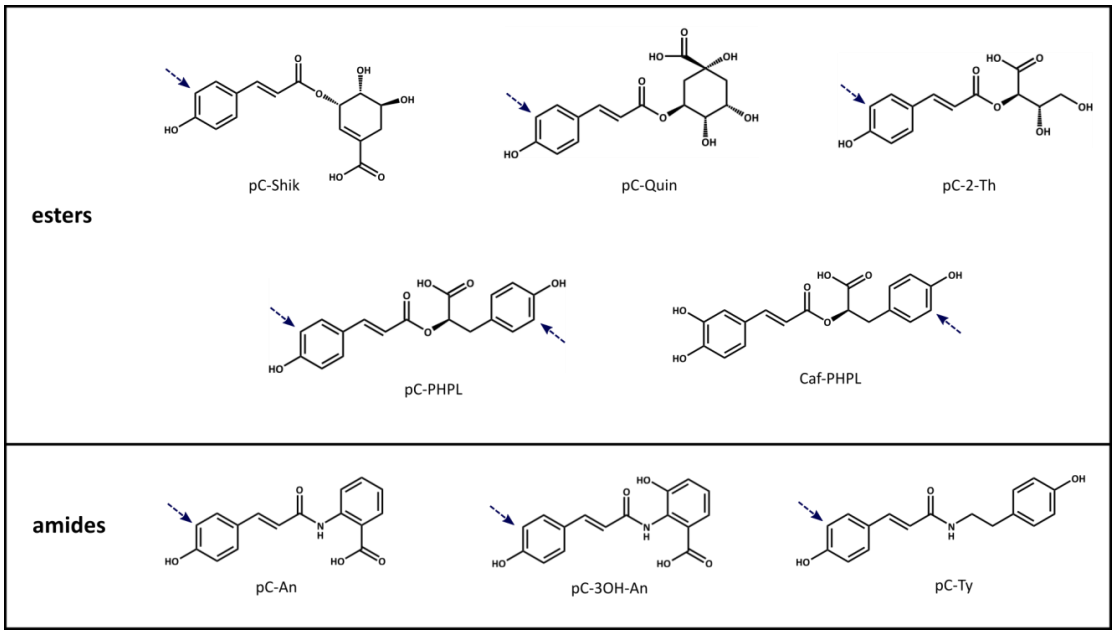


Figure 74 Synthesized or isolated 4-coumaroyl (pC) and caffeoyl (Caf) esters or amides. Possible hydroxylation sites are marked by arrows.

The precursors of rosmarinic acid, pC-4-hydroxyphenyllactic acid (pC-PHPL) and Caf-4-hydroxyphenyllactic (Caf-PHPL), were isolated from *Melissa officinalis* suspension cultures treated with the cytochrome P450 inhibitor tetcyclacis, as described in chapter II.4.3.1. After isolation, the extract was purified by TLC (Fig. 75). This led to several elution fractions, which were checked for purity on HPLC. The concentration was determined photometrically (pC-PHPL: 312 nm; Caf-PHPL: 328 nm) and calculated using the extinction coefficients for pC-PHPL (11.94 $mM^{-1}cm^{-1}$) and Caf-PHPL (9.15 $mM^{-1}cm^{-1}$). Only the purest fractions were used in enzyme assays.

pC-shikimate (pC-Shik), pC-quinate (pC-Quin) and pC-3´-hydroxyanthranilate (pC-3OH-An) were obtained from enzyme assays as described in chapter II.4.3.2. While pC-Quin and pC-3OH-An were used directly without any further purification, pC-Shik was isolated and purified by Olga Haag by HPLC.

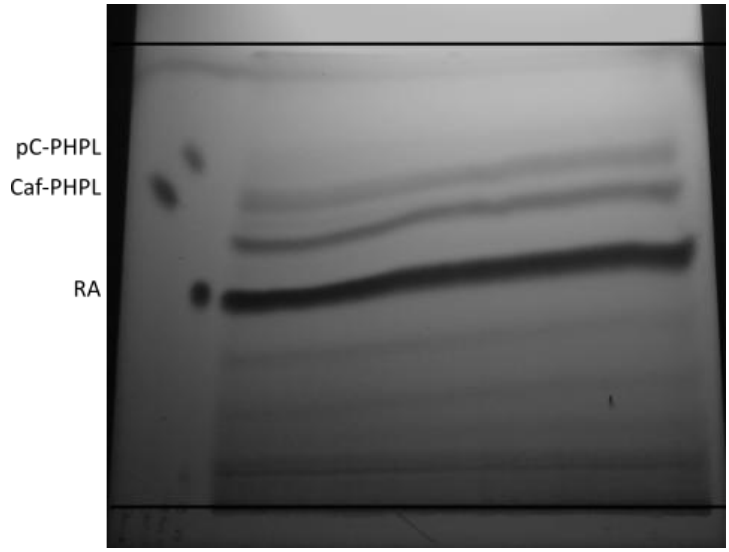


Figure 75 TLC of an ethanolic extract of *Melissa officinalis* treated with the cytochrome P450 inhibitor tetcyclacis. pC-PHPL, Caf-PHPL and rosmarinic acid (RA) standards were applied on the left side for determination of the R_f values. EtOAc:chloroform:formic acid in a ratio of 5:4:1 was used as mobile phase and silica gel F_{254nm} was used as stationary phase, detection at 254 nm.

pC-anthranilate (pC-An) and pC-tyramine (pC-Ty) were synthesized by a coupling reaction (Chapter II.4.3.3). The chemical synthesis was done in the group of Prof. Dr. Schlitzer (Universität Marburg), under supervision of Rolf Erik Emmerich and with the help of Jan Walter. The substances were purified by column chromatography and were checked by ^{1}H-NMR (Appendix VIII.3). While pC-Ty was synthesized in a one step reaction using EDC and triethylamine, this reaction failed for pC-An. After purification pC-Ty was obtained as a white solid (~ 20 mg/ > 10 %). pC-An was synthesized according to Alber et al. (2019). The first step was the formation of (*E*)-3-(4-acetoxyphenyl) acrylic acid by incubation of 4-coumaric acid with acetic anhydride. The product was filtered off and dried to get a white solid (6.3 g/100 %). The substance was checked by ^{1}H-NMR. Afterwards (*E*)-3-(4-acetoxyphenyl) acrylic acid was incubated with oxalyl chloride to obtain 4-(acetyloxy)-(*E*)-cinnamoyl chloride which was used directly in the next step, the coupling to anthranilic acid. The resulting 4-(acetyloxy)-cinnamoylanthranilic acid was purified and checked by ^{1}H-NMR (77 mg/ ≥ 50 %). At last the acyl group was removed by incubation with NaOH. The resulting pC-An was purified to obtain a yellow/orange solid (35 mg/ 67 %).

pC-2-threonic acid (pC-2-Th) was kindly provided by H. Renault, University of Strasbourg.

7.3 Expression of AaCYP98 in *Saccharomyces cerevisiae*

The first attempt of heterologous expression was carried out using *Saccharomyces cerevisiae*. Primers were designed with restriction sites for EcoRI (5´) and NotI (3´) and the stop codon was removed to ensure the attachment of a FLAG-tag at the *C*-terminus already present on the expression vector pESC$_{-ura}$. After PCR-amplification (Fig. 71d) the fragment was first inserted into the vector pDrive and multiplied in *E. coli* EZ for checking the correctness of the sequence. Hereafter the sequence was checked. Then both, expression vector and fragment in pDrive, were digested with EcoRI and NotI. After insertion of the sequence into pESC$_{-ura}$ and multiplication in *E. coli* EZ the sequence was checked again. The plasmid uptake was checked by colony-PCR (Fig. 76a) after transformation of the yeast strains BY4741, BY4742 and CB018. Expression in SCG$_{-ura}$ medium failed to produce the expected protein as shown by immunoblotting (Fig. 76b). Furthermore no product formation could be detected in enzyme assays using microsomes with pC-PHPL and Caf-PHPL as substrates.

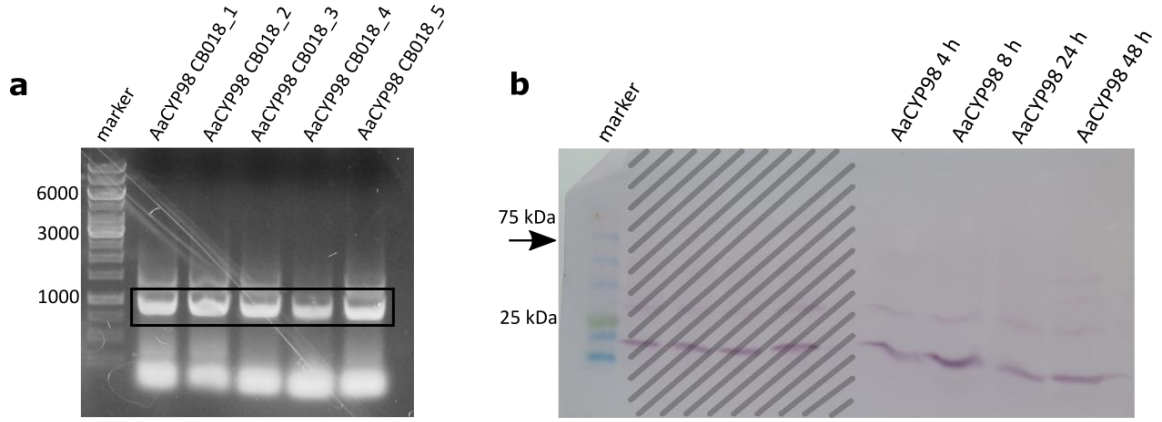


Figure 76 Expression of AaCYP98 in *S. cerevisiae* strain CB018. a Colony-PCR of five different transformants; **b** Western blot analysis with anti-FLAG-antibody. Samples were taken after 4, 8, 24 and 48 h of incubation in SCG$_{-ura}$-medium. Collected samples were adjusted to an OD_{600} of 4. Hatched areas belong to other experiments.

7.4 Expression of AaCYP98 in *Physcomitrella patens*

For expression in *P. patens* the restriction sites were changed to SalI (5´) and EcoRI (3´) and equipped with *C*-terminal 6xHis codons. The sequence was then inserted into the entry vector pENTR™1A. With the help of a LR recombination reaction the sequence was

then ligated into pTHUbiGate. After the vector had been multiplied, *Physcomitrella* was transformed. Six stable transformants were obtained after selection on hygromycin. All six were cultivated, but transformant 1 did not grow. The others were tested for AaCYP98 by immunoblotting with an anti-His-antibody (Fig. 77a). While the wildtype did not show any signal at around 60 kDa, transformant 3 showed the strongest signal and transformant 6 a faint one. Therefore all further experiments were performed with transformant 3.

At first, a culture characterisation over 21 days was performed. The transformed Pp_AaCYP98 cultures showed no difference to the *P. patens* wild type cultures regarding the fresh weight and pH of the medium (Fig. 77b). Both cultures reached a fresh weight of around 1.2 g per flask after 21 days and the medium pH increased from pH 6.3 to 7.1.

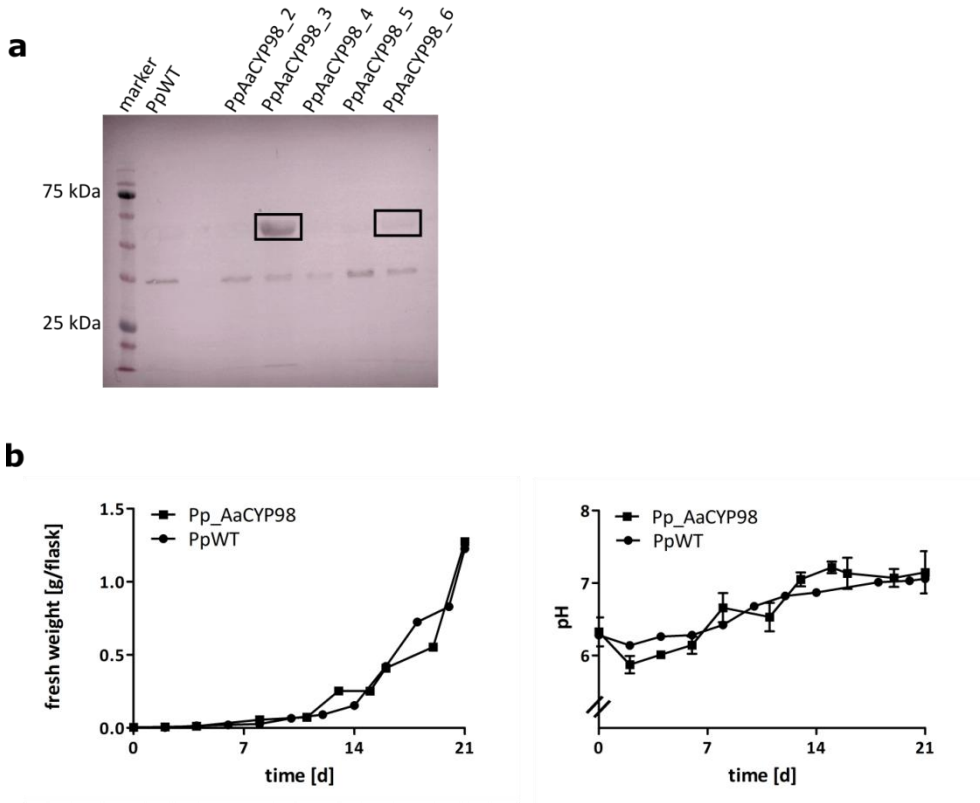


Figure 77 Expression of AaCYP98 in *P. patens*. a Western blot analysis of Pp_AaCYP98 transformants with anti-His-antibody. *P. patens* was cultivated for 12 days, afterwards a crude protein extract (2 mg/ml) was prepared. Crude protein extract of *P. patens* wild type served as a negative control; **b** Culture characterization of Pp_AaCYP98 and PpWT over 21 days. Samples were taken every second to third day. Fresh weight [mg/flask] is displayed on the left side, pH of the culture medium on the right side (mean ± SD, n=3).

A crude protein extract was prepared from the cells after 12 d of cultivation. Enzyme assays with pC-PHPL, Caf-PHPL, pC-Shik all led to an increasing peak over time at a retention time of about 6.75 min. Assays without any substrate showed, that this peak also appeared at a much higher rate in the transformed culture compared to the wildtype (Fig. 78a). Analysis of these assays by LC-MS revealed that the peak had two potential masses of m/z 308.04 or 444.08 [M-H]. The compound could not yet be identified.

Another problem was the decreasing substrate concentration over time, while there was no product formation, for example with Caf-PHPL and pC-Shik (Fig. 78b). On the other hand, also Caf-2-Th/Caf-4-Th, compound(s) present in the crude extract and potential product(s) after hydroxylation of pC-2-Th/pC-4-Th, decreased over time. For this reason potentially formed products could not be unequivocally detected.

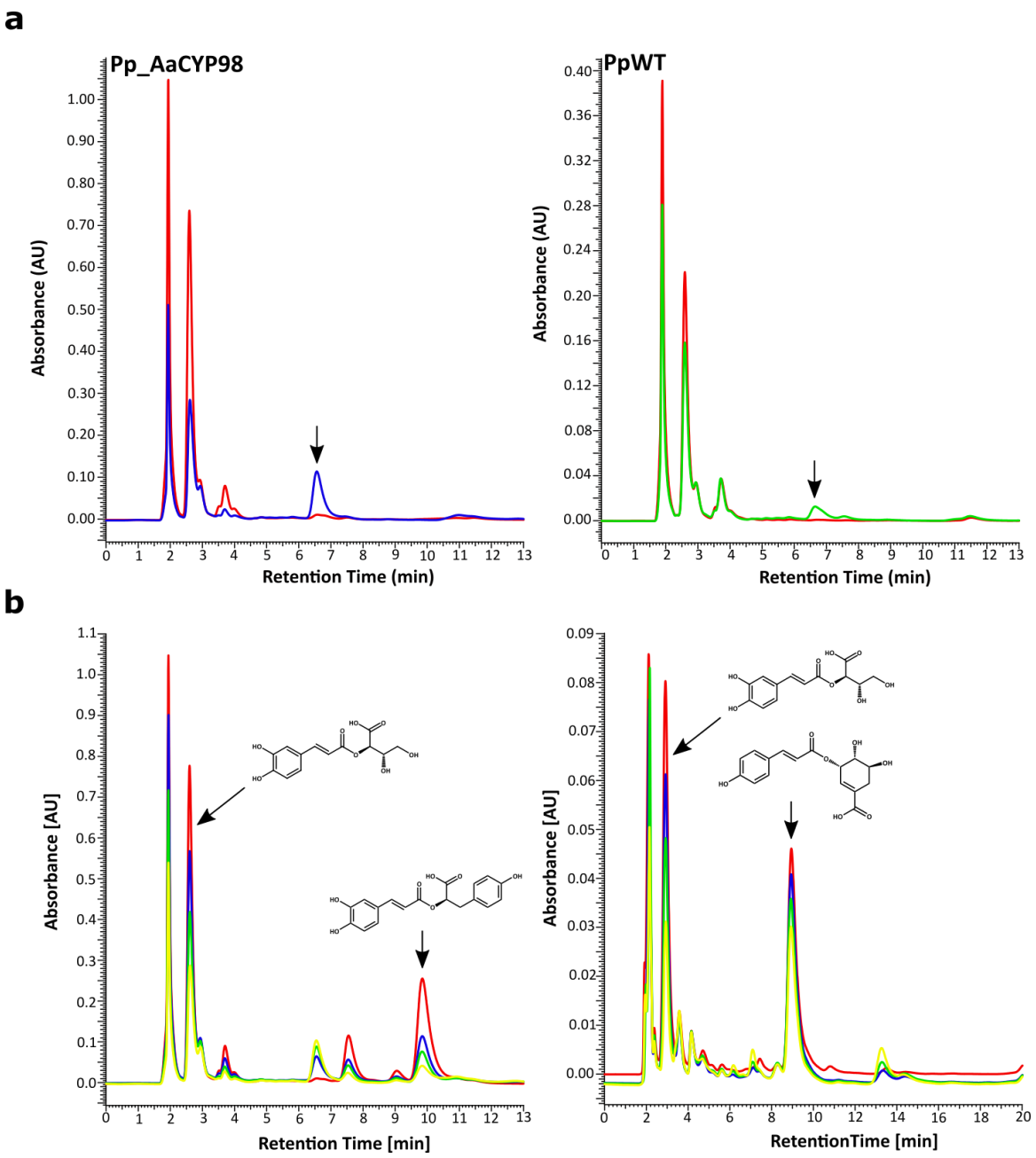


Figure 78 Enyzme assays of Pp_AaCYP98 and PpWT. a HPLC chromatograms of assays incubated with crude protein extract (200 µg protein) of Pp_AaCYP98 (left) or PpWT (right) and NADPH. The assay incubated for 0 min is displayed in red and for 60 min in blue (Pp_AaCYP98) or green (PpWT). The unknown product is marked with an arrow; **b** Decreasing substrates and potential products incubated with crude protein extract (200 µg protein) of Pp_AaCYP98. The chromatogram on the left shows an enzyme assay with Caf-PHPL incubated for 0 (red), 15 (blue), 30 (green) and 60 min (yellow) at 25 °C. The chromatogram on the right shows an enzyme assay with pC-Shik incubated for 0 (red), 15 (blue), 30 (green) and 60 min (yellow) at 25 °C. The two substrates as well as the potential product Caf-2-Th decreased, without any other product formation.

7.5 Expression of codon-optimized AaCYP98 in *Saccharomyces cerevisiae*

Since expression in *P. patens* led to some difficulties, the *AaCYP98* sequence was codon optimized (*coAaCYP98*) for the expression in yeast. This was carried out commercially by GeneCust. The sequence was equipped with restriction sites for EcoRI (5´) and NotI (3´), the stop codon was removed and the codon-optimized sequence was delivered in the vector pUC57. The new sequence revealed a decreased GC-content (Fig. 79). A closer look at the nucleotide sequence showed that often nucleotides (which allowed a change) were changed to adenine or thymine:

5´ATGGCAGAATTGGTTGTTAGAGCTGTTTTGGCTTTGATTGCTGTTGCTGTTGCTGCTGGTTTT
TTGTTTTATAGAGCTAGATGGTCTAGATATAAGTTGCCACCAGGTCCAAAGCCAAGATTGTTTGT
TGGTAATTTGTTGGATGTTACTGCTGTTAGATTTAAATGTTTCGAACAATGGGCTAAGGTTTATG
GTCCAGTTATGTCAGTTTGGATGGGTGCTACTTTGAATGTTGTTGTTTCTTCTGCTGAAGGTGCT
AAAGAAGTTTTGAAAGATAAGGATCAATCCTTGGCTGCTAGAGCAAGATCTAGAGCTGCTGCTAA
ATTTTCTAGAGATGGTCAAGATTTGATCTGGGCTGATTATGGTCCACATTATGTTAAGGTTAGAA
AAGTTTGTAACGTCGAATTGTTTACTCCAAAGAGATTGGAAGCTTTGAGACCAATTAGACAAGAT
GAAGTAAGTGCTATGGTTCATAAAATCCATAGAGATGCTCCAAAGGGTCCAGTCAATCCAAAAAA
ATATTTGTCTGCTATGGCTTTTAATAATATTACTAGATTGGCTTTTGGTAAAAGATTTGTTGATG
ATGAAGGTAAAATTGATTCTCAAGGTGTTGAATTTAAAGAAGTTATTGGTCAAGGTATGAAATTG
GGTGCTTCTTTGAAAATTTCTGAACATATTCCATCTATTAGATGGATGTTTCCATTGCAAAATGA
AGAATTTGCTAAACATGGTGCTAGAAGAGATACTTTGACTAGAACTATTATGAAAGAACATGCTG
CTTTGAGAGAAACTTTGGGTCCACAATCTCATTTTGTTGATGCTTTGTTGGGTTTGCAAAACAA
TATGATTTGTCTGAAACTACTATTATTGGTTTGTTGTGGGATATGGTTACTGCTGGTATGGATAC
TGTTGCTATTACTGTTGAATGGGCTTTGGCTGAATTGATTAGAAATCCAGAAGTTCAAAGAAAAG
CTCAAGAAGAATTGGATGCTGTTGTTGGTGATGAAACTGTTTTGACTGAAGCTCATTTTTCTAAA
TTGCCATATTTGACTGCTTTGACTAAAGAATCTTTGAGATTGCATCCACCAACTCCATTGATGTT
GCCACATAAATGTACTGAAGATGTTAAAATTCAAGGTTATGATATTCCAAAAGGTACTGTTGTTC
ATTGTAATATTTATGTTATTGCTAGAGATCCAGCTGTTTGGAAAGAACCATTGGCTTTTAGACCA
GAAAGATTTTTGGAACAAGATGTTGATATTAAAGGTCATGATTTTAGATTGTTGCCATTTGGTGC
TGGTAGAAGAGTTTGTCCAGGTGCTCAATTGGGTTTGAATATGGTTCAATTGATGATTGGTAGAT
TGTTGCATCAATTTTCTTGGTCTCCACCAGAAGGTGTTAGAGCTGAAGATATTTCTTTGGAAGAA
AGATCTGGTATTGTTACTTATATGGCTAAATCTTTGGAAGCTGTTGCTACTCCAAGATTGCCACA
AGCTTTGTATGCTAGAAATGTT 3´

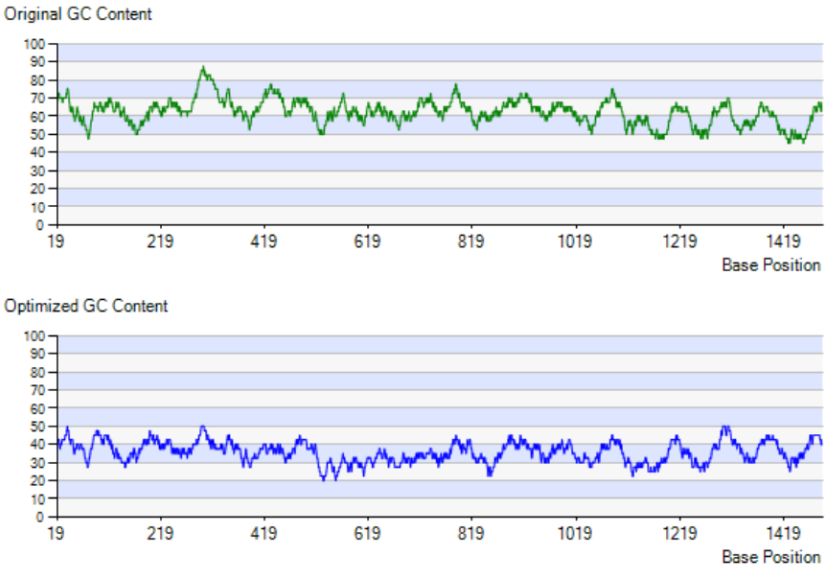


Figure 79 GC-content of original AaCYP98 and codon-optimized AaCYP98. The graphic was provided by GeneCust.

coAaCYP98 was ligated into the MCSI from the expression vector pESC$_{-ura}$. Furthermore, a sequence encoding a CPR from *Coleus blumei* (CbCPR) characterized by Eberle et al. (2009) was ligated into MCSII or MSCII was left empty.

The plasmid without the reductase was inserted into the yeast strains BY4742, BY4741, CB018 and InvSc1. After colony-PCR (Fig. 80a) with primers directed against the MCSI from pESC$_{-ura}$ a signal was detected at the expected size of approximately 1800 bp. It was strongest for the transformants BY4742 and InvSc1. BY4741 and CB018 only showed faint or no bands. Next, the protein was expressed and detected by Western blotting with an anti-FLAG-antibody. Like the result of colony-PCR, BY4742 followed by InvSc1 showed the highest protein formation at around 60 kDa. No stained protein bands were seen in BY4741 and CB018 extracts (Fig. 80b).

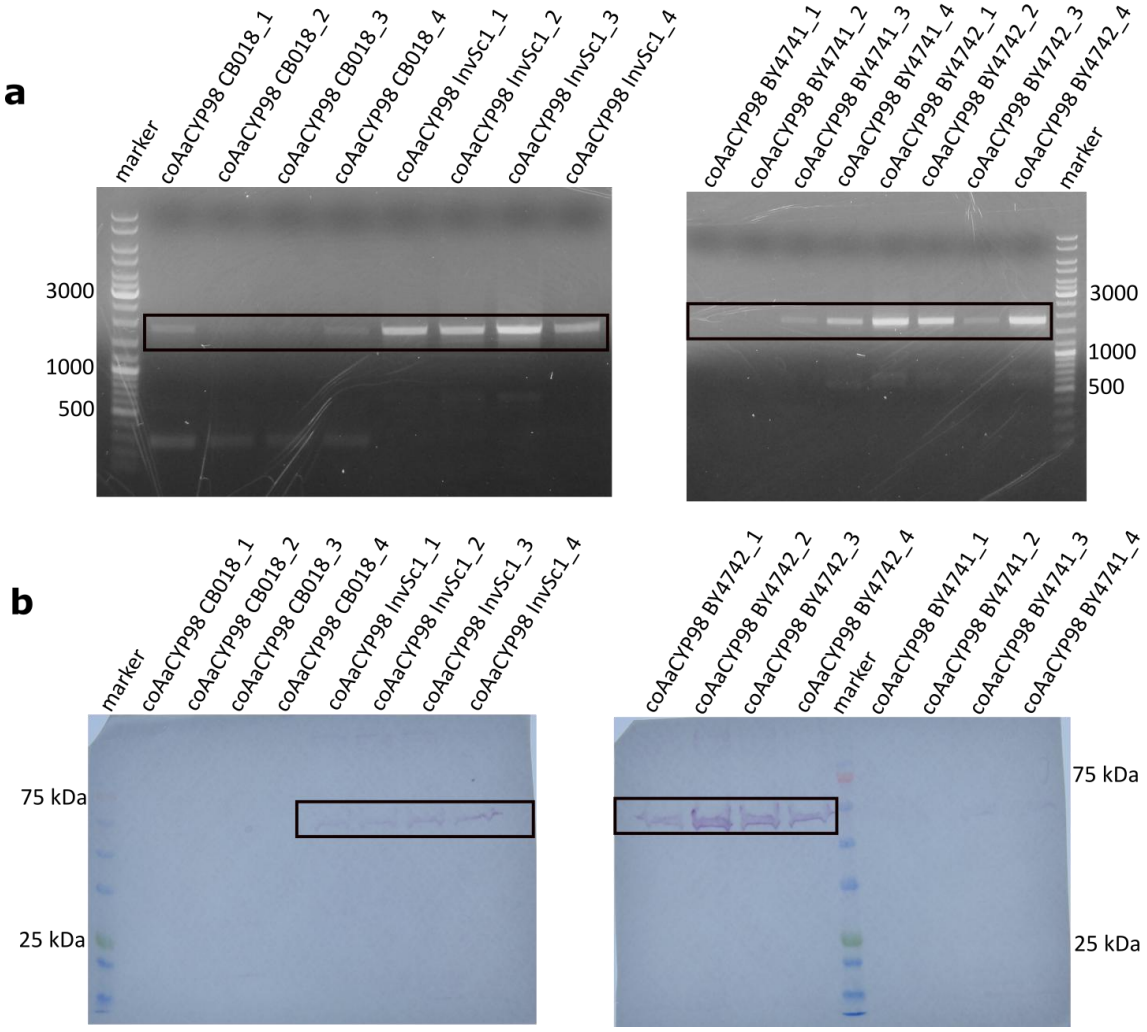


Figure 80 Expression of coAaCYP98 in different *S. cerevisiae* strains. a Colony-PCR of *coAaCYP98* expressed in the yeast strains CB018, InvSc1, BY4741 or BY4742. For each strain four transformants were tested; **b** Western blot analysis of coAaCYP98 with anti-FLAG-antibody. Four transformants of each yeast strain were checked (same transformants as tested by colony-PCR). Samples were taken after 48 h of incubation in SCG_{-ura}-medium and had a OD_{600} of 5.

For this reason the plasmid carrying *coAaCYP98* and *Cb_CPR* (*coAaCYP98+CbCPR*) was only inserted into BY4742 and InvSc1. Colony-PCR was carried out using primers against both MCSI and MCSII (Fig. 81a). The amplification of *coAaCYP98* (around 1800 bp) worked much better than *Cb_CPR* (around 2600 bp), but all transformants showed a signal for both genes. Immunoblotting with anti-FLAG-antibody (Fig. 81b) displayed a slight signal mostly in BY4742 and even fainter in InvSC1. A Western blot using anti-Myc-antibody, against CbCPR, was not attempted.

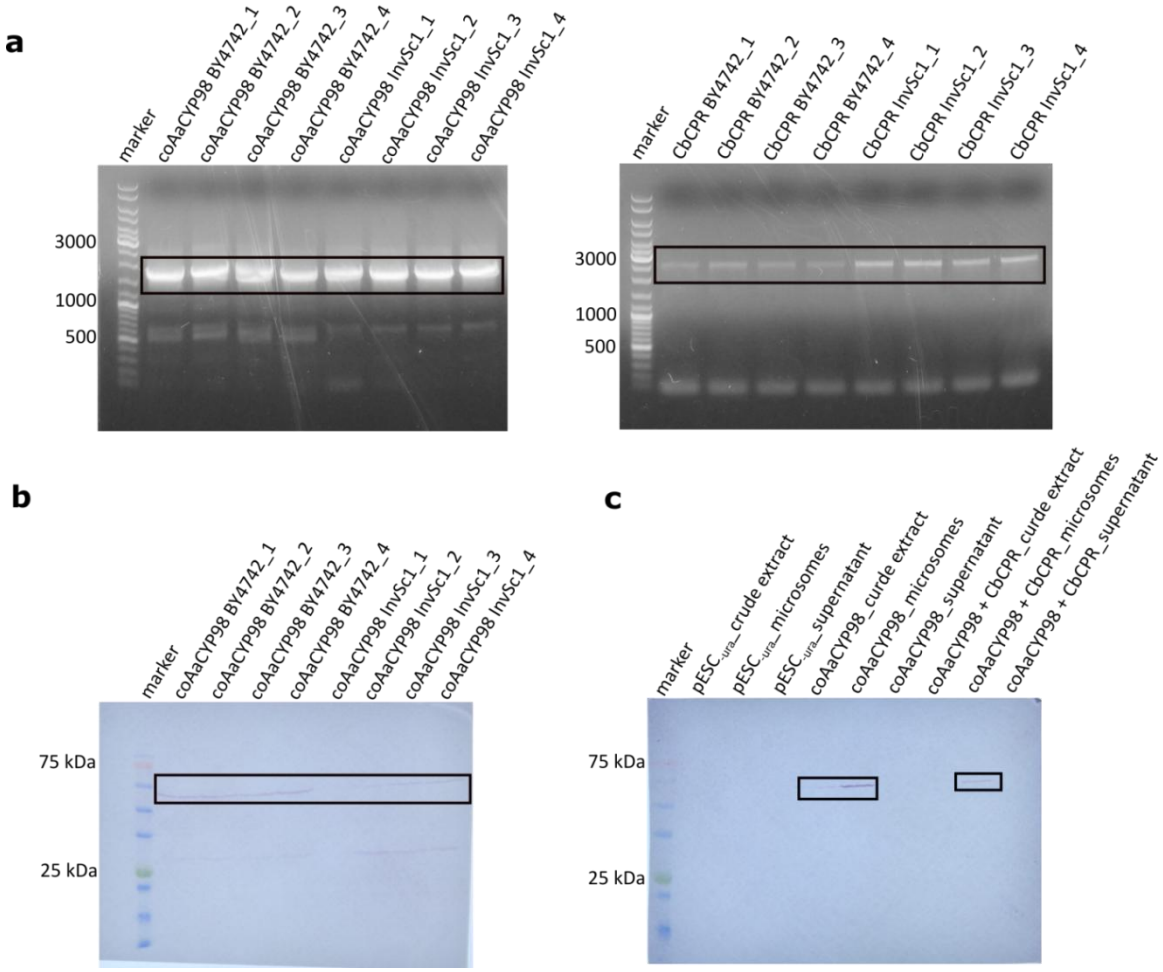


Figure 81 Expression of coAaCYP98+CbCPR in *S. cerevisiae* strains BY4742 and InvSc1. a Colony-PCR of *coAaCYP98* in MCSI (left) and *CbCPR* in MCSII (right). Four transformants of each strain were checked; **b** Western blot analysis of coAaCYP98 with anti-FLAG-antibody. Four transformants of each strain were checked and samples were taken after 48 h of incubation in SCG_{-ura}-medium and had an OD_{600} of 5; **c** Western blot analysis of crude protein extracts, microsomes and supernatant of coAaCYP98, coAaCYP98+CbCPR and BY4742 transformed with the empty vector $pESC_{-ura}$. Anti-FLAG-antibody was used for detection.

Overall the transcription and translation seemed to be better for the plasmid carrying only *coAaCYP98* and was best in BY4742. Therefore BY4742 was chosen as yeast strain for expression. Transformation of BY4742 with the empty vector served as control for all further experiments.

After another round of expression, proteins were isolated and microsomes were prepared. Although the protein concentration was rather low, protein localization was shown to be in the microsomal fraction and not in the supernatant (Fig. 81c). Enzyme

assays using crude protein extracts either from coAaCYP98 or coAaCYP98+CbCPR with Caf-PHPL, pC-Quin or pC-2-Th did not show any product formation after 60 min (Fig. 82).

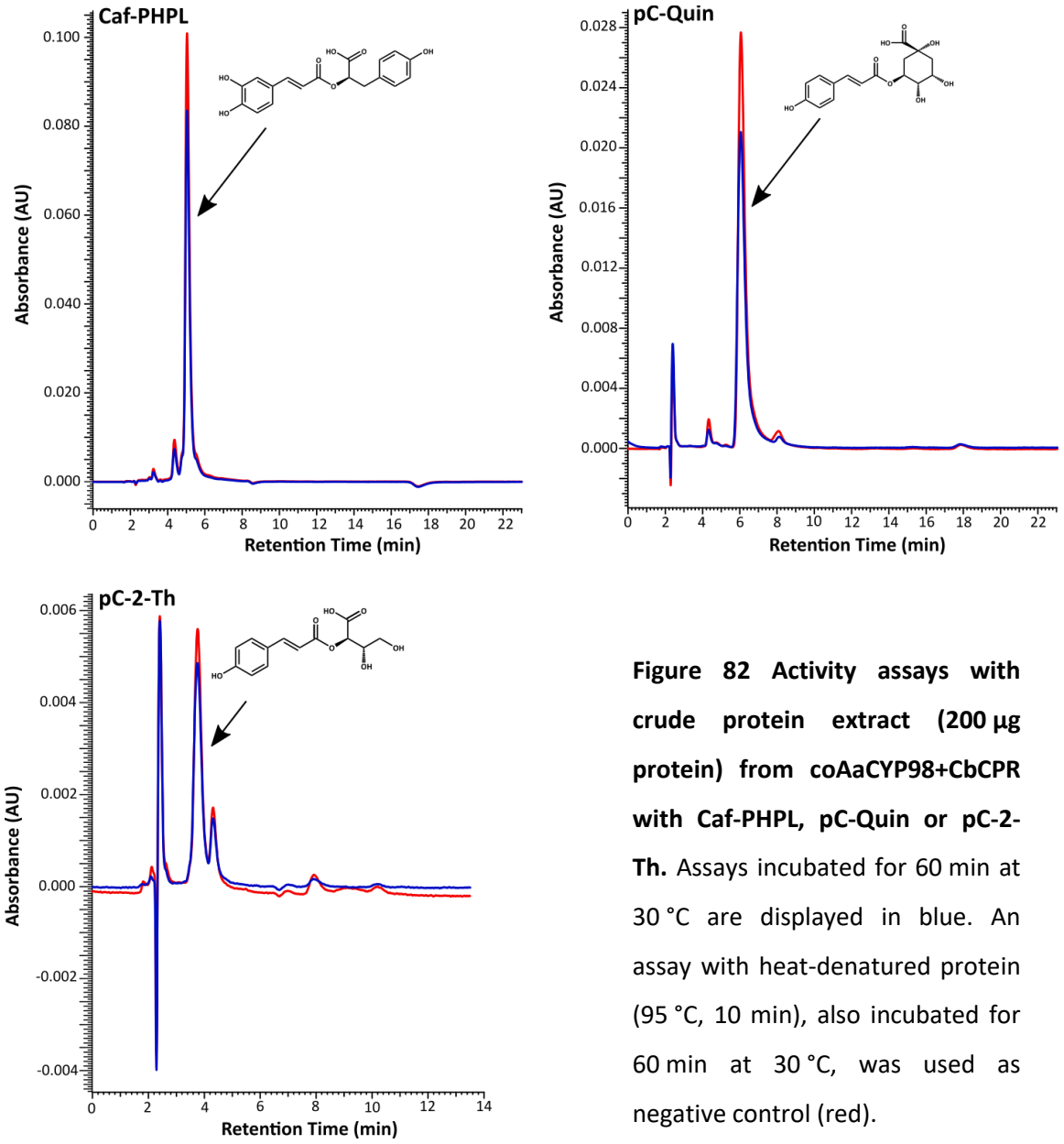


Figure 82 Activity assays with crude protein extract (200 µg protein) from coAaCYP98+CbCPR with Caf-PHPL, pC-Quin or pC-2-Th. Assays incubated for 60 min at 30 °C are displayed in blue. An assay with heat-denatured protein (95 °C, 10 min), also incubated for 60 min at 30 °C, was used as negative control (red).

Only the substrate pC-3OH-An led to a small increasing product peak with crude protein extract from coAaCYP98 (Fig. 83a). This peak did not appear using crude protein extract from the empty vector control transformant (Fig. 83b). Analysis of these assays by LC-MS proved the existence of Caf-3OH-An with a mass of m/z 314.07 [M-H] (Fig. 83c). Product formation was increased by 13-fold by using crude protein extract from coAaCYP98+CbCPR instead of only coAaCYP98 (Fig. 83d).

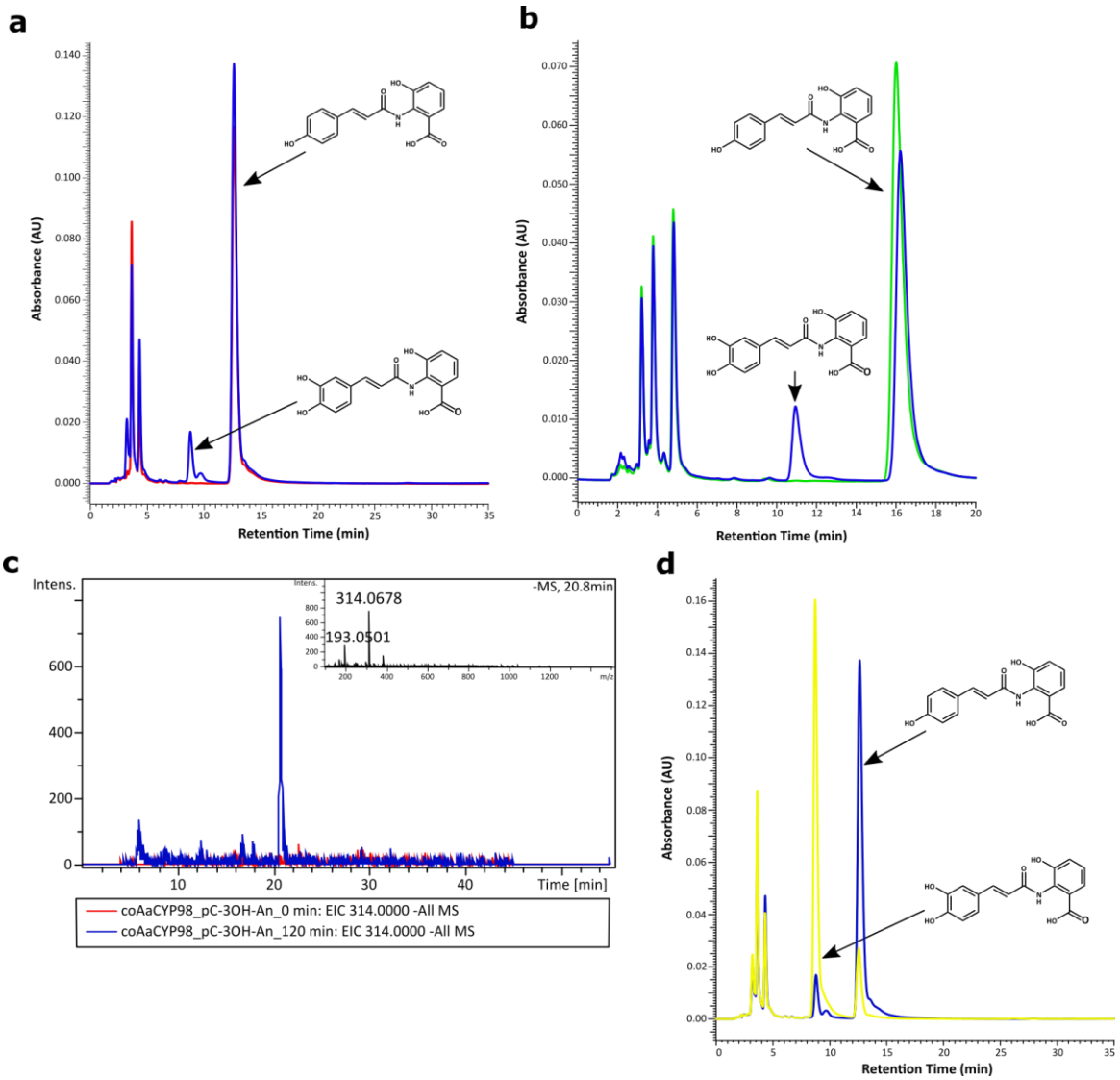
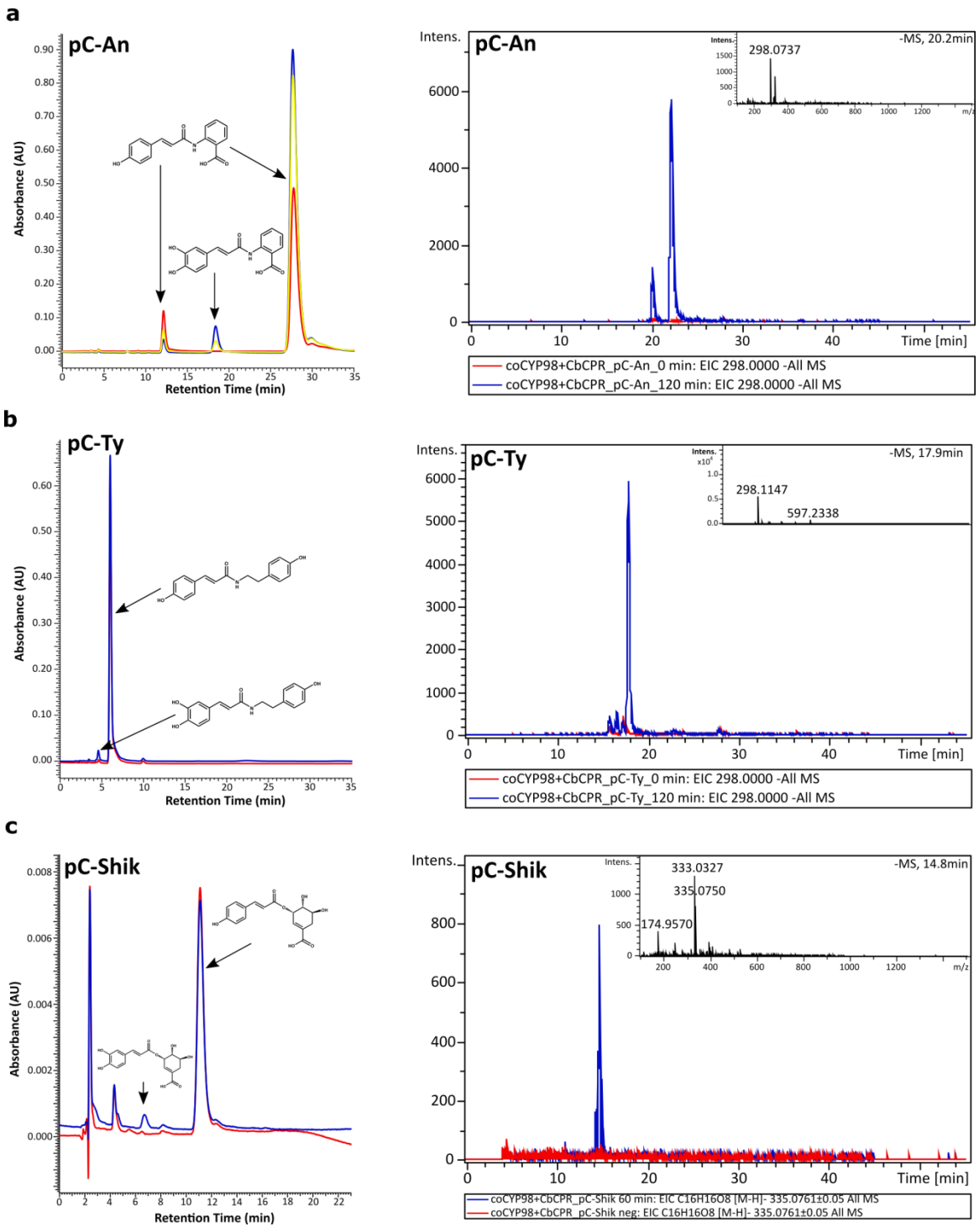


Figure 83 Activity of crude protein extract (200 µg protein) of coAaCYP98, coAaCYP98+CbCPR with pC-3OH-An analyzed by HPLC (333 nm) or LC-MS. a pC-3OH-An incubated with coAaCYP98 for 0 min (red) or 120 min (blue) at 25 °C; **b** pC-3OH-An incubated for 120 min at 25 °C with coAaCYP98 (blue) or crude protein extract from the empty vector control (green). In this case, the isocratic elution was performed with 45 % MeOH + 0.01 % H_3PO_4 instead of 50 % MeOH + 0.01 % H_3PO_4; **c** LC-MS-analysis of coAaCYP98 incubated with pC-3OH-An for 0 min (red) or 120 min (blue). Chromatograms show the EIC of Caf-3OH-An in the negative mode. The exact mass of Caf-3OH-An is m/z 314.07 [M-H]; **d** Enzyme assay with pC-3OH-An after 120 min at 25 °C with either coAaCYP98 (blue) or coAaCYP98+CbCPR (yellow).

Furthermore, pC-An, pC-Ty, pC-Shik and pC-PHPL were also accepted. This could only be observed using coAaCYP98+CbCPR. Masses of Caf-An (m/z 298.07 [M-H]), Caf-Ty (m/z 298.11 [M-H]), Caf-Shik (m/z 335.06 [M-H]) Caf-PHPL or pC-DHPL (m/z 343.08 [M-H]) were again checked by LC-MS (Fig. 84a-d). In this case, heat-denatured protein (5 min at

95 °C) or incubation for 0 min was used as a negative control. Enzyme assays with crude protein extract of the empty vector control showed no additional peak after 30, 60 and 120 min with pC-Ty or pC-An (Fig. 84e).



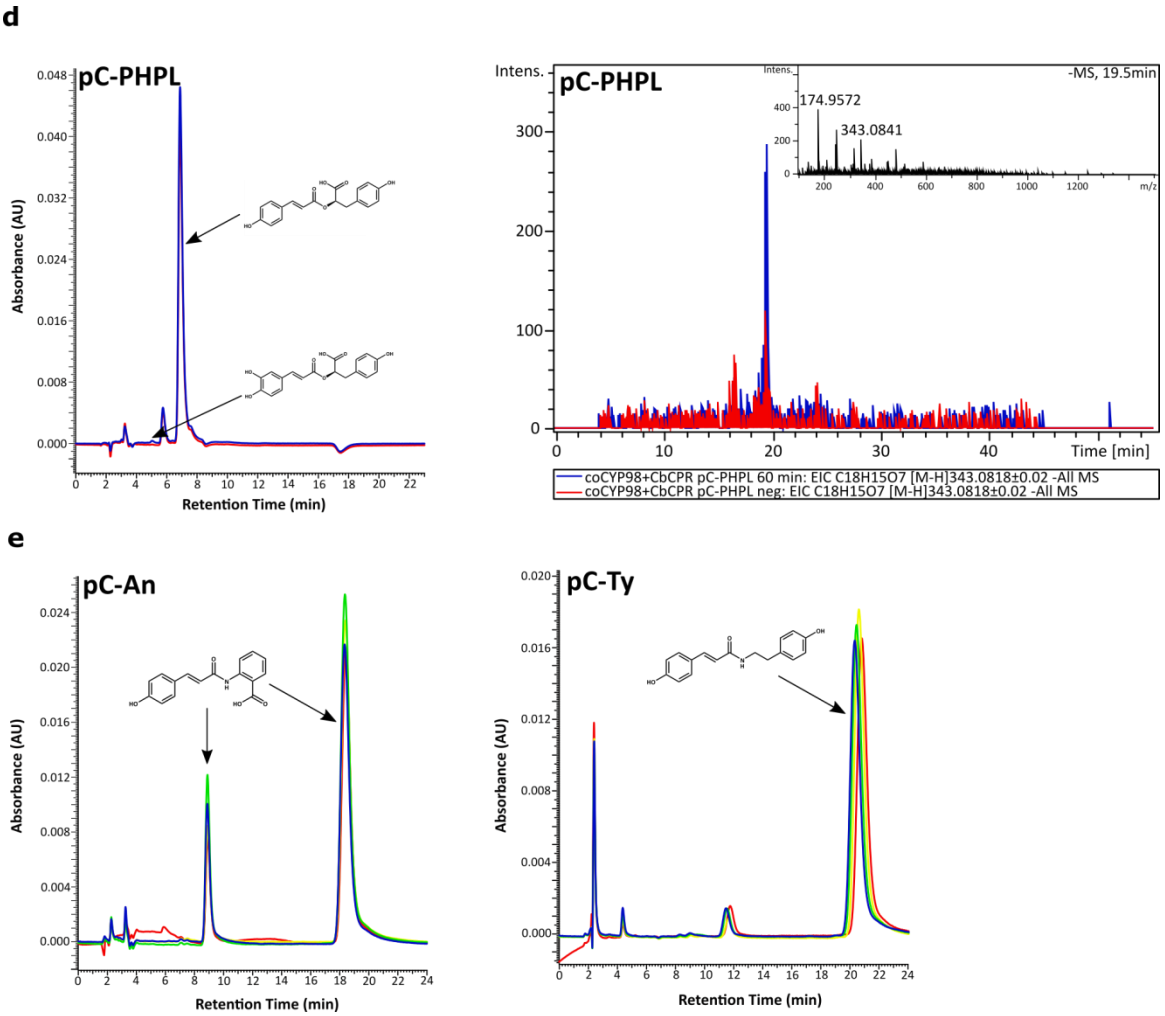


Figure 84 Activity of coAaCYP98+CbCPR (200 µg protein) with different esters and amides of 4-coumaric acid. a pC-An. On the left side the HPLC chromatogram (333 nm) is displayed after incubation for 0 min (red), 60 min (yellow) or 120 min (blue) at 25 °C. Both pC-An peaks are marked by arrows. The EIC and mass of Caf-An obtained by LC-MS analysis is displayed on the right. Again 0 min is displayed in red and 120 min in blue; **b** pC-Ty. 0 min reaction time is displayed in red and 120 min at 25 °C in blue in both HPLC (333 nm, left) and LC-MS (EIC, right) chromatograms. The mass of the produced Caf-Ty can be found in the inset in the LC-MS chromatogram; **c** pC-Shik. In this case, heat-denatured protein (5 min, 95 °C) served as a negative control, displayed in red. The assay after 60 min at 30 °C is displayed in blue. HPLC chromatogram at 333 nm is shown on the left side and the EIC with the corresponding mass of Caf-Shik from LC-MS analysis is displayed on the right; **d** pC-PHPL. HPLC chromatogram at 333 nm (left) and LC-MS analysis (right). Assay incubated for 60 min at 30 °C is displayed in blue and negative control (heat-denatured protein) in red. The corresponding mass of Caf-PHPL/pC-DHPL is shown in the inset in the LC-MS chromatogram; **e** HPLC analysis (333 nm) of pC-An (left) and pC-Ty (right) incubated with crude protein extract (200 µg protein) from the empty vector control. Assays were incubated for 0 (red), 30 (yellow), 60 (green) or 120 min (blue) at 30 °C. Again, both pC-An peaks are marked by arrows.

Due to unknown reasons pC-An formed two peaks in the HPLC chromatogram. The peak corresponding to the standard was found at a retention time of around 18 min and the second smaller peak had a retention time of approximately 9 min in HPLC analysis (using an isocratic elution with 50 % MeOH + 0.01 % H_3PO_4). Both peaks displayed the same mass in LC-MS analysis. The small peak had an area of around 20-25 % compared to the big peak. This was also observed in assays using the crude protein extract from the empty vector control (Fig. 84e). The relative amount remained unchanged in enzyme assays using buffers with different pH-values. By increasing the temperature the amount decreased to 5-10 % at 60 °C. In controls using only buffer instead of crude protein extract the small peak had only 5 % and this value remained the same at room temperature or 60 °C. The same peak pattern was observed in a much smaller amount (2 %) in the diluted stock solution dissolved in 50 % MeOH. In addition, also Caf-An formed two peaks (again both with the exact same mass in LC-MS analysis). Therefore both product peaks were summed for enzyme characterization.

Since CPR transfers electrons via FAD and FMN to the cytochrome, FAD and FMN were added to assays but this did not show any difference to assays without. Enzyme assays using crude protein extract, microsomes or supernatant indicated the same effect as with AaC4H (Chapter III.1.5.1). Assays with microsomes had only 25 % residual activity. Using NADH or NADP as cosubstrate led to approximately 5 % of Caf-3OH-Aa and less than 10 % Caf-An compared to assays with NADPH.

7.5.1 Temperature and pH-optimum of AaCYP98

pC-An was used as substrate for the determination of temperature and pH-optima. Since no standard of Caf-An was available at this time, the specific activity could not be calculated and only area was plotted; both Caf-An peaks were taken into account. For both optima two different transformants of coAaCYP98+CbCPR in BY4742 were expressed and crude protein extracts were isolated. To account for technical variation three replicates were made for each data point. AaCYP98 showed a temperature optimum at 30 °C and a pH optimum of pH 7.1 using a buffer with pH 7 (Fig. 85).

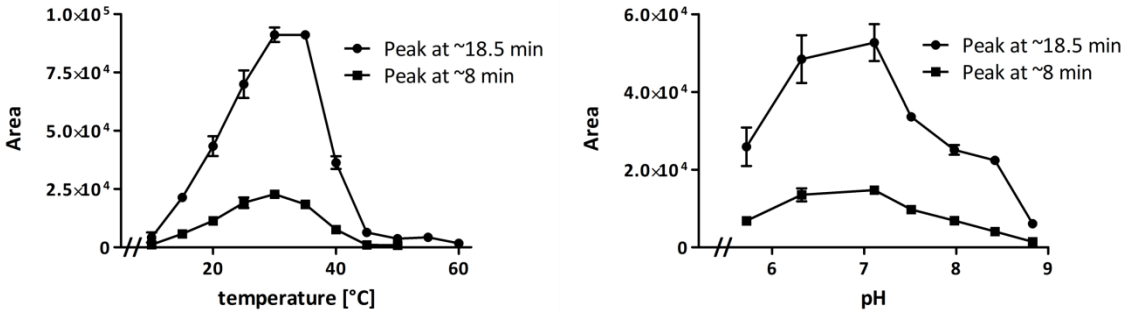
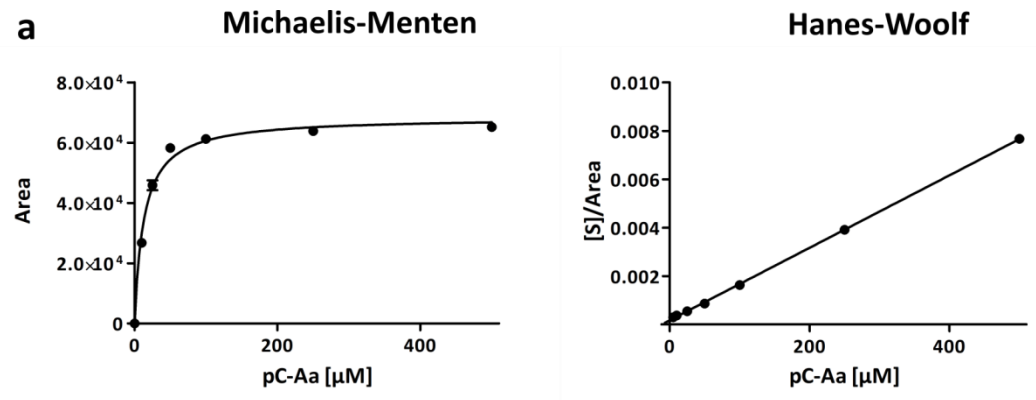


Figure 85 Temperature (left) and pH-optimum (right) of coAaCYP98+CbCPR. Values are mean of six independent determinations ± standard deviation.

7.5.2 AaCYP98: reaction kinetics

To this date, a standard of pC-3OH-An was not available. Pure pC-Shik was only available at amounts not sufficient for determination of kinetic data. pC-PHPL was not checked, since product formation was barely detectable after incubation for 60 min. Therefore K_m-values were only determined for pC-An, pC-Ty and the cosubstrate NADPH. To ensure biological variation, crude protein extract was obtained from three independent protein isolations from two different yeast transformants of coAaCYP98+CbCPR. Three technical replicates were made for each concentration. The results were analyzed with the GraphPad Prism 5 software using Michaelis-Menten, Lineweaver-Burk (not shown) and Hanes-Woolf models. For pC-An both product peaks of Caf-An were added up. The substrate saturation curve of pC-An led to an apparent K_m-value of 6.4 ± 0.8 µM. AaCYP98 displayed a K_m-value of 36.4 ± 3.6 µM for pC-Ty. NADPH AaCYP98 displayed a K_m-value of 53.2 ± 4.0 µM for NADPH (Fig. 86; Table 18).



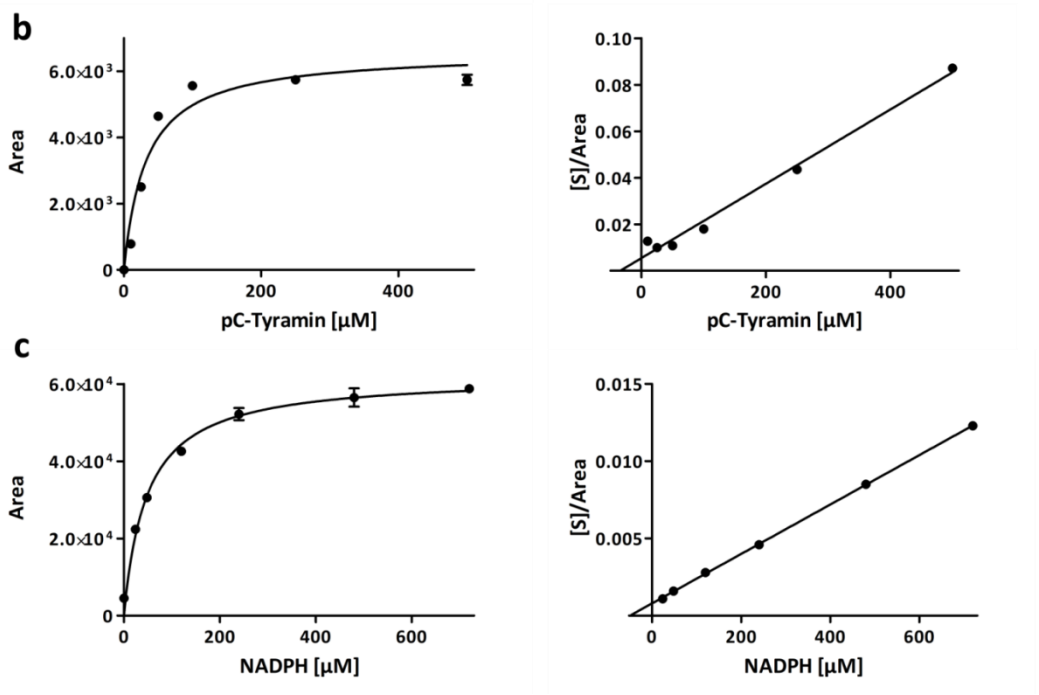


Figure 86 Kinetic data for coAaCYP98+CbCPR displayed by Michaelis-Menten and Hanes-Woolf diagrams. a pC-An; **b** pC-Ty; **c** NADPH. Graphs represent the mean values of one biological replicate with three technical replicates. Error bars represent the standard deviation.

Table 18 Kinetic values of AaCYP98 for p-coumaroylanthranilic acid (pC-An), p-coumaroyltyramine (pC-Ty) and NADPH. Values were calculated with the help of Michaelis-Menten (MM), Lineweaver-Burk (LB) and Hanes-Woolf plots (HW). Mean values of 9 determinations ± standard error.

		AaCYP98 K_m [µM]
pC-An	MM	6.4 ± 0.8
	LB	6.6 ± 0.6
	HW	5.5 ± 1.6
pC-Ty	MM	36.4 ± 3.6
	LB	44.6 ± 2.5
	HW	34.3 ± 0.7
NADPH	MM	53.2 ± 4.0
	LB	46.5 ± 2.9
	HW	50.6 ± 3.4

8 AaAp626

8.1 Identification and isolation of AaAp626 from *Anthoceros agrestis*

On the search for another *CYP98* candidate gene, *AaCYP98* was used as template for a BLASTp search in an *Anthoceros punctatus* database. Scaffold 626 was chosen and at first primers against a partial sequence were designed. After amplification using cDNA, two fragments were observed at the expected size of around 650 bp (Fig. 87a). After purification, ligation into pDrive, transformation of *E. coli* EZ, plasmid multiplication and isolation, the sequence of both fragments was determined. Both sequences were identified as *AaAp626*, but one of them had an additional intron. This was presumably because the cDNA had been contaminated with some remaining gDNA. 5´-RACE-PCR was performed (Fig. 87b) and the sequence was completed. The full-length nucleotide sequence consisted of 1590 bp which could be translated to a protein of 529 amino acid residues with a molecular mass of 59.53 kDa.

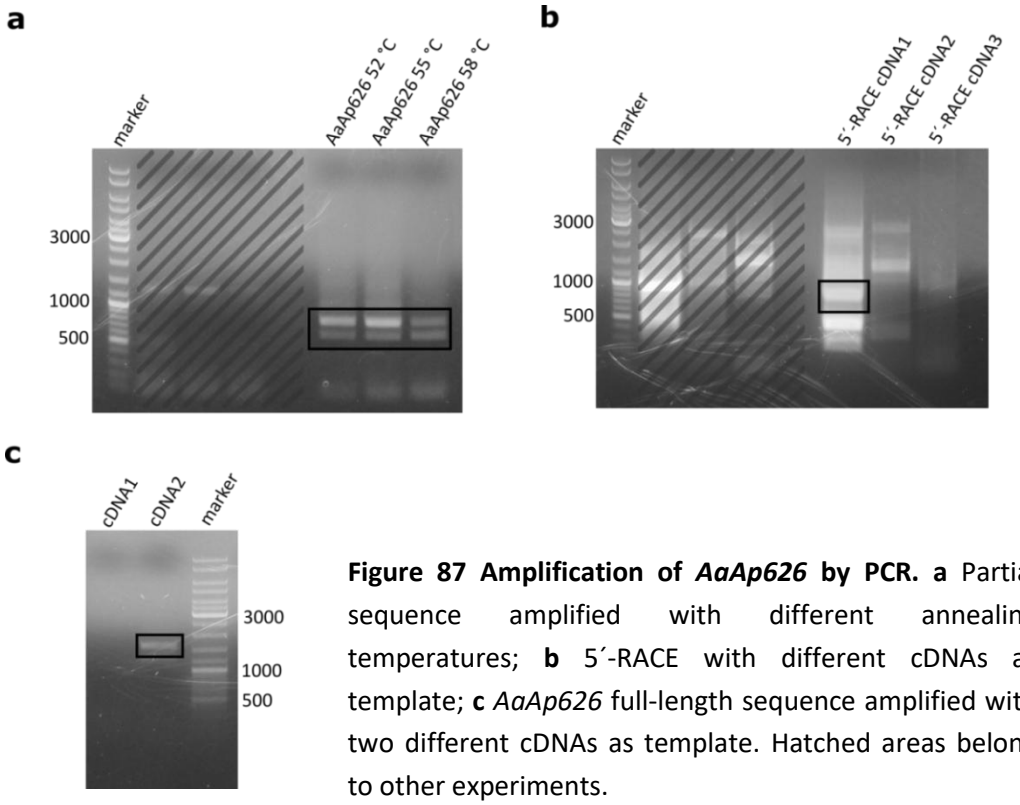


Figure 87 Amplification of *AaAp626* by PCR. a Partial sequence amplified with different annealing temperatures; **b** 5´-RACE with different cDNAs as template; **c** *AaAp626* full-length sequence amplified with two different cDNAs as template. Hatched areas belong to other experiments.

BLASTp algorithm, using the amino acid sequence, revealed a high identity to flavonoid-3-monooxygenases and CYP71A1, but all similar sequences had only been predicted by automated computational analysis (Table 19). Nevertheless, all general characteristics of canonical cytochrome P450 enzymes were identified (Fig. 88) like the proline-rich region, the O_2-binding region, the PERF motif and the heme-binding cysteine motif (Schuler 1996; Mizutani et al. 1997; Chapple 1998; Werck-Reichhart et al. 2002).

Table 19 Protein BLAST result of AaAp626. https://blast.ncbi.nlm.nih.gov/Blast.cgi 17.10.2019

Description	Organism	Identity	E value	Accession
PREDICTED: cytochrome P450 71A1-like	*Nelumbo nucifera*	40 %	6e-119	XP_010273844.1
PREDICTED: flavonoid 3'-monooxygenase-like isoform X2	*Pyrus x bretschneideri*	39 %	6e-118	XP_009337219.1
cytochrome P450 71A1-like	*Elaeis guineensis*	40 %	6e-117	XP_010911364.2
flavonoid 3'-monooxygenase-like	*Prunus avium*	40 %	1e-116	XP_021804836.1
flavonoid 3'-monooxygenase-like	*Physcomitrella patens*	38 %	5e-116	XP_024374554.1

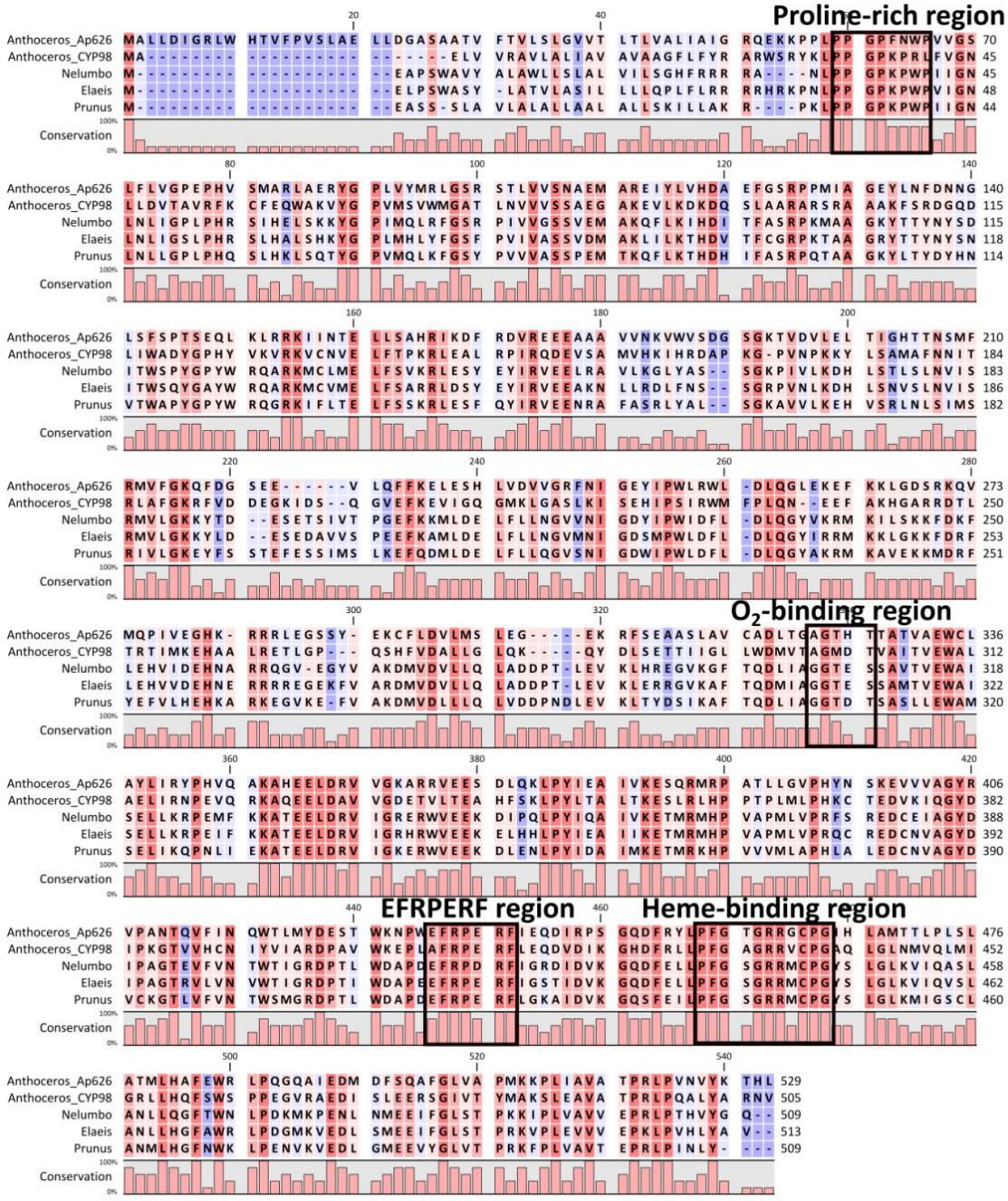


Figure 88 Alignment of AaAp626 amino acid sequence with AaCYP98 (this work) and three cytochrome P450 71A1-like or flavonoid 3'-monooxygenase-like sequences from *Nelumbo nucifera* (XP_010273844.1), *Elaeis guineensis* (XP_010911364.2) and *Pyrus x bretschneideri* (XP_009337219.1). Highly conserved amino acids are marked red, low conservation is marked blue. Typical conserved sequence motifs are marked by boxes.

8.2 Expression of AaAp626 in *Physcomitrella patens*

Because the expression of both AaC4H and AaCYP98 failed in yeast, but worked in *P. patens*, the expression was performed in *Physcomitrella* directly. The nucleotide sequence was equipped with *C*-terminal 6xHis codons followed by the stop codon and restriction sites for NotI (5´) and XhoI (3´). Then the sequence was amplified by PCR (Fig. 87c) and entered into pENTR™1A. Again two fragments were amplified, the bigger one with the additional intron, but only the intron-free sequence was further used. After insertion of the sequence into pTHUbiGate the plasmid was transferred into *P. patens*. 14 stable transformants were obtained after double selection with hygromycin. All of them were checked for protein production by Western blot analysis after expression for 12 days. The untransformed Pp wildtype served as negative control. Figure 89 shows a strong signal at the expected size of around 60 kDa in transformants 3, 9, 11, 13 and 14 and a faint signal in 2 and 4. The wildtype did not show any stained protein. For further expression 9 and 13 were used. Protein isolation was carried out after protein expression for 12 d in BCD-media.

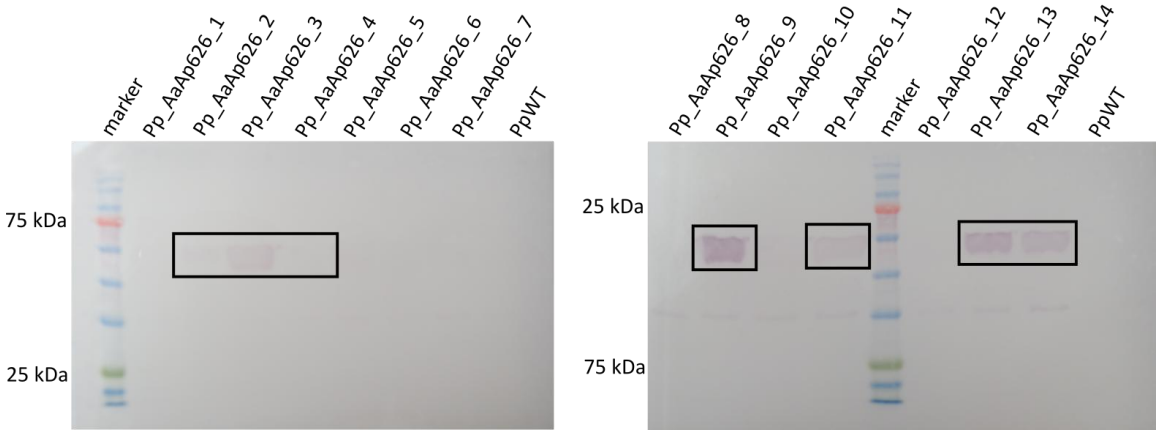


Figure 89 Western blot analysis of AaAp626 expressed in *P. patens* with anti-His-antibody. Crude protein extracts of the 14 stable transformants had a concentration between 1.0-2.3 mg/ml. Crude protein extract of PpWT had a concentration of 1.1 mg/ml. 20 µl of each sample was analyzed by SDS-PAGE.

Enzyme activity was checked using crude protein extract and the substrates galangin, kaempferol, naringenin and pC-PHPL (Fig. 90). Again Pp wildtype served as a control, since *Physcomitrella* likely has an own flavonoid 3'-hydroxylase. None of the added

substrates led to a higher product formation compared to the wildtype. This was also checked by LC-MS and none of the occurring masses corresponded to an expected product.

Until now, the function of this protein is unknown and some further substrate testing is necessary.

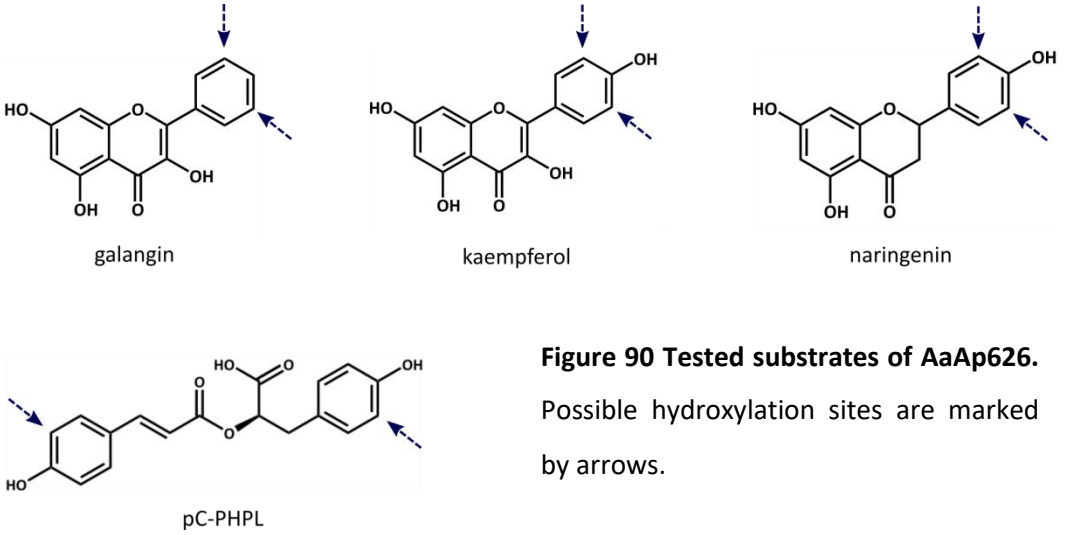


Figure 90 Tested substrates of AaAp626. Possible hydroxylation sites are marked by arrows.

IV Discussion

The production of phenolic compounds had enormous advantages for plants in conquering dry habitats. Phenolic compounds can act for example as natural sun screens or cell wall reinforcements (Rensing 2018). Many of them are derived from the phenylpropanoid pathway. So far hornworts were the only group of the bryophytes in which no enzyme of the phenylpropanoid pathway had yet been isolated and characterized (de Vries et al. 2017; Renault et al. 2019). Only the first two enzymes PAL and C4H have been characterized in *Anthoceros agrestis* suspension cultures (Petersen 2003; Pezeshki 2016).

1 Sequence analysis and phylogeny

Genes encoding a PAL have already putatively been identified in microorganisms (Barros et al. 2016) and clear evidence of plant *PAL* were found in Streptophyta (de Vries et al. 2017; Renault et al. 2019). Currently, experiments with a heterologously expressed PAL from *Anthoceros* are carried out in our working group (Pezeshki et al., unpublished data).

In this work a sequence encoding a cinnamic acid 4-hydroxylase was identified. C4H belongs to the cytochrome P450 enzymes (CYP73) and is located in the plant ER (Hannemann et al. 2007; Gamenara et al. 2013). The evolutionarily first *C4H* homologs were found in bryophytes, while they were absent in green algae including Klebsormidiales and Zygnematales (Ehlting et al. 2006; Renault et al. 2019). Furthermore, a duplication event occurred prior to the split between gymnosperms and angiosperms, dividing *CYP73* genes into class I and II. Due to a gene loss, class II *C4H*s are not found in some conifers and Brassicaceae. *C4H*s from bryophytes, lycophytes and monilophytes could be assigned to neither of the two groups. In bryophytes C4Hs from *P. patens* (Pp3c4_21680V3.1) and the liverworts *Marchantia* and *Plagiochasma* have already been heterologously expressed (Liu et al. 2017; Renault et al. 2017b). In this thesis, an alignment of different C4Hs with AaC4H (Fig. 14) showed a high identity of the conserved domains throughout the plant kingdom (Hasemann et al. 1995; Halkier 1996; Werck-Reichhart et al. 2002). Phylogenetic analysis of different C4H amino acid

sequences suggested the existence of two major clades (Fig. 15). The first comprised liverworts and mosses and the second included sequences from lycophytes, hornworts, ferns, gymnosperms and angiosperms. This rather confirms the assumption that hornworts are the youngest group of the bryophytes (Qiu et al. 2006; Ligrone et al. 2012a; Ruhfel et al. 2014), but does only partly support the statement that bryophytes might have developed independently from vascular plants (Renzaglia et al. 2000; Nishiyama et al. 2004; Wickett et al. 2014).

The second characterized enzyme in this work was a NADPH-dependent CPR. This enzyme transports two electrons from NADPH to the P450 and is also located in the ER. CPRs can be found ubiquitously in living organisms and together with the P450 they can be divided into 9 classes based on their partners present during the electron transfer (Hannemann et al. 2007). In plants, they mostly belong to the second class, consisting of two membrane-bound proteins, the P450 and a CPR. All functional motifs of CPRs are conserved throughout the plant kingdom, based on an alignment of AaCPR with other plant CPRs (Fig. 32). Often, the membrane anchor in the 5´-region was slightly shorter in mosses, hornworts and ferns compared to gymnosperms and angiosperms. Ro et al. (2002) divided plant CPRs into two classes, based on the length and amino acid conservation of the *N*-terminal amino acid sequence. AaCPR could not be assigned to either of the two classes. This opens up the possibilty of more than two classes of membrane anchors occurring in plant CPRs. In the phylogenetic analysis of different CPR amino acid sequences the branching appeared different compared to the C4Hs. Here, moss CPRs have their own clade, while in the other clade, starting with liverworts, the branches are deverging from another up to monocotyledons and dicotyledons (Fig. 33).

The next step in the phenylpropanoid pathway is the activation of 4-coumaric acid by a 4CL. Homologs of *4CL* were already found in the green algae *Mesostigma* and *Klebsormidium* (de Vries et al. 2017; Renault et al. 2019). Four genes from *Physcomitrella patens* encoding a putative 4CL have been heterologously expressed. They converted preferably 4-coumaric acid, but also accepted cinnamic acid, caffeic acid and ferulic acid (Silber et al. 2008). Scaffold 6378 from the *Anthoceros punctatus* database was used as a template to identify a gene encoding Aa4CL. Two potential start codons were found after RACE-PCR. After an alignment with different plant 4CLs it was

assumed that the shorter sequence (Aa4CL_2) was the actual coding sequence (Fig. 42). Aa4CL had a high identity to other 4CLs in the conserved sequence motifs of box I and box II (Conti et al. 1997; Stuible and Kombrink 2001). Even the 12 amino acids, which are supposed to act as the 4CL substrate specificity code, were always identical (Schneider et al. 2003). Again the phylogenetic analysis, using the maximum-likelihood method of the MEGA X software package, revealed *Anthoceros* as the youngest group of bryophytes according to the 4CL protein sequences (Fig. 43). There were again two major clades. The first is divided into two groups, the lower one starts with Physomitrella and the branches are deverging from another in order from moss, lycophyte, hornwort to gymnosperms. Interestingly, the lycophyte sequence was placed between the sequences from mosses and hornworts; this supports the results published by Li et al. (2015), that the sequences of bryophytes and lycophytes are closely related. In the upper part angiosperms are deverging from the fern *Dryopteris*. The second branch contains only sequences from angiosperms which supports the thesis of a duplication event occurring in seed plants before the split of gymnosperms and angiosperms (Li et al. 2015).

A second enzyme with 4-coumaric acid activating properties was found with Aa4HBCL. Plant CoA-ligases activating benzoic acid derivatives are to date rarely characterized. Genes encoding benzoate:CoA-ligases (BCL) were identified in *Arabidopsis thaliana* and *Clarkia breweri* (Beuerle and Pichersky 2002; Kliebenstein et al. 2007). The BCL from *Arabidopsis* is supposed to be involved in glucosinolate biosynthesis. A cinnamoyl CoA-ligase with benzoyl-CoA forming properties was characterized from *Petunia hybrida* (Klempien et al. 2012). Two other enzymes were characterized biochemically: a 3-hydroxybenzoate coenzyme A ligase from *Centaurium erythraea* acts in xanthone biosynthesis (Barillas and Beerhues 1997) and a BCL from *Pyrus pyrifolia* contributes to the biosynthesis of biphenyl phytoalexins (Saini et al. 2020). The amino acid sequence of Aa4HBCL from *Anthoceros agrestis* displayed the highest identities to 4CL-like sequences from the lycophyte *Selaginella* and the moss *Physcomitrella* (Table 13). Heterologous expression revealed that this enzyme actually preferred benzoic acid derivatives over cinnamic acid derivatives. The alignment of Aa4HBCL with other plant 4CLs and (hydroxy)benzoic acid CoA-ligases (Fig. 52) demonstrated that the sequence motifs of

box I and box II were still highly conserved within CoA-ligases with different functions (Conti et al. 1997; Stuible and Kombrink 2001). This is partly also reflected in the 4CL substrate specificity code (Schneider et al. 2003). While most of the 12 amino acids are conserved in Aa4HBCL and the 4CLs, three changes are very conspicuous: N288T, G380A and V390I or V390A. These changes were observed in all of the three aligned benzoic acid activating enzymes. Therefore it could be possible, that this is enough to change substrate preference. By site-specific mutagenesis this hypothesis could easily be verified in future experiments.

After activation of 4-coumaric acid, the 4-coumaroyl moiety of pC-CoA is often coupled with another substrate. This is catalyzed by hydroxycinnamoyltransferase (HCT). At present, HCTs from *Anthoceros agrestis* are identified and characterized in our group by Lucien Ernst and Maike Petersen. In the plant kingdom, first identified HCT homologs – although without functional characterization – can be found in the green algae *Klebsormidium* and *Nitella* (de Vries et al. 2017; Renault et al. 2019).

Hydroxylation of HCT products by the cytochrome P450 enzyme CYP98 was also investigated in this work. The only CYP98 homolog in green algae was found in *Klebsormidium* (Renault et al. 2019) and a single homolog was identified in bryophytes, lycophytes, monilophytes and gymnosperms. In angiosperms, several copies can be found, due to many independent gene duplications and losses (Alber et al. 2019). A suitable CYP98 candidate gene was found with scaffold 26091 from the *A. agrestis* database. A BLASTp search (Table 17) revealed high identities to characterized CYP98s from *P. patens* (Renault et al. 2017a) and *P. taeda* (Anterola et al. 2002). As for AaC4H, sequence motifs generally found in plant P450 were also highly conserved in AaCYP98 (Fig. 72) (Hasemann et al. 1995; Halkier 1996; Werck-Reichhart et al. 2002). The phylogenetic analysis of various CYP98 amino acid sequences revealed that the bryophyte enzymes from *Anthoceros* and *Physcomitrella* (Renault et al. 2017a) were located on a separate branch (Fig. 73). This was also the case for the monocot enzymes (Morant et al. 2007; Escamilla-Trevino et al. 2014) and the two *Arabidopsis* CYP98 producing tricoumaroyl or triferuloyl spermidines (Matsuno et al. 2009).

Another potential CYP98 candidate sequence was *AaAp626*. The sequence was obtained by alignment of *AaCYP98* in the *Anthoceros punctatus* database. After RACE-PCR, protein BLAST of the full-length sequence showed, that this protein might be a flavonoid 3-monooxygenase (F3H) instead of a CYP98 (Table 19). Both enzymes belong to the largest set of plant P450s, the CYP71 clan. Nevertheless, the alignment affirmed the theory of Alber et al. (2019), that only one CYP98 homolog can be found in bryophytes and gymnosperms. All conserved motifs of plant P450 were found in the protein sequence of AaAp626 (Fig. 88) (Hasemann et al. 1995; Halkier 1996; Werck-Reichhart et al. 2002).

At last also another potential 4CL candidate sequence was found in *Aa20832*. The amino acid sequence of Aa20832 was aligned with Aa4CL, Aa4HBCL and the two protein sequences with the highest identities in a BLASTp search from *P. patens* and *S. moellendorffii* (Fig. 62). These sequences were both annotated as 4CL-like. As for Aa4CL, Aa20832 had two potential start codons, but the shorter sequence had a better fit to the other aligned sequences. While amino acids in the motif box I were almost identical for all analyzed sequences, some amino acids in box II differed in Aa20832 and the two 4CL-like sequences compared to Aa4CL and Aa4HBCL. For example the cysteine in position 501 was a tryptophan in Aa20832 and the other two 4CL-like sequences. Except for the first two of the six amino acids in box II, all varied in at least one of the five CoA-ligase sequences. Moreover, six of the 12 amino acids encoding the 4CL substrate specificity code differed to Aa4CL, but these changes were not explicitly the same in all of the 4CL-like sequences. Unlike Aa4CL and Aa4HBCL, the other three sequences shared a peroxisomal signal sequence at the *C*-terminus. All of them could be assigned to the peroxisomal sequence 1 (PTS1) (Reumann 2004; Reumann et al. 2016). The last three amino acids of each sequence belonged to the major canonical PTS1 tripeptides. Secondary structure prediction with two different programs (Phyre^2 and predictprotein.org) revealed the presence of at least one transmembrane helix and confirmed the localization in peroxisomes (Fig. 69). Both predictions forecasted a transmembrane helix around the amino acids 282-298. Phyre^2 additionally predicted another one between amino acids 129-137. In *Arabidopsis* two long-chain acyl-CoA synthetases were identified, both bound to peroxisomal membranes and involved in the oxidation of fatty acids (Fulda et al. 2002). Therefore it was not so surprising, that

Aa20832 also had a high identity (42 %) to the OPC-8:0 CoA-ligase 1 isoform 1 from *Dorcoceras hygrometricum*. In *Arabidopsis*, three fatty acid or OPC:CoA-ligases (At1g20510, At4g05160 and At5g63380) were identified, all of them were localized in peroxisomes and contributed to jasmonic acid biosynthesis (Schneider et al. 2005; Koo et al. 2006; Kienow et al. 2008). Interestingly, At4g05160 and At5g63380 were, at first, wrongly annotated as 4CL-like proteins.

Except for the phylogenetic analysis of AaCYP98, all other ancestral trees support the thesis that hornworts are the youngest bryophyte (Nishiyama et al. 2004; Karol et al. 2010; Ruhfel et al. 2014). Nickrent et al. (2000), Nishiyama et al. (2004), Wickett et al. (2014) and many more share the opinion, that mosses and liverworts share a clade. This was supported by phylogenetic analysis of C4H and 4CL. For CPR all bryophytes starting with mosses emerge from another. No liverwort CYP98 was characterized to this date. The phylogenetic analysis of different CYP98 tends more to the thesis that bryophytes might have involved independently from vascular plants (Renzaglia et al. 2000; Nishiyama et al. 2004; Wickett et al. 2014). Although it is unknown, why the monocot CYP98 was on a separate branch. Only the phylogenetic tree of different C4Hs supported one of the three most likely trees of Puttick et al. (2018), with mosses and liverworts evolving independently and horworts being the youngest bryophyte.

2 Heterologous expression of cytochrome P450 monooxygenases and NADPH:cytochrome P450 reductase

2.1 Expression in *Saccharomyces cerevisiae*

2.1.1 AaC4H, AaCYP98 and AaCPR

For the cytochrome P450 enzymes AaC4H and AaCYP98, as well as AaCPR, heterologous expression was at first attempted in *S. cerevisiae*, since yeast is still the most often used expression system for plant membrane proteins (Renault et al. 2017a; Dai et al. 2019; Hausjell et al. 2019). For this, *AaC4H* and *AaCYP98* were ligated into the expression vector $pESC_{-ura}$ (with an integrated sequence encoding a *C*-terminal FLAG-Tag) and *AaCPR* was integrated into pYES2 after the sequence was equipped with a sequence translated to six histidine residues in front of the stop codon. pYES2 was chosen instead

of pESC$_{-ura}$ for *AaCPR*, because Western blot analysis with anti-FLAG-antibody led to many unspecific bands. Different yeast strains (BY4741, BY4742 and CB018) were transformed with these plasmids. Although all genes were verified in the yeast strains by colony-PCR, none of them showed any signal of expression in Western blot analysis. Moreover, no enzyme activity was detectable. For AaCYP98 this was assumed, since only pC-PHPL and Caf-PHPL were tested and only pC-PHPL was later accepted to a very low extent by the codon optimized AaCYP98 in combination with CbCPR. In addition AaC4H and AaCYP98 were tested after extraction of the microsomal fraction. It was later revealed, that the heterologously expressed proteins lost a high percentage of their activity using the membrane fraction. Only for AaCPR crude protein extract was tested, but no difference was observed compared to the empty vector control. There are many reasons why expression of the proteins was not working in yeast. One reason might be that the codon usage of *A. agrestis* differed to *S. cervisiae*. The codon usage of *AaC4H*, *AaCYP98* and *AaCPR* was compared to the one in yeast (codon usage database: www.kazusa.or.jp). For example, *Anthoceros* preferred the codon triplet GAG with a frequency of 52/1000 while GAA was only used in 13/1000. For yeast, it is exactly the opposite, with 19/1000 for GAG and 47/1000 for GAA. In yeast, the most often used triplets for leucine are TTG (27/1000) and TTA (26/1000). *Anthoceros* almost not uses TTA (0.6/1000) but instead prefers CTG (25/1000). The different codon usage could also be related to the higher GC content of the hornwort genes. While yeast has an average GC content of 39 % (Nakase and Komagata 1971), *AaC4H* had 57 %, *AaCYP98* 62 % and *AaCPR* 53 %. Nevertheless it is possible that a small amount of heterologous protein might have been present, which, however, could not be detected.

2.1.2 Codon optimized AaCYP98

To overcome this issue, codon optimization was one method of choice (Perlak et al. 1991; Haas et al. 1996; Rouwendal et al. 1997). This was done for *AaCYP98* by the company GeneCust. Thus, the GC-content was adjusted from 62 % to 36 % in the *coAaCYP98* nucleotide sequence (Fig. 79). This sequence was integrated into the MCSI of the expression vector pESC$_{-ura}$. In addition, the *CPR* from *Coleus blumei* was inserted into the MCSII of pESC$_{-ura}$ or MCSII was left empty. The *CPR* from *Coleus blumei* was chosen, since heterologous expression of AaCPR failed in yeast without codon optimization and

CbCPR had already been heterologously expressed (in yeast) and characterized in our group by Eberle et al. (2009). Therefore a plasmid with *CbCPR* was already available and only the restriction sites had to be changed in order to subclone the sequence into $pESC_{-ura}$. Four different yeast strains were transformed with the plasmid carrying only *coAaCYP98* (BY4741, BY4742, CB018 and InvSc1) and two strains (BY4742 and InvSc1) were transformed with the plasmid carrying *coAaCYP98+CbCPR*. In both cases coAaCYP98 could be detected after Western blotting with anti-FLAG-antibody (Fig. 80 and 81). For both plasmids, expression in BY4742 worked best. In that sense, codon optimization really was beneficial for protein expression. Assays were performed mostly using crude protein extract. Unlike soluble proteins, purification (for example by metal chelate chromatography) of membrane bound proteins is a tough undertaking. In most cases a solubilisation of the protein from the microsomal fraction with detergents is necessary. Then the enzyme has to be purified and afterwards integrated into artificial membranes. However, this often leads to loss of function (Gabriac et al. 1991; Durst and O'Keefe 1995). Because of this, purification was not attempted. Another problem occured during the isolation of the microsomal fraction, but this topic will be discussed later. Activity of coAaCYP98 without CbCPR could only be determined for pC-3OH-An. By expression of coAaCYP98 in combination with CbCPR activity of coAaCYP98 could additionally be determined for the substrates pC-An, pC-Ty, pC-Shik and pC-PHPL (Fig. 83 and 84). Increasing hydroxylation activity of heterologously expressed P450s by coexpression of a plant CPR was frequently observed (Ro et al. 2002; Kim et al. 2005; Jennewein et al. 2005; Eberle et al. 2009). For pC-3OH-An the formation of Caf-3OH-An was increased 13-fold in coAaCYP98+CbCPR compared to coAaCYP98. A possible explanation could be that the yeast CPRs were fully occupied and could not transfer sufficient electrons to the overexpressed P450 or the interaction of plant P450 to a yeast CPR did not work correctly. The additional plant CPR was able to overcome both assumed problems. pC-3OH-An, pC-Shik and pC-Quin were obtained by enzyme assays from HCTs from *A. agrestis* and *Sarcandra glabra* (identified and characterized by Paul Bömeke, Lucien Ernst and Maike Petersen). Since pC-Quin was not accepted by coAaCYP98+CbCPR, purification of pC-Quin for characterization was unnecessary. pC-Shik was isolated and purified by Olga Haag in a small amount. AaCYP98 presumably showed the highest conversion with the substrate pC-3OH-An. Until now, purification of

of high amounts of pC-3OH-An failed. Therefore unfortunately no quantitative data for the activity with this substrate can be given here. pC-An and pC-Ty were chemically synthesized. While only one step was necessary for pC-Ty, pC-An was synthesized by several steps essentially according to Alber et al. (2019). In the future, pC-3OH-An may also be produced by a similar chemical synthesis like pC-An.

AaCYP98 displayed a similar substrate preference as the CYP98 from the moss *P. patens* (Alber et al. 2019). PpCYP98 also had the highest identity to AaCYP98 in a protein BLAST search (Table 17). According to Alber et al. (2019), moss, fern and lycopod CYP98 mostly favoured pC-An. In this work it was proven, that pC-3OH-An might even be a better substrate, but this substrate was not tested by Alber et al. (2019). In *A. agrestis* (characterized by Petersen and Ernst in our group, unpublished data) as well as *P. patens* (Eudes et al. 2016) a HCT was found, preferably producing pC-3OH-An but lacking activity for the production of pC-An. The same is true for HST from *Coleus blumei* which, besides shikimate, accepted 3-hydroxyanthranilate but not anthranilate (Sander and Petersen 2011). Therefore a higher activity of AaCYP98 towards pC-3OH-An was not unexpected. In plants, caffeoylanthranilates act for example as phytoalexins (Ishihara et al. 1999; Okazaki et al. 2004; Ahuja et al. 2012). The respective compounds have, however, not been described to be present in *Anthoceros agrestis* (Trennheuser 1992).

While coAaCYP98 did not show any conversion of pC-2-Th, PpCYP98 did convert this substrate in small amounts (Renault et al. 2017a; Alber et al. 2019). Knockout of PpCYP98 revealed caffeoylthreonate to be the major metabolite missing, while anthranilate conjugates were not detected in either the knockout mutant or PpWT (Renault et al. 2017a). Therefore it is unknown, which role anthranilate conjugates might play in bryophytes. Both *Anthoceros* and *Physcomitrella* HCTs produced pC-Shik (own unpublished data; Eudes et al. 2016) and the CYP98s were able to hydroxylate pC-Shik to Caf-Shik which is regarded a step in monolignol biosynthesis (Schoch et al. 2001). Nevertheless, true lignin is absent in bryophytes and only the pre-lignins caffeoylthreonic acid and caffeoylanthranilic acid were found in mosses and G- and GS-lignin-like units were detected in mosses and hornworts (Ligrone et al. 2008; Renault et al. 2017a; Renault et al. 2019). Since both AaCYP98 and PpCYP98 share a similar function

and PpCYP98 is considered to be involved in the biosynthesis of suberin and cutin, this might also be the case for AaCYP98 (Renault et al. 2017a).

The kinetic values of coAaCYP98+CbCPR were determined for the chemically synthesized substrates pC-An and pC-Ty and, moreover, for the cosubstrate NADPH (Fig. 86;Table 18). AaCYP98 had a higher affinity for pC-An with a K_m-value of 6.4 ± 0.8 µM. The K_m-value for pC-Ty was determined to be higher, namely at 36.4 ± 3.6 µM. A comparison with other plant CYP98 was unfortunately not possible, as there are currently no kinetic values available for these substrates. Only the K_m-value of NADPH (53.2 ± 4.0 µM) was compared to the heterologously expressed CbCYP98 (Eberle et al. 2009). In this study a coexpression with CbCPR was also conducted. Therefore both enzymes were (for the most part, because expression was taken out in different yeast strains) supplied by the same CPR. This was also demonstrated by the same K_m-value of CbCPR of 50 µM for NADPH (Eberle et al. 2009). This publication also showed that the additional plant CbCPR had a much higher affinity for NADPH than the yeast CPRs with 215 µM. NADPH is the preferred electron donor for the majority of cytochrome P450 enzymes. Using NADH or NADP instead of NADPH as cosubstrate led to only 5-10 % product formation in coAaCYP98+CbCPR assays incubated with pC-An or pC-3OH-An.

For pC-3OH-An and pC-Shik kinetic values still have to be determined. Since isolation of pC-Shik from HCT enzyme assays generally is no problem, this will be solved in time. For pC-3OH-An the next step will be the chemical synthesis, since isolation from enzyme assays was problematic, because the substance was not stable (data not shown). Even though pC-PHPL was available as a purified substance, determination of K_m-values for pC-PHPL will be difficult, because product formation generally was very low. Moreover, it was not quite clear, which substance (Caf-PHPL or pC-DHPL) was produced. Both as well as RA can theoretically be formed by CYP98. This has to be clarified after assay optimization.

2.1.3 Chimeric CbCPR-AaCPR

For heterologous expression of AaCPR in *S. cerevisiae* a different approach was chosen. This was based on the publication of Batard et al. (2000). This group also had problems with expression of monocot P450s and CPR in yeast because of the high GC-content.

They were able to overcome this issue by self codon optimization of the first 18, 39 or 111 bp with the help of a megaprimer. Decreasing the GC-content in the 5´-end resulted in a stronger expression in yeast. Especially in the *AaCPR* nucleotide sequence, the first 100 bp displayed an even higher GC-content (67 %) as already obvious in the full-length sequence (53 %). Thus, this seemed to be a reasonable approach for the expression in yeast. However, the 5'-optimization was solved in a different way: Instead of using a megaprimer, the first 277 bp of *AaCPR* were exchanged with 382 bp from the *CPR* of *Coleus blumei* (*CbCPR*) (Fig. 7). CbCPR had already successfully been expressed in yeast by Eberle et al. (2009). Both sequences shared a restriction site for EcoRV between the highly conserved FMN and P450-binding regions. As a result, the GC content of the first 100 base pairs was reduced from the original 67 % to 46 %. Afterwards the nucleotide sequence of the chimeric CPR, still equipped with a sequence encoding a *C*-terminal 6xHis-tag, was again inserted into pYES2 and the yeast strain CB018 was transformed. Although the chimeric gene was identified by colony-PCR, unfortunately, the protein was not detectable by Western blots. Moreover, the crude protein extract of the chimeric CPR transformant displayed the same CPR activity as the crude protein extract of the empty vector control (Fig. 35). Therefore the question arises whether a decreased GC-content in the 5´-region might not have been sufficient for a successful expression of AaCPR. This experiment was only conducted using the yeast strain CB018. As already seen for coAaCYP98, CB018 was not ideal for expression, therefore other yeast strains should have been tested. However, heterologous expression of AaCPR was solved in a different way.

2.2 Expression in *E. coli*

For AaC4H and AaCPR, a new expression organism was considered. Expression in *E. coli* seemed to be an alternative, since it offered a very fast and cheap approach of protein production. Furthermore, several researchers could already demonstrate *E. coli* as a suitable expression host for CYPs and CPR (Hotze et al. 1995; Leonard et al. 2005; Leonard et al. 2007; Quinlan et al. 2007). The disadvantage is that the monoxygenase cannot be expressed alone because of the lack of the membrane-bound P450 redox partner proteins in *E. coli*, therefore fusion with a (plant) CPR was essential. Different constructs were designed: AaC4H was generally fused with its *C*-terminus to AaCPR

(without its *N*-terminal membrane anchor), but differed in the length of the full-length (fus1) or 5'-truncated (fus2, fus3) AaC4H and the *N*-terminal addition of mammalian peptide ε in fus2 and fus3 (Fig. 8). These constructs were inserted into the expression vector pET15b and the *E. coli* strains BL21RIPL and SoluBL21 were transformed. First experiments with the fus-constructs showed no fusion protein of the expected size of 131.85 kDa (fus1), 132.23 kDa (fus2) and 128.79 kDa (fus3) on Western blots with anti-His-antibody, instead two fragments around 63 kDa and 75 kDa were visible (predominantly in BL21RIPL). While no C4H activity was detectable, an increased CPR activity was present compared to the empty vector control. This was observed for both *E. coli* strains, in crude protein extract from BL21RIPL there was no difference between the tested constructs, in SoluBL21 fus3 displayed a higher activity than fus1. After these experiments it was noticed that the ribosome binding site (rbs) in front of the transcript was missing due to the use of XbaI. Nevertheless, all of the *E. coli* strains were able to express a soluble AaCPR. Therefore the constructs were checked for the same rbs sequence (like the one on pET15b). The fusion constructs comprised four rbs sequences. Only two of them led to a protein in the correct reading frame, the other two were followed by an early stop codon and led only to short peptides. The two proteins had calculated masses of 88 and 62 kDa. The bands on the Western blot of BL21RIPL could probably represent these proteins.

Constructs with the additional rbs neither displayed a C4H or CPR enzyme activity, nor showed a signal on Western blots after expression in BL21RIPL or SoluBL21. All experiments were repeated with the *E. coli* strain C41(DE3); this strain is described to be better suited for expression of toxic proteins. Protein detection with anti-His-antibody showed faint bands at the expected size of the fusion constructs and an additional band at approximately 75 kDa (Fig. 37). Again, higher CPR activity compared to the empty vector control was measurable. This was predominantly observed in the supernatant after separation of the membrane fraction. Only the membrane fraction of the fus3 transformant had a slightly higher activity than the empty vector control. The 75 kDa stained fragment on Western blots was again assigned to a soluble AaCPR and the marginally higher CPR activity of fus3 in the membrane fraction was supposed to be caused by the fusion protein. Moreover, protein extracts from C41(DE3) showed a low

conversion of *t*-cinnamic acid in C4H activity assays. C4H activity was increased using the membrane fraction. Here, the shortest construct (fus3) led to the best product yield, while no product was detectable for fus2 and the empty vector control (Fig. 20). In similar experiments of Leonard et al. (2007), the fusion protein corresponding to fus2 was the most successful. Therefore, the 5´-region of AaC4H might be less important for enzyme acitivity. The overall reactivity of the fusion protein, however, was comparatively low and optimization would be necessary in order to achieve higher hydroxylation activities.

The used *E. coli* strains all had different benefits: BL21RIPL contains extra copies of the argU, ileY, leuW and proL tRNA genes. This strain is used for expression of heterologous proteins from organisms that have either AT- or GC-rich genomes (Agilent, BL21-CodonPlus Competent Cells instruction manual). This was the case for enzymes from *A. agrestis*. The strain SoluBL21 was generated to overcome toxicity of proteins expressed in BL21(DE3) host strains, resulting in the production of the target proteins in a soluble form (amsbio, SoluBL21™Competent *E. coli* user manual). It is unknown, whether the inserted sequences had a toxic effect on the bacteria. The strain C41(DE3) is specialized for the effective production of toxic and membrane proteins (Lucigen, OverExpress® Chemically Competent cells instruction manual). This was the case for the fusion constructs. Thus it was reasonable, that the membrane-bound fusion proteins could only be expressed in C41(DE3). The soluble proteins Aa4HBCL and Aa20832 (discussed later) were only expressed in the *E. coli* strain SoluBL21 and no protein formation was detectable after transformation of *E. coli* BL21RIPL cells, unlike the expression of the C4H-CPR fusion proteins without the rbs, where expression was less in SoluBL21 compared to BL21RIPL. This demonstrated the importance choosing the right expression strain.

It is unknown, whether AaCPR was only expressed in C41(DE3) because of the internal rbs or wheter the CPR was cut off from the monooxygenase after translation. The second could have possibly been prevented by a linker sequence between AaC4H and AaCPR (Leonard et al. 2007). However, this had the advantage of a soluble CPR which could be purified by metal chelate chromatography. The purified soluble AaCPR displayed an apparent K_m-value of 9.2 ± 0.3 µM for NADPH and a V_{max} of 49.2 ±

3.9 mkat/kg (Fig. 39; Table 9). These values were comparable to those for *Gossypium hirsutum* (Gh) and *Withania somnifera* (Ws) CPRs which had also been purified and heterologously expressed in *E. coli* (Yang et al. 2010; Rana et al. 2013). The respective K_m-values for NADPH were 4.6 ± 0.2 µM (GhCPR1), 5.6 ± 0.6 µM (GhCPR2), 5.06 ± 0.30 µM (WsCPR1) and 6.48 ± 0.33 µM (WsCPR2). The V_{max}-values ranged between 50 and 180 mkat/kg.

2.3 Expression in *Physcomitrella patens*

Expression of the P450 enzymes AaC4H, AaCYP98 and AaAp626 in addition to AaCPR was also conducted in the moss *Physcomitrella patens*, because of its close relationship and moreover similar gene structure, similar GC-content and the ability of posttranslational modifications. As a novel expression system *Physcomitrella patens* has already been used for the heterologous expression of enzymes involved in plant specialised metabolism (Anterola et al. 2009; Bach et al. 2014; Zhan et al. 2014; Pan et al. 2015; Khairul Ikram et al. 2017). Due to the high rates of homologous recombination, stable transformants were obtained, making the permanent use of selective media unnecessary (Schaefer et al. 1991). Compared to *S. cerevisiae* and *E. coli*, *P. patens* had the disadvantage that the transformation and expression is time-consuming. Besides, *P. patens* had own homologs of the expressed proteins (as registered in Phytozome) and for this reason, all experiments were made with the *P. patens* transformants and the *Physcomitrella* wild-type in parallel. After the second selection three stable transformants were obtained for AaC4H, six for AaCYP98, fourteen for AaAp626 and five for AaCPR. But only a few transformants showed protein formation verified on Western blots, probably because the gene was not integrated and, unfortunately, they did not lose the plasmid after the second period of regeneration. To discriminate aginst these cells, cultures were were kept on media without selective pressure for a longer period of time, resulting in the loss of the plasmid (false positive transformant).

2.3.1 AaC4H

One stable transformant showing AaC4H protein formation was further characterized. This transformant produced at least double to triple amounts of 4-coumaric acid in *in vitro* assays compared to wildtype (Fig. 22). For *P. patens* six potenial C4H genes were

found in the genome (Phytozome). Only the activity of PpCYP73A48 (named PpC4H_2 in this work) was proven by Renault et al. (2017b). This was also the gene with the highest expression values displayed in Phytozome in all tested media and developmental stages. Moreover, two of the six genes were not expressed (*Pp3c13_14870V3.1.p* and *Pp3c12_6560V3.1.p*) and two other genes (*Pp3c16_23740V3.1.p* and *Pp3c3_17840V3.1.p*) displayed only a very low expression with less than 0.5 fragments per kilobase million (FPKM) on BCD under continuous light. Only *Pp3c25_10190V3.1.p* (named PpC4H_1 in this work) and PpC4H_2 displayed higher values with a FPKM of 2.9 and 60.4 (Phytozome).

A problem not only with AaC4H expressed in *P. patens* but also with coAaCYP98 expressed in *S. cerevisiae* was the loss of function after isolation of the membrane fractions (Fig. 24). First attempts for the preparation of microsomes were performed with either $MgCl_2$ or polyethylene glycol with NaCl to aggregate the microsomes which were then precipitated by centrifugation for 20 min at 60000 g (Pompon et al. 1996; Petersen 2003). Furthermore, microsomes were prepared by ultracentrifugation of the crude protein extract for 1 h at 100000 g (Benveniste et al. 1986; Abas and Luschnig 2010). At last, the membrane fraction was isolated at 21000 g according to the MCF protocol of Abas and Luschnig (2010). All methods used had the same effect on AaC4H and AaCYP98: Both enzymes had only 25 % activity left in the microsomal fraction compared to the crude protein extract in *in vitro* assays. For both proteins it was proven, however, that they were located in the membrane fraction. To restore the function of Pp_AaC4H different experiments were carried out. C4H assays using microsomes from Pp_AaC4H in combination with either Pp_AaC4H crude protein extract/supernatant or Pp_AaCPR microsomes/supernatant led to no change. A decrease of activity was observed in the microsomes after using increasing centrifugation speeds (5000 to 60000 g after precipitation with $MgCl_2$). Different methods of resuspending the membranes had no effect. It seemed, that the protein was irreversibly denatured. Own experimental failure was excluded after the result could be reproduced by Maike Petersen (data not shown) and microsomes isolated directly from *Anthoceros agrestis* also showed a reduced activity (around 40 %) compared to the crude protein extract in a C4H activity assay. It was not tested, whether this effect occurred only using enzymes

from bryophytes or whether it was a problem of all plant P450s. Thus it remains unclear, why nearly all publications concerning plant P450 work with the microsomal fraction in *in vitro* assays (e.g. Sullivan and Zarnowski 2010; Liu et al. 2017; Renault et al. 2017a; Alber et al. 2019). Of course, specific activity was increasing after removal of the soluble proteins, but since the P450s were not pure, specific activity was still calculated based on a broad range of membrane proteins. For this reason all characterizations in this work were carried out with crude protein extract instead of the microsomal fraction.

The pH-optimum of both Pp_AaC4H and PpWT was determined to be at pH 7 (Fig. 25). Typical for C4H (Werck-Reichhart 1995) and the majority of other cytochrome P450 enzymes from higher plants is a range between pH 7.0 to 7.5. Also the temperature optimum was in the expected range of C4H (25-30 °C) with 25 °C for Pp_AaC4H and PpWT. Both values of Pp_AaC4H correspond to the results for C4H measured in microsome fractions from *Anthoceros agrestis* suspension cultures (Petersen 2003). The K_m for cinnamic acid was at 17.3 ± 2.5 µM for Pp_AaC4H and 25.1 ± 4.0 µM for Pp_WT showing a higher affinity of AaC4H for cinnamic acid than the PpC4H(s) (Fig. 26; Table 5). For most C4Hs the apparent K_m for cinnamic acid ranges between 2 and 30 µM (Werck-Reichhart 1995). For C4H(s) from *Anthoceros agrestis* suspension cultures the measured K_m-value of cinnamic acid was 5 µM (Petersen 2003). For *A. agrestis* it is unknown, whether more than one gene encoding a C4H is present. Three heterologously expressed C4Hs from liverworts displayed K_m-values ranging between 0.7 and 1.7 µM for cinnamic acid (Liu et al. 2017). The liverwort enzymes also had some activity towards 3-hydroxycinnamic acid resulting in the formation of caffeic acid. This was also tested for Pp_AaC4H and PpWT. The crude protein extract was incubated with 4-coumaric acid to produce caffeic acid. The product peak had only an area of 0.7 % compared to the incubation with cinnamic acid and was only detectable in PpWT. Moreover, 3-coumaric acid was tested. This led to a caffeic acid concentration of 0.06 nmol/assay after 60 min for both Pp_AaC4H and PpWT. Therefore it was assumed, that this reaction had to be attributed to a *Physcomitrella* enzyme and not to the additional AaC4H. Benzoic acid was not converted to 4-hydroxybenzoic acid by either Pp_AaC4H or PpWT.

Using NADH instead of NADPH resulted in a formation of only 20 % 4-coumaric acid. NADH as a comparatively potent electon donor has been reported for the C4H-CPR

enzyme complex from *Anthoceros agrestis* (isolated from suspension cultures) (Petersen 2003). Both, Pp_AaC4H and Pp_WT displayed almost the same K_m-value for NADPH with 88.0 ± 9.8 µM (Pp_AaC4H) and 92.3 ± 5.4 µM (PpWT) (Fig. 26; Table 5). This was expected since the CPRs from *P. patens* provided the electrons also for AaC4H. This also indicates that the additional AaC4H worked freely with the PpCPRs and the additional enzyme had no negative impact on the P450 redox partner. In suspension cultures of *A. agrestis* the C4H(s) displayed a K_m of 60 µM for NADPH (Petersen 2003).

Tissue samples of Pp_AaC4H and PpWT were collected over three weeks and fresh weight, pH value of the remaining medium and phenolic content was determined (Fig. 27). Over three weeks the fresh weight increased to a final weight of 1.2 to 1.4 g/flask for both cultures. Therefore the weight was doubled almost every second day and even after three weeks the fresh weight still increased. This illustrated how sparingly the cells use the nutrients in the medium and effectively reproduce by photosynthesis, since the BCD medium contained only some macro- and microelements but no sugar. Cove (2005) described that *Physcomitrella* cultures can grow for at least two years without aging, but without light they die within a few months. The pH-value of the medium increased from pH 6.5 to 7.8 in Pp_AaC4H and 7.1 in PpWT. At first the pH dropped after two days in both cultures, since phosphate was taken up by the cells.

After 0, 7, 14 and 21 days the phenolic content was determined, but the amount barely changed and was always around 650-950 µg/g fresh weight for both Pp_AaC4H and PpWT (Fig. 27). Thus, the introduction of an additional active C4H did not result in a higher content of phenolic compounds. A similar effect was observed in transgenic tobacco plants (Sewalt et al. 1997; Blount et al. 2000). While down-regulation of C4H resulted in a reduced accumulation of caffeic acid esters, overexpression of C4H, on the other hand, did not result in an increased accumulation of phenolic compounds and lignins. The overall phenolic content was rather low compared to other samples from medicinal plants ranging from 2100 to 3700 µg/g of fresh weight (Kaur and Mondal 2014).

The tissue samples of Pp_AaC4H were used for quantitative real-time PCR determinations targeting *AaC4H*, *PpC4H_1*, *PpC4H_2* and the reference gene *St-P 2a*

(used for normalisation) (Fig. 30). Although primer sets were already available for potential reference genes from *Physcomitrella* (Le Bail et al. 2013), *St-P 2a* was the only fragment amplified, that did not show any unspecific amplicons. *PpC4H_1* was constantly expressed at a low level, which reflected the expression data in different media available on Phytozome. AaC4H and *PpC4H_2* had similar expression patterns, both having their maximum expression between day 4 to 6. Afterwards both rates decreased again and reached the same level as on day 0 after 10 (*AaC4H*) and 12 days (*PpC4H_2*). For *PpC4H_2* this corresponds to the expression data provided by Phytozome, that expression is high in protonema cells and comparably low in the gametophyte. It seemed, that after 12 days protonemata had completely transformed into gametophyte tissue. Thus, the same could have happened with *AaC4H*. If so, it would be interesting to know, whether this effect could also be observed in *Anthoceros agrestis* itself, since the protonema phase in hornworts covers only a marginal period of the developmental cycle compared to mosses (Renzaglia et al. 2000). Compared to *PpC4H_1*, both *AaC4H* and *PpC4H_2* were always present at a higher level. *PpC4H_2* mRNA was present fourfold compared to *PpC4H_1*, which also represents the higher expression rate of *PpC4H_2* displayed in Phytozome. mRNA of *AaC4H* was present at least 270-fold compared to *PpC4H_1*. The highest difference was observed after 4 and 19 days with 2000- to 4000-fold. In comparison to the two tested *Physcomitrella C4H*s *AaC4H* had extremely high expression levels, which reflects the effectiveness of the maize ubiquitin promoter (present on the expression vector) even in a lower plant (Schaefer 2002). However, the very high expression levels were not reflected in enzyme activity data. Here the differences between PpWT and Pp_AaC4H were considerably lower. This could be explained either by low translation rates of the *AaC4H*-mRNA or by the formation of non-functional AaC4H proteins. Another reason might be the restriction of the electron transfer capacity of the *Physcomitrella* CPR. The same was already observed with coAaCYP98 expression in yeast. Here, activity was increased 13-fold after coexpression of CbCPR. To overcome this issue, expression rates of CPR could be increased – either from *Physcomitella* itself or from *Anthoceros agrestis* – by using strong promoters for these genes as well.

2.3.2 AaCYP98

As for AaC4H, also AaCYP98 was expressed in *P. patens*. These experiments were conducted before the codon optimization of AaCYP98. Two stable transformants showed AaCYP98 protein formation and transformant 3 was used for further characterization. At first, tissue samples of Pp_AaCYP98 were collected over three weeks. As for Pp_AaC4H, Pp_AaCYP98 displayed no difference to the PpWT concerning fresh weight and the pH of the medium (Fig. 77). Since all measured parameters were essentially the same in Pp_AaC4H, Pp_AaCYP98 and PpWT, the additional genes might have had no negative effects on the plant. After protein isolation Pp_AaCYP98 was tested with the substrates pC-PHPL, Caf-PHPL, pC-Shik and pC-2-Th. For some substrates, e.g. Caf-PHPL and pC-Shik, it was observed, that the substrates just decreased over time (Fig. 78). It is known, that hydroxycinnamic acid esters can be enzymatically cleaved by esterases or the reverse reaction of HCTs (Hoffmann et al. 2004; Vanholme et al. 2013; Ha et al. 2016; Werner 2018). The same effect was observed for Caf-2-Th and Caf-4-Th, naturally occurring substances in the crude protein extract of *P. patens* and products of PpCYP98 (Renault et al. 2017a; Alber et al. 2019) as well as potential products of AaCYP98. Interestingly, an additional peak appeared at the same retention time for all tested substrates. Although this peak was also present in PpWT, the production was higher using Pp_AaCYP98. This peak also occurred after incubation of Pp_AaCYP98 crude protein extract with NADPH without any substrate. After LC-MS analysis two potential masses were detected. The first, m/z 308.04 [M-H], would correspond to the molecular formula of $C_{15}H_3N_9$, $C_{16}H_9N_2O_5$ or $C_{14}H_7N_5O_4$. The second mass peak, m/z 444.08 [M-H] might correspond to a molecular formula of $C_{18}H_{15}N_5O_9$ or $C_{32}H_{13}O_3$. All molecules could not be assigned and therefore it is unknown which substance was produced. Since many issues appeared over time using *P. patens* as the expression host for AaCYP98, expression was attempted again in yeast with a codon optimized AaCYP98, as already discussed above.

2.3.3 AaAp626

Another potential CYP98 candidate sequence (*AaAp626*) had a higher identity to flavonoid 3´-monooxygenases. Although this enzyme might not be involved in the

phenylpropanoid pathway, an attempt was nevertheless made to express the enzyme heterologously. Since expression in yeast without codon optimization was unsuccessful for AaC4H and AaCYP98, AaAp626 was directly expressed in *Physcomitrella patens*. Seven of the fourteen stable transformants showed a signal after detection with anti-His-antibody (Fig. 89). Crude protein extract of Pp_AaAp626 incubated with the substrates galangin, kaempferol, naringenin and pC-PHPL did not result in product formation compared to PpWT. Flavonoid 3´-hydroxylases belong to the CYP75 clan (Castellarin et al. 2006). CYP75A are flavonoid 3´,5´-hydroxylases while CYP75B are flavonoid 3´-hydroxylases and both P450s use dihydroflavonols as substrate (Nelson 2006b). Nelson (2006b) furthermore claimed, that the CYP75, as well as the CYP81 and CYP93 families are not present in the moss *Physcomitrella* and first homologs of CYP75 were found in pine. But he did not exclude CYP75 in other bryophytes, since flavonoids were found in some bryophytes (Asakawa 2000; Asakawa et al. 2013). Besides, in the meantime 19 or respectively nine potential candidate genes encoding a flavonoid 3´-monoxygenase or flavonoid 3´,5´-hydroxylase were identified in *Physcomitrella patens* (Phytozome, Rensing et al. 2008) and 21 or respectively 11 candidate genes were identified in *Anthoceros angustus* (Zhang et al. 2020). Until now, flavonoids were not detected in hornworts (Yonekura-Sakakibara et al. 2019; Davies et al. 2020). Thus, activity of AaAp626 towards the tested flavonoids would have been surprising.

Protein BLAST also showed a high identity of AaAp626 to the CYP71A1 clan (Table 19). CYP71A1 is supposed to play a role in the oxidation of monoterpenoids (Hamberger and Bak 2013). Functionally characterized enzymes from *Catharanthus roseus* and *Nepeta racemosa* catalyzed the hydroxylation of geraniol and nerol at position C-10 (Meehan and Coscia 1973; Hallahan et al. 1994). A CYP71A1 homolog of *Persea americana* catalysed 2,3- or 6,7-epoxidation of the same substrates (Hallahan et al. 1994). In addition, CYP71A1 from avocado had a *p*-chloro-*N*-methylaniline demethylase acitivity (Bozak et al. 1992). In rice CYP71A1 encodes a tryptamine 5-hydroxylase, which catalyses conversion of tryptamine to serotonin (Lu et al. 2018). In *Anthoceros angustus* 11 genes were identified as CYP71A1-like (Zhang et al. 2020).

Hence, AaAp626 still needs to be tested with many different substrates and it should be considered whether a heterologous expression in yeast with a codon optimized

sequence should be performed to exclude a simultaneous reaction of moss proteins. Thus, the activity could probably be attributed exclusively to AaAp626.

2.3.4 AaCPR

At last, also AaCPR was expressed in *P. patens*. One of the three stable transformants showing AaCPR protein formation was further characterized. However, Pp_AaCPR displayed no higher activity compared to PpWT (Fig. 40). This was attributed to the fact that *P. patens* has four potential CPRs (Phytozome: Pp3c20_9680V3.1; Pp3c24_17560V3.1; Pp3c8_19940V3.1 and Pp3c14_22890V3.1). Under continuous light, all of them show a FPKM at least over 1 in BCD. Pp3c20_9680V3.1 displayed the highest expression with a FPKM of 10 (Phytozome). For this reason, the difference in activity of the additional AaCPR might have been too low compared to the *Physcomitrella* CPRs. Due to unknown reasons, absorption decreased in cytochrome c reductase assays with the microsomal fraction after addition of NADPH, while the supernatant still had almost the same activity as the crude protein extract. Perhaps chlorophyll, predominantely found in the microsomes and higher concentrated compared to the crude protein extract, interfered. Western blot analysis of Pp_AaCPR also showed, that the protein was located in the membrane fraction. To exclude a soluble PpCPR, localization prediction was performed for the four potential genes (DeepLoc and PredictProtein Open) and there was a high probability for all of them being anchored to the ER. Perhaps the CPRs were denatured during the process of microsome extraction. It was not possible that the protein was cleaved off from the membrane anchor since AaCPR was still detected in the membrane fraction by a *C*-terminal tag. This could have only happend to the *Physcomitrella* CPRs. Since AaCPR could already be charaterized after expression in *E. coli*, this issue was not further dealt with.

3 Heterologous expression of CoA-ligases

Besides membrane-bound proteins, soluble enzymes all belonging to the family of CoA-ligases were investigated in this work. The predominant goal was to find a gene encoding a 4CL, the last enzyme in the core phenylpropanoid pathway. Four candidate sequences were selected: Aa19917, Aa20832, Aa35279 and AaAp6378. Although

different primer combinations were used, *Aa19917* was the only sequence, which could not be amplified using cDNA as template. On the other hand, amplification was no problem using gDNA, leading to fragments, persumably, all of them with one or more introns of approximately 500 bp (Fig. 70). Since the full-length sequence was unknown, amplification of the 5´- and 3´-ends would have been necessary. This was most effectively done in our lab with the SMARTer® RACE 5'/3' kit (Takara/Clontech) and cDNA as template. As this was supposably impossible, RAGE-PCR could have been used for amplification of the genomic ends with gDNA as a template (Cormack and Somssich 1997). After identification of the introns (NetPlantGene server), a fusion-PCR would have been the next step (Cha-Aim et al. 2012). But since all other potential 4CL candidate sequences were amplified with cDNA and *Aa19917* is most probably silenced, this gene was not further processed.

3.1 Aa4CL

AaAp6378 was the first of the three remaining candidate sequences that was expressed in *E. coli* SoluBL21. Since there were two possible start codons, both sequences were edited and ligated into pRSETc. The selection of the expression vector, the used *E. coli* strain and the purification procedure was based on the work of Weitzel (2009). Both heterologously expressed proteins (Aa4CL_1 and Aa4CL_2) demonstrated an activity towards 4-coumaric, caffeic, cinnamic, ferulic, isoferulic, 2-coumaric and 3-coumaric acid (Fig. 44 and 46). On the other hand, they lacked affinity for sinapic acid and all tested benzoic acid derivatives (Fig. 44 and 48). The missing acceptance for sinapic acid was not surprising, since the last two of the twelve amino acids encoding the 4CL specificity code, valine and leucine, were both present. If only one amino acid is deleted, the enzyme aquires the new function towards sinapic acid (Lindermayr et al. 2002; Schneider et al. 2003). From an alignment of different plant 4CLs, it was presumed, that the shorter Aa4CL_2 might be the more likely 4CL sequence (Fig. 42). This was also proven by the higher activity of Aa4CL_2 compared to Aa4CL_1. Aa4CL displayed the highest specific activity towards the substrates ferulic acid and 4-coumaric acid (Fig. 47). In 2002, Lindemayr et al. announced, that 4CLs differed substantially in substrate specificity and preference. For both, Aa4CL_1 and Aa4CL_2, cinnamic acid was the substrate with the lowest turnover. While the conversion of cinnamic acid depends

primarily on the hydrophobicity of the substrate binding pocket, the activation of ferulic acid and other substituted cinnamic acid derivatives is controlled by size exclusion (Schneider et al. 2003). For example, ferulic acid was not accepted, when the amino acids methionine^{343} and lysine^{370} were present (based on the protein sequence of Aa4CL_1). In both cases amino acids with shorter side chains appeared in Aa4CL with proline^{343} and methionine^{370}, thus, it was not surprising, that substituted cinnamic acid derivatives were activated. With a K_m-value of around 9.9 ± 1.8 µM for Aa4CL_1 and 7.5 ± 1.8 µM for Aa4CL_2 isoferulic acid displayed the highest affinity. For this substrate kinetic parameters for a comparison were only available from *Petroselinum crispum* (26 µM; Knobloch and Hahlbrock 1977) and *Glycine max* (100 µM for isoenzyme 1 and 150 µM for isoenzyme 2; Knobloch and Hahlbrock 1975). Aa4CL had almost the same affinity for 4-coumaric acid (12.8 ± 4.0 µM for Aa4CL_1 and 11.8 ± 4.4 µM for Aa4CL_2). This was in the range between 11 µM for *Pinus taeda* (Chen et al. 2014) and 16 µM for *Physcomitrella patens* 4CL2 (Silber et al. 2008). For caffeic acid the K_m-values of different 4CLs ranged from 1 µM (*Nicotiana tabacum*; Li and Nair 2015) to 725 µM (*Physcomitrella patens* 4CL3; Silber et al. 2008). With 15.5 ± 1.9 µM (Aa4CL_1) and 20.5 ± 4.6 µM (Aa4CL_2) the 4CL from *Anthoceros* was within this range. This was also the case for ferulic acid, ranging from 2.2 µM (*Oryza sativa* 4CL3; Gui et al. 2011) to 800 µM (*Physcomitrella patens* 4CL3; Silber et al. 2008). Cinnamic acid was the substrate with the lowest affinity with K_m-values ranging between 218.2 ± 74.7 µM for Aa4CL_1 and 216.7 ± 86.1 µM for Aa4CL_2. This was also in the range, 9.4 µM (*Oryza sativa* 4CL1; Gui et al. 2011) to 6.63 mM (*Arabidopsis thaliana*; Ehlting et al. 1999), of published K_m-values for this substrate. Only for cinnamic acid the low conversion rate corresponds to the comparably high K_m-value. The specific activity was only partially reflected in the substrate affinity, therefore catalytic efficiencies (K_{cat}/K_m) were calculated. These values reflected the order determined from the kinetic values with acceptance in the following order for Aa4CL_2: isoferulic > 4-coumaric > caffeic > ferulic > cinnamic acid and Aa4CL_1: 4-coumaric > isoferulic > caffeic > ferulic > cinnamic acid (Fig. 50; Table 11). With 227.1 ± 31.8 µM (Aa4CL_1) and 151.1 ± 25.3 (Aa4CL_2) the K_m-value for ATP was comparably high, but also in wild type and mutant enzymes of *A. thaliana* values ranged between 151 µM and 1.50 mM (Stuible et al. 2000). In *Forsythia suspensa* and *Glycine max* the K_m-value for CoA ranged between 3.2 µM and 7 µM (Gross and Zenk 1974;

Knobloch and Hahlbrock 1975). This value was higher for the analyzed enzyme, ranging between 10.2 ± 0.2 µM (Aa4CL_1) and 14.6 ± 3.2 µM (Aa4CL_2).

3.2 Aa4HBCL

Aa4HBCL was found as another CoA-ligase with activating properties towards cinnamic acid derivatives. In contrast to Aa4CL, Aa4HBCL preferably accepted benzoic acid derivatives (Fig. 58). Moreover, only the cinnamic acid derivatives isoferulic, 4-coumaric acid, cinnamic acid and caffeic acid (in order of preference) and additionally 2-coumaric and 3-coumaric acid were accepted (detected by LC-MS) (Fig. 54 and 57). The enzyme lacked affinity for sinapic acid and ferulic acid. This supports the hypothesis of Lindemayr et al. (2002) that 4CLs within a species differ in their substrate specificity and preference. On the other hand, both hypotheses regarding the activation of sinapic acid and substituted cinnamic acid derivatives were partially not valid for Aa4HBCL (Lindermayr et al. 2002; Schneider et al. 2003). If at least one of the last two amino acids of the 4CL substrate specificity code, valine and leucine, were deleted sinapic acid should be accepted. Although the valine was exchanged to an isoleucine and the leucine was changed to a threonine in Aa4HBCL, sinapic acid was not accepted by the enzyme. Furthermore, the acceptance of other substituted cinnamic acid derivatives was considered to be based on size exclusion. In the case of Schneider et al. (2003) ferulic acid was accepted, if the amino acids methionine and lysine (fourth and fifth amino acid of the 4CL substrate specificity code) were exchanged with amino acids with shorter side chains. Both, Aa4HBCL and Aa4CL shared the amino acids proline and methionine at these positions. While Aa4CL accepted both, ferulic acid and isoferulic acid, Aa4HBCL lacked affinity for ferulic acid but activated isoferulic acid. Thus, there might still be some unsolved questions concerning the 4CL substrate specificity code. Aa4HBCL preferably activated benzoic acid derivatives, primarily 4HBA and BA. Only salicylic acid, 3-aminosalicylic acid and vanillic acid were not accepted. Comparable results could also be observed for 3-hydroxybenzoate:CoA-ligase of *Centaurium erythraea* (Barillas and Beerhues 1997). However, this enzyme only accepted 3HBA, BA and 4HBA. BCLs from *Arabidopsis thaliana* and *Pyrus pyrifolia* preferably accepted benzoic acid, but also activated other monohydroxylated and monoaminated benzoic acid derivatives (Kliebenstein et al. 2007; Saini et al. 2020). Another BCL from *Clarkia breweri* was only

tested with benzoic acid (Beuerle and Pichersky 2002). Only one other plant CoA-ligase is currently published from *Petunia hybrida*, activating both cinnamic acid and benzoic acid derivatives, but with a much higher affinity towards cinnamic acids (Klempien et al. 2012).

The determination of activated benzoic acids turned out to be difficult. Only the formation of 4-hydroxybenzoyl-CoA could directly be measured photometrically at 300 nm (Biegert et al. 1993). At first, the others were measured by an indirect method, based upon the reduction of ATP to AMP (Fig. 55). This method was not sensitive enough for most of the tested substrates. Reproducing the kinetic parameters for 4HBA with this method was also not possible, since a difference in ATP concentration was not measurable in an assay saturated with ATP after a short reaction time to ensure initial reaction velocities. Another method tested with 4HBA and BA was an indirect assay based on the consumption of CoA (data not shown). By addition of Ellman's reagent (5,5'-dithio-bis-[2-nitrobenzoic acid]) reduced cysteines and other free sulfhydryls in solutions can be measured, in this case the unbound CoA (Shepherd and Garland, 1969). This, however, did not lead to any decrease of absorption after incubation of 4HBA or BA with Aa4HBCL for 0, 60 or 120 min. The best option would probably have been to transfer the LC-MS runs to a normal HPLC together with detection at the optimal wavelengths (as recorded by LC-MS). But since no standard solution was available for all products, only K_m could have been determined.

Thus, kinetic parameters were only determined for the main substrate 4HBA as well as the cosubstrates ATP and CoA (Fig. 60; Table 14). The K_m for 4HBA was 664.2 ± 1.5 µM this could only be compared to badA from *Rhodopseudomonas palustris* (158 µM) (Thornburg et al. 2015), since, at present, there were no available kinetic parameters for plant BCLs. The benzoate activating enzyme from *C. breweri* (Beuerle and Pichersky 2002) displayed a K_m-value of 95 µM for ATP. With 1.2 ± 0.1 mM Aa4HBCL had a much lower affinity. Also the K_m-value of 13 µM from *C. breweri* for CoA is much lower than the K_m-value of Aa4HBCL of 247.8 ± 19.9 µM. Kinetic values for isoferulic acid, 4-coumaric acid, cinnamic acid, caffeic acid as well as all accepted benzoic acid derivatives still need to be determined. Although the specific acitivties of both enzymes, Aa4CL and

Aa4HBCL, were in the same range, around 30 mkat/kg, for the two cosubtrates ATP and CoA, they were used with a much higher affinity by Aa4CL.

To date, the role of Aa4HBCL in the hornwort is quite unclear. A possible hint was found in the dissertation of Trennheuser (1992), where he elucidated the structure of many secondary metabolites found in *Anthoceros agrestis* suspension cultures. These included amides of glutamic acid with 4-hydroxybenzoic acid, protocatechuic acid, 4-coumaric acid, isoferulic acid and vanillic acid. These substances could have probably been produced by an HCT using glutamic acid and the appropriate CoA-ester. Aa4HBCL was able to activate all these substrates except vanillic acid. Methylation of protocatechuoylglutamate might be the reaction establishing the vanillic acid moiety observed in *Anthoceros*. Nevertheless, the participating HCT still needs to be found. Only in *Theobroma cacao*, *p*-coumaroylglutamate and caffeoylglutamate were identified amongst other amino acid conjugates (Stark and Hofmann 2005). In other plants, benzoyl-CoA producing CoA-ligases are involved in glucosinolate biosynthesis, xanthone biosynthesis or the biosynthesis of taxol and biphenyl phytoalexins (Barillas and Beerhues 1997; Walker and Croteau 2000; Kliebenstein et al. 2007; Saini et al. 2020).

3.3 Aa20832

The last potential 4CL candidate investigated in this work was Aa20832. Assays with crude protein extract and purified protein of the heterologously expressed Aa20832 did not show any product formation with either cinnamic or benzoic acid derivatives (Fig. 64 and 65). Only the formation of AMP could be determined, no matter which substrate was used (Fig. 68). This function towards ATP was also observed with Aa4HBCL and Aa4CL in a smaller scale. Thus it was assumed, that the purified enzyme was still active, although a potential transmembrane helix was identified in a secondary structure prediction (Phyre^2 and predictprotein.org). This could have probably been clarified by Western blot analysis of the membrane fraction. Since *E. coli* does not have peroxisomes, perhaps the heterologous enzyme might still be soluble. Specific antibodies against Aa20832 would therefore be necessary to detect the localisation of the protein in *Anthoceros*.

Because of the peroxisomal signal sequence and the protein BLAST result, fatty acids and OPC derivatives were considered as potential substrates for Aa20832 (Fulda et al. 2002; Schneider et al. 2005; Koo et al. 2006; Kienow et al. 2008; Reumann et al. 2016). Thus, Aa20832 was incubated with palmitic acid and stearic acid. These assays were analyzed by the indirect method and - as already mentioned above - Aa20832 displayed an ATPase activity even without a substrate (Fig. 66). The two fatty acids displayed a stronger decrease of ATP concentration over time than methanol, sinapic acid or salicylic acid. Consequently a different method needs to be considered. Activity of enzymes from *Arabidopsis thaliana* were either identified in a way similar to the indirect method used in this work, by adding phosphoenolpyruvate, myokinase, pyruvate kinase and lactate dehydrogenase (Koo et al. 2006; Kienow et al. 2008) or by mass spectrometric analysis (Schneider et al. 2005). The second should be considered for Aa20832. Putative product peaks could then be further characterized by MS/MS analysis. Since substrates of Aa20832 are still unknown, nothing can be predicted about the function of this enzyme.

4 Summary and outlook

The phenylpropanoid pathway and derived metabolic pathways lead to a multitude of different compounds, e.g. coumarins, flavonoids, isoflavonoids, lignans and lignin, cutin and suberin as well as hydroxycinnamic acid esters and amides (e.g. chlorogenic acid and rosmarinic acid) (Vogt 2010). Genes encoding C4H, 4CL, CYP98, 4HBCL and CPR were identified, heterologously expressed and biochemically characterized in this thesis. Moreover, two proteins with yet unknown function, a P450 and another CoA-ligase, were expressed in *P. patens* and *E. coli*.

Since the heterologous expression of the native P450 enzymes and CPR failed in yeast, various strategies have been pursued to address this issue. While some of them still failed, for example a chimeric CPR from *Coleus* and *Anthoceros*, most of them led to a detectable protein formation. Probably the fastest and simplest option was the codon optimization of *AaCYP98*. AaC4H and AaCPR were fused and expressed in *E. coli*. *Physcomitrella patens* was introduced as a suitable expression host for P450s. Overall, the heterologous expression in the moss was a good alternative compared to yeast.

However, the big disadvantage was that PpWT always had to be carried along because *Physcomitrella patens* has own copies of most of the genes under investigation. To overcome this issue, targeted knockouts would be necessary (Cove 1992; Cove and Knight 1993; Schaefer and Zrÿd 1997). For *AaCYP98* this would not be too difficult, since only one homolog is present in *Physcomitrella* and ideally *AaCYP98* could just be exchanged with *PpCYP98*, but for all other genes investigated in this survey, often more than one homolog was present. All of them could be deleted in one step by the CRISPR system, but this is not yet established in our lab. Nevertheless, this has already been performed by other researchers in *Physcomitrella* and other mosses (Lopez-Obando et al. 2016; Nomura et al. 2016). Thus, the moss will move into even greater consideration as an expression host compared to yeast and other diploid plants in the future.

Targeted knockout directly in *Anthoceros agrestis* would also be an interesting alternative. Protoplastation as the first step of transformation of the hornwort was also tried in this work (data not shown). While the protoplastation of *Anthoceros* suspension cells was successful, the cells were not able to reproduce on solid CBM medium. Compared to *Physcomitrella*, protoplasts were smaller and mostly colourless, as they had barely any chloroplasts. Thus it would be neccessary to use a carbohydrate source in the regeneration medium (perhaps even at a higher concentration than in CBM medium) to support growth. Clearly, many other factors would need to be adjusted, for example viscosity of the medium, ventilation and light conditions. If the cells could be regenerated, it would be interesting to see, whether the cells would regenerate as callus cells or a gametophyte would grow again. Up to date, no transformation protocol is available for the hornwort and it is unknown, whether *Anthoceros* will display the same rates of homologous recombination as *Physcomitrella* to obtain stable transformants. However, it will only be a matter of time before *Anthoceros* will be fully recognized as a model plant.

Two CoA-ligases, activating different cinnamic acid derivatives, were found with Aa4CL and Aa4HBCL. Both differed substantially in their substrate preference: Aa4CL only accepted hydroxycinnamic acids but no benzoic acid derivatives. The other CoA-ligase preferably activated benzoic acid or monohydroxylated benzoic acids (except for salicylic acid) and many other benzoic acid derivatives.

Nevertheless, many questions remain unanswered. Initially, this work was planned to further clarify rosmarinic acid biosynthesis in *Anthoceros agrestis*. Rosmarinic acid (RA), an ester of caffeic acid and 3,4-dihydroxyphenyllactic acid, was found in more than 40 plant families, including hornworts, ferns, monocots and dicots (Harborne 1966; Takeda et al. 1990; Petersen and Simmonds 2003; Petersen 2013). *Anthoceros agrestis* can accumulate up to 5 % of the dry weight as rosmarinic acid (Vogelsang et al. 2006). Yet, the biosynthetic pathway for RA in the hornwort is unclear. It is uncertain whether the ability to synthesize rosmarinic acid was already present when tracheophytes and anthocerotophytes were separated and was then lost in many taxa or whether the ability developed independently in the individual taxa. The latter was further supported by this work. Although all involved enzymes of the phenylpropanoid metabolism, as well as the two enzymes tyrosine aminotransferase (TAT) and hydroxyphenylpyruvate reductase (HPPR) for the synthesis of 4-hydroxyphenyllactic acid were found in the hornwort (Elke Bauerbach, Tobias Busch, Maike Petersen, Soheil Pezsehki, Ina Warmbier, Julia Wohl), the key enzyme rosmarinic acid synthase (RAS) has not been found yet. Enzyme assays with the crude protein extract of *Anthoceros agrestis* incubated with Caf-CoA and PHPL did not lead to the expected product Caf-PHPL (Pezeshki 2016). In Lamiaceae, the next step of rosmarinic acid biosynthesis is the hydroxylation of pC-PHPL at position 3 and 3´ by a CYP98. Although AaCYP98 was able to hydroxylate pC-PHPL and probably also pC-DHPL, the conversion was neglectable in comparison to the large amounts of RA accumulated in *Anthoceros*. AaCYP98 was not able to hydroxylate Caf-PHPL to rosmarinic acid, this supports the hypothesis of Petersen (1997) and Eberle et al. (2009) that actually two P450s might be involved in hydroxylation. Since there is probably only one homolog of CYP98 present in the bryophyte (Alber et al. 2019), at least this step will probably differ to the rosmarinic biosynthesis found in Lamiaceae (Petersen et al. 1993; Eberle et al. 2009). Catalyzing the hydroxylation of pC-3OH-An and pC-An, AaCYP98 is probably involved in the biosynthesis of cutin and suberin, like the CYP98 found in *Physcomitrella patens* (Renault et al. 2017a). Thus it remains unknown, how the hornwort is able to produce RA. Alternatively, Caf-3OH-An and/or Caf-An might function as donor for caffeoyl moieties to transfer them to pHPL.

Besides RA, many other metabolites (some presumably derived from RA) were found in the hornwort (Trennhäuser 1992), including rosmarinic acid methylester, megacerotonic acid, 3-hydroxymegacerotonic acid, anthocerodiazonin and anthocerotonic acid. Until now, the biosynthesis of these, partly very complex, structures has not yet been investigated. Besides various esters and amides of 4-coumaric and caffeic acid he also identified derivatives of isoferulic acid while similar adducts with ferulic acid were missing. This might be a specific feature of hornworts or lower plants since isoferulic acid is no predominant compound in gymnosperms and angiosperms whereas ferulic acid and its derivatives are more important, e.g. as coniferyl alcohol unit (G-unit) in lignin formation.

Unexpectively, the biosynthesis of various amides of glutamic acid (namely amides with 4-hydroxybenzoic acid, protocatechuic acid, vanillic acid, isoferulic acid and 4-coumaric acid) found in *Anthoceros agrestis* might have been discovered. In this case Aa4HBCL might play a role in the activation of the benzoic acid or cinnamic acid derivatives. Nevertheless, the HCT, forming the amide, still needs to be identified. Moreover it is unknown, whether hydroxylation events occur before or after the coupling. Since vanillic acid was not activated by Aa4HBCL, methylation might occur in later steps (Fig. 91).

Until now, hornworts were the only gap within the bryophytes in understanding the phenylpropanoid metabolism on a genomic level (de Vries et al. 2017; Renault et al. 2019). Based on this work and the work of many other researchers in this working group (Elke Bauerbach, Lucien Ernst, Maike Petersen, Soheil Pezsehki and Ina Warmbier; many unpublished results) the core phenylpropanoid pathway and enzymes of subsequent biosynthetic pathways of *Anthoceros agrestis* could be identified and characterized. These results provide a good basis to gain more insight into the function and evolution of the plant phenylpropanoid biosynthetic pathway.

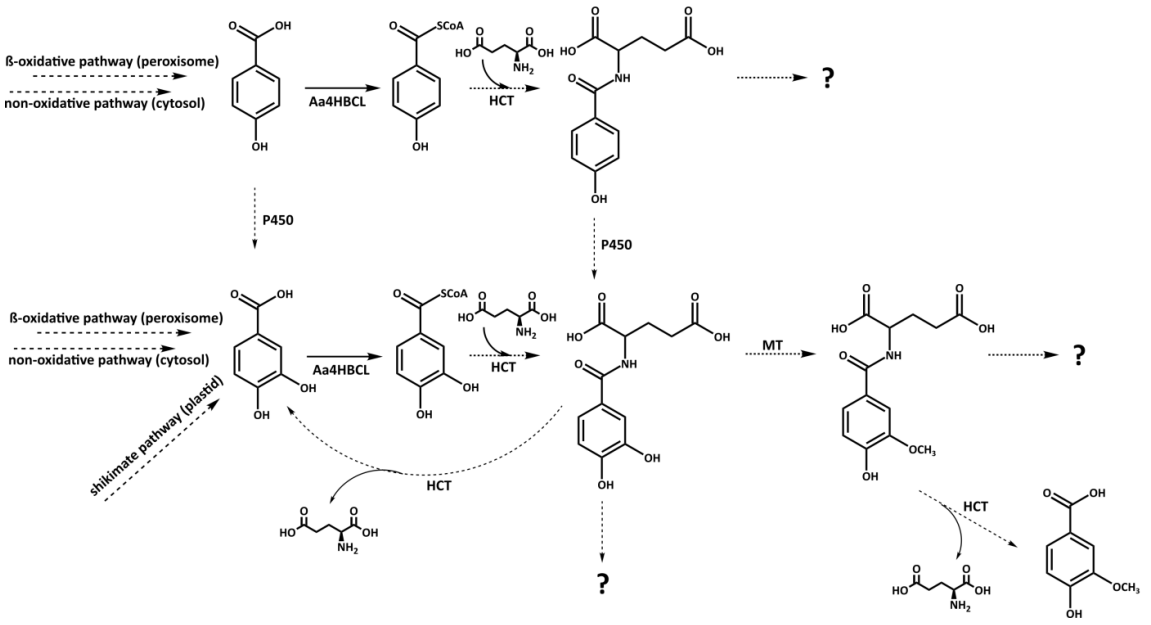


Figure 91 Assumed biosynthetic pathway of amides of glutamic acid found in *Anthoceros agrestis*. The hypothetical biosynthetic pathway is based on the results of this work and the publications of Trennheuser (1992) and Widhalm and Dudareva (2015). The abbreviations are HCT – hydroxycinnamoyltransferase, P450 – cytochrome P450 monooxygenase, MT - methyltransferase

V Summary

An important group of plant compounds are phenolics, leading to many different secondary metabolites, e.g. flavonoids, lignans and lignin, cutin and suberin as well as hydroxycinnamic acid esters and amides (e.g. chlorogenic acid and rosmarinic acid). They can act for example as natural sun screens or cell wall reinforcements, essential for the conquering of dry habitats [1]. Until now, hornworts were the only gap amongst the bryophytes in understanding the phenylpropanoid metabolism on a genomic level [2, 3].

Cinnamic acid 4-hydroxylase (C4H) from *Anthoceros agrestis* was heterologously expressed in *Escherichia coli* and *Physcomitrella patens*. For expression in *E. coli AaC4H* was fused to *AaCPR*, a NADPH:cytochrome P450 reductase. The resulting fusion protein produced only a small amount of 4-coumaric acid. Additionally, AaCPR was present as a soluble protein and was therefore characterized after purification by metal chelate chromatography. *P. patens* was introduced as a suitable expression host for plant P450s. In real time PCR experiments it was shown, that expression of *AaC4H* was between 270 to 3700-fold compared to two potential *C4H*s from *P. patens*. Protein extracts from transformed cultures revealed the formation of double to triple the amount of 4-coumaric acid and increased the affinity for cinnamic acid compared to the wild-type control.

Two CoA-ligases, activating different (hydroxy)cinnamic and (hydroxy)benzoic acid derivatives, were found with 4-coumarate CoA-ligase (Aa4CL) and 4-hydroxybenzoate CoA-ligase (Aa4HBCL). Both differed substantially in their substrate preference. Aa4CL accepted 4-coumaric, caffeic, cinnamic, ferulic, isoferulic, 2-coumaric and 3-coumaric acid, but lacked affinity for sinapic acid and benzoic acid derivatives. The other CoA-ligase preferably activated benzoic acid or monohydroxylated benzoic acids and several other benzoic acid derivatives (except for salicylic acid, 3-aminosalicylic acid and vanillic acid) but also demonstrated activity towards cinnamic, 2-coumaric and 3-coumaric, 4-coumaric, caffeic and isoferulic acid.

Besides Aa4CL and Aa4HBCL a third potential CoA-ligase was expressed in *E. coli*. The amino acid sequence of Aa20832 revealed a peroxisomal signal sequence (PTS1) and two potential transmembrane helices were identified by secondary structure prediction.

The protein did not activate any tested (hydroxy)cinnamic or (hydroxy)benzoic acids but demonstrated a high ATPase activity. Aa20832 might have an activity towards fatty acids, but this was not clearly proven.

Another cytochrome P450 characterized in this work was AaCYP98, a hydroxycinnamoyl ester/amide 3-hydroxylase. The native AaCYP98 was expressed in *P. patens* and a codon optimized sequence was expressed in *S. cerevisiae*. AaCYP98 was able to hydroxylate 4-coumaroyl-3´-hydroxyanthranilic acid, 4-coumaroylanthranilic acid, 4-coumaroyltyramine, 4-coumaroylshikimic acid and 4-coumaroyl-4´-hydroxyphenyllactic acid but did not accept 4-coumaroylquinic acid, 4-coumaroyl-2´-threonic acid and caffeoyl-4´-hydroxyphenyllactic acid. Since these substrates were not commercially available most of them were either isolated from plant cell cultures or synthesized enzymatically or chemically. Activity of AaCYP98 was increased 13-fold after coexpression with a CPR from *Coleus blumei*.

At last, the cytochrome P450 AaAp626 was expressed in *P. patens*. In a BLASTp search, the sequence demonstrated the highest identities towards putative flavonoid 3´-hydroxylases and CYP71A1. No activity was observed in enzyme assays with 4-coumaroyl-4´-hydroxyphenyllactic acid and the three flavonoids galangin, kaempferol and naringenin. Thus, the function of this protein remains unclear.

In summary, based on this work the last two enzymes of the core phenylpropanoid pathway (AaC4H and Aa4CL) and enzymes of subsequent biosynthetic pathways of the hornwort *Anthoceros agrestis* (AaCYP98 and Aa4HBCL), as well as a NADPH-dependent CPR (AaCPR) were identified and characterized. Moreover, two proteins with yet unknown function, a P450 and another CoA-ligase, were heterologously expressed in *P. patens* and *E. coli*. These results provide a good foundation to gain more insight into the function and evolution of the plant phenylpropanoid biosynthetic pathway.

[1] Rensing (2018) Current opinion in plant biology 42: 49-54
[2] de Vries et al. (2017) Plant and Cell Physiology 58: 934-945
[3] Renault et al. (2019) Current opinion in biotechnology 56: 105-111

VI Zusammenfassung

Eine wichtige Gruppe pflanzlicher Inhaltstoffe sind phenolische Verbindungen, welche zu vielen unterschiedlichen Sekundärmetaboliten umgewandelt werden können, z.B. Flavonoiden, Lignanen und Lignin, Cutin und Suberin sowie Hydroxyzimtsäureestern und -amiden (z.B. Chlorogensäure und Rosmarinsäure). Diese Substanzen können zum Beispiel als natürlicher Sonnenschutz oder als Zellwandverstärkung dienen, welche für das Leben oberhalb der Wasseroberfläche unerlässlich sind [1]. Unter den Bryophyten waren die Hornmoose bisher die einzige Abteilung, bei der es keine Informationen über den Phenylpropanoid-Stoffwechsels auf genomischer Ebene gab [2, 3].

Die Zimtsäure 4-Hydroxylase aus *Anthoceros agrestis* wurde heterolog in *Escherichia coli* und *Physcomitrella patens* exprimiert. Zur Expression in *E. coli* wurde *AaC4H* mit *AaCPR*, einer NADPH:Cytochrom-P450-Reduktase, fusioniert. Das resultierende Fusionsprotein produzierte nur eine geringe Menge an 4-Cumarsäure. Zusätzlich lag AaCPR als lösliches Protein vor und konnte daher nach Aufreinigung über Affinitätschromatographie charakterisiert werden. *P. patens* wurde als geeigneter Expressionswirt pflanzlicher CYPs vorgestellt. In RT-qPCR Experimenten konnte gezeigt werden, dass die Expression von *AaC4H* zwischen 270 bis 3700-fach im Vergleich zu zwei potentiellen *C4H*s aus *P. patens* war. Extrakte aus transformierten Kulturen zeigten die Bildung der doppelten bis dreifachen Menge an 4-Cumarsäure und erhöhten die Affinität für Zimtsäure im Vergleich zur Wildtyp-Kontrolle.

Mit einer 4-Cumarat CoA-Ligase (Aa4CL) und einer 4-Hydroxybenzoat CoA-Ligase (Aa4HBCL) wurden zwei Enzyme entdeckt, welche in der Lage waren, verschiedene Zimtsäure- und Benzoesäurederivate zu aktivieren. Dabei unterschieden sich beide erheblich in ihrer Substratpräferenz. Aa4CL aktivierte ausschließlich (Hydroxy-) Zimtsäuren, akzeptierte aber keine Sinapinsäure. Aa4HBCL aktivierte vorzugsweise Benzoesäure, monohydroxylierte Benzoesäuren und viele andere Benzoesäurederivate (außer Salicylsäure, 3-Aminosalicylsäure und Vanillinsäure), zeigte aber auch Aktivität mit (Hydroxy-)Zimtsäuren (außer Ferulasäure und Sinapinsäure).

Neben Aa4CL und Aa4HBCL wurde eine dritte potenzielle CoA-Ligase in *E. coli* exprimiert. Die Aminosäuresequenz von Aa20832 zeigte eine peroxisomale

Signalsequenz (PTS1) und zwei potenzielle Transmembran-Helices. Das Protein aktivierte weder (Hydroxy-)Zimtsäuren noch (Hydroxy-)Benzoesäuren, zeigte aber eine hohe ATPase-Aktivität. Es konnte nicht eindeutig nachgewiesen werde, ob Aa20832 eine Aktivität gegenüber Fettsäuren aufweist.

Ein weiteres in dieser Arbeit charakterisiertes Enzym war AaCYP98, eine Hydroxycinnamoylester/amid 3-Hydroxylase. Eine native AaCYP98 wurde in *P. patens* exprimiert und eine Codon-optimierte Sequenz der AaCYP98 in *S. cerevisiae*. AaCYP98 war in der Lage, 4-Cumaroyl-3´-Hydroxyanthranilat, 4-Cumaroylanthranilat, 4-Cumaroyltyramin, 4-Cumaroylshikimat und 4-Cumaroyl-4´-Hydroxyphenyllactat zu hydroxylieren, nicht jedoch 4-Cumaroylchinat, 4-Cumaroyl-2´-Threonat und Caffeoyl-4´-Hydroxyphenyllactat. Da diese Substrate nicht kommerziell erhältlich waren, wurden die meisten von ihnen entweder aus pflanzlichen Zellkulturen isoliert oder enzymatisch sowie chemisch synthetisiert. Durch Koexpression einer CPR aus *Coleus blumei* konnte die Aktivität der AaCYP98 13-fach gesteigert werden.

Das Cytochrom P450 Enzym AaAp626 wurde in *P. patens* exprimiert. In einer BLASTp-Suche zeigte die Sequenz die höchste Identität gegenüber putativen Flavonoid 3'-Hydroxylasen und CYP71A1. In Enzymtests mit verschiedenen Substraten konnte aber keine Aktivität beobachtet werden. Daher bleibt die Funktion dieses Proteins ungewiss.

Zusammenfassend wurden in dieser Arbeit die beiden letzten Enzyme des zentralen Phenylpropanoid-Stoffwechsels (AaC4H und Aa4CL), Enzyme nachfolgender Biosynthesewege (AaCYP98 und Aa4HBCL) sowie eine NADPH-abhängige CPR (AaCPR) aus dem Hornmoos *Anthoceros agrestis* identifiziert und charakterisiert. Darüber hinaus wurden zwei Proteine mit noch unbekannter Funktion, ein CYP und eine weitere CoA-Ligase, heterolog in *P. patens* und *E. coli* exprimiert. Diese Ergebnisse bieten eine gute Grundlage, um einen tieferen Einblick in die Funktion und Evolution des pflanzlichen Phenylpropanoid-Biosynthesewegs zu gewinnen.

[1] Rensing (2018) Current opinion in plant biology 42: 49-54
[2] de Vries et al. (2017) Plant and Cell Physiology 58: 934-945
[3] Renault et al. (2019) Current opinion in biotechnology 56: 105-111

VII References

Abas L, Luschnig C (2010) Maximum yields of microsomal-type membranes from small amounts of plant material without requiring ultracentrifugation. Anal Biochem 401:217-227. doi.org/10.1016/j.ab.2010.02.030

Adams ZP, Ehlting J, Edwards R (2019) The regulatory role of shikimate in plant phenylalanine metabolism. J Theor Biol 462:158-170. doi.org/10.1016/j.jtbi.2018.11.005

Ahuja I, Kissen R, Bones AM (2012) Phytoalexins in defense against pathogens. Trends Plant Sci 17:73-90. doi.org/10.1016/j.tplants.2011.11.002

Aigner S, Remias D, Karsten U, Holzinger A (2013) Unusual phenolic compounds contribute to ecophysiological performance in the purple-colored green alga *Zygogonium ericetorum* (Zygnematophyceae, Streptophyta) from a high-alpine habitat. J Phycol 49:648-660. doi.org/10.1111/jpy.12075

Alber AV, Renault H, Basilio-Lopes A, Bassard JE, Liu Z, Ullmann P et al. (2019) Evolution of coumaroyl conjugate 3-hydroxylases in land plants: lignin biosynthesis and defense. Plant J 99:924-936. doi.org/10.1111/tpj.14373

Altman FP (1976) Tetrazolium salts and formazans. Prog Histochem Cyto 9:1-51. doi.org/10.1016/S0079-6336(76)80015-0

An JH, Lee GY, Jin-Won JUNG, Weontae LEE, Kim YS (1999) Identification of residues essential for a two-step reaction by malonyl-CoA synthetase from *Rhizobium trifolii*. Biochem J 344:159-166. doi.org/10.1042/bj3440159

Anterola AM, Jeon JH, Davin LB, Lewis NG (2002) Transcriptional Control of Monolignol Biosynthesis in *Pinus taeda* factors affecting monolignol ratios and carbon allocation in phenylpropanoid metabolism. J Biol Chem 277:18272-18280. doi.org/ 10.1074/jbc.M112051200

Anterola A, Shanle E, Perroud PF, Quatrano R (2009) Production of taxa-4(5),11(12)-diene by transgenic *Physcomitrella patens*. Transgenic Res 18:655. doi.org/10.1007/s11248-009-9252-5

Asakawa Y, Toyota M, Tori M, Hashimoto T (2000) Chemical structures of macrocyclic bis (bibenzyls) isolated from liverworts (Hepaticae). J Spectrosc 14:149-175. doi.org/10.1155/2000/570265

Asakawa Y, Ludwiczuk A, Nagashima F (2013) Chemical constituents of bryophytes: bio- and chemical diversity, biological activity, and chemosystematics (Vol. 95). Springer-Verlag, Wien. doi.org/10.1007/978-3-7091-1084-3

Bach SS, King BC, Zhan X, Simonsen HT, Hamberger B (2014) Heterologous stable expression of terpenoid biosynthetic genes using the moss *Physcomitrella patens*. In Rodríguez-Concepción M (ed) Plant isoprenoids. Methods in molecular biology (Methods and protocols, Vol. 1153), pp 257-271. Humana Press, New York, NY. doi.org/10.1007/978-1-4939-0606-2_19

Barillas W, Beerhues L (1997) 3-Hydroxybenzoate: coenzyme A ligase and 4-coumarate: coenzyme A ligase from cultured cells of *Centaurium erythraea*. Planta 202:112-116. doi.org/10.1007/s004250050109

Barros J, Serrani-Yarce JC, Chen F, Baxter D, Venables BJ, Dixon RA (2016) Role of bifunctional ammonia-lyase in grass cell wall biosynthesis. Nat Plants 2:1-9. doi.org/10.1038/nplants.2016.50

Batard Y, Hehn A, Nedelkina S, Schalk M, Pallett K, Schaller H, Werck-Reichhart D (2000) Increasing expression of P450 and P450-reductase proteins from monocots in heterologous systems. Arch Biochem Biophys 379:161-169. doi.org/10.1006/abbi.2000.1867

Becker B (2013) Snow ball earth and the split of Streptophyta and Chlorophyta. Trends Plant Sci 18:180-183. doi.org/10.1016/j.tplants.2012.09.010

Benveniste I, Gabriac B, Durst F (1986) Purification and characterization of the NADPH-cytochrome P-450 (cytochrome c) reductase from higher-plant microsomal fraction. Biochem J 235:365-373. doi.org/10.1042/bj2350365

Benveniste I, Lesot A, Hasenfratz MP, Kochs G, Durst F (1991) Multiple forms of NADPH-cytochrome P450 reductase in higher plants. Biochem Biophys Res Co 177:105-112. doi.org/10.1016/0006-291X(91)91954-B

Bernhardt R (2006) Cytochromes P450 as versatile biocatalysts. J Biotechnol 124:128-145. doi.org/10.1016/j.jbiotec.2006.01.026

Bertani G (1951) Studies on lysogenesis I.: the mode of phage liberation by lysogenic *Escherichia coli*. J Bacteriol 62:293.

Beuerle T, Pichersky E (2002) Purification and characterization of benzoate: coenzyme A ligase from *Clarkia breweri*. Arch Biochem Biophys 400:258-264. doi.org/10.1016/S0003-9861(02)00026-7

Biegert T, Altenschmidt U, Eckerskorn C, Fuchs G (1993) Enzymes of anaerobic metabolism of phenolic compounds: 4-Hydroxybenzoate-CoA ligase from a denitrifying *Pseudomonas species.* Eur J Biochem 213:555-561. doi.org/10.1111/j.1432-1033.1993.tb17794.x

Blount JW, Korth KL, Masoud SA, Rasmussen S, Lamb C, Dixon RA (2000) Altering expression of cinnamic acid 4-hydroxylase in transgenic plants provides evidence

for a feedback loop at the entry point into the phenylpropanoid pathway. Plant Physiol 122:107-116. doi.org/10.1104/pp.122.1.107

Bowman JL, Kohchi T, Yamato KT, Jenkins J, Shu S, Ishizaki K et al. (2017) Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. Cell 171:287-304. doi.org/10.1016/j.cell.2017.09.030

Bozak KR, O'Keefe DP, Christoffersen RE (1992) Expression of a ripening-related avocado (*Persea americana*) cytochrome P450 in yeast. Plant Physiol 100:1976-1981. doi.org/10.1104/pp.100.4.1976

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248-254. doi.org/10.1016/0003-2697(76)90527-3

Castellarin SD, Di Gaspero G, Marconi R, Nonis A, Peterlunger E, Paillard S et al. (2006) Colour variation in red grapevines (*Vitis vinifera* L.): genomic organisation, expression of flavonoid 3'-hydroxylase, flavonoid 3', 5'-hydroxylase genes and related metabolite profiling of red cyanidin-/blue delphinidin-based anthocyanins in berry skin. Bmc Genomics 7:12. doi.org/10.1186/1471-2164-7-12

Cha-aim K, Hoshida H, Fukunaga T, Akada R (2012) Fusion PCR via novel overlap sequences. In: Peccoud J (ed) Gene Synthesis. Methods in Molecular Biology (Methods and Protocols, Vol. 852), pp 97-110. Humana Press. doi.org/10.1007/978-1-61779-564-0_8

Chang KH, Xiang H, Dunaway-Mariano D (1997) Acyl-adenylate motif of the acyl-adenylate/thioester-forming enzyme superfamily: a site-directed mutagenesis study with the *Pseudomonas* sp. strain CBS3 4-chlorobenzoate: coenzyme A ligase. Biochemistry 36:15650-15659. doi.org/10.1021/bi971262p

Chang Y, Graham SW (2011) Inferring the higher-order phylogeny of mosses (Bryophyta) and relatives using a large, multigene plastid data set. Am J Bot 98:839-849. doi.org/10.3732/ajb.0900384

Chapple C (1998) Molecular-genetic analysis of plant cytochrome P450-dependent monooxygenases. Annu Rev Plant Biol 49:311-343. doi.org/10.1146/annurev.arplant.49.1.311

Chen AH, Chai YR, Li JN, Chen L (2007) Molecular cloning of two genes encoding cinnamate 4-hydroxylase (C4H) from oilseed rape (*Brassica napus*). BMB Rep 40:247-260. doi.org/10.5483/BMBRep.2007.40.2.247

Chen HY, Babst BA, Nyamdari B, Hu H, Sykes R, Davis MF et al. (2014) Ectopic expression of a loblolly pine class II 4-coumarate: CoA ligase alters soluble phenylpropanoid metabolism but not lignin biosynthesis in *Populus*. Plant Cell Physiol 55:1669-1678. doi.org/10.1093/pcp/pcu098

Cheynier V, Comte G, Davies KM, Lattanzio V, Martens S (2013) Plant phenolics: recent advances on their biosynthesis, genetics, and ecophysiology. Plant Physiol Biochem 72:1-20. doi.org/10.1016/j.plaphy.2013.05.009

Chomczynski P, Sacchi N (1987) The single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. Anal Biochem 162:156-159. doi.org/10.1038/nprot.2006.83

Clé C, Hill LM, Niggeweg R, Martin CR, Guisez Y, Prinsen E, Jansen MA (2008) Modulation of chlorogenic acid biosynthesis in *Solanum lycopersicum*; consequences for phenolic accumulation and UV-tolerance. Phytochemistry 69:2149-2156. doi.org/10.1016/j.phytochem.2008.04.024

Comino C, Hehn A, Moglia A, Menin B, Bourgaud F, Lanteri S, Portis E (2009) The isolation and mapping of a novel hydroxycinnamoyltransferase in the globe artichoke chlorogenic acid pathway. BMC Plant Biol 9:30. doi.org/10.1186/1471-2229-9-30

Conti E, Franks NP, Brick P (1996) Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes. Structure 4:287-298. doi.org/10.1016/S0969-2126(96)00033-0

Conti E, Stachelhaus T, Marahiel MA, Brick P (1997) Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. EMBO J 16:4174-4183. doi.org/10.1093/emboj/16.14.4174

Cormack RS, Somssich IE (1997) Rapid amplification of genomic ends (RAGE) as a simple method to clone flanking genomic DNA. Gene 194:273-276. doi.org/10.1016/S0378-1119(97)00205-9

Cove DJ (1992) Regulation of development in the moss *Physcomitrella patens*. In: Russo VEA, Brody S, Cove D, Ottolenghi S (eds) Development: The molecular genetic approach. Springer, Berlin, Heidelberg, pp 179-193. doi.org/10.1007/978-3-642-77043-2_13

Cove DJ, Knight CD (1993) The moss *Physcomitrella patens*, a model system with potential for the study of plant reproduction. Plant Cell 5:1483-1488. doi.org/10.1105/tpc.5.10.1483

Cove D (2005) The moss *Physcomitrella patens*. Annu Rev Genet 39:339-358. doi.org/10.1146/annurev.genet.39.073003.110214

Cove DJ, Perroud PF, Charron AJ, McDaniel SF, Khandelwal A, Quatrano RS (2009a) Culturing the moss *Physcomitrella patens*. Cold Spring Harbor Protocols, pdb-prot5136. doi.org/10.1101/pdb.prot5136

Cove DJ, Perroud PF, Charron AJ, McDaniel SF, Khandelwal A, Quatrano RS (2009b) Isolation and regeneration of protoplasts of the moss *Physcomitrella patens*. Cold Spring Harbor Protocols, pdb-prot5140. doi.org/10.1101/pdb.prot5140

Cove DJ, Perroud PF, Charron AJ, McDaniel SF, Khandelwal A, Quatrano RS (2009c) Transformation of the moss *Physcomitrella patens* using direct DNA uptake by protoplasts. Cold Spring Harbor Protocols, pdb-prot5143. doi.org/10.1101/pdb.prot5143

Cove DJ, Perroud PF, Charron AJ, McDaniel SF, Khandelwal A, Quatrano RS (2009d) The moss *Physcomitrella patens*: a novel model system for plant development and genomic studies. Cold Spring Harbor Protocols, pdb-emo115. doi.org/10.1101/pdb.emo115

Crouse J, Amorese D (1987) Ethanol precipitation: ammonium acetate as an alternative to sodium acetate. Focus 9:3-5.

Dai Z, Liu Y, Sun Z, Wang D, Qu G, Ma X et al. (2019) Identification of a novel cytochrome P450 enzyme that catalyzes the C-2α hydroxylation of pentacyclic triterpenoids and its application in yeast cell factories. Metab Eng 51:70-78. doi.org/10.1016/j.ymben.2018.10.001

Davies K, Jibran R, Zhou Y, Albert N, Brummell D, Jordan B et al. (2020) The evolution of flavonoid biosynthesis: a bryophyte perspective. Front Plant Sci 11:7. doi.org/10.3389/fpls.2020.00007

de Vries J, Stanton A, Archibald JM, Gould SB (2016) Streptophyte terrestrialization in light of plastid evolution. Trends Plant Sci 21:467-476. doi.org/10.1016/j.tplants.2016.01.021

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. doi.org/10.1093/pcp/pcx037

Delwiche CF, Cooper ED (2015) The evolutionary origin of a terrestrial flora. Curr Biol 25:899-910. doi.org/10.1016/j.cub.2015.08.029

Denisov IG, Baas BJ, Grinkova YV, Sligar SG (2007) Cooperativity in cytochrome P450 3A4 linkages in substrate binding, spin state, uncoupling, and product formation. J Biol Chem 282:7066-7076. doi.org/10.1074/jbc.M609589200

Duckett JG, Renzaglia KS (1988) Ultrastructure and development of plastids in bryophytes. Advances in Bryology 3:33-93.

Durst F, O'Keefe DP (1995) Plant cytochromes P450: an overview. Drug metabolism and drug interactions 12:171-188. doi.org/10.1515/DMDI.1995.12.3-4.171

Dymond JS (2013) *Saccharomyces cerevisiae* growth media. Method Enzymol 533:191-204. Academic Press. 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. doi.org/10.1007/s11103-008-9420-7

Ehlting J, Büttner D, Wang Q, Douglas CJ, Somssich IE, Kombrink E (1999) Three 4-coumarate: coenzyme A ligases in *Arabidopsis thaliana* represent two evolutionarily divergent classes in angiosperms. Plant J 19:9-20. doi.org/10.1046/j.1365-313X.1999.00491.x

Ehlting J, Hamberger B, Million-Rousseau R, Werck-Reichhart D (2006) Cytochromes P450 in phenolic metabolism. Phytochem Rev 5:239-270. doi.org/10.1007/s11101-006-9025-1

Emiliani G, Fondi M, Fani R, Gribaldo S (2009) A horizontal gene transfer at the origin of phenylpropanoid metabolism: a key adaptation of plants to land. Biol Direct 4:7. doi.org/10.1186/1745-6150-4-7

Engel PP (1968) The induction of biochemical and morphological mutants in the moss *Physcomitrella patens*. American J Bot 55:438-446. doi.org/10.1002/j.1537-2197.1968.tb07397.x

Escamilla-Treviño LL, Shen H, Hernandez T, Yin Y, Xu Y, Dixon RA (2014) Early lignin pathway enzymes and routes to chlorogenic acid in switchgrass (*Panicum virgatum* L.). Plant Mol Biol 84:565-576. doi.org/10.1007/s11103-013-0152-y

Espiñeira JM, Uzal EN, Ros LG, Carrión JS, Merino F, Barceló AR, Pomar F (2011) Distribution of lignin monomers and the evolution of lignification among lower plants. Plant Biol 13:59-68. doi.org/10.1111/j.1438-8677.2010.00345.x

Eudes A, Pereira JH, Yogiswara S, Wang G, Teixeira Benites V, Baidoo EE et al. (2016) Exploiting the substrate promiscuity of hydroxycinnamoyl-CoA: shikimate hydroxycinnamoyl transferase to reduce lignin. Plant Cell Physiol 57:568-579. doi.org/10.1093/pcp/pcw016

Ferrer J, Austin MB, Stewart C, Noel JP (2008) Structure and function of enzymes involved in the biosynthesis of phenylpropanoids. Plant Physiol Biochem 46:356–370. doi.org/10.1016/j.plaphy.2007.12.009

Fowke LC, Pickett-Heaps JD (1969) Cell division in Spirogyra. ii. Cytokinesis. J Phycol 5:273-281. doi.org/10.1111/j.1529-8817.1969.tb02614.x

Franke R, Humphreys JM, Hemm MR, Denault JW, Ruegger MO, Cusumano JC, Chapple C (2002) The *Arabidopsis* REF8 gene encodes the 3-hydroxylase of phenylpropanoid metabolism. Plant J 30:33-45. doi.org/10.1046/j.1365-313X.2002.01266.x

Fulda M, Shockey J, Werber M, Wolter FP, Heinz E (2002) Two long-chain acyl-CoA synthetases from *Arabidopsis thaliana* involved in peroxisomal fatty acid β-oxidation. Plant J 32:93-103. doi.org/10.1046/j.1365-313X.2002.01405.x

Gabriac B, Werck-Reichhart D, Teutsch H, Durst F (1991) Purification and immunocharacterization of a plant cytochrome P450: the cinnamic acid 4-hydroxylase. Arch Biochem Biophys 288:302-309. doi.org/10.1016/0003-9861(91)90199-S

Galway ME, Hardham AR (1991) Immunofluorescent localization of microtubules throughout the cell cycle in the green alga *Mougeotia* (Zygnemataceae). Am J Bot 78:451-461. doi.org/10.1002/j.1537-2197.1991.tb15211.x

Gamenara D, Seoane G, Méndez PS, de María PD (2013) Redox biocatalysis: fundamentals and applications. Wiley, Hoboken, New Jersey. doi.org/10.1002/9781118409343

Gang DR, Beuerle T, Ullmann P, Werck-Reichhart D, Pichersky E (2002) Differential production of *meta* hydroxylated phenylpropanoids in sweet basil peltate glandular trichomes and leaves is controlled by the activities of specific acyltransferases and hydroxylases. Plant Physiol 130:1536-1544. doi.org/10.1104/pp.007146

Gensel PG, Edwards D (2001) Plants invade the land: evolutionary and environmental perspectives. Columbia University Press, New York. doi.org/10.7312/gens11160

Gertlowski C, Petersen M (1993) Influence of the carbon source on growth and rosmarinic acid production in suspension cultures of *Coleus blumei*. Plant Cell Tiss Org 34:183-190. doi.org/10.1007/BF00036100

Gerrienne P, Gonez P (2011) Early evolution of life cycles in embryophytes: a focus on the fossil evidence of gametophyte/sporophyte size and morphological complexity. J Syst Evol 49:1-16. doi.org/10.1111/j.1759-6831.2010.00096.x

Gietz RD, Schiestl RH (2007) High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat Protoc 2:31-34. doi.org/10.1038/nprot.2007.13

Goiris K, Muylaert K, Voorspoels S, Noten B, De Paepe D, E Baart GJ, De Cooman L (2014) Detection of flavonoids in microalgae from different evolutionary lineages. J Phycol 50:483-492. doi.org/10.1111/jpy.12180

Gravot A, Larbat R, Hehn A, Lievre K, Gontier E, Goergen JL, Bourgaud F (2004) Cinnamic acid 4-hydroxylase mechanism-based inactivation by psoralen derivatives: cloning

and characterization of a C4H from a psoralen producing plant—*Ruta graveolens*—exhibiting low sensitivity to psoralen inactivation. Arch Biochem Biophys 422:71-80. doi.org/10.1016/j.abb.2003.12.013

Gross GG, Zenk MH (1974) Isolation and properties of hydroxycinnamate: CoA ligase from lignifying tissue of *Forsthia*. Eur J Biochem 42:453-459. doi.org/10.1111/j.1432-1033.1974.tb03359.x

Guengerich FP (2001) Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. Chem Res Toxicol 14:611-650. doi.org/10.1021/tx0002583

Gui J, Shen J, Li L (2011) Functional characterization of evolutionarily divergent 4-coumarate: coenzyme A ligases in rice. Plant Physiol 157:574-586. doi.org/10.1104/pp.111.178301

Ha CM, Escamilla-Trevino L, Yarce JCS, Kim H, Ralph J, Chen F, Dixon RA (2016) An essential role of caffeoyl shikimate esterase in monolignol biosynthesis in *Medicago truncatula*. Plant J 86:363-375. doi.org/10.1111/tpj.13177

Haas E, Horecker BL, Hogness TR (1940) The enzymatic reduction of cytochrome c. Cytochrome c reductase. J Biol Chem 136:747-774.

Haas J, Park EC, Seed B (1996) Codon usage limitation in the expression of HIV-1 envelope glycoprotein. Curr Biol 6:315-324. doi.org/10.1016/S0960-9822(02)00482-7

Halkier BA, Møller BL (1991) Involvement of cytochrome P-450 in the biosynthesis of Dhurrin in *Sorghum bicolor* (L.) Moench. Plant Physiol 96:10-17. doi.org/10.1104/pp.96.1.10

Halkier BA (1996) Catalytic reactivities and structure/function relationships of cytochrome P450 enzymes. Phytochemistry 43:1-21. doi.org/10.1016/0031-9422(96)00263-4

Hallahan DL, Lau SMC, Harder PA, Smiley DW, Dawson GW, Pickett JA et al. (1994) Cytochrome P-450-catalysed monoterpenoid oxidation in catmint (*Nepeta racemosa*) and avocado (*Persea americana*); evidence for related enzymes with different activities. Biochim Biophys Acta 1201:94-100. doi.org/10.1016/0304-4165(94)90156-2

Hamberger B, Bak S (2013) Plant P450s as versatile drivers for evolution of species-specific chemical diversity. Philos Tr Roy Soc B: 368:20120426. doi.org/10.1098/rstb.2012.0426

Hanahan D (1983) Studies on transformation of *Escherichia coli* with plasmids. J Mol Biol 166:557-580. doi.org/10.1016/S0022-2836(83)80284-8

Hannemann F, Bichet A, Ewen KM, Bernhardt R (2007) Cytochrome P450 systems—biological variations of electron transport chains. Biochim Biophys Acta 1770:330-344. doi.org/10.1016/j.bbagen.2006.07.017

Harborne JB (1966) Caffeic acid ester distribution in higher plants. Z Naturforsch B 21:604-605. doi.org/10.1515/znb-1966-0634

Hasemann CA, Kurumbail RG, Boddupalli SS, Peterson JA, Deisenhofer J (1995) Structure and function of cytochromes P450: a comparative analysis of three crystal structures. Structure 3:41-62. doi.org/10.1016/S0969-2126(01)00134-4

Hausjell J, Schendl D, Weissensteiner J, Molitor C, Halbwirth H, Spadiut O (2019) Recombinant production of a hard-to-express membrane-bound cytochrome P450 in different yeasts–comparison of physiology and productivity. Yeast. doi.org/10.1002/yea.3441

Hobbs KTC, Tartoff K (1987) Improved media for growing plasmid and cosmid clones. Focus 9:9-12.

Hodgson AV, Strobel HW (1996) Characterization of the FAD binding domain of cytochrome P450 reductase. Arch Biochem Biophys 325:99-106. doi.org/10.1006/abbi.1996.0012

Hoffmann L, Besseau S, Geoffroy P, Ritzenthaler C, Meyer D, Lapierre C et al. (2004) Silencing of hydroxycinnamoyl-coenzyme A shikimate/quinate hydroxycinnamoyltransferase affects phenylpropanoid biosynthesis. Plant Cell 16:1446-1465. doi.org/10.1105/tpc.020297

Hotze M, Schröder G, Schröder J (1995) Cinnamate 4-hydroxylase from *Catharanthus roseus* and a strategy for the functional expression of plant cytochrome P450 proteins as translational fusions with P450 reductase in *Escherichia coli*. FEBS Lett 374:345-350. doi.org/10.1016/0014-5793(95)01141-z

Hu WJ, Kawaoka A, Tsai CJ, Lung J, Osakabe K, Ebinuma H, Chiang VL (1998) Compartmentalized expression of two structurally and functionally distinct 4-coumarate: CoA ligase genes in aspen (*Populus tremuloides*). P Natl Acad Sci USA 95:5407-5412. doi.org/10.1073/pnas.95.9.5407

Hubbard PA, Shen AL, Paschke R, Kasper CB, Kim JJP (2001) NADPH-cytochrome P450 oxidoreductase structural basis for hydride and electron transfer. J Biol Chem 276:29163-29170. doi.org/10.1074/jbc.M101731200

Ishihara A, Ohtsu Y, Iwamura H (1999) Biosynthesis of oat avenanthramide phytoalexins. Phytochemistry 50:237-242. doi.org/10.1016/S0031-9422(98)00535-4

Jennewein S, Park H, DeJong JM, Long RM, Bollon AP, Croteau RB (2005) Coexpression in yeast of *Taxus* cytochrome P450 reductase with cytochrome P450 oxygenases

involved in Taxol biosynthesis. Biotechnol Bioeng 89:588-598. doi.org/10.1002/bit.20390

Jennings AC (1981) The determination of dihydroxy phenolic compounds in extracts of plant tissues. Anal Biochem 118:396-398. doi.org/10.1016/0003-2697(81)90600-X

Jensen K, Møller BL (2010) Plant NADPH-cytochrome P450 oxidoreductases. Phytochemistry 71:132-141. doi.org/10.1016/j.phytochem.2009.10.017

Kahn RA, Durst F (2000) Function and evolution of plant cytochrome P450. Recent Adv Phytochem 34:151-190. doi.org/10.1016/S0079-9920(00)80007-6

Karamat F, Olry A, Doerper S, Vialart G, Ullmann P, Werck-Reichhart D et al. (2012) CYP98A22, a phenolic ester 3'-hydroxylase specialized in the synthesis of chlorogenic acid, as a new tool for enhancing the furanocoumarin concentration in *Ruta graveolens*. BMC Plant Biol 12:152. doi.org/10.1186/1471-2229-12-152

Karol KG, Arumuganathan K, Boore JL, Duffy AM, Everett KD, Hall JD et al. (2010) Complete plastome sequences of *Equisetum arvense* and *Isoetes flaccida*: implications for phylogeny and plastid genome evolution of early land plant lineages. BMC Evol Biol 10:321. doi.org/10.1186/1471-2148-10-321

Karplus PA, Daniels MJ (1991) Atomic structure of ferredoxin-$NADP^+$ reductase: prototype for a structurally novel flavoenzyme family. Science 251:60-66. doi.org/10.1126/science.1986412

Kaur S, Mondal P (2014) Study of total phenolic and flavonoid content, antioxidant activity and antimicrobial properties of medicinal plants. J Microbiol Exp 1:00005. doi.org/10.15406/jmen.2014.01.00005

Kawai S, Mori A, Shiokawa T, Kajita S, Katayama Y, Morohoshi N (1996) Isolation and analysis of cinnamic acid 4-hydroxylase homologous genes from a hybrid aspen, *Populus kitakamiensis*. Biosci Biotechnol Biochem 60:1586-1597. doi.org/10.1271/bbb.60.1586

Kenrick P, Crane PR (1997a) The origin and early evolution of plants on land. Nature 389:33. doi.org/10.1038/37918

Kenrick P, Crane PR (1997b) The origin and early diversification of land plants. Smithsonian Institution, Washington DC. doi.org/10.1017/S0016756899242367

Kenrick P, Wellman CH, Schneider H, Edgecombe GD (2012) A timeline for terrestrialization: consequences for the carbon cycle in the Palaeozoic. Philos T Roy Soc B 367:519-536. doi.org/10.1098/rstb.2011.0271

Kenrick P (2017) How land plant life cycles first evolved. Science 358:1538-1539. doi.org/10.1126/science.aan2923

Khairul Ikram NKB, Beyraghdar Kashkooli A, Peramuna AV, van der Krol AR, Bouwmeester H, Simonsen HT (2017) Stable production of the antimalarial drug artemisinin in the moss *Physcomitrella patens*. Front Bioeng Biotechnol 5:47. doi.org/10.3389/fbioe.2017.00047

Kienow L, Schneider K, Bartsch M, Stuible HP, Weng H, Miersch O et al. (2008) Jasmonates meet fatty acids: functional analysis of a new acyl-coenzyme A synthetase family from *Arabidopsis thaliana*. J Exp Bot 59:403-419. doi.org/10.1093/jxb/erm325

Kim DH, Kim BG, Lee HJ, Lim Y, Hur HG, Ahn JH (2005) Enhancement of isoflavone synthase activity by co-expression of P450 reductase from rice. Biotechnol Lett 27:1291-1294. doi.org/10.1007/s10529-005-0221-7

Kim J, Choi B, Natarajan S, Bae H (2013) Expression analysis of kenaf cinnamate 4-hydroxylase (C4H) ortholog during developmental and stress responses. Plant Omics 6:65.

Klempien A, Kaminaga Y, Qualley A, Nagegowda DA, Widhalm JR, Orlova I et al. (2012) Contribution of CoA ligases to benzenoid biosynthesis in petunia flowers. Plant Cell 24:2015-2030. doi.org/10.1105/tpc.112.097519

Kliebenstein DJ, D'Auria JC, Behere AS, Kim JH, Gunderson KL, Breen JN et al. (2007) Characterization of seed-specific benzoyloxyglucosinolate mutations in *Arabidopsis thaliana*. Plant J 51:1062-1076. doi.org/10.1111/j.1365-313X.2007.03205.x

Knobloch KH, Hahlbrock K (1975) Isoenzymes of p-coumarate: CoA ligase from cell suspension cultures of *Glycine max*. Eur J Biochem 52:311-320. doi.org/10.1111/j.1432-1033.1975.tb03999.x

Knobloch KH, Hahlbrock K (1977) 4-Coumarate: CoA ligase from cell suspension cultures of *Petroselinum hortense* Hoffm: partial purification, substrate specificity, and further properties. Arch Biochem Biophys 184:237-248. doi.org/10.1016/0003-9861(77)90347-2

Koo AJ, Chung HS, Kobayashi Y, Howe GA (2006) Identification of a peroxisomal acyl-activating enzyme involved in the biosynthesis of jasmonic acid in *Arabidopsis*. J Biol Chem 281:33511-33520. doi.org/10.1074/jbc.M607854200

Koopmann E, Hahlbrock K (1997) Differentially regulated NADPH: cytochrome P450 oxidoreductases in parsley. P Natl Acad Sci USA 94:14954-14959. doi.org/10.1073/pnas.94.26.14954

Koopmann E, Logemann E, Hahlbrock K (1999) Regulation and functional expression of cinnamate 4-hydroxylase from parsley. Plant Physiol 119:49-56. doi.org/10.1104/pp.119.1.49

Koprivova A, Altmann F, Gorr G, Kopriva S, Reski R, Decker EL (2003) *N*-glycosylation in the moss *Physcomitrella patens* is organized similarly to higher plants. Plant Biol 5:582-591. doi.org/10.1055/s-2003-44721

Labeeuw L, Martone PT, Boucher Y, Case RJ (2015) Ancient origin of the biosynthesis of lignin precursors. Biol Direct 10:23. doi.org/10.1186/s13062-015-0052-y

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680-685. doi.org/10.1038/227680a0

Lang D, Ullrich KK, Murat F, Fuchs J, Jenkins J, Haas FB et al. (2018) The *Physcomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution. Plant J 93:515-533. doi.org/10.1111/tpj.13801

Le Bail A, Scholz S, Kost B (2013) Evaluation of reference genes for RT qPCR analyses of structure-specific and hormone regulated gene expression in *Physcomitrella patens* gametophytes. PloS one 8:e70998 doi.org/10.1371/journal.pone.0070998

Lee D, Douglas CJ (1996) Two divergent members of a tobacco 4-coumarate: coenzyme A ligase (4CL) gene family (cDNA structure, gene inheritance and expression, and properties of recombinant proteins). Plant Physiol 112:193-205. doi.org/10.1104/pp.112.1.193

Leitch IJ, Bennett MD (2007) Genome size and its uses: the impact of flow cytometry. Flow cytometry with plant cells: analysis of genes, chromosomes and genomes 153-176. doi.org/10.1002/9783527610921.ch7

Lenton TM, Crouch M, Johnson M, Pires N, Dolan L (2012) First plants cooled the Ordovician. Nat Geosci 5:86. doi.org/10.1038/ngeo1390

Leonard E, Yan Y, Koffas MA (2005) Functional expression of a P450 flavonoid hydroxylase for the biosynthesis of plant-specific hydroxylated flavonols in *Escherichia coli*. Metab Eng 8:172-181. doi.org/10.1016/j.ymben.2005.11.001

Leonard E, Koffas MA (2007) Engineering of artificial plant cytochrome P450 enzymes for synthesis of isoflavones by *Escherichia coli*. Appl Environ Microbiol 73:7246-7251. doi.org/10.1128/AEM.01411-07

Li Y, Im Kim J, Pysh L, Chapple C (2015) Four isoforms of *Arabidopsis* 4-coumarate: CoA ligase have overlapping yet distinct roles in phenylpropanoid metabolism. Plant Physiol 169:2409-2421. doi.org/10.1104/pp.15.00838

Li Z, Nair SK (2015) Structural basis for specificity and flexibility in a plant 4-coumarate: CoA ligase. Structure 23:2032-2042. doi.org/10.1016/j.str.2015.08.012

Ligrone R, Carafa A, Duckett JG, Renzaglia KS, Ruel K (2008) Immunocytochemical detection of lignin-related epitopes in cell walls in bryophytes and the charalean alga *Nitella*. Plant Syst Evol 270:257-272. doi.org/10.1007/s00606-007-0617-z

Ligrone R, Duckett JG, Renzaglia KS (2012a) Major transitions in the evolution of early land plants: a bryological perspective. Ann Bot 109:851-871. doi.org/10.1093/aob/mcs017

Ligrone R, Duckett JG, Renzaglia KS (2012b) The origin of the sporophyte shoot in land plants: a bryological perspective. Ann Bot 110:935-941. doi.org/10.1093/aob/mcs176

Lindermayr C, Möllers B, Fliegmann J, Uhlmann A, Lottspeich F, Meimberg H, Ebel J (2002) Divergent members of a soybean (*Glycine max* L.) 4-coumarate: coenzyme A ligase gene family: Primary structures, catalytic properties, and differential expression. Eur J Biochem 269:1304-1315. doi.org/10.1046/j.1432-1033.2002.02775.x

Liu Z, Tavares R, Forsythe ES, André F, Lugan R, Jonasson G et al. (2016) Evolutionary interplay between sister cytochrome P450 genes shapes plasticity in plant metabolism. Nat Commun 7:13026. doi.org/10.1038/ncomms13026

Liu XY, Yu HN, Gao S, Wu YF, Cheng AX, Lou HX (2017) The isolation and functional characterization of three liverwort genes encoding cinnamate 4-hydroxylase. Plant Physiol Biochem 117:42-50. doi.org/10.1016/j.plaphy.2017.05.016

Lopez-Obando M, Hoffmann B, Géry C, Guyon-Debast A, Téoulé E, Rameau C et al. (2016) Simple and efficient targeting of multiple genes through CRISPR-Cas9 in *Physcomitrella patens*. G3: Genes, Genomes, Genetics 6:3647-3653. doi.org/10.1534/g3.116.033266

Lu AY, Coon MJ (1968) Role of hemoprotein P-450 in fatty acid ω-hydroxylation in a soluble enzyme system from liver microsomes. J Biol Chem 243:1331-1332.

Lu AY, Junk KW, Coon MJ (1969) Resolution of the cytochrome P-450-containing ω-hydroxylation system of liver microsomes into three components. J Biol Chem 244:3714-3721.

Lu HP, Luo T, Fu HW, Wang L, Tan YY, Huang JZ et al. (2018) Resistance of rice to insect pests mediated by suppression of serotonin biosynthesis. Nat Plants 4:338-344. doi.org/10.1038/s41477-018-0152-7

Mahesh V, Million-Rousseau R, Ullmann P, Chabrillange N, Bustamante J, Mondolot L et al. (2007) Functional characterization of two p-coumaroyl ester 3’-hydroxylase genes from coffee tree: evidence of a candidate for chlorogenic acid biosynthesis. Plant Mol Biol 64:145-159. doi.org/10.1007/s11103-007-9141-3

Martone PT, Estevez JM, Lu F, Ruel K, Denny MW, Somerville C, Ralph J (2009) Discovery of lignin in seaweed reveals convergent evolution of cell-wall architecture. Curr Biol 19:169-175. doi.org/10.1016/j.cub.2008.12.031

Matsuno M, Nagatsu A, Ogihara Y, Ellis BE, Mizukami H (2002) CYP98A6 from *Lithospermum erythrorhizon* encodes 4-coumaroyl-4'-hydroxyphenyllactic acid 3-hydroxylase involved in rosmarinic acid biosynthesis. FEBS Lett 514:219-224. doi.org/10.1016/S0014-5793(02)02368-2

Matsuno M, Compagnon V, Schoch GA, Schmitt M, Debayle D, Bassard JE et al. (2009) Evolution of a novel phenolic pathway for pollen development. Science 325:1688-1692. doi.org/10.1126/science.1174095

Meehan TD, Coscia CJ (1973) Hydroxylation of geraniol and nerol by a monooxygenase from *Vinca rosea*. Biochem Bioph Res Co 53:1043-1048. doi.org/10.1016/0006-291X(73)90570-6

Meijer AH, Cardoso MIL, Voskuilen JT, de Waal A, Verpoorte R, Hoge JHC (1993) Isolation and characterization of a cDNA clone from *Catharanthus roseus* encoding NADPH: cytochrome P-450 reductase, an enzyme essential for reactions catalysed by cytochrome P-450 mono-oxygenases in plants. Plant J 4:47-60. doi.org/10.1046/j.1365-313X.1993.04010047.x

Meunier B, de Visser SP, Shaik S (2004) Mechanism of oxidation reactions catalyzed by cytochrome P450 enzymes. Chem Rev 104:3947-3980. doi.org/10.1021/cr020443g

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. doi.org/10.1104/pp.113.3.755

Mizutani M, Ohta D (1998) Two Isoforms of NADPH: Cytochrome P450 Reductase in *Arabidopsis thaliana*: Gene Structure, Heterologous Expression in Insect Cells, and Differential Regulation. Plant Physiol 116:357-367. doi.org/10.1104/pp.116.1.35

Moglia A, Comino C, Portis E, Acquadro A, De Vos RC, 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. doi.org/10.1007/s00299-009-0695-1

Morant M, Bak S, Møller BL, Werck-Reichhart D (2003) Plant cytochromes P450: tools for pharmacology, plant protection and phytoremediation. Curr Opin Biotech 14:151-162. doi.org/10.1016/S0958-1669(03)00024-7

Morant M, Schoch GA, Ullmann P, Ertunç T, Little D, Olsen CE et al. (2007) Catalytic activity, duplication and evolution of the CYP98 cytochrome P450 family in wheat. Plant Mol Biol 63:1-19. doi.org/10.1007/s11103-006-9028-8

Murugan K, Krishnan R (2013) Phytochemical analysis, in vitro antifungal activity and mode of action of ethanolic extract of *Marchantia linearis* Lehm & Lindenb. A bryophyte. World J Pharm Pharm Sci 2:3650-66.

Nair PM, Vining LC (1965) Cinnamic acid hydroxylase in spinach. Phytochemistry 4:161-168. doi.org/10.1016/S0031-9422(00)86159-2

Nakase T, Komagata K (1971) Significance of DNA base composition in the classification of yeast genus *Saccharomyces*. J Gen Appl Microbiol 17:227-238. doi.org/10.2323/jgam.17.227

Narayanasami R, Horowitz PM, Masters BSS (1995) Flavin-binding and protein structural integrity studies on NADPH-cytochrome P450 reductase are consistent with the presence of distinct domains. Arch Biochem Biophys 316:267-274. doi.org/10.1006/abbi.1995.1037

Nelson DR (2006a) Cytochrome P450 Nomenclature, 2004. In: Phillips IR, Shephard EA (eds) Cytochrome P450 Protocols. Methods in Molecular Biology (Vol. 320). Humana Press, Totowa, NJ, pp. 1-10. doi.org/10.1385/1-59259-998-2:1

Nelson DR (2006b) Plant cytochrome P450s from moss to poplar. Phytochem Rev 5:193-204. doi.org/10.1007/s11101-006-9015-3

Ni ZY, Li B, Neumann MP, Lü M, Fan L (2014) Isolation and expression analysis of two genes encoding cinnamate 4-hydroxylase from cotton (*Gossypium hirsutum*). J Integr Agr 13:2102-2112. doi.org/10.1016/S2095-3119(13)60643-7

Nickrent DL, Parkinson CL, Palmer JD, Duff RJ (2000) Multigene phylogeny of land plants with special reference to bryophytes and the earliest land plants. Mol Biol Evol 17:1885-1895. doi.org/10.1093/oxfordjournals.molbev.a026290

Nielsen KA, Møller BL (2005) Cytochrome P450s in plants. In: Ortiz de Montellano PR (ed) Cytochrome P450. Springer, Boston, MA, pp. 553-583. Springer, Boston, MA. doi.org/10.1007/0-387-27447-2_12

Niklas KJ, Cobb ED, Matas AJ (2017) The evolution of hydrophobic cell wall biopolymers: from algae to angiosperms. J Exp Bot 68:5261-5269. doi.org/10.1093/jxb/erx215

Nishiyama T, Wolf PG, Kugita M, Sinclair RB, Sugita M, Sugiura C et al. (2004) Chloroplast phylogeny indicates that bryophytes are monophyletic. Mol Biol Evol 21:1813-1819. doi.org/10.1093/molbev/msh203

Nomura T, Sakurai T, Osakabe Y, Osakabe K, Sakakibara H (2016) Efficient and heritable targeted mutagenesis in mosses using the CRISPR/Cas9 system. Plant Cell Physiol 57:2600-2610. doi.org/10.1093/pcp/pcw173

Ohta D, Mizutani M (2004) Redundancy or flexibility: molecular diversity of the electron transfer components for P450 monooxygenases in higher plants. Front Biosci 9:1587-1597. doi.org/10.2741/1356

Okazaki Y, Isobe T, Iwata Y, Matsukawa T, Matsuda F, Miyagawa H et al. (2004) Metabolism of avenanthramide phytoalexins in oats. Plant J 39:560-572. doi.org/10.1111/j.1365-313X.2004.02163.x

Omura T, Sato R (1964a) The carbon monoxide-binding pigment of liver microsomes I. Evidence for its hemoprotein nature. J Biol Chem 239:2370-2378.

Omura T, Sato R (1964b) The carbon monoxide-binding pigment of liver microsomes II. Solubilization, purification, and properties. J Biol Chem 239:2379-2385.

Pan XW, Han L, Zhang YH, Chen DF, Simonsen HT (2015) Sclareol production in the moss *Physcomitrella patens* and observations on growth and terpenoid biosynthesis. Plant Biotechnol Rep 9:149-159. doi.org/10.1007/s11816-015-0353-8

Perlak FJ, Fuchs RL, Dean DA, McPherson SL, Fischhoff DA (1991) Modification of the coding sequence enhances plant expression of insect control protein genes. P Natl Acad Sci USA 88:3324-3328. doi.org/10.1073/pnas.88.8.3324

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 C 43:501-504. doi.org/10.1515/znc-1988-7-804

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. doi.org/10.1007/BF00201337

Petersen M (1997) Cytochrome P450-dependent hydroxylation in the biosynthesis of rosmarinic acid in *Coleus*. Phytochemistry 45:1165-1172. doi.org/10.1016/S0031-9422(97)00135-0

Petersen M (2003) Cinnamic acid 4-hydroxylase from cell cultures of the hornwort *Anthoceros agrestis*. Planta 217:96-101. doi.org/10.1007/s00425-002-0960-9

Petersen M, Simmonds MS (2003) Rosmarinic acid. Phytochemistry 62:121-125. doi.org/10.1016/S0031-9422(02)00513-7

Petersen M (2013) Rosmarinic acid: new aspects. Phytochem Rev 12:207-227. doi.org/10.1007/s11101-013-9282-8

Petersen M (2016) Hydroxycinnamoyltransferases in plant metabolism. Phytochem Rev 15:699-727. doi.org/10.1007/s11101-015-9417-1

Pezeshki S (2016) Biosynthese von Kaffeesäuremetaboliten im Ackerhornmoos *Anthoceros agrestis* und im Kleinen Blasenmützenmoos *Physcomitrella patens* Doctoral thesis, Philipps-Universität Marburg. doi.org/10.17192/z2016.0084

Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT–PCR. Nucleic Acids Res 29:e45-e45. doi.org/10.1093/nar/29.9.e45

Pichrtová M, Remias D, Lewis LA, Holzinger A (2013) Changes in phenolic compounds and cellular ultrastructure of Arctic and Antarctic strains of *Zygnema* (Zygnematophyceae, Streptophyta) after exposure to experimentally enhanced UV to PAR ratio. Microb Ecol 65:68-83. doi.org/10.1007/s00248-012-0096-9

Pickett-Heaps JD, Wetherbee R (1987) Spindle function in the green alga *Mougeotia*: absence of anaphase A correlates with postmitotic nuclear migration. Cell Motil Cytoskel 7:68-77. doi.org/10.1002/cm.970070109

Pompon D, Louerat B, Bronine A, Urban P (1996) Yeast expression of animal and plant P450s in optimized redox environments. Method Enzymol 272:51-64. Academic Press. doi.org/10.1016/S0076-6879(96)72008-6

Porter TD, Kasper CB (1986) NADPH-cytochrome P-450 oxidoreductase: flavin mononucleotide and flavin adenine dinucleotide domains evolved from different flavoproteins. Biochemistry 25:1682-1687. doi.org/10.1021/bi00355a036

Proust H, Honkanen S, Jones VA, Morieri G, Prescott H, Kelly S et al. (2016) RSL class I genes controlled the development of epidermal structures in the common ancestor of land plants. Curr Biol 26:93-99. doi.org/10.1016/j.cub.2015.11.042

Pu G, Wang P, Zhou B, Liu Z, Xiang F (2013) Cloning and characterization of *Lonicera japonica p*-coumaroyl ester 3-hydroxylase which is involved in the biosynthesis of chlorogenic acid. Biosci Biotech Biochem 77:1403-1409. doi.org/10.1271/bbb.130011

Puttick MN, Morris JL, Williams TA, Cox CJ, Edwards D, Kenrick P et al. (2018) The interrelationships of land plants and the nature of the ancestral embryophyte. Curr Biol 28:733-745. doi.org/10.1016/j.cub.2018.01.063

Quinlan RF, Jaradat TT, Wurtzel ET (2007) *Escherichia coli* as a platform for functional expression of plant P450 carotene hydroxylases. Arch Biochem Biophys 458:146-157. doi.org/10.1016/j.abb.2006.11.019

Qiu YL, Li L, Wang B, Chen Z, Knoop V, Groth-Malonek M et al. (2006) The deepest divergences in land plants inferred from phylogenomic evidence. P Natl Acad Sci USA 103:15511-15516. doi.org/10.1073/pnas.0603335103

Qiu YL, Taylor AB, McManus HA (2012) Evolution of the life cycle in land plants. J Syst Evol 50:171-194. doi.org/10.1111/j.1759-6831.2012.00188.x

Ravichandran KG, Boddupalli SS, Hasermann CA, Peterson JA, Deisenhofer J (1993) Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's. Science 261:731-736. doi.org/10.1126/science.8342039

Rana S, Lattoo SK, Dhar N, Razdan S, Bhat WW, Dhar RS, Vishwakarma R (2013) NADPH-cytochrome P450 reductase: molecular cloning and functional characterization of two paralogs from *Withania somnifera* (L.) Dunal. PLoS One 8:e57068. doi.org/10.1371/journal.pone.0057068

Renault H, Alber A, Horst NA, Lopes AB, Fich EA, Kriegshauser L et al. (2017a) A phenol-enriched cuticle is ancestral to lignin evolution in land plants. Nat Commun 8:14713. doi.org/10.1038/ncomms14713

Renault H, De Marothy M, Jonasson G, Lara P, Nelson DR, Nilsson I et al. (2017b) Gene duplication leads to altered membrane topology of a cytochrome P450 enzyme in seed plants. Mol Biol Evol 34:2041-2056. doi.org/10.1093/molbev/msx160

Renault H, Werck-Reichhart D, Weng JK (2019) Harnessing lignin evolution for biotechnological applications. Curr Opin Biotech 56:105-111. doi.org/10.1016/j.copbio.2018.10.011

Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H et al. (2008) The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. Science 319:64-69. doi.org/10.1126/science.1150646

Rensing SA (2018) Great moments in evolution: the conquest of land by plants. Curr Opin Plant Biol 42:49-54. doi.org/10.1016/j.pbi.2018.02.006

Renzaglia KS (1978) Comparative morphology and developmental anatomy of the Anthocerotophyta. J Hattori Bot Lab 44.

Renzaglia KS, Joel Duff R, Nickrent DL, Garbary DJ (2000) Vegetative and reproductive innovations of early land plants: implications for a unified phylogeny. Philos T Roy Soc B 355:769-793. doi.org/10.1098/rstb.2000.0615

Renzaglia KS, Villarreal JC, Duff RJ (2008) New insights into morphology, anatomy, and systematics of hornworts. In: Goffinet B, Shaw Aj (eds) Bryophyte biology. Cambridge University Press, Cambridge, pp 139-171. doi.org/10.1017/CBO9780511754807.004

Reski R (1999) Molecular genetics of *Physcomitrella*. Planta 208:301-309. doi.org/10.1007/s004250050563

Reski R, Parsons J, Decker EL (2015) Moss-made pharmaceuticals: from bench to bedside. Plant Biotechnol J 13:1191-1198. doi.org/10.1111/pbi.12401

Reski R, Bae H, Simonsen HT (2018) *Physcomitrella patens*, a versatile synthetic biology chassis. Plant Cell Rep 37:1409-1417. doi.org/10.1007/s00299-018-2293-6

Reumann S (2004) Specification of the peroxisome targeting signals type 1 and type 2 of plant peroxisomes by bioinformatics analyses. Plant Physiol 135:783-800. doi.org/10.1104/pp.103.035584

Reumann S, Chowdhary G, Lingner T (2016) Characterization, prediction and evolution of plant peroxisomal targeting signals type 1 (PTS1s). Biochim Biophys Acta Molecular Cell Research 1863:790-803. doi.org/10.1016/j.bbamcr.2016.01.001

Ro DK, Ehlting J, Douglas CJ (2002) Cloning, functional expression, and subcellular localization of multiple NADPH-cytochrome P450 reductases from hybrid poplar. Plant Physiol 130:1837-1851. doi.org/10.1104/pp.008011

Rosco A, Pauli HH, Priesner W, Kutchan TM (1997) Cloning and heterologous expression of NADPH-cytochrome P450 reductases from the Papaveraceae. Arch Biochem Biophys 348:369-377. doi.org/10.1006/abbi.1997.0374

Rouwendal GJ, Mendes O, Wolbert EJ, De Boer AD (1997) Enhanced expression in tobacco of the gene encoding green fluorescent protein by modification of its codon usage. Plant Mol Biol 33:989-999. doi.org/10.1023/A:1005740823703

Ruhfel BR, Gitzendanner MA, Soltis PS, Soltis DE, Burleigh JG (2014) From algae to angiosperms–inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. BMC Evol Biol 14:23. doi.org/10.1186/1471-2148-14-23

Russell DW, Conn EE (1967) The cinnamic acid 4-hydraxylase of pea seedlings. Arch Biochem Biophys 122:256-258. doi.org/10.1016/0003-9861(67)90150-6

Russell DW (1971) The metabolism of aromatic compounds in higher plants X. Properties of the cinnamic acid 4-hydroxylase of pea seedlings and some aspects of its metabolic and developmental control. J Biol Chem 246:3870-3878.

Saini SS, Gaid M, Sircar D (2020) Benzoate-CoA ligase contributes to the biosynthesis of biphenyl phytoalexins in elicitor-treated pear cell cultures. Plant Cell Rep 39:207-215. doi.org/10.1007/s00299-019-02484-0

Sander M, Petersen M (2011) Distinct substrate specificities and unusual substrate flexibilities of two hydroxycinnamoyltransferases, rosmarinic acid synthase and hydroxycinnamoyl-CoA: shikimate hydroxycinnamoyl-transferase, from *Coleus blumei* Benth. Planta 233:1157-1171. doi.org/10.1007/s00425-011-1367-2

Schaefer D, Zryd JP, Knight CD, Cove DJ (1991) Stable transformation of the moss *Physcomitrella patens*. Mol Gen Genet 226:418-424. doi.org/10.1007/BF00260654

Schaefer DG, Zrÿd J (1997) Efficient gene targeting in the moss *Physcomitrella patens*. Plant J 11:1195-1206. doi.org/10.1046/j.1365-313X.1997.11061195.x

Schaefer DG (2002) A new moss genetics: targeted mutagenesis in *Physcomitrella patens*. Annu Rev Plant Biol 53:477-501. doi.org/10.1146/annurev.arplant.53.100301.135202

Schneider K, Hövel K, Witzel K, Hamberger B, Schomburg D, Kombrink E, Stuible HP (2003) The substrate specificity-determining amino acid code of 4-coumarate: CoA ligase. P Natl Acad Sci USA 100:8601-8606. doi.org/10.1073/pnas.1430550100

Schneider K, Kienow L, Schmelzer E, Colby T, Bartsch M, Miersch O et al. (2005) A new type of peroxisomal acyl-coenzyme A synthetase from *Arabidopsis thaliana* has the catalytic capacity to activate biosynthetic precursors of jasmonic acid. J Biol Chem 280:13962-13972. doi.org/10.1074/jbc.M413578200

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. doi.org/10.1074/jbc.M104047200

Schoch GA, Morant M, Abdulrazzak N, Asnaghi C, Goepfert S, Petersen M et al. (2006) The *meta*-hydroxylation step in the phenylpropanoid pathway: a new level of complexity in the pathway and its regulation. Environ Chem Lett 4:127-136. doi.org/10.1007/s10311-006-0062-1

Schuler MA (1996) Plant cytochrome P450 monooxygenases. Crit Rev Plant Sci 15:235-284. doi.org/10.1080/07352689609701942

Schuler MA, Werck-Reichhart D (2003) Functional genomics of P450s. Annu Rev Plant Biol 54:629-667. doi.org/10.1146/annurev.arplant.54.031902.134840

Schween G, Gorr G, Hohe A, Reski R (2003) Unique tissue-specific cell cycle in *Physcomitrella*. Plant Biol 5:50-58. doi.org/10.1055/s-2003-37984

Sewalt VJ, Ni W, Blount JW, Jung HG, Masoud SA, Howles PA et al. (1997) Reduced lignin content and altered lignin composition in transgenic tobacco down-regulated in expression of L-phenylalanine ammonia-lyase or cinnamate 4-hydroxylase. Plant Physiol 115:41-50. doi.org/10.1104/pp.115.1.41

Shaik S, Kumar D, de Visser SP, Altun A, Thiel W (2005) Theoretical perspective on the structure and mechanism of cytochrome P450 enzymes. Chem Rev 105:2279-2328. doi.org/10.1021/cr030722j

Shen H, Mazarei M, Hisano H, Escamilla-Trevino L, Fu C, Pu Y et al. (2013) A genomics approach to deciphering lignin biosynthesis in switchgrass. Plant Cell 25:4342-4361. doi.org/10.1105/tpc.113.118828

Shepherd D, Garland PB (1969) Citrate synthase from rat liver:[EC 4.1. 3.7 Citrate oxaloacetage-lyase (CoA-acetylating)]. Method Enzymol 13:11-16. Academic Press. doi.org/10.1016/0076-6879(69)13006-2

Shet MS, Sathasivan K, Arlotto MA, Mehdy MC, Estabrook RW (1993) Purification, characterization, and cDNA cloning of an NADPH-cytochrome P450 reductase from mung bean. P Natl Acad Sci 90:2890-2894. doi.org/10.1073/pnas.90.7.2890

Shockey JM, Fulda MS, Browse J (2003) *Arabidopsis* contains a large superfamily of acyl-activating enzymes. Phylogenetic and biochemical analysis reveals a new class of acyl-coenzyme A synthetases. Plant Physiol 132:1065-1076. doi.org/10.1104/pp.103.020552

Silber MV, Meimberg H, Ebel J (2008) Identification of a 4-coumarate: CoA ligase gene family in the moss, *Physcomitrella patens*. Phytochemistry 69:2449-2456. doi.org/10.1016/j.phytochem.2008.06.014

Smith GC, Tew DG, Wolf CR (1994) Dissection of NADPH-cytochrome P450 oxidoreductase into distinct functional domains. P Natl Acad Sci USA 91:8710-8714. doi.org/10.1073/pnas.91.18.8710

Sono M, Roach MP, Coulter ED, Dawson JH (1996) Heme-containing oxygenases. Chem Rev 96:2841-2888. doi.org/10.1021/cr9500500

Soriano G, Cloix C, Heilmann M, Núñez-Olivera E, Martínez-Abaigar J, Jenkins GI (2018) Evolutionary conservation of structure and function of the UVR 8 photoreceptor from the liverwort *Marchantia polymorpha* and the moss *Physcomitrella patens*. New Phytol 217:151-162. doi.org/10.1111/nph.14767

Soriano G, Del-Castillo-Alonso MÁ, Monforte L, Núñez-Olivera E, Martínez-Abaigar J (2019) Phenolic compounds from different bryophyte species and cell compartments respond specifically to ultraviolet radiation, but not particularly quickly. Plant Physiol Biochem 134:137-144. doi.org/10.1016/j.plaphy.2018.07.020

Stark T, Hofmann T (2005) Isolation, structure determination, synthesis, and sensory activity of N-phenylpropenoyl-L-amino acids from cocoa (*Theobroma cacao*). J Agr Food Chem 53:5419-5428. doi.org/10.1021/jf050458q

Staswick PE, Tiryaki I (2004) The oxylipin signal jasmonic acid is activated by an enzyme that conjugates it to isoleucine in *Arabidopsis*. Plant Cell 16:2117-2127. doi.org/10.1105/tpc.104.023549

Stöckigt J, Zenk MH (1975) Chemical syntheses and properties of hydroxycinnamoyl-coenzyme A derivatives. Z Naturforsch C 30:352-358. doi.org/10.1515/znc-1975-5-609

Stuible HP, Büttner D, Ehlting J, Hahlbrock K, Kombrink E (2000) Mutational analysis of 4-coumarate: CoA ligase identifies functionally important amino acids and verifies its close relationship to other adenylate-forming enzymes. FEBS Lett 467:117-122. doi.org/10.1016/S0014-5793(00)01133-9

Stuible HP, Kombrink E (2001) Identification of the substrate specificity-conferring amino acid residues of 4-coumarate: coenzyme A ligase allows the rational design of mutant enzymes with new catalytic properties. J Biol Chem 276:26893-26897. doi.org/10.1074/jbc.M100355200

Sullivan ML, Zarnowski R (2010) Red clover coumarate 3'-hydroxylase (CYP98A44) is capable of hydroxylating *p*-coumaroyl-shikimate but not *p*-coumaroyl-malate: implications for the biosynthesis of phaselic acid. Planta 231:319-328. doi.org/10.1007/s00425-009-1054-8

Szövényi 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 15:98. doi.org/10.1186/s12870-015-0481-x

Takeda R, Hasegawa J, Shinozaki M (1990) The first isolation of lignans, megacerotonic acid and anthocerotonic acid, from non-vascular plants, Anthocerotae (hornworts). Tetrahedron Lett 31:4159-4162. doi.org/10.1016/S0040-4039(00)97569-5

Tam THY, Catarino B, Dolan L (2015) Conserved regulatory mechanism controls the development of cells with rooting functions in land plants. P Natl Acad Sci USA 112:E3959-E3968. doi.org/10.1073/pnas.1416324112

Thornburg CK, Wortas-Strom S, Nosrati M, Geiger JH, Walker KD (2015) Kinetically and crystallographically guided mutations of a benzoate Coa ligase (BadA) elucidate mechanism and expand substrate permissivity. Biochemistry 54:6230-6242. doi.org/10.1021/acs.biochem.5b00899

Trennheuser F (1992) Phytochemische Untersuchung und in vitro Kultur ausgewählter Vertreter der Anthocerotopsida. Doctoral thesis, Universität des Saarlandes, Saarbrücken

Troitsky AV, Ignatov MS, Bobrova VK, Milyutina IA (2007) Contribution of genosystematics to current concepts of phylogeny and classification of bryophytes. Biochemistry (Moscow) 72:1368-1376. doi.org/10.1134/S0006297907120115

Tsugawa H, Nakabayashi R, Mori T, Yamada Y, Takahashi M, Rai A et al. (2019) A cheminformatics approach to characterize metabolomes in stable-isotope-labeled organisms. Nat Methods 16:295. doi.org/10.1038/s41592-019-0358-2

Urban P, Mignotte C, Kazmaier M, Delorme F, Pompon D (1997) Cloning, yeast expression, and characterization of the coupling of two distantly related *Arabidopsis thaliana* NADPH-cytochrome P450 reductases with P450 CYP73A5. J Biol Chem 272:19176-19186. doi.org/10.1074/jbc.272.31.19176

Vanholme R, Cesarino I, Rataj K, Xiao Y, Sundin L, Goeminne G et al. (2013) Caffeoyl shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway in *Arabidopsis*. Science 341:1103-1106. doi.org/10.1126/science.1241602

Vermilion JL, Coon MJ (1978) Identification of the high and low potential flavins of liver microsomal NADPH-cytochrome P-450 reductase. J Biol Chem 253:8812-8819.

Villarreal JC, Renner SS (2012) Hornwort pyrenoids, carbon-concentrating structures, evolved and were lost at least five times during the last 100 million years. P Natl Acad Sci USA 109:18873-18878. doi.org/10.1073/pnas.1213498109

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. doi.org/10.1007/s00425-005-0089-8

Vogt T (2010) Phenylpropanoid Biosynthesis. Mol Plant 3:2–20. doi.org/10.1093/mp/ssp106

Walker K, Croteau R (2000) Taxol biosynthesis: molecular cloning of a benzoyl-CoA: taxane 2α-*O*-benzoyltransferase cDNA from *Taxus* and functional expression in *Escherichia coli*. P Natl Acad Sci USA 97:13591-13596. doi.org/10.1073/pnas.250491997

Wang M, Roberts DL, Paschke R, Shea TM, Masters BSS, Kim JJP (1997) Three-dimensional structure of NADPH–cytochrome P450 reductase: prototype for FMN- and FAD-containing enzymes. P Natl Acad Sci USA 94:8411-8416. doi.org/10.1073/pnas.94.16.8411

Watenpaugh KD, Sieker LC, Jensen LH (1973) The binding of riboflavin-5'-phosphate in a flavoprotein: flavodoxin at 2.0-Å resolution. P Natl Acad Sci USA 70:3857-3860. doi.org/10.1073/pnas.70.12.3857

Watkins PA (1997) Fatty acid activation. Prog Lipid Res 36:55-83. doi.org/10.1016/S0163-7827(97)00004-0

Weitzel C (2009) Rosmarinsäure-Biosynthese in Suspensionskulturen von *Melissa officinalis* L. Doctoral thesis, Philipps-Universität Marburg. doi.org/10.17192/z2009.0715

Wellman CH, Strother PK (2015) The terrestrial biota prior to the origin of land plants (embryophytes): a review of the evidence. Palaeontology 58:601-627. https://doi.org/10.1111/pala.12172

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. doi.org/10.1105/tpc.109.073528

Werck-Reichhart D (1995) Cytochromes P450 in phenylpropanoid metabolism. Drug metabolism and drug interactions 12:221-244. doi.org/10.1515/DMDI.1995.12.3-4.221

Werck-Reichhart D, Feyereisen R (2000) Cytochromes P450: a success story. Genome Biol 1:reviews3003-1. doi.org/10.1186/gb-2000-1-6-reviews3003

Werck-Reichhart D, Bak S, Paquette S (2002) Cytochromes P450. *Arabidopsis* book 1:e0028. doi.org/10.1199/tab.0028

Werner V (2018) Untersuchungen zur Bildung von Hydroxyzimtsäureestern in *Ocimum basilicum*, *Cichorium intybus* und *Actaea racemosa*. Doctoral thesis, Philipps-Universität Marburg. doi.org/10.17192/z2019.0053

Wickett NJ, Mirarab S, Nguyen N, Warnow T, Carpenter E, Matasci N et al. (2014) Phylotranscriptomic analysis of the origin and early diversification of land plants. P Natl Acad Sci USA 111:E4859-E4868. doi.org/10.1073/pnas.1323926111

Widhalm JR, Dudareva N (2015) A familiar ring to it: biosynthesis of plant benzoic acids. Mol Plant 8:83-97. doi.org/10.1016/j.molp.2014.12.001

Williams CH, Kamin H (1962) Microsomal triphosphopyridine nucleotide-cytochrome c reductase of liver. J Biol Chem 237:587-595.

Wohl J, Petersen M (2020) Functional expression and characterization of cinnamic acid 4-hydroxylase from the hornwort *Anthoceros agrestis* in *Physcomitrella patens*. Plant Cell Rep 39:597–607. doi.org/10.1007/s00299-020-02517-z

Xu H, Park NI, Li X, Kim YK, Lee SY, Park SU (2010) Molecular cloning and characterization of phenylalanine ammonia-lyase, cinnamate 4-hydroxylase and genes involved in flavone biosynthesis in *Scutellaria baicalensis*. Bioresource Technol 101:9715-9722. doi.org/10.1016/j.biortech.2010.07.083

Xu J, Ding Z, Vizcay-Barrena G, Shi J, Liang W, Yuan Z et al. (2014) Aborted Microspores acts as a master regulator of pollen wall formation in *Arabidopsis*. Plant Cell 26:1544-1556. doi.org/10.1105/tpc.114.122986

Yamaoka S, Nishihama R, Yoshitake Y, Ishida S, Inoue K, Saito M et al. (2018) Generative cell specification requires transcription factors evolutionarily conserved in land plants. Curr Biol 28:479-486. doi.org/10.1016/j.cub.2017.12.053

Yang CQ, Lu S, Mao YB, Wang LJ, Chen XY (2010) Characterization of two NADPH: cytochrome P450 reductases from cotton (*Gossypium hirsutum*). Phytochemistry 71:27-35. doi.org/10.1016/j.phytochem.2009.09.026

Yonekura-Sakakibara K, Higashi Y, Nakabayashi R (2019) The origin and evolution of plant flavonoid metabolism. Front Plant Sci 10:943. doi.org/10.3389/fpls.2019.00943

Yoshikawa M, Luo W, Tanaka G, Konishi Y, Matsuura H, Takahashi K (2018) Wounding stress induces phenylalanine ammonia lyases, leading to the accumulation of phenylpropanoids in the model liverwort *Marchantia polymorpha*. Phytochemistry 155:30-36. doi.org/10.1016/j.phytochem.2018.07.014

Zenk MH (1979) Recent work on cinnamoyl CoA derivatives. In: Swain T, Harbone JB, Van Sumere CF (eds) Biochemistry of Plant Phenolics. Recent Advances in Phytochemistry, Vol. 12, pp 139-176. Springer Boston, MA. doi.org/10.1007/978-1-4684-3372-2_5

Zhan X, Zhang YH, Chen DF, Simonsen HT (2014) Metabolic engineering of the moss *Physcomitrella patens* to produce the sesquiterpenoids patchoulol and α/β-santalene. Front Plant Sci 5:636. doi.org/10.3389/fpls.2014.00636

Zhao Q, Modi S, Smith G, Paine M, McDonagh PD, Wolf CR, Tew D, Lian LY, Roberts GC, Driessen HP (1999) Crystal structure of the FMN-binding domain of human cytochrome P450 reductase at 1.93 Å resolution. Protein Sci 8:298-306. doi.org/10.1110/ps.8.2.298

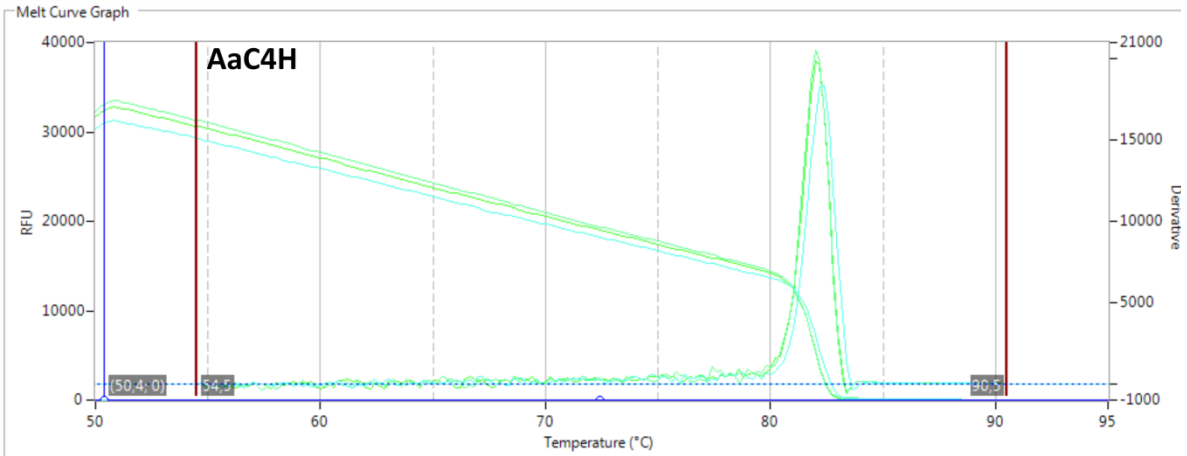
Zhang J, Fu XX, Li RQ, Zhao X, Liu Y, Li MH et al. (2020) The hornwort genome and early land plant evolution. Nat Plants 6:107-118. doi.org/10.1038/s41477-019-0588-4

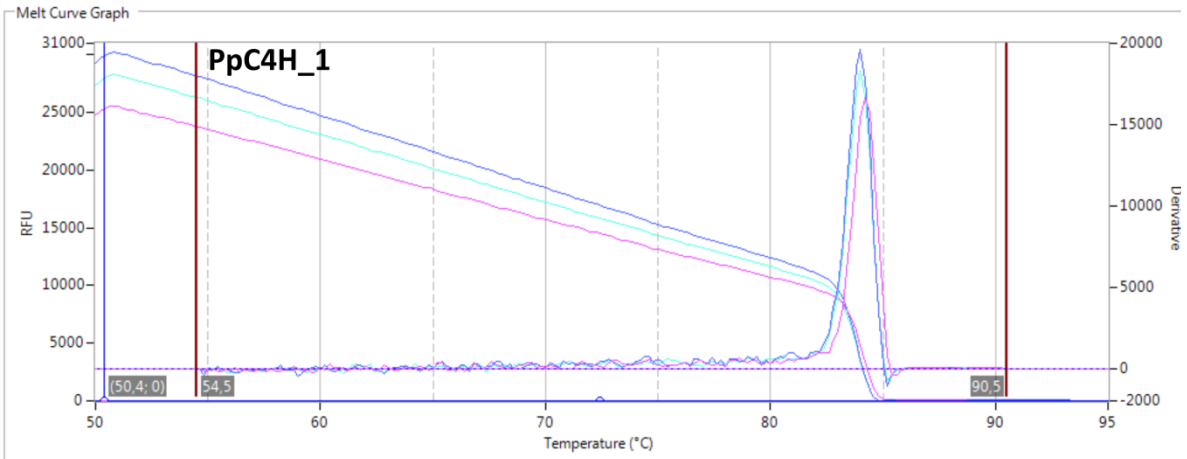
Zinsmeister HD, Becker H, Eicher T (1991) Bryophytes, a source of biologically active, naturally occurring material? Chemistry 30:130-147. doi.org/10.1002/anie.199101301

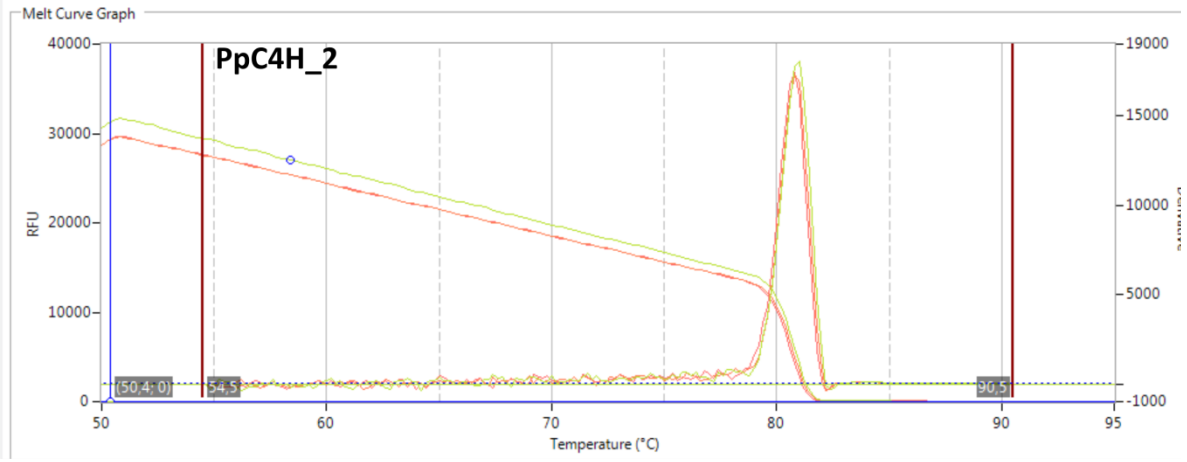
VIII Appendix

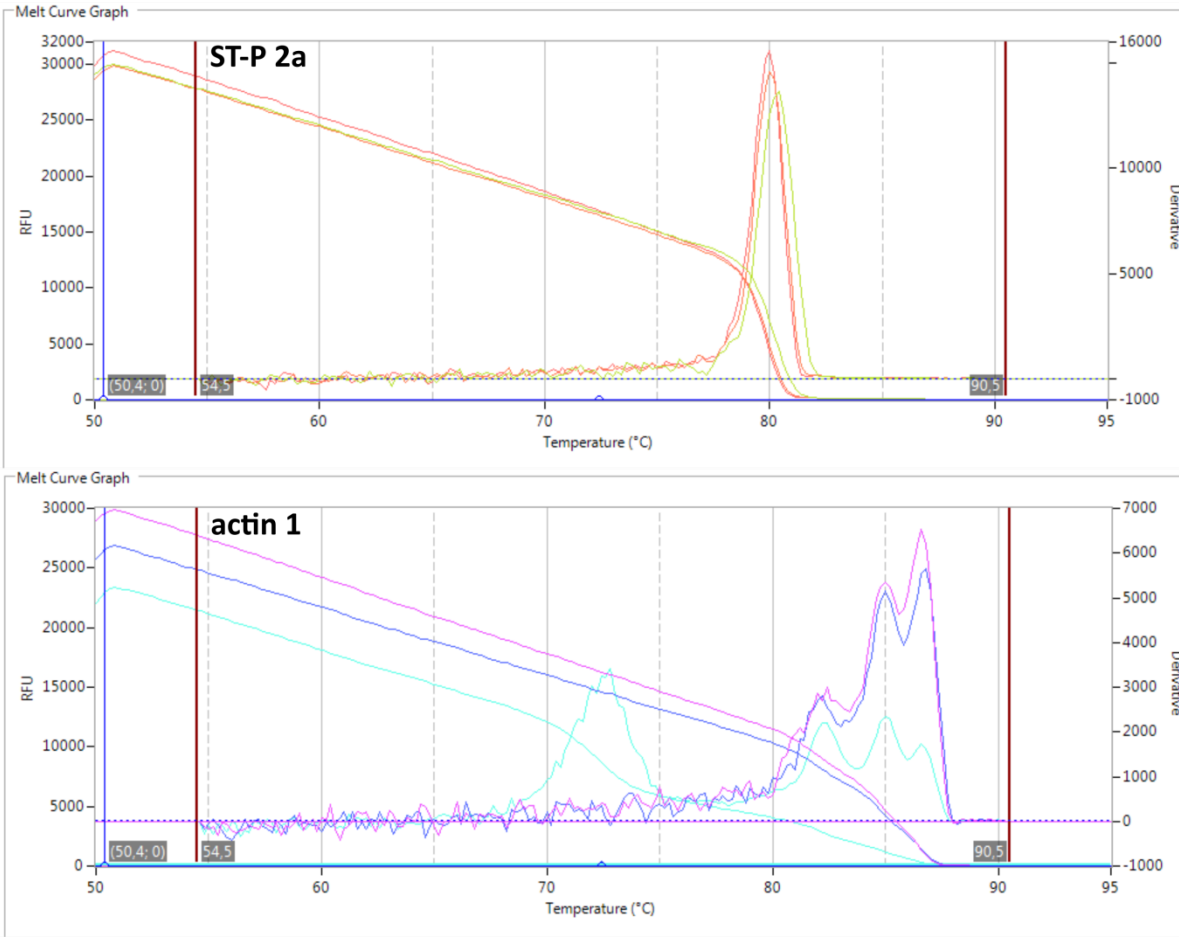
1 Melt curve data

Graphs display the melt curves of three technical replicates.








Melt Curve Graph
ST-P 2a
RFU
Derivative
Temperature (°C)
(50,4; 0)
54,5
90,5
Melt Curve Graph
actin 1
RFU
Derivative
Temperature (°C)
(50,4; 0)
54,5
90,5

2 Mean Cq values

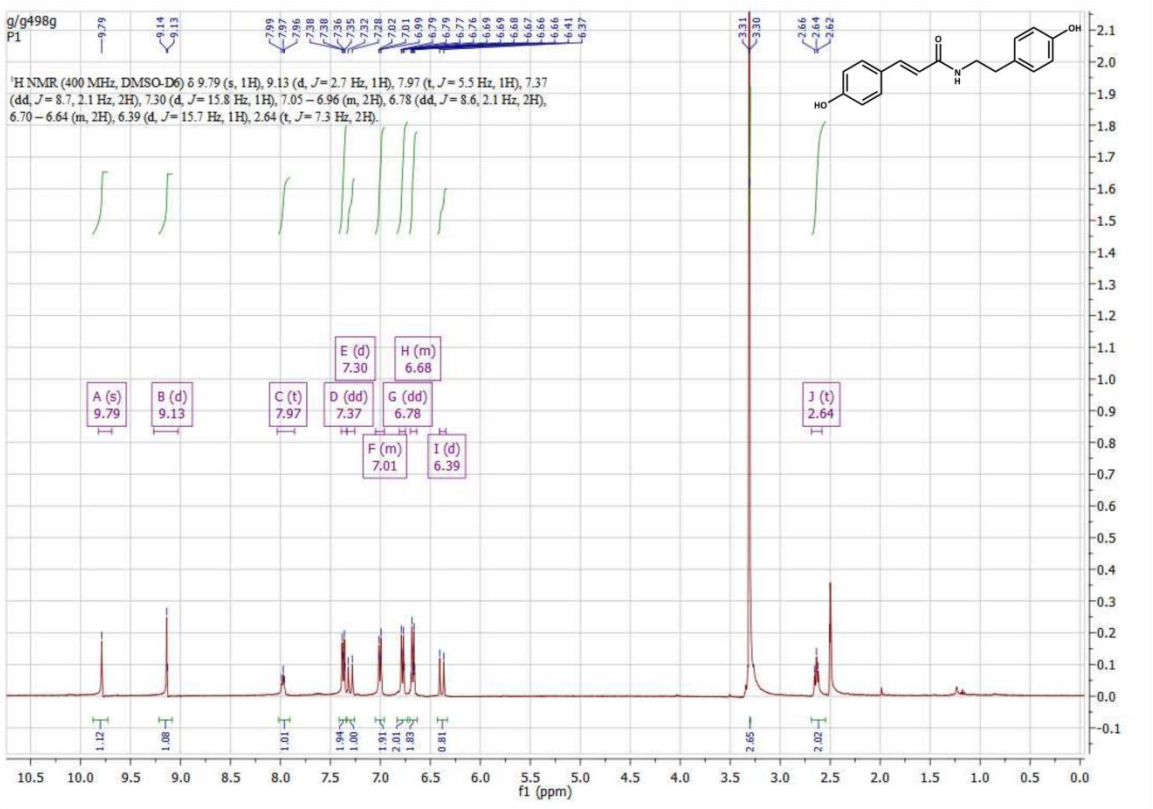
Each value represents the mean of two technical replicates.

St-P 2a											
day	**0**	**2**	**4**	**6**	**8**	**10**	**12**	**14**	**16**	**19**	**21**
RNA_1_1.1	26.14	25.37	24.57	23.99	24.49	23.47	26.05	24.43	23.12	23.25	23.48
RNA_1_1.2	26.37	24.99	24.34	23.47	24.35	23.57	25.98	24.09	22.98	23.55	23.70
RNA_1_1.3	26.36	25.58	24.59	24.28	25.31	23.55	26.34	24.45	23.41	23.86	23.44
RNA_2_1	26.83	25.29	25.16	24.89	24.39	23.48	25.00	25.21	25.11	22.59	24.55
RNA_2_2	26.01	26.07	25.75	24.59	24.81	23.32	27.38	25.28	24.05	23.56	
RNA_2_3	25.70	25.80	24.80	25.03	24.40	23.92	26.24	25.16	24.04	23.34	23.86
RNA_1_2.1	27.47	25.67	24.53	24.53	24.85	22.34	24.74	24.41	23.83	22.79	24.05
RNA_1_2.2	27.14	25.91	24.69	24.88	24.42	23.62	25.71	24.56	24.75	23.22	24.69
AaC4H											
day	**0**	**2**	**4**	**6**	**8**	**10**	**12**	**14**	**16**	**19**	**21**
RNA_1_1.1	25.31	22.96	21.13	20.60	22.53	23.21	23.56	24.00	22.01	20.50	21.14
RNA_1_1.2	25.54	22.77	21.26	20.89	22.60	23.68	23.25	23.75	21.17	20.60	21.22
RNA_1_1.3	25.77	22.82	21.35	20.91	22.87	24.02	23.81	23.59	21.45	21.02	20.68
RNA_2_1	24.72	22.74	22.69	20.98	21.41	21.55		25.29	23.16	21.13	23.25
RNA_2_2	23.30	22.85	22.25	22.12	21.72	20.92	22.93	21.07	20.56	21.02	21.78
RNA_2_3	24.14	23.18	21.70	22.02	21.36	22.39	23.09	23.02	23.12	20.77	21.44
RNA_1_2.1	26.26	23.04	20.95	22.58	21.72	21.08	21.14	24.57	23.47	20.25	22.40
RNA_1_2.2	26.32	23.33	21.89	22.52	22.93	21.19	21.96	25.49	24.09	21.15	23.46
PpC4H_1											
day	**0**	**2**	**4**	**6**	**8**	**10**	**12**	**14**	**16**	**19**	**21**
RNA_1_1.1	30.94	30.33	29.86	29.05	29.67	29.57	32.88	31.45	29.22	29.69	29.83
RNA_1_1.2	32.04	30.65	30.14	29.60	30.44	30.06	31.72	31.47	29.95	30.82	30.35
RNA_1_1.3	32.45	31.14	30.04	30.10	30.44	30.13	31.26	31.09	29.51	30.15	29.54
RNA_2_1	32.88	32.60	32.24	30.38	31.03	30.11	29.38	30.99	30.69	30.22	30.44
RNA_2_2	31.61	29.47	31.42	31.38	30.12	28.91	33.15	32.56	31.20	31.25	31.15
RNA_2_3	31.53	31.75	31.01	30.74	29.40	28.93	31.62	32.61	31.17	31.53	30.48
RNA_1_2.1		31.46	31.25	29.51	30.26	27.28	28.63	32.20	31.41	31.16	30.59
RNA_1_2.2		30.86	29.13	29.85	29.30	27.27	29.45	32.89	31.49	30.79	30.32
PpC4H_2											
day	**0**	**2**	**4**	**6**	**8**	**10**	**12**	**14**	**16**	**19**	**21**
RNA_1_1.1	31.22	29.35	27.25	26.09	27.62	27.12	31.59	29.69	28.17	28.32	29.30
RNA_1_1.2	30.66	28.28	26.84	26.14	27.40	26.81	30.13	30.29	27.36	28.23	28.45
RNA_1_1.3	31.94	29.22	27.74	26.80	28.10	27.71	31.17	29.58	27.93	28.33	29.30
RNA_2_1	34.85	29.62	29.56	28.02	28.39	25.62	27.05	30.10	31.17	29.23	30.06
RNA_2_2	31.45	29.75	29.11	28.21	27.26	26.96		31.37	29.74	29.68	30.66
RNA_2_3	30.43	29.86	28.70	28.63	27.23	26.48	31.63	31.65	30.00	29.80	30.42
RNA_1_2.1	33.46	29.58	27.84	27.60	26.36	24.98	27.13	30.51	30.57	28.06	29.65
RNA_1_2.2	34.69	30.50	29.13	28.18	28.20	26.10	28.38	30.13	32.41	28.99	32.99

3 NMR data

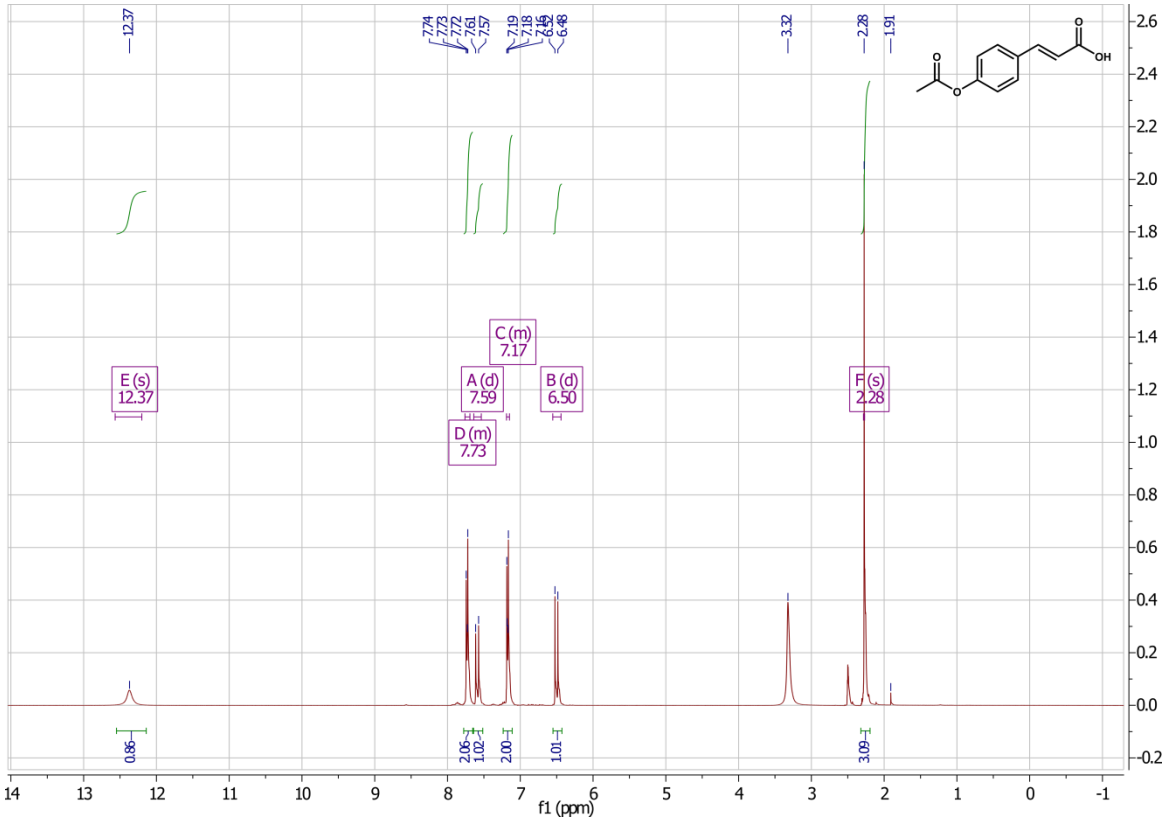
NMR data for pC-Ty:

^{1}H-NMR: (400 MHz, DMSO-d_6); δ [ppm] = 9.79 (s, 1H, OH), 9.13 (s, 1H, OH), 7.97 (t, J = 5.5 Hz, 1H, NH), 7.37 (dd, J = 8.7, 2.1 Hz, 2H, H_{Ar}), 7.30 (d, J = 15.8 Hz, 1H, H_{Olefin}), 7.05 – 6.96 (m, 2H, H_{Ar}), 6.78 (dd, J = 8.6, 2.1 Hz, 2H, H_{Ar}), 6.70 – 6.64 (m, 2H, H_{Ar}), 6.39 (d, J = 15.7 Hz, 1H, H_{Olefin}), 3.31 (s, 2H, CH_2), 2.64 (t, J = 7.3 Hz, 2H, $C_{Ar}CH_2$).



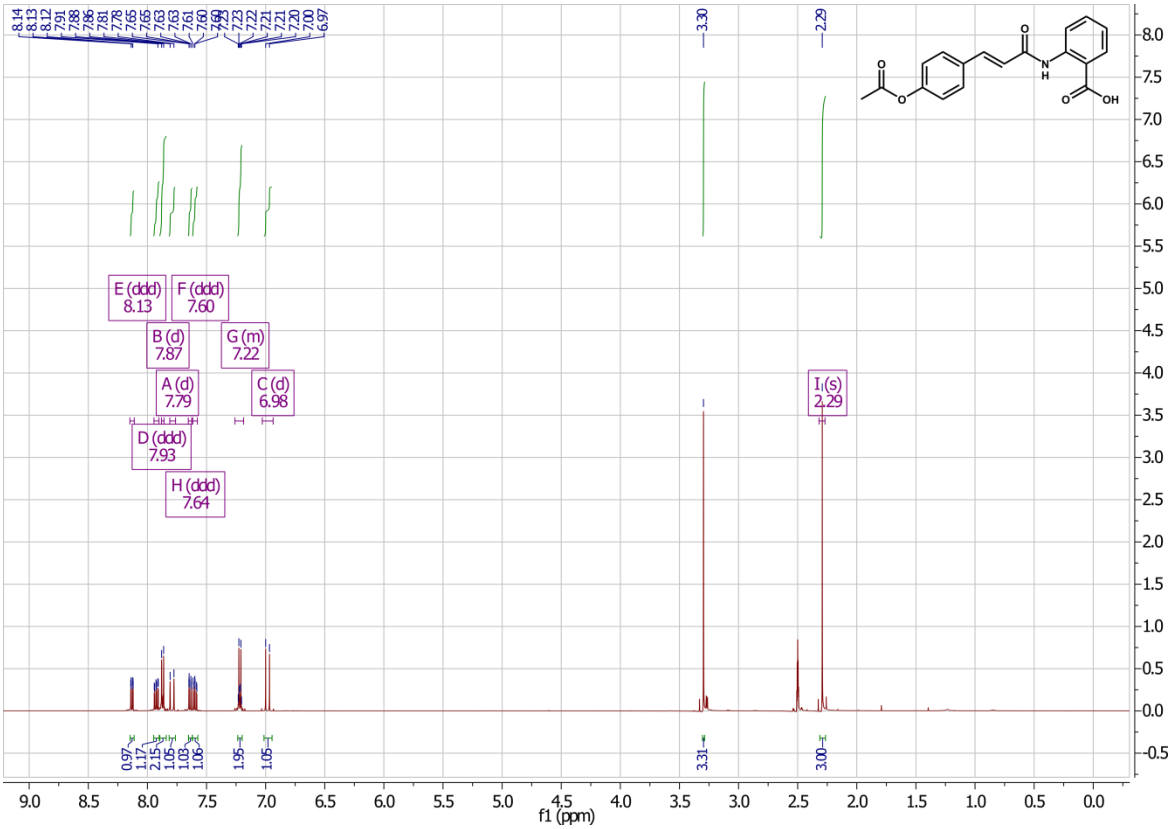
NMR data for (*E*)-3-(4-acetoxyphenyl) acrylic acid:

^{1}H-NMR: (400 MHz, DMSO-d_6); δ [ppm] = 12.37 (s, 1H, COO*H*), 7.76 – 7.69 (m, 2H, H_{Ar}), 7.59 (d, J = 16.0 Hz, 1H, H_{Olefin}), 7.19 – 7.15 (m, 2H, H_{Ar}), 6.50 (d, J = 16.0 Hz, 1H, H_{Olefin}), 2.28 (s, 3H, CH_3).



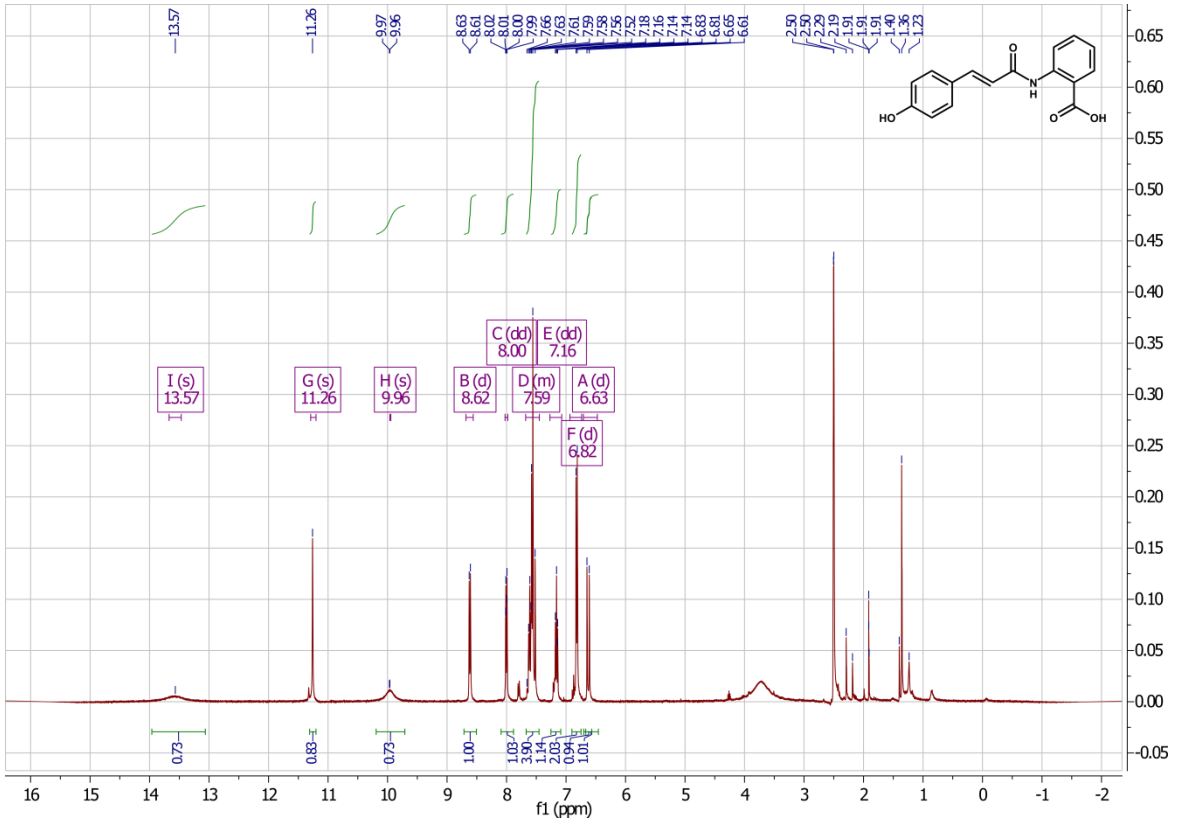
NMR data for 4-(acetyloxy)-cinnamoylanthranilic acid:

^{1}H-NMR: (500 MHz, DMSO-d_6); δ [ppm] = 8.13 (ddd, J = 7.8, 1.5, 0.5 Hz, 1H, H_{Ar}), 7.93 (ddd, J = 8.1, 7.3, 1.5 Hz, 1H, H_{Ar}), 7.87 (d, J = 8.5 Hz, 2H, H_{Ar}), 7.79 (d, J = 16.2 Hz, 1H, H_{Olefin}), 7.64 (ddd, J = 8.1, 1.1, 0.5 Hz, 1H, H_{Ar}), 7.60 (ddd, J = 7.8, 7.3, 1.1 Hz, 1H, H_{Ar}), 7.26 – 7.19 (m, 2H, H_{Ar}), 6.98 (d, J = 16.2 Hz, 1H, H_{Olefin}), 2.29 (s, 3H, CH_3).



NMR data for pC-An:

^{1}H-NMR: (400 MHz, DMSO-d_6); δ [ppm] = 13.57 (s, 1H, COOH), 11.26 (s, 1H, NH), 9.97 (s, 1H, OH), 8.62 (d, J = 8.4 Hz, 1H, H_{Ar}), 8.00 (dd, J = 7.9, 1.6 Hz, 1H, H_{Ar}), 7.58 (m, 4H, H_{Ar} und H_{Olefin}), 7.16 (dd, J = 10.7, 4.5 Hz, 1H, H_{Ar}), 6.82 (d, J = 8.5 Hz, 2H, H_{Ar}), 6.63 (d, J = 15.6 Hz, 1H, H_{Olefin}).



4 Abbreviations

2HBA	2-hydroxybenzoic acid, salicylic acid
2,4-D	2,4-dichlorophenoxyacetic acid
3HBA	3-hydroxybenzoic acid
4CL	4-coumaric acid CoA-ligase
4HBA	4-hydroxybenzoic acid
4HBCL	4-hydroxybenzoic acid CoA-ligase
aa	amino acid
Aa	*Anthoceros agrestis*
AAEs	acyl-activating enzymes
AaHCT6	hydroxycinnamoyltransferase 6 from *A. agrestis*
AcOH	acetic acid
ADP	adenosine diphosphate
AMP	adenosine monophosphate
appr.	approximately
APS	ammonium persulphate
AtBNB	basic helix-loop-helix transcription factor from *A. thaliana*

ATP	adenosine triphosphate
BA	benzoic acid
badA	benzoic acid CoA-ligase
BCIP	5-bromo-4-chloro-3-indolyl phosphate
BCL	benzoic acid CoA-ligase
bHLH	basic helix-loop-helix transcription factor
bp	base pairs
BSA	bovine serum albumin
C3H	cinnamic acid 3-hydroxylase
C4H	cinnamic acid 4-hydroxylase
CA	cinnamic acid
Caf	caffeic acid
Caf-2-Th	caffeoyl-2-threonic acid
Caf-3OH-An	caffeoyl-3-hydroxyanthranilic acid
Caf-4-Th	caffeoyl-4-threonic acid
Caf-An	caffeoylanthranilic acid
Caf-PHPL	caffeoyl-4-hydroxyphenyllactic acid
Caf-Quin	caffeoylquinic acid, chlorogenic acid
Caf-Shik	caffeoylshikimic acid
Caf-Ty	caffeoyltyramine
Cb	*Coleus blumei, Plectranthus scutellarioides*
Cb+AaCPR	chimeric CPR of *C. blumei* and *A. agrestis*
cDNA	complementary DNA
CoA	coenzyme A
coAaCYP98	codon optimized CYP98 from *A. agrestis*
coAaCYP98+CbCPR	coexpression of coAaCYP98 and CbCPR
CPR	NADPH:cytochrome P450 reductase
Cq	cycle quantification
CYP	cytochrome P450 monooxygenase
CYP98	hydroxycinnamoyl ester/amide 3-hydroxylase
DCM	dichloromethane
DIECA	sodium diethyldithiocarbamate
DHPL	3,4-dihydroxyphenyllactic acid
DMF	dimethylformamide
dNTP	desoxyribonucleoside triphosphate
DTT	dithiothreitol
ε	mammalian peptide MALLLAVF
E	amplification efficiency
EB	extraction buffer
EDC*HCl	1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride

EDTA	ethylenediaminetetraacetic acid
EIC	extracted ion chromatogram
ER	endoplasmic reticulum
ESI	electrospray ionization
EtOAc	ethyl acetate
EtOH	ethanol
F3M	flavonoid-3/3´-monooxygenase
FAD	flavin adenine dinucleotide
FLAG-tag	protein tag with the sequence DYKDDDDK
Fld	FMN-containing bacterial flavodoxin
FMN	flavin mononucleotide
FNR	FAD-containing ferredoxin-$NADP^+$ reductase
FPKM	fragments per kilobase million
fus1	AaC4H [1-524] fused with AaCPR [38-682]
fus2	AaC4H [ε:7-524] fused with AaCPR [38-682]
fus3	AaC4H [ε:39-524] fused with AaCPR [38-682]
FW	fresh weight
G6PDH	glucose-6-phosphate dehydrogenase
G-unit	coniferyl alcohol unit
gDNA	genomic DNA
Gh	*Gossypium hirsutum*
GSP	gene specific primer
H	hornwort
H-unit	4-coumaryl alcohol unit
HCl	hydrochloric acid
HCT	hydroxycinnamoyltransferase
His-tag	protein tag with the sequence HHHHHH
HPLC	high-performance liquid chromatography
HPPR	hydroxyphenylpyruvate reductase
HW	Hanes-Woolf
IAA	indole-3-acetic acid
IPTG	isopropyl-β-D-thiogalactopyranoside
k_{cat}	turnover number
K_m	kinetic value
KPi buffer	potassium phosphate buffer
L	liverwort
LB medium	lysogeny broth
LB	Lineweaver-Burk
LC-MS	liquid chromatography–mass spectrometry
LiAc	lithium acetate

M	moss
m/z	mass-to-charge ratio
MCS	multiple cloning site
MeOH	methanol
MES	2-(*N*-morpholino)ethanesulfonic acid
MM	Michaelis-Menten
MpBNB	basic helix-loop-helix transcription factor from *M. polymorpha*
mRNA	messenger RNA
MT	methyltransferase
Myc-tag	protein tag with the sequence EQKLISEEDL
NAA	1-naphthaleneacetic acid
NADH	nicotinamide adenine dinucleotide (reduced)
NADPH	nicotinamide adenine dinucleotide phosphate (reduced)
NBT	nitro-blue tetrazolium chloride
NEt_3	triethylamine
NTA	nitrilotriacetic acid
NMR	nuclear magnetic resonance
OD_{600}	optical density at 600 nm
ODS	octadecylsilane
OPC-8:0	3-oxo-2(2´-[Z]-pentenyl)cyclopentane-1-octanoic acid
P450	cytochrome P450 monooxygenase
PAGE	polyacrylamide gel electrophoresis
PAL	L-phenylalanine ammonia-lyase
pC	4-coumaric acid
pC-2-Th	4-coumaroyl-2-threonic acid
pC-4-Th	4-coumaroyl-4-threonic acid
pC-3OH-An	4-coumaroyl-3-hydroxyanthranilic acid
pC-An	4-coumaroylanthranilic acid
pC-CoA	4-coumaroyl-coenzyme A
pC-PHPL	4-coumaroyl-4-hydroxyphenyllactic acid
pC-Quin	4-coumaroylquinic acid
pC-Shik	4-coumaroylshikimic acid
pC-Ty	4-coumaroyltyramine
PEG	polyethylene glycol
PCR	polymerase chain reaction
PheA	phenylalanine-activating subunit
PHPL	4-hydroxyphenyllactic acid
Pp	*Physcomitrella patens*
Pp_AaC4H	AaC4H expressed in *P. patens*
Pp_AaCPR	AaCPR expressed in *P. patens*

Pp_AaCYP98	AaCYP98 expressed in *P. patens*
Pp_AaAp626	AaAp626 expressed in *P. patens*
PpC4H_1	potential C4H from *P. patens*, Pp3c25_10190V3.1
PpC4H_2	CYP73A48, Pp3c4_21680V3.1
PRMB	protoplast regeneration medium, bottom layer
PRMT	protoplast regerneration medium, top layer
PTS1	peroxisomal signal sequence 1
PVDF	polyvinylidene difluoride
RA	rosmarinic acid
RACE	rapid amplification of cDNA ends
RAGE	rapid amplification of gDNA ends
RAS	rosmarinic acid synthase
rbs	ribosomal binding site
R_f	retention factor
rpm	revolutions per minute
rRNA	ribosomal RNA
RT	room temperature
RT qPCR	qualitative real-time polymerase chain reaction
S-unit	sinapyl alcohol unit
SCD	synthetic complete minimal defined medium with glucose
SCG	synthetic complete minimal defined medium with galactose
SD	standard deviation
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SE	standard error
SOC	super optimal broth with catabolite repression
St-P 2a	serine threonine protein phosphatase 2a
TAL	L-tyrosine ammonia-lyase
TAT	tyrosine aminotransferase
TB	Terrific Broth
TEMED	tetramethylethylenediamine
THF	tetrahydrofuran
TLC	thin-layer chromatography
TM-helix	transmembrane helix
Tris	tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
T_m	melting temperature
UPM	universal primer mix
UV	ultraviolet
V_{max}	maximum reaction velocity

VP	vascular plants
Ws	*Withania somnifera*
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

5 Nucleotides and amino acids

Nucleotides

Nucleotide	Symbol
adenine	A
cytosine	C
guanine	G
thymine	T
uracil	ura

Amino acids

Amino acid	3-Letter	1-Letter	Amino acid	3-Letter	1-Letter
alanine	Ala	A	arginine	Arg	R
asparagine	Asn	N	aspartic acid	Asp	D
cysteine	Cys	C	glutamine	Gln	Q
glutamic acid	Glu	E	glycine	Gly	G
histidine	His	H	isoleucine	Ile	I
leucine	Leu	L	lysine	Lys	K
methionine	Met	M	phenylalanine	Phe	F
proline	Pro	P	serine	Ser	S
threonine	Thr	T	tryptophan	Trp	W
tyrosine	Tyr	Y	valine	Val	V