Phenylalanine ammonia-lyases and 4-coumaric acid coenzyme A ligases in *Chara braunii*, *Marchantia polymorpha* and *Physcomitrium patens* as extant model organisms for plant terrestrialization

Dissertation zur Erlangung des

Doktorgrades der Naturwissenschaften (Dr. rer. nat.)

dem Fachbereich der Pharmazie

der Philipps-Universität Marburg

vorgelegt von

Christoph Michael Schwarze

aus Marburg

Marburg/Lahn 2024

Dem Fachbereich Pharmazie der Philipps-Universität Marburg als Dissertation

eingereicht am 25.10.2024

Erstgutachterin: Prof. Dr. Maike Petersen

Zweitgutachter: Prof. Dr. Raphael Reher

Tag der mündlichen Prüfung: 11.12.2024

Hochschulkennziffer: 1180

Eidesstattliche Erklärung

Ich versichere, dass ich meine Dissertation

Phenylalanine ammonia-lyases and 4-coumaric acid coenzyme A ligases in *Chara braunii*, *Marchantia polymorpha* and *Physcomitrium patens* as extant model organisms for plant terrestrialization

selbstständig 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 24.10.2024

Acknowledgements

The work at hand could only be accomplished thanks to the excellent supervision of Prof. Dr. Maike Petersen who was always ready to engage in conversation, both of scientific and empathetic nature. With her genuine, down-to-earth personality and the endeavour of always being accessible, she succeeded in providing a pleasant and fulfilling atmosphere. For this, I am very grateful and would like to cordially thank her and wish her all the best for the future!

I would furthermore like to thank all my colleagues for the time we have spent together, for aid and occasional tomfoolery and I wish everyone perseverance and success for the life ahead.

My heartfelt thanks also to all members of the MadLand community, especially Prof. Dr. Stefan Rensing and Prof. Dr. Jan de Vries who were immediately involved in providing materials and scientific expertise to support our project.

Finally, I would like to thank my parents and my girlfriend Tanja, who all participated eagerly in cooling off the volcano when things occasionally did not go as smoothly as planned in the laboratory.

Publications and Presentations

2024 Publication The Plant Journal

Phenylalanine ammonia-lyases and 4-coumaric acid coenzyme A ligases in *Chara braunii*, *Marchantia polymorpha* and *Physcomitrium patens* as extant model organisms for plant terrestrialization

2024 Poster presentation MadLand Annual Meeting, Göttingen

Phenylalanine ammonia-lyases and 4-coumaric acid coenzyme A ligases in *Chara braunii*, *Marchantia polymorpha* and *Physcomitrium patens* as extant model organisms for plant terrestrialization

2023 Scientific lecture Section of plant natural products of the German Botanical Society, Hartenrod

Characterization of phenylalanine ammonia-lyase and 4-coumarate CoA-ligase in early-diverged plants

2023 Scientific lecture MadLand Annual Meeting, Feldberg (Schwarzwald)

Characterization of phenylalanine ammonia-lyase and 4-coumarate CoA-ligase in early-diverged plants

2022 Scientific lecture MadLand Annual Meeting, Feldberg (Schwarzwald)

The role of phenylalanine ammonia-lyase and 4-coumarate CoA-ligase in early-diverged plants

2021 Scientific lecture MadLand Annual Meeting, Feldberg (Schwarzwald)

Characterization of phenylalanine ammonia-lyase from *Chara braunii, Physcomitrium patens* and *Marchantia polymorpha*

Table of contents

1 Introduction

1.1 The phenylpropanoid pathway

Since the beginnings of the evolution of life on earth, organisms had to enhance certain abilities in order to adapt to new environmental conditions or gain novel functions that warranted survival and enabled versatility. About 500 million years ago, aquatic plants, such as green algae were able to colonize dry land and therefore lay the foundation for an unprecedented evolutionary progression. In order to withstand the intricacies associated with their new environment, plants had to develop sophisticated methods to cope with stresses, such as UV light, herbivores and drought. For this purpose, the phenylpropanoid pathway provided a number of compounds, serving as a starting point for the synthesis of specialized metabolites that helped plants master the transition from water to land. One of the most important derivatives of the phenylpropanoid pathway for tracheophytes is lignin, which consists of multiple units of 4-hydroxyphenylpropanoid derivatives that accumulate in secondarily thickened cell walls. Lignin provides a strengthening function for the plants to grow upwards and stabilize, and it prevents herbivores from inflicting damage on the plant (Ralph et al. 2004; Vanholme et al. 2010; Yadav et al. 2020; Domozych and Bagdan 2022; Pfeifer et al. 2023).

The phenylpropanoid pathway is divided into two parts: the general and the lignin-specific phenylpropanoid pathway (Figure 1) (Labeeuw et al. 2015; Barros and Dixon 2020). The general phenylpropanoid pathway begins with the amino acid L-phenylalanine that is transformed by the enzyme phenylalanine ammonia-lyase (PAL) into *t*-cinnamic acid. PAL has catalytic properties for both L- and D-phenylalanine and L-tyrosine. These substances are derived from arogenate (Jung et al. 1986; Cho et al. 2007; Barros and Dixon 2020), which evolves at the end of the shikimate pathway. For this purpose, chorismate is turned into prephenate via Claisen rearrangement catalyzed by chorismate mutase, whereas prephenate in turn is then transaminated into arogenate via prephenate aminotransferase. Arogenate can either be precursor for L-Phe (via arogenate dehydratase) or L-Tyr (via arogenate dehydrogenase) (Cotton and Gibson 1965; Herrmann and Weaver 1999). *t*-Cinnamic acid formed from phenylalanine is then hydroxylated into 4-coumaric acid by cinnamic acid

4-hydroxylase (C4H), whereas L-Tyr can be directly transformed into 4-coumaric acid by phenylalanine/tyrosine ammonia-lyase (PTAL) (Barros and Dixon 2020). 4-Coumaric acid is subsequently activated into its CoA thioester by 4-coumarate CoA-ligase (4CL) before the lignin-specific pathway begins. The CoA thioester is reduced to the respective aldehyde via cinnamoyl-CoA reductase (CCR), which is subsequently dehydrogenated via cinnamyl alcohol dehydrogenase (CAD) to the respective cinnamyl alcohol. From here, the synthesis of the lignin monomers begins (Guo et al. 2010; Xie et al. 2018; Vanholme et al. 2019).

The biosynthesis of monolignols has been discovered in a number of plants not containing lignin, such as the rhodopyte *Calliarthron cheilosporioides* where G, H and S lignin have been identified. In this particular case, it has been postulated that convergent evolution might be the reason for the presence of lignin within the cell wall (Martone et al. 2009). Furthermore, 4-coumaric acid respectively sequences encoding 4CL have been found in the green algae *Chlorella vulgaris*, *Haematococcus pluviaris* and *Tetraselmis suecica*, the haptophytes *Diacronema lutheri* and *Emiliana huxleyi* and the diatom *Phaeodactylum tricornutum* (Labeeuw et al. 2015). Polymers that resemble lignin have also been discovered in the cell wall of the charalean alga *Nitella* (Ligrone et al. 2008). From the monolignols *p*-coumaryl alcohol (H-unit), coniferyl alcohol (G-unit) and sinapyl alcohol (S-unit) conventional lignin is formed (Xie et al. 2018; Vanholme et al. 2019). Lignin is considered the second most abundant biopolymer on earth (Boerjan et al. 2003), highlighting its importance as a key factor in the conquest of land. G lignins have been discovered in both ferns and lycophytes (Rencoret et al. 2021). The monolignols themselves have toxic properties and are glucosylated by UDP-glucose coniferyl alcohol glucosyltransferase resulting in the respective monolignol glucosides, 4-coumaryl alcohol glucoside, coniferin, and syringin (Whetten and Sederoff 1995).

The phenylpropanoid pathway is essential for a wide variety of specialized metabolites, such as flavonoids derived from chalcones via chalcone synthase and stilbenes, such as resveratrol synthesized via stilbene synthase (Liu et al. 2021). Flavonoids have antioxidant properties and reduce UV-induced stress in plants (Tavares et al. 2013), while the stilbene resveratrol from *Vitis vinifera* has shown antifungal effects against *Botrytis cinerea*, the cause of grey mould (Adrian et al. 1997).

Chalcones are precursors for anthocyanins and tannins (Dixon et al. 2002). Anthocyanins are water-soluble, relatively unstable compounds that are responsible for natural colour in plants

and also induce health-beneficial effects (Castañeda-Ovando et al. 2009). Tannins also play an important role as defence compounds against herbivores, showing antioxidant as well as prooxidant properties, the latter leading to the accumulation of semiquinone radicals and quinones. Quinones are capable of generating high amounts of reactive oxygen species that cause oxidative stress in herbivores. Some studies have also suggested tannins to be inhibitors of protein digestion (DeGabriel et al. 2009; Barbehenn and Constabel 2011).

Another substance with disease-mitigating effects derived from the phenylpropanoid pathway is salicylic acid that is constructed via decarboxylation of *t*-cinnamic acid to benzoic acid and subsequent 2-hydroxylation (Lee et al. 1995). Alternatively, salicylic acid can be synthesized via isochorismate (Lefevere et al. 2020). Salicylic acid represents an advantageous specialized compound for plants because of its beneficial effects on plant immunity against pathogen infections (Peng et al. 2021).

Furthermore, the group of coumarins is derived from the phenylpropanoid pathway. Starting from *t*-cinnamic acid via *o*-coumaric acid, coumarin is formed; if hydroxylation occurs in *p*-position, the pathway leads to further coumarin derivatives, such as scopoletin, esculetin and umbelliferone. Umbelliferone is then further prenylated to furanocoumarins like psoralen and bergapten (Matern 1991; Bourgaud et al. 2006). Furanocoumarins prominently occur in the family Apiaceae and especially members of the genus *Heracleum* can cause serious skin irritations, which represents yet another defence mechanism against predators (Jakubska-Busse et al. 2013). In pharmacotherapy, coumarin derivatives, such as phenprocoumon and warfarin, are common as oral anticoagulants (Ufer 2005).

Another important substance derived from the phenylpropanoid pathway is chlorogenic acid, an ester of caffeic and quinic acids, which is also an important antioxidant and UV-shielding substance as well as a protectant against pathogens (Niggeweg et al. 2004; Clé et al. 2008). Its positive effects on the damage inflicted by reactive oxygen species (ROS) are well known (Xu et al. 2012). Chlorogenic acid as a component of coffee has been found to reduce blood pressure, implying a possible benefit for cardiovascular disease patients (Mubarak et al. 2012).

Structurally similar to chlorogenic acid is another ester of caffeic acid called rosmarinic acid, which was first isolated from rosemary (*Salvia rosmarinus*, syn. *Rosmarinus officinalis*, Lamiaceae) in 1958 (Scarpati and Oriente 1958). Rosmarinic acid is an ester of caffeic acid that

evolves after esterification of 4-coumaric acid from 4-coumaroyl-CoA derived from L-Phe and *p*-hydroxyphenyllactic acid derived from L-Tyr (Ellis and Towers 1970; Petersen and Alfermann 1988). It is most commonly found in the families Lamiaceae (subfamily Nepetoideae) and Boraginaceae (Petersen and Simmonds 2003) and has been ascribed to 'Labiatengerbstoffe', tannin-like compounds derived from Lamiaceae (Litvinenko et al. 1975; Petersen 2013). It has distinct antimicrobial, antioxidant and anti-inflammatory properties as well as positive effects on blood glucose, liver function and immune system (Hitl et al. 2021). To date, there have been no reports of rosmarinic acid in any of the three organisms under investigation in this work, the moss *Physcomitrium patens*, the liverwort *Marchantia polymorpha* and the green alga *Chara braunii*, but it has been thoroughly examined in several members of the Anthocerotaceae, e. g., *Anthoceros agrestis* (Busch and Petersen 2021; Ernst et al. 2022). The first discovery of rosmarinic acid in hornworts has been made about 35 years ago along with two other derivatives, megacerotonic and anthocerotonic acids (Takeda et al. 1990).

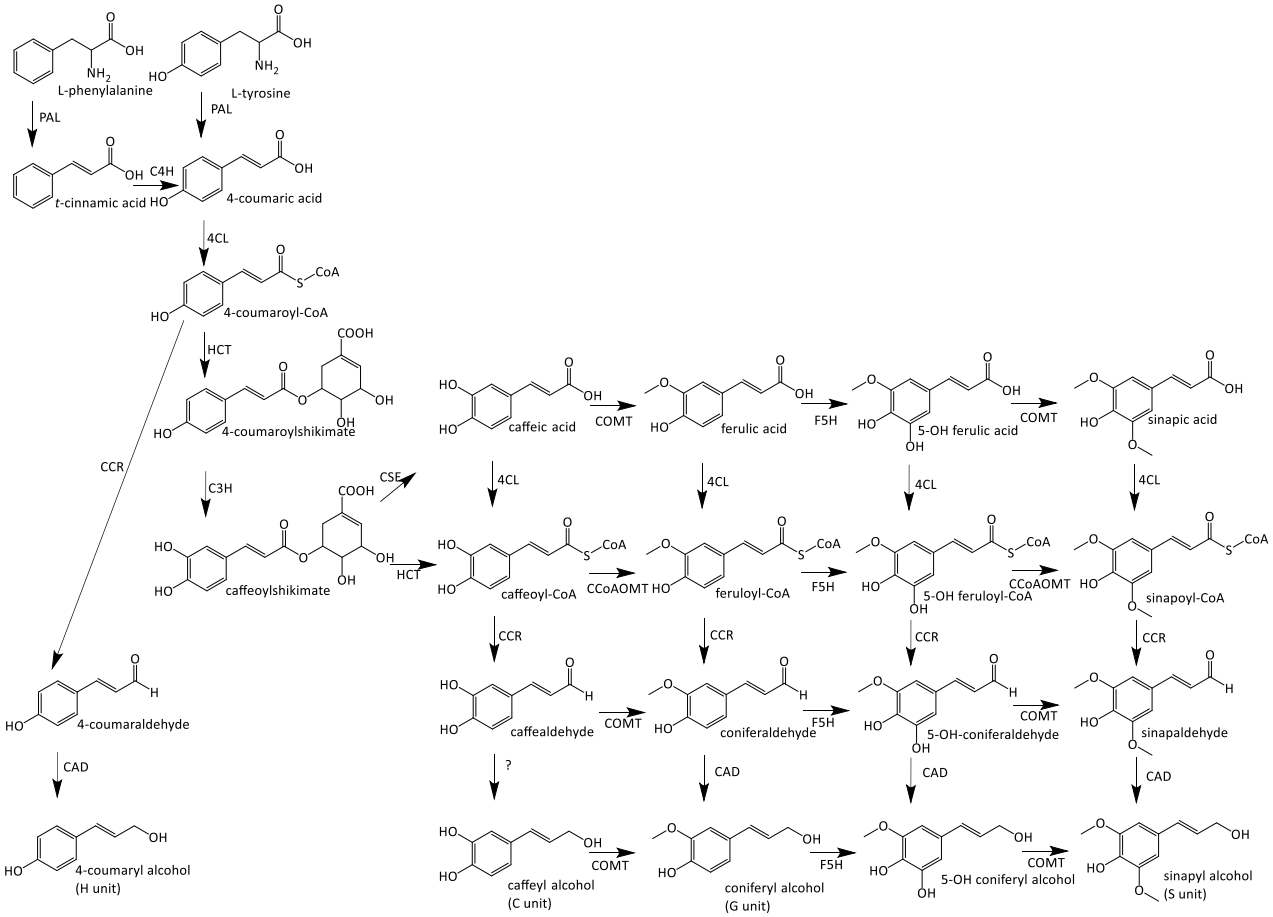


Figure 1 Lignin biosynthesis pathway in plants according to Vanholme et al. (2019).

1.2 The evolution of land plants

At a certain point in the Silurian period about 430 million years ago, due to photosynthetic activity of algae and cyanobacteria, the atmospheric oxygen level transcended 2 %, which was sufficient for facilitating terrestrial aerobic cellular respiration and gave rise to the first plants colonizing dry land (Nabors 2007).

Plants capable of producing multicellular embryos that reside in the female gametophyte, namely bryophytes (mosses) and vascular plants, are called embryophytes (land plants, also known as Embryophyta). Embryophytes are said to have evolved somewhere in the middle Cambrian/early Ordovician period and are putatively derived from a single streptophyte algal progenitor closely related to Zygnematophyceae (Nabors 2007; Wickett et al. 2014; de Vries and Archibald 2018; Morris et al. 2018).

The Streptophyta are composed of the monophyletic embryophytes and the paraphyletic streptophyte algae divided into the paraphyla KCM (Klebsormidiophyceae, Chlorokybophyceae and Mesostigmatophyceae) and ZCC (Zygnematophyceae, Coleochaetophyceae and Charophyceae). ZCC streptophyte algae together with Embryophyta combine to form the monophyletic clade Phragmoplastophyta (Fürst-Jansen et al. 2020). Streptophyta together with Chlorophyta form the clade Viridiplantae (also called Chloroplastida), the so-called “green line”. Viridiplantae, Rhodophyta and Glaucophyta represent the three branches of the Archaeplastida - for the respective phylogenetic tree, see Figure 2 (Jackson et al. 2015; Fürst-Jansen et al. 2020).

The emergence of the photosynthetically active Archaeplastida was putatively made possible by the endosymbiotic uptake of a cyanobacterium by a heterotrophic protist, which resulted in the formation of photosynthetic eukaryotes. This genesis of eukaryotes is said to have occurred approximately 1.6 billion years ago (Bengtson et al. 2017; de Vries and Gould 2018; Fürst-Jansen et al. 2020). Archaeplastida are inclined to accumulate starch as carbohydrate storage (Cenci et al. 2014).

Some members of terrestrial or freshwater algae are especially outstanding in terms of their survival strategies, i.e., the Klebsormidiophyceae being capable of withstanding extreme

temperatures (ranging from Antarctic climate to the dry heat of the Chilean Atacama Desert) and showing remarkably meticulous body plans (Bierenbroodspot et al. 2024).

As mentioned above, the clade of embryophytes consists of vascular and non-vascular plants (see also Figure 2). Vascular plants are known as Tracheophyta or tracheophytes and incorporate all plants harbouring vascular bundles. These can be further divided into seedless vascular plants (such as lycophytes (Lycopodiopsida), ferns (Polypodiopsida) and horsetails (Equisetopsida)) and spermatophytes (seed plants) (Edwards 2003; Edwards et al. 2006; Nabors 2007; Ponce De León 2024).

The clade of non-vascular embryophytes consists of the three taxonomic divisions of Anthocerotophyta (hornworts), Marchantiophyta (liverworts) and Bryophyta (mosses), each of which consisting of different species (Hedges 2002; Nabors 2007). These divisions represent the closest segregated clade to tracheophytes among the streptophytes. Cell wall analysis studies revealed a close relationship between Bryophyta and Marchantiophyta and Anthocerotophyta being in a phylogenetic branch of its own (Jackson et al. 2015; Fürst-Jansen et al. 2020; Pfeifer et al. 2022). For the exact phylogenetic relationship between the different members of the Embryophyta, see Figure 2.

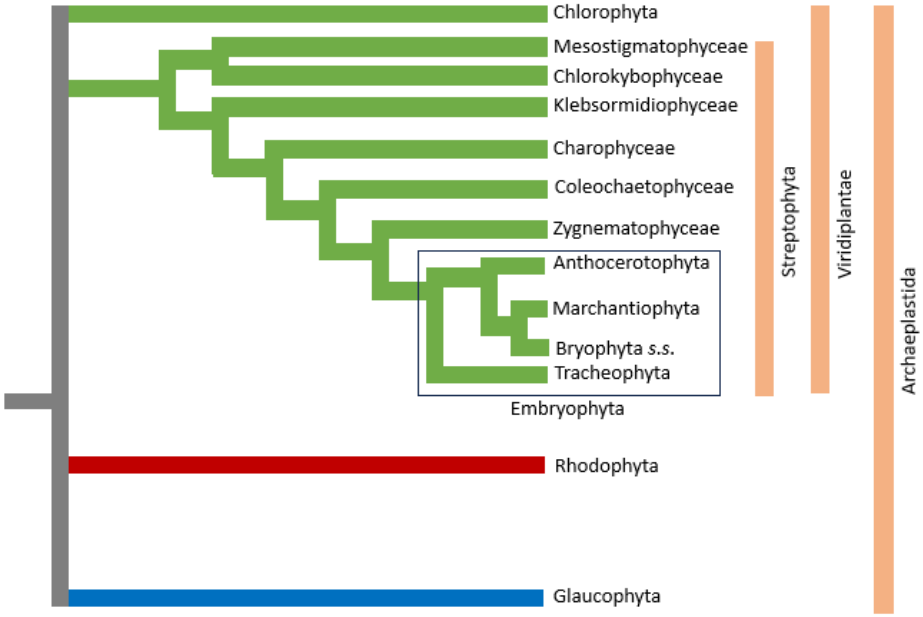


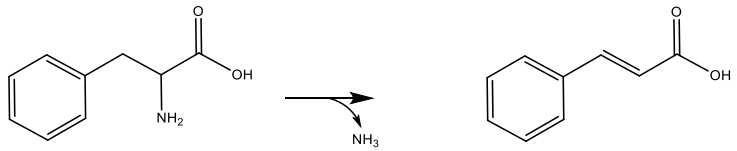
Figure 2 Phylogenetic tree incorporating the different classes of Archaeplastida with special focus on the embryophytes (Jackson et al. 2015; Fürst-Jansen et al. 2020).

1.3 Investigated enzymes

1.3.1 Phenylalanine ammonia-lyase

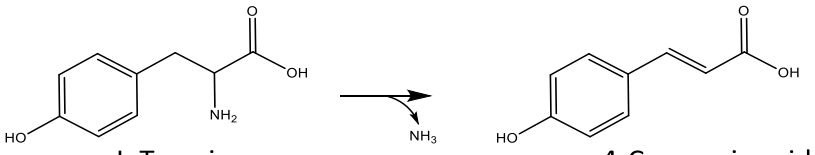
Phenylalanine ammonia-lyase (PAL) is the first enzyme in the phenylpropanoid pathway and catalyzes the reaction of L-Phe to *t*-cinnamic acid under elimination of ammonia. L-Phe is the best accepted substrate, but there is also susceptibility for D-Phe also being converted into *t*-cinnamic acid and L-Tyr, which is transformed into 4-coumaric acid. Furthermore, L-histidine can be transformed via histidine ammonia-lyase (HAL) into urocanic acid (Barros and Dixon 2020). The reaction of HAL has also been researched in this work on account of its comparability in terms of substrate size and physiological effects. Also, it is believed that HAL plays a vital role in the development of PAL and might have even been a precursor enzyme of PAL, seeing how its structures are very similar (Ritter and Schulz 2004). The respective reactions are shown in Figure 3.

a)



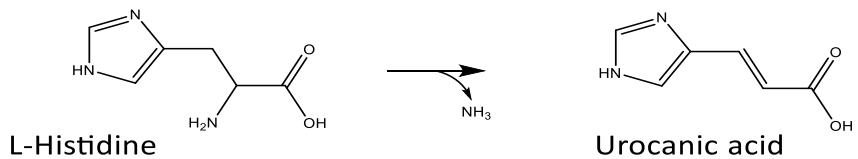
L/D-Phenylalanine *t*-Cinnamic acid

b)



L-Tyrosine 4-Coumaric acid

c)



L-Histidine Urocanic acid

Figure 3 a) Reaction of L-/D-Phe to *t*-cinnamic acid b) reaction of L-Tyr to 4-coumaric acid c) reaction of L-His to urocanic acid.

Aside from PAL putatively capable of transforming D-Phe and L-Tyr in addition to L-Phe, there exist enzymes with high selectivity towards L-Tyr, called tyrosine ammonia-lyases (TAL) and phenylalanine/tyrosine ammonia-lyases (PTAL). An active TAL with focus on transformation of L-Tyr has to date not been discovered in either fungi or plants, but in bacteria, whereas PTAL, displaying equal susceptibility ratio towards L-Tyr and L-Phe, is mainly found in Poaceae. HAL has as of yet not been discovered in plants, but appears in animals, where it catalyzes the deamination of L-His to urocanic acid, representing an absorbing agent of UV irradiation, as well as a modulator of the immune system via T-cell activation (Hart et al. 2019; Barros and Dixon 2020).

PALs, TALs, PTALs and HALs usually possess a characteristic functional group known as MIO (3,5-dihydro-methylidene-4H-imidazol-4-one). This particular structure is formed via autocatalysis from the three adjacent amino acids L-alanine, L-serine and L-glycine (ASG) at the active site of the enzyme during protein folding (Schwede et al. 1999; Baedeker and Schulz 2002; Ritter and Schulz 2004). For the respective reaction, see Figure 4. The conserved ASG-motif is found in all active PALs, as well as in TAL, PTAL, HAL and phenylalanine/tyrosine aminomutases (PAM/TAM). Aromatic ammonia lyases and PAM/TAM are therefore designated as MIO-dependent enzymes (Steele et al. 2005; Brack et al. 2024).

PAL was first described in 1961 in *Hordeum vulgare* (Koukol and Conn 1961) and has since been extensively researched in terms of its function as entry point enzyme of the phenylpropanoid pathway.

In humans, a defect of phenylalanine hydroxylase (PAH), an enzyme catalysing the tetrahydrobiopterin-dependent conversion of L-Phe to L-Tyr, can lead to the genetic disease phenylketonuria that causes serious mental disorders including typical symptoms of attention deficit hyperactivity disorder (Cannon Homaei et al. 2022) and microcephaly (Rouse et al. 2000). As a treatment method, the enzyme therapy with the medication Pegvaliase, which contains as active ingredient a pegylated PAL, can be applied to significantly reduce blood phenylalanine levels (Hydery and Coppenrath 2019).

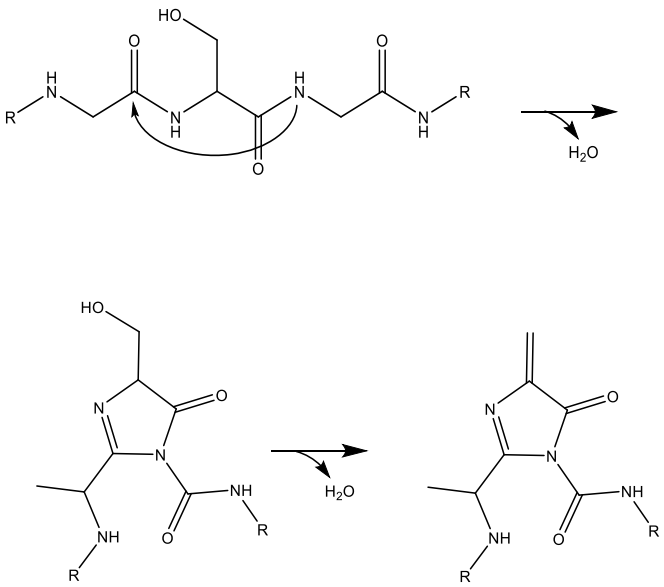


Figure 4 Autocatalytic formation of MIO from the three amino acids L-alanine, L-serine and L-glycine (ASG) (Baedeker and Schulz 2002).

There are two postulated reaction mechanisms for the interaction with MIO and the substrate during deamination (Bartsch and Bornscheuer 2009; Barros and Dixon 2020). The first is a Friedel-Crafts reaction (see Figure 5) with MIO performing an electrophilic attack on the aromatic ring of the substrate, which leads to the ring forfeiting its aromaticity and to the formation of a delocalized carbocation on the ring. The acidity of the β-proton is increased by these ongoings and can be removed with the help of a base like the hydroxy group of tyrosine. In the following, ammonia is released and by restoration of the original conformation of MIO the aromaticity of the substrate's ring is also reconstituted (Barros and Dixon 2020). The difficulty here is the abstraction of a formerly non-acidic β-proton while not deprotonating the more acidic α-ammonium group, which must later be released in the form of ammonia (Rétey 1996; Poppe and Rétey 2005).

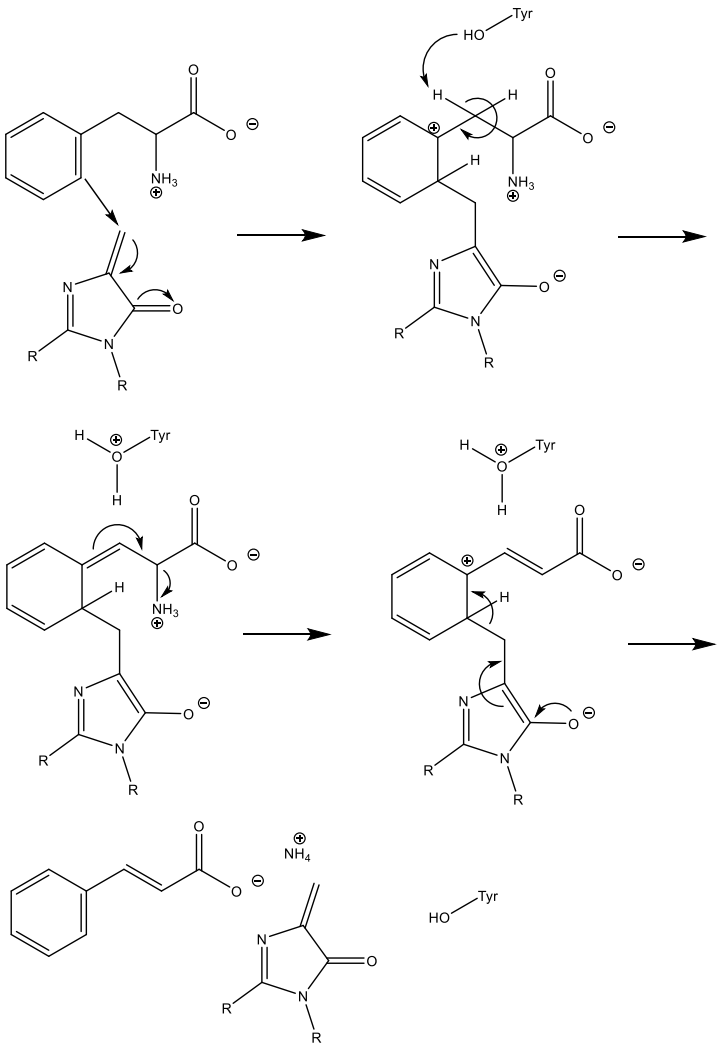


Figure 5 Proposed mechanism for the elimination of ammonia as Friedel-Crafts reaction (Barros and Dixon 2020).

The second proposed mechanism is an E_1cB elimination (see Figure 6), which is widely accepted as the most plausible of reactions for the catalysis of the deamination (Barros and Dixon 2020). In this reaction, MIO performs a nucleophilic attack on the amino group of the respective substrate, which leads to the connection of MIO to the substrate at the amino group. The free electron pair of a base like the hydroxy group of tyrosine can now remove the β-proton having experienced an increase in acidity, which leads to the formation of an unstable carbanion. The collapse of this intermediate leads to the elimination of MIO still connected to the amino group and the formation of the double bond between the α- and β-carbon atoms. The balancing of the electron arrangement of MIO now leads to the elimination of ammonia (Calabrese et al. 2004; Lovelock et al. 2014).

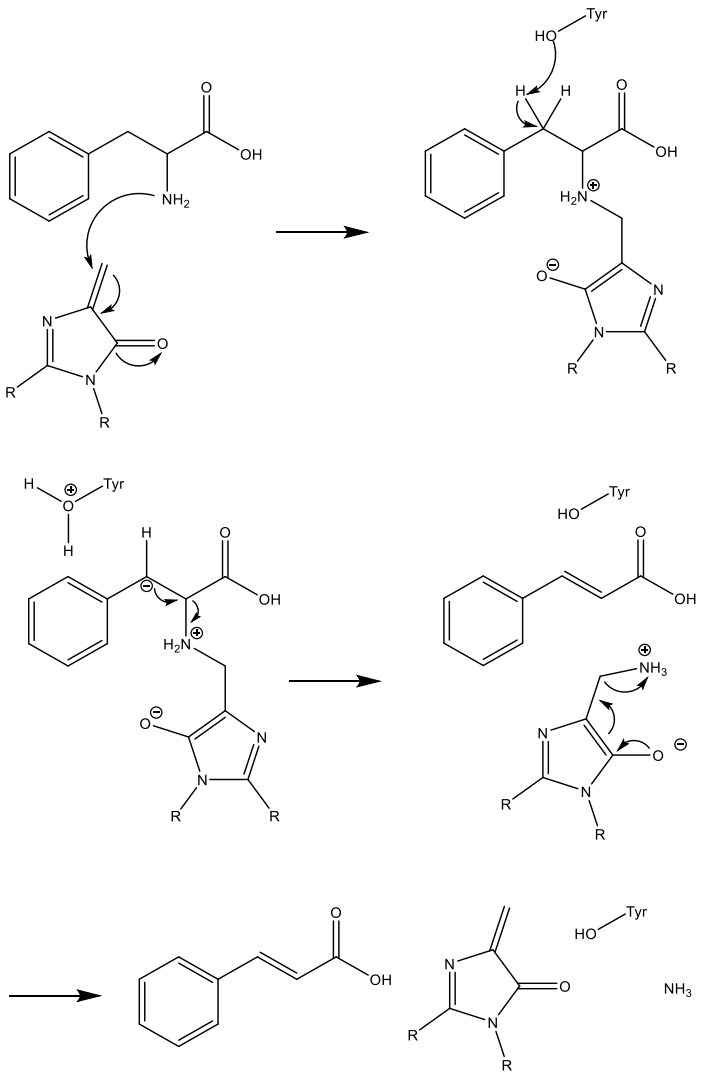


Figure 6 Proposed mechanism for the elimination of ammonia as E_1cB elimination.

Based on its structure, it is likely that PAL has derived from HAL, an enzyme with analogous mechanism of action, catalysing the reaction of L-His to urocanic acid under elimination of ammonia (Ritter and Schulz 2004). The structural similarity has been corroborated by crystallographic evaluation of PAL from *Rhodosporidium toruloides* (Calabrese et al. 2004). In general, PAL is longer in the *N*-terminus than HAL and possesses a larger internal α-helix (Ritter and Schulz 2004). HAL has to date not been described in land plants (Barros and Dixon 2020), whereas in humans it is encoded by the HAL-gene. A decrease in its activity leads to the benign metabolic disorder histidinemia (Suchi et al. 1995).

Further crystallographic experiments of PAL have been performed with isoforms from both cyanobacteria and angiosperms. The crystal structures from the cyanobacteria *Nostoc*

punctiforme [pbd-code 2NYF] and *Anabaena variabilis* [pbd-codes 2NYN, 3CZO, 5LTM] reveal major structural differences between eukaryotic and prokaryotic PALs (Moffitt et al. 2007; Wang et al. 2008; Weise et al. 2017). For eukaryotes, the crystal structure for PAL from *Sorghum bicolor* [pdb-code 6AT7] has been investigated (Jun et al. 2018), as well as from *Petroselinum crispum* [pdb-codes 1W27, 6F6T, 6HQF, 6H2O, 6RGS] (Ritter and Schulz 2004; Nagy et al. 2019; Bata et al. 2021).

Research into PAL derived from loblolly pine (*Pinus taeda*) showed similar enzymatic properties to angiosperm PAL but revealed an absence of isoenzymes in contrast to angiosperm PALs representing a possible difference in the organization and regulation of PAL activity in gymnosperms (Whetten and Sederoff 1992).

Phylogenetic analyses have suggested a possible scenario for the emergence of land plant PALs as having originated in bacteria, probably involved in the formation of antimicrobial compounds, and having been symbiotically obtained by a member of the fungal lineage from the respective bacterium via horizontal gene transfer (HGT). From the fungus, PAL was then putatively transferred to an ancestor of land plants via ancient symbiosis, creating new possibilities in terms of adaptation to novel environments (Emiliani et al. 2009). The propensity of arbuscular mycorrhizal symbiosis between plants and soil fungi is common to most extant plant clades, including extant members of bryophytes (Sgroi and Paszkowski 2020). The fossil record of plant-fungal interaction dates back approximately 407 million years to the early Devonian and has been preserved in the sedimentary deposit known as the Rhynie chert, a tropical wetland having been located on the southern margin of the palaeocontinent Laurussia. The reliability of exact characterization of fossilized material is difficult, however, because fossilization of plant tissue requires special conditions, such as rapid permineralization (Strullu-Derrien et al. 2018). Furthermore, it is often complicated to give clear insight into the origins of evolutionary novelties, taking into consideration that many plants in question have become extinct and only extant organisms are available for investigation (Jablonski and Shubin 2015; Strullu-Derrien et al. 2016).

1.3.2 4-Coumarate CoA-ligase

4-Coumarate CoA-ligase (4CL) is the third enzyme in the phenylpropanoid pathway catalysing the reaction of 4-coumaric acid to 4-coumaroyl-CoA. Further substrates are caffeic, ferulic, cinnamic, sinapic and 4-hydroxybenzoic acids that can also be transformed to their respective CoA thioesters. The respective reactions are shown in Figure 7. For the esterification, a number of co-substrates are required, namely coenzyme A (CoA), ATP and $MgCl_2$. In the process of esterification, AMP derived from ATP is coupled to the carboxylic group of the substrate and subsequently released as CoA is linked, forming the thioester. This two-step reaction is common for all members of the ANL (adenylating enzymes) superfamily (Schneider et al. 2003; Shockey et al. 2003; Gulick 2009; Shockey and Browse 2011; Wohl and Petersen 2020).

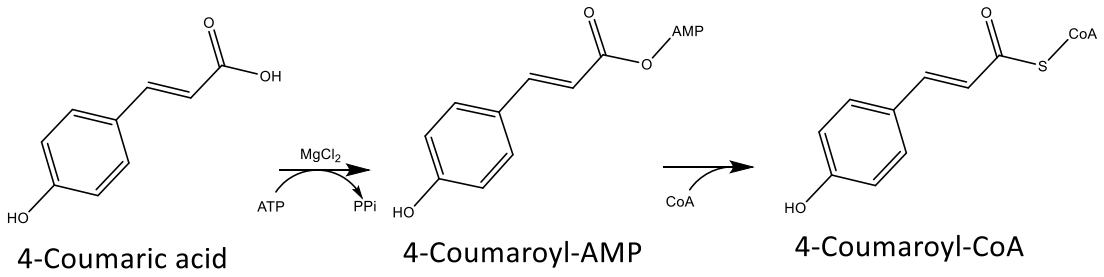
In 1970, the reaction towards a CoA thioester of *t*-cinnamic acid was first proven to be dependent on the addition of CoA (Walton and Butt 1970). A year later, cinnamyl-CoA synthetase activity has been described in extracts from the leaves of spinach beet (*Beta vulgaris*) by direct spectrophotometric assays (Walton and Butt 1971). The first report of a fungus capable of catalysing the activation of hydroxycinnamic acids to the respective CoA thioesters has been made from the basidiomycete *Polyporus hispidus* (Vance et al. 1975), while the reaction of caffeate to caffeoyl-CoA has been observed in the prokaryote *Acetobacterium woodii* (Hess et al. 2011).

4CL belongs to the family of acyl-activating enzymes that also incorporates acyl-CoA synthetases, luciferases and non-ribosomal peptide synthetases (Conti et al. 1996; Cukovic et al. 2001). While not especially congruent in terms of sequence identity, these enzymes share a conserved 12-amino acid motif (box I) that represents the binding domain of AMP derived from ATP during acyl activation (Cukovic et al. 2001; Shockey et al. 2003; Shockey and Browse 2011). 4CLs possess an additional conserved motif (box II) designating them as such: the amino acid sequence GEICIRG (Stuible and Kombrink 2001). Additionally, substrate discrimination is postulated to be determined by a specific array of amino acid residues that are located between the AMP-binding domain (box I) and the GEICIRG motif (box II). Especially the large sinapic acid is sterically difficult in terms of positioning within the active site of the protein, being blocked by two amino acid residues (most likely Val-355 and Leu-356) that interfere with the methoxy groups of this particular substrate (Schneider et al. 2003). Crystallographic experiments have shown that there is also a significant level of discrimination for ferulic acid

in between the different isoforms of 4CL derived from *Arabidopsis thaliana* that is also supposed to be tied to the region between box I and box II (Ehlting et al. 2001). Hu et al. (2010) have crystallized 4CL from *Populus tomentosa* in different varieties, as unmodified (apo) form, as well as complexed with AMP and with the intermediate adenosine 5'-(3-(4-hydroxyphenyl)propyl)phosphate (APP), respectively [pdb-codes 3A9U, 3A9V]. Their experiments have shown that 4CL consists of two distinctive globular domains and furthermore revealed the position of both the substrate binding pocket and the AMP binding pocket to be located in a cavity on the significantly larger *N*-terminus [*N*-terminal domain], while the shorter *C*-terminus [*C*-terminal domain] harbours the vital amino acid residues necessary for reactivity. In the case of Pt4CL 1, residue Lys-523 is crucial for adenylation, while residues Lys-438 and Gln-443 are responsible for the creation of the thioester via formation of hydrogen bonds. The substrate specificity, on the other hand, is not correlated to interactions of hydrogen bonds, but is rather determined by the size and structure of the substrate binding pocket (Hu et al. 2010). Further crystallographic experiments have been conducted with 4CL from *Nicotiana tabacum* [pdb-codes 5BSM, 5BSR, 5BST, 5BSU, 5BSV, 5BSW], revealing the four distinct binding pockets harbouring hydroxycinnamate, AMP, pyrophosphate and CoA, respectively. The investigations into the different crystal structures of Nt4CL 2 have shown that no distinct change occurs in the formation of the hydroxycinnamate- and AMP-binding pockets between the formation of adenylate- and thioester-group, thus putatively ruling out correlation between substrate binding and domain alternation (Li and Nair 2015).

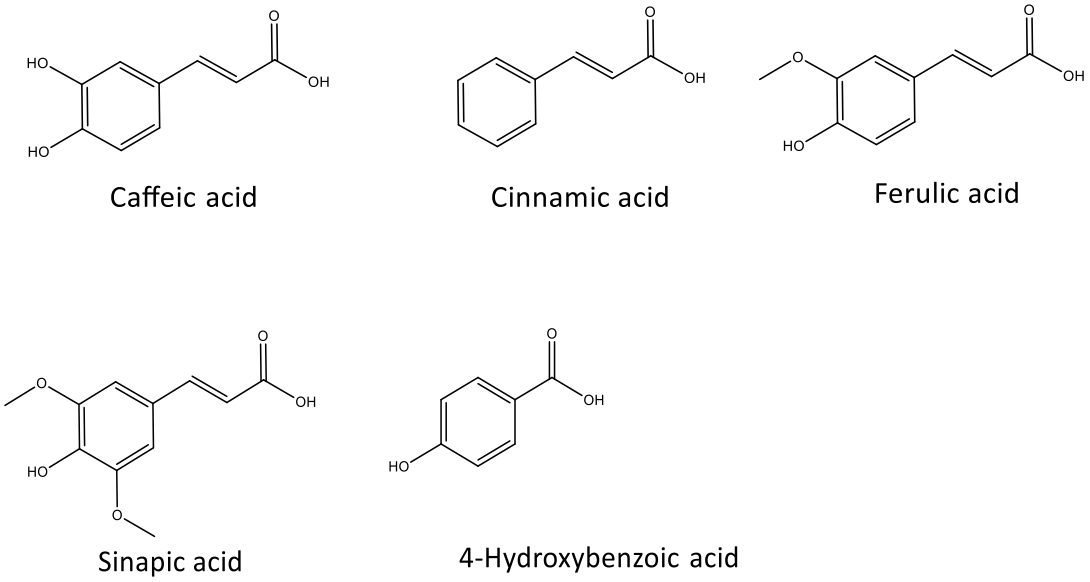
4CL experiments have been performed with four isoforms from *P. patens* that were cloned, heterologously expressed and partially characterized by Silber et al. (2008). Furthermore, a hornwort 4CL derived from *Anthoceros agrestis* has been characterized as well as a similar enzyme susceptible for 4-hydroxybenzoic acid (Wohl and Petersen 2020). For members of the liverworts, one isoform from *Plagiochasma appendiculatum* has been isolated (Gao et al. 2015) as well as four isoforms from *Marchantia paleacea* (Gao et al. 2024). Whole or partial sequences containing members of the 4CL/ACOS5-like gene family have also been found in *Klebsormidium nitens* (Klebsormidiophyceae) as well as in *Zygnema circumcarinatum* and *Spirogyra pratensis* (both members of the Zygnematophyceae) (de Vries et al. 2021).

a)



4-Coumaric acid 4-Coumaroyl-AMP 4-Coumaroyl-CoA

b)



Caffeic acid Cinnamic acid Ferulic acid

Sinapic acid 4-Hydroxybenzoic acid

Figure 7 a) Reaction of 4-coumaric acid to 4-coumaroyl-CoA b)-f) further substrates of 4CL.

1.4 Organisms under investigation

1.4.1 *Chara braunii*

Chara braunii is a multicellular green alga from the family Characeae within the paraphyletic group of the charophytes (see Figure 8). Members of the Characeae appear ubiquitously on earth in rivers and lakes with the exception of Antarctica (Beilby 2019). Until recently, it had been believed that charophyte freshwater algae are the closest living relatives of land plants (Sato et al. 2014) until Zygnematophyceae were revealed to be most closely connected to the embryophytes (de Vries and Archibald 2018).

The unicellular flagellate *Mesostigma* had previously been identified via phylogenetic analyses as the earliest branch of the charophyte lineage (Karol et al. 2001). Both groups collectively form the group of the streptophytes (see Figure 2). Together with Coleochaetophyceae and Zygnematophyceae as well as the embryophytes, Charophyceae represent the clade Phragmoplastophyta characterized by the presence of the phragmoplast, a precursor of the cell wall plate, that is made up of microtubules (Nishiyama et al. 2018).

C. braunii is inclined to grow in shallow waters, such as ponds, lakes and flooded field areas (Holzhausen et al. 2022). The monoecious alga possesses complex thalli with dendritic branching. It grows from a terminal apical cell and is made up of a central stem with branchlets regularly growing forth from axial nodes (Nabors 2007; Heß et al. 2023). *C. braunii* features vegetative cells encapsulating oogonia and antheridia (Nabors 2007). Sexual reproduction is carried out by male (antheridia) and female (oogonia) gametangia, both of which are derived from short shoots from the nodal cells of branchlets (Holzhausen 2024). When the antheridium matures, it releases spermatozoids that swim to the oogonium in order to fuse with ova, which then produce oospores. Following meiosis, the oospores germinate, growing haploid cells and forming protonemata. This eventually leads to the formation of thalli (Sato et al. 2014).

It has been shown that environmental conditions, such as light, temperature and salinity have a significant influence on the process of fertilization for *C. braunii*. It has been proposed that, when exposed to equally high temperatures during day and night instead of lower temperature at night, the oogonia remain in an unfertilized state (Holzhausen 2024).

The genome of *C. braunii* has been presented in 2018 by Nishiyama et al. (2018) who have also postulated a putative loss of morphological complexity in members of the Zygnematophyceae that evolved later, yet are structurally less sophisticated, while Charophyceae possess more advanced body plans (Nishiyama et al. 2018). *C. braunii* has also been found to be more closely related to *Nitella* species than to *Chara* species (Schubert et al. 2018).

Because of the developmental connection between Charales and evolutionarily younger plants, investigations with *Chara* species have been conducted considering the putative occurrence of rosmarinic acid, however, the search did not yield any positive results (Petersen 2013).

Figure 8 *Chara braunii* contained in a glass with its culture medium.

1.4.2 *Marchantia polymorpha*

Marchantia polymorpha is a liverwort from the family Marchantiaceae within the Marchantiales (see Figure 9) that is widely distributed and has already been mentioned by the ancient Greeks as herbal remedy. The term "liverwort" is derived from the common belief based on the doctrine of signatures that the liver-shaped thallus of *Marchantia* indicates some sort of healing property for liver diseases (Shimamura 2016). The first fossils found among the embryophytes were fragments of fossilized liverworts dating back around 450 million years (Nabors 2007). *Marchantia polymorpha* resides in cool, moist places like forests and is rather large compared to other members of the Marchantiales. As an interesting characteristic of *M. polymorpha*, its gametophores possess the shape of little trees, heaving antheridia and archegonia approximately 1 cm above the rest of the thallus. *M. polymorpha* is dioecious, possessing both male and female gametophytes. The top of the antheridiophore is flat, possessing small cavities, in which the antheridia harbouring the sperm cells are located. When water drops on top of its surface, it carries the sperm cells to the archegoniophore that has the archegonia located at its undersurface, where fertilization can then occur. The zygote forms within the archegonium and eventually breeds the sporophyte, undergoing meiosis and releasing hundreds to thousands of spores (Nabors 2007; Kato et al. 2015).

M. polymorpha has become a significant model organism and its genome has been analysed in 2017, revealing a low genetic redundancy in most regulatory pathways. Repetitive DNA

elements represent 22 % of the autosomal genome, which is significantly lower than the percentage of repetitive DNA elements in *Physcomitrium patens*, which is 48 % (Rensing et al. 2008; Bowman et al. 2017; Naramoto et al. 2022). In the hornwort *A. agrestis*, however, repetitive elements constitute only 6.98 %, which is even lower than for *M. polymorpha* (Szövényi 2016).

M. polymorpha contains several specialized metabolites, such as sesquiterpenoids and the macrocyclic bisbibenzyl ether marchantin A, that have been verified in the thalli (Tanaka et al. 2016). Marchantin A is said to possess antiprotozoic activities (Jensen et al. 2012) and experiments with the same substance derived from *Marchantia emarginata* have shown apoptosis-inducing effects in human breast cancer cells (Huang et al. 2010).

a) **b)**

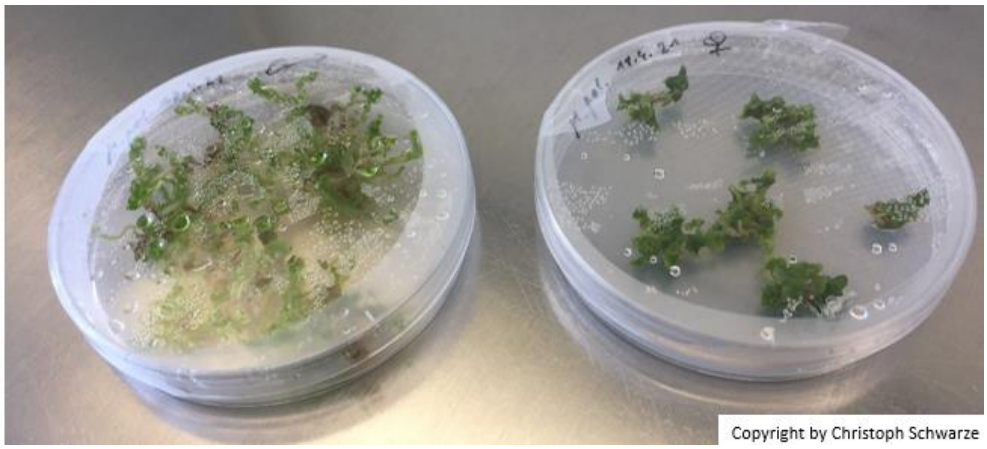


Figure 9 Sterile cultures of *Marchantia polymorpha* on B5 medium. a) male gametophytes b) female gametophytes.

1.4.3 *Physcomitrium patens*

Physcomitrium patens (see Figure 10) is a bryophyte from the family Funariaceae within the Funariales having arisen approximately ten million years ago (Rensing et al. 2020). It is native to Europe, North America and East Asia (Medina et al. 2015). Among the main classes of bryophytes, *P. patens* is a member of the Bryopsida, the class that contains most members of the bryophytes (Nabors 2007). *P. patens* has been established as a significant model organism based on laboratory cultures derived from a single spore of a sample having been collected by the British botanist H. Whitehouse at a site in Gransden Wood (Huntingdonshire, UK) in the

year 1962. Since then, it has been an indispensable system for the study of plant morphology, phytohormone effects, nutrition, and the impact of environmental influence on land plant evolution (Rensing et al. 2020). Intentional exposure to UV-B irradiance and pathogens (the hemibiotroph fungus *Colletotrichum gloeosporioides* causing the plant disease anthracnose) has been shown to lead to an upregulation of certain genes of the phenylpropanoid pathway in *P. patens* (Wolf et al. 2010; Otero-Blanca et al. 2021; Reboledo et al. 2021).

As is the case with all land plants, mosses like *P. patens* possess a haplodiplontic life cycle, both generations, the haploid gametophyte and the diploid sporophyte, are multicellular (Rensing et al. 2020). The sexual life cycle of bryophytes is determined by an alternation of generations between a diploid sporophyte and a haploid gametophyte, the latter being the dominant part and the former being in a state of dependency towards it in terms of sustenance. The gametophytes develop sexual organs (gametangia). Male gametangia (antheridia) produce sperm cells or male gametes via mitosis, whereas female gametangia (archegonia) form the ovum or female gamete. The archegonium has the form of a bottle, with the ovum developing at the bottom. After fertilization, the sporophyte develops on the gametophyte. Fertilization is facilitated when drops of water filled with sperm cells fall on top of archegonia and get stuck so that a single sperm cell coalesces with an ovum, leading to the development of a diploid zygote. The sporophyte that grows out of the zygote within the archegonium is supplied with water and nutrients by the gametophyte until it matures and breeds the sporangium at its top, which then produces haploid spores via meiosis. Spores that are released from the sporangium fall to the ground and germinate, leading to the development of protonemata, a filamentous structure that grows into a gametophore, a structure containing gametangia of its own (Nabors 2007).

The genome of *P. patens* has been fully sequenced and was first published in 2008 (Rensing et al. 2008). Genomic studies have also led to the conclusion that *P. patens* has undergone genome duplication, which may have contributed to its versatility in terms of metabolism and the high level of conservation of many crucial genes (Rensing et al. 2007). Furthermore, *P. patens* has become a significant asset in the scope of molecular farming and the production of recombinant biopharmaceutical products since they combine advantages of both microbial systems and mammalian cell systems, being easy in the handling and safe in terms of post-translational modifications. Also, they minimize the risk of contamination by infectious agents

derived from cell lines or culture media. For this purpose, *P. patens* is used as manufacturer of recombinant proteins in bioreactors (Decker and Reski 2008; Reski et al. 2015). Its relatively simple morphology and easy handling make it an ideal model organism (Prigge and Bezanilla 2010).

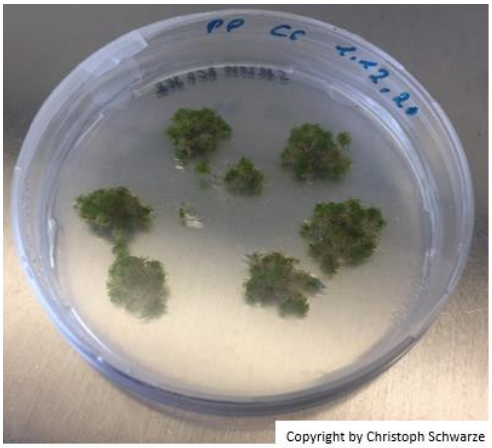


Figure 10 Sterile culture of *Physcomitrium patens* on BCD medium.

1.5 Aim of this work

C. braunii has become an increasingly researched organism since deciphering of its genome, albeit having been subject of research as long as 250 years ago (Corti 1774). Despite Charophyceae not being the most closely related group to embryophytes in phylogenetic analyses it has been considered to be the closest living relative to land plants some time ago until phylogenetic analysis revealed Zygnematophyceae to be the sister group to embryophytes (Kurtović et al. 2024). Zygnematophyceae are considered to be the family of streptophytes with the highest diversity of species (Wodniok et al. 2011). Since this project started, much research has been done on species of Zygnematophyceae shedding new light on the origins of land plants and revealing its unique cell morphology (Zhou and Schwartzenberg 2020).

The fact that *C. braunii* is therefore evolutionarily more distanced from land plants made it intriguing to determine whether the basic conditions for the transition to land were already present, and in which manner of way genes discovered in the *Chara* genome would actually come to play a role in the development of specialized compounds.

From a phylogenetic point of view, *M. polymorpha* was another ideal choice, being in the bryophyte clade together with *P. patens* and *A. agrestis*. Recent studies have highlighted that comparison of the cell wall composition of hornworts, liverworts, and mosses indicates a closer relationship between liverworts and mosses as opposed to liverworts and hornworts (Pfeifer et al. 2022). The genus *Marchantia* has not been investigated much in terms of enzymatic behaviour until very recently. In two other species, *M. emarginata* and *M. paleacea*, three genes of flavonoid *O*-glucosyltransferase have been functionally characterized (Yuan et al. 2021) and, recently, four isoforms of 4CL have been heterologously expressed and investigated in terms of their activity and substrate affinity (Gao et al. 2024). For a liverwort of a different genus, *Plagiochasma appendiculatum*, genes for both, PAL and 4CL could be retrieved. So, it was likely that *M. polymorpha* would also yield candidates for both enzymes.

Despite *P. patens* constituting an affluent and intensively researched organism and PAL being one of the best characterized enzymes, there had never been a collaborative biochemical study of PAL derived from *P. patens* to date. Therefore, investigating the very basis of the phenylpropanoid pathway with *P. patens* as model organism was a logical resolution. 4CLs derived from *P. patens* had been expressed and partially researched by Silber et al. (2008) but not all-encompassing in terms of catalytic behaviour and lacking characterization of Pp4CL 5.

Eventually, the goal was to discover putative genes for PAL and 4CL from *C. braunii, M. polymorpha* and *P. patens*, conduct heterologous expression and characterize the corresponding proteins with biochemical methods, analysing their catalytic properties. Verification of substrate transformation would help to gain further insight into the development of bryophytes as well as promote the understanding of metabolic pathways in extant members of bryophytic embryophytes. In order to categorize new findings and elucidate relationships between the investigated organisms and enzymes on the basis of their amino acid sequences and putative catalytic domains, phylogenetic analyses were to be conducted. This would also serve to support further understanding of the catalytic behaviour and substrate susceptibilities of the examined proteins and provide insight into the development of the first plants to cultivate dry land millions of years ago.

2 Materials and methods

2.1. Plant material and cultures

Sterile cultures of *Physcomitrium patens*, *Marchantia polymorpha* and *Chara braunii* have been provided by the group of Prof. Stefan Rensing (Department of Biology, Philipps-Universität Marburg*).* The cultures of *P. patens* were cultivated on solid BCD medium. *M. polymorpha* was established on solid ½ Gamborg B5 plates (Gamborg et al. 1968). Both cultures were subcultured regularly after a period of 3 months. For the cultivation of *C. braunii*, a modified culture method according to Andrews et al. (1984) and Wüstenberg et al. (2011) was deployed. Subculturing of *C. braunii* took place after a period of 2 weeks.

2.2 Culture media

2.2.1 Cultivation media for *E. coli*

2.2.1.1 Lysogeny Broth (LB)

LB medium (Bertani 1951) is used as the standard cultivation medium for *E. coli*. It is applied for overnight cultures with both the purpose of plasmid isolation and protein expression. Also, it serves as the medium for cultivation plates. In the latter case, 1 % agar have to be added. The pH is adjusted to 7.0 with NaOH. LB is composed as follows:

Composition for 1 l LB (liquid):

10 g tryptone/peptone
5 g yeast extract
10 g NaCl

Prior to autoclaving, both liquid and solid LB were divided onto flasks in volumes of 100 ml. Solid LB hardens and needs to be liquified in the microwave and adapted with 100 mg l^{-1} ampicillin prior to moulding plates. In case of the presence of a lacZ cassette on the plasmid in order to conduct blue-white screening, 50 µl l^{-1} 100 mM IPTG (filter-sterilized) and 160 µl l^{-1} X-Gal (50 mg l^{-1} in DMSO) need to be added.

2.2.1.2 Terrific Broth (TB)

For protein expression in *E. coli*, TB medium (Hobbs and Tartoff 1987) is implemented. Its advantage is its capability of providing stable conditions during the logarithmic growth phase of bacteria cells for a longer duration of time compared with other media and therefore guaranteeing a higher yield of protein. Its composition is as follows:

Composition for 1 l:

12 g tryptone/peptone
24 g yeast extract
4 ml glycerol
Adjust to 900 ml with deionized water

Prior to autoclaving, 90 ml TB medium were distributed onto Erlenmeyer flasks. Afterwards, 10 ml KPi buffer (0.17 M KH_2PO_4 and 0.72 M K_2HPO_4) were added. The buffer had also been autoclaved.

2.2.1.3 Super Optimal Broth with Catabolite repression (SOC)

SOC medium, an optimized version of SOB, supplemented by glucose (Hanahan 1983), is added to suspensions of competent cells having undergone heat-shock transformation for reconstitution. It is composed as follows:

Composition for 100 ml:

2 g tryptone/peptone
0.5 g yeast extract
50 mg NaCl

The medium was adjusted to pH 7.0 with NaOH and autoclaved before 2.5 M (end concentration) KCl (filter-sterilized), 20 mM (end concentration) glucose and 10 mM (end concentration) $MgSO_4$ were appended.

2.2.2 Media for moss cultures

2.2.2.1 BCD

BCD (Cove et al. 2009a) is the standard cultivation medium for cultures of *P. patens*. It consists of a compilation of different stock solutions and trace elements listed in Table 1.

Table 1 List of components for BCD.

Name of stock solution	Component needed for 1 l stock solution
Solution B	25 g $MgSO_4 * H_2O$
Solution C	25 g $KH_2PO_4 * H_2O$ pH 6.5
Solution D	101 g KNO_3
Hoaglands trace elements (TES)	614 mg H_3BO_3
	55 mg $Al_2(SO_4)_3 * K_2SO_4 * 24\ H_2O$
	55 mg $CoCl_2 * 4\ H_2O$
	55 mg $CuSO_4 * 5\ H_2O$
	55 mg $ZnSO_4 * 7\ H_2O$
	28 mg KBr
	28 mg LiCl
	28 mg KI
	28 mg $SnCl_2$
	389 mg $MnCl_2 * H_2O$

1 l of BCD is composed as follows:

10 ml Solution B
10 ml Solution C
10 ml Solution D
1 ml TES
219 mg $CaCl_2$
12.8 mg $FeSO_4 * 7\ H_2O$
17.5 mg Na_2-EDTA

All components were mixed and dissolved in water (final volume 1 l). 50 ml medium each were distributed onto Erlenmeyer flasks to be autoclaved. For solid BCD, 7 g l^{-1} agar was added to the mix.

2.2.2.2 ½ B5

The standard medium for cultures of *M. polymorpha* is ½ B5. The recipe is a variation of (Gamborg et al. 1968). For the ingredients of Gamborg B5, see Table 2. The composition for 1 l ½ B5 is as follows:

1.53 g Gamborg B5 (Duchefa Biochemie)
1 g glucose
Ad 1 l water

For pouring plates for growing cultures, 100 ml medium were deployed in Erlenmeyer flasks and 1.4 g agar were added before autoclaving. The pH is adjusted to 5.5 with KOH. Plates were poured after autoclaving.

Table 2 Composition of Gamborg B5.

Micro elements	**mg/l**
$CuSO_4 * 5\ H_2O$	0.025
$CoCl_2 * 6\ H_2O$	0.025
KI	0.75
H_3BO_3	3.00
FeNaEDTA	36.70
$ZnSO_4 * 7\ H_2O$	2.00
$Na_2MoO_4 * 2\ H_2O$	0.25
$MnSO_4 * H_2O$	10.00
Macro elements	**mg/l**
$CaCl_2$	113.23
KNO_3	2500.00
NaH_2PO_4	130.44
$MgSO_4$	121.56
$(NH_4)_2SO_4$	134.00

2.2.3 Medium for *Chara braunii*

The medium used for cultivation of *C. braunii* was composed as follows:

10 ml garden compost (sieved)
0.1 ml lime
100 ml quartz sand
appr. 50 ml water

First, the compost was carefully distributed at the bottom of a glass vessel (Weck®, 500 ml), so that the ground was completely covered. Then, the lime was poured on top of the compost and the quartz sand was slowly added and about 50 ml of water were carefully dripped to the mixture until the entire mass was imbued. The lid was put on the glass and the glass containing the finished medium was autoclaved. The medium was let stand for 24 h before being autoclaved a second time. Prior to subculturing of *C. braunii*, the lid was taken off and double-autoclaved sterile water was carefully poured on top of the matrix, until the glass was about 70-80 % filled. Then the alga could be placed with the help of sterile tweezers into the

middle of the matrix, with the lower parts of the thallus reaching to the compost layer at the bottom.

2.2.4 Harvesting of plant material

Fresh plant material was harvested at the laminar flow bench using sterile tweezers and falcons, in which the parts were placed. Plant material was weighed, and between 50 and 100 mg were distributed onto sterile reaction tubes and frozen in liquid nitrogen. Prior to RNA- or gDNA-extraction, an appropriate amount of frozen plant material (between 50 and 100 mg) was ground using a sterile mortar and pestle under continuous addition of liquid nitrogen to prevent thawing. The ground material was either directly used for RNA- or gDNA-extraction or stored at -80 °C.

2.3 Methods in molecular biology

2.3.1 gDNA isolation

Extraction of gDNA was conducted according to the method of Rogers and Bendich (1985). Plant material was harvested under aseptic conditions and avoidance of skin contact and frozen in liquid nitrogen. The plant material was ground in a sterile pre-cooled mortar, steadily applying liquid nitrogen, and the samples were subsequently divided onto sterile 2 ml tubes and frozen in liquid nitrogen. Only autoclaved devices and pipette tips were used. Approximately 50 to 100 mg of plant material were gathered in each tube and 300 µl 2x CTAB buffer (2 % (w/v) cetyltrimethylammonium bromide (CTAB), 100 mM Tris/HCl pH 8, 1.4 M NaCl, 1 % (w/v) polyvinylpyrrolidone), having been pre-heated at 65 °C, were added. The samples were incubated for 10 min at 65 °C under occasional inversion. The samples were then cooled on ice for an additional 10 min and 300 µl chloroform were appended. The tubes were centrifuged for 5 min at 17000 g and 4 °C. In the meantime, new 1.5 ml tubes had been prepared with 30 µl pre-heated (65 °C) 10x CTAB buffer (10 % (w/v) CTAB, 0.7 M NaCl) and the supernatant of the chloroform extraction was added to the buffer. Afterwards, additional 300 µl chloroform were appended, and the tubes were gently mixed before being centrifuged for 5 min at 17000 g and 4 °C. Meanwhile, new 1.5 ml tubes with 250 µl CTAB precipitation buffer (1 % (w/v) CTAB, 50 mM Tris/HCl pH 8, 10 mM EDTA pH 8) were arranged and 250 µl of the aqueous supernatant pipetted into the tubes, followed by gentle inversion. The tubes were

centrifuged for another 10 min at 17000 g and 4 °C and the supernatant discarded. The pellet was resuspended in 200 µl High Salt TE buffer (10 mM Tris/HCl, 1 mM EDTA pH 8, 1 M NaCl) and 400 µl 99 % ethanol were added. After gentle inversion, the tubes were incubated at -20 °C for 15 min, allowing the gDNA to precipitate. The tubes were centrifuged for 15 min at 11000 g and 4 °C and the resulting pellet was washed twice with 70 % ethanol. The pellet was then left to dry at 37 °C, resuspended in 20 µl 0.1x TE + 100 µg ml^{-1} RNAse A and incubated for 30 min at 37 °C. After photometrically determining amount and purity of the gDNA, the samples were stored at -80 °C.

2.3.2 RNA isolation

For the extraction of total RNA from crude plant material, the method of Chomczynski and Sacchi (1987) was deployed. Plant material was harvested under aseptic conditions and avoidance of skin contact and frozen in liquid nitrogen. The samples were then ground under continuous application of liquid nitrogen using a pre-cooled mortar and pestle. All devices had been heated at 200 °C for 2 h prior to processing, and only double autoclaved pipette tips were applied. After the ground samples had been divided into quantities of 50 to 100 mg each and placed in different sterile 2 ml tubes, the material was again frozen in liquid nitrogen. 500 µl of `Solution D´ (4 M guanidinium thiocyanate solution in 25 mM citrate buffer pH 7.0 + 0.5 % lauroylsarcosine) were added to each sample and incubated on ice. Afterwards, 50 µl 2 M sodium acetate pH 4 were added and mixed. 500 µl phenol saturated with citrate-buffer pH 2 were added under the vent, and additional 100 µl ice-cold chloroform were applied. The samples were gently mixed, incubated on ice for a period of 15 min and afterwards centrifuged at 17000 g and 4 °C for 10 min. 400 µl of the aqueous upper phase were transferred into fresh sterile tubes and 400 µl ice-cold isopropanol were appended. The samples were then inverted gently and deposited for 15 min at -20 °C in order to precipitate the RNA. Afterwards, the samples were centrifuged for 10 min at 17000 g at 4 °C. The supernatant was discarded, and the pellet washed with 500 µl 70 % ethanol by centrifugation for 5 min at 11000 g and 4 °C and then a second time with 500 µl 96 % ethanol (5 min, 11000 g, 4 °C). The supernatant was once again discarded and the pellet was left to dry at 37 °C. The RNA was then dissolved in 20 µl sterile water at 50 °C. The amount of resulting RNA was determined photometrically by applying 1 µl RNA-sample and 99 µl sterile water and measuring the absorption at 260 nm and 280 nm, respectively. Substantial samples with high levels of purity ($A_{260/280}$ ~ 2.0) were stored

at -80 °C and processed further. The integrity of the RNA was ascertained via agarose gel electrophoresis. For this purpose, the chamber, in which electrophoresis was performed as well as all surrounding laboratory environment, were thoroughly cleaned in order to prevent contamination with RNases. All the used materials were sterilized, and only double autoclaved equipment was applied.

2.3.3 cDNA synthesis

cDNA for polymerase chain reaction was synthesized using the RevertAid First Strand cDNA synthesis Kit (Fermentas). 2 µg RNA were adjusted to a total volume of 11 µl with sterile H_2O in a sterile 1.5 ml reaction tube. Per tube, 1 µl oligo$(dT)_{18}$ primer, 4 µl 5x reaction buffer, 2 µl 10 mM dNTP mix, 1 µl RiboLock RNAse inhibitor and 1 µl RevertAid reverse transcriptase were added, the final volume being 20 µl. The samples were gently shaken and centrifuged for 1 min at 5000 g and room temperature before being incubated for 1 h at 42 °C. The reaction was stopped by heating the samples for 5 min at 70 °C. cDNA samples were stored at -80 °C.

2.3.4 Polymerase chain reaction

Partial and full-length sequences were amplified by standard PCR. Coding sequences to be further processed into proteins were amplified using a polymerase with high-fidelity and proofreading function in order to minimize the error rate of the resulting amplified sequences. For coding sequences to be transformed into the corresponding enzymes via protein expression, specialized primers with restriction sites were designed. Prior to PCR, the absence of introns in the coding sequences was ascertained with the help of NetGene2.
Amplification of partial DNA sequences was conducted using GoTaq polymerase (Promega) under application of the following protocol:

1 µl cDNA/gDNA (template)
0.5 µl primer 1 (10 or 100 mM)
0.5 µl primer 2 (10 or 100 mM)
5 µl GoTaq buffer (5x)
3 µl $MgCl_2$ (25 mM)
0.5 µl dNTP mix (10 mM)
0.1 µl GoTaq polymerase (5 U/µl)
14.4 µl H_2O

Full-length sequences destined to be converted into proteins were amplified applying either AccuPrime™ polymerase or KOD polymerase. Primers with restriction sites were designed at the Eurofins website and adjusted to similar annealing temperatures before being ordered for synthesis from Eurofins. All primers used in this work are listed in Table 7.

For polymerase chain reaction using KOD polymerase, the following protocol was applied:

1 µl cDNA/gDNA (template)
0.75 µl primer 1 (10 or 100 mM)
0.75 µl primer 2 (10 or 100 mM)
2.5 µl 10x KOD buffer
1.5 µl $MgSO_4$ (25 mM)
2.5 µl dNTP mix (10 mM)
0.5 µl KOD Hot Start DNA Polymerase
15.5 µl H_2O

For polymerase chain reaction using AccuPrime™ polymerase, the following protocol was applied:

1 µl cDNA/gDNA (template)
0.5 µl primer 1 (10 or 100 mM)
0.5 µl primer 2 (10 or 100 mM)
2.5 µl 10x AccuPrime™ buffer
0.1 µl AccuPrime™ polymerase
20.4 µl H_2O

The respective cycler programs for the different PCR experiments are listed in Table 6 (see 2.10 PCR programs).

2.3.5 Agarose gel electrophoresis

RNA as well as DNA samples were separated by agarose gel electrophoresis. 1 g agarose was dissolved in 100 ml 1x TAE buffer (20 mM acetic acid, 1 mM Na_2-EDTA, 40 mM Tris) by heating in a microwave. The hot solution was cooled for approximately 20 min and, being still warm and liquid, about 60 ml of it cast into the designated form under application of 1 µl of a 1 % ethidium bromide solution. The gel was left to harden for about 30 min and afterwards put into an electrophoresis chamber containing 1x TAE buffer. The samples were now pipetted into the slots of the gel, the first being the marker (GeneRuler™), of which 4 µl were deployed. The other samples were dyed with 3 µl 6x loading dye (0.03 % bromophenol blue, 0.03 % xylene

cyanol, 60 mM EDTA in 60 % glycerol). Electrophoresis took place at 130 V for approximately 35 min and the gel was examined under a blue-green LED light looking through an amber filter.

2.3.6 Isolation and purification of resulting DNA fragments

Fragments of about the correct sizes resulting either from PCR or restriction digests had to be cut out of the gel using a scalpel blade and cleaned up applying a NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel). The gel slices were dissolved in 200 µl binding buffer NT for every 100 mg gel at 65 °C. The liquefied samples were then pipetted onto NucleoSpin silica filter tubes and centrifuged for 2 min at 11000 g and the flow-through discarded. The filter was washed applying 600 µl NT3 wash buffer and again centrifuging for 2 min at 11000 g. After discarding the flow-through, the tubes were again centrifuged analogously in order to dry the membrane and 30 µl H_2O was added and incubated for 5 min before being centrifuged into a sterile 1.5 ml tube. PCR products (as well as products of restriction digests) were stored at -20 °C and the silica membranes were washed for further experiments.

2.3.7 Fusion PCR

For coding sequences unobtainable via conventional PCR due to one or more introns interrupting the reading frame, fusion PCR according to Bryksin and Matsumura (2010) was conducted. For a sequence with one intron, two primer pairs were designed: For the first part of the sequence, the forward primer started at the beginning of the coding sequence and the reverse primer directly before the beginning of the intron. For the second fragment, the forward primer started directly behind the intron and the reverse primer at the end of the entire sequence. The first PCR yielded the two separate fragments and a subsequent PCR linked the fragments together, whereby the intron was eliminated. For this method, it was paramount that the two primers set around the intron overlapped the coding ends (both forward and reverse) of the gene sequence so that when undergoing a subsequent PCR, the intron was excluded and the two fragments joined together. For the fusion PCR, the two fragments were used as templates and the shorter fragment was added abundantly. After successful fusion, the amplified complete coding region was further processed.

2.3.8 Ligation

2.3.8.1 Ligation into pDRIVE

Ligation into pDRIVE was conducted according to the Qiagen protocol: 2 µl of the insert were added to 1 µl pDRIVE vector and 2.5 µl 2x Ligation Master Mix in a sterile 1.5 ml tube. The tube was then put into the refrigerator overnight, so that ligation could happen under conditions of about 8 °C.

2.3.8.2 Ligation into expression plasmid

For ligation into restriction sites, the expression vector pRSET C (Invitrogen) was applied. Beforehand, a restriction digest under application of the respective restriction enzymes had been performed, and the resulting fragments were cut from the gel (using a scalpel blade) and purified before being introduced to the plasmid. The vector itself was prepared with the same restriction enzymes so that sticky ends were created (see 2.3.15).

Ligation into pRSET C was conducted according to the following scheme:

2 µl DNA fragment
1 µl plasmid
1 µl T4 ligase buffer (10x)
1 µl T4 DNA ligase
5 µl H_2O

Ligation samples were left standing in the refrigerator overnight for ligation to happen at 8 °C and further processed the following day.

2.3.9 Production of LB agar plates for transformation purposes

Two types of agar plates were applicable: for clones transformed with pRSET C, 100 ml solid LB were liquefied by heating in a microwave and 500 µl 20 mg ml^{-1} ampicillin were added before the mixture was cast on Petri dishes to solidify. For this, about half of the bottom of each dish was covered with LB solution and the mixture was carefully spread until the entire plate bottom was covered with a thin layer of LB medium and no bubbles or contaminations were visible. Afterwards, the plates were let stand under the laminar flow cabinet in order to harden and cool. Clones transformed with pDRIVE underwent blue-white-screening in order to determine, which clone incorporated DNA fragments. For this, 160 µl X-Gal (80 µg ml^{-1}) and

50 µl IPTG (12 µg ml^{-1}) were additionally added to the LB + ampicillin mixture to cast agar plates. Blue-white screening is a method to identify successful incorporation of an insert into a cloning plasmid. This particular cloning plasmid carries at its multiple cloning site the gene for β-galactosidase (*lacZ*), which is deactivated when a transgene is inserted into the multiple cloning site. β-Galactosidase is a glycosidase that cleaves X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside), resulting in galactose and 5-bromo-4-chloro-3-hydroxyindole. The latter product spontaneously dimerizes and oxidizes to 5,5'-dibromo-4,4'-dichloro-indigo, an insoluble blue substance. Blue colonies can definitively be excluded as candidates for successful transformation because their plasmid still carries an active *lacZ* gene and is therefore still capable of cleaving X-Gal. White colonies have either lost this ability because of successful insertion of the target gene or because of insertion of different DNA or no DNA at all, which can also occur from time to time. Because of this, restriction digest has to be performed eventually, in order to select the correct clones (Ullmann et al. 1967; Koenen et al. 1982; Sambrook and Russell 2001).

2.3.10 Production of chemically competent *E. coli* cells

Essentially, there are two types of *E. coli* strain that were used for plasmid multiplication and verification, as well as protein expression. *E. coli* EZ were used for multiplication and verification purposes, while SoluBL21 were used for protein expression. For multiplication purposes, *E. coli* cells from a glycerol stock of the respective strain were picked with a sterile toothpick and cultivated overnight at 37 °C, 220 rpm in a glass tube containing 2 ml liquid LB medium. In the case of *E. coli* EZ, additional 25 µg tetracycline were added. After 14 to 18 h of incubation, 2 ml of the overnight culture were added to 100 ml liquid LB medium with an equivalent amount of antibiotic and incubated at 37 °C, 220 rpm until an optical density of 0.4-0.6 (OD_{600}) was reached. This measurement was conducted with the Eppendorf BioPhotometer at 600 nm. The cells were then transferred into 50 ml tubes and centrifuged at 3000 g, 4 °C for 10 min and the supernatant discarded. The cell pellets were resuspended in 10 ml cold 100 mM $CaCl_2$, combined and centrifuged for 12 min at 2500 g, 4 °C and the supernatant was discarded. The pellet was once again resuspended in 2 ml cold 100 mM $CaCl_2$ and incubated at 4 °C for 20 min. The suspension was centrifuged at 2500 g, 4 °C for 12 min, the supernatant discarded and the cells resuspended in 2 ml 100 mM $CaCl_2$ supplemented with

15 % glycerol. The bacteria suspension was aliquoted into sterile 1.5 ml reaction tubes in fractions of 150 µl each and frozen in liquid nitrogen to be stored at -80 °C.

2.3.11 Transformation of *E. coli*

5 µl ligation mix (2.3.8.2) were heated for 15 min at 65 °C in order to stop the ligation process. Meanwhile, tubes containing 150 µl frozen competent *E. coli* stored at -80 °C were gently thawed on ice and the 5 µl ligation mix were added to 70 µl *E. coli* suspension, the entire process taking place at 0 °C. After gentle mixing, the tubes were put back on ice for 30 min. Afterwards, the tubes were incubated for 90 s at 42 °C (heat shock) and 150 µl of SOC medium were appended and the suspension incubated at 37 °C for another 30 min. The mixture was streaked onto a pre-warmed LB agar plate (LB medium containing 20 mg ml^{-1} ampicillin) and the sealed plate was incubated at 37 °C overnight.

2.3.12 Overnight cultures

The plates were examined for colonies and the most auspicious were picked with a sterile toothpick, put in a sterile glass tube containing 4 ml liquid LB medium and 20 µl 20 mg ml^{-1} ampicillin and incubated overnight (about 18 h) at 37 °C and 220 rpm. The following day, cultures containing the correct insert could be used to establish glycerol stocks for long-term use by mixing 425 µl suspension culture with 75 µl of sterile glycerol and freezing the sample in liquid nitrogen to be stored at -80 °C.

2.3.13 Plasmid isolation from *E. coli*

Plasmids can be isolated by plasmid extraction. The extraction was carried out according to the protocol of the QIAprep® Spin Miniprep Kit by Qiagen. For this, 1.5 ml cell suspension were pipetted into sterile 2 ml tubes and centrifuged for 5 min at 11000 g. The supernatant was discarded and another 1.5 ml of cell suspension were pipetted on top of the cell pellet to be centrifuged again at the same conditions. After discarding the supernatant, the pellet was resuspended in 250 µl buffer P1 (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 100 µg ml^{-1} RNase A). Then, 250 µl buffer P2 (200 mM NaOH, 1 % SDS) were added and inverted ten times. Next, 350 µl buffer N3 (4.2 M guanidinium-HCl, 0.9 M potassium acetate pH 4.8) were appended and inverted, followed by 10 min centrifugation at 11000 g. The supernatant was transferred onto QIAprep® Spin columns and centrifuged analogously. After removal of the flow-through, the

columns were washed with 500 µl buffer PB (5 M guanidinium-HCl, 30 % 2-propanol) and subsequently with 750 µl buffer PE (10 mM Tris-HCl pH 7.5, 80 % ethanol) by centrifugation at 11000 g for 1 min, both times discarding the flow-through. The columns were dried by centrifugation at 11000 g for 1 min and removal of residual flow-through. The columns were racked on fresh sterile 1.5 ml tubes, 50 µl of sterile water were pipetted on top of the silica membrane and left to incubate for 5 min. To elute the plasmid, the tubes were centrifuged at 11000 g for 1 min and the plasmid preparation stored at -20 °C.

2.3.14 Restriction digest

DNA digest via restriction endonucleases could be used for cloning purposes or as a method to ascertain the insertion of a certain DNA sequence into a plasmid. All restriction enzymes were manufactured by Fermentas/ThermoFisher and their respective reaction conditions were abstracted from the manufacturer's protocol. In case of normal digestion of a gene ligated into the pDRIVE vector (Qiagen), the following protocol was applied:

3 µl plasmid preparation
0.5 µl EcoRI
1.5 µl EcoRI buffer (10x)
10 µl sterile water

For a digest of a plasmid equipped with restriction sites with the goal of protein expression, the following protocol was implemented:

14 µl plasmid preparation
1.0 µl restriction enzyme 1
1.0 µl restriction enzyme 2
4.0 µl buffer

A digest performed with the goal of protein expression was incubated for at least 2 h before agarose gel electrophoresis and clean-up of the digestion product. Electrophoresis can be used as a method to identify the digested samples that resulted in the correct insert. For that purpose, 5 µl of the digestion product were mixed with 3 µl 6x loading dye (0.03 % bromophenol blue, 0.03 % xylene cyanol, 60 mM EDTA in 60 % glycerol) and electrophoresis was conducted at 110 V for approximately 1 h. After a fragment on the correct height was visible on the gel, the product of the plasmid isolation fitting the visible band was selected and its concentration and possible contamination determined via photometric measurement. The

sample was concentrated up to 80 to 120 ng μl^{-1} and 12 µl total were transferred into a sterile reaction tube to be sequenced commercially at Microsynth Seqlab (Göttingen, Lower Saxony), in order to verify the entire sequence of the DNA fragment in question.

As a last verification step prior to expression, a control digest was performed with the plasmid, applying the resulting product of the final plasmid isolation. The protocol for the control digest was as follows:

3 µl plasmid preparation
0.5 µl EcoRI
1.5 µl EcoRI buffer (10x)
10 µl sterile water

2.3.15 Preparation of the expression plasmid pRSET C

For expression of recombinant proteins, the pRSET C vector was applied. This plasmid possesses a T7-promotor guaranteeing a high expression rate and the T7-gene 10 sequence that provides better protein stability. It furthermore carries the *N*-terminal His-Tag (6x His), which is necessary for protein purification following expression. *E. coli* EZ were transformed with the vector and cultivated/isolated analogously to 2.3.11 to 2.3.13 before restriction digest was performed (using the same enzymes that were applied for the respective insertion).

2.3.16 Dephosphorylation with CIAP (Calf Intestinal Alkaline Phosphatase)

In case of an insert containing two identical restriction sites, the plasmid had to be dephosphorylated prior to ligation. For this purpose, the plasmid was digested with the respective restriction enzyme and processed as follows:

10 µl digested plasmid
2 µl 10x buffer for CIAP (MBI Fermentas)
7 µl H_2O
1 µl CIAP (Calf Intestinal Alkaline Phosphatase, MBI Fermentas)

The initial solution was incubated for 5 min at 37 °C in the case of sticky before inactivation occurred at 65 °C for 15 min. Samples were stored at -20 °C.

2.4 Processing of genes, protein expression and verification

2.4.1 Codon-optimization

For proteins that were not to be attained with molecular biological methods, the opportunity to synthesize an optimized gene sequence was taken. For this purpose, the nucleotide sequence was adjusted to the special requirements of the expression host *E. coli*. The reason for this deliberate change was the fact that *E. coli* possesses a different codon usage than *P. patens*, *M. polymorpha* or *C. braunii*. In order to enhance the rate of expression of recombinant proteins, the amino acid codons whose tRNAs are most common in *E. coli* were applied during the process of optimization (Chung and Lee 2012; Fu et al. 2020). Synthesis and sequence optimization were performed commercially by BioCat GmbH. Restriction sites and additional base pairs to keep the open reading frame intact could also be installed in accordance with the respective requirements. The synthesized sequences were delivered inserted into the plasmid pUC57. Prior to ordering the DNA sequences to be synthesized, the absence of introns in the coding sequences had to be ascertained implementing NetGene2.

2.4.2 Further processing of synthesized genes

The resulting products arrived in the lab in reaction tubes containing 5 µg of freeze-dried plasmid. The plasmids were resuspended in 20 to 40 µl demineralized water and stored at -20 °C. 1 µl resuspended sample was used to conduct transformation, mixing it with 70 µl competent *E. coli* and incubating on ice for 30 min followed by heat shock for 90 s. Afterwards, the process of transformation, cultivation and plasmid isolation was conducted analogously to 2.3.11 to 2.3.13.

2.4.3 Expression of recombinant proteins

After confirming the amplified gene to be the requested one, it was ligated into the expression vector under application of the T4 ligase and afterwards transferred into competent *E. coli* and (after being plated onto solid LB medium plates containing 20 mg ml^{-1} ampicillin) incubated at 37 °C overnight. For cultivation, single colonies were taken directly from the plates using a sterile toothpick and placed in a sterile glass tube containing 4 ml LB medium and 20 µl 20 mg ml^{-1} ampicillin. The bacteria were incubated overnight at 37 °C at 220 rpm. The following day, 2 ml of the bacteria culture were transferred into a flask containing 90 ml TB medium

containing 500 µl 20 mg ml^{-1} ampicillin and 10 ml 10x potassium phosphate buffer according to Lessard (2013). The rest of the bacteria culture was used to generate glycerol stocks for long-term use, stored at -80 °C. The flasks were sealed and incubated at 37 °C and 220 rpm until an optical density of 0.4 to 0.6 (OD_{600}) was attained. The culture was then supplemented with 100 µl 1 M isopropyl-β-D-galactopyranoside (IPTG), setting protein expression in motion. For this purpose, the cultures were incubated at 25 °C and 220 rpm overnight. The next day, the cultures were centrifuged at 4 °C, 11000 g for 10 min and the supernatant discarded. The cell pellet was frozen in liquid nitrogen before being carefully thawed again on ice and 4 ml per gram cell pellet resuspension buffer (0.1 M potassium phosphate buffer pH 8) were added. After gentle mixing, 50 mg lysozyme were added and incubated for 30 min on ice. Afterwards, the cells were disrupted using ultrasonication on ice (4x 30 s, 100 %, 0.3 cycles), followed by centrifugation at 4 °C, 11000 g for 10 min. The supernatants of every sample were combined and the cell residue discarded. The supernatant could be stored as crude protein extract at -80 °C.

2.4.4 Purification of recombinant proteins by metal chelate chromatography

Prior to clean-up, a dispensable column was equipped with 1 ml ROTI®Garose-His/Ni Beads, discarding the flow-through, and leaving a Ni-NTA resin on the matrix of the column. The resin was equilibrated with binding buffer (50 mM potassium phosphate buffer pH 8, 300 mM NaCl, 10 mM imidazole). The crude protein extract, pre-adjusted to 10 mM imidazole and 300 mM NaCl, was applied to the column (which was tightly sealed) before being incubated on ice on a rocking platform for 1 h. The flow-through was discarded, and the column was washed five times with 2 ml wash buffer (50 mM potassium phosphate buffer pH 8, 300 mM NaCl, 20 mM imidazole) before the protein was eluted three times by applying 1 ml each elution buffer (50 mM potassium phosphate buffer pH 8, 300 mM NaCl, 250 mM imidazole).

2.4.5 Desalting via PD-10-columns

The elution fractions were combined and 2.5 ml of the elution were applied to PD-10 columns (pre-equilibrated with water and the respective elution buffer) in order to remove salt and imidazole according to the manufacturer's instructions. After removal of the flow-through, 3.5 ml buffer (the respective buffer ultimately used for enzyme characterization) were used to elute the protein. The protein fractions were combined, aliquoted and stored at -80 °C.

2.4.6 Determination of protein concentration

Determination of protein concentrations was conducted according to Bradford (1976). For this, Bradford reagent was produced 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 then filtered twice and stored at 4 °C. For the assay, five acrylic cuvettes were prepared with 980 µl Bradford reagent each. A clock was then set and at exact intervals of 15 s in between pipetting the samples were given to the Bradford solution as follows: 20 µl of the respective buffer as blank sample, 20 µl of protein solution (2 samples) and 20 µl of 1 mg ml^{-1} bovine serum albumin (2 samples) as positive control. The absorbance was measured at 595 nm after 20 min of incubation and the mean of the measurements was used to determine the total protein concentration necessary for characterization.

2.4.7 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was conducted in order to separate the proteins according to their molecular weight and verify the correct size of the harvested and purified protein. The principle of the method is the different rate of migration of proteins through an applied electrical field using a gel as sieving matrix. For this purpose, samples containing 15 µl of the following fractions were prepared with 5 µl Laemmli buffer each (in order to remove the intrinsic charge of the protein): crude protein extract, flow-through, washing fractions 1 to 5, elution and PD-10 fraction. Laemmli buffer consists of 0.8 g SDS, 4 ml glycerol, 1 ml β-mercaptoethanol, 1 mg bromophenol blue, 2.4 ml 1 M Tris-HCl pH 6.8 and 2.5 ml water. The reducing agent β-mercaptoethanol is used to disrupt the tertiary structure, and SDS establishes coherent negative charge on the proteins. Due to SDS binding uniformly to the linear protein (about 1.4 g SDS per g protein), the protein's charge becomes proportional to the molecular weight. SDS-PAGE was performed according to Laemmli (1970) applying a separating and a stacking gel with the following components:

Separating gel:

1.55 ml 1.5 M Tris-HCl buffer pH 8.8
1.8 ml H_2O
2.55 ml Rotiphorese® Gel 30
250 µl SDS (10 % w/v)
10 µl TEMED
43.5 µl APS (10 % w/v)

Stacking gel:

625 µl 1.5 M Tris-HCl buffer pH 8.8
1.4 ml H_2O
375 µl Rotiphorese® Gel 30
100 µl SDS (10 % w/v)
5 µl TEMED
20 µl APS (10 % w/v)
2.5 µl bromophenol blue

The glass plates were placed in the bracket, closely stuck together but with appropriate spacers for the gel in between, and the separating gel was mixed, applying APS last. The gel was quickly poured between the plates in order to polymerize and a layer of isopropanol was pipetted on top of it to balance the border between separating and stacking gel and remove residual bubbles. After polymerization was complete, the isopropanol was poured out of the chamber and thoroughly removed using filter paper and the stacking gel was mixed (again applying APS as last component) and pipetted on top of the separating gel. Before the stacking gel was polymerized, the comb was added. After polymerization was complete, the contraption was put into the electrophoresis chamber and the chamber was filled with electrophoresis buffer (192 mM glycine, 25 mM Tris, 0.1 % SDS, pH 5.3). The comb was removed carefully, and the pockets were straightened and washed with buffer to remove gel residues. The protein samples of 15 µl each were mixed with 5 µl Laemmli buffer and incubated at 95 °C for 5 min before being centrifuged and pipetted into the gel pockets. The protein marker (5 µl Roti®-Mark TRICOLOR) was applied next to the PD-10 fraction. The electrophoresis was performed at 150 V and 80 mA for 90 min. The gel was then removed from the glass plates and the separating gel was further used for Western blot analysis.

2.4.8 Western Blot and immunolabeling

In order to detect the presence of specific proteins in complex solutions, Western Blot analysis can be conducted according to Mahmood and Yang (2012) following SDS-PAGE. Afterwards, the target protein can be bound by specific antibodies and eventually visualized with the help of colour or chemoluminescence detection. After implementation of electrophoresis and separation of proteins by size, the separating gel is further used for Western blotting by being carefully placed on top of a membrane and enclosed from both sides in a layer of filter paper. The membrane (Roti®-PVDF 2.0, 0.2 µm pore size, Roth) is incubated for 15 s in methanol and

then for 2 min in water, while the separating gel is equilibrated in transfer buffer (Towbin buffer: 25 mM Tris, 192 mM glycine, 20 % (v/v) methanol, pH 8.3) for 10 to 30 min (Towbin et al. 1979). After washing, the membrane was also equilibrated in transfer buffer, as well as the filter paper (cut to the size of 7 x 8 cm). In order not to contaminate the experiment with endogenous proteins, it was necessary to wear gloves.

The blotting chamber was prepared putting the first part of filter papers carefully on top of one another, eliminating water bubbles and inconsistencies between the layers. Then, the membrane was meticulously placed on the top layer, followed by the separating gel and the remaining part of the filter paper. The lid of the blotting chamber was put over the stack and closed tightly before the start of the blotting procedure at 100 V and 100 W for 1.5 h. The current was determined by the size of the membrane (length [cm] x width [cm] x 2 mA). The membrane was afterwards washed with TBS-T buffer (10 mM Tris-HCl, 0.9 % NaCl, 0.05 % Tween 20, pH 7.4) three times for 5 min before blocking with 5 % milk powder in TBS-T for 2 h or alternatively overnight. The membrane was again washed five times before the first antibody (mouse anti-6x His, diluted 1:10000 in TBS-T supplemented with 1 % milk powder) was added and incubated for 90 min. After removal of the first antibody, the membrane was washed eight times with TBS-T for 5 min each and the secondary antibody (anti-mouse IgG coupled with alkaline phosphatase, diluted 1:20000 in TBS-T supplemented with 1 % bovine serum albumin) was appended and incubated for 90 min. After removal of the secondary antibody, the membrane was again washed five times with TBS-T. The membrane was then placed in substrate buffer (100 mM Tris-HCl, 100 mM NaCl, 5 mM $MgCl_2$, pH 9.5) for 5 min and equilibrated before staining was performed applying a staining solution consisting of 80 µl BCIP (5-bromo-4-chloro-3-indoxyl phosphate; 20 mg ml^{-1} in 100 % dimethylformamide) and 60 µl NBT (nitro blue tetrazolium; 50 mg ml^{-1} in 70 % dimethylformamide) dissolved in 10 ml substrate buffer. Staining was conducted for 15 min and stopped by washing with water (three times). A positive result was corroborated when the membrane showed intense purple colouration due to the alkaline phosphatase coupled to the secondary antibody separating a phosphate moiety from BCIP and producing an indigo dye. The indoxyl group then reduced a tetrazolium salt (NBT) leading to formation of an insoluble formazan that coupled with the indigo dye to produce a coloured precipitate (Altman 1976).

As an alternative to the aforementioned procedure, Western Blot can be conducted according to the protocol for the Immobilon®-P transfer membrane (EMD Millipore Corporation).

Following electrophoresis, the blotting chamber was prepared as follows: first two sheets of filter paper wetted in anode buffer I (0.3 M Tris-HCl, 10 % MeOH, pH 10.4), followed by one sheet of filter paper wetted in anode buffer II (25 mM Tris-HCl, 10 % MeOH, pH 10.4). Next, the membrane was placed with the rough surface facing downwards and the gel was carefully placed on top of the membrane (before blotting, the membrane had been placed in MeOH for 15 s in order to gain hydrophobicity to avoid the binding of unwanted proteins and then in water for 2 min before having been incubated in anode buffer II). Lastly, three sheets of filter paper wetted in cathode buffer (25 mM Tris-HCl, 40 mM glycine, 10 % MeOH, pH 9.4) were placed on top of the gel and blotting was commenced using the same conditions as mentioned above (210 V, 265 mA, 100 W) for the duration of 40 min. Between first and second antibodies, only three washing steps with TBS-T were performed and the immunostaining was conducted according to (Altman 1976).

2.4.9 Trichloroacetic acid precipitation

For samples that showed a rather low protein concentration, trichloroacetic acid (TCA) precipitation could be performed in order to concentrate the amount of protein to a more detectable level. For this purpose, the sample was added to the same volume of 40 % TCA and incubated at 4 °C for 60 min in order to precipitate. The sample was then centrifuged for 20 min at 17000 g, 4 °C and washed twice with 80 % acetone. The solvent was evaporated, and the precipitate dried before resuspension in 20 µl 0.1 M Na_2CO_3 + 10 % saccharose. The resuspended concentrated sample could then be applied onto a SDS gel.

2.5 Biochemical characterization of enzymes

2.5.1 Photometric determination of enzyme activities

2.5.1.1 Reactivity tests PAL

Prior to the actual characterization, a preliminary test was performed to verify the enzyme's activity. For this purpose, 2, 5 and 10 % (of assay volume in the cuvette) crude extract was applied and measured photometrically, adding buffer (0.5 M Tris-HCl pH 7.5) and substrate (L-Phe for reactivity tests). A stock solution containing 10 mM L-Phe dissolved in 0.5 M Tris-HCl pH 7.5 was prepared and the concentrations of L-Phe for the test were varied between 1 and 5 mM. If the crude extract showed reactivity, the same volumes of desalted and purified protein were applied and the test repeated.

2.5.1.2 Enzyme assays PAL

Standard PAL assays contained desalted and purified protein solution (PD-10-fraction), buffer (0.5 M Tris-HCl pH 7.5) and substrate (10 mM L-/D-Phe; 10 mM L-Tyr; 10 mM L-His; 10 mM L-DOPA; each dissolved in 0.5 M Tris-HCl pH 7.5). For measurements with L-His, 5 mM DTT were added as well as 100 µM $MnCl_2$ (Rechler 1969; Klee 1972; Brand and Harper 1976). Both DTT and $MnCl_2$ were dissolved in 0.5 M Tris-HCl pH 7.5. Assays were performed at 40 °C and absorption was recorded at 290/309/277 nm for Phe/Tyr/His. Enzyme activities were calculated using extinction coefficients as follows: *t*-cinnamic acid $\varepsilon_{290\,nm}$ = 8,747 cm^2 $mmol^{-1}$, 4-coumaric acid $\varepsilon_{309\,nm}$ = 10,872 cm^2 $mmol^{-1}$, urocanic acid $\varepsilon_{277\,nm}$ = 10,349 cm^2 $mmol^{-1}$ in 0.5 M Tris-HCl buffer. The amount of protein solution necessary for conducting the experiment was determined via Bradford assay and protein masses per cuvette varied from 3 µg to 50 µg, depending on the isoform's susceptibility toward the respective substrate. Prior to the actual measurement, the samples were prepared and pre-incubated for 180 s. Because of the persistence of linearity of PAL experiments, incubation could be processed with the complete samples containing protein solution and ready to be started immediately after the 180 s. For each PAL measurement, a cuvette containing a reference was prepared. This reference consisted either of one cuvette for the entire measurement batch or one reference cuvette for each sample cuvette. References contained buffer (0.5 M Tris-HCl pH 7.5) and protein solution, but no substrate. Reference cuvettes were newly supplied for every new measurement.

2.5.1.3 Reactivity tests 4CL

For 4CL, activity tests were also performed, in this case with the two substrates 4-coumaric and caffeic acids at a concentration of 200 to 500 µM and with 2, 5 and 10 % (of assay volume in the cuvette) of crude protein extract as well as PD-10 fraction. For the exact assay composition, see 2.5.1.4.

2.5.1.4 Enzyme assays 4CL

Standard 4CL assays contained desalted and purified protein solution (PD-10-fraction), buffer (0.1 M KPi pH 7.5), substrate (phenolic acid), 2.5 mM ATP, 2.5 mM $MgCl_2$, 0.1 mM CoA and a compensational volume of equal parts methanol and water to balance the volume of the substrate solution in order to maintain a steady concentration of methanol while varying the substrate concentration. The phenolic acids were dissolved in equal parts of methanol and water. Their initial substrate concentration was 10 mM, assay concentrations varied between 0 and 500 µM. The end concentration of methanol per cuvette was 2.5 %, a higher amount could compromise the measurement. 4CL activity was photometrically determined by measuring formation of the resulting CoA thioester from the respective carboxylic acid. The measured substrates were 4-coumaric, caffeic, cinnamic, ferulic, sinapic and 4-hydroxybenzoic acids with their respective CoA thioester's extinction coefficients as follows: 4-coumaroyl-CoA $\varepsilon_{333\,nm}$ = 21 cm^2 $mmol^{-1}$, caffeoyl-CoA $\varepsilon_{346\,nm}$ = 18 cm^2 $mmol^{-1}$, cinnamoyl-CoA $\varepsilon_{311\,nm}$ = 22 cm^2 $mmol^{-1}$, feruloyl-CoA $\varepsilon_{345\,nm}$ = 19 cm^2 $mmol^{-1}$, sinapoyl-CoA $\varepsilon_{352\,nm}$ = 20 cm^2 $mmol^{-1}$, 4-hydroxybenzoyl-CoA $\varepsilon_{300\,nm}$ = 13 cm^2 $mmol^{-1}$ (Stöckigt and Zenk 1975; Zenk 1979; Biegert et al. 1993). Assays were incubated at 40 °C and the measurements were started by adding CoA to the reaction. The total volume in the cuvettes was either 1000 µl or 700 µl each. Because this type of reaction showed only very transient linearity, a quick working procedure was required. The samples were therefore incubated for 180 s before addition of CoA, quickly mixed and the measurement was started immediately afterwards. Usually, the linearity of the experiment was stable for less than 120 s, so only this area could be adduced for calculation of specific activities. For every 4CL measurement, a reference was implemented, either one reference cuvette per measured sample or one reference for one complete batch of measurement. Because of the decrease of protein activity and possible evaporation of liquid, especially with rising assay temperature, it was necessary to prepare new references after one completed measurement. References contained protein solution, ATP, $MgCl_2$, buffer (0.1 M KPi

pH 7.5), substrate (in the same concentration as the sample) and the respective compensational volume of water/methanol. Instead of CoA, references contained an equal amount of water. In order to determine the influence of ATP and CoA on 4CL measurements, referential experiments for both co-substrates were conducted with the most efficient 4CL (Mp4CL 3). ATP/$MgCl_2$ concentrations were varied between 0.625 and 12.5 mM per cuvette, CoA concentrations ranged from 15 to 700 µM (see 3.4.5.4).

2.5.2 Evaluation of measured data

2.5.2.1 Determination of kinetic parameters according to Michaelis-Menten

For every kinetic experiment the specific activities [mkat kg^{-1}] were determined and the Michaelis-Menten method was applied in order to determine K_m [µM] and V_{max} [mkat kg^{-1}], which were subsequently used to calculate the turnover number k_{cat} [s^{-1}] as well as the catalytic efficiency k_{cat}/K_m [s^{-1} mol^{-1} l]. The K_m determines the affinity of the enzyme to the substrate, while V_{max} states the maximum rate of turnover. The turnover number indicates how many molecules of a certain substrate are being processed by the enzyme in 1 second, while the catalytic efficiency determines the overall efficaciousness of the enzyme in terms of kinetics and substrate susceptibility. A high catalytic efficiency indicates an effective enzyme.

The physical principle of the photometric measurement is the Beer-Lambert law that describes the decrease in irradiation intensity and wavelength when passing through an absorbing agent. The decrease of both units is exponentially proportional to the concentration of the dissolved agent, and the distance the ray has to travel through the medium (layer thickness):

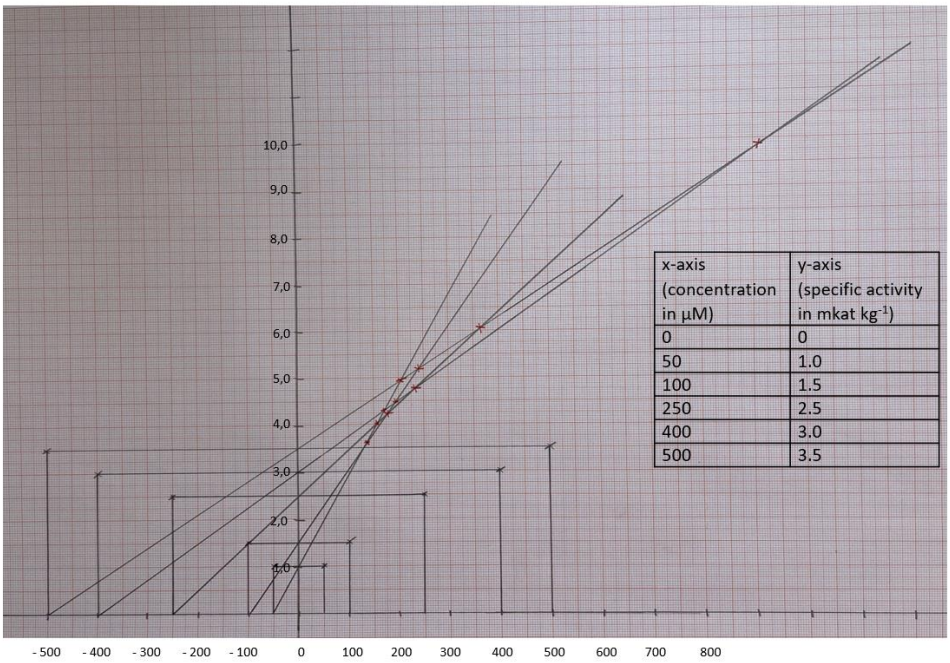
$$I = I(0)e^{-\varepsilon dc}$$

I = irradiation intensity
$I(0)$ = irradiation intensity at the start of the experiment
ε = extinction coefficient [mol^{-1} cm^{-1} l]
d = layer thickness [cm]
c = concentration [mol l^{-1}]

The quotient $I/I(0)$ is called transmission, the reciprocal $I(0)/I$ is called opacity. Usually, both values are determined in percent (Bisswanger 2015).

2.5.2.2 Application of Cornish-Bowden method

The Cornish-Bowden method (Cornish-Bowden and Eisenthal 1978) is a non-calculative method to graphically determine K_m and V_{max} for Michaelis-Menten graphs that do not reach saturation. This can happen when a substrate is only soluble up to a certain concentration, at which saturation is not yet attainable. If, however, an inclination towards saturation is visible, the Cornish-Bowden method can be deployed. This was done for several kinetics in this work on coordinate paper as follows. First, all the measuring points of the graph were mirrored over the y-axis into the second quadrant of a coordinate system and the original and the reflected point were connected with a line parallel to the x-axis. Then, a perpendicular was let fell from both points, resulting in two x-intercepts and constructing a rectangle. A line was drawn through the x-intercept from the second quadrant to the y-intercept of the connecting line between the measuring point and its mirror image. This procedure was repeated for each of the measuring points of the respective graph, resulting in a certain number of intersections of the drawn lines located in the first quadrant. All these intersections were then used to, firstly, let fall another perpendicular, each of which creating a new x-intercept and, secondly, draw a parallel to the x-axis, each resulting in a new y-intercept. The median of the x-intercepts represents the K_m, while the median of the y-intercepts represents the V_{max} to be determined. An exemplary implementation of the Cornish-Bowden method, using fictitious values for the concentration in µM (depicted on the x-axis) and the specific activity in mkat kg^{-1} (depicted on the y-axis) and conducted on coordinate paper, is shown in Figures 11 and 12.



x-axis (concentration in µM)	y-axis (specific activity in mkat kg^{-1})
0	0
50	1.0
100	1.5
250	2.5
400	3.0
500	3.5

Figure 11 Exemplary depiction of Cornish-Bowden method. The five pairs of values are shown in a table to the right of Figure 11. The intercepts in the first quadrant are depicted in red.

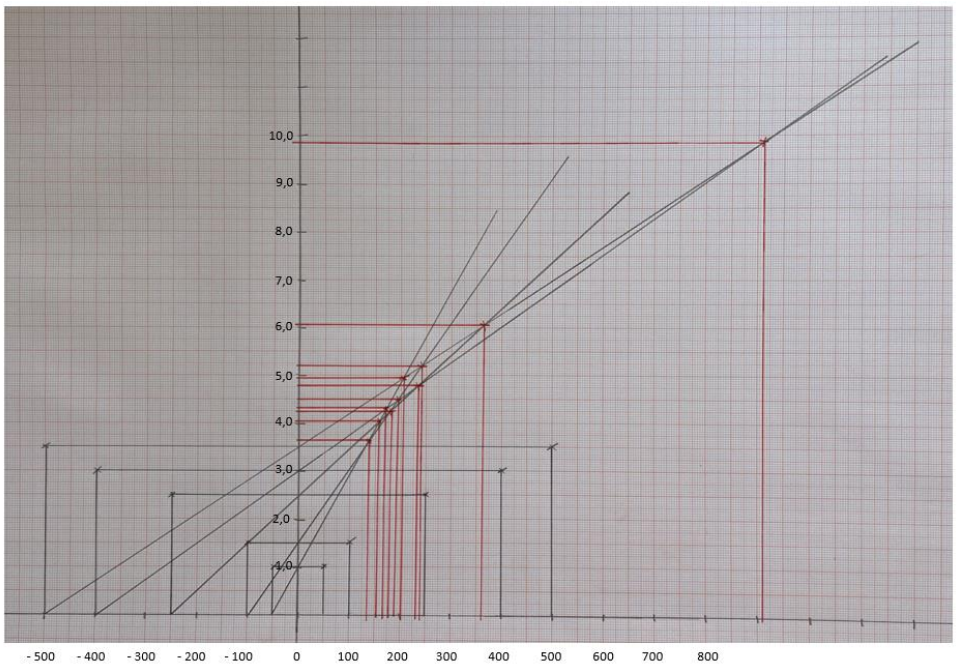


Figure 12 Continuation of Cornish-Bowden method with the same exemplary kinetic values shown in Figure 11. Perpendiculars to both axes are depicted in red. The median of the x-intercepts represents K_m, in this case the median of the x-axis intercepts of perpendiculars 5 and 6 = 195 µM. The median of the y-intercepts represents V_{max}, in this case the median of the y-axis intercepts of perpendiculars 5 and 6 = 4.65 mkat kg^{-1}.

2.5.3 Temperature optima

In order to determine optimal temperature conditions for some representative isoforms of PAL respectively 4CL, temperature optima were measured. For this, concentrations of 2 mM for PAL and 500 µM for 4CL were chosen, and a photometric measurement was conducted analogously to the procedures described under 2.5.1.2 and 2.5.1.4. For evaluation purposes, the temperature was shown on the x-axis, while the specific activity was depicted on the y-axis. For PAL measurements, L-Phe was used as substrate, whereas 4-coumaric acid was applied for experiments with 4CL.

2.6 Verification of measured data

2.6.1 Empty vector control

In order to guarantee that the vector itself did not feign any kind of activity by its own presence, an empty vector control for both PAL and 4CL was performed. For this purpose, pRSET C-harbouring *E. coli* SoluBL21 were cultivated, and protein expression was conducted. Following metal chelate chromatography, a standard PAL assay was implemented using the protein extract from empty vector controls as substitute for purified enzyme. Three subsequent measurements were applied with a concentration of 2 mM L-Phe at 40 °C. To ascertain that the vector also did not interfere with 4CL measurements, an analogous procedure was implemented for 4CL tests. Protein extract from empty vector controls was used as substitute for the enzyme using 4-coumaric, caffeic, ferulic and cinnamic acids as substrate, respectively. For each tested substrate, three measurements were performed with a concentration of 400 µM and a temperature of 40 °C.

2.6.2 Sample preparation for PAL assay verification via HPLC and LCMS

Substrate transformation from the specific reactions with PAL was corroborated via HPLC. For this, two different kinds of samples were implemented: a negative control and a sample that contained a complete enzyme assay having been incubated for 1 respectively 12 h. The negative control also incorporated all components of the assay, but it was stopped with 6 N HCl prior to addition of the substrate in order to prevent any unwanted reactions from occurring. The positive control was stopped with 6 N HCl after incubation. The relation of HCl to the complete volume of the assay was 1:4. Reaction mixtures were added to an equal

volume of ethyl acetate and mixed before being centrifuged for 20 min at 17000 g and room temperature. The top layer (ethyl acetate phase) was pipetted into a fresh reaction tube. Again, an equivalent amount of ethyl acetate was added to the aqueous phase, mixed thoroughly and again centrifuged under the same conditions. The top layer was transferred to fresh reaction tubes and evaporated. The residue was resuspended in 100 µl 20 to 45 % methanol in water with 0.01 % *o*-phosphoric acid. 50 µl were conveyed to clean vials and transferred to HPLC. The applied device was an HPLC of the type Hitachi Chromaster with isocratic elution at 30 or 45 % methanol acidified with 0.01 % *o*-phosphoric acid at a flow rate of 1 ml/min and 35 °C on a Hypersil ODS column (250 x 4 mm; pre-column: 20 x 4 mm; particle size 5 µm; Dr. Maisch GmbH). Authentic standards (10 mM *t*-cinnamic acid, 10 mM 4-coumaric acid) were used as comparative agents. Detection of the peaks took place at the following wavelengths: 278 nm (L-Phe → *t*-cinnamic acid), 310 nm (L-Tyr → 4-coumaric acid), 277 nm (L-His → urocanic acid).

2.6.3 Sample preparation for 4CL assay verification via LCMS

4CL reactivity was verified implementing LCMS analysis. Because of the difficulty of corroborating the presence of the CoA thioesters, an indirect method of verification was chosen: after the standard 4CL assay was incubated for 1 h, 5 µl 10 mM 4-hydroxyphenyllactate (pHPL) as well as 5 µl (1 µg) heterologously synthesized and purified rosmarinic acid synthase (RAS) derived from *Coleus blumei* were added to the enzyme assays for all isoforms of 4CL and incubated for 1 h. Afterwards, the subsequent reaction leading to the corresponding pHPL-ester was verified via LCMS. Four different types of samples were prepared. A positive sample containing all components of the respective 4CL assay plus 5 µl RAS and 5 µl 10 mM pHPL and three negative samples containing either inactive 4CL combined with active RAS, inactive RAS combined with active 4CL or both enzymes having been inactivated. After incubation, all samples were stopped with 6 N HCl and extracted twice with an equal amount of ethyl acetate and centrifuged for 10 min at 17000 g and 4 °C. The combined ethyl acetate phases were evaporated and the residues resuspended in 45 % methanol in water with 0.01 % *o*-phosphoric acid. Standards used for verification consisted of 25 µM 4-coumaroyl-4'-hydroxyphenyllactate respectively 25µM caffeoyl-4'-hydroxyphenyllactate.

2.6.4 Reactivity verification via liquid chromatography/mass spectrometry (LCMS)

Liquid chromatography was conducted using HPLC (Agilent 1260 series) with a Multospher 120 RP18 column (CS Chromatographie Service GmbH). The dimension was 250 x 2 mm, the particle size 5 µm. Two solvents were applied: A = 0.1 % (v/v) aqueous formic acid and B = acetonitrile with 0.1 % (v/v) formic acid. Measurements could be conducted either in the short or long method with the following gradients (Tables 3 and 4):

Table 3 Short method for reactivity verification applying HPLC (Agilent 1260 series) with a Multospher 120 RP18 column.

Time (min)	A (%)	B (%)
0:00	95.00	5.00
10:00	0.00	100.00
15:00	0.00	100.00
15:10	95.00	5.00
20:00	95.00	5.00
flow rate: 0.5 ml min^{-1}; temperature: 25 °C		

Table 4 Long method for reactivity verification applying HPLC (Agilent 1260 series) with a Multospher 120 RP18 column.

Time (min)	A (%)	B (%)
0:00	95.00	5.00
40:00	0.00	100.00
45:00	0.00	100.00
45:10	95.00	5.00
55:00	95.00	5.00
flow rate: 0.25 ml min^{-1}; temperature: 20 °C		

Detection was performed by a mass spectrometer of the type micrOTOF-Q III with ESI source (Bruker Daltonics) calibrated with 5 mM sodium formate under application of the negative mode. Calibration reached from m/z 100 to m/z 1000 in order to analyse higher masses.

2.7 Construction of phylogenetic trees

In order to issue a structured compilation of the genetic comparability of the species-specific protein isoforms of a respective enzyme, a phylogenetic analysis was performed. For this purpose, amino acid sequences of the enzymes under investigation were retrieved from databases and with the help of the program MEGA, a phylogenetic tree (Maximum Likelihood Algorithm) was constructed (MEGA 11, 1000 bootstraps). This analysis visualizes the similarity and comparability of the enzymes' amino acid sequences amongst one another. A list of all organisms from which were derived the different protein sequences for PAL and 4CL is shown in Table 5.

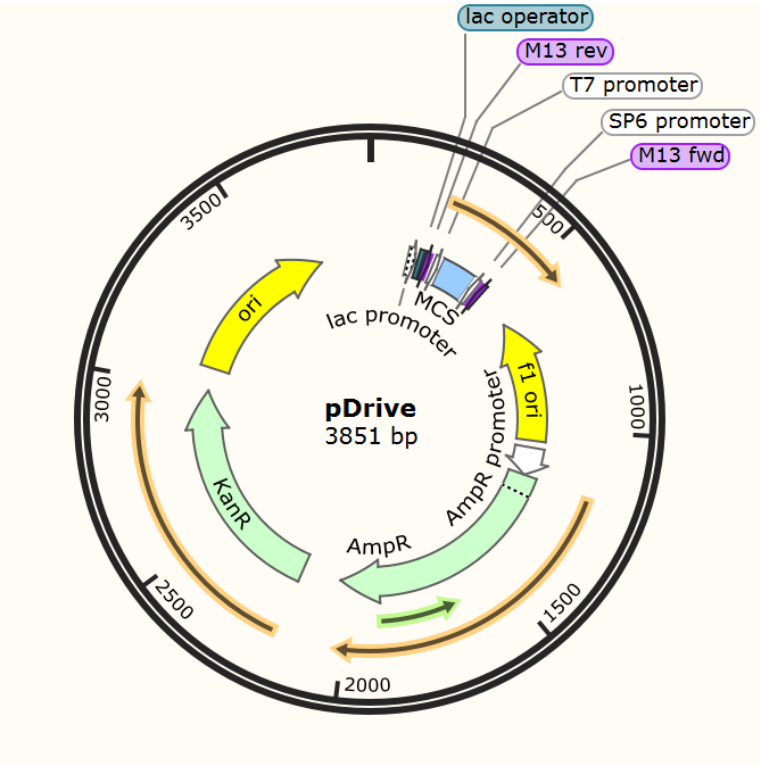
Table 5 List of species and abbreviations used for the construction of the phylogenetic trees of PAL and 4CL.

Abbreviation	**Species**
Allsat	*Allium sativum*
Antagr	*Anthoceros agrestis*
Aratha	*Arabidopsis thaliana*
Asahet	*Asarum heterotropoides*
Asasie	*Asarum sieboldii*
Betpla	*Betula platyphylla*
Betvul	*Beta vulgaris*
Bradis	*Brachypodium distachyon*
Camsin	*Camellia sinensis*
Chabra	*Chara braunii*
Citret	*Citrus reticulata*
Cofara	*Coffea arabica*
Cucsat	*Cucumis sativus*
Cunlan	*Cunninghamia lanceolata*
Daucar	*Daucus carota*
Diptri	*Diphasiastrum tristachyum*
Dryfra	*Dryopteris fragrans*
Elagui	*Elaeis guineensis*
Ephsin	*Ephedra sinica*
Equarv	*Equisetum arvensis*
Erygut	*Erythranthe guttata*
Eucglo	*Eucalyptus globulus*
Eucrob	*Eucalyptus robusta*
Ginbil	*Ginkgo biloba*
Hibcan	*Hibiscus cannabinus*
Isolac	*Isoetes lacustris*
Klenit	*Klebsormidium nitens*
Lacsat	*Lactuca sativa*

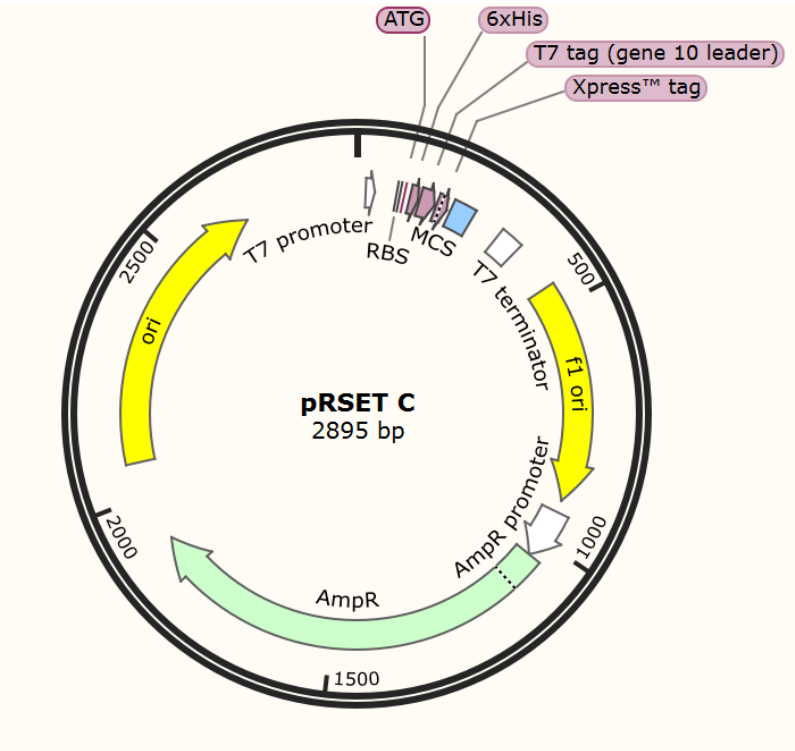
Larkae	*Larix kaempferi*
Lilhyb	*Lilium* hybrid
Litery	*Lithospermum erythrorhizon*
Marpol	*Marchantia polymorpha*
Medsat	*Medicago sativa*
Meloff	*Melissa officinalis*
Musacu	*Musa acuminata*
Nelnuc	*Nelumbo nucifera*
Nictab	*Nicotiana tabacum*
Phypat	*Physcomitrium patens*
Pintae	*Pinus taeda*
Plaapp	*Plagiochasma appendiculatum*
Poptre	*Populus tremuloides*
Poptri	*Populus trichocarpa*
Psinud	*Psilotum nudum*
Rubida	*Rubus idaeus*
Selmoe	*Selaginella moellendorffii*
Strspi	*Struthiopteris spicant*
Triaes	*Triticum aestivum*
Trisub	*Trifolium subterraneum*
Vitvin	*Vitis vinifera*
Zeamay	*Zea mays*
Zosmar	*Zostera marina*

2.8 Vector maps

pDRIVE (Qiagen):



pRSET C (Invitrogen):



2.9 List of reagents and materials

2.9.1 List of restriction enzymes and buffers for restriction digest

Restriction enzyme/buffer	**Activity**	**Supplier**
BamHI	10 U/µl	Thermo Scientific
BglII	10 U/µl	Thermo Scientific
EcoRI	10 U/µl	Thermo Scientific
10X EcoRI buffer with BSA		Thermo Scientific
HindIII	10 U/µl	Thermo Scientific
KpnI	10 U/µl	Thermo Scientific
NcoI	10 U/µl	Fermentas
PvuII	10 U/µl	Thermo Scientific
XhoI	10 U/µl	Thermo Scientific
10X Buffer R with BSA		Fermentas
10X Buffer Tango with BSA		Thermo Scientific

2.9.2 List of chemicals

Chemical	**Particulars**	**Producer**
1-butanol		Roth
1,4-dithiothreitol		Roth
3-hydroxybenzoic acid		Acros Organics
4-coumaric acid		Sigma-Aldrich, Fluka, Merck
4-hydroxybenzoic acid		Merck
DL-4-hydroxyphenyllactic acid (pHPL)		Sigma-Aldrich
5-bromo-4-chloro-3-indoxyl-β-D-galactopyranoside (X-gal)		Roth
5-bromo-4-chloro-3-indoxyl phosphate (BCIP)		Roth
acetic acid (glacial)		Roth
acetone		Roth
acetonitrile		Roth
acrylamide 30 %/bisacrylamide 0.8 % (37.5: 1)	Rotiphorese® Gel 30	Roth
activated carbon		Merck
adenosine-5'-triphosphate (ATP), sodium salt		Roth
agar		Cero
agarose (NEEO Ultra-Qualität)		Roth
amidoblack		Roth
ammonium persulfate (APS)		Sigma-Aldrich
ammonium sulphate		Roth
ampicillin, sodium salt		Roth
L-ascorbic acid, sodium salt		Roth
boric acid		Roth

bovine serum albumin (BSA)		Roth
bromophenol blue		Merck
buffer solutions pH 4.0, pH 7.0, pH 10.0	for calibration of pH electrode	Roth, Chemsolute
caffeic acid		Roth
caffeoyl-CoA	chemical synthesis according to Stöckigt and Zenk (1975)	AG Petersen
caffeoyl-4'-hydroxyphenyllactate	obtained from *Melissa officinalis*	AG Petersen
calcium chloride dihydrate		Sigma-Aldrich
cetyltrimethylammonium bromide (CTAB)		Roth
chloroform		Roth
chlorogenic acid		Roth
t-cinnamic acid		Merck
cinnamoyl-CoA	chemical synthesis according to Stöckigt and Zenk (1975)	AG Petersen
citric acid		Fluka
cobalt(II) chloride hexahydrate		Merck
coenzyme A (trihydrogen diphosphate)		AnGene, AA BLOCKS Inc.
cupric(II) sulfate pentahydrate		Fluka
Coomassie® Brilliant Blue G250		Serva
Coomassie® Brilliant Blue R250		Serva
4-coumaroyl-CoA	chemical synthesis according to Stöckigt and Zenk (1975)	AG Petersen
4-coumaroyl-4'-hydroxyphenyllactate	isolated and purified from Melissa officinalis	AG Petersen
cyclohexane		Fisher Scientific
dichloromethane		Roth
dimethylsulfoxide (DMSO)		Roth
dipotassium hydrogen phosphate trihydrate		Roth
disodium ethylenediaminetetraacetate dihydrate (EDTA-Na_2)		Roth
dNTPs (dATP, dCTP, dGTP, dTTP)		Fermentas
L-DOPA		Sigma-Aldrich
ethanol 96 %		Roth
ethidiumbromide		AppliChem
ethidiumbromide solution (0.025 %)		Roth
ethyl acetate		Fisher Scientific
ethylenediaminetetraacetic acid ferric, sodium salt (EDTA-Fe(III)-Na)		Sigma-Aldrich

feruloyl-CoA	chemical synthesis according to Stöckigt and Zenk (1975)	AG Petersen
ferulic acid		Roth
formic acid		Roth
D-(+)-galactose		Roth
garden compost	acquired at garden center	Compo
D-(+)-glucose		Roth
glycerol		Roth
L-glycine		Roth
guanidine hydrochloride solution 6.7 M		Roth
guanidine thiocyanate		Roth
L-histidine		Roth
hydrochloric acid		Roth
imidazole		Roth
iron(II) sulphate heptahydrate		Fluka
isopropanol		Roth
isopropyl-β-thiogalactopyranoside (IPTG)		Roth
lime	acquired at garden center	Schola Premium Gärtnerkalk
lithium chloride		Roth
lysozyme		Roth
magnesium chloride hexahydrate		Roth
magnesium sulfate heptahydrate		Duchefa
manganese chloride tetrahydrate		Roth
β-mercaptoethanol		Fisher Scientific
methanol		Fisher Scientific
N,*N*-dimethylformamide (DMF)		Merck
nitro blue tetrazolium (NBT)		Roth
nitrogen (liquid)		Linde
N-lauroylsarcosine sodium salt		Sigma-Aldrich
o-phosphoric acid (85 %)		Roth
D-phenylalanine		Sigma
L-phenylalanine		Roth
phenol (citrate buffer saturated)		Sigma-Aldrich
polysorbate 20 (Tween 20)		Sigma-Aldrich
polyvinylpyrrolidone (PVP) 40000		Sigma-Aldrich
potassium bromide		Roth
potassium chloride		Roth
potassium dihydrogen phosphate		Roth
potassium iodide		Merck
potassium nitrate		Roth
quartz sand		Roth
rosmarinic acid	isolated and purified from Melissa officinalis	AG Petersen

sinapic acid		Fluka
sinapoyl-CoA	chemical synthesis according to Stöckigt and Zenk (1975)	AG Petersen
skimmed milk powder		Spinnrad
sodium acetate trihydrate		Merck
sodium bicarbonate		Merck
sodium carbonate		Fluka
sodium chloride		Roth
sodium citrate		Roth
sodium dihydrogen phosphate monohydrate		Merck
sodium dodecyl sulfate	SDS Pellets	Roth
sodium hydroxide		Roth
sodium molybdate dihydrate		Fluka
sucrose	(aquired at the supermarket)	Diadem
tetracycline, hydrochloride		Sigma Chemical
tetramethylethylenediamine (TEMED)		Roth
tin(II) chloride		Roth
toluol		Roth
trifluoroacetic acid (TFA)		Roth
Triton® X-114		Fluka
tris(hydroxymethyl)aminomethane (TRIS)		Roth
tryptone/peptone		Roth
D-tyrosine		Roth
L-tyrosine		Merck
vanillic acid		Merck
xylene cyanol		Fluka
yeast extract		Roth
zinc chloride		Merck
zinc(II) sulphate heptahydrate		Duchefa

2.9.3 List of reagents and kits

Product	**Particulars**	**Producer**
AccuPrime™ *Taq* DNA Polymerase, High Fidelity		Invitrogen
Buffer T4 DNA Ligase		Thermo Scientific
Calf intestine alkaline phosphatase (CIAP)	1 U/µl	Fermentas
10X buffer for CIAP		Fermentas
Gene Ruler™ 1 kb DNA Ladder Mix	0.5 µg/µl	Thermo Scientific
Gene Ruler™ DNA Ladder Mix	0.5 µg/µl	Thermo Scientific

Goat Anti-Mouse IgG Fc, Alkaline Phosphatase (AP)	secondary antibody conjugate, affinity purified	Invitrogen (REF: A16087)
KOD Hot Start DNA Polymerase	1 U/µl	Novagen
Mouse anti-6xHis-tag monoclonal antibody	primary antibody	Invitrogen (REF: MA1-21315)
NucleoSpin® Gel and PCR Clean-up		Macherey-Nagel
Phusion® High Fidelity DNA Polymerase	2 U/µl	New England Biolabs
pRSET C		Invitrogen
QIAGEN® PCR Cloning Kit		Qiagen
QIAprep® Spin Miniprep Kit		Qiagen
Revert Aid First Strand cDNA Synthesis Kit		Thermo Scientific
RNase A	10 mg ml^{-1}	Fermentas
ROTI®Garose-His/Ni Beads		Roth
ROTI®Mark TRICOLOR		Roth
ROTIPHORESE® Gel 30 (37.5: 1)		Roth
T4 DNA Ligase		Thermo Scientific

2.9.4 List of devices, instruments and analytical consumables

Instrument	**Product**	**Producer/Distributer**
-80 °C freezer	C585 Innova	New Brunswick Scientific
agarose gel documentation chamber	FAS-Digi	Nippon Genetics
agarose gel electrophoresis chamber	multiSUB Midi	Cleaver Scientific
autoclave	AL02-02-100	Advantage-Lab
	VX-95	Systec
blotting chamber	Semi-Dry-Blotter PROfessional	Roth
blotting paper 0.35 mm	Type BP002	Hahnemühle
cell culture shakers	10X 400	Gallenkamp
	Ecotron	Infors HT
centrifuges	3-30KS	Sigma
	Heraeus Fresco 17	Thermo Scientific
	Heraeus Pico 17	Thermo Scientific
	Centrifuge 5415 D	Eppendorf
	Micro centrifuge SD	Roth
cuvettes	cuvettes, acrylic	Sarstedt
	cuvettes, polystyrene	Sarstedt

PD-10 columns Sephadex G-25M		GE Healthcare
freeze drier	Alpha 1-2 LDPlus	Christ
gas burner	Flammy S	Schütt
HPLC	D-2500 Chromato-Integrator	Merck-Hitachi
	L-4000 UV Detector	Merck-Hitachi
	L-6200A Intelligent Pump	Merck-Hitachi
	L-4250 UV-Vis Detector	Merck-Hitachi
	SpectraSystem SCM1000	Thermo Electron Corporation
	SpectraSystem UV1000	Thermo Electron Corporation
	SpectraSystem P4000	Thermo Electron Corporation
	Chromaster 5160 Pump + Organizer	VWR-Hitachi
	Chromaster 5280 Auto Sampler	VWR-Hitachi
	Chromaster 5310 Column oven	VWR-Hitachi
	Chromaster 5430 Diode Array Detector	VWR-Hitachi
ice machine	RF-0244A	Manitowoc
laminar air flow	Laminar Flow Workstation	Microflow
	Scanlaf Mars	Labogene
LC-MS	Agilent 1260 Series	Agilent
	micrOTOF-Q III MS with ESI source	Bruker Daltonics
LC-MS column	Multospher 120 RP18 column	CS-Chromatographie Service
magnetic stirrer	MR 3001	Heidolph
microwave		Bosch
mixer	Vortex-Genie 2	Scientific Industries
oven	U40	Memmert
PCR thermocycler	MyCycler	Bio-Rad
	Eppendorf Mastercycler gradient	Eppendorf
pH meter	FiveEasy + electrode	Mettler Toledo
photometer	BioPhotometer	Eppendorf
	Specord 200 plus	Analytik Jena
pipettes	Research plus (2 to 20 µl, 20 to 200 µl, 100 to 1000 µl, 1 to 10 ml)	Eppendorf
	Pipetman (0.2 to 2 µl, 2 to 20 µl, 20 to 200 µl, 100 to 1000 µl)	Gilson

pipette tips		Sarstedt
power source	EV3020	Consort
	EV2310	Consort
	E835	Consort
protein electrophoresis chamber	Mini-Protean 2 Cell	Bio-Rad
reaction tubes		Sarstedt
rocking platform	Rocker 35 EZ	Labnet International
	Duomax1030	Heidolph
rotary evaporator	Rotavapor RE120	Büchi
scales	EG300-3M	Kern
	440-35A	Kern
	440-47	Kern
	Explorer EX225D	Ohaus
	H64	Mettler
thermomixer	Thermomixer comfort	Eppendorf
	uniTHERMIX 2	LLG Labware
TLC plates	DC Cellulose F	Merck
	DC Kieselgel 60 F_{254}	Merck
transfer membrane	Immobilon®-P 0.45 µm	Millipore
	ROTI®PVDF 0.45 µm	Roth
	PVDF 0.2 transfer membrane (0.2 µm pore size)	Serva
ultrapure water	OmniaPure	Stakpure GmbH
ultrasonic bath	Sonorex Super RK 510 H	Bandelin
ultrasonic processor	UP 200 s	Dr. Hielscher GmbH
vacuum centrifuge	RVC 2-18 Cdplus	Christ
vacuum pump	MZ 2C NT	Vacuubrand
	Drehschieberpumpe P 4 Z	Ilmvac GmbH
vessels/lids for *C. braunii* growth	Rundrand Glas 100	Weck
water bath	E-PA/KUE	Julabo Labortechnik GmbH

2.9.5 Implemented software

Software	**Developer/Provider**
AspectUV 1.5	Analytik Jena
Blastn	NCBI
Blastp	NCBI
Bruker Data Analysis	Bruker Corporation
ChemDraw Professional 20	PerkinElmer Informatics
Citavi 6	Swiss Academic Software
CLC Sequence Viewer	CLC bio
Chromaster System Manager d-7000	Hitachi

EMBOSS Needle	EMBL-EBI
Expasy protparam	Swiss Institute of Bioinformatics (SIB)
Expasy translate	Swiss Institute of Bioinformatics (SIB)
GraphPad Prism 10	GraphPad
HPLC System Manager 4.0	Hitachi
MarpolBase	Genome Informatics Laboratory
MEGA X	MEGA, Pennsylvania State University
Microsoft Office 2010	Microsoft Corporation
NetGene2	DTU Health Tech (Denmark)
OrcAE	bioinformatics Gent
ORFfinder Home	NCBI
PhycoCosm	Joint Genome Institute (JGI)
Phytozome v13	Joint Genome Institute (JGI)
PyMOL	DeLano Scientific LLC, Schrödinger Inc.
SnapGene Viewer	SnapGene®
UniversalProtein Resource (Uniprot)	UniProt consortium
WinASPECT PLUS 4.2.9.0	Analytik Jena

2.10 PCR programs

Table 6 PCR programs and assay compositions for the PAL isoforms that were attained via molecular biological methods.

Name of the gene	**Source (cDNA/gDNA)**	**Polymerase**	**Assay composition**	**Program**
PpPAL 1 fw fragment	gDNA	KOD	- 1 µl gDNA *Pp* - 0.75 µl fw primer 1 - 0.75 µl rv primer 1 - 0.5 µl KOD - 2.5 µl KOD buffer - 2.5 µl dNTP's - 1.5 µl $MgSO_4$ - 15.5 µl H_2O	Cycle 1: 95 °C for 120 s, 95 °C for 20 s, 61-66 °C for 20 s, 72 °C for 25 s Cycle 2-40: 95 °C for 20 s, 61-66 °C for 20 s, 72 °C for 25 s
PpPAL 1 rv fragment	gDNA	KOD	- 1 µl gDNA *Pp* - 0.75 µl fw primer 2 - 0.75 µl rv primer 2 - 0.5 µl KOD - 2.5 µl KOD buffer - 2.5 µl dNTP's - 1.5 µl $MgSO_4$ - 15.5 µl H_2O	Cycle 1: 95 °C for 120 s, 95 °C for 20 s, 61-66 °C for 20 s, 72 °C for 25 s Cycle 2-40: 95 °C for 20 s, 61-66 °C for 20 s, 72 °C for 25 s
PpPAL 1 (full length sequence)	gDNA	AccuPrime™	- 1 µl fragment 1 (706 bp) - 1 µl fragment 2 (1517 bp)	Cycle 1: 95 °C for 120 s, 95 °C for 20 s, 61-66 °C for 20 s, 72 °C for 25 s

			- 0.5 µl fw primer 1 - 0.5 µl rv primer 1 - 0.1 µl AccuPrime™ - 2.5 µl AccuPrime™ buffer - 19.4 µl H_2O	Cycle 2-40: 95 °C for 20 s, 61-66 °C for 20 s, 72 °C for 25 s
PpPAL 2	gDNA	AccuPrime™	- 1 µl cDNA *Pp* - 0.5 µl fw primer - 0.5 µl rv primer - 0.1 µl AccuPrime™ - 2.5 µl AccuPrime™ buffer - 20.4 µl H_2O	Cycle 1: 95 °C, 180 s, 61-66 °C 60 s, 68 °C 150 s Cycle 2-39: 95 °C 60 s, 61-66 °C 60 s, 68 °C 150 s Cycle 40: 95 °C 60 s, 61-66 °C 60 s, 68 °C 300 s
PpPAL 3	gDNA	AccuPrime™	- 1 µl cDNA *Pp* - 0.5 µl fw primer - 0.5 µl rv primer - 0.1 µl AccuPrime™ - 2.5 µl AccuPrime™ buffer - 20.4 µl H_2O	Cycle 1: 95 °C for 120 s, 95 °C for 60 s, 58-64 °C for 60 s, 68 °C for 150 s Cycle 2-39: 95 °C for 60 s, 58-64 °C for 60 s, 68 °C for 150 s Cycle 40: 95 °C for 60 s, 58-64 °C for 60 s, 68 °C for 300 s
PpPAL 4	cDNA	AccuPrime™	- 1 µl cDNA *Pp* - 0.5 µl fw primer - 0.5 µl rv primer - 0.1 µl AccuPrime™ - 2.5 µl AccuPrime™ buffer - 20.4 µl H_2O	Cycle 1: 95 °C for 120 s, 95 °C for 60 s, 61-66 °C for 60 s, 68 °C for 150 s Cycle 2-39: 95 °C for 60 s, 61-66 °C for 60 s, 68 °C for 150 s Cycle 40: 95 °C for 60 s, 61-66 °C for 60 s, 68 °C for 300 s
CbPAL 1	cDNA	KOD	- 1 µl cDNA *Cb* - 0.75 µl fw primer - 0.75 µl rv primer - 0.5 µl KOD - 2.5 µl KOD buffer - 2.5 µl dNTP's - 1.5 µl $MgSO_4$ - 15.5 µl H_2O	Cycle 1: 95 °C for 120 s, 95 °C for 20 s, 64-69 °C for 20 s, 72 °C for 25 s Cycle 2-40: 95 °C for 20 s, 64-69 °C for 20 s, 72 °C for 25 s

MpPAL 1	gDNA	KOD	- 1 µl gDNA *Mp* - 0.75 µl fw primer - 0.75 µl rv primer - 0.5 µl KOD - 2.5 µl KOD buffer - 2.5 µl dNTP's - 1.5 µl $MgSO_4$ - 15.5 µl H_2O	Cycle 1: 95 °C for 120 s, 95 °C for 20 s, 57-63 °C for 20 s, 72 °C for 25 s Cycle 2-40: 95 °C for 20 s, 57-63 °C for 20 s, 72 °C for 25 s
MpPAL 2	gDNA	KOD	- 1 µl gDNA *Mp* - 0.75 µl fw primer - 0.75 µl rv primer - 0.5 µl KOD - 2.5 µl KOD buffer - 2.5 µl dNTP's - 1.5 µl $MgSO_4$ - 15.5 µl H_2O	Cycle 1: 95 °C for 120 s, 95 °C for 20 s, 57-63 °C for 20 s, 72 °C for 25 s Cycle 2-40: 95 °C for 0:20 min, 57-63 °C for 25 s, 72 °C for 25 s

2.11 Primer list

Table 7 List of applied primers with their respective restriction sites. Delineation also contains the primer names, sequences, annealing temperatures and restriction sites/enzymes. Restriction sites within the sequences are written in italics.

Name	**Sequence**	**Annealing temperature [°C]**	**Restriction enzyme**	**Target gene**
PAL000825f	ATGCCGATGGCAATATCAGGG	59.8	-	PpPAL 1
PAL000825r	TCAGCATGAGAAGGGTCCCG	61.4	-	PpPAL 1
PpPAL000825_fus-r1	TGGGCCTGGCAGTGAGGAGACCAGCGATGT	80.2	-	PpPAL 1 (fragment 1 rv)
PpPAL000825_fus-f2	TCCCGCTGTCGTACATCGCTGGTCTCCTCA	80.2	-	PpPAL 1 (fragment 2 fw)
PpPAL000825-VL-f	ATT*AGATCT*AATGCCGATGGCAATATCAGGG	65.5	BglII	PpPAL 1 (full length fw)
PpPAL000825-VL-r	AT*GAATTC*TCAGCATGAGAAGGGTCCCG	66.6	EcoRI	PpPAL 1 (full length rv)
PAL003442f	ATGGCACCACGAGCTGGG	60.5	-	PpPAL 2
PAL003442r	CTAGCAAGAGGTGCGAGCG	61.0	-	PpPAL 2

PpPAL003442-VL-f	ATT*AGATCT*AATGGCACCACGAGCTGGG	66.6	BglII	PpPAL 2
PpPAL003442-VL-r	AT*GAATTC*CTAGCAAGAGGTGCGAGCG	66.5	EcoRI	PpPAL 2
PAL014234f	ATGGCACCACGAGCTGA	55.2	-	PpPAL 3
PAL014234r	CTAGCAAGAGGCGCGAG	57.6	-	PpPAL 3
PpPAL014234-VL-f	ATT*AGATCT*AATGGCACCACGAGCTGA	63.4	BglII	PpPAL 3
PpPAL014234-VL-r	AT*GAATTC*CTAGCAAGAGGCGCGAG	64.6	EcoRI	PpPAL 3
PAL017163f	ATGGCACCACGAGCTGGG	60.5	-	PpPAL 4
PAL017163r	CTAGCAAGAGGCGCGAGAGA	61.4	-	PpPAL 4
PpPAL017163-VL-f	ATT*AGATCT*AATGGCACCACGAGCTGGG		BglII	PpPAL 4
PpPAL017163-VL-r	AT*GAATTC*CTAGCAAGAGGCGCGAGAGA		EcoRI	PpPAL 4
PAL018389f	ATGTTGATGAGCAAGAGGCCCA	60.3	-	PpPAL 5
PAL018389r	TTATACACACAACGACCCCGGC	62.1	-	PpPAL 5
PAL000816f	ATGCCGATGGCAATGCCA	56.0	-	PpPAL 6
PAL000816r	TCAGCATGAGAAGGGTCCCG	61.4	-	PpPAL 6
PAL003543f	ATGGCACCACGAGCTGGG	60.5	-	PpPAL 7
PAL003543r	CTAGCAAGAGGCTCGAGCC	61.0	-	PpPAL 7
PAL024089f	ATGGCACCACGAGCTGGG	60.5	-	PpPAL 8
PAL024089r	TCAACTAACAAGAGGCGCCAAC	60.3	-	PpPAL 8
PAL025848f	ATGGAATGCAGCCACCTC	56.0	-	PpPAL 9
PAL025848r	CTAGCAAGAGGCGCGAG	57.6	-	PpPAL 9
Chabra_PAL1263-f	ATGGCCCTGCTGGAGCCAGC	65.6	-	CbPAL 1
Chabra_PAL1263-r	TCATGAATAAAGCTGCCACGACTCCTTCC	66.7	-	CbPAL 1
Chabra_PAL3261-f	ATGTTCATCGGCGACCACATCAAGC	64.6	-	CbPAL 2

Chabra_PAL3 261-r	CTAATATAAACCTTC CAGGTTCAGGTGC	63.7	-	CbPAL 2
CBPAL2_Fw1	ATT*AGATCT*AATGAA AAAGCAATTTGAGTA	64.9	BglII	CbPAL 2
CBPAL2_Fw2	ATT*AGATCT*AATGCA GAGCGGTGAGCAG	65.1	BglII	CbPAL 2
CBPAL2_Rv	AT*GAATTC*CTAATAT AAACCTTCCAGGTTC	64.7	EcoRI	CbPAL 2
chabra-fw	ATT*AGATCT*AATGGC CCTGCTGGAGCCAGC	69.5	BglII	CbPAL 1
chabra-rv	AT*GAATTC*TCATGAA TAAAGCTGCCACGAC TCCTTCCG	70.5	EcoRI	CbPAL 1
CB1fw	ATT*GGATCC*AATGGC CCTGCTGGAGCCAGC	72.2	BamHI	CbPAL 1
Chara1-fw	ATT*GGATCC*TAATGG CCCTGCTGGAGCCAG C	72.1	BamHI	CbPAL 1
MpPAL(A)f	ATT*GGATCC*TAATGA ACGAATTCTTCTCGT C	64.2	BamHI	MpPAL 1
MpPAL(A)r	AT*AAGCTT*TTACGCT ATTGTTTTGGAAGGC	62.7	HindIII	MpPAL 1
MpPAL(B)f	ATT*GGATCC*TAATGA TGAACGAATTCTTC	61.0	BamHI	MpPAL 2
MpPAL(B)r	AT*GGATCC*TTAAACT ACAATTTTGCAAGG	61.0	BamHI	MpPAL 2
MpPAL€f	ATT*CTCGAG*TAATGA TGAACGAAATCTTCT CG	64.4	XhoI	MpPAL 5
MpPAL€r	AT*CAGCTG*TTATATT ACGATTCTAGACGGC T	64.2	PvuII	MpPAL 5
MpPAL(D)f	ATT*GGATCC*TAATGG TGGTCGAGACTGGA GC	69.5	BamHI	MpPAL 6
MpPAL(D)r	AT*CAGCTG*TTAGAAC GGGCCCGGGG	69.5	PvuII	MpPAL 6
MpPAL€f	ATT*GGATCC*TAATGG CGGCCATGGTGAT	66.6	BamHI	MpPAL 7
MpPAL€r	AT*GAATTC*TCATACA CGAGGCGCGTCGA	66.6	EcoRI	MpPAL 7
MpPAL(G)f	ATT*GGATCC*TAATGA TGCAAGTGAGAGAC TC	65.5	BamHI	MpPAL 8
MpPAL(G)r	AT*GGTACC*CTATACT ACAATCCCGCAGT	65.1	KpnI	MpPAL 8

Pp4CL1-f	ATT*GGATCC*AATGAG TGCGGGCATTGCCGA	69.5	BamHI	Pp4CL 1
Pp4CL1-r	AT*CCATGG*CTACACC TTGTTTCTGAGATCCT TC	68.2	NcoI	Pp4CL 1
Pp4CL2-f	ATT*GGATCC*AATGTC TCCTAGTTTGCTCC	65.3	BamHI	Pp4CL 2
Pp4CL2-r	AT*CCATGG*CTATACT TTGTTTCTTAGATCCT T	63.1	NcoI	Pp4CL 2
Pp4CL3-f	ATT*GGATCC*AATGTC TCCCAGTGTGATGTC GGA	69.5	BamHI	Pp4CL 3
Pp4CL3-r	AT*CCATGG*CTATACC TTGTTTCTCAGGTCCT TCC	69.5	NcoI	Pp4CL 3
Pp4CL4-f	ATT*GGATCC*AATGTC TCCTAGCATAATCTC GG	66.9	BamHI	Pp4CL 4
Pp4CL4-r	AT*CCATGG*CCTAGAA CTTCAAGGTGATATC CTTT	67.1	NcoI	Pp4CL 4
PAL-seq	CGCTTCCTGAATGCG GGAGTCCT	66.0	-	PpPAL 2 PpPAL 3 PpPAL 4 PpPAL 7 PpPAL 8 PpPAL 9
PAL-seq2	TGCCGCTCCGAGGCA CCATC	65.5	-	PpPAL 1 PpPAL 6
PAL-seq3	CGGCCTGCTTACCGG TCGC			PpPAL 9
M.pol.PALA_ mitte	CCGGCATTCGGTGG GCAATCC	26.1	-	MpPAL 1
M.pol.PALB_ mitte	GGTCTGCTTACCGGC CGTCCC	25.3	-	MpPAL 2

2.12 Genes ordered at BioCat GmbH

Table 8 List of genes ordered directly at BioCat GmbH. Also listed are restriction sites/enzymes and the corresponding target gene.

Name of gene	Restriction enzymes	Restriction sites	Target gene
Pp4CL1 in pUC57	EcoRI BglII	G_AATTC A_GATCT	Pp4CL 1
Pp4CL2 in pUC57	EcoRI BglII	G_AATTC A_GATCT	Pp4CL 2
Pp4CL3 in pUC57	EcoRI BglII	G_AATTC A_GATCT	Pp4CL 3
Pp4CL4 in pUC57	EcoRI BglII	G_AATTC A_GATCT	Pp4CL 4
Pp3c3_24370 in pUC57	EcoRI BglII	G_AATTC A_GATCT	Pp4CL 5
CBR_g21135 in pUC57	EcoRI BglII	G_AATTC A_GATCT	Cb4CL 1
CBR_g39225 in pUC57	EcoRI BglII	G_AATTC A_GATCT	Cb4CL 2
Mapoly0014s0059 in pUC57	EcoRI BglII	G_AATTC A_GATCT	Mp4CL 1
Mapoly0002s0100 in pUC57	EcoRI BglII	G_AATTC A_GATCT	Mp4CL 2
Mapoly0197s0014 in pUC57	EcoRI BglII	G_AATTC A_GATCT	Mp4CL 3
Mapoly0136s0004 in pUC57	EcoRI BglII	G_AATTC A_GATCT	Mp4CL 4
Mapoly0005s0086 in pUC57	EcoRI BglII	G_AATTC A_GATCT	MpPAL 3
Mapoly0142s0036 in pUC57	EcoRI BglII	G_AATTC A_GATCT	MpPAL 4
Pp4CL018389 in pUC57	EcoRI BglII	G_AATTC A_GATCT	PpPAL 5
CbPAL2k in pUC57	EcoRI BglII	G_AATTC A_GATCT	CbPAL 2 (short)
CbPAL2l in pUC57	EcoRI BglII	G_AATTC A_GATCT	CbPAL 2 (long)
Cb4CL1_mut in pUC57	EcoRI BglII	G_AATTC A_GATCT	Cb4CL 1 (inserted GEICIRG motif)
Cb4CL2_mut in pUC57	EcoRI BglII	G_AATTC A_GATCT	Cb4CL 2 (inserted GEICIRG motif)

3 Results

In the following, the genes that yielded proteins derived from *Chara braunii*, *Marchantia polymorpha* and *Physcomitrium patens* to be biochemically characterized are systematically listed depending on their origin organism. Afterwards, the characterized enzymes are subsequently listed, sorted by isoform and organism. The system of numbering for the different isoforms has been developed after characterization in order to maintain a steady chronological order of the viable enzymes as opposed to the ones that proved inoperative. The numbering for these last-mentioned enzymes continued the order at the point where the quantity of active isoforms was complete. Isoforms that were not examined also were not numbered.

3.1 General list of investigated proteins and putative PAL candidates

In the downstream tables, there are four different types of enzymes:

- expressed proteins that showed activity
- expressed proteins that showed no activity
- investigated genes that did not yield a protein
- gene annotations that were not investigated in this work.

The proteins that showed activity are marked in bold writing with **(+)**, the ones that were expressed but did not show any significant activity are marked in bold writing with **(-)**. All genes that have been obtained via molecular biological means in the laboratory are shown in red (bold writing), whereas genes that have been synthesized are shown in blue (bold writing). Investigated yet inoperative genes are shown in black and without bold writing. Annotations that were not examined are left without special designation in the section “name in this work”.

In the case of PAL, the selection was based on the presence of the ASG motif putatively responsible for the formation of the catalytic MIO group. Annotations that did not contain the motif were not investigated further. In the case of 4CL, annotations that implied a “real” 4CL were chosen for investigation, whereas annotations as “4CL-like” were left out.

All database entry numbers, annotations and names of the genes are listed for PAL in Table 9 and for 4CL in Table 10.

Table 9 Phytozome/OrcAE entries for the different annotations of PAL from *Chara braunii*, *Marchantia polymorpha* (v3.1) and *Physcomitrium patens* (v3.3). Further listed are the respective annotations, the convenient names of the isoforms given in Petersen lab, as well as the state of process.

Chara braunii		
OrcAE entry	**Annotated as**	**Name in this work**
CBR_g34530	HAL	**CbPAL 1 (+)**
CBR_g66119 (short)	HAL	**CbPAL 2 short (-)**
CBR_g66119 (long)	HAL	**CbPAL 2 long (-)**
Marchantia polymorpha		
Phytozome entry (v3.1)	**Annotated as**	**Name in this work**
Mapoly0070s0061	PAL	**MpPAL 1 (+)**
Mapoly0070s0071	PAL	**MpPAL 2 (+)**
Mapoly0005s0086	HAL	**MpPAL 3 (+)**
Mapoly0142s0036	PAL/HAL	**MpPAL 4 (-)**
Mapoly0005s0089	HAL	
Mapoly0009s0173	PAL	MpPAL 8
Mapoly0014s0211	PAL	MpPAL 6
Mapoly0044s0114	PAL	MpPAL 5
Mapoly0070s0065	HAL	
Mapoly0070s0066	HAL	
Mapoly0070s0068	PAL	The nucleotide sequence is identical to MpPAL 2
Mapoly0070s0069	HAL	
Mapoly0132s0049	PAL	MpPAL 7
Mapoly0066s0077	-	No ammonia-lyase
Physcomitrium patens		
Phytozome entry (v3.3)	**Annotated as**	**Name in this work**
Pp3c1_18940	HAL	**PpPAL 1 (+)**
Pp3c2_30610	HAL	**PpPAL 2 (+)**
Pp3c10_21810	HAL	**PpPAL 3 (+)**
Pp3c13_9000	HAL	**PpPAL 4 (+)**
Pp3c14_11870	HAL	**PpPAL 5 (-)**
Pp3c1_18830	HAL	PpPAL 6
Pp3c2_32410	HAL	PpPAL 7
Pp3c19_13690	HAL	PpPAL 8
Pp3c21_7680	HAL	PpPAL 9
Pp3c13_12190	HAL	
Pp3c21_7670	HAL	
Pp3c24_1311	HAL	
Pp3c1_18680	HAL	
Pp3c3_1810	PAL	
Pp3c3_8360	PAL	
Pp3c17_7320	PAL	
Pp3cc21_7640	HAL	
Pp3c22_17320	HAL	

Pp3c23_4200	PAL	
Pp3c26_2420	HAL	
Pp3c26_14610	HAL	

Table 10 Phytozome/PhycoCosm entries for the different annotations of 4CL from *Chara braunii*, *Marchantia polymorpha* (v3.1) and *Physcomitrium patens* (v3.3). Annotations, chronological numeration and activity status are provided for each isoform.

Chara braunii		
PhycoCosm/Phycocosm entry	**Annotated as**	**Name in this work**
CBR_21135	*o*-succinylbenzoic acid CoA-ligase	**Cb4CL 1 (+)**
CBR_39225	4CL	**Cb4CL 2 (+)**
Marchantia polymorpha		
Phytozome entry (v3.1)	**Annotated as**	**Name in this work**
Mapoly0014s0059	4CL-like	**Mp4CL 1 (+)**
Mapoly0002s0100	4CL	**Mp4CL 2 (+)**
Mapoly0197s0014	4CL	**Mp4CL 3 (+)**
Mapoly0136s0004	4CL	**Mp4CL 4 (+)**
Physcomitrium patens		
Phytozome entry (v3.3)	**Annotated as**	**Name in this work**
Pp3c19_13170	4CL	**Pp4CL 1 (+)**
Pp3c18_6360	4CL	**Pp4CL 2 (+)**
Pp3c21_15460	4CL	**Pp4CL 3 (+)**
Pp3c22_15350	4CL	**Pp4CL 4 (+)**
Pp3c3_24370	4CL	**Pp4CL 5 (-)**
Pp3c22_9051	4CL-like	
Pp3c19_10110	4CL-like	
Pp3c8_730	4CL-like	
Pp3c3_37980	4CL-like	
Pp3c1_9620	4CL-like	
Pp3c222_9050	4CL-like	

3.2 Preparation of pRSET C for expression

In order to provide a plasmid for cultivation purposes with the goal of protein expression, pRSET C was cultivated and linearized using the respective restriction enzymes. The restriction digest is shown in Figure 13. Restriction enzymes had to be adapted depending on the requested gene. The initial sample was retrieved from a stock sample kept in the lab of AG Petersen (originally acquired from Invitrogen). With this sample, overnight cultures were prepared and plasmid isolation was performed. The vector was linearized by restriction

digestion (see 2.3.15), yielding bands at approximately 3000 bp that were cut from the gel and purified. The plasmid-containing *E. coli* were stored in the form of glycerol stocks at -80 °C.

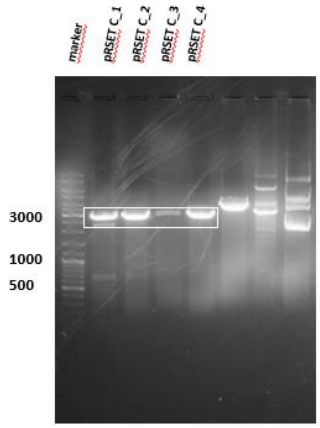


Figure 13 Restriction digest of pRSET C (in this case with BglII and EcoRI). The bands that were further processed are marked by a white box.

3.3. Amplification of coding DNA sequences, expression and verification of recombinant proteins

3.3.1 PAL from *Chara braunii*

Sequences for PAL derived from *C. braunii* were retrieved by searching the PhycoCosm database using the sequence of AaPAL 1 (AaPAL 1 Genbank MN378319) from *Anthoceros agrestis* as template. The search resulted in two very similar sequences (CBR_g57646 = Genbank accession GBG92888.1, CBR_g34530 = Genbank accession GBG82247.1), the differences only consisting of a small number of unidentified nucleotides. The corresponding proteins proposed by BlastP were > 95 % identical and seemingly not of eukaryotic origin. The sequence CBR_g34530 was amplified and was eventually translated into an active protein. Another sequence was suggested by Jan de Vries (Georg-August-Universität Göttingen), which was unusually long (3261 bp) and contained a potential bacterial HAL-like sequence 3'-part, while the 5'-part shared high similarity with a putative histidine tRNA-ligase. The 3'-part featured two putative start codons, resulting in two possible sequences of different length, possibly encoding a PAL/HAL-protein. Both of these sequences were treated as an independent gene respectively protein, in the following referred to as CbPAL 2s (short version) and CbPAL 2l (long version).

3.3.1.1 CbPAL 1

PCR with cDNA using KOD polymerase was performed based on the scaffold sequence CBR_g34530, yielding a DNA amplicon of 1584 base pairs length. The corresponding band was cut from the agarose gel and cleaned up before ligation into pDRIVE overnight. Afterwards, transformation with *E. coli* EZ was performed, followed by cultivation and plasmid isolation according to 2.3.12 to 2.3.13. Restriction digest with BglII and EcoRI and subsequent agarose gel electrophoresis yielded an insert with the right size that was cut from the gel, cleaned up according to 2.3.6 and ligated into pRSET C (expression plasmid, linearized beforehand with BglII and EcoRI), as described under 2.3.8.2. After transformation of *E. coli* EZ, cultivation and control digest, the plasmid was transferred into *E. coli* SoluBL21 and converted into a protein of 739 amino acids with a molecular mass of 56.97 kDa (61.71 kDa, including the His-Tag). For the PCR and restriction digest of CbPAL 1, see Figure 14. The protein was isolated using reusable His-Tag columns and desalted via PD-10 columns. Elution was performed using 3.5 ml 0.5 Tris HCl pH 8.5. Protein expression and His-tag purification were performed as described in 2.4.3 to 2.4.5. To subsequently validate the protein's presence, SDS-PAGE and Western blot were conducted according to 2.4.7 to 2.4.8, showing a faint but recognizable band at the appropriate height (see Figure 15).

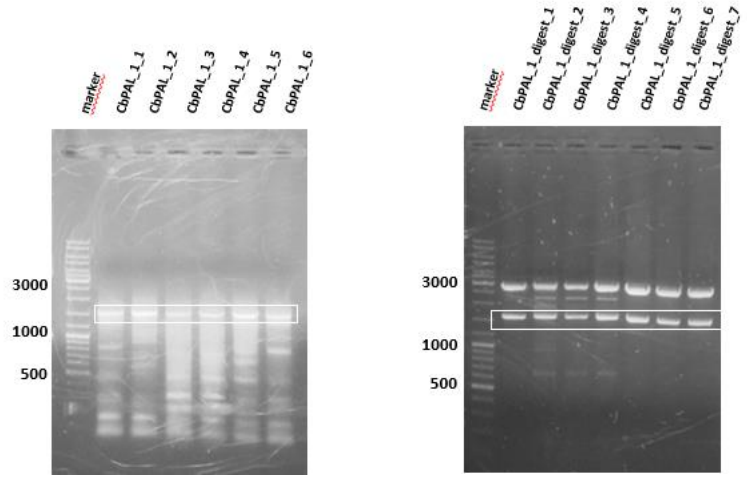


Figure 14 PCR of CbPAL 1 (left) and corresponding restriction digest of pRSET C containing CbPAL 1. PCR product CbPAL_1_1 was ligated and digest product CbPAL_1_digest_5 was used for expression. Bands on the correct height (PCR product on the left, released insert on the right) are highlighted by white boxes.

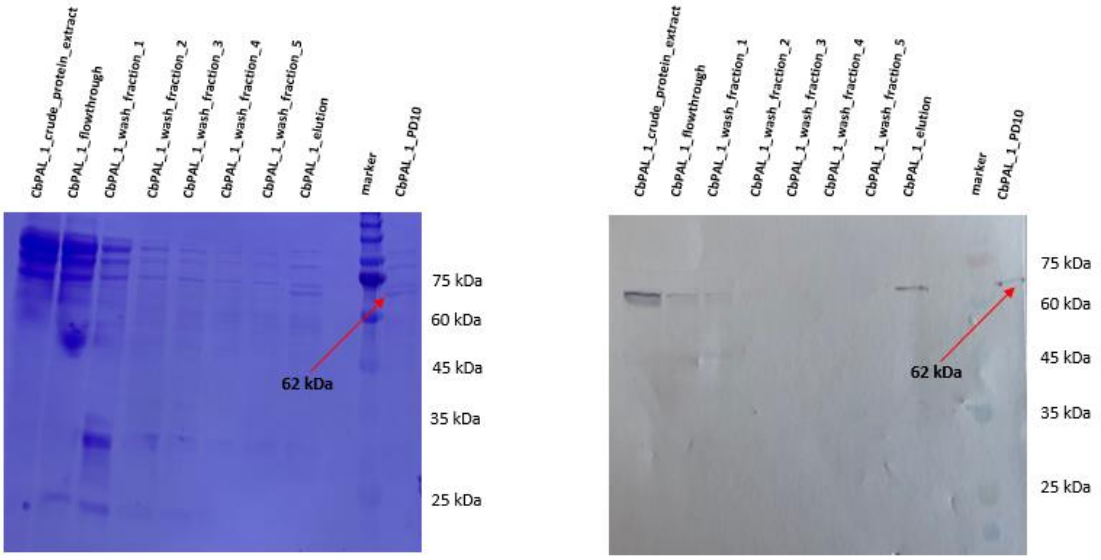


Figure 15 SDS-PAGE (left) and Western blot (right) of CbPAL 1 following metal chelate chromatography (applied samples from left to right: crude protein extract, flow-through, wash fractions 1 to 5, elution, marker, PD-10).

3.3.1.2 CbPAL 2

CbPAL 2 has been worked at in two different versions, a longer and a shorter variation of the sequence. Both versions of the gene were ordered codon-optimized at BioCat GmbH and arrived in the plasmid pUC57-BsaI freeze-dried to be resuspended in water (40 µl). 1 µl was used to conduct transformation, the transformed bacteria were cultivated and the plasmids isolated and submitted to restriction digest (restriction sites were BglII and EcoRI). The released insert was ligated into pRSET C for expression purposes and further treated analogously to CbPAL 1. The short version of the gene consisted of 1506 nucleotides and the corresponding protein possessed a molecular mass of approximately 59.97 kDa (including the His-Tag), while the nucleotide length of the longer version of CbPAL 2 was 1692 base pairs with a protein mass of about 67.08 kDa (including the His-Tag).

After expression, both protein varieties were cleaned up applying reusable His-Tag columns and cleansed from all salt and imidazole residues using PD-10 columns. Elution was performed with 3.5 ml 0.5 Tris-HCl pH 8.5. Immediately following elution, aliquots of the PD-10 fractions of both CbPAL 2s and CbPAL 2l underwent TCA precipitation, concentrating the protein solution 10:1. SDS-PAGE electrophoresis and Western blot were performed, yet only CbPAL 2s showed distinct bands at the correct heights of approximately 60 kDa. Bands were especially visible in

the crude protein extract and flow-through fractions, as well as in wash fraction 1 and the TCA precipitate. Restriction digests for the short and long version of CbPAL 2 are shown in Figures 16 and 17, respectively. Pictures of successful Western blot following SDS-PAGE gel electrophoresis for CbPAL 2s are depicted in Figure 18.

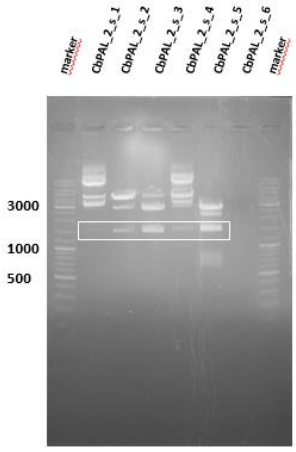


Figure 16 Restriction digest of pRSET C with CbPAL 2 (short version). The white box marks the correct height of the insert bands. Restriction product CbPAL_2_s_5 was further processed.

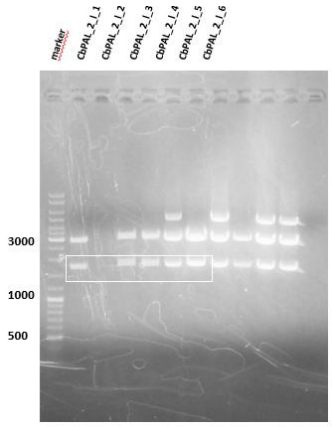


Figure 17 Restriction digest of pRSET C with CbPAL 2 (long version). The white box marks bands on the right height. Restriction product CbPAL_2_l_3 was further processed.

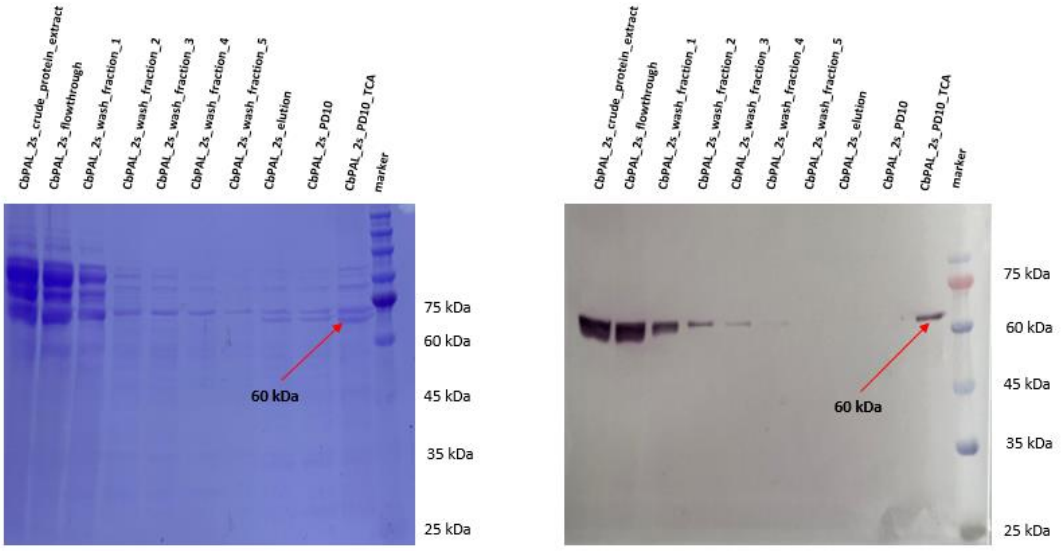


Figure 18 SDS-PAGE (left) and Western blot (right) of CbPAL 2 (short version) following metal chelate chromatography (applied samples from left to right: crude protein extract, flow-through, wash fractions 1 to 5, elution, PD-10, TCA precipitation of PD-10 (concentrated 10:1), marker).

3.3.2 PAL from *Marchantia polymorpha*

Sequences for PAL from *M. polymorpha* were retrieved from Phytozome 13 with AaPAL 1 as bait sequence (AaPAL 1 Genbank MN378319), the search yielded a total of fourteen hits. Seven sequences were annotated definitively as PAL: Mapoly0070s0061 (MpPAL 1), Mapoly0070s0071 (MpPAL 2), Mapoly0044s0114 (MpPAL 5), Mapoly0014s0211 (MpPAL 6), Mapoly0132s0049 (MpPAL 7), Mapoly0009s0173 (MpPAL 8) and Mapoly0070s0068. The nucleotide sequence of the last named, however, shared 100 % identity to MpPAL 2 and was therefore disregarded. Mapoly0142s0036 (MpPAL 4) showed ambiguous annotations, both as PAL and as HAL. Five sequences were definitively annotated as HAL, Mapoly0005s0086 (MpPAL 3), Mapoly0070s0065, Mapoly0070s0066, Mapoly0070s0069 and Mapoly0005s0089, and one sequence, Mapoly0066s0077, appeared on the annotations list yet without being designated as ammonia-lyase (see Table 8).

The seven sequences annotated as PAL were processed further, eventually resulting in two genes (Mapoly0070s0061 (MpPAL 1), Mapoly0070s0071 (MpPAL 2)), being amplified before protein expression was carried out. One of the sequences annotated as HAL, Mapoly0005s0086 (MpPAL 3) eventually proved to possess PAL activity and also susceptibility towards D-Phe, as well as L-Tyr. From the other sequences, Mapoly0142s0036 (MpPAL 4) was

also chosen for amplification, but did not show activity. For none of the other PAL annotations, any positive result in terms of presence or activity could be ascertained.

3.3.2.1 MpPAL 1

On the basis of the scaffold sequence Mapoly0070s0061, PCR with cDNA using KOD polymerase was performed. PCR yielded an amplicon of 2202 base pairs length, which was cleaned up, digested and ligated into pRSET C. Transformation with *E. coli* EZ was conducted, and the bacteria were cultivated before plasmid isolation and restriction digest verifying the presence of the insert. Both PCR and restriction digest are shown in Figure 19. Expression in *E. coli* SoluBL21 yielded a protein of 733 amino acids with a molecular mass of 80.05 kDa (84.78 kDa, including the His-Tag). The protein was purified using reusable His-Tag columns and desalted via PD-10 columns. Elution was performed using 3.5 ml 0.5 Tris-HCl pH 8.5. SDS-PAGE electrophoresis and Western blot validated the protein's presence, showing a vague band at approximately 85 kDa (see Figure 20).

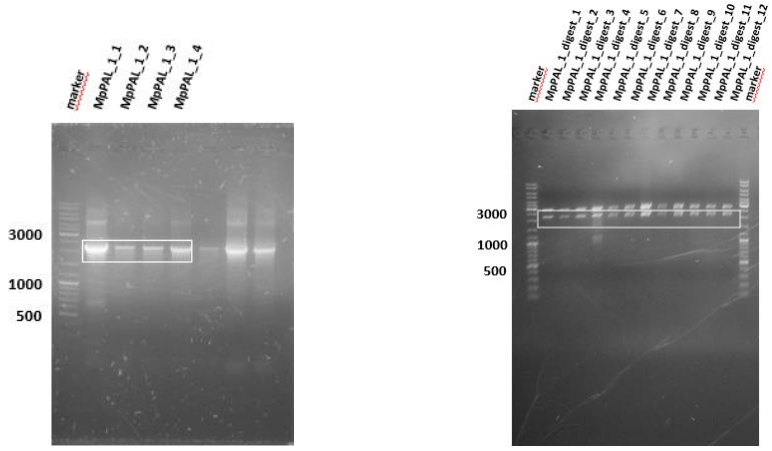


Figure 19 PCR of MpPAL 1 (left) and corresponding restriction digest of pRSET C containing insert (right). Bands identifying the corresponding PCR product respectively digest product are marked by white boxes. MpPAL_1_1 was used for ligation, whereas MpPAL_1_digest_7 was chosen for expression purposes.

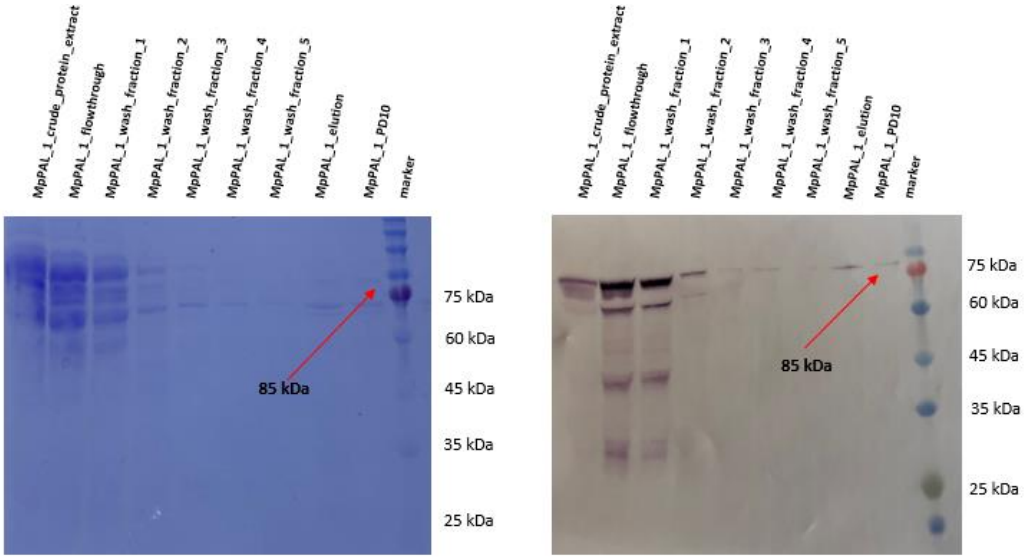


Figure 20 SDS-PAGE (left) and Western blot (right) of MpPAL 1 following metal chelate chromatography (applied samples from left to right: crude protein extract, flow-through, wash fractions 1 to 5, elution, PD-10, marker).

3.3.2.2 MpPAL 2

Based on the scaffold sequence Mapoly0070s0071, PCR with cDNA using KOD polymerase was executed. PCR yielded an amplicon of 2199 base pairs length (see Figure 21), which was processed analogously to MpPAL 1 (see 3.3.2.1). Because BamHI represented both restriction sites, the plasmid had to be dephosphorylated with CIAP (Calf intestinal Alkaline Phosphatase) before the insert could be ligated. After ligation into pRSET C and further cultivation, restriction digest was conducted in order to verify to presence of the gene (see Figure 21). Expression of the inserted sequence in *E. coli* resulted in a protein of 732 amino acids with a molecular mass of 80.40 kDa (85.14 kDa, including the His-Tag). The protein was purified using reusable His-Tag columns and desalted via PD-10 columns. Elution was performed using 3.5 ml 0.5 Tris-HCl pH 8.5. To confirm the presence and correct size of the protein, SDS-PAGE electrophoresis with subsequent Western blot was conducted. This did not yield a positive result, probably due to the low concentration of protein in the PD-10-fraction. The test was repeated three times with the same batch of protein solution and additional two times with a different expression, but no band at the correct height could be retrieved. In order to verify the presence of the (clearly active) protein, TCA precipitation was implemented, which led to a positive result (see 3.3.7). The successful SDS-PAGE electrophoresis and Western blot are shown in Figure 55.

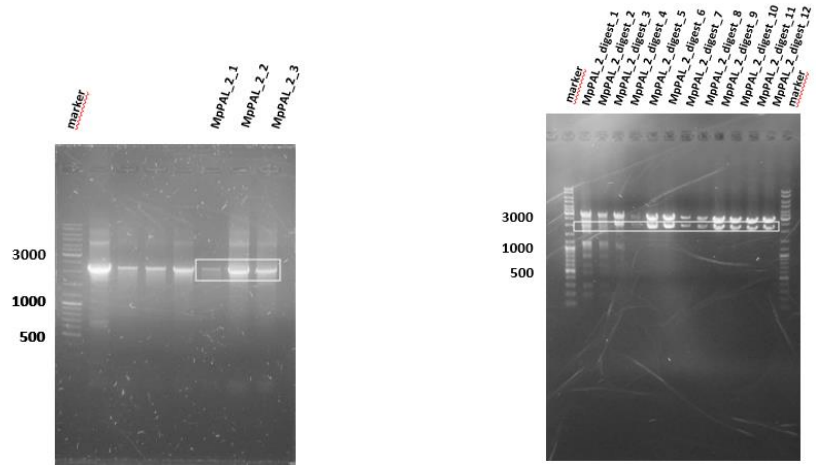


Figure 21 PCR of MpPAL 2 (left) and corresponding restriction digest of pRSET C containing MpPAL 2 as insert (right). Corresponding bands are framed by white boxes. MpPAL_2_2 was ligated into pRSET C and MpPAL_2_digest_5 was used for protein expression.

3.3.2.3 MpPAL 3

The gene sequence of MpPAL 3 contained an intron and the coding sequence was obtained via synthesis from BioCat GmbH and delivered codon-optimized on the basis of the scaffold sequence Mapoly0005s0086. The length of the complete open reading frame without the intron was 2097 nucleotides. Restriction sites ordered for synthesis were BglII and EcoRI. 1 µl of the plasmid solution was used to conduct transformation using *E. coli* EZ before cultivation and plasmid isolation with subsequent restriction digest. After verification via agarose gel electrophoresis, the insert was introduced into pRSET C and processed further, analogously to MpPAL 1 (see 3.3.2.1). The inserted sequence was subsequently transcribed and translated by *E. coli* SoluBL21 (for the control digest prior to expression, see Figure 22) into a protein of 699 amino acids with a molecular mass of 75.61 kDa (80.35 kDa, including the His-Tag). This protein was then cleaned up using reusable His-Tag columns and desalted via PD-10 columns. Elution was performed using 3.5 ml 0.5 Tris-HCl pH 8.5. The protein's presence was proven by SDS-PAGE electrophoresis and Western blot yielding very distinctive bands of about 80 kDa size in the fractions of the crude protein extract and the flow-through, in wash fraction 1 as well as in the elution and PD-10 fractions (see Figure 23).

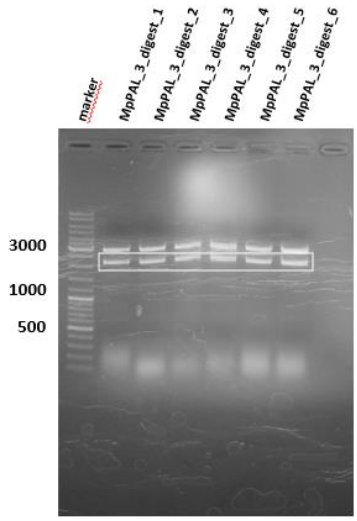


Figure 22 Restriction digest of pRSET C containing MpPAL 3 as insert. The resulting bands on the correct heights are marked by a white box. MpPAL_3_digest_6 was chosen for protein expression.

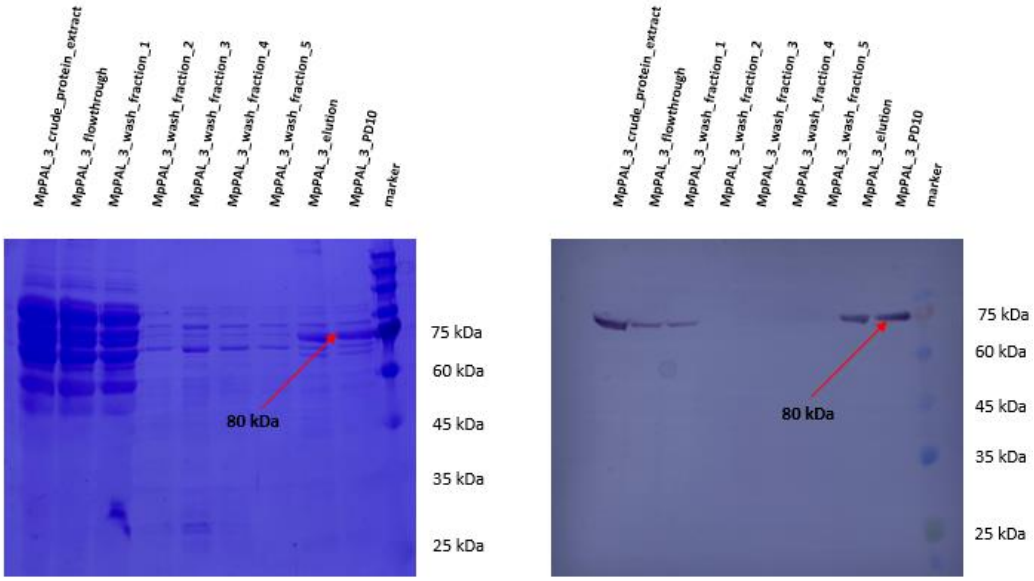


Figure 23 SDS-PAGE (left) and Western blot (right) of MpPAL 3 following metal chelate chromatography (applied samples from left to right: crude protein extract, flow-through, wash fractions 1 to 5, elution, PD-10, marker).

3.3.2.4 MpPAL 4

The coding sequence for MpPAL 4 was codon-optimized and synthesized by BioCat GmbH and arrived in freeze-dried form. *E. coli* EZ were transformed with 1 µl of the resuspended plasmid,

followed by cultivation and plasmid isolation. Afterwards, restriction digest with BglII and EcoRI was implemented, and agarose gel electrophoresis yielded bands at approximately 2200 bp. The released insert was ligated into pRSET C and further processed until restriction digest was performed (see Figure 24). The coding sequence for MpPAL 4 had a length of 2175 bp. Protein expression was performed after transformation of *E. coli* SoluBL 21 analogously to MpPALs 1 to 3 and SDS-PAGE electrophoresis with subsequent Western blot was conducted. The expected protein consisted of 725 amino acids and possessed a molecular mass of 84.12 kDa (including the His-Tag). Western blot, after several repetitions, yielded no positive result. TCA precipitation was performed in a ratio of 10:1, not resulting in any improvement.

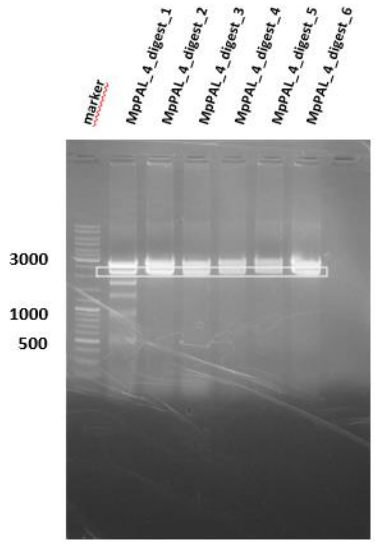


Figure 24 Restriction digest of pRSET C containing MpPAL 4. Digest product MpPAL_4_digest_2 was used for protein expression. Bands that represent the corresponding insert are framed by a white box.

3.3.2.5 MpPALs 5 to 8

Primers for MpPALs 5 to 8 were ordered and PCR was conducted applying both cDNA and gDNA. All three polymerases (GoTaq, AccuPrime™ and KOD) were tested for all isoforms in at least three separate experiments. In no case could any positive result be deduced after verification via agarose gel electrophoresis. These suggested MpPAL sequence from databases thus were not processed further.

3.3.3 PAL from *Physcomitrium patens*

The nucleotide sequences for PpPAL were searched for on the website Phytozome 13 using PAL 1 derived from *Anthoceros agrestis* (AaPAL 1 Genbank MN378319) as template. Investigations yielded 32 hits, most of them, however, annotated as HAL. Some sequences did not display the expected nucleotide length to qualify as a complete PAL/HAL gene. In order to rule out sequences appearing twice and thus limiting the number of genes to be examined, several analyses using Blast and Clustal Omega were conducted to discover similarities and repetitions. Searches for "histidine ammonia-lyase" did not lead to additional scaffold sequences. Eventually, nine promising isoforms were found for *P. patens*: Pp3c1_18940 (PpPAL 1), Pp3c2_30610 (PpPAL 2), Pp3c10_21810 (PpPAL 3), Pp3c13_9000 (PpPAL 4), Pp3c14_11870 (PpPAL 5), Pp3c1_18830 (PpPAL 6), Pp3c2_32410 (PpPAL 7), Pp3c19_13690 (PpPAL 8) and Pp3c21_7680 (PpPAL 9). These nine isoforms were chosen for further investigation. Four of them (PpPALs 1 to 4) could be amplified via molecular biological methods and one (PpPAL 5) was synthesized by BioCat GmbH, having been successfully amplified with gDNA but not cDNA and not possessing an open reading frame. In the case of isoforms 6 to 9, no positive result for molecular biological investigations could be attained.

3.3.3.1 PpPAL 1

Based on the scaffold sequence Phypa_000825, PCR was performed with both cDNA and gDNA. It was not possible to amplify the coding sequence using cDNA and because the sequence contained an intron of 114 base pairs length, fusion PCR was conducted as described under 2.3.7. Two separate PCRs yielded two fragments of 706 bp and 1517 bp, respectively. Both fragments were subsequently linked together via fusion PCR, resulting in a total nucleotide sequence length of 2223 bp. All three PCRs were performed using AccuPrime™ polymerase. After ligation into pDRIVE, multiplication in *E. coli* EZ and plasmid isolation, restriction digest was performed using EcoRI and BglII. The subsequent PCRs and the restriction digest of PpPAL 1 are depicted in Figure 25. The released insert was ligated into pRSET C (having been linearized with EcoRI and BglII prior to ligation) and multiplied in *E. coli* EZ for plasmid isolation and control digest. 5 µl of plasmid solution were used for transformation of *E. coli* SoluBL21 and protein expression was conducted. Expression yielded a protein of 740 amino acids with a molecular mass of 80.30 kDa (85.04 kDa, including the His-Tag), which was cleaned up using reusable His-Tag columns and desalted as well as

cleansed of imidazole via PD-10 columns. Elution was performed using 3.5 ml 0.5 Tris-HCl pH 8.5. To confirm the protein's presence, SDS-PAGE electrophoresis with subsequent Western blot was conducted, yielding clear bands in the fractions of crude protein extract, elution and PD-10. SDS gel and Western blot are shown in Figure 26.

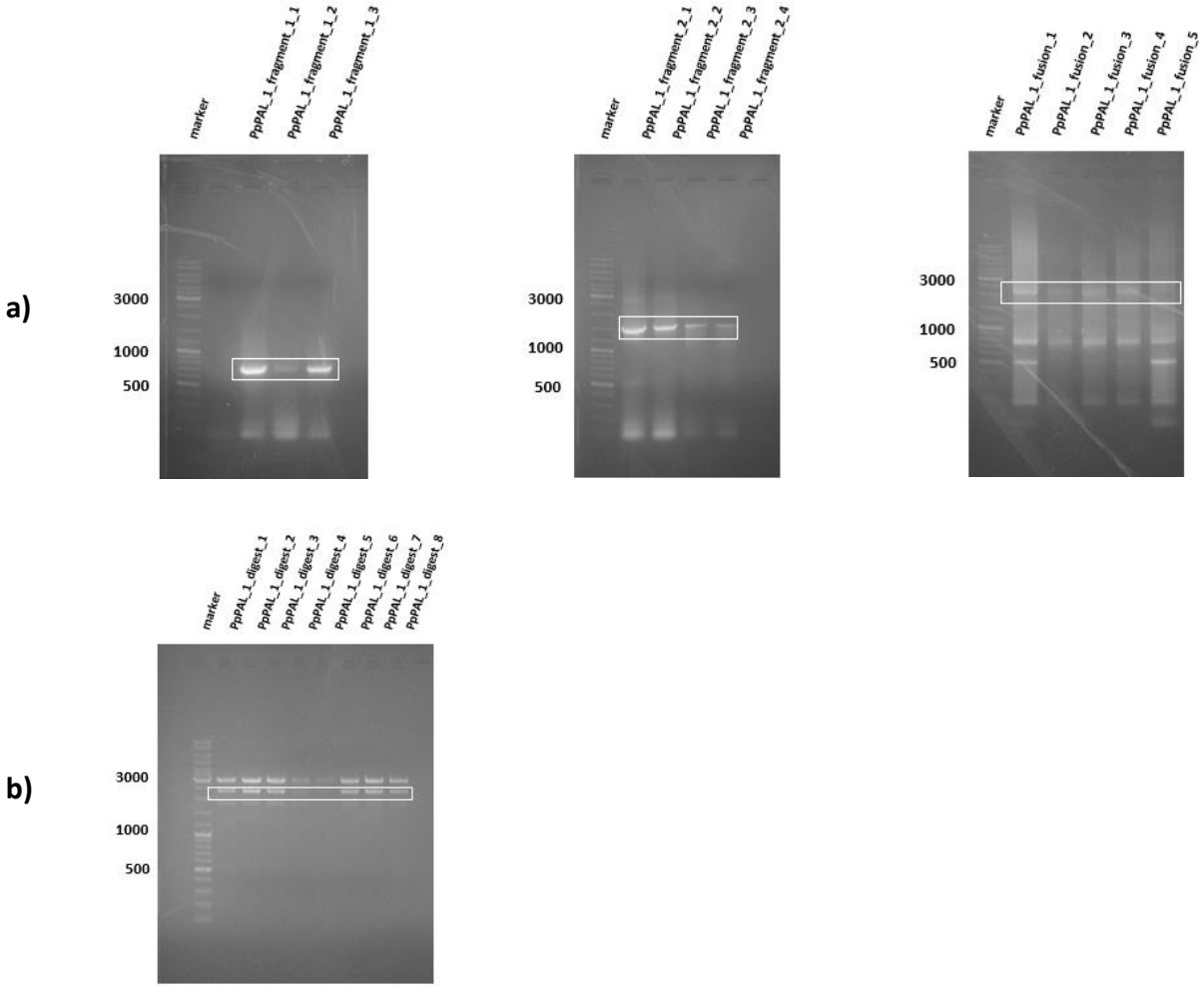


Figure 25 a) Fusion-PCR of PpPAL 1 (from left to right: amplification of fragment 1, amplification of fragment 2, fusion of both fragments); White boxes mark the bands to be further processed. PpPAL_1_fragment_1_1 and PpPAL_1_fragment_2_1 were chosen for fusion PCR, PCR product PpPAL_1_fusion_1 was chosen for ligation. b) restriction digest of pRSET C containing the complete sequence of PpPAL 1. The white box frames the bands representing the corresponding coding DNA. PpPAL_1_digest_2 was used for protein expression.

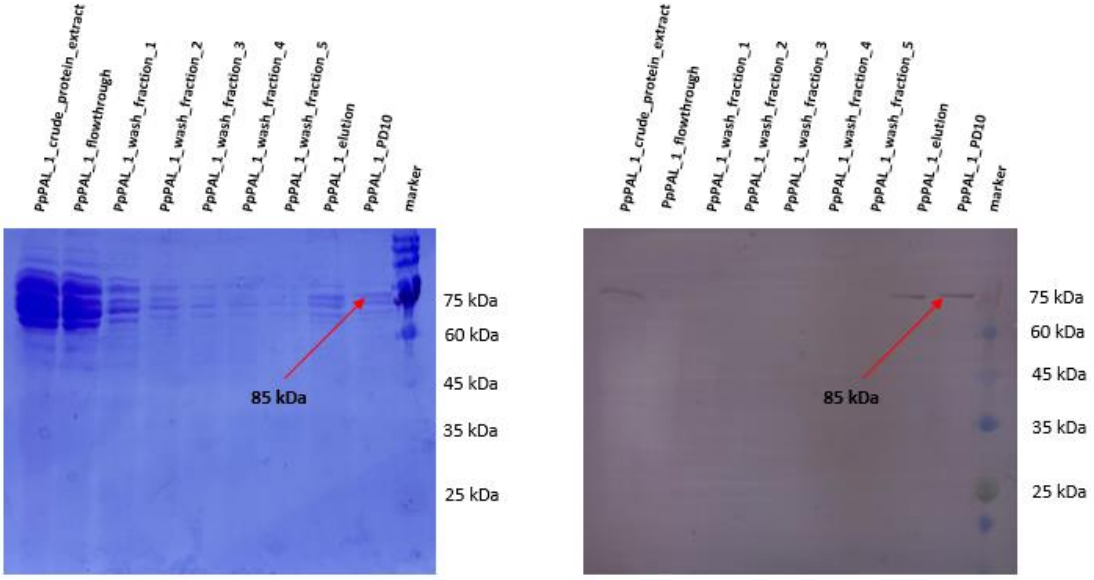


Figure 26 SDS-PAGE (left) and Western blot (right) of PpPAL 1 following metal chelate chromatography (applied samples from left to right: crude protein extract, flow-through, wash fractions 1 to 5, elution, PD-10, marker).

3.3.3.2 PpPAL 2

Based on the scaffold sequence Phypa_003442, PCR was performed with both cDNA and gDNA. PCR with gDNA using AccuPrime™ polymerase yielded an amplicon of 2217 base pairs length, which was processed analogously to PpPAL 1 as described in 3.3.3.1 (restriction sites were BglII and EcoRI) and yielded a protein of 738 amino acids with a molecular mass of 79.85kDa (84.59 kDa, including the His-Tag). Analysis regarding introns had been performed prior to PCR and shown that an open reading frame was present and no introns were located within the sequence. PCR and restriction digest of PpPAL 2 are depicted in Figure 27. The protein was cleaned up using reusable His-Tag columns and desalted via PD-10 columns. Elution was performed using 3.5 ml 0.5 Tris-HCl pH 8.5. To consolidate the presence of the protein, SDS-PAGE electrophoresis with subsequent Western blot was conducted. The blot yielded distinctive bands at the correct height in all fractions. For SDS gel and Western blot of PpPAL 2, see Figure 28.

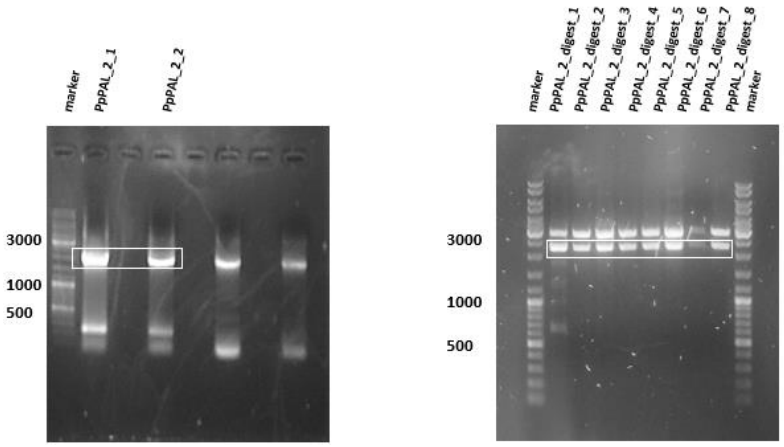


Figure 27 PCR of PpPAL 2 (left) and corresponding restriction digest of pRSET C with PpPAL 2 (right). White boxes frame the PCR products respectively bands representing the corresponding amplified sequence. PpPAL_2_1 was ligated into pRSET C and PpPAL_2_digest_4 was used for protein expression.

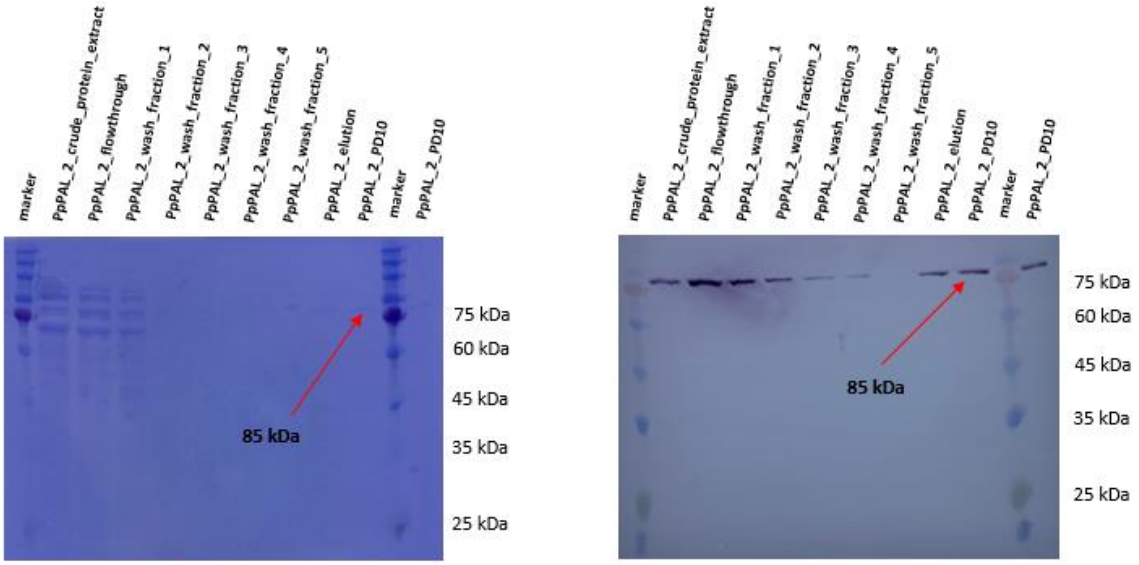


Figure 28 SDS-PAGE (left) and Western blot (right) of PpPAL 2 following metal chelate chromatography (applied samples from left to right: marker, crude protein extract, flow-through, wash fractions 1 to 5, elution, PD-10, marker, PD-10).

3.3.3.3 PpPAL 3

Based on the scaffold sequence Phypa_014234, PCR was conducted applying both cDNA and gDNA. PCR with gDNA using AccuPrime™ polymerase yielded an amplicon of 2220 base pairs length. Analysis of the gene sequence prior to PCR revealed that no introns interrupted the

coding sequence. The amplified DNA was processed analogously 3.3.3.1 (restriction sites were BglII and EcoRI) and protein expression yielded a protein of 739 amino acids with a molecular mass of 80.04 kDa (84.78 kDa, including the His-Tag). Depictions of PCR and restriction digest for PpPAL 3 are shown in Figure 29. The protein was purified by His-tag metal-chelate chromatography applying reusable columns and desalted using PD-10 columns. Elution was conducted using 3.5 ml 0.5 Tris-HCl pH 8.5. SDS-PAGE electrophoresis and Western blot were performed, yielding clear bands in all fractions, confirming the protein's presence (Figure 30).

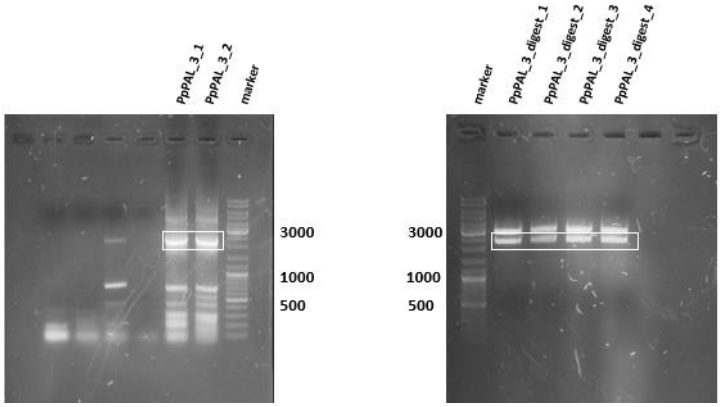


Figure 29 PCR of PpPAL 3 (left) and corresponding restriction digest of pRSET C with PpPAL 3 (right). Bands representing the PCR product respectively the released insert are marked by white boxes. PpPAL_3_1 was ligated and PpPAL_3_digest_4 was chosen for protein expression.

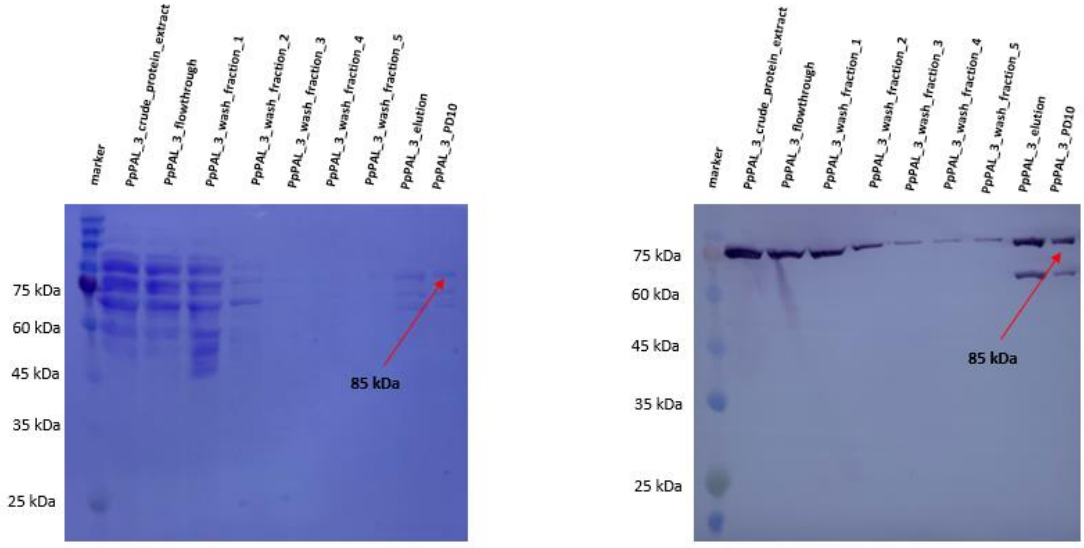


Figure 30 SDS-PAGE (left) and Western blot (right) of PpPAL 3 following metal chelate chromatography (applied samples from left to right: marker, crude protein extract, flow-through, wash fractions 1 to 5, elution, PD-10).

3.3.3.4 PpPAL 4

On the basis of the scaffold sequence Phypa_017163, PCR was performed with cDNA using AccuPrime™ polymerase and yielding an amplified DNA of 2217 base pairs length. Transformation of *E. coli* EZ as well as SoluBL 21 with the amplicon in pRSET C, cultivation and plasmid isolation were performed as described under 3.3.3.1 analogously to PpPAL 1 (with restriction sites BglII and EcoRI). Restriction digest was performed yielding bands of the correct size. PCR and restriction digest for PpPAL 4 are shown in Figure 31. The coding sequence was processed into a protein of 738 amino acids with a molecular mass of 80.06 kDa (84.80 kDa, including the His-Tag). The protein was cleaned up applying His-Tag columns and desalted and cleansed of imidazole via PD-10 columns. Applying 3.5 ml 0.5 Tris-HCl pH 8.5, the protein was eluted. To then confirm the protein's presence, SDS-PAGE electrophoresis with subsequent Western blot was conducted, showing clear bands at the correct height (see Figure 32).

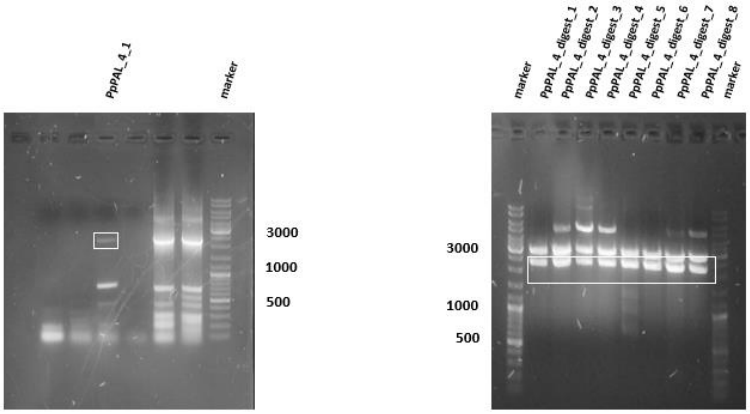


Figure 31 PCR of PpPAL 4 (left) and corresponding restriction digest of pRSET C containing the sequence of PpPAL 4 as insert (right). White boxes frame the PCR product respectively the bands representing the released insert. PCR product PpPAL_4_1 was ligated into pRSET C and PpPAL_4_digest_6 was chosen for protein expression.

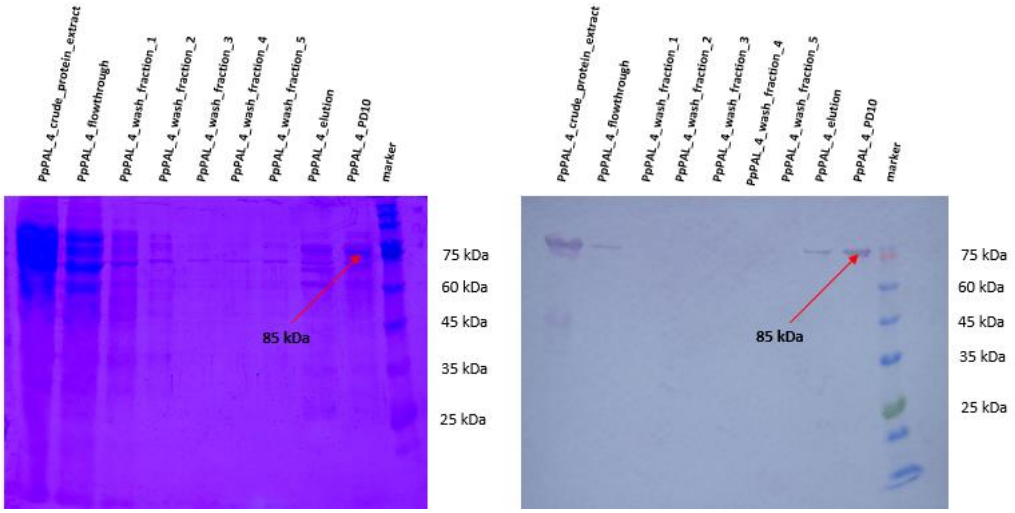


Figure 32 SDS-PAGE (left) and Western blot (right) of PpPAL 4 following metal chelate chromatography (applied samples from left to right: crude protein extract, flow-through, wash fractions 1 to 5, elution, PD-10, marker).

3.3.3.5 PpPAL 5

PpPAL 5 was investigated with both gDNA and cDNA on the basis of the scaffold sequence PpPAL018389. PCR with gDNA followed by agarose gel electrophoresis yielded a solitary band at a height of approximately 2800 bp, which correlated with the number of nucleotides in the postulated gene sequence that contained one intron. Disregarding the intron, the gene still constituted the longest sequence out of all PAL isoforms (2274 bp). The restriction digest resulting in the band of ca. 2800 bp in size is shown in Figure 33. The PCR in question was carried out applying gDNA, so that the intron was still in place, and no bands could be attained using cDNA. Because the PCR was performed with GoTaq polymerase possessing no proofreading capacity, the resulting band was not processed further. Applying other polymerases, a positive PCR could not be achieved and because of the additional impossibility of obtaining the gene free of the intron from cDNA, it was ordered to be codon-optimized and synthesized commercially by BioCat GmbH.

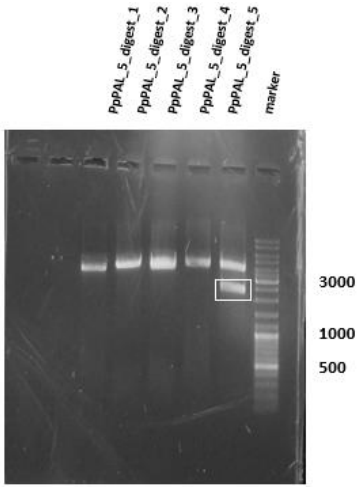


Figure 33 Restriction digest of pDRIVE containing PpPAL 5 with EcoRI. The white box frames the band of ca. 2800 bp in size which was further processed.

BglII and EcoRI were chosen as restriction sites. 1 µl of the delivered plasmid was applied to conduct transformation of *E. coli* EZ prior to cultivation and plasmid isolation with subsequent restriction digest. After verification via agarose gel electrophoresis, the synthesized DNA was processed further, analogously to PpPAL 1. The eventual restriction digest is shown in Figure 34. The introduced DNA was then transformed into a protein of 758 amino acids with a molecular mass of 82.11 kDa (86.84 kDa, including the His-Tag). The protein was cleaned up using reusable His-Tag columns and desalted via PD-10 columns. Elution was performed using 3.5 ml 0.5 Tris-HCl pH 8.5. SDS-PAGE gel electrophoresis and subsequent Western blot were conducted. The Western blot showed distinctive bands for the crude extract and flow-through, as well as the first wash fractions. Bands for both elution and PD-10 were very faint yet unequivocal and at the correct height, thus corroborating the presence of PpPAL 5 (see Figure 35).

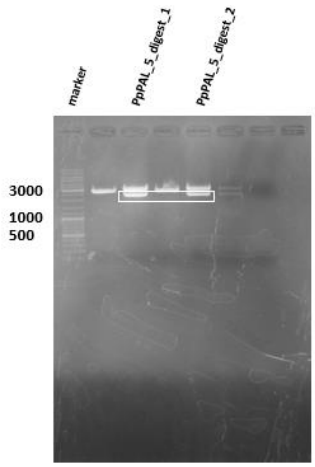


Figure 34 Restriction digest of pRSET C containing PpPAL 5. The white box frames the bands on the correct height to be further processed. PpPAL_5_digest_1 was chosen for protein expression.

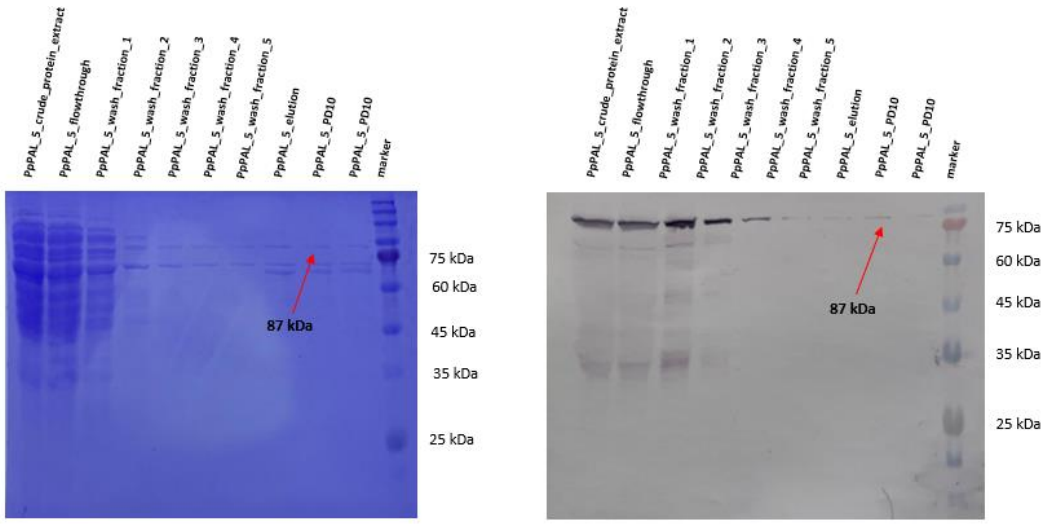


Figure 35 SDS-PAGE (left) and Western blot (right) of PpPAL 5 following metal chelate chromatography (applied samples from left to right: crude protein extract, flow-through, wash fractions 1 to 5, elution, PD-10, PD-10, marker).

3.3.3.6 PpPALs 6 to 9

For the remaining four isoforms derived from *P. patens*, PCR with both cDNA and gDNA was performed, implementing GoTaq, AccuPrime™ and KOD polymerase, respectively. None of the respective experiments yielded positive results after at least three different trials.

3.3.4 4CLs 1 and 2 from *Chara braunii*

Both genes for the two isoforms Cb4CL 1 and Cb4CL 2 were discovered searching the PhycoCosm database using the sequence of Aa4CL 1 (GenBank MN922305) as bait. The respective scaffold numbers were CBR_g21135 (Cb4CL 1) and CBR_g39225 (Cb4CL 2). Additional text search was carried out using the phrases "4-coumaric acid/4-coumarate CoA-ligase". Cb4CL 1 was annotated as "*o*-succinylbenzoic acid CoA-ligase". The genomic sequences contained eleven (Cb4CL 1) respectively two (Cb4CL 2) introns. The two amino acid sequences shared a similarity of < 30 % to Aa4CL 1 and also very low identity/similarity to one another. The GEICIRG motif (box I) putatively imperative for correct function of the enzyme was found to be distorted in Cb4CL 2 and completely missing in Cb4CL 1. The AMP-binding domain (box II) was discovered in both sequences, however, in both cases corrupted. The typical amino acid residues responsible for substrate recognition located between boxes I and II were found to be more or less absent (see alignment Figure 69).

The coding sequences for Cb4CLs 1 and 2 were synthesized by BioCat GmbH and arrived codon-optimized for the expression in *E. coli* on the basis of their respective scaffold sequences. Restriction sites for ligation into pRSET C were BglII and EcoRI. *E. coli* EZ were transformed with pDRIVE containing the synthesized sequences and cultivated analogously to 2.3.12 before plasmid isolation according to 2.3.13 was performed. Following restriction digest with BglII and EcoRI, the released DNA was ligated into pRSET C (expression plasmid) and transformation and cultivation were conducted. A control digest indicated the presence of the inserted sequence. Protein expression was conducted according to 2.4.3 followed by protein purification with His-tag columns. PD-10 columns were used for desalinization and removal of imidazole. Elution was performed using 3.5 ml 0.1 M KPi pH 7.5. SDS-PAGE electrophoresis with subsequent Western blot was conducted in order to confirm the protein's presence. The nucleotides sequences as well as the protein sizes and their molecular weights (with and without the His-Tag) are listed in Table 11. The restriction digests for Cb4CLs 1 and 2 are depicted in Figures 36 and 37, whereas SDS gels and Western blots are shown in Figures 38 and 39, respectively.

Table 11 Nucleotide sequence lengths, protein sizes and molecular weights (with and without His-Tag) for Cb4CL 1 and Cb4CL 2.

Cb4CL isoform	Nucleotide sequence length [bp]	Protein size [aa]	Molecular weight [kDa]	Molecular weight including His-Tag [kDa]
Cb4CL 1	1632	543	59.80	64.54
Cb4CL 2	1656	551	58.30	63.03

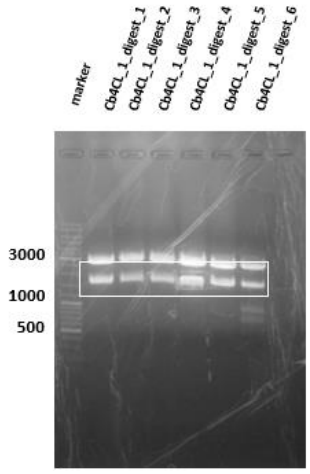


Figure 36 Restriction digest of pRSET C containing Cb4CL 1. The white box frames the bands on the correct heights to be processed further. Cb4CL_1_digest_4 was chosen for protein expression.

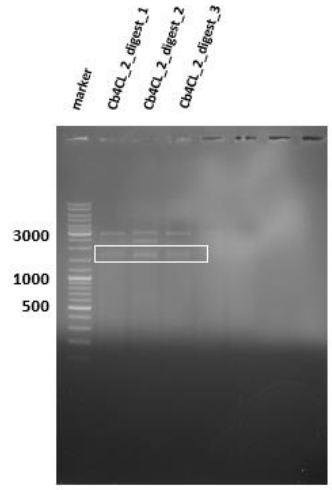


Figure 37 Restriction digest of pRSET C with insert Cb4CL 2. Bands representing genes on the right heights are marked by a white box. Cb4CL_2_digest_3 was further processed for protein expression.

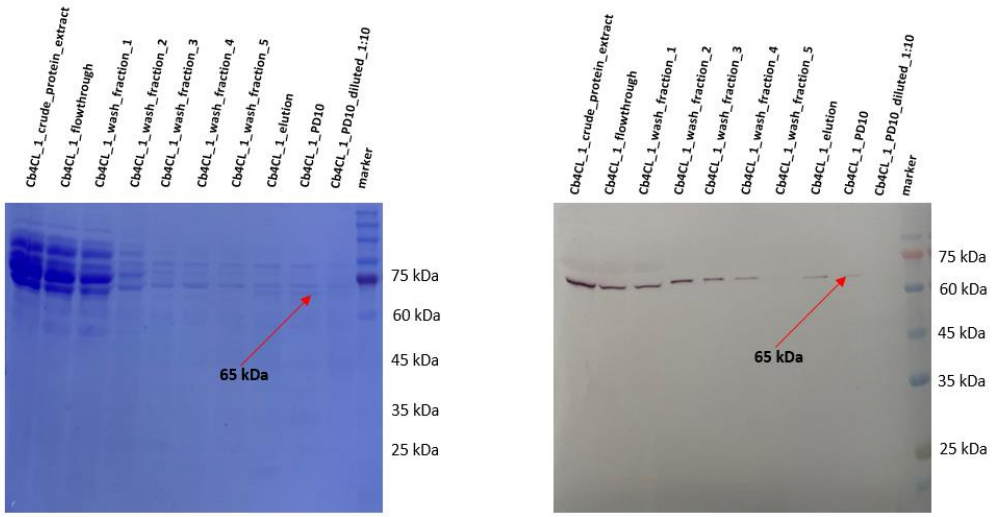


Figure 38 SDS-PAGE (left) and Western blot (right) of Cb4CL 1 following metal chelate chromatography (applied samples from left to right: crude protein extract, flow-through, wash fractions 1 to 5, elution, PD-10, PD-10 diluted 1:10, marker).

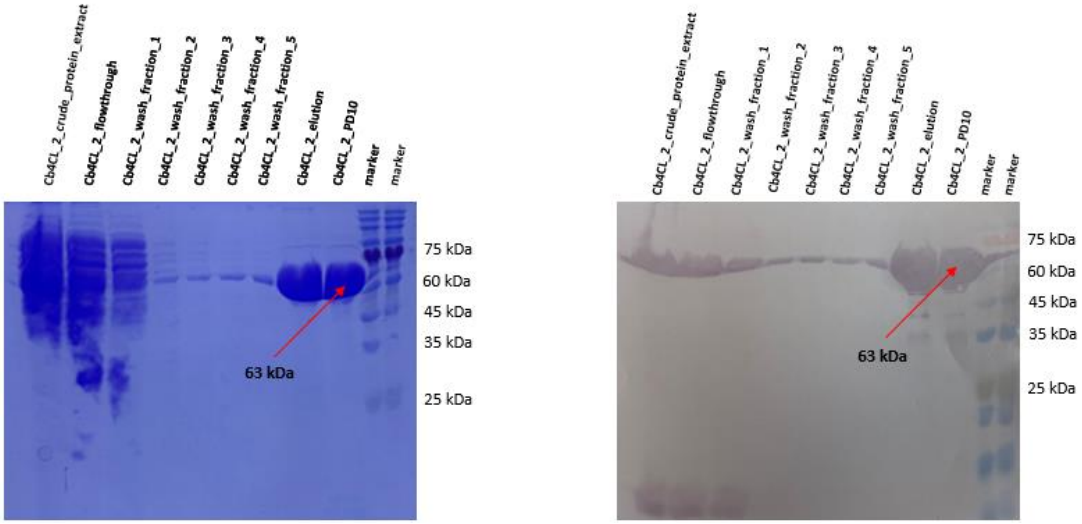


Figure 39 SDS-PAGE (left) and Western blot (right) of Cb4CL 2 following metal chelate chromatography (applied samples from left to right: crude protein extract, flow-through, wash fractions 1 to 5, elution, PD-10, marker, marker).

3.3.5 4CLs 1 to 4 from *Marchantia polymorpha*

The genes for all four isoforms of 4CL derived from *M. polymorpha* were retrieved from the Phytozome 13 database using Aa4CL 1 (GenBank MN922305) as bait sequence. The search yielded three sequences, Mapoly0002s0100 (Mp4CL 2), Mapoly0197s0014 (Mp4CL 3) and Mapoly0136s0004 (Mp4CL 4), annotated as 4CL and one sequence, Mapoly0014s0059 (Mp4CL 1), annotated as 4CL-like. Mp4CLs 2 to 4 contained five introns, whereas Mp4CL 1 contained six introns.

All four gene sequences were codon-optimized for the expression in *E. coli* and synthesized by BioCat GmbH in accordance with their respective scaffold sequences. Following ligation, transformation and restriction digest were performed, and recombinant proteins were expressed according to 2.4.3. The agarose gels of the restriction digests of Mp4CLs 1 to 4 are shown in Figure 40. Purification of the proteins was accomplished with the help of His-tag columns and desalination was performed afterwards, implementing reusable PD-10 columns. The proteins were eluted using 3.5 ml 0.1 M KPi pH 7.5 and the protein concentration was determined via Bradford assay. To check the protein's presence, SDS-PAGE electrophoresis with subsequent Western blot was performed. The protein sizes and their molecular weights (with and without the His-Tag) for the four isoforms of Mp4CL are listed in Table 12. Western blot proved that all four isoforms were successfully expressed, as can be interpreted from the presence of bands at the correct heights in SDS gels and blots, respectively. For SDS gels and Western blots for all four Mp4CL isoforms, see Figures 41 to 44.

Table 12 Nucleotide sequence lengths, protein sizes and molecular weights (with and without His-Tag) for Mp4CLs 1 to 4.

Mp4CL isoform	**Nucleotide sequence length [bp]**	**Protein size [aa]**	**Molecular weight [kDa]**	**Molecular weight including His-Tag [kDa]**
Mp4CL 1	1734	577	62.74	67.42
Mp4CL 2	1683	560	61.45	66.13
Mp4CL 3	1650	549	59.12	63.80
Mp4CL 4	1659	553	61.20	65.88

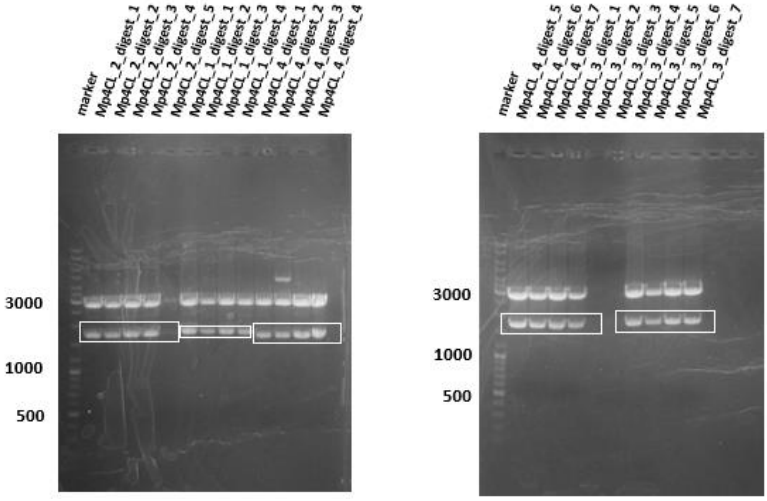


Figure 40 Restriction digest of pRSET C containing Mp4CLs 1 to 4. Shown are two agarose gels of the same digest. White boxes define the genes belonging together. First gel, from left to right: Mp4CL 2, Mp4CL 1, Mp4CL 4; second gel, from left to right: Mp4CL 4, Mp4CL 3. Further processed for protein expression were the following digest products: Mp4CL_1_digest_1, Mp4CL_2_digest_1, Mp4CL_3_digest_4 and Mp4CL_4_digest_4.

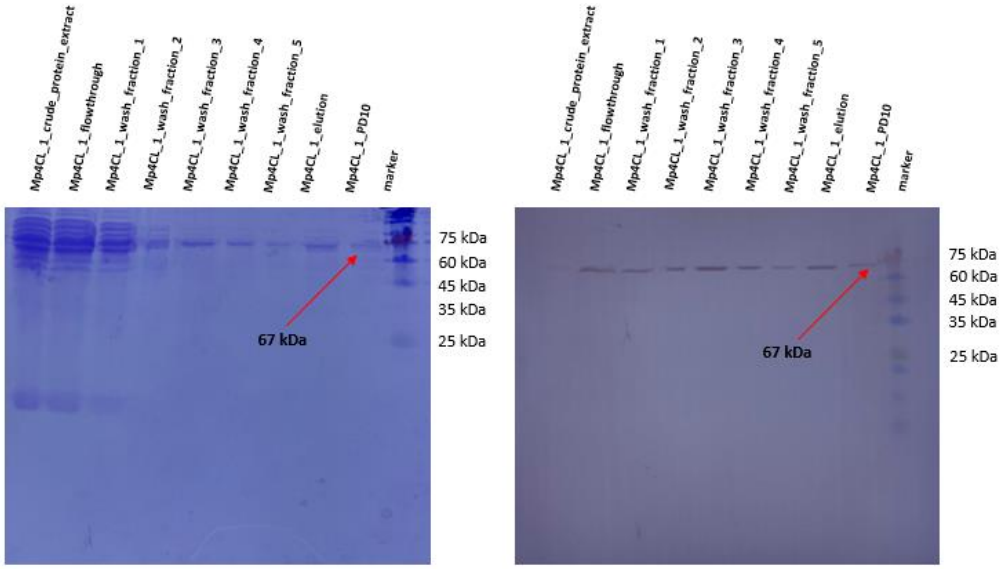


Figure 41 SDS-PAGE (left) and Western blot (right) of Mp4CL 1 following metal chelate chromatography (applied samples from left to right: crude protein extract, flow-through, wash fractions 1 to 5, elution, PD-10, marker).

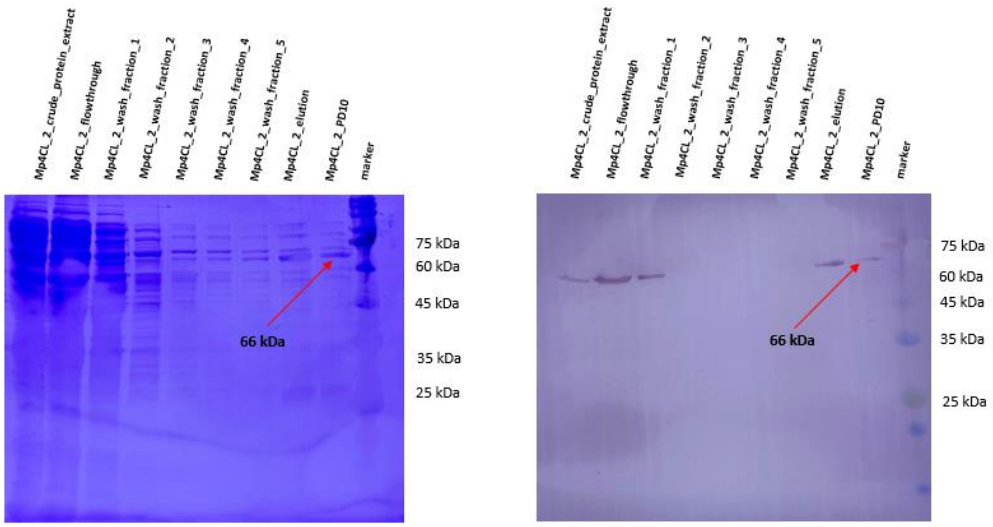


Figure 42 SDS-PAGE (left) and Western blot (right) of Mp4CL 2 following metal chelate chromatography (applied samples from left to right: crude protein extract, flow-through, wash fractions 1 to 5, elution, PD-10, marker).

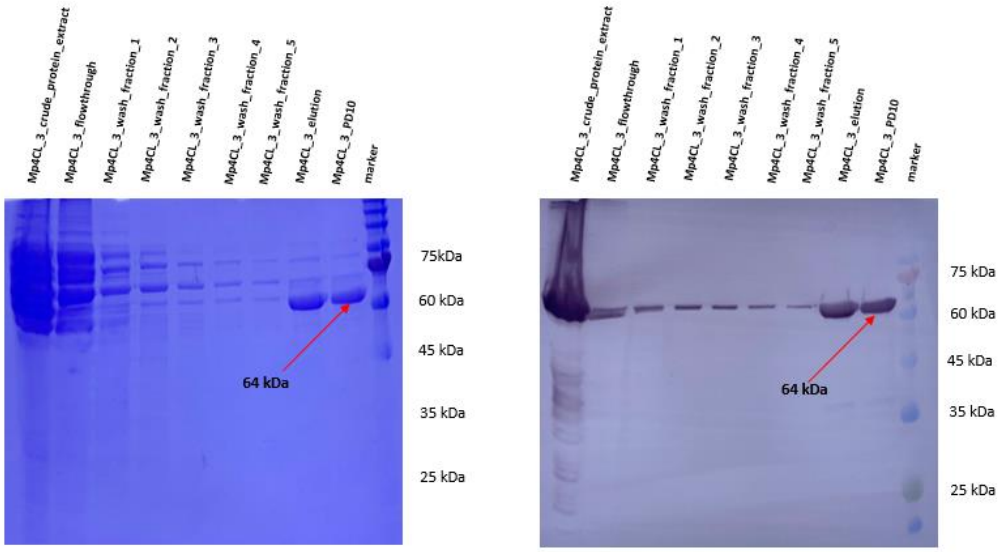


Figure 43 SDS-PAGE (left) and Western blot (right) of Mp4CL 3 following metal chelate chromatography (applied samples from left to right: crude protein extract, flow-through, wash fractions 1 to 5, elution, PD-10, marker).

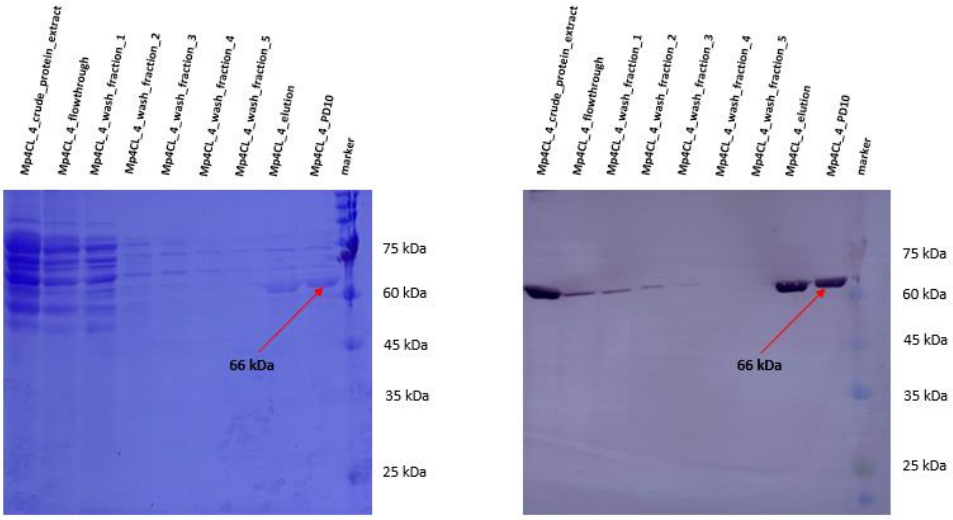


Figure 44 SDS-PAGE (left) and Western blot (right) of Mp4CL 4 following metal chelate chromatography (applied samples from left to right: crude protein extract, flow-through, wash fractions 1 to 5, elution, PD-10, marker).

3.3.6 4CLs 1 to 5 from *Physcomitrium patens*

Sequences for 4CLs derived from *P. patens* were searched for using the Phytozome 13 database. For this purpose, the coding sequence of 4CL 1 from *A. agrestis* was used as bait. Investigations yielded five sequences annotated as genuine 4CL and six annotated as 4CL-like. Four of the genuine 4CLs, Pp3c19_13170 (Pp4CL 1), Pp3c18_6360 (Pp4CL 2), Pp3c21_1540 (Pp4CL 3) and Pp3c22_15350 (Pp4CL 4), had already been described by Silber et al. (2008). The fifth, Pp3c3_24370 (Pp4CL 5), was investigated for the first time. Pp4CL 4 is composed of four exons, whereas Pp4CLs 1, 2 and 3 are composed of five. The sequence for Pp4CL 5 contained a total of seven introns. The search yielded an additional six sequences, harbouring between six and sixteen introns (see Table 9), all of which were annotated as “4CL-like”. For Pp4CLs 1 to 4, primers with restriction sites (in all cases BglII and EcoRI) were synthesized. Fresh plant material was harvested under sterile conditions and frozen in liquid nitrogen before RNA isolation was performed according to Chomczynski and Sacchi (1987). The integrity of the RNA was confirmed via agarose gel electrophoresis and the concentration was determined by photometric measurement at 260 and 280 nm. Applying 2 µg of the respective RNA per sample, cDNA synthesis was conducted. Next, PCR was performed with KOD and AccuPrime™ polymerase according to 2.3.4, not yielding any positive result for any of the five isoforms using cDNA.

gDNA of *P. patens* was isolated as described under 2.3.1. PCR was conducted with KOD polymerase using gDNA and applying primers designated for Pp4CLs 1 to 4. PCR yielded bands for isoforms 2, 3 and 4, but not for Pp4CL 1. The bands for Pp4CL 2 and Pp4CL 4 were very faint yet visible, whereas Pp4CL 3 yielded a distinct band at a height of approximately 2400 bp (see Figure 45). Despite successful PCR, verifying the natural existence of three of the Pp4CLs in *Physcomitrium patens*, the genes still seemed to contain introns, which could be inferred by the length of the obtained gene. The band for Pp4CL 2 was at approximately 2200 bp, which correlated with its expected position when including the introns (2126 bp total). For Pp4CL 4, the height was approximately 2700 bp, which was also in accordance with the sequence length including the introns (2742 bp). The band for Pp4CL 3 was located at approximately 2400 bp, a height too low to be considered a positive outcome, considering how the expected position including the introns would have to be 3216 bp. Further experiments with gDNA did not lead to clearer results and the inability to obtain the coding sequences from cDNA made it necessary to acquire the sequences via commercial synthesis.

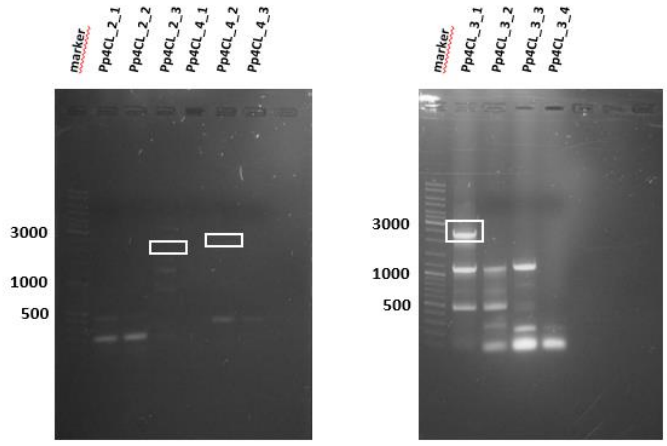


Figure 45 Agarose gel electrophoresis with PCR products of Pp4CL 2 and Pp4CL 4 (left) and Pp4CL 3 (right). Faint bands for Pp4CL 2 (Pp4CL_2_3) and Pp4CL 4 (second sample) are marked by white boxes, respectively; clear band for Pp4CL 3 (first sample) also marked by a white box.

The coding sequences were synthesized by BioCat GmbH and arrived codon-optimized for the expression in *E. coli* to be processed further. Restriction sites were BglII and EcoRI. The gene for Pp4CL 5 and its annotation as “real” 4CL was discovered after examinations of the other four isoforms were already complete, and it was ordered directly from BioCat GmbH (restriction sites were BglII and EcoRI). 1 µl of the resuspended solution of the synthesized DNA

in the plasmid pUC57 was applied for transformation of competent *E. coli* before cultivation and plasmid isolation. After restriction digest, the released insert DNA was then ligated into an expression plasmid (pRSET C) and further processed as described in 2.3.11 to 2.3.14 before protein expression according to 2.4.3 was performed. The agarose gels for the restriction digest prior to protein expression for all four isoforms are depicted in Figures 46 to 49. The protein was cleaned up applying reusable His-tag columns and desalted, implementing PD-10 columns. Elution was performed with 3.5 ml 0.1 M KPi pH 7.5. To confirm the protein's presence for all five isoforms, SDS-PAGE electrophoresis with subsequent Western blot was conducted. The protein sizes and their molecular weights (with and without the His-Tag) are listed in Table 13. For Pp4CL 3 and 5, TCA precipitation according to 2.4.9 was performed prior to SDS-PAGE since the protein concentration was evidently insufficient. In the case of Pp4CLs 1 to 4, successful Western blot corroborated the proper expression of the respective proteins (see Figures 50 to 53). Following expression and verification, the enzyme activities were photometrically characterized. For Pp4CL 5, no successful SDS-PAGE gel or Western blot could be experimentally generated and this particular isoform also showed no signs of activity (see also 3.4.6.1). The restriction digest prior to protein expression for Pp4CL 5 is shown in Figure 54.

Table 13 Nucleotide sequence lengths, protein sizes and molecular weights (with and without His-Tag) for Pp4CLs 1 to 5.

Pp4CL isoform	**Nucleotide sequence length [bp]**	**Protein size [aa]**	**Molecular weight [kDa]**	**Molecular weight including His-Tag [kDa]**
Pp4CL 1	1923	641	69.97	74.70
Pp4CL 2	1755	585	63.90	68.64
Pp4CL 3	1710	570	62.31	67.04
Pp4CL 4	1728	576	63.12	67.86
Pp4CL 5	1524	508	56.16	60.90

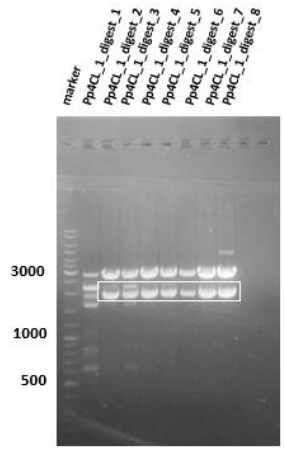


Figure 46 Restriction digest of pRSET C with insert Pp4CL1. The white box marks the further processed bands at the correct height. Pp4CL_1_digest_7 was used for protein expression.

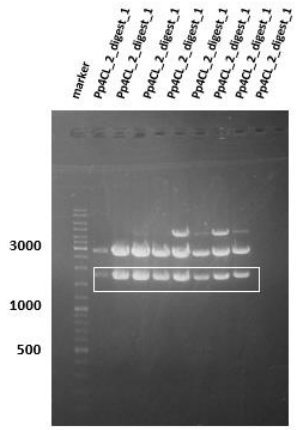


Figure 47 Restriction digest of pRSET C with Pp4CL 2. The white box frames the bands to be processed further. Pp4CL_2_digest_2 was chosen for protein expression.

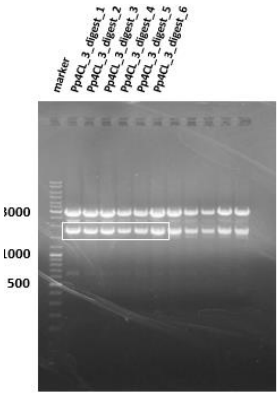


Figure 48 Restriction digest of pRSET C Pp4CL 3. Bands on the correct heights are marked by a white box. Pp4CL3_digest_2 was used to conduct protein expression.

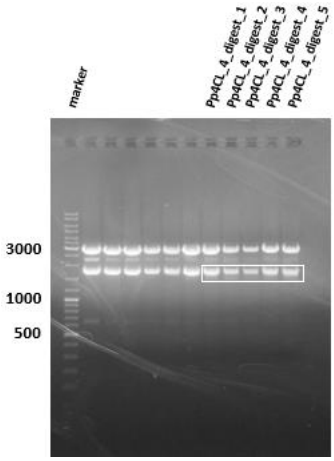


Figure 49 Restriction digest of pRSET C containing Pp4CL 4. The white box frames the bands on the right heights to be processed further. Pp4CL_4_digest_1 was processed further.

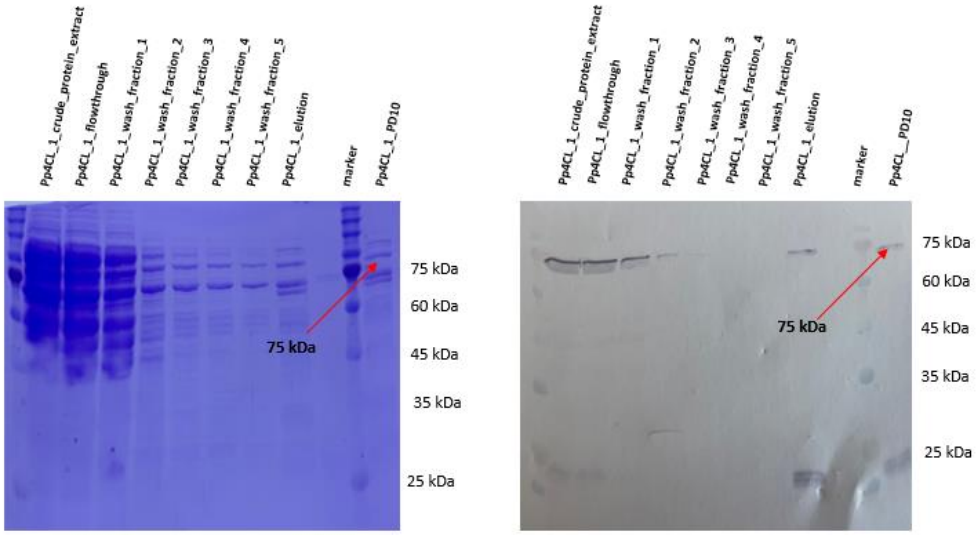


Figure 50 SDS-PAGE (left) and Western blot (right) of Pp4CL 1 following metal chelate chromatography (appended samples from left to right: crude protein extract, flow-through, wash fractions 1 to 5, elution, marker, PD-10).

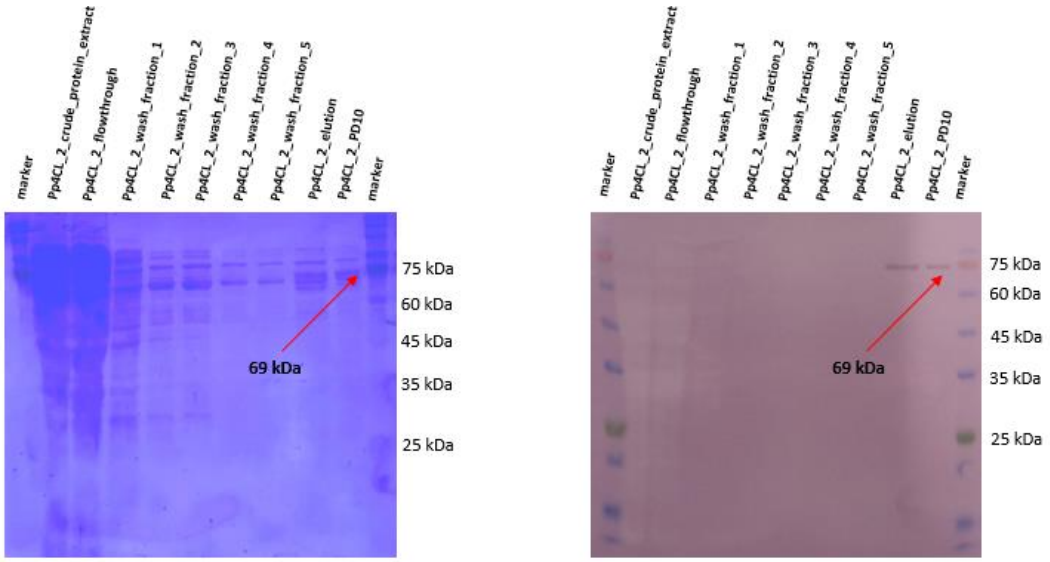


Figure 51 SDS-PAGE (left) and Western blot (right) of Pp4CL 2 following metal chelate chromatography (implemented samples from left to right: marker, crude protein extract, flow-through, wash fractions 1 to 5, elution, PD-10, marker).

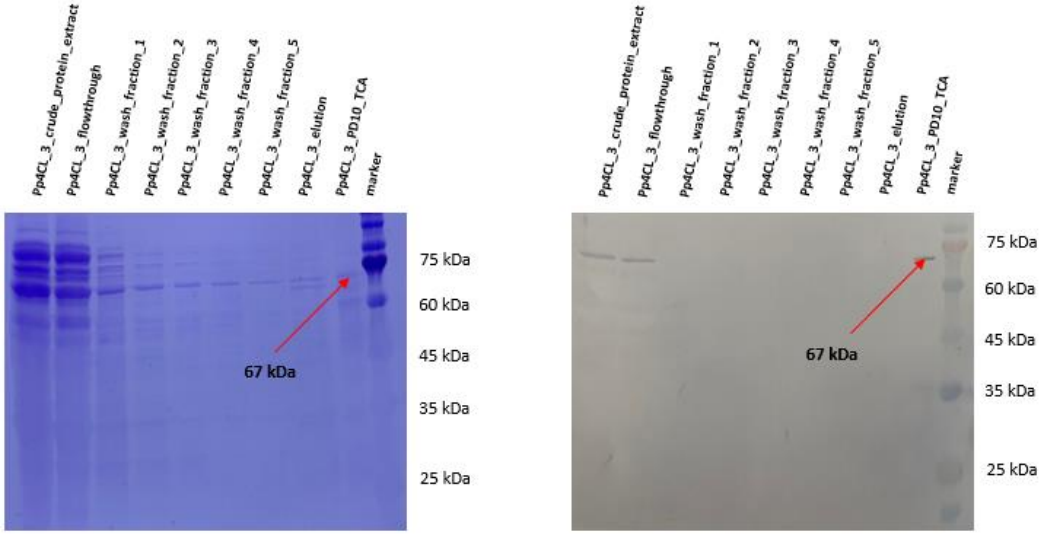


Figure 52 SDS-PAGE (left) and Western blot (right) of Pp4CL 3 following metal chelate chromatography (applied samples from left to right: crude protein extract, flow-through, wash fractions 1 to 5, elution, PD-10-fraction as TCA precipitate, marker).

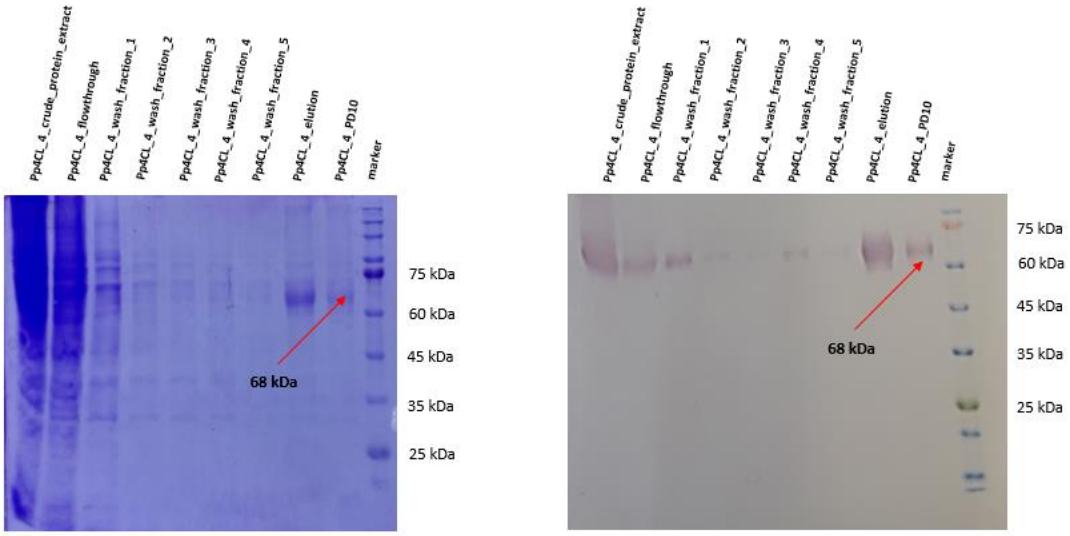


Figure 53 SDS-PAGE (left) and Western blot (right) of Pp4CL 4 following metal chelate chromatography (appended samples from left to right: crude protein extract, flow-through, wash fractions 1 to 5, elution, PD-10, marker).

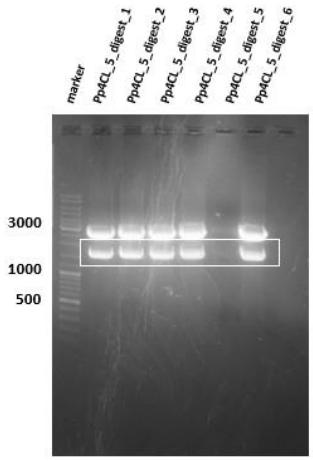


Figure 54 Restriction digest of pRSET C containing Pp4CL 5 as insert. The white box frames the bands on the right heights to be processed further. Pp4CL_5_digest_4 was used for protein expression.

3.3.7 TCA precipitation with MpPAL 2 and Pp4CL 5

In the case of MpPAL 2 and Pp4CL 5, no successful Western blot was viable, presumably due to low protein concentration within the samples. Because of this, TCA precipitation according to 2.4.9 was performed for both enzymes. In this case, the concentrations per sample were increased tenfold. For this particular experiment, only the PD-10-fractions were used.

Following TCA precipitation, SDS-PAGE electrophoresis with subsequent Western blot was conducted. The SDS gel showed bands on the correct height for both proteins, yet the blot validated only the presence of MpPAL 2. For Pp4CL 5, a band appeared on the blot, but it was located far too high in terms of protein size to be considered as an actual verification of Pp4CL 5. SDS gel and Western blot are depicted in Figure 55.

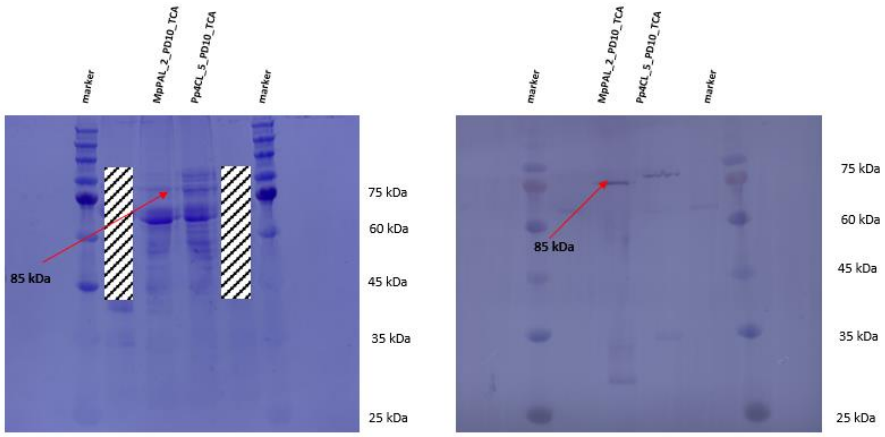


Figure 55 SDS-PAGE (left) and Western blot (right) applying TCA precipitated His-tag-purified and desalted protein in a relation of 10:1 to the original PD-10 fraction (following metal chelate chromatography) of MpPAL 2 and Pp4CL 5. Hatched areas belong to different experiments.

3.4 Photometrical characterization of enzyme activities and comparative experiments

3.4.1 Characterization of PAL from *Chara braunii*

3.4.1.1 Characterization of CbPAL 1

CbPAL 1 was photometrically characterized according to 2.5.1.2 applying the substrates L- and D-Phe, as well as L-Tyr and L-His. The only tested substrate, for which CbPAL 1 showed acceptance was L-Phe, while not being susceptible for D-Phe, L-Tyr, L-His or L-DOPA. Experiments with L-His were conducted applying 100 µM $MnCl_2$ and 5 mM DTT/glutathione. The measurement of L-Phe revealed a rather high affinity (low K_m) but the specific activity was low, resulting in a weak turnover and a scant catalytic efficiency of slightly above 500 s^{-1} mol^{-1} l. Raw data were used to establish Michaelis-Menten and Hanes-Woolf kinetics in order to determine K_m and V_{max} and calculate k_{cat} as well as k_{cat}/K_m (catalytic efficiency). The corresponding graph is shown in Figure 56 and the kinetic data of the measurement with L-Phe are displayed in Table 14.

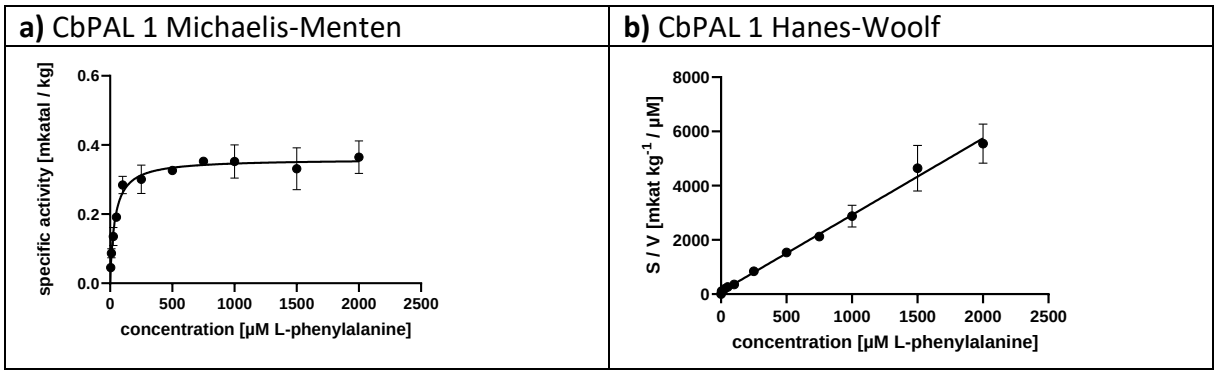


Figure 56 Reaction of CbPAL 1 with L-Phe displayed in Michaelis-Menten (a) and Hanes-Woolf (b) diagrams. Graphs represent the mean values of one biological replicate with four technical replicates. Error bars represent the standard error.

Table 14 Kinetic values for CbPAL 1 with L-Phe determined via Michaelis-Menten (MM) respectively Hanes-Woolf (HW) methods. Values were calculated, directly consulting data from Michaelis-Menten/Hanes-Woolf plots.

Enzyme	Substrate	K_m [µM]	V_{max} [mkat kg^{-1}]	k_{cat} [s^{-1}]	k_{cat}/K_m [s^{-1} mol^{-1} l]
CbPAL 1	L-Phe (MM)	38.41 ± 5.08	0.36 ± 0.02	0.02	520.70
	L-Phe (HW)	35.68 ± 13.06	0.36 ± 0.03	0.02	560.54

3.4.1.2 Characterization of CbPAL 2

Both varieties of CbPAL 2 (long and short version) were tested with L-/D-Phe, L-Tyr, L-His and L-DOPA, but no reactivity could be detected. For CbPAL 2s, however, a successful Western blot proved the expression of the protein.

3.4.1.3 Comparison of CbPAL 1 and 2

An alignment was constructed using CLC Sequence Viewer for CbPAL 1 and both the short and the long version of CbPAL 2, showing that CbPAL 1 was in terms of sequence length located between CbPAL 2s and CbPAL 2l. The catalytic motif ASG was found to be present in all three sequences. Furthermore, the enzymes were examined in terms of their identities/similarities on amino acid level using EMBOSS Needle and revealed comparable identities of slightly under 30 % of CbPAL 1 towards both versions of CbPAL 2 and similarities of approximately 47 to 50 %. CbPAL 2s and 2l were naturally highly identical/similar to one another. In Table 15, the identity/similarity comparison is depicted, the alignment of the three CbPALs is shown in

Figure 57. For all three isoforms/versions of CbPAL it is conspicuous that identities and similarities to AaPAL 1 (having been used as scaffold sequence for the database searches) are especially low, with similarities in all cases less than 40 % and identities ranging only at about 20 %.

Table 15 Identities/similarities [%] of the PAL isoforms derived from *Chara braunii* on amino acid level. Calculations were conducted using EMBOSS Needle.

	CbPAL 1	**CbPAL 2s**	**CbPAL 2l**
CbPAL 1	100.0/100.0	29.5/50.3	28.2/47.0
CbPAL 2s		100.0/100.0	89.0/89.0
CbPAL 2l			100.0/100.0

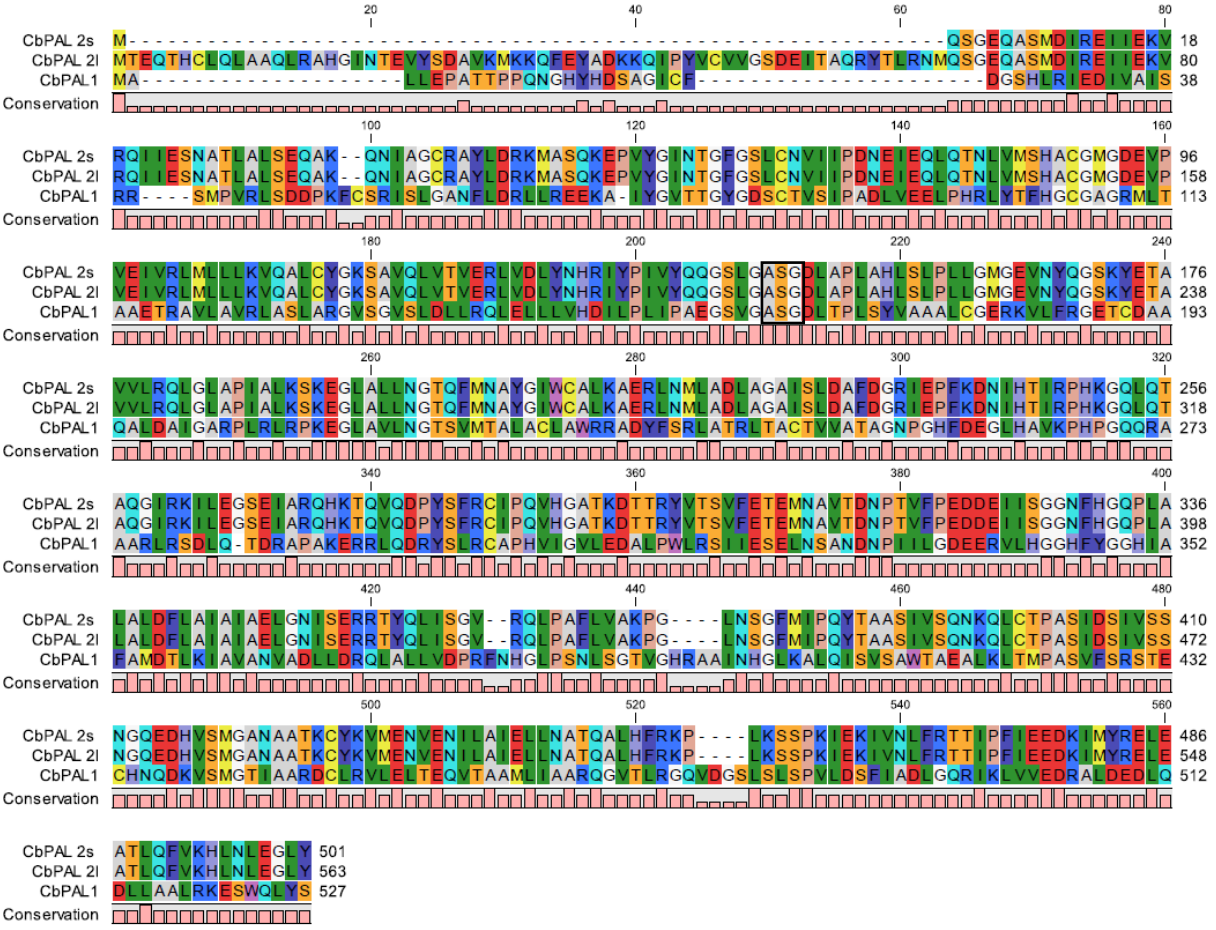


Figure 57 Alignment of CbPAL 1, CbPAL 2s and CbPAL 2l. The amino acid residues ASG most likely responsible for the formation of the MIO group are framed by a black box. The alignment was performed with CLC Sequence Viewer version 8.0 (www.qiagenbioinformatics.com).

3.4.2 Characterization of PAL from *Marchantia polymorpha*

3.4.2.1 Characterization of MpPALs 1 to 3

MpPALs 1 to 3 were photometrically characterized according to 2.5.1.2 applying the substrates L- and D-Phe as well as L-Tyr and L-His. Implementing raw data derived from the measurements, Michaelis-Menten and Hanes-Woolf kinetics were constructed in order to directly determine K_m and V_{max} and then calculate k_{cat} as well as k_{cat}/K_m. Isoforms 1 and 2 were only susceptible towards L-Phe, while MpPAL 3 was not only acceptant of D-Phe and L-Tyr, but also showed a significantly higher V_{max} for L-Phe. The unusually high turnover resulted in by far the highest catalytic efficiency for MpPAL 3 among all PAL isoforms. The susceptibility of MpPAL 3 for D-Phe and L-Tyr was far lower than for L-Phe. For L-His, there was no reactivity with any of the isoforms. Reactivity tests were conducted using 100 µM $MnCl_2$ and 5mM DTT/glutathione as cofactors. Experiments with L-DOPA also yielded no positive results for MpPALs 1 to 3. The oxidation reaction of L-DOPA made the measurement difficult and required protective measures to be taken.

All graphs resulting from Michaelis-Menten and Hanes-Woolf kinetics of MpPAL 1 to 3 tested with L-Phe are shown in Figure 58, while the graphs for MpPAL 3 resulting from experiments with D-Phe and L-Tyr are shown in Figure 59. All kinetic parameters are displayed in Table 16.

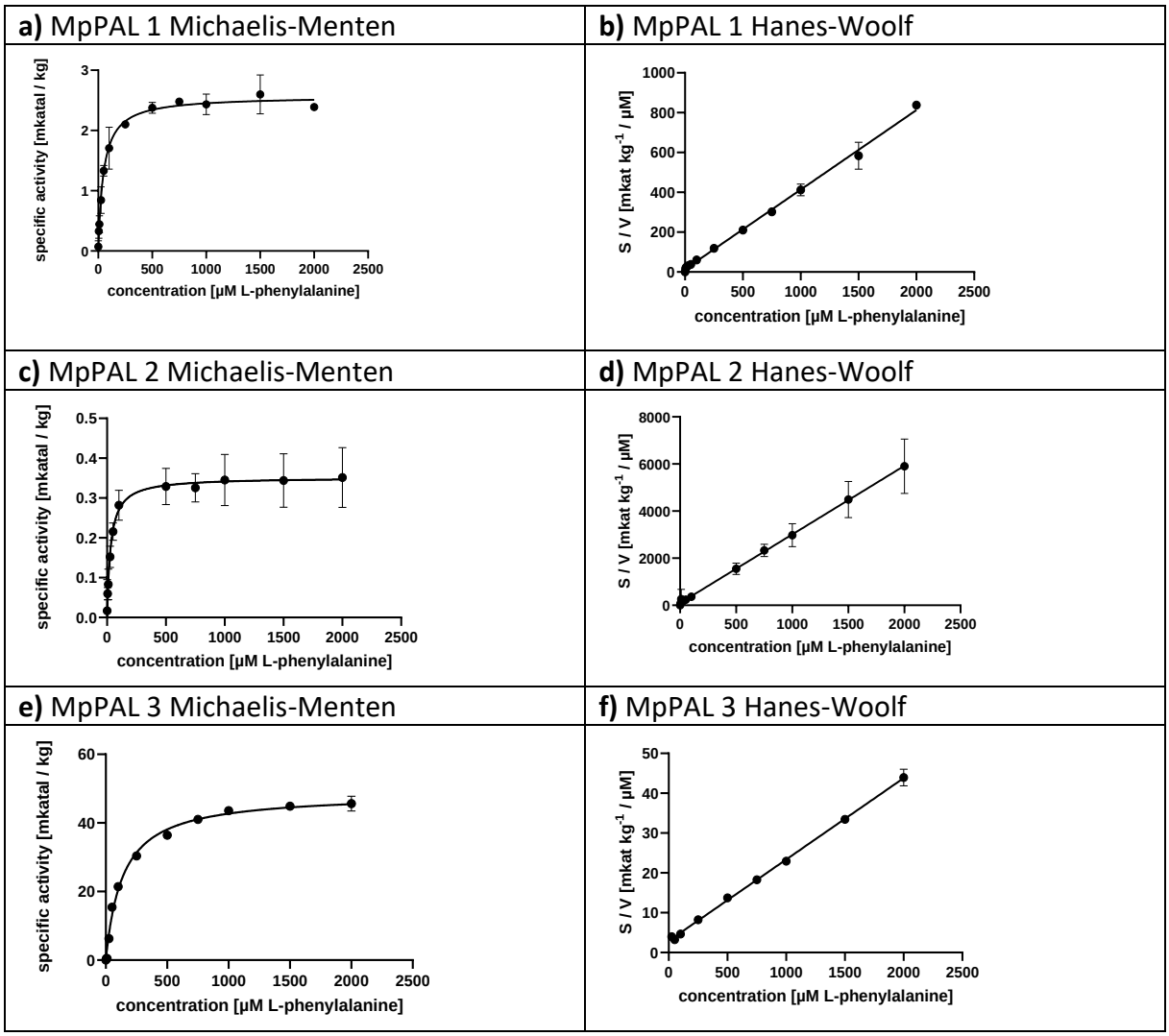


Figure 58 Reaction of MpPAL 1 with L-Phe displayed in Michaelis-Menten (a) and Hanes-Woolf (b) diagrams. Graphs represent the mean values of one biological replicate with four technical replicates. Error bars represent the standard error. Reaction of MpPAL 2 with L-Phe displayed in Michaelis-Menten (c) and Hanes-Woolf (d) diagrams. Graphs represent the mean values of one biological replicate with eight technical replicates. Error bars represent the standard error. Reaction of MpPAL 3 with L-Phe displayed in Michaelis-Menten (e) and Hanes-Woolf (f) diagrams. The graph represents the mean values of one biological replicate with six technical replicates. Error bars represent the standard error.

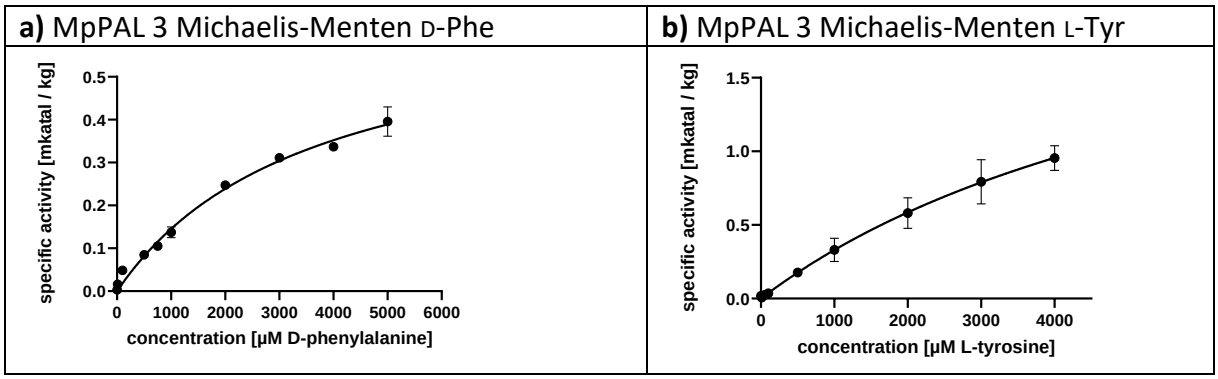


Figure 59 a) Reaction of MpPAL 3 with D-Phe (no saturation attained, therefore determination of kinetic parameters via Cornish-Bowden plot). The graph represents the mean values of one biological replicate with six technical replicates. Error bars represent the standard error. b) Reaction of MpPAL 3 with L-Tyr (no saturation attained, therefore determination of kinetic parameters via Cornish-Bowden plot, as described in Figures 11 and 12). The graph represents the mean values of one biological replicate with six technical replicates. Error bars represent the standard error.

Table 16 Kinetic values for MpPALs 1 to 3 with L-Phe, D-Phe and L-Tyr, corresponding to the graphs shown in Figures 58 and 59. Values with standard errors were calculated directly using the data from the Michaelis-Menten (MM)/Hanes-Woolf (HW) plots. Values without standard errors were determined with the help of the Cornish-Bowden method (see 2.5.2.2).

Enzyme	Substrate	K_m [µM]	V_{max} [mkat kg^{-1}]	k_{cat} [s^{-1}]	k_{cat}/K_m [s^{-1} mol^{-1} l]
MpPAL 1	L-Phe (MM)	50.39 ± 8.30	2.57 ± 0.05	0.22	4365.95
	L-Phe (HW)	35.66 ± 7.27	2.51 ± 0.05	0.21	5888.95
MpPAL 2	L-Phe (MM)	34.33 ± 7.04	0.36 ± 0.04	0.03	873.87
	L-Phe (HW)	39.08 ± 16.68	0.35 ± 0.02	0.03	767.66
MpPAL 3	L-Phe (MM)	138.30 ± 3.00	48.63 ± 0.45	3.91	28271.87
	L-Phe (HW)	143.13 ± 6.49	48.99 ± 0.73	3.94	27527.42
	D-Phe	2100.0	0.54	0.04	19.05
	L-Tyr	7190.0	2.70	0.22	30.60

3.4.2.2 Characterization of MpPAL 4

MpPAL 4 was tested with L-/D-Phe, L-Tyr and L-His, but no significant reactivity could be ascertained. Because of the absence of adequate proof of successful protein expression by positive Western blot, no definite statement considering reactivity could be made.

3.4.2.3 Comparison of MpPALs 1 to 4

For all four isoforms of MpPAL, an alignment was constructed, revealing especially high similarity between isoforms 1 and 2 and moderate similarity of the latter two towards MpPAL 3. MpPAL 4 showed great incongruity towards the other three isoforms. The catalytical ASG motif was present in all four MpPALs. The amino acid sequences of MpPALs 1 to 4 underwent comparative analysis with the help of EMBOSS Needle in order to determine identities/similarities on amino acid level. The juxtaposition revealed the highest percentage of identities for MpPAL 1 and 2, followed by MpPAL 3 (being slightly more identical to MpPAL 2 than MpPAL 1). In the case of MpPAL 4, the identities were in all cases only about 30 % (similarities about 45 %), making it highly dissimilar. The alignment is shown in Figure 60 and identities/similarities are shown in Table 17.

Table 17 Identities/similarities [%] in the amino acid sequences between characterized PALs derived from *Marchantia polymorpha* (applied program: EMBOSS Needle).

	MpPAL 1	**MpPAL 2**	**MpPAL 3**	**MpPAL 4**
MpPAL 1	100.0/100.0	88.3/93.6	73.1/81.2	32.6/44.8
MpPAL 2		100.0/100.0	73.8/82.5	33.3/45.4
MpPAL 3			100.0/100.0	33.1/46.5
MpPAL 4				100.0/100.0

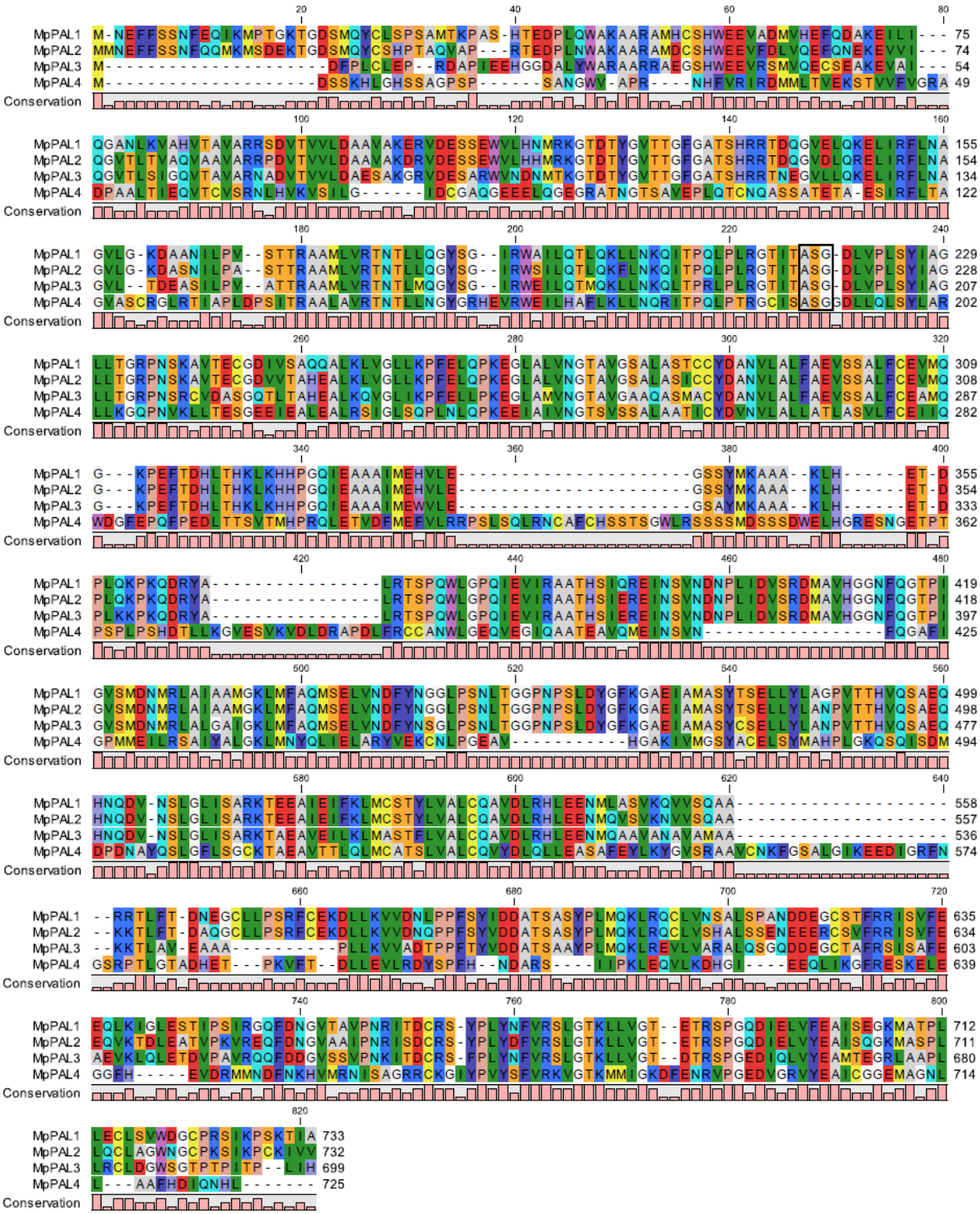


Figure 60 Alignments of PALs 1 to 4 from *M. polymorpha*. The ASG motif putatively responsible for the formation of the MIO group is marked by a black box. The alignment was performed with CLC Sequence Viewer version 8.0 (www.qiagenbioinformatics.com).

3.4.2.4 Temperature optimum of MpPAL 3

For MpPAL 3 as representative of all MpPALs, a temperature optimum was measured, in order to evaluate the influence of incubation temperature on the enzyme activities. For this purpose, 3 µg enzyme per cuvette were applied and temperatures ranging from 20 to 70 °C were measured with L-Phe as substrate in a PAL assay as described in 2.5.1.2. The incubation time was 180 s and a substrate concentration of 2 mM was chosen. The resulting temperature optimum was based on the evaluation of four technical replicates and showed that the highest specific activity is reached at 55 °C, after which a rapid decrease in specific activity can be observed. The corresponding graph is shown in Figure 61.

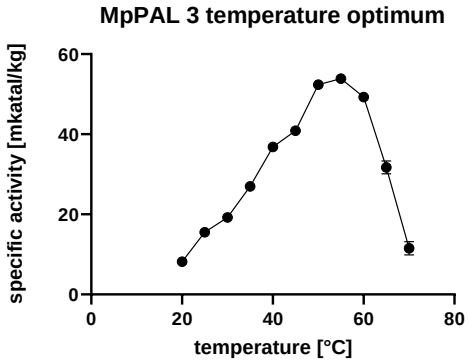


Figure 61 Temperature optimum of MpPAL 3 with L-Phe. Temperatures of 20 to 70 °C were examined. Error bars represent the standard error for four technical replicates.

3.4.3 Characterization of PAL from *Physcomitrium patens*

3.4.3.1 Characterization of PpPALs 1 to 4

PpPALs were photometrically characterized according to 2.5.1.2. The applied substrates were L- and D-Phe as well as L-Tyr and L-His. Raw data derived from the measurements were used to establish Michaelis-Menten and Hanes-Woolf kinetics in order to determine K_m and V_{max} and calculate k_{cat} as well as k_{cat}/K_m. For the respective measurements, between three and six technical replicates were compiled. For all four isoforms, it could be shown that L-Phe was by far the best accepted substrate, while D-Phe and L-Tyr were transformed to a significantly lower degree (the catalytic efficiencies for these two substrates varied between 1 and 22 s^{-1} mol^{-1} l). For L-Phe, it was shown that PpPAL 1 had the highest affinity (lowest K_m) and PpPAL 2 showed the strongest turnover yet the weakest affinity. The highest catalytic efficiency was also attained with PpPAL 1. All measurements of PpPALs 1 to 4 with L-Phe are displayed in Figure 62, while experiments with L-Tyr are shown in Figures 65 and experiments with D-Phe are shown in Figures 63 and 64. All kinetic data derived from the measurements with PpPALs are depicted in Table 18.

For L-His, no susceptibility could be deduced for any of the four PpPALs. The assay was conducted using 5 mM DTT (dithiothreitol) and 100 µM $MnCl_2$ (Klee 1972; Brand and Harper 1976). As an alternative to DTT, 5 mM glutathion was applied.

L-DOPA also showed no apparent transformation and, due to its instability, the photometric measurement was difficult, as L-DOPA quickly oxidized and the solution turned brown. Protective measures against oxidation were taken by wrapping the reaction vessels in aluminium foil to keep out the light, and the measurement was conducted as quickly as possible.

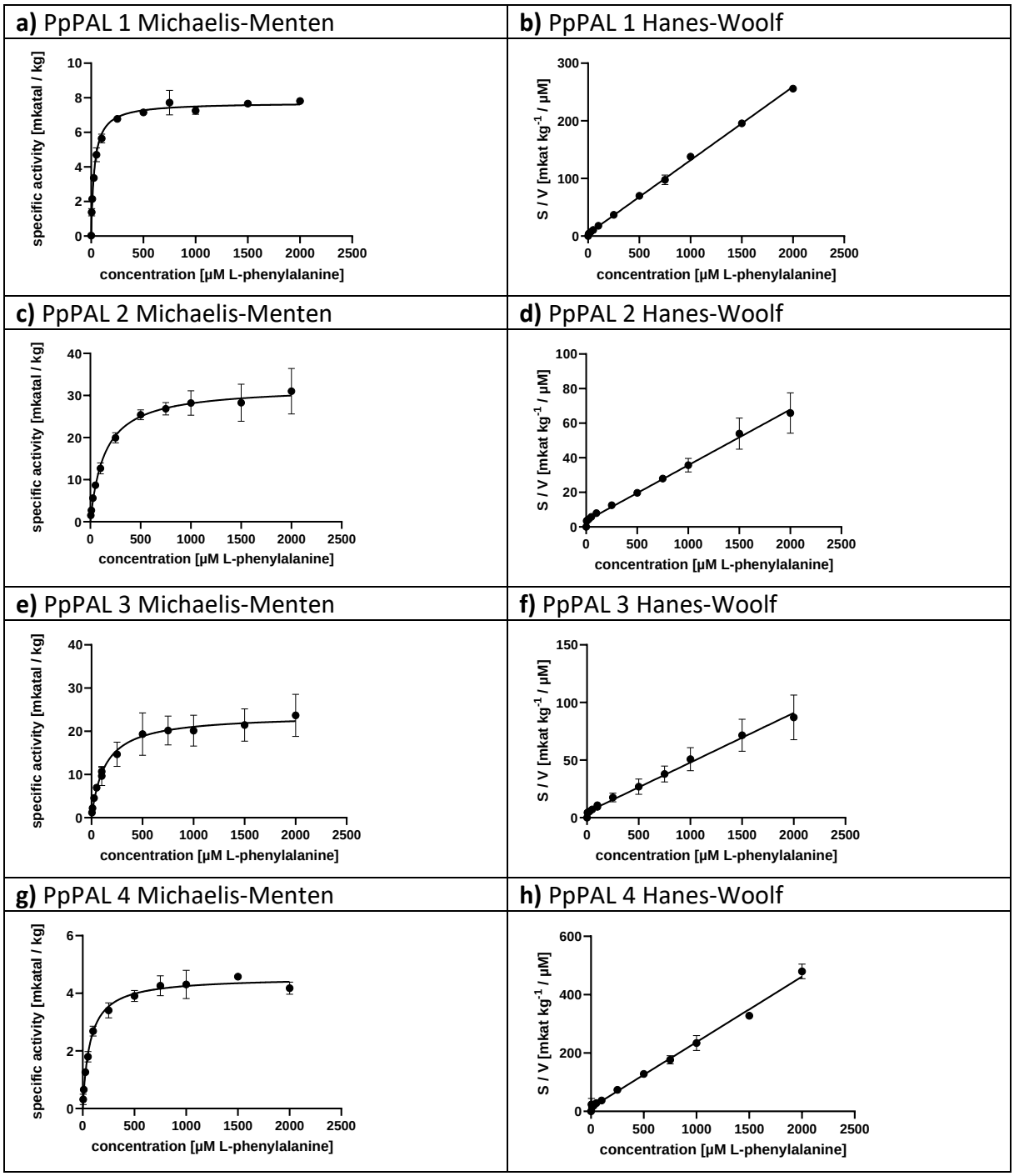


Figure 62 Reactions with L-Phe displayed in Michaelis-Menten (left) and Hanes-Woolf (right) diagrams for PpPAL 1 (a, b), PpPAL 2 (c, d), PpPAL 3 (e, f) and PpPAL 4 (g, h). Graphs represent the mean values of one biological replicate with four technical replicates for PpPAL 1 and three technical replicates for PpPALs 2, 3 and 4. Error bars represent the standard error.

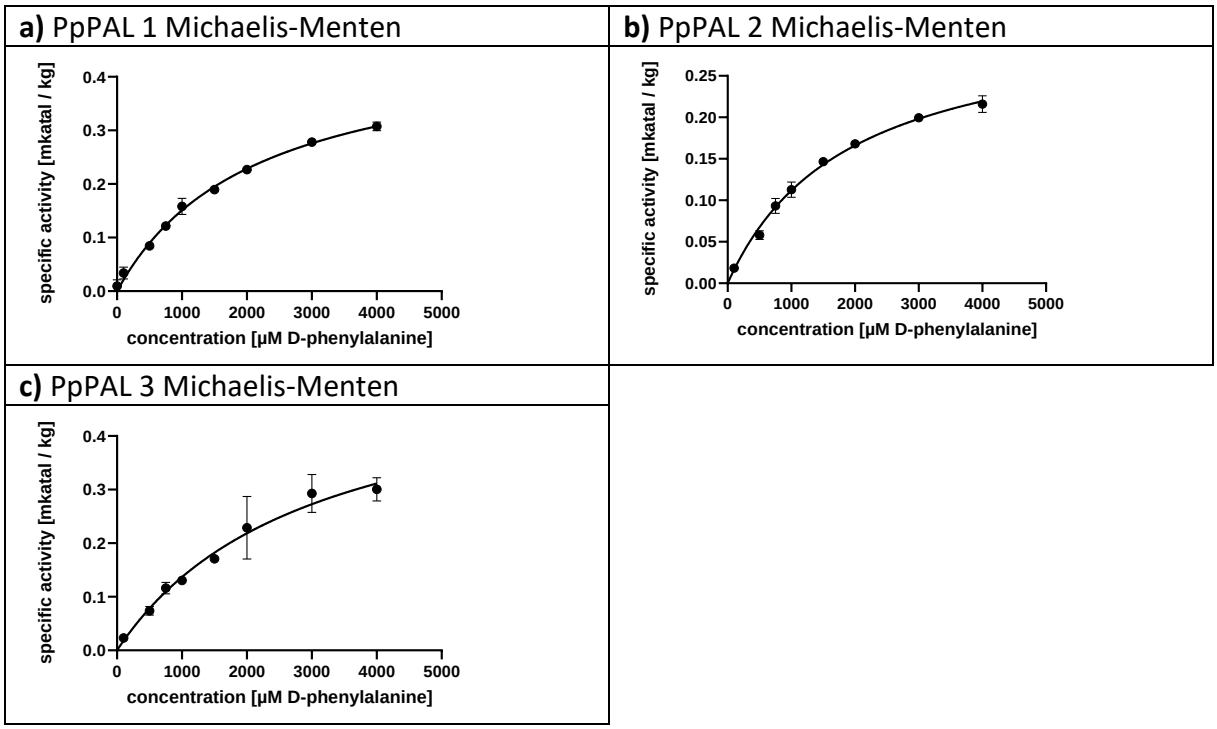


Figure 63 Reactions of PpPALs 1 (a), PpPAL 2 (b) and PpPAL 3 (c) with D-Phe (no saturation attained, therefore determination of kinetic parameters via Cornish-Bowden plot).

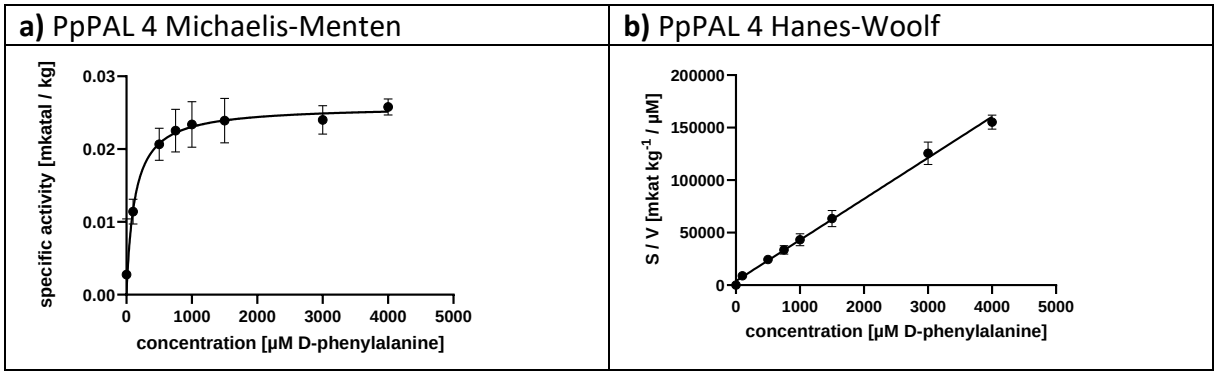


Figure 64 Reaction of PpPAL 4 with D-Phe displayed in Michaelis-Menten (a) and Hanes-Woolf (b) diagrams. Graphs represent the mean values of one biological replicate with four technical replicates. Error bars represent the standard error.

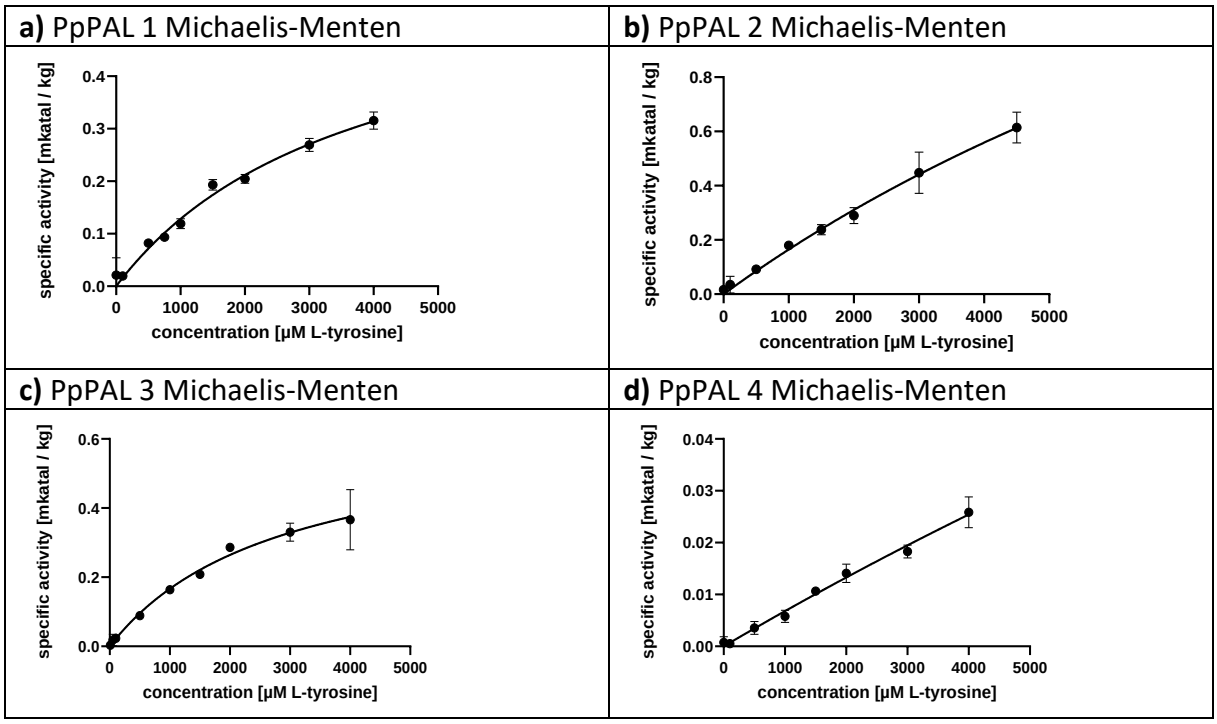


Figure 65 Reactions of PpPAL 1 (a), PpPAL 2 (b), PpPAL 3 (c) and PpPAL 4 (d) with L-Tyr (no saturation attained, therefore determination of kinetic parameters via Cornish-Bowden plot). Graphs represent the mean values of one biological replicate with four technical replicates for PpPAL 1, three technical replicates for PpPAL 2 and PpPAL 4 and six technical replicates for PpPAL 3. Error bars represent the standard error.

Table 18 Kinetic values for PpPALs 1 to 4 with L-Phe, D-Phe and L-Tyr, corresponding to the graphs shown in Figures 62 to 65. Values with standard deviations/errors were calculated directly using the data from the Michaelis-Menten (MM)/Hanes-Woolf (HW) plots. Values without standard deviations were determined via Cornish-Bowden.

Enzyme	Substrate	K_m [µM]	V_{max} [mkat kg^{-1}]	k_{cat} [s^{-1}]	k_{cat}/K_m [s^{-1} mol^{-1} l]
PpPAL 1	L-Phe (MM)	31.19 ± 0.95	7.73 ± 0.07	0.66	21160.63
	L-Phe (HW)	31.66 ± 1.95	7.84 ± 0.02	0.67	21162.35
	D-Phe	2300.0	0.47	0.04	18.50
	L-Tyr	3200.0	0.54	0.05	15.63
PpPAL 2	L-Phe (MM)	145.23 ± 20.71	32.28 ± 2.55	2.73	18797.77
	L-Phe (HW)	112.81 ± 23.74	31.63 ± 2.92	2.68	23756.76
	D-Phe	2100.0	0.34	0.03	15.76
	L-Tyr	3180.0	0.80	0.07	22.01
PpPAL 3	L-Phe (MM)	131.17 ± 11.12	23.84 ± 2.70	2.02	15399.86
	L-Phe (HW)	114.57 ± 14.31	23.89 ± 2.66	2.03	17718.43
	D-Phe	3050.0	0.53	0.04	13.11
	L-Tyr	2680.0	0.61	0.05	18.66
PpPAL 4	L-Phe (MM)	71.86 ± 5.76	4.56 ± 0.17	0.39	5427.22
	L-Phe (HW)	61.98 ± 11.25	4.48 ± 0.14	0.38	6131.01
	D-Phe (MM)	126.58 ± 15.26	0.03 ± 0.001	0.002	15.80
	D-Phe (HW)	97.80 ± 18.33	0.03 ± 0.001	0.002	20.44
	L-Tyr	2300.0	0.03	0.003	1.30

3.4.3.2 Characterization of PpPAL 5

PpPAL 5 was investigated in terms of substrate susceptibility with L- and D-Phe as well as L-Tyr and L-His. There was no perceptible turnover for any of the substrates detectable, despite the fact that PpPAL 5 was apparently expressed due to the Western blot yielding a positive result.

3.4.3.3 Comparison of PpPALs 1 to 5

For all five PpPAL isoforms, an alignment was constructed showing high consensus between isoforms 2, 3 and 4, sharing more than 95 % identity. PpPAL 1 was considerably different in terms of sequence identity yet comparable in terms of sequence length, while PpPAL 5 was the most different in terms of both sequence identity and length, the sequence being about 20 amino acids longer at the beginning. The conserved ASG motif was present in all five isoforms at the end of the first third of the amino acid sequence. In order to provide a statement of amino acid sequence identities and similarities, calculations in % were performed with the five PpPAL isoforms. This analysis revealed striking identities and similarities between PpPALs 2, 3 and 4 of > 97 %. Significantly less congruence was found for PpPAL 1 with PpPALs 2, 3 and 4 and PpPAL 5 was the least identical/similar to the other four isoforms with identities of around 65 % and similarities slightly under 80 %. It can, however, be concluded that the five isoforms of PpPAL exhibit rather high similarities amongst each other. The alignment is depicted in Figure 66 and the identity/similarity test is shown in Table 19.

Table 19 Identities/similarities on amino acid level [%] between the five different isoforms of PAL derived from *Physcomitrium patens* (calculations conducted with EMBOSS Needle).

	PpPAL 1	**PpPAL 2**	**PpPAL 3**	**PpPAL 4**	**PpPAL 5**
PpPAL 1	100.0/100.0	72.7/84.2	72.8/84.2	72.5/84.0	67.9/78.2
PpPAL 2		100.0/100.0	99.3/99.6	97.4/98.1	65.5/78.4
PpPAL 3			100.0/100.0	97.2/98.1	65.6/77.9
PpPAL 4				100.0/100.0	64.9/77.9
PpPAL 5					100.0/100.0

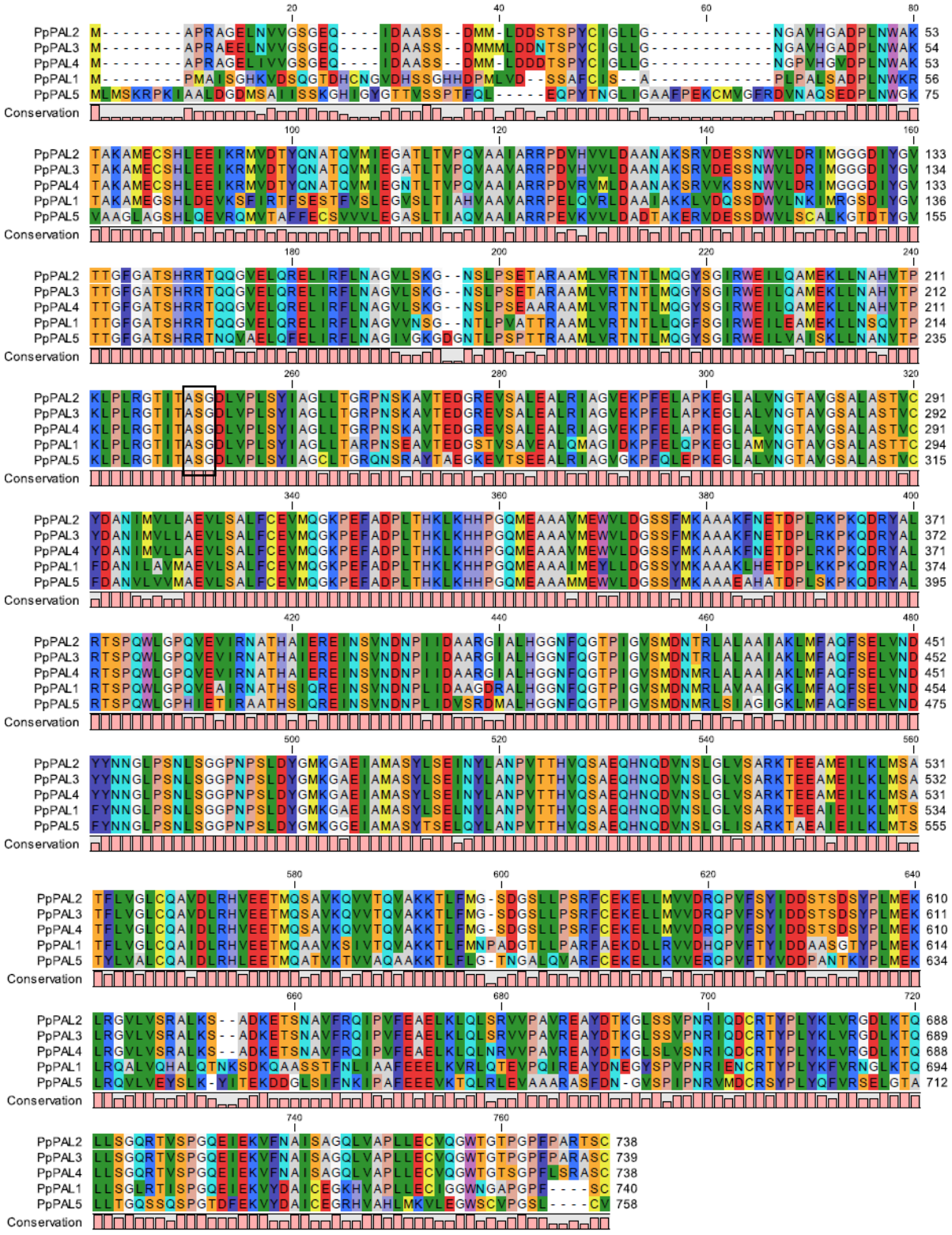


Figure 66 Alignment of the five PpPAL isoforms performed with CLC Sequence Viewer version 8.0 (www.qiagenbioinformatics.com). The ASG motif putatively responsible for the formation of the catalytic MIO group is marked by a black box.

3.4.3.4 Temperature optimum of PpPAL 2

As reference for the influence of temperature on enzyme activities, a temperature optimum was experimentally determined with L-Phe as substrate. For this, 10 µg enzyme per cuvette of PpPAL 2 (as representative isoform for PpPAL) were applied in a classic PAL assay according to 2.5.1.2 and incubated for 180 s at each temperature (ranging from 20 to 70 °C in steps of 5 °C). All measurements were conducted with 2 mM substrate concentration and four technical replicates. It could be determined that the temperature at which the highest enzyme activity could be observed, was 60 °C. Reaching higher temperatures, the activity started falling again until, at approximately 70 °C, the efficiency of the PpPAL 2 had significantly decreased. The corresponding graph is shown in Figure 67.

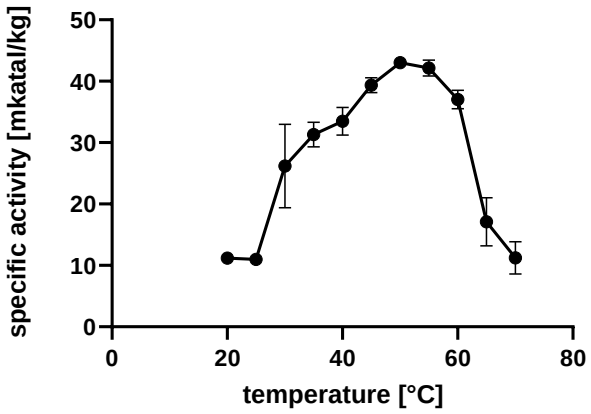


Figure 67 Temperature optimum of PpPAL 2 with L-Phe. Temperatures of 20 to 70 °C were measured. Error bars represent the standard error for four technical replicates.

3.4.4 Characterization of 4CLs 1 and 2 from *Chara braunii*

3.4.4.1 Enzymatic characterization of Cb4CLs 1 and 2

For both isoforms of 4CL derived from *C. braunii*, measured at a concentration of 400 µM with 4-coumaric, caffeic, ferulic, cinnamic, sinapic and 4-hydroxybenzoic acids, the specific activities ranged from insignificant to moderate, but the construction of kinetics was not possible with any of the tested substrates. The highest specific activity could be attained with caffeic acid for both isoforms (appr. 0.6 mkat kg^{-1}), the second highest was reached with 4-hydroxybenzoic acid (appr. 0.35 mkat kg^{-1} for Cb4CL 1 and appr. 0.2 mkat kg^{-1} for Cb4CL 2). The specific activities for the other substrates, including 4-coumaric acid, ranged between 0.1 and 0.2 mkat kg^{-1} for Cb4CL 1 and for Cb4CL 2, only the experiment with 4-coumaric acid surpassed 0.1 mkat kg^{-1}. All specific activities for both isoforms of Cb4CL are depicted in form of a bar chart in Figure 68.

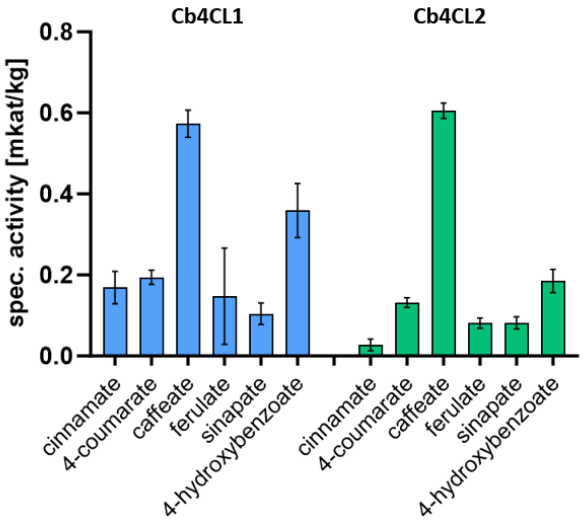


Figure 68 Activities of Cb4CLs 1 and 2 with cinnamic, 4-coumaric, caffeic, ferulic, sinapic and 4-hydroxybenzoic acids at a substrate concentration of 400 µM. Error bars represent the standard deviation of 6 replicates (Schwarze and Petersen 2024).

3.4.4.2 Comparison of Cb4CLs 1 and 2

For both isoforms of Cb4CL, an alignment and an analysis in terms of amino acid sequence identity/similarity (using EMBOSS Needle) was conducted. The alignment showed that both sequences were of comparable length and did not contain either a complete box I or box II motif (Cukovic et al. 2001; Stuible and Kombrink 2001; Shockey et al. 2003; Shockey and

Browse 2011). The GEICIRG motif was disrupted in Cb4CL 2 and missing in Cb4CL 1. The amino acid residues said to be responsible for substrate specificity (Schneider et al. 2003) were also not the genuine ones. Sequence identity was with 23.4 % extremely low and also the similarity level of 35.8 % did not imply a lot of congruence between the two isoforms. The respective alignment is shown in Figure 69.

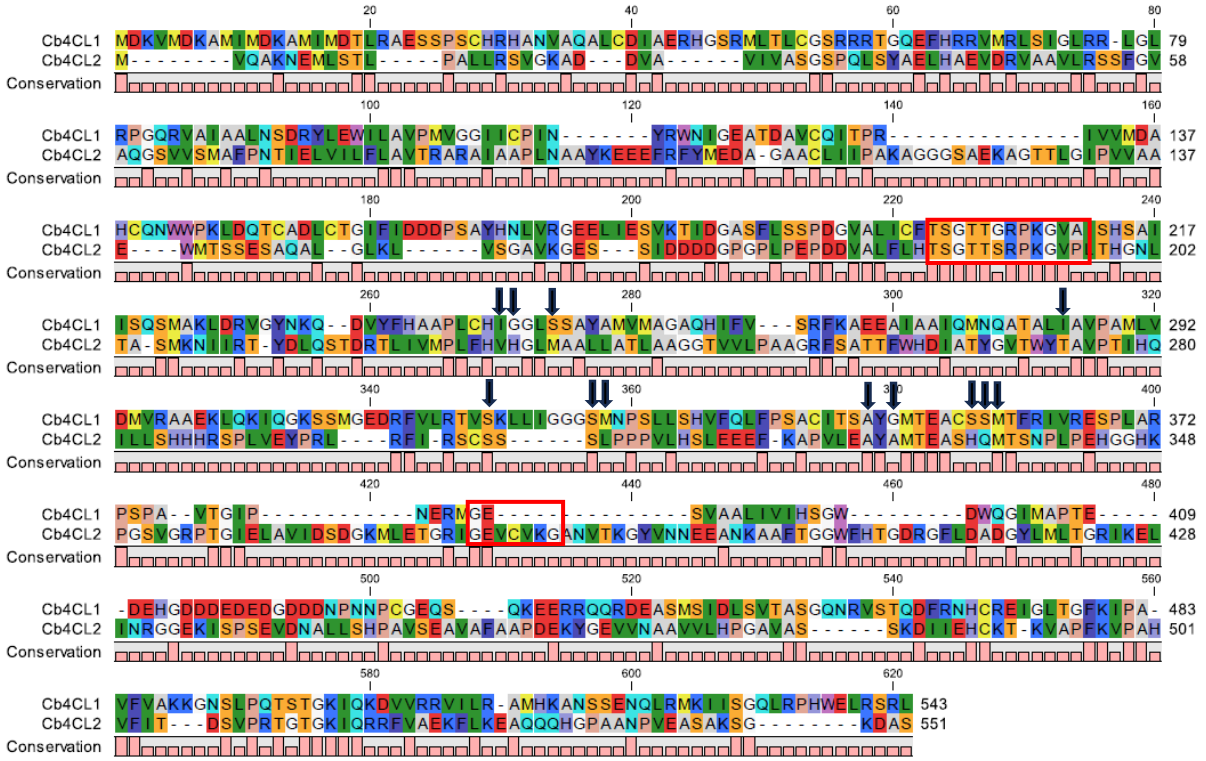


Figure 69 Alignment of 4CLs 1 and 2. from *Chara braunii*. Box I (AMP-binding domain) and box II motif (4CL functionality domain) are framed by red boxes. Amino acid residues supposedly responsible for substrate specificity (Schneider et al. 2003) are marked by black arrows. The alignment was performed with CLC Sequence Viewer version 8.0 (www.qiagenbioinformatics.com).

3.4.5 Characterization of 4CLs 1 to 4 from *Marchantia polymorpha*

3.4.5.1 Enzymatic characterization of Mp4CLs 1 to 4

Mp4CLs 1 to 4 were photometrically characterized according to 2.5.1.4. The investigated substrates were 4-coumaric, caffeic, ferulic, cinnamic, sinapic and 4-hydroxybenzoic acids. Isoforms 1, 2 and 4 proved to be too low in terms of activity to adequately construct Michaelis-Menten kinetics. Mp4CL 3 was extraordinarily active and showed by far the highest catalytic efficiencies out of all the investigated enzymes, both PALs and 4CLs. For 4-coumaric acid, the catalytic efficiency exceeded 2.5 million s^{-1} mol^{-1} l (Michaelis-Menten). The turnover rate for 4-coumaric acid (Michaelis-Menten) was also the highest (5.0 s^{-1}) and the K_m was very low (1.96 µM), indicating a high affinity towards this particular substrate. Also, caffeic and ferulic acids were accepted by Mp4CL 3 to a very high degree, with catalytic efficiencies of appr. 300000 and 870000 s^{-1} mol^{-1} l, respectively (Michaelis-Menten). The reaction with cinnamic acid revealed a high turnover but a significantly lower affinity (catalytic efficiency of 38000 s^{-1} mol^{-1} l). For both sinapic and 4-hydroxybenzoic acids, there was no significant reactivity detectable in Mp4CL 3. Mp4CLs 1 and 2 accepted caffeic acid as the best substrate and showed no acceptance towards cinnamic acid. The specific activities for caffeic acid ranged from appr. 0.15 to 0.2 mkat kg^{-1}, whereas all other accepted substrates had specific activities distinctly below 0.5 mkat kg^{-1}. Mp4CL 4 showed the highest acceptance towards 4-coumaric acid (about 0.5 mkat kg^{-1}), followed by cinnamic acid (about 0.3 mkat kg^{-1}). It should be noted that the specific activity for 4-hydroxybenzoic acid for Mp4CL 4 was with appr. 0.2 mkat kg^{-1} at the same level as for caffeic acid with isoforms 1 and 2. It was also the only of the four isoforms that accepted 4-hydroxybenzoic acid to a noticeable degree. Caffeic acid was also accepted by Mp4CL 4 (also about 0.2 mkat kg^{-1}). All specific activities are depicted in form of a bar chart for Mp4CLs 1, 2 and 4 in Figure 71. The substrate saturation curves are shown in Figure 70. The kinetic data derived from the measurements with Mp4CL 3 are depicted in Table 20.

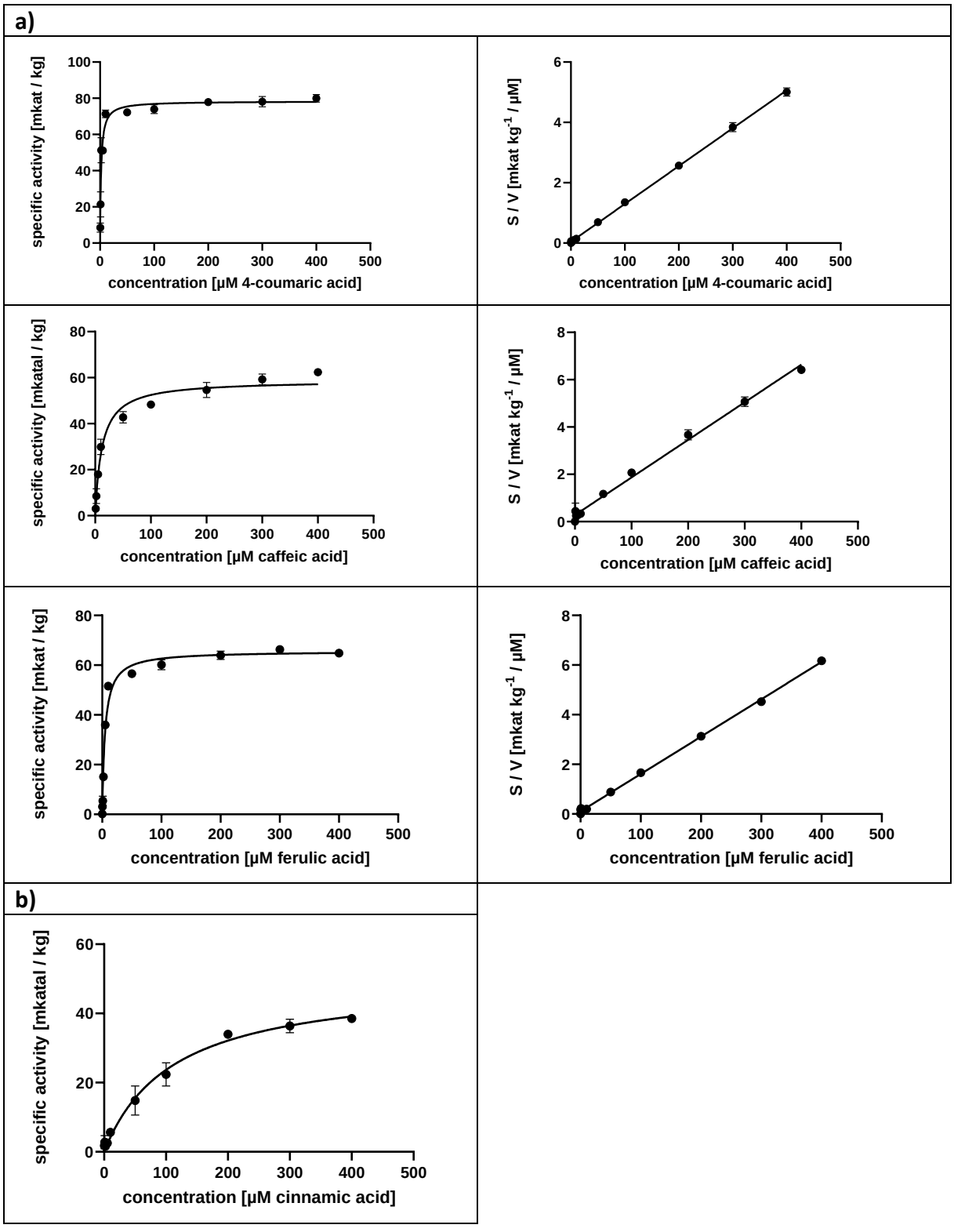


Figure 70 a) Reactions of Mp4CL 3 with 4-coumaric, caffeic and ferulic acids acid displayed in Michaelis-Menten (left) and Hanes-Woolf (right) diagrams. b) Reaction of Mp4CL 3 with cinnamic acid (no saturation attained, therefore determination of kinetic parameters via Cornish-Bowden plot). Graphs represent the mean values of one biological replicate with six technical replicates. Error bars represent the standard error.

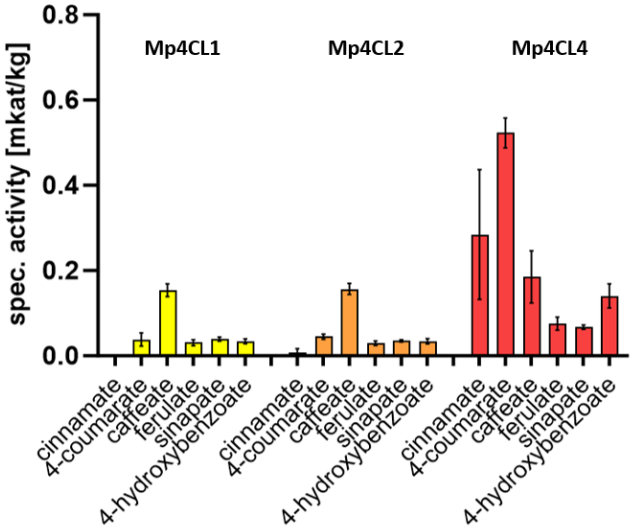


Figure 71 Specific activities Mp4CLs 1, 2 and 4 with cinnamic, 4-coumaric, caffeic, ferulic, sinapic and 4-hydroxybenzoic acids at a substrate concentration of 400 µM with 6 technical replicates shown as bar chart. Error bars represent the standard deviation (Schwarze and Petersen 2024).

Table 20 Kinetic values for Mp4CL 3 with 4-coumaric, caffeic, ferulic and cinnamic acids, the displayed parameters belong to the graphs shown in Figure 70. Values with standard errors were calculated directly using the data from the Michaelis-Menten (MM) and Hanes-Woolf (HW) plots. Values without standard errors were determined with the help of the Cornish-Bowden method.

Enzyme	Substrate	K_m [µM]	V_{max} [mkat kg^{-1}]	k_{cat} [s^{-1}]	k_{cat}/K_m [s^{-1} mol^{-1} l]
Mp4CL 3	4-coumaric acid (MM)	1.96 ± 0.14	78.38 ± 0.69	5.0	2551020.41
	4-coumaric acid (HW)	3.27 ± 0.27	79.72 ± 0.77	5.01	1532110.09
	ferulic acid (MM)	4.84 ± 0.07	65.66 ± 0.39	4.19	865702.48
	ferulic acid (HW)	6.98 ± 0.70	66.48 ± 0.45	4.24	607449.86
	cinnamic acid	75.0	44.60	2.85	38000.00
	caffeic acid (MM)	12.30 ± 0.79	58.75 ± 0.74	3.75	304878.05
	caffeic acid (HW)	17.27 ± 1.52	62.73 ± 0.61	4.00	231615.52

3.4.5.2 Comparison of Mp4CLs 1 to 4

For all four isoforms of Mp4CL, an alignment and an identity/similarity comparison were created. The alignment revealed that isoforms 1 and 3 exhibited an intact AMP-binding domain, and only Mp4CL 3 possessed a pristine GEICIRG motif. The other Mp4CLs displayed mostly intact motifs, with one (in Mp4CLs 2 and 4) respectively two (in Mp4CL 1) amino acids difference (see alignment in Figure 72). Similarities between the different sequences varied between 52 and 61 %, while identities were generally < 40 %. For Mp4CL 3, the amino acids putatively responsible for substrate specificity mostly resembled those conserved in Pp4CLs 1 to 4 (see Figure 79), with the exception of a valine residue replacing a methionine residue (see Figure 72). Mp4CLs 1 and 4 shared 6 mutual amino acid residues, whereas Mp4CL 2 shared 7 with Pp4CLs 1 to 4. The alignment is depicted in Figure 72, while the identity/similarity comparison is shown in Table 21.

Table 21 Identities/similarities [%] for the four isoforms of 4CL derived from *Marchantia polymorpha* calculated with EMBOSS Needle.

	Mp4CL 1	**Mp4CL 2**	**Mp4CL 3**	**Mp4CL 4**
Mp4CL 1	100.0/100.0	35.2/52.8	40.3/59.5	36.6/55.9
Mp4CL 2		100.0/100.0	36.8/55.0	39.6/61.1
Mp4CL 3			100.0/100.0	41.2/59.9
Mp4CL 4				100.0/100.0

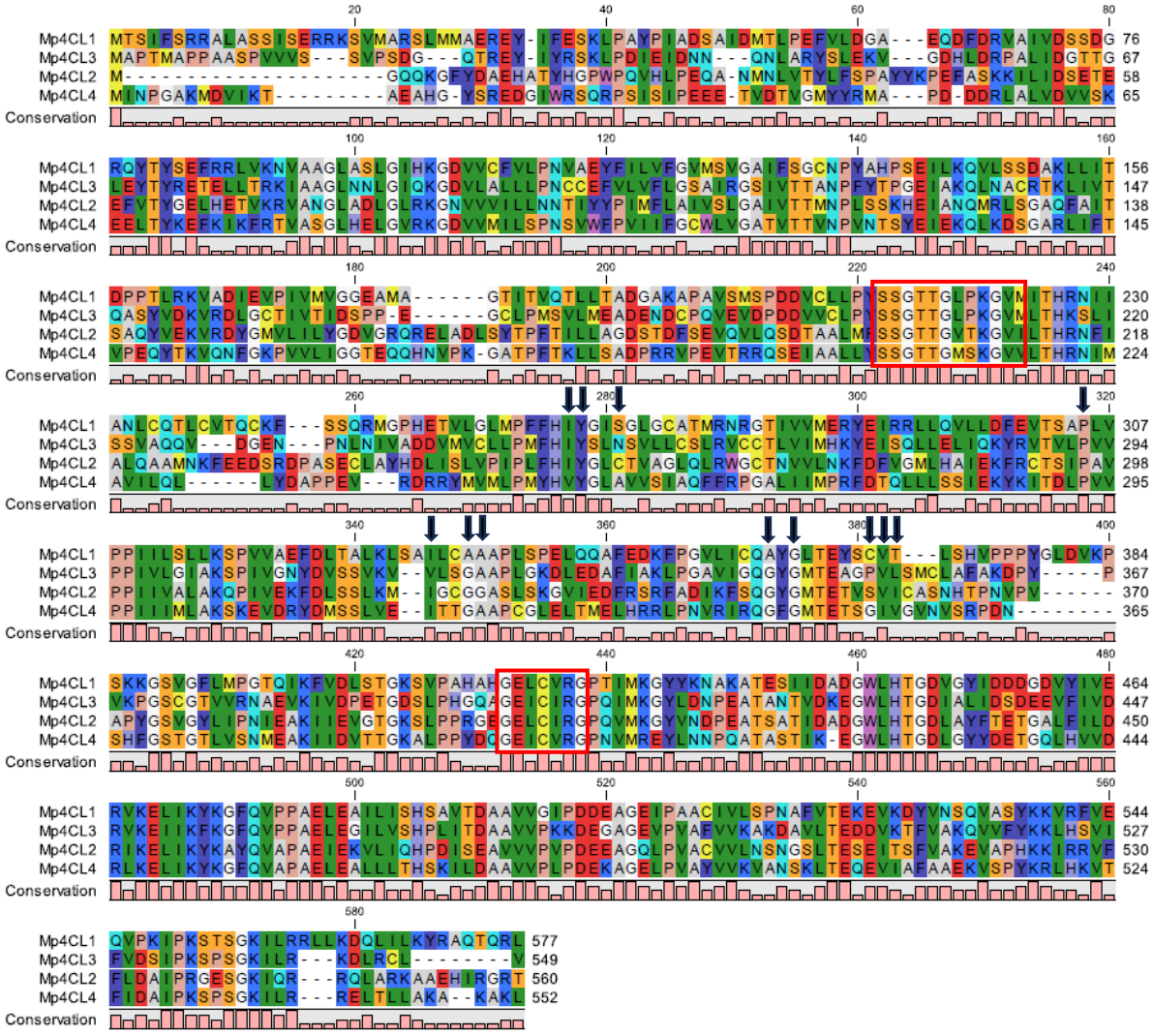


Figure 72 Alignment of amino sequences for 4CL 1 to 4 from *Marchantia polymorpha*. The AMP-binding domain (box I motif) as well as the 4CL functionality domain (box II motif) are marked by red boxes. Amino acid residues putatively responsible for substrate specificity (Schneider et al. 2003) are marked by black arrows. The alignment was performed with CLC Sequence Viewer version 8.0 (www.qiagenbioinformatics.com).

3.4.5.3 Temperature optimum of Mp4CL 3

As a reference of Mp4CL activity under continuous change of temperature, Mp4CL 3 was chosen (it also being the only one of the four expressed isoforms, with which sufficient kinetics could be constructed) and measured photometrically with 4-coumaric acid as described under 2.5.1.4. The evaluated temperatures ranged from 5 to 70 °C (continually increasing the temperature in steps of 5 °C) and the chosen substrate concentration was 400 µM 4-coumaric acid. Four technical replicates were measured. The resulting temperature optimum was determined to be 55 °C, after which a rapid decrease of activity could be observed, probably due to decay of the protein. Up to the respective peak, the activity rose constantly (see Figure 73).

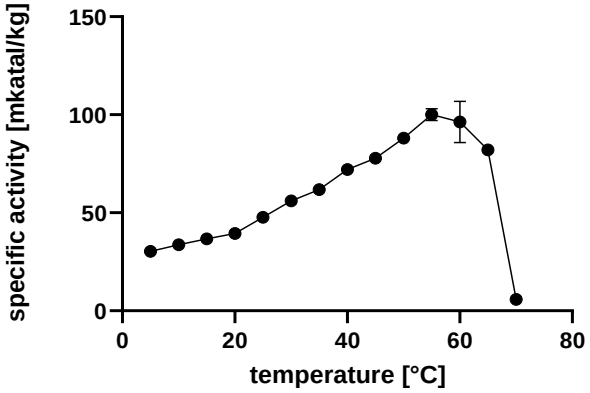


Figure 73 Temperature optimum of Mp4CL 3 with 400 µM 4-coumaric acid (four technical replicates). Error bars represent the standard error.

3.4.5.4 Referential measurement of Mp4CL 3 with CoA and ATP/$MgCl_2$

In order to validate the influence of the co-substrates ATP/$MgCl_2$ and CoA on 4CL measurements, the most efficient isoform, Mp4CL 3, was chosen for comparative measurements with 300 µM 4-coumaric acid as substrate. In the case of ATP (which was used as the main varied substrate in this experiment), the concentration of $MgCl_2$ was varied accordingly. The substrate saturation curves for both measurements are depicted in Figure 74 and the kinetic parameters derived from the Michaelis-Menten method are shown in Table 22.

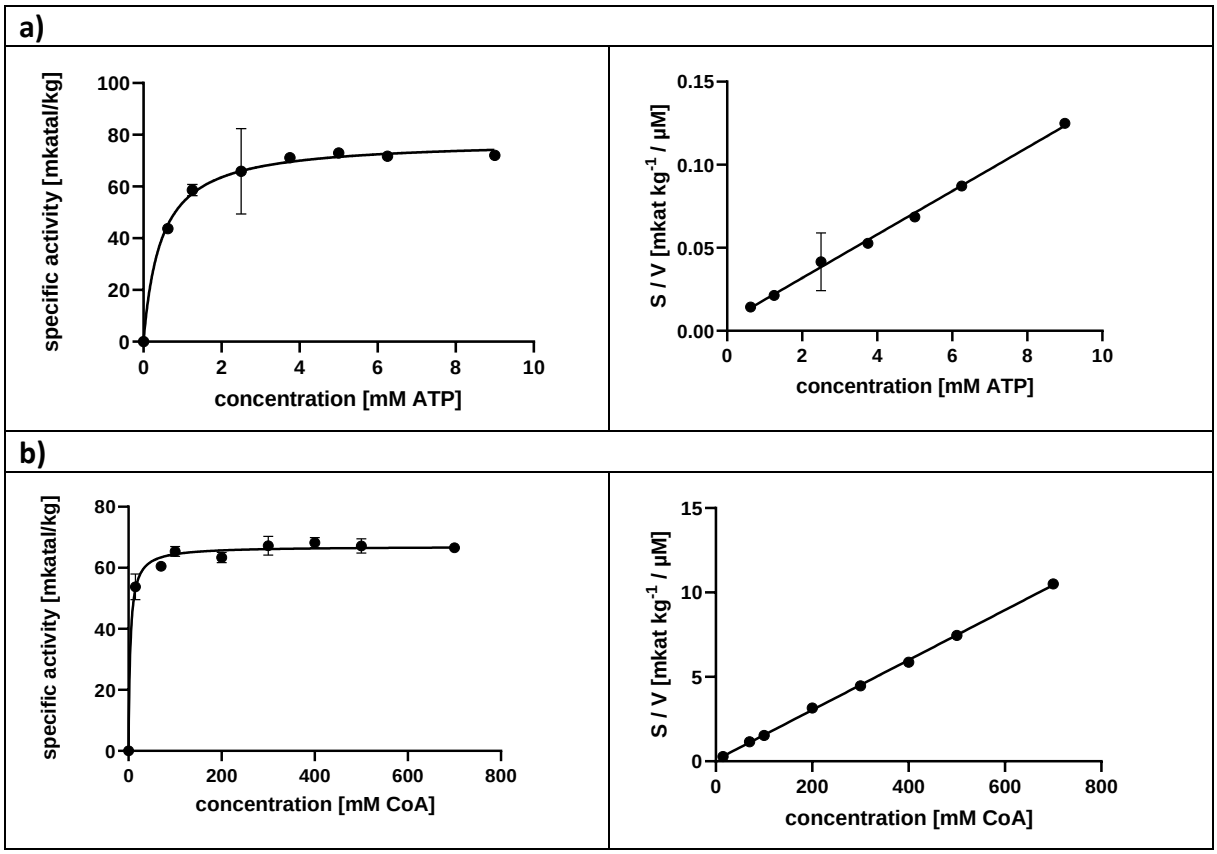


Figure 74 a) Substrate saturation curves of Mp4CL 3 for ATP/$MgCl_2$ with 4-coumaric acid (300 µM) depicted as Michaelis-Menten (left) and Hanes-Woolf (right) diagrams. ATP/$MgCl_2$ concentrations ranged from 0.625 mM to 12.5 mM. b) Substrate saturation curves of Mp4CL 3 for CoA with 4-coumaric acid (300 µM) depicted as Michaelis-Menten (left) and Hanes-Woolf (right) diagrams. CoA concentrations varied from 15 to 700 µM. For both experiments, six technical replicates were measured. Error bars represent the standard error.

Table 22 Kinetic values for Mp4CL 3 with 4-coumaric acid (300 µM) and ATP/$MgCl_2$ respectively CoA as substrates. Values were determined with the help of the Michaelis-Menten method.

Enzyme	Substrate	K_m [µM]	V_{max} [mkat kg^{-1}]	k_{cat} [s^{-1}]	k_{cat}/K_m [s^{-1} mol^{-1} l]
Mp4CL 3	ATP	0.45 ± 0.01	77.93 ± 1.01	4.97	11044444.44
	CoA	4.02 ± 0.51	67.02 ± 0.20	4.28	1064676.62

3.4.6 Characterization of 4CLs 1 to 5 from *Physcomitrium patens*

3.4.6.1 Enzymatic characterization of Pp4CLs 1 to 5

Pp4CLs 1 to 4 were photometrically characterized according to 2.5.1.4. The investigated substrates were 4-coumaric, caffeic, ferulic, cinnamic, sinapic and 4-hydroxybenzoic acids. The substrate concentrations ranged from 0 to 500 µM and were conducted with at least six technical replicates. Michaelis-Menten kinetics were then constructed on the basis of the photometric raw data in order to determine K_m and V_{max} and to calculate k_{cat} and k_{cat}/K_m. Measurements with 4-coumaric and caffeic acids gained the highest catalytic efficiencies, and these two can therefore be considered the most effective of the tested substrates. Measurements with ferulic and cinnamic acids had to be evaluated with the help of Cornish-Bowden plots according to Cornish-Bowden and Eisenthal (1978) because saturation was not reached. There was no significant activity detectable for both, sinapic and 4-hydroxybenzoic acid for any of the isoforms. Pp4CL 4 was the most active of the investigated isoforms, showing the highest catalytic efficiencies, especially for 4-coumaric and caffeic acids. The catalytic efficiencies for ferulic and cinnamic acids were about tenfold lower compared to 4-coumaric and caffeic acids, but still the highest of the investigated isoforms. Pp4CL 3 showed the highest V_{max} of the investigated reactions of Pp4CL with 4-coumaric acid, but the affinity was the lowest, resulting in a moderately high catalytic efficiency. In terms of overall catalytic properties, Pp4CL 1 was the weakest, but the affinities for all substrates were by far the least distinct with Pp4CL 3 (highest K_m values for all substrates). In the case of Pp4CL 5, there was no activity detectable for any of the tested substrates, which were 4-coumaric, caffeic, ferulic, cinnamic, sinapic and 4-hydroxybenzoic acids. All graphs of the corresponding substrate saturation experiments are shown in Figures 75 to 78 and the kinetic parameters for Pp4CL measurements are listed in Table 23.

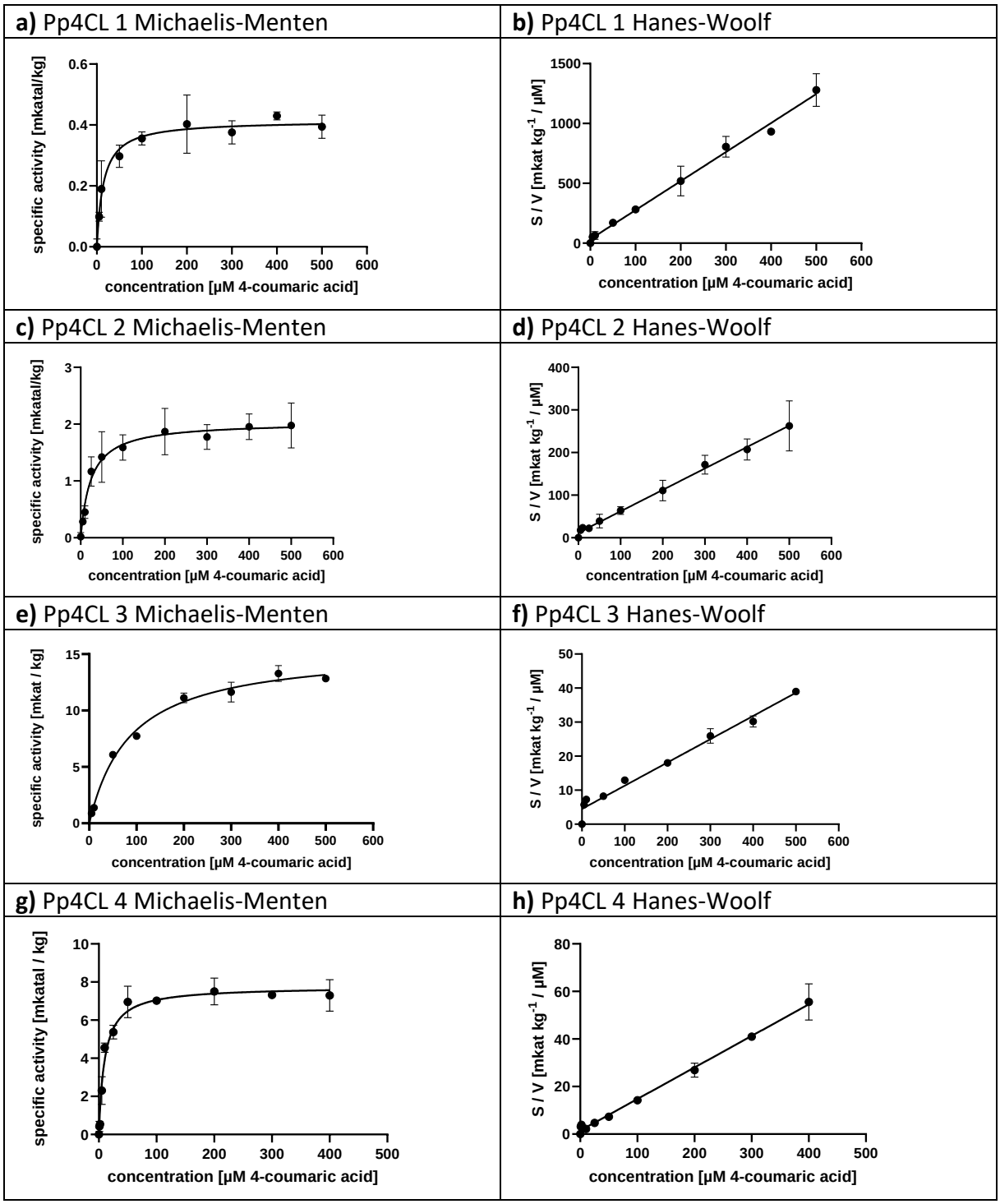


Figure 75 Reactions of Pp4CL 1 (a, b), Pp4CL 2 (c, d), Pp4CL 3 (e, f) and Pp4CL 4 (g, h) with 4-coumaric acid displayed in Michaelis-Menten (left) and Hanes-Woolf (right) diagrams. Graphs represent the mean values of one biological replicate with six technical replicates. Error bars represent the standard error.

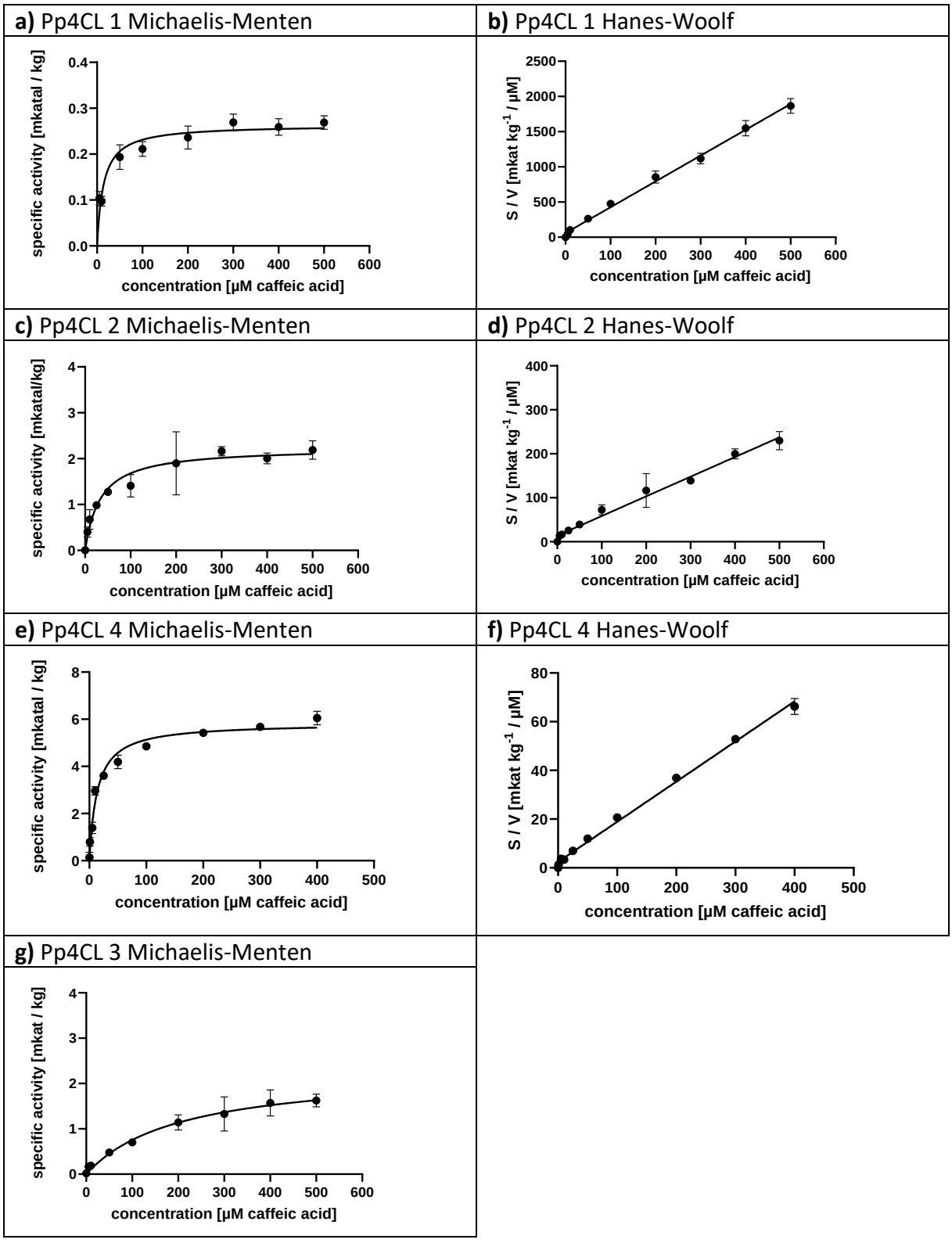


Figure 76 Reactions of Pp4CLs 1 (a, b), Pp4CL 2 (c, d) and Pp4CL 4 (e, f) with caffeic acid displayed in Michaelis-Menten (left) and Hanes-Woolf (right) diagrams. g) Reaction of Pp4CL 3 with caffeic acid (no saturation attained, therefore determination of kinetic parameters via Cornish-Bowden plot). Graphs represent the mean values of one biological replicate with six technical replicates. Error bars represent the standard error.

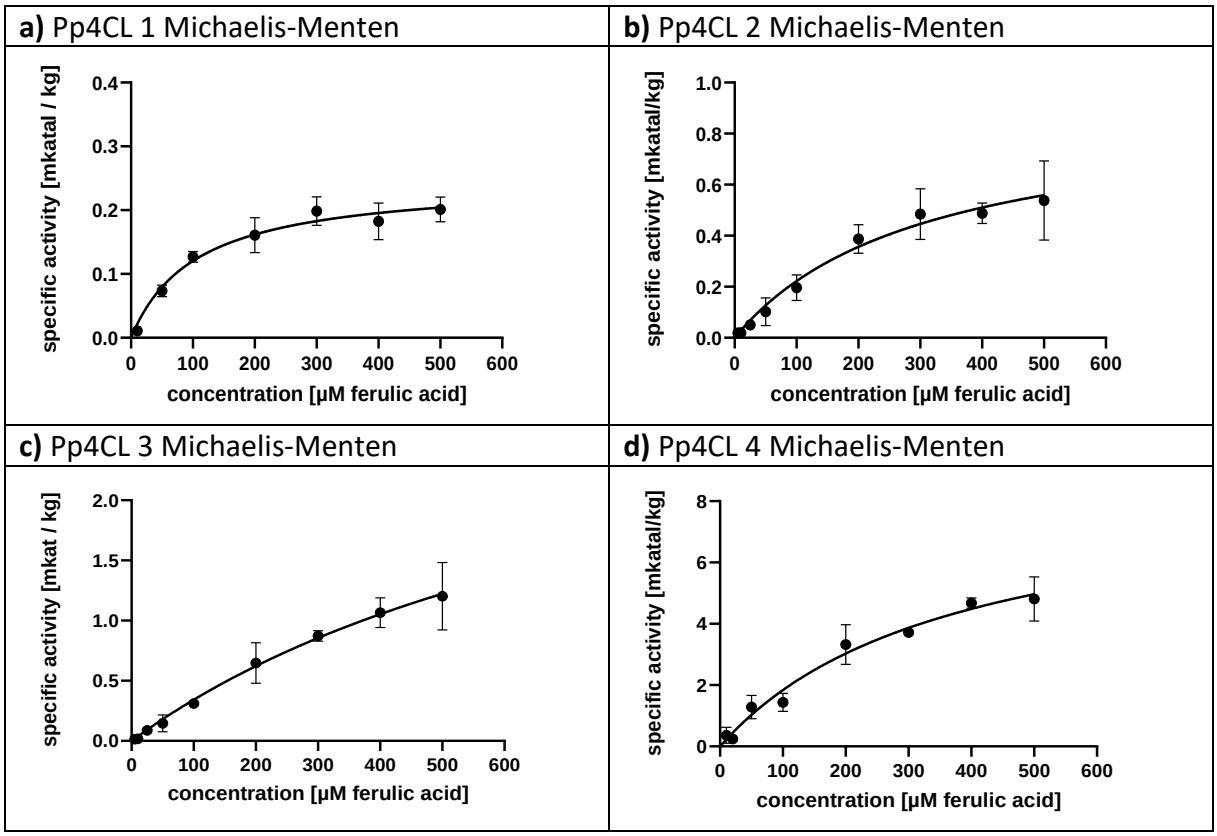


Figure 77 Reactions of Pp4CL 1 (a), Pp4CL 2 (b), Pp4CL 3 (c) and Pp4CL 4 (d) with ferulic acid (no saturation attained, therefore determination of kinetic parameters via Cornish-Bowden plot). Graphs represent the mean values of one biological replicate with six technical replicates. Error bars represent the standard error.

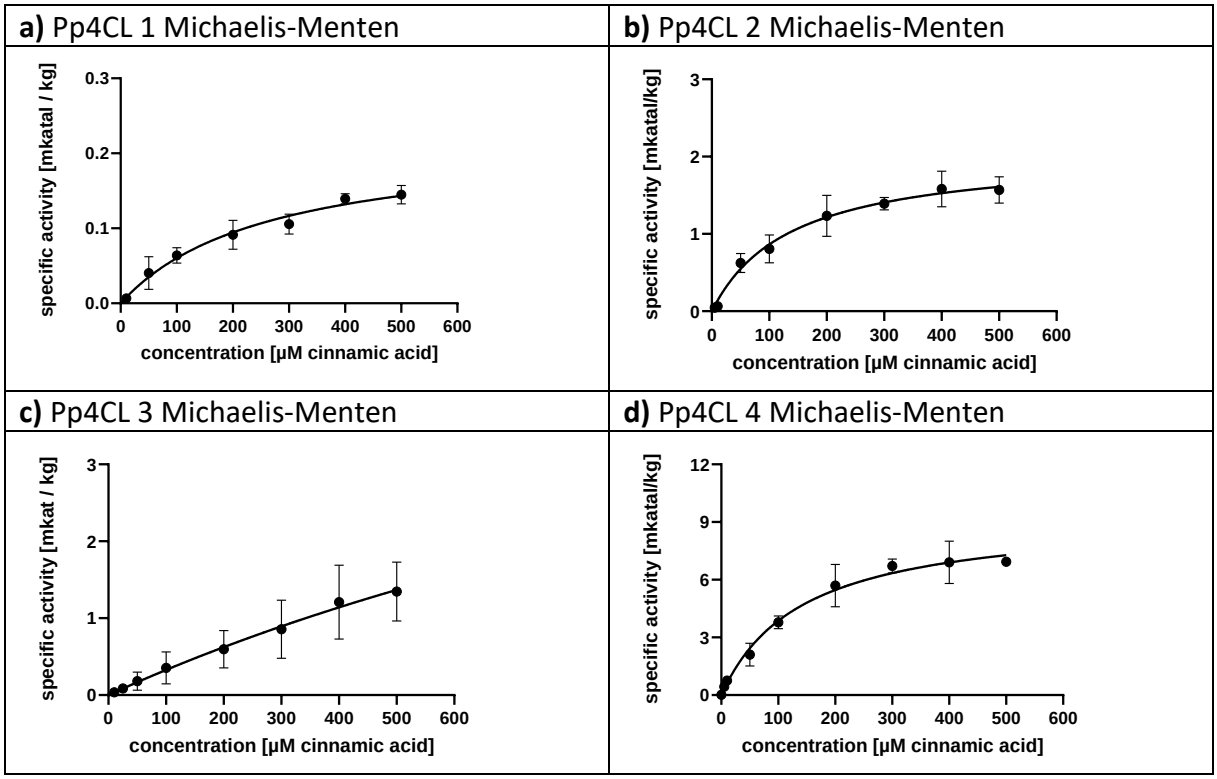


Figure 78 Reactions of Pp4CL 1 (a), Pp4CL 2 (b), Pp4CL 3 (c) and Pp4CL 4 (d) with cinnamic acid (no saturation attained, therefore determination of kinetic parameters via Cornish-Bowden plot). Graphs represent the mean values of one biological replicate with six technical replicates for Pp4CLs 1, 2 and 4 and twelve technical replicates for Pp4CL 3. Error bars represent the standard error.

Table 23 Kinetic values for Pp4CLs 1 to 4 with 4-coumaric, caffeic, ferulic and cinnamic acids, parameters belonging to the graphs shown in Figures 75 to 78. Values with standard errors were calculated directly using the data from the Michaelis-Menten (MM) and Hanes-Woolf (HW) plots. Values without standard errors were determined with the help of the Cornish-Bowden method.

Enzyme	Substrate	K_m [µM]	V_{max} [mkat kg^{-1}]	k_{cat} [s^{-1}]	k_{cat}/K_m [s^{-1} mol^{-1} l]
Pp4CL 1	4-coumaric acid (MM)	16.53 ± 3.20	0.42 ± 0.01	0.03	1814.88
	4-coumaric acid (HW)	13.60 ± 3.54	0.41 ± 0.01	0.03	2205.88
	ferulic acid	102.0	0.24	0.02	196.08
	cinnamic acid	340.0	0.26	0.02	58.82
	caffeic acid (MM)	14.43 ± 2.02	0.26 ± 0.01	0.02	1386.0
	caffeic acid (HW)	16.32 ± 2.18	0.27 ± 0.01	0.02	1225.49
Pp4CL 2	4-coumaric acid (MM)	27.31 ± 2.06	3.04 ± 0.09	0.21	7689.49
	4-coumaric acid (HW)	23.75 ± 4.30	2.02 ± 0.12	0.14	5894.74
	ferulic acid	420.0	0.97	0.07	166.67
	cinnamic acid	123.0	1.94	0.13	1056.91
	caffeic acid (MM)	29.80 ± 1.46	2.12 ± 0.11	0.15	5033.56
	caffeic acid (HW)	30.87 ± 2.78	2.25 ± 0.06	0.15	4859.09
Pp4CL 3	4-coumaric acid (MM)	87.10 ± 2.65	15.46 ±0.12	1.04	16073.48
	4-coumaric acid (HW)	65.76 ± 1.10	14.64 ± 0.04	0.98	14902.68
	ferulic acid	1375.0	4.15	0.28	203.64
	cinnamic acid	860.0	3.5	0.23	267.44
	caffeic acid	156.0	2.10	0.14	126.35
Pp4CL 4	4-coumaric acid (MM)	9.67 ± 0.62	7.77 ± 0.08	0.53	54808.69
	4-coumaric acid (HW)	11.43 ± 2.05	7.58 ± 0.23	0.51	44619.42
	ferulic acid	210.0	6.80	0.46	2190.48
	cinnamic acid	95.0	8.30	0.56	5894.74
	caffeic acid (MM)	13.80 ± 0.54	5.84 ± 0.06	0.40	28985.51
	caffeic acid (HW)	14.67 ± 0.61	6.07 ± 0.07	0.41	27948.19

3.4.6.2 Comparison of Pp4CLs 1 to 5

For all five isoforms of PpPAL, a sequence alignment (Figure 79) as well as a comparison of identities/similarities of the amino acid sequences (Table 24) was created. Both analyses revealed moderately high identities and high similarities between Pp4CLs 1 to 4, yet very low identity (< 30 %) and low similarity (< 45 %) of Pp4CL 5 to the other four isoforms. In Pp4CL 1 to 4, the GEICIRG motif and the AMP-binding domain were found intact, whereas in Pp4CL 5 both were found to be present, yet corrupted. Six out of twelve of the amino acid residues putatively responsible for substrate specificity (Schneider et al. 2003) were dissimilar in Pp4CL 5, the other four isoforms featured all twelve congruently. The 12 amino acid residues (IYNPMGAGGPVL) are with the exception of the methionine residue identical to the residues in Mp4CL 3, which harbours a valine residue in this particular position.

Table 24 Identities/similarities of investigated 4CLs derived from *P. patens* on amino acid level [%]. Calculations were performed using EMBOSS Needle.

	Pp4CL 1	**Pp4CL 2**	**Pp4CL 3**	**Pp4CL 4**	**Pp4CL 5**
Pp4CL 1	100.0/100.0	75.2/81.4	75.0/81.9	72.5/81.4	26.6/39.1
Pp4CL 2		100.0/100.0	84.3/91.1	78.3/87.9	29.5/42.7
Pp4CL 3			100.0/100.0	80.4/89.4	29.9/43.9
Pp4CL 4				100.0/100.0	29.5/43.7
Pp4CL 5					100.0/100.0

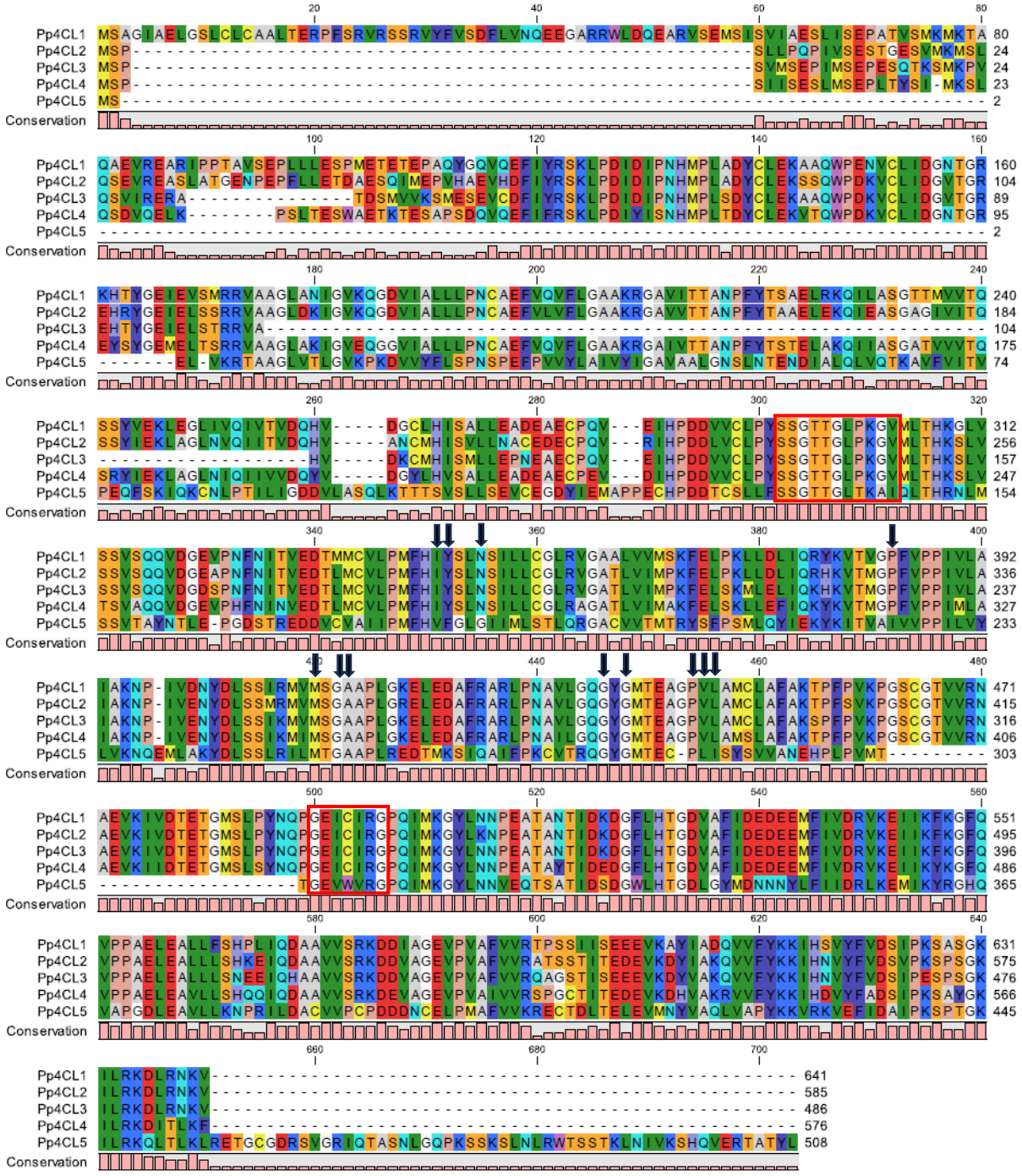


Figure 79 Alignment of amino acid sequences for the five Pp4CL isoforms. The two conserved motifs are marked by black boxes (box I: AMP binding domain as described by Shockey et al. (2003) and Shockey and Browse (2011); box II: 4CL functionality motif as described by Stuible and Kombrink (2001)). The amino acids putatively responsible for substrate specificity as described by Schneider et al. (2003) are marked by black arrows. The alignment was performed with CLC Sequence Viewer version 8.0. (www.qiagenbioinformatics.com).

3.5 Verification of product formation

3.5.1 Verification of PAL products formed from L-Phe, D-Phe and L-Tyr

In order to verify the substrate transformation reactions mentioned in 3.4.1 to 3.4.3, a standard PAL assay was conducted for all eight isoforms as described in 2.5.1.2. For the reactions with L- and D-Phe as well as L-Tyr, activity could be corroborated via HPLC showing the formation of *t*-cinnamic acid from Phe and 4-coumaric acid from Tyr. The highest transformation peaks were attained with MpPAL 3 for all three substrates. L-Phe transformation with MpPAL 3 was similar to PpPALs 3 and 4. In the following, all PAL verification reactions are listed consecutively, showing the respective diagrams (Figures 80 to 82).

It can be seen that, for measurements with L-Phe, PpPALs 2, 3, and 4 show substrate transformation peaks ten to fifteen times higher than that of the standard (25 µM *t*-cinnamic acid), whereas the peak of PpPAL 1 is about three times as high. MpPAL 1 shows a transformation peak comparable to the peak of the standard, while those of MpPAL 2 and CbPAL 1 are significantly smaller. The transformation peak of MpPAL 3 was approximately fifteen times higher than the peak of the standard, showing transformation comparable to that of PpPALs 2, 3 and 4. All HPLC verification graphs of measurements with L-Phe are shown in Figure 80.

D-Phe transformation could be verified for all four PpPALs, as well as for MpPAL 3. In the case of PpPALs 1, 2 and 4, the transformation peak was significantly lower than the peak of the standard (25 µM *t*-cinnamic acid), whereas PpPAL 3 yielded a peak comparable to that of the standard. For MpPAL 3, the transformation was far more considerable, resulting in a peak ten times higher than that of the standard. For the graphs of D-Phe transformation verification, see Figure 81.

Transformation of L-Tyr could be verified for all four PpPALs, MpPAL 1 and MpPAL 3. As was the case with D-Phe, PpPAL 3 and MpPAL 3 showed considerable substrate transformation, the peaks being twenty to thirty times as high as that of the standard of 25 µM 4-coumaric acid. The other four samples displayed low transformation, the peaks being significantly smaller than that of the standard. In the case of the reaction of MpPAL 1 with L-Tyr, HPLC verification proved acceptance, although this isoform did not possess enough activity for proper kinetics. For all L-Tyr verification measurement graphs, see Figure 82.

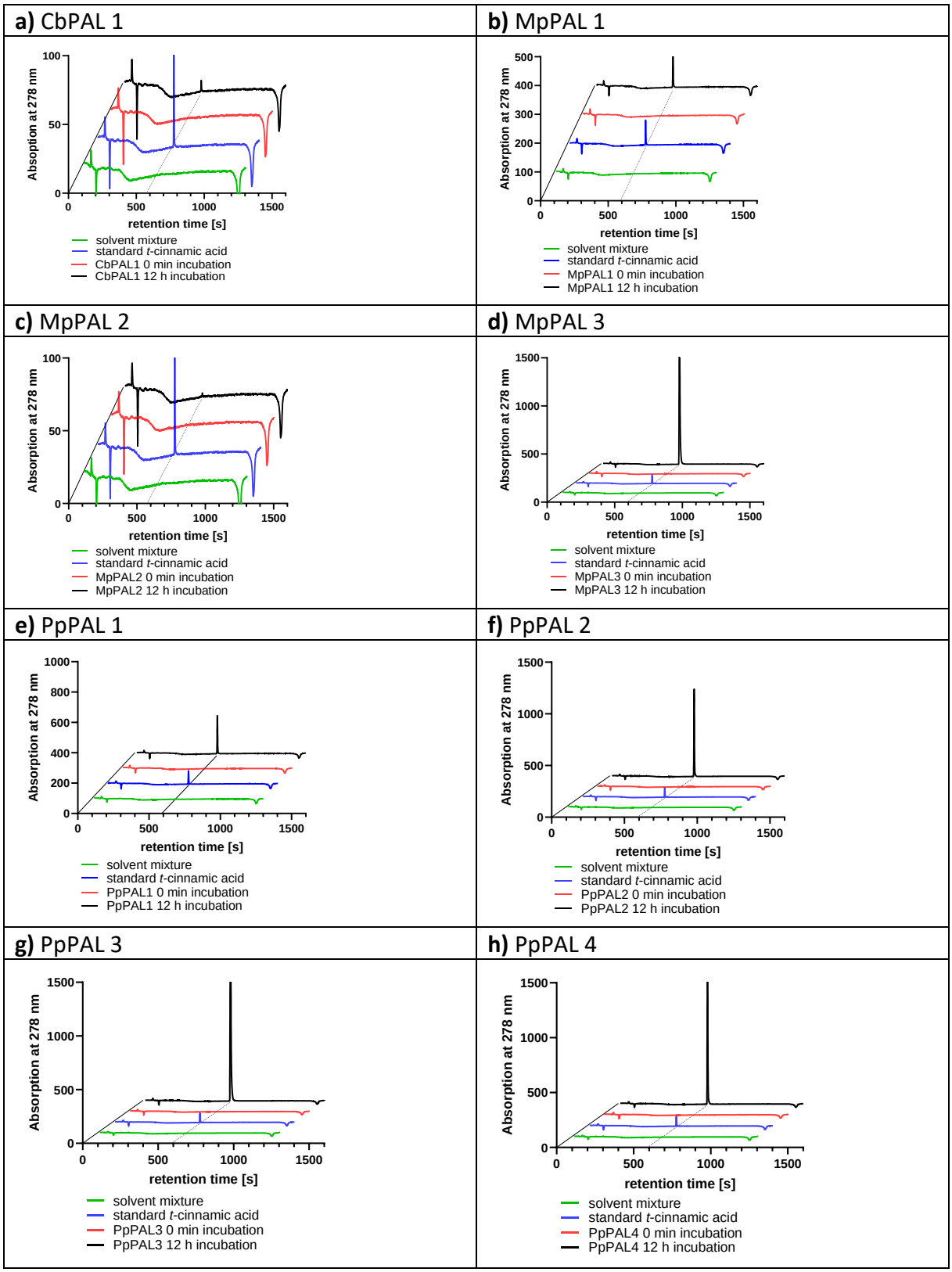


Figure 80 Verification of PAL measurements with L-Phe for CbPAL 1 (a), MpPALs 1 to 3 (b to d) and PpPALs 1 to 4 (e to h). Green line marks the solvent, blue line marks the standard (25 µM *t*-cinnamic acid), red line marks the enzyme assay having been incubated for 0 min, black line marks the enzyme assay having been incubated for 12 h.

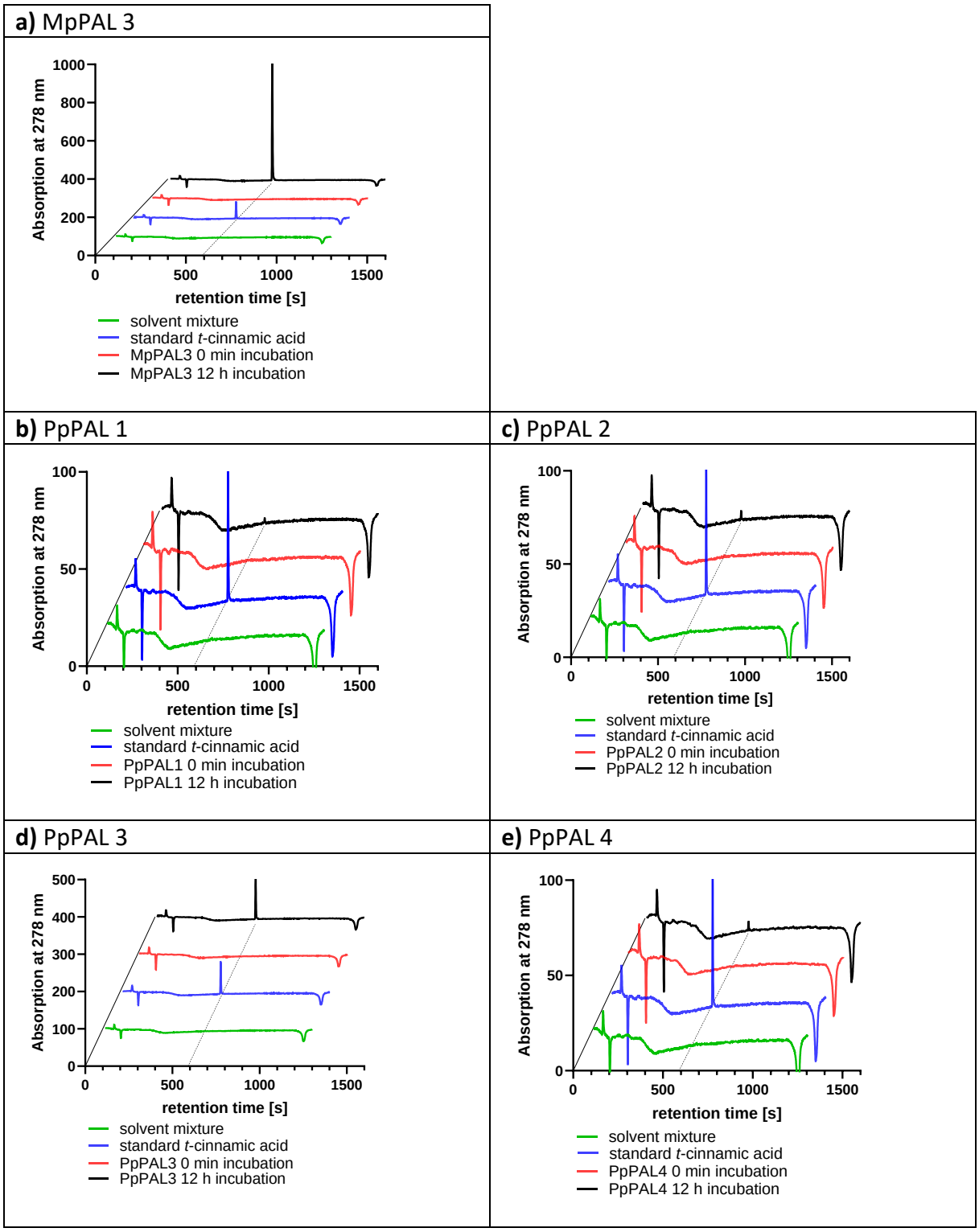


Figure 81 Verification of PAL measurements with D-Phe for MpPAL 3 (a) and PpPALs 1 to 4 (b to e). Green line represents the solvent, blue line represents the standard (25 µM *t*-cinnamic acid), red line represents the enzyme assay stopped instantly (0 min incubation) and black line represents the enzyme assay following incubation for 12 h.

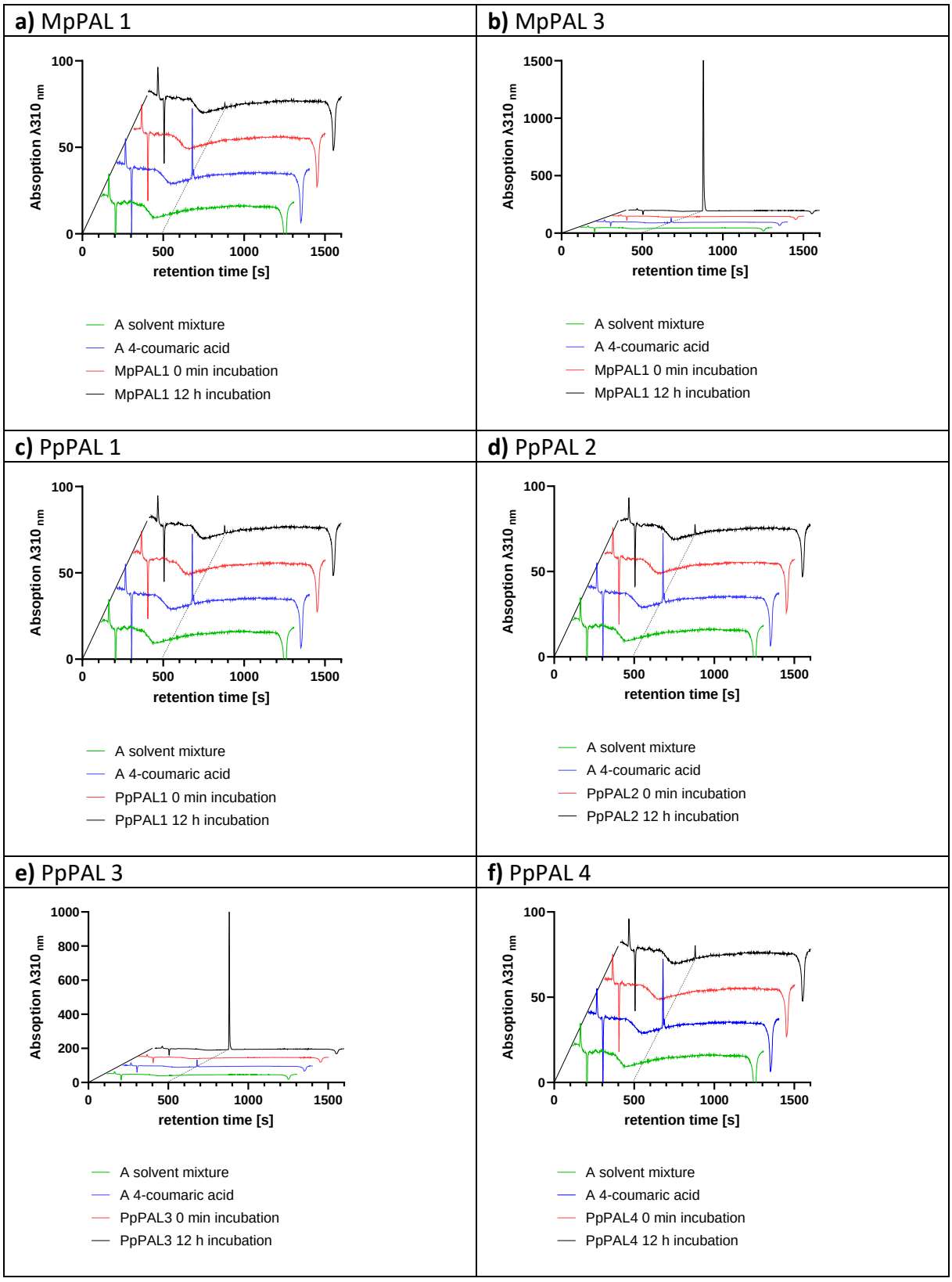


Figure 82 Verification of PAL measurements with L-Tyr for MpPALs 1 and 3 (a and b) and PpPALs 1 to 4 (c to f). Green line marks the solvent, blue line marks the standard (25 µM 4-coumaric acid), red line marks the enzyme assay incubated for 0 min, black line marks the enzyme assay incubated for 12 h.

3.5.2 Verification of PAL measurements with L-His

All eight isoforms of PAL were intensively examined with L-His as substrate. The reaction, first tested with the most efficacious PpPAL, isoform 2, seemed to yield activity. This assumption arose based on the measuring curve seemingly rising slowly, which usually implies an increase in substrate transformation. It is crucial to understand that, when measuring reactions with very low affinities and reactivities, it is quite difficult to differentiate between a genuine reaction taking place or a feigned reaction that is really nothing but a consecutively rising slope on the computer screen. This is an optically confusing phenomenon, possibly elicited by slow incalescence of the cuvette or a miniscule side reaction that otherwise would not be further noticed. In this particular situation, the high sensitivity of the photometer evokes this false conclusion. Such low, but visible slopes in fact do exist; sometimes, substrate transformation occurs on such low concentration levels that minimal rises of the measuring curve can be observed. However, this circumstance makes it necessary to conduct verification measurements (via LCMS or HPLC) in order to exclude false reactivity. To exhaustively clarify whether L-His was transformed by PpPAL, two schemes were developed. First, a widespread assay was conceptualized incorporating all components of His-measurements in different variations of composition (instead of 5 mM DTT, 5 mM glutathione was used). In this measurement, buffer was also tested against buffer and mere water against water. The assay was designed as summarized in Table 25.

This widespread investigation did not consolidate any kind of reactivity for L-His. In some cases, the slight slope was visible, but no discernible difference could be shown to the measurement of mere deionized water applied to both measuring and reference cuvette. Because of this unsatisfactory result, all eight PAL isoforms underwent HAL assays according to 2.5.1.2 with all necessary components and subsequent HPLC analyses with preparations as described under 2.6.2 followed. Subsequent evaluation unambiguously proved that urocanic acid was not formed by any of the eight PALs.

Table 25 Comparative measurement of PpPAL 2 with L-His in different assay compositions. B = buffer (0.5 M Tris-HCl pH 7.5); E = 50 µg/ml enzyme (PpPAL 2); M = 100 mM $MnCl_2$; G = 5 mM glutathione; H = 2 mM L-His. Both 277 nm and 290 nm were tested. All measurements were conducted with two technical replicates per wavelength. The temperature was 40 °C.

Contents of measuring cuvette	Contents of reference cuvette	Wavelength (nm)
B	B	277
E + B + M	E + B + M	277
E + B + M + G	E + B + M + G	277
E + B + M + G + H	E + B + M + G + H	277
E + B + M + H	E + B + M	277
E + B + M + G	E + B + M	277
E + B + M + G + H	E + B + M + G	277
B + H	B + H	277
B + H + G	B + H + G	277
water	water	277
B	B	290
E + B + M	E + B + M	290
E + B + M + G	E + B + M + G	290
E + B + M + G + H	E + B + M + G + H	290
E + B + M + H	E + B + M	290
E + B + M + G	E + B + M	290
E + B + M + G + H	E + B + M + G	290
B + H	B + H	290
B + H + G	B + H + G	290
water	water	290

3.5.3 Comparative PAL measurement applying 5mM DTT

In order to reassess the effects of DTT on PAL/HAL experiments, PpPAL 3 was chosen for a comparative measurement at a concentration of 1 mM L-Phe both with and without 5 mM DTT as cofactor. The resulting specific activities of both measurements (three technical replicates) were compared, and it could be calculated that addition of DTT evoked a significant decrease in the specific activity (about 25 %).

3.5.4 Verification of 4CL measurements

Verification of 4CL measurements was performed according to 2.6.4. Because of the difficulty of identifying peaks confirming the emergence of CoA thioesters in the presence of a peak from the respective hydroxycinnamic acid, an indirect method using rosmarinic acid synthase (RAS) derived from *Coleus blumei* was chosen. After incubation of the standard 4CL assay, RAS and 10 mM 4-hydroxyphenyllactate (pHPL) were added to the reaction and incubated for 1

hour (see also 2.6.3). Then, verification of the corresponding pHPL-ester was conducted via LCMS. The substrates with the highest catalytic efficiencies were applied for the respective isoform. It was shown that all Pp4CLs as well as Mp4CL 3 accepted 4-coumaric acid best. Substrate transformation for Pp4CL 1 yielded a peak of the same height than that of the standard (25 µM 4-coumaroyl-pHPL), while transformation of the other four tested isoforms resulted in peaks far higher than the peak of the standard. For the verification graphs of 4CL measurements conducted with 4-coumaric acid (Pp4CLs 1 to 4, Mp4CL 3), see Figure 84. For isoforms 1 and 2 from *C. braunii* and Mp4CL 4, caffeic acid showed the highest specific activities, but the transformation peaks obtained from LCMS were present, but very low when compared with the peak of the standard (25 µM caffeoyl-pHPL). For the graphs displaying substrate transformation of Cb4CL 1, Cb4CL 2 and Mp4CL 4 with caffeic acid, see Figure 83.

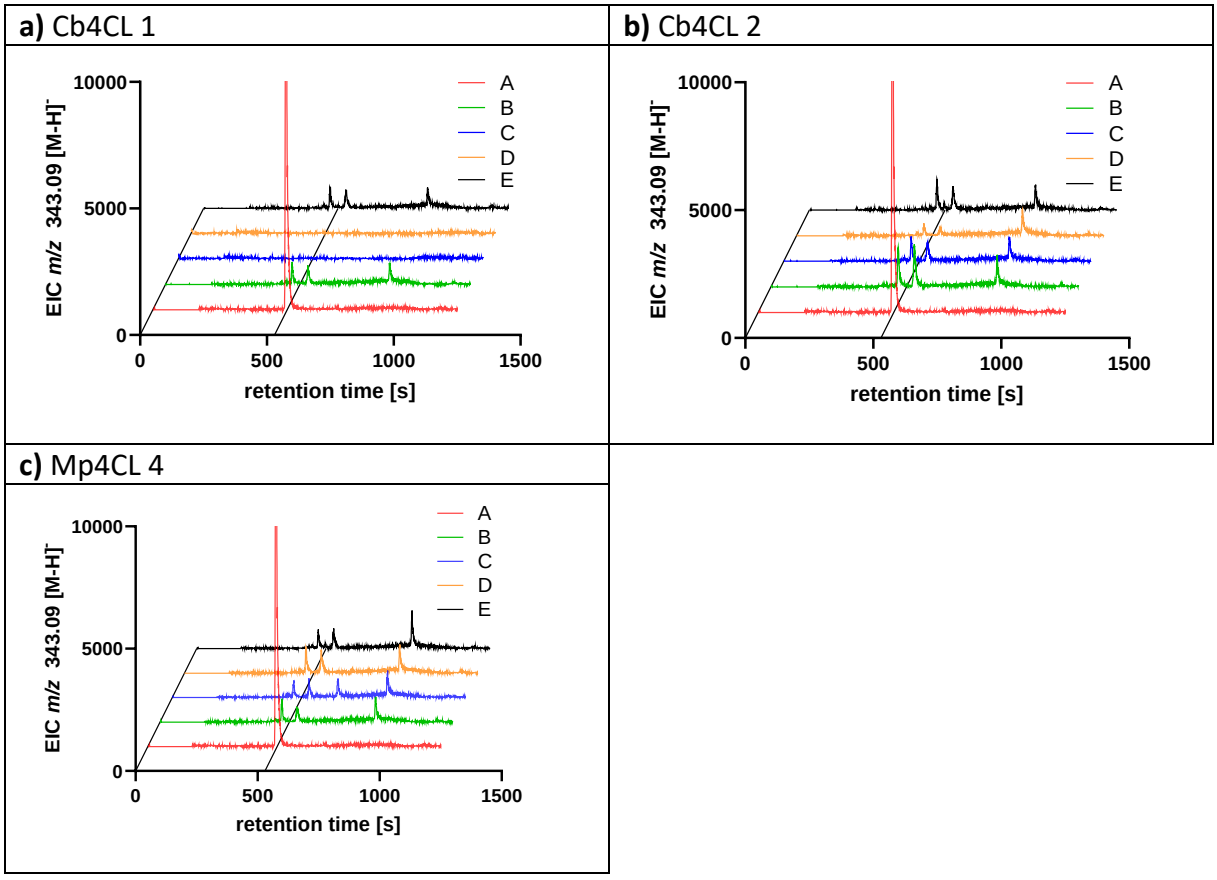


Figure 83 Verification of 4CL activity with caffeic acid for Cb4CLs 1 and 2 (a and b) and Mp4CL 4 (c). Red line (A): standard (25 µM caffeoyl-pHPL), green line (B): active 4CL with inactivated RAS, blue line (C): inactivated 4CL with active RAS, orange line (D): inactivated 4CL with inactivated RAS, black line (E): active 4CL with active RAS.

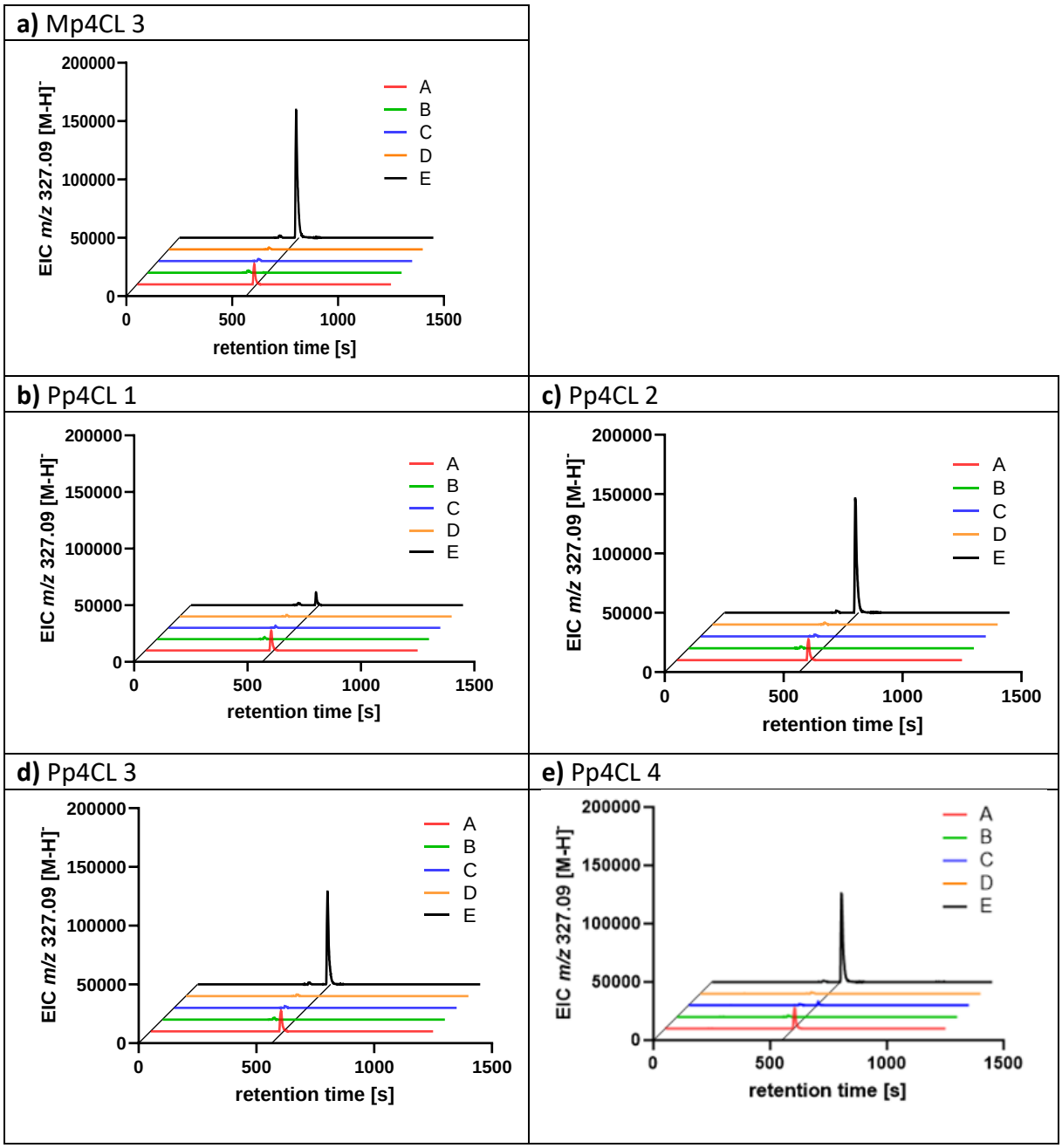


Figure 84 Verification of 4CL activity with 4-coumaric acid for Mp4CL 3 (a) and Pp4CLs 1 to 4 (b to e). Red line (A): standard (25 µM 4-coumaroyl-pHPL), green line (B): active 4CL with inactivated RAS, blue line (C): inactivated 4CL with active RAS, orange line (D): inactivated 4CL with inactivated RAS, black line (E): active 4CL with active RAS.

3.6 Phylogenetic analyses of PAL

The phylogenetic analysis (Maximum Likelihood Algorithm, 1000 bootstraps) for all isoforms of PAL was conducted using MEGA 11 according to 2.7. The PALs investigated were all derived from members of the Streptophyta and, with the exception of CbPALs 1 and 2 and a putative PAL from *Klebsormidium nitens*, all belonged to the embryophytes. The resulting tree showed a close phylogenetic relation between almost all members of the angiosperms, with one entire clade for the monocots and one for the dicots. The only isoform that did not appear among the other annotated isoforms of its kind was PAL 1 derived from *Nelumbo nucifera*, it instead appeared in a clade together with gymnosperm PALs. This particular clade also encompassed two PALs derived from the lycophytes *Selaginella moellendorffii* and *Isoetes lacustris*, respectively. The neighbouring clade to this last described contains four miscellaneous PALs derived from lycophytes/polypodiophytes or gymnosperms, as well as almost all investigated bryophyte PALs. The two isoforms of PAL derived from *Anthoceros agrestis* described in Pezeshki et al. (2022) appear also in this group. Another isolated clade of the tree contained two isoforms derived from polypodiophytes and two isoforms derived from the monocots.

Closer investigations into bryophyte phylogeny of PAL showed a distinct similarity between all five isoforms of PpPAL, being located directly adjacent to one another in the same group. Of the four PALs derived from *M. polymorpha*, PALs 1 and 2 shared the highest similarity, whereas MpPAL 3 was located on a neighbouring branch yet still highly similar.

MpPAL 4, however, was extremely dissimilar to the others, being located in a separate clade of its own, sharing no apparent similarity with any other isoform in the tree. This is especially striking, considering that MpPAL 4 also showed no enzymatic reactivity.

For CbPAL, both isoforms were located directly adjacent to one another and closely related to a putative PAL from *K. nitens*. The area of most condensed agglomeration in terms of comparability for the investigated enzymes included the aforementioned clade encompassing all five PpPALs and MpPALs 1 to 3, PALs 1 and 2 derived from *A. agrestis* being located on the neighbouring branch of the same clade. The phylogenetic tree of PAL is shown in Figure 85.

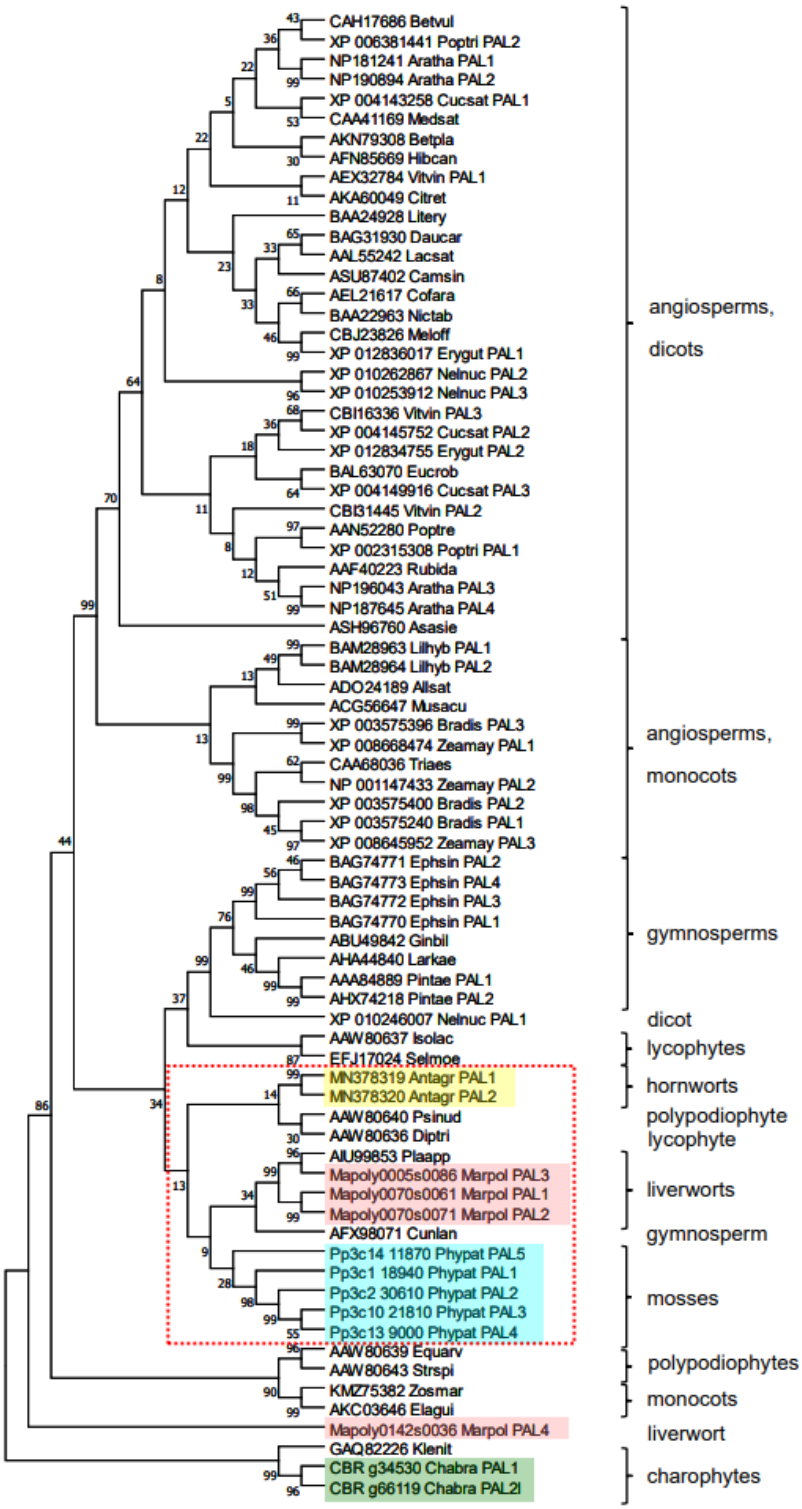


Figure 85 Phylogenetic analysis of PAL sequences created with the Maximum Likelihood algorithm with 1000 bootstraps of the MEGA 11 software (Schwarze and Petersen 2024). Sequences of enzymes relevant in this work are shown in the following colours: green – *C. braunii*, red – *M. polymorpha*, blue – *P. patens*, yellow – *A. agrestis*. The red box signifies the closely related bryophyte PALs with the exception of MpPAL 4.

3.7 Phylogenetic analyses of 4CL

Phylogenetic analyses were conducted for a number of isoforms of 4CL according to 2.7. Compared to the analysis of the same kind performed for PAL, this phylogeny was significantly more heterogenous in terms of similarity. No unanimous order could be established for gymnosperms and angiosperms, and the investigated isoforms of this work were distributed amongst the different clades and also the compartmentalization was distinctly more chaotic. In the upper half of the tree (depicted in Figure 86), gymnosperms, angiosperms and bryophytes appeared alternatingly and no real order could be deduced. Among the investigated isoforms of bryophytic origin, the only condensed zone of similarity included isoforms 1 to 4 derived from *P. patens* and Mp4CL 3 as well as the isoform derived from *A. agrestis* (having been characterized by Wohl and Petersen (2020)). Aa4CL 1 shared the highest comparability with Mp4CL 3. Pp4CL 5 was located on a branch of its own in a different clade, far distanced from Pp4CLs 1 to 4, directly adjacent to 4CL derived from the liverwort *Plagiochasma appendiculatum*. Both of these isoforms, however, appeared in a branch of their own and are apparently not closely related to any of the other isoforms in the tree. Mp4CLs 1, 2 and 4 were distributed loosely across the upper third of the tree, apparently not sharing higher similarities either. Mp4CL 1 appeared in a clade together with 4CLs 1 and 2 derived from *C. braunii*, which were also located close to one another, yet on different branches. Mp4CLs 2 and 4 were situated on different branches of the clade that also harboured Pp4CL 5. All three isoforms were, however, significantly distanced from one another and apparently shared no higher similarity.

In summary, it is conspicuous to observe that all the isoforms of significant activity are adjacent to one another (including Aa4CL 1), whereas the isoforms that display no obvious reactivity are scattered loosely across the phylogenetic tree with no apparent similarity to one another. The close phylogenetic relation between Pp4CLs 1 to 4 and Mp4CL 3 might indicate their close catalytic relatedness.

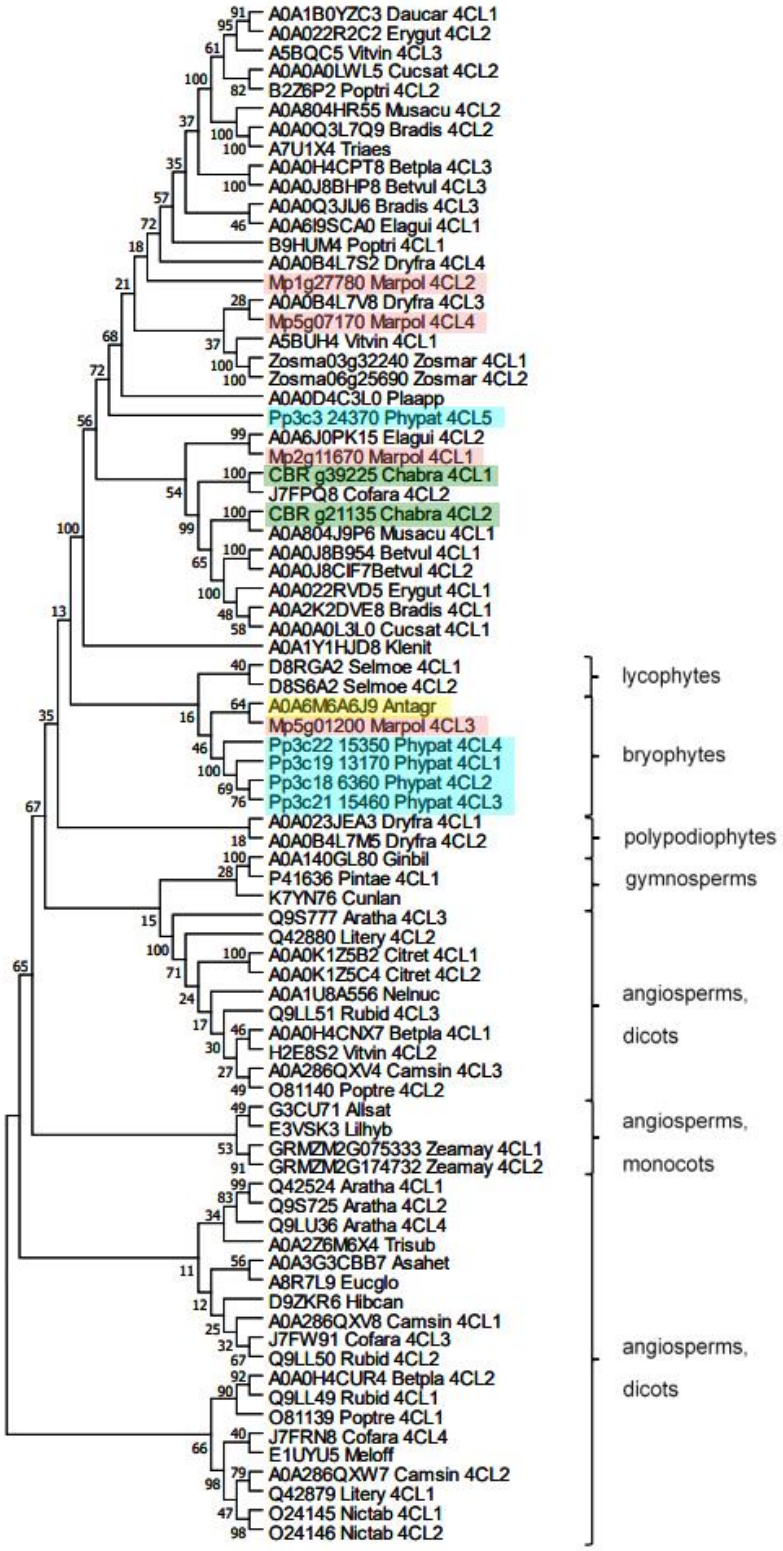


Figure 86 Phylogenetic tree of 4CL sequences created with the Maximum Likelihood algorithm with 1000 bootstraps of the MEGA 11 software (Schwarze and Petersen 2024). Sequences representing enzymes relevant in this work are shown in the following colours: green – *C. braunii*, red – *M. polymorpha*, blue – *P. patens*, yellow – *A. agrestis*.

4 Discussion

4.1 The role of specialized metabolism during plant terrestrialization

With ongoing immersion into the origins of land plant evolution, the understanding of the emergence of specialized metabolism has become increasingly important. This is especially relevant regarding the fact that specialized plant metabolism is closely connected to plant terrestrialization (Dadras et al. 2023). It is furthermore relevant to determine which events encouraged the evolution of the immense biodiversity of the embryophytes that account for the majority of the earth's biomass (Becker et al. 2020; Rieseberg et al. 2023).

The general phenylpropanoid metabolism is a crucial pathway for the development of specialized compounds that support land plant survival by providing strengthening elements for cell wall stability and desiccation avoidance as well as protective metabolites that help to shield off UV light and repel predators (Renault et al. 2017; Wohl and Petersen 2020).

Regarding plant terrestrialization, the phenylpropanoid pathway is an important area of investigation in terms of its key enzymes and their activities, as well as a means to provide possible answers to some of the questions involving the transition to land. Here, it is necessary to stress that much work has already been put into finding active isoforms from all kinds of eukaryotic as well as prokaryotic enzymes and phylogenetic experiments have been conducted both for the different enzymes derived from the phenylpropanoid metabolism and for the investigated classes of plants harbouring this pathway (Dixon et al. 2002).

When investigating phenylalanine ammonia-lyase, the first enzyme in the general phenylpropanoid metabolism, the question arises whether this essential enzyme can be found in extant members of streptophyte algae having at some point millions of years ago facilitated their transition to land.

Phenylalanine ammonia-lyase is known to carry out the first reaction step of the general phenylpropanoid pathway, the deamination of L-Phe leading to *t*-cinnamic acid, but it can also catalyze the deamination of L-Tyr, leading to 4-coumaric acid. Apart from land plants, where PAL appears ubiquitously, it occasionally plays a vital role in the metabolism of certain prokaryotes. In the marine bacterium *Streptomyces maritimus*, a PAL has been discovered that catalyzes the first step in the biosynthesis of benzoyl-CoA that originates from *t*-cinnamic acid

via β-oxidation and is then further transformed via polyketide biosynthesis into the bacteriostatic agent enterocin (Xiang and Moore 2005). In *Sorangium cellulosum*, benzoyl-CoA derived from *t*-cinnamic acid is further transformed into the 18-membered macrolide Soraphen A (Hill et al. 2003) while in *Streptomyces verticillatus*, PAL is involved in the biosynthesis of the carboxamide cinnamamide (Bezanson et al. 1970). Usually, prokaryotic PALs are significantly shorter in terms of their amino acid sequence when compared to eukaryotic isoforms. A gene encoding PAL that is about the same size as the isoform derived from *S. maritimus* (slightly over 500 amino acids) has been described in the gram-negative prokaryote *Photorhabdus luminescens* from the family Enterobacteriaceae. Biochemical investigations have shown that this isoform, though most likely a HAL-candidate, effectively transforms L-Phe into *t*-cinnamic acid. TAL-activity, on the other hand, could not be detected. It has furthermore been proven that the respective gene is required for the production of stilbene by *P. luminescens* (Williams et al. 2005).

In recent times there have been a number of finds for putative PAL gene candidates in streptophyte algae, such as *Klebsormidium nitens* (de Vries et al. 2017) and *Mesotaenium endlicherianum* (de Vries et al. 2021) giving rise to the assumption that these extant members of streptophyte algae might have already carried the ability to produce phenylpropanoids. PAL 1 derived from *Chara braunii* having been described in this work belongs in this context and deserves special attention on account of it being the only example of a functional PAL from a streptophyte alga to date (Schwarze and Petersen 2024). The fact that CbPAL 1 effectively transforms L-Phe to *t*-cinnamic acid (yet does not transform L-His to urocanic acid) indicates that the general purpose of the phenylpropanoid metabolism as a pathway towards specialized metabolites capable of enhancing plant resilience was already present prior to the actual terrestrialization process. This process has occurred most likely in a member of the Zygnematophyceae (de Vries and Archibald 2018). Considering the high specificity of CbPAL 1 for L-Phe as well as its comparatively low catalytic efficiency, these observations, when compared with MpPALs and PpPALs, might indicate an evolutionary optimization process towards more efficacious enzymes. MpPAL 3 and PpPALs 1 to 4 not only show higher catalytic efficiencies, but also transform D-Phe and L-Tyr.

Aside from streptophytic algae, there have been studies with Rhodophyta leading to the discovery of G, H and S lignin in the cell wall of the calcified red algae *Calliarthron cheilosporioides* (Martone et al. 2009) and *Calliarthron tuberculosum* (Xue et al. 2022). In the

case of *C. cheilosporioides*, two possible reasons for the presence of lignin have been suggested: either the red alga acquired lignin biosynthesis via convergent evolution or the discovered monolignols have derived from terrestrial wood decay and accumulated in the red alga (Renault et al. 2019). In *C. tuberculosum*, extensive research into the transcriptomic dataset led to the discovery of putative candidates for five genes that play major roles in the phenylpropanoid pathway: 4CL, 4-coumaroyl-CoA reductase (CCR), cinnamyl alcohol dehydrogenase (CAD), caffeoyl-CoA *O*-methyltransferase (CCoAOMT) and caffeoylshikimic acid esterase (CSE). CAD is known to be a marker of lignification (Boerjan et al. 2003; Soares et al. 2014). Several other steps of the pathway, on the other hand, could not be detected, such as the enzymes PAL, TAL, PTAL, hydroxycinnamoyltransferase (HCT), caffeic acid *O*-methyltransferase (COMT), cinnamic acid 3-/4-hydroxylase (C3H/C4H) or ferulate 5-hydroxylase (F5H). In the case of PAL, TAL and PTAL, fragmented gene candidates were found, however, these sequences are not considered to be very plausible on a genomic level (Xue et al. 2022).

Because of this incomplete array of proven enzyme candidates and the evolutionary distance between Rhodophyta and Embryophyta, convergent evolution has been suggested as an explanation for the lignified cell walls of red algae (Labeeuw et al. 2015). Convergent evolution correlated with syringyl lignin biosynthesis had already been demonstrated in the lycophyte *Selaginella*, which uses a dual *meta*-hydroxylase, Sm F5H, for S lignin biosynthesis, as opposed to angiosperm S lignin biosynthesis, which occurs through a different pathway with participation of two phenylpropanoid *meta*-hydroxylases, C3'H and F5H (Weng et al. 2010). F5H is the only gene from the phenylpropanoid pathway, of which no homologs have been found in the genome of *Physcomitrium patens* (Ye and Zhong 2022). Because of the exceptional complexity of lignin biosynthesis, it is, however, debatable whether *Calliarthron* and the terrestrial plants have developed this pathway completely independently (Martone et al. 2009).

Convergent evolution has also been researched in Funariaceae, the family of *P. patens*, and evidence has been found for the parallel development of alleles in members of *Physcomitrium* (Beike et al. 2014).

In the streptophyte alga *C. braunii*, two sequences annotated as HAL have been analysed regarding their activity and substrate susceptibility in this work. The unequivocal reactivity with L-Phe as well as the absence of substrate transformation for L-His show that this particular

enzyme can be classified rather as a PAL catalysing the first step towards the biosynthesis of lignin. Investigation of the genome of *C. braunii* in OrcAE (https://bioinformatics.psb.ugent.be/orcae) did not yield any positive hits for cinnamic acid 4-hydroxylase (C4H), the second enzyme of the phenylpropanoid pathway, which transforms *t*-cinnamic acid into 4-coumaric acid.

Having found evidence for the existence as well as activity of two isoforms of 4CL in *C. braunii*, the apparent absence of C4H is all the more enigmatic, especially when considering that there has been no significant TAL/PTAL-activity detected with CbPAL 1, which would have provided an alternative way for the production of 4-coumaric acid.

All of these finds can be correlated with the hypothesis that streptophyte algae already possessed a number of characteristic genetic traits necessary for the synthesis of lignin that would develop later in streptophyte evolution (de Vries et al. 2017). Charophyceae are said to possess compounds similar to lignin, whereas the question of the presence of “true” lignin can as of yet not finally be answered (Martone et al. 2009; Labeeuw et al. 2015). In this context, the discovery of an active PAL isoform in *C. braunii* that has been derived directly from plant material via molecular biological means constitutes an isolated case as of now. Following terrestrialization, a putative optimization of the effectiveness of enzymatic activities and the subsequent diversification of the specialized metabolite pool is likely to have occurred.

4.2 Investigations into phenylalanine ammonia-lyase

The fact that CbPAL 1 was also annotated as HAL required thorough examination regarding possible acceptance of L-His, especially, when considering the fact that CbPAL 1 represents the putatively least elaborate PAL isoform investigated in this work and would therefore be the most likely candidate for HAL activity. This deduction was based on the assumption that PAL is derived from HAL (Ritter and Schulz 2004). To date, there have not been many reports of apparent HAL activity outside of prokaryotes; the only report from a tracheophyte comes from *Vicia faba* describing the requirement to add a sulfhydryl reagent as well as a divalent metal cation (in this case Mn^{2+}) to the assay in order to attain reactivity (Kamel and Maksoud 1978). The necessity of a divalent cation for HAL activity (but not PAL activity) has already been described by Klee (1972) with HAL from *Pseudomonas*.

In this work, the measurement with L-His was conducted adding both Mn^{2+} and a sulfhydryl agent, the latter varying between dithiothreitol (DTT), mercaptoethanol and glutathione. None of the applied cofactors induced any kind of optimization, on the contrary, a specific comparative experiment (see 3.5.3) revealed the activity-diminishing properties of DTT in a regular PAL measurement. In this particular case, the specific activity deteriorated significantly after addition of the sulfhydryl agent. Another impairing factor was the fact that after a short amount of time, DTT underwent oxidation, which led to a distinct browning of the sample solution and a detrimental change of the absorption rate, thus preventing any diligent evaluation of the experiment.

In contrast to that, there is one report of histidase activity in the unicellular green alga *Chlamydomonas reinhardtii* where the measurement was conducted applying Mn^{2+}, which created an increase of enzymatic activity of about 20 % (Hellio et al. 2004). The fact that the HAL assay (implementing a sulfhydryl agent and Mn^{2+}) works has been consolidated in our lab by Janik Marks using a putatively bacterial HAL (unpublished data).

In order to categorize all isoforms regarding their catalytic behaviour, the K_m values were compared with those of reported PALs from the BRENDA enzyme database (https://www.brenda-enzymes.org/), which ranged from 6.5 µM to 75.6 mM for L-Phe and from 97 µM to 7.8 mM for L-Tyr. Based on comparison with results from the database, the PAL isoforms investigated in this work can be considered catalytically equal to seed plant PALs.

The isoforms of PAL derived from *M. polymorpha* behave heterogeneously in terms of their catalytic properties. Isoforms 1 and 2 are very selective towards L-Phe at a low turnover rate (catalytic efficiencies of ~ 4400 s^{-1} mol^{-1} l and 900 s^{-1} mol^{-1} l, respectively), whereas MpPAL 3 displays medium affinity towards L-Phe with a K_m comparable to PpPALs 2 and 3. MpPAL 3 is also capable of transforming D-Phe and L-Tyr, which makes it unique when comparing MpPALs with each other. Also, it is noteworthy that measurements with MpPAL 3 generated the highest catalytic efficiencies for all three substrates out of all investigated PAL isoforms, making it the most efficacious PAL of this study. This also correlates with the verification of substrate transformation performed as HPLC measurements and described in 3.5.1, where MpPAL 3 displayed the highest transformation peaks for all three tested substrates. The catalytic efficiencies for all five PAL isoforms accepting D-Phe and L-Tyr are, however, relatively insignificant as they fluctuate merely between 1 and 30 s^{-1} mol^{-1} l. For L-Phe, MpPAL 3 achieved

a catalytic efficiency of ~ 28000 s^{-1} mol^{-1} l, which is significantly higher than for the most effective PpPAL (isoform 1, catalytic efficiency of ~ 21000 s^{-1} mol^{-1} l) (Schwarze and Petersen 2024). What is particularly noteworthy in the context of catalytic efficaciousness is the fact that MpPAL 3 is clearly the most effective isoform for L-Phe derived from *M. polymorpha* in spite of it being the only one of the four investigated MpPALs annotated as HAL and no catalytic reactivity could be observed for L-His. When compared to a characterized PAL from the liverwort *Plagiochasma appendiculatum*, the affinities for L-Phe and L-Tyr are comparable to MpPAL, whereas the catalytic efficiencies are in both cases significantly higher for PaPAL (Yu et al. 2014). The transcriptome of *M. polymorpha* has been examined and three isoforms for C4H have been identified so that the first three subsequent enzymes in the phenylpropanoid pathway are present in *M. polymorpha* (Takahashi and Asakawa 2017). Recently, it has been shown that a deficiency in C4H in *P. patens, M. polymorpha* and *A. agrestis*, artificially caused by either *CYP73* inactivation or inhibitor treatment, leads to a diminution of phenylpropanoid derivatives as well as a compromised development of plant growth (Knosp et al. 2023).

The bryophyte *P. patens* possesses lignin-like structures, yet does not undergo lignification (Espiñeira et al. 2011). In this work, five isoforms of PpPAL have been expressed into recombinant proteins and biochemically characterized. All of these enzymes had initially been annotated as HAL (see Table 8). Measurements with L-His were carried out according to 2.5.1.2 and after apparent absence of transformation, a verification experiment according to 3.5.2 was conducted in order to definitively exclude HAL-activity.

In four of the investigated isoforms, PpPALs 1 to 4, the transformation of L-Phe, D-Phe and L-Tyr was successfully substantiated, hereby proving the first step of the general phenylpropanoid metabolism to be present and also providing an independent path towards 4-coumaric acid. The catalytic efficiencies ranged from approximately 5400 to 21000 s^{-1} mol^{-1} l with L-Phe (K_m ranging from 31 to 145 µM), thus indicating a higher efficaciousness for PpPAL when compared to the PAL derived from *C. braunii*. Measurements with D-Phe and L-Tyr resulted in much lower catalytic efficiencies, which is in accordance with the findings from other organisms. The catalytic properties of PpPALs 1 to 4 are at approximately the same level with those of tracheophytes having been investigated in the past. PALs 1 to 4 from *Petroselinum crispum* showed rather low and comparable K_m values for L-Phe ranging from 15 to 24.5 µM, which speaks for a high affinity, whereas L-Tyr measurements revealed

considerably weaker affinities of 2600 to 7800 µM (Appert et al. 1994). Isoforms 1 to 4 derived from *Arabidopsis thaliana* show a significantly lower affinity (higher K_m) than the PpPALs, ranging from 64 to 71 µM for isoforms 1, 2 and 4, while isoform 3 is even less affine with a K_m of 2560 µM. Again, L-Tyr proved much less efficacious a substrate than L-Phe (Cochrane et al. 2004).

PpPAL 5 was investigated with all putative substrates in the highest possible concentrations yet did not show any transformation. The fact that a positive Western blot consolidated the presence of the putatively expressed protein gives rise to the assumption that either the expression process somehow damaged the protein to the point of becoming inoperative or that PpPAL 5 was simply not operational from the beginning. Phylogenetic analyses revealed a high similarity between the five PAL isoforms derived from *P. patens*, including PpPAL 5 having shown no reactivity. The close structural relatedness implies at least a theoretical PAL identity, and it has also been shown that the imperative characteristic ASG motif is present in PpPAL 5 (see alignment Figure 66). Expression of PpPAL 5 has been performed several times, and crucial parameters have been diversified (incubation of the His-tag column on ice for twice as long; change of the buffer from 0.5 M Tris-HCl pH 8.5 to 50 mM 0.1 M KPi pH 7.5; three additional washing steps), but to no avail.

A similar situation was observed with CbPAL 2, which had yielded a positive Western blot (performed with the short version) but proved inoperative when conducting enzyme kinetics. Regarding the fact that the sequence of CbPAL 2 was not unambiguously ascribable to active PAL/HAL (the 5'-part being conspicuously similar to a putative histidine tRNA-ligase) and the identities/similarities with the active CbPAL 1 being very low, it is not unlikely that no activity should be detected.

Among the hornworts, two isoforms of PAL derived from *Anthoceros agrestis* have been biochemically investigated regarding their catalytic properties. Both enzymes showed high affinity towards L-Phe, the K_m values were altogether lower than for the average PpPAL (AaPAL 1: 39 µM; AaPAL 2: 18 µM), the turnover, however, was also lower for AaPAL, which resulted in catalytic efficiencies comparable to PpPALs 1 to 4. There was also activity detectable with L-Tyr for both isoforms from *A. agrestis* revealing very similar results for all kinetic parameters compared to PpPALs 1 to 4 (Pezeshki et al. 2022).

Among the monocotyledonous plants, measurements have been conducted with PAL isoforms from *Sorghum bicolor* as well as *Zea mays*. PAL isoforms from *S. bicolor* revealed activity for L-Phe and L-Tyr, providing a way to the emergence of 4-coumaric acid without the use of C4H. Investigations into the catalytic behaviour of SbPAL have led to the observation that the turnover is higher for L-Phe, whereas the affinity is higher for L-Tyr (lower K_m for L-Tyr). For PAL from *Z. mays*, similar observations were made, as the K_m value of ZmPAL 1 for L-Tyr is lower (the affinity therefore higher) than for L-Phe, but the turnover is about 10 times higher for L-Phe, resulting in comparable catalytic efficiencies for both substrates (Rösler et al. 1997; Zang et al. 2015; Jun et al. 2018).

Considering the K_m values retrieved from the BRENDA database, it can be asserted that PALs derived from *P. patens* are comparable to angiosperm PALs in terms of affinity and catalytic behaviour (Schwarze and Petersen 2024).

The molecular masses of the investigated isoforms vary between 80 and 87 kDa for PpPALs and MpPALs. For the considerably smaller CbPAL 1 and the two versions of CbPAL 2 the molecular masses range from 60 to 67 kDa.

The fact that most of the PAL isoforms derived from the investigated bryophytes also transform L-Tyr, further consolidates the thesis that with ongoing terrestrialization the enzymatic properties increase and diversify. There have as of now been no findings of enzymes that are chiefly affine towards L-Tyr, except in bacteria (Barros and Dixon 2020). The two organisms *Rhodobacter capsulatus* and *Saccharothrix espanaensis* have proven to harbour isoforms with principal TAL activity and subordinate PAL activity (Kyndt et al. 2002; Berner et al. 2006) but no dedicated TAL could be identified in plants. In contrast to this, PTALs from Poaceae are found to be quite actively participating in the production of lignin via TAL activity (Barros and Dixon 2020).

Whereas enzymatic effectiveness seems to be improving as terrestrialization progresses and vascular plants develop, there is phylogenetic evidence that most intense diversification had already commenced prior to the development of angiosperms. In this particular phylogenetic analysis, gymnosperm PAL genes cluster in three different clades, the earliest of which is postulated to predate the emergence of vascular plants, whereas the other two are said to have originated via gene duplication in an ancestral organism to seed plants prior to the diversification of gymnosperms and angiosperms (Bagal et al. 2012).

Another intriguing approach involved L-DOPA, which is synthesized from L-Tyr and structurally close to caffeic acid, as deamination of L-DOPA would lead directly to its formation. This would present an alternative pathway to caffeic acid, which is otherwise synthesized from cinnamic acid via C4H/C3'H. It has already been shown that L-DOPA is accumulated in large quantities in several plants that play an important role as traditional agricultural crops, such as *Vicia faba* from the family Fabaceae, *Sinapis arvense* from the family Apiaceae or *Hordeum vulgare* from the family Poaceae, where it is said to be involved in the repellence of herbivores (Soares et al. 2014). Plants from the family Poaceae are noteworthy for producing almost half of their lignin from L-Tyr. Grass PTALs are postulated to have emerged 50 to 70 million years ago (Barros and Dixon 2020).

Experiments with L-DOPA have been conducted utilizing all investigated PAL isoforms, none of which showed any substrate transformation. The handling of L-DOPA is rather difficult due to its instability and quick oxidation, making the experiments more laborious and especially the photometric determination difficult. At the current state, however, it can be ruled out, that the investigated PALs are susceptible towards L-DOPA.

4.3 Investigations into 4-coumarate CoA-ligase

Further downstream the phenylpropanoid pathway, 4-coumaric acid synthesized via C4H (or, in an alternative reaction, via TAL directly from L-Tyr) is transformed by 4CL to the respective coenzyme A thioester. Activation as CoA thioester is well-known from primary metabolism. Acyl-activating enzymes are a large superfamily catalysing the formation of CoA thioesters in a two-step reaction. First, the acid is coupled with AMP derived from ATP releasing pyrophosphate. Secondly, AMP is replaced by CoA, which forms the respective thioester (Shockey and Browse 2011). 4CLs possess two characteristic motifs: boxes I and II. Box I represents the AMP-binding domain, which is made up of a conserved motif of twelve amino acid residues (SSGTTGLPKGV) and is found in all members of the adenylate-forming family (Lavhale et al. 2018). This motif appears unaltered in six of the eleven investigated 4CL isoforms in this work: Pp4CLs 1 to 4 as well as Mp4CLs 1 and 3. In the remaining five isoforms, between two and four amino acids differ from the genuine motif, which could putatively be correlated to their significantly decreased activities. The second conserved motif, box II (GEICIRG), is specific for 4CLs and is only completely conserved in Pp4CLs 1 to 4 and Mp4CL 3. These are

the five isoforms with the most distinct catalytic activity, which implies a coherence between the condition of this particular motif and the efficiency of the respective enzyme. The role of GEICIRG is not conclusively clarified, and there has been much debate about its function in the past. Especially the cysteine residue in the box II motif has been suggested to be vital for catalytic functionality, yet mutational alterations of this particular residue have only moderately altered enzyme activity (Gocht and Marahiel 1994; Stuible et al. 2000; Lavhale et al. 2018).

Pp4CL 5, Mp4CLs 1, 2 and 4 and Cb4CLs 1 and 2 proved less efficient in terms of catalytic properties and in the case of these six isoforms, the GEICIRG motif is degenerated or, in case of Cb4CL 1, evidently not present. All these isoforms show moderate or no activity and with the exception of Mp4CL 1 (which possesses the highest specific activities of the six inefficient isoforms) none of them has an intact box I motif. It is noteworthy that both isoforms of Cb4CL display comparable specific activities (in both cases for caffeic acid) despite the fact that Cb4CL 1 seemingly does not possess a genuine box II motif.

In terms of size, most 4CL isoforms have approximately the same molecular masses, ranging from 63 to 69 kDa. The only exception is Pp4CL 1 with a molecular mass of about 75 kDa.

4CLs derived from *C. braunii* were shown to possess not enough activity to perform reasonable kinetics with, as opposed to CbPAL 1, being moderate in terms of turnover but clearly susceptible towards L-Phe and showing distinct affinity. It has to be emphasized that for Cb4CLs, the GEICIRG motif was at best distorted (in the case of Cb4CL 2) or completely missing (in the case of Cb4CL 1) and the AMP-binding domain was not immaculate either, an initial situation that did not support any declaration of reactivity. The only substrate that showed moderate acceptance was caffeic acid, attaining specific activities of up to 0.6 mkat kg^{-1}. A unique feature is the fact that both Cb4CLs turn over 4-hydroxybenzoic acid as second-best substrate, however, with distinctly lower specific activities.

Considering the fact that Cb4CL 1 was annotated as *o*-succinylbenzoic acid CoA-ligase, it would make for an interesting investigation to reassess the potential susceptibility towards *o*-succinylbenzoic acid.

The affinity of the investigated isoforms of Pp4CL is highest for caffeic and 4-coumaric acid, usually, 4-coumaric acid is accepted a little better (only Pp4CL 1 accepts caffeic acid better) and the K_m values are mostly very low, which speaks for a high affinity. The exception here is Pp4CL 3, which displays moderate affinity (K_m of 87.10 µM) for 4-coumaric acid but balances this with

a high V_{max} of 15.46 mkat kg^{-1} and a strong turnover number of 1.04 s^{-1}. This results in a considerable catalytic efficiency of ~ 16000 s^{-1} mol^{-1} l making it the most unconventional of the investigated Pp4CLs in terms of catalytic behaviour. Pp4CL 3 is also particularly unaccepting of caffeic acid, possessing a K_m of 156.0 (determined via Cornish-Bowden method) and not able to reach saturation by experimental means. Ferulic and cinnamic acids are the least accepted substrates and only Pp4CL 4 reaches sizable catalytic efficiencies for both substrates (2190.48 and 5894.74 s^{-1} mol^{-1} l, respectively). Both K_m and V_{max} had to be determined via the Cornish-Bowden method because saturation was not reached with 500 µM substrate. The particular problem in cases where no saturation could be attained was the fact that above a substrate concentration of ca. 500 µM, the absorption rates of both substrate and resulting reaction product were too high to conduct an appropriate and comparable experiment with due to disturbance of the photometric measurement.

Pp4CL 4 proved to be the most effective isoform derived from *P. patens* as it was highly efficient for both 4-coumaric and caffeic acids with catalytic efficiencies of 54808.74 and 28985.51 s^{-1} mol^{-1} l, respectively. It was also the Pp4CL isoform with the most constantly substantial turnover rates, all of which ranged around 0.5 s^{-1}. These numbers correlate with the turnover dates generated by the measurements of 4CL 1 derived from *Plagiochasma appendiculatum* and a number of its mutants, which range between 0.12 and 0.78 s^{-1} (Gao et al. 2015). Investigations into the BRENDA database revealed turnover numbers much higher with 4-coumaric acid as substrate, however, the examined plants are in most cases members of the seed plants (spermatophytes) and usually distinctly lignified (*Sorbus aucuparia, Populus tomentosa, Pinus taeda*). Turnover numbers for 4CL from *P. tomentosa* range as high as 88.68 s^{-1}, there is, however, no mention of the applied method in the report, so it is not clear whether a comparable method of measuring absorbance or an end point determination was appended (Rao et al. 2015). Furthermore, bryophytes and liverworts, naturally, do not require such extensive production capacity of lignin, a fact that again corroborates the development of subsequently more efficient enzymes as terrestrialization and necessity of acquired resilience progresses.

In the case of 4CL derived from *P. patens*, isoforms 1 to 4 had previously been identified by Silber et al. (2008) and three of them had been partially characterized, however, no kinetic experiments were performed. It was shown that 4-coumaric acid was the best accepted substrate and that, besides ferulic and caffeic acids, cinnamic acid is also accepted by Pp4CL.

These results correspond to those in this work. The numbering of the four enzymes differs from the numbering applied in this work as follows: Pp4CL 1 = Pp4CL 2 (Silber et al. 2008); Pp4CL 2 = Pp4CL 1 (Silber et al. 2008); Pp4CL 3 = Pp4CL 4 (Silber et al. 2008); Pp4CL 4 = Pp4CL 3 (Silber et al. 2008). In the case of Pp4CL 1 (Pp4CL 2 in Silber et al. (2008)) the version of the protein described in this work is 56 amino acid residues shorter than in Silber et al. (2008). Pp4CL 5 has not been investigated by Silber et al. (2008).

It is unknown whether Pp4CL 5 was regularly expressed in *E. coli* and just did not show any significant activity, or whether an undefined problem prevented correct expression. Because of the fact that no positive Western blot could be attained for this isoform, it remains an open question whether Pp4CL 5 is completely inactive or not.

The most salient characteristic of 4CL derived from *M. polymorpha* is the fact that only one of four isoforms possesses noticeable activity and that this particular isoform is abundantly powerful, even when compared to angiosperm 4CLs. Mp4CL 3 is by far the most active enzyme of this study and reaches catalytic efficiencies that exceed 1 million s^{-1} mol^{-1} l. For 4-coumaric acid (Michaelis-Menten), the K_m is particularly low (1.96 µM) and the catalytic efficiency is 2551020.41 s^{-1} mol^{-1} l, the highest out of all investigated reactions. It also has to be emphasized that Mp4CL 3 accepts and transforms ferulic acid almost three times as well as caffeic acid, which is also a unique feature when comparing affinities and activities of 4CLs in general. The other three Mp4CLs possess activities that range from very low to insignificant. Mp4CLs 1 and 2 show turnover, but not enough for construction of kinetics, cinnamic acid does not seem to be accepted at all. Caffeic acid appears to be the only substrate with a clearly defined acceptance, but the specific activity is still extremely low. Mp4CL 4 turns over 4-coumaric acid to a degree comparable to the turnover of caffeic acid by Cb4CLs 1 and 2. Cinnamic acid is the second-best accepted substrate. These findings of one especially active and three very weak isoforms of 4CL for *M. polymorpha* correspond to recently published results for four isoforms of 4CL derived from *Marchantia paleacea* that also feature one very active and three considerably weaker isoforms (Gao et al. 2024). The thesis of this extreme difference of activity is corroborated by the fact that all four isoforms yielded a positive Western blot. This means that at least the expression was conducted successfully and most likely indicates a very low activity, and that the great disparity represents the natural situation.

4.4 Assessment of phylogenetic analyses

Phylogenetic investigations described in 3.6 and depicted in Figure 85 have led to the conclusion that the investigated isoforms of PAL derived from *M. polymorpha* and *P. patens* mostly cluster together in one subclade encompassing also hornwort, lycophyte and polypodiophyte PAL, as well as one gymnosperm isoform (derived from *Cunninghamia lanceolata*). The other subclade of the respective clade contains the gymnosperm PALs, whereas the angiosperm PALs are situated in a clade of their own (segmented into monocots and dicots). The two isoforms from *C. braunii* are located in a separate branch, together with a putative PAL from *Klebsormidium nitens*, the three of them not sharing significant similarity with the rest of the examined isoforms. The most enigmatic isoform is MpPAL 4, being completely isolated and seemingly unrelated to the other MpPALs. This particular PAL did not show any signs of activity and was also not detectable via SDS-PAGE electrophoresis and Western blot. Whether any kind of correlation can be deduced from this, is not known. PpPALs 1 to 5 are directly adjacent to one another, however, close phylogenetic relatedness does not necessarily represent a means to corroborate comparability of catalytic behaviour. This must be especially taken into account, considering that PpPAL 5, though phylogenetically similar to the other PpPALs, did not show activity. Also, MpPALs 1 and 2 are situated adjacent to MpPAL 3, which shows far higher catalytic efficiencies and is distinctly more versatile in terms of substrate susceptibility. On the other hand, a tendency of congregation for the more active isoforms is nonetheless visible.

Similar conclusions can be drawn from evaluating phylogenetic analyses of 4CL (see 3.7 and Figure 86). Here, the most immediate observation was the close approximation of the five 4CL isoforms of this work that showed the most significant activities, Mp4CL 3 and Pp4CLs 1 to 4, in one subclade. Also, Aa4CL 1 was located in closest proximity. The rest of the investigated isoforms of this work, Cb4CLs 1 and 2, Mp4CLs 1, 2 and 4, as well as Pp4CL 5 (none of which were distinctly active) were scattered all over the different subclades. It was also noticeable that 4CL isoforms were not as easily allocatable in terms of the different divisions, as opposed to PAL isoforms that could easily be attributed to angiosperms, gymnosperms etc.

It is probable that a connection between activeness in general and phylogenetic similarity exists, which would be a logical assumption, considering that certain conserved motifs play a crucial role in the catalytic activity of the investigated enzymes, especially for 4CL.

Fundamentally, the most efficacious isoforms of both, PAL and 4CL are clearly those that also share the most phylogenetic similarity.

4.5 Coherence and outlook

The studies described in this work have shown that the green alga *C. braunii* already possessed an array of meticulous traits that would later in evolution become absolutely paramount for the transition to land. It has been proven that this extant member of the charophytes is in fact capable of performing the first and putatively third enzymatic reaction in the phenylpropanoid pathway towards phenolic metabolites. Taking into consideration the fact that, to date, no findings of genes coding for C4H in *C. braunii* have emerged (Rieseberg et al. 2023), it is necessary to investigate further in this direction. It has to be noted that 4-coumaric acid, a precursor molecule for a great variety of specialized compounds, has been detected in a number of green algae and other algal species (Labeeuw et al. 2015). *t*-Cinnamic acid has even been detected in a number of prokaryotes, such as soil bacteria and in certain fungi (Niklas et al. 2017). There has been much debate about the role of charophyte algae as paraphyletic group to land plants (Lewis and McCourt 2004) and structural similarities to Mesostigmatophyceae have been highlighted (Melkonian 1989).

Taking into consideration that the Zygnematophyceae, among the charophyte algae, represent the closest living relative to all land plants (Zhou and Schwartzenberg 2020), it would be meaningful to investigate possible PAL and 4CL activities from members of this family, e.g., *Mesotaenium, Zygnema* or *Spirogyra*. On the basis of the investigations that put *C. braunii* in close phylogenetic proximity to *Nitella* spec. where lignin-like structures have been discovered within the cell walls (Ligrone et al. 2008; Schubert et al. 2018), it would make for a reasonable investigation into potential properties of *C. braunii* to synthesize comparable metabolites.

The activity and rather high affinity towards L-Phe of CbPAL 1 poses the question whether activity can be found in extant organisms that have evolved even much earlier in time, such as cyanobacteria. In these photosynthetically active bacteria, PALs have already been crystallized (Moffitt et al. 2007) implying a potential readiness for synthesizing phenylpropanoids. Regarding their capability of producing some of the most powerful toxins like the alkaloid anatoxin A (Méjean et al. 2014), it seems natural for cyanobacteria to produce specialized compounds from different metabolic pathways. Cyanobacteria are suggested to have been

contributors to the development of Archaeplastida in that they have provided the ability to conduct photosynthesis and transferred this trait onto eukaryotic organisms that would in turn later become photosynthetic eukaryotes. This thesis is corroborated by phylogenetic evidence linking the freshwater cyanobacterium *Gloeomargarita lithophora* to plastids and depicting it as phylogenetically close to glaucophytes, the first archaeplastid lineage to evolve (de Vries and Archibald 2017; Ponce-Toledo et al. 2017). Cyanobacteria have performed photosynthesis for at least 2.7 billion years (Badger and Price 2003) before this ability was conveyed via endosymbiosis into a unicellular eukaryote by incorporating the cyanobacterium, an event that gave rise to the emergence of eukaryotic algae approximately 1.5 billion years ago (Bachy et al. 2022). This particular eukaryote was, in light of energy metabolism, a facultative anaerobe already equipped with a mitochondrion (Martin et al. 2015). Cyanobacteria also play a role as the photosynthetic component and therefore supplier of energy in biological soil crusts, small but versatile microcosmoses that can incorporate all kinds of bacterial, algal or bryophytic organisms and that display abundant areas of biodiversity (Garcia-Pichel 2023). Soil crusts are prone to represent vital habitats in areas of extreme terrestrial conditions where the subsistence of vascular plants is limited by low water supply and hot temperatures. Under such conditions, soil crusts are distinguished by stabilization abilities and water retention properties, which shows remarkable adaptability (Mikhailyuk et al. 2015). Secondary metabolism is also present, however, the role of specialized compounds, for instance cyanobacterial toxins, has not been clarified exhaustively (Garcia-Pichel 2023). In this context, it would make for an interesting investigation to find out whether the phenylpropanoid pathway also plays a vital role in the adaptation and survival of bryophytes associated with soil crusts under extraordinary conditions.

After thorough examination, it can be stated that the investigated PALs in general behave in accordance with what would be expected from PALs derived from eukaryotic organisms. The kinetic parameters, when compared to those from already investigated PAL isoforms from different organisms, roughly represent the evolutionary line of development of Archaeplastida. The examined PALs from *C. braunii*, *M. polymorpha* and *P. patens* are very rigid in their substrate susceptibilities and most of them either only turn over L-Phe or transform it as their best accepted substrate. Considering the fact that the transformation of L-Phe to *t*-cinnamic acid is the first and therefore an especially important reaction in the general

phenylpropanoid pathway, it is logical that this reaction be highly conserved and selective for L-Phe.

The most unexpected result for PAL measurements is the complete absence of isoforms that turn over L-His, which had to be expected as the main substrate for most examined enzymes, due to the fact that the majority of enzymes, later turning out to be PALs, were annotated as HALs. After several different approaches and variations of putative cofactors, L-His could be conclusively excluded as substrate for all tested PALs and the initial annotation as HAL can likely be denied for all thus designated isoforms.

The investigated isoforms of 4CL mostly act in accordance with expectation. The five isoforms that show activity distinctly enough for kinetics turn over 4-coumaric and caffeic acids well and, proportionately to their respective efficaciousness, transform ferulic and cinnamic acids less efficiently.

The general tendency seems to be that enzymes derived from *P. patens* show analogous catalytic properties as well as affinities, and the differences in catalytic efficiencies (for the active isoforms) are in most cases not tremendous. At least the ranges are comparable in this organism. In *M. polymorpha*, on the other hand, it can be concluded that not only are both PAL and 4CL very differently equipped in terms of catalytic effectiveness, but some isoforms are almost on the point of inoperability while others belong to the most efficient enzymes derived from non-seed plants. This gives rise to the assumption that, in the case of *M. polymorpha*, one isoform might be representative for the entire function of the enzyme, catalyzing the respective reaction alone, the other isoforms not displaying relevant enough activity to be considered crucial.

Of course, all this can only be stated from the in vitro point of view. Reactivity of such low degree as seen for some of the PAL and 4CL isoforms in this work would probably not be salient under natural conditions, seeing how, in some cases, the substrate concentrations reached unnaturally high levels. It would make sense that reactions as crucial as those at the beginning of the phenylpropanoid pathway appear conserved by a number of different isoforms in order to safeguard the emergence of vital specialized compounds, but it is still possible that some organisms gaining this function possess only one isoform that is responsible for the particular reaction.

The precise inducements for the different isoforms being so inflexible in their choice of substrate can in the case of 4CL putatively be correlated to the high level of conservation of the amino acid residues most likely responsible for substrate specificity and the intactness of the box I and II motifs is also likely to play a vital role in this context (Stuible and Kombrink 2001). For PAL, it is more difficult to discern a possible reason for the even more narrow range of substrate acceptance since the only traceable constant within the sequence is the ASG motif, which, however, does not discriminate between different ligands. It would make sense to either generate the crystal structures for the different isoforms or perform simulations with an existing crystal structure as template and engage in docking processes in order to distinguish between the propensities of the enzymes. This would also help to provide insight into the sizes of the active sites of the isoforms of PAL and 4CL and identify steric hindrances, e.g., in the case of sinapic acid as a substrate for 4CL, where the sheer size of the molecule represents an obstacle to fit into the active site (Schneider et al. 2003).

It is a long way from the first reaction of the general phenylpropanoid pathway to the different lignin molecules and other specialized compounds further downstream, in the specialized phenylpropanoid pathway. With so many different enzymes catalysing various reactions and the distribution of this metabolic pathway stretching across a great variety of organisms, it is a great challenge to find evolutionary patterns and clarify all the enzymes' behaviour. Therefore, it is imperative to continuously search for connections and discover new aspects that might support further understanding of the role that phenylpropanoid derivatives might have played during streptophytic conquest of land. On the basis of the biochemical data presented in this work, comprehensive research into all of these characteristics as well as into the phylogenetic and catalytic correlations of the aqueous organisms and the first streptophytes to cultivate dry land might ensue.

5 Summary

Plant terrestrialization represents one of the most momentous events in palaeobotany, necessitating remarkable adaptability and resilience of organisms having been adjusted to marine conditions for millions of years. A most significant epiphenomenon along the way of land colonization was the emergence of new pathways and novel abilities within plant specialized metabolism that facilitated embryophytic subsistence on land. The phenylpropanoid pathway introduced an array of metabolic compounds and processes leading to crucial mechanisms of protection and self-preservation for plants, including lignification of cell walls, setup of water conducting elements and herbivore repellence (Agorio et al. 2024; Kunz et al. 2024). This pathway is divided into the general and the lignin-specific phenylpropanoid pathway (Labeeuw et al. 2015; Barros and Dixon 2020).

Phenylalanine ammonia-lyase (PAL), the first enzyme in the general phenylpropanoid pathway, catalyzes the reaction of the amino acid L-phenylalanine via elimination of ammonia to *t*-cinnamic acid (Barros and Dixon 2020). This imperative reaction has been found to be conserved in extant members of the embryophytes as well as in the green alga *Chara braunii*, a member of the Charophyceae. In *C. braunii*, the transformation of L-Phe occurs with high substrate specificity, yet low turnover, whereas in extant members of bryophytes, the versatility as well as catalytic efficaciousness of PAL increase. PALs derived from the liverwort *Marchantia polymorpha* and the moss *Physcomitrium patens* are capable of not only transforming L-Phe to a higher degree than CbPAL, but also possess affinity towards D-phenylalanine and L-tyrosine. Here, MpPAL 3 possesses with 28000 s^{-1} mol^{-1} l the highest catalytic efficiency for L-Phe, whereas in PpPALs 1 to 4 the catalytic efficiency varies between 5000 and 22000 s^{-1} mol^{-1} l. The catalytic efficiencies for D-Phe and L-Tyr are significantly lower (Schwarze and Petersen 2024). The transformation of L-Tyr leads directly to 4-coumaric acid, which is otherwise produced via cinnamic acid 4-hydroxylase (C4H) from *t*-cinnamic acid (Barros and Dixon 2020). Intrinsic TAL-activity of PAL thus represents an alternative pathway towards this particular product. The increase in catalytic efficiency with ongoing evolution, as well as the more widespread orientation towards different substrates might indicate that the investigated CbPAL could represent a less elaborated form of PAL with the putative capability of evolving into more efficient enzymes.

The third enzyme in the general phenylpropanoid pathway, 4-coumarate CoA-ligase (4CL), transforms 4-coumaric acid further into 4-coumaroyl-CoA. 4CL is capable of catalysing the same reaction with a number of different substrates, such as caffeic, ferulic, cinnamic, sinapic and 4-hydroxybenzoic acids (Schneider et al. 2003; Wohl and Petersen 2020). Several isoforms of 4CL have been discovered in all three investigated organisms. In *C. braunii*, activity is present, especially for caffeic acid, but not to a significantly high degree. In contrast to that, kinetic experiments could be performed for 4-coumaric, caffeic, ferulic and cinnamic acids with 4CL isoforms derived from *M. polymorpha* and *P. patens*. In *M. polymorpha*, the majority of catalytic functionality seems to be carried out by Mp4CL 3, being extraordinarily active and especially affine towards 4-coumaric acid (with a catalytic efficiency of more than 2.5 million s^{-1} mol^{-1} l). The other three isoforms display no substantial catalytic activities. In *P. patens,* four isoforms have been kinetically examined, revealing rather heterogenous catalytic behaviour. The most active isoform, Pp4CL 4, reached catalytic activities of appr. 55000 s^{-1} mol^{-1} l for 4-coumaric acid and appr. 29000 s^{-1} mol^{-1} l for caffeic acid, respectively (Schwarze and Petersen 2024).

The biochemical investigations of *C. braunii*, *M. polymorpha* and *P. patens* have shown that the metabolic toolkit for the production of specialized phenolic compounds necessary for plant survival on dry land is at least in some parts already present in extant members of streptophytes that are not yet part of the land plants. It has furthermore provided evidence that the variance and extent of these abilities have ostensibly increased as evolution continued. With the help of metabolites derived from the lignin-specific phenylpropanoid pathway, streptophytic organisms could unfold their potential to enhance vital characteristics, a development that contributed substantially to plant terrestrialization and further evolution on dry land.

6 Zusammenfassung

Die Terrestrialisierung von Pflanzen stellt eines der bedeutungsträchtigsten Ereignisse der Paläobotanik dar, welches den Organismen, die Millionen Jahre in marinen Bedingungen beheimatet waren, bemerkenswertes Anpassungsvermögen und Resilienz abverlangte. Ein signifikantes Begleitphänomen auf dem Weg zur Kolonisation des Landes war die Entstehung von neuartigen Stoffwechselwegen und Fähigkeiten innerhalb des pflanzlichen Sekundärmetabolismus, welche den Embryophyta das Überleben an Land erleichterten. Der Phenylpropanstoffwechsel führte eine Reihe neuer metabolischer Substanzen und Prozesse ein, die zu wichtigen Schutz- und Selbsterhaltungsmechanismen für Pflanzen führten, unter anderem Lignifizierung von Zellwänden, Etablierung von Wasserleitelementen und Abwehr von Herbivoren (Agorio et al. 2024; Kunz et al. 2024). Dieser Stoffwechselweg ist unterteilt in den allgemeinen sowie den Lignin-spezifischen Phenylpropanstoffwechsel (Labeeuw et al. 2015; Barros and Dixon 2020).

Phenylalanin Ammoniak-Lyase (PAL), das erste Enzym des allgemeinen Phenylpropanstoffwechsels, katalysiert die Reaktion der Aminosäure L-Phenylalanin über eine Abspaltung von Ammoniak zur *t*-Zimtsäure (Barros and Dixon 2020). Diese essentielle Reaktion zeichnet sich durch Konservierung sowohl in rezenten Embryophyta als auch in der Grünalge *Chara braunii* aus der Gruppe der Charophyceae, aus. In *C. braunii* findet die Umwandlung von L-Phe mit hoher Substratspezifität, jedoch geringem Umsatz statt, wohingegen sich in rezenten Bryophyta eine höhere Versatilität sowie katalytische Aktivität beobachten lassen. PALs aus dem Lebermoos *Marchantia polymorpha* sowie dem Laubmoos *Physcomitrium patens* sind nicht nur in der Lage, L-Phe in höherem Maße als CbPAL umzusetzen, sondern besitzen auch Affinität zu D-Phenylalanin und L-Tyrosin. Dabei kommt MpPAL 3 mit ca. 28000 s^{-1} mol^{-1} l für L-Phe die höchste katalytische Effizienz zu, während diese in PpPALs 1 bis 4 zwischen 5000 und 22000 s^{-1} mol^{-1} l liegt. Die katalytischen Effizienzen für D-Phe und L-Tyr sind deutlich geringer (Schwarze and Petersen 2024). Die Umwandlung von L-Tyr führt direkt zur 4-Cumarsäure, welche ansonsten über das Enzym Zimtsäure 4-Hydroxylase (C4H) aus *t*-Zimtsäure gebildet wird (Barros and Dixon 2020). Somit repräsentiert intrinsische TAL-Aktivität von PAL einen alternativen Stoffwechselweg hin zu diesem speziellen Produkt. Die Zunahme der katalytischen Effizienz mit fortschreitender Evolution sowie die breitere Orientierung hin zu

verschiedenen Substraten könnten implizieren, dass es sich bei der untersuchten CbPAL um eine weniger elaborierte Version von PAL, mit der möglichen Fähigkeit, sich in effizientere Enzyme weiterzuentwickeln, handelt.

Das dritte Enzym im allgemeinen Phenylpropanstoffwechsel, 4-Cumarat CoA-Ligase (4CL), wandelt 4-Cumarsäure weiter zu 4-Cumaroyl-CoA um. 4CL ist in der Lage, dieselbe Reaktion mit einer Reihe verschiedener Substrate zu katalysieren, unter anderem Kaffee-, Ferula-, Zimt-, Sinapin- und 4-Hydroxybenzoesäure (Schneider et al. 2003; Wohl and Petersen 2020). Verschiedene Isoformen von 4CL wurden in allen drei genannten Organismen gefunden. In *C. braunii* wurde Aktivität gefunden, insbesondere mit Kaffeesäure, allerdings nicht in besonders hohem Maße. Demgegenüber konnten mit 4CL-Isoformen aus *M. polymorpha* und *P. patens* kinetische Experimente mit 4-Cumar-, Kaffee-, Ferula- und Zimtsäure durchgeführt werden. Im Fall von *M. polymorpha* scheint Mp4CL 3 für den Großteil der katalytischen Funktion verantwortlich zu sein, wobei es sich hier um ein ausgesprochen aktives Enzym mit besonders hoher Affinität zu 4-Cumarsäure (katalytische Effizienz von mehr als 2.5 Millionen s^{-1} mol^{-1} l) handelt. Für die anderen drei Isoformen fand sich keine signifikante katalytische Aktivität. Vier Isoformen aus *P. patens* wurden kinetisch untersucht, wobei sich recht unterschiedliche katalytische Verhaltensweisen abzeichneten. Die aktivste Isoform, Pp4CL 4, erreichte katalytische Effizienzen von 55000 s^{-1} mol^{-1} l für 4-Cumarsäure und 29000 s^{-1} mol^{-1} l für Kaffeesäure (Schwarze and Petersen 2024).

Die biochemischen Untersuchungen für *C. braunii*, *M. polymorpha* and *P. patens* haben gezeigt, dass zumindest ein Teil des metabolischen Bausatzes zur Produktion der zum Überleben der Pflanzen an Land benötigten Sekundärmetaboliten bereits in rezenten Streptophyta vorhanden ist, welche noch nicht zu den Landpflanzen gezählt werden. Darüber hinaus haben die Untersuchungen dazu beigetragen, zu zeigen, dass sich Versatilität und Umfang dieser Fähigkeiten augenscheinlich im Laufe der Evolution erhöht haben. Mit Hilfe von Sekundärmetaboliten aus dem Lignin-spezifischen Phenylpropanstoffwechsel konnten streptophytische Organismen ihr Potential, überlebenswichtige Eigenschaften zu verbessern, frei entfalten, was in erheblichem Maße zur Terrestrialisierung von Pflanzen und ihrer weiteren Entwicklung an Land beitrug.

7 References

Adrian M, Jeandet P, Veneau J, Weston LA, Bessis R (1997) Biological Activity of Resveratrol, a Stilbenic Compound from Grapevines, Against *Botrytis cinerea*, the Causal Agent for Gray Mold. J Chem Ecol 23:1689–1702. https://doi.org/10.1023/B:JOEC.0000006444.79951.75

Agorio A, Mena E, Rockenbach MF, Ponce De León I (2024) The evolution of plant responses underlying specialized metabolism in host-pathogen interactions. Phil Trans R Soc Lond B 379:20230370. https://doi.org/10.1098/rstb.2023.0370

Altman FP (1976) Tetrazolium salts and formazans. Prog Histochem Cytochem 9:1–56. https://doi.org/10.1016/S0079-6336(76)80015-0

Andrews M, McInroy S, Raven JA (1984) Culture of *Chara hispida*. Brit Phycol J 19:277–280. https://doi.org/10.1080/00071618400650291

Appert C, Logemann E, Hahlbrock K, Schmid J, Amrhein N (1994) Structural and catalytic properties of the four phenylalanine ammonia-lyase isoenzymes from parsley (*Petroselinum crispum* Nym.). Eur J Biochem 225:491–499. https://doi.org/10.1111/j.1432-1033.1994.00491.x

Bachy C, Wittmers F, Muschiol J, Hamilton M, Henrissat B, Worden AZ (2022) The Land-Sea Connection: Insights Into the Plant Lineage from a Green Algal Perspective. Annu Rev Plant Biol 73:585–616. https://doi.org/10.1146/annurev-arplant-071921-100530

Badger MR, Price GD (2003) CO2 concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution. J Exp Bot 54:609–622. https://doi.org/10.1093/jxb/erg076

Baedeker M, Schulz GE (2002) Autocatalytic peptide cyclization during chain folding of histidine ammonia-lyase. Structure 10:61–67. https://doi.org/10.1016/S0969-2126(01)00692-X

Bagal UR, Leebens-Mack JH, Lorenz WW, Dean JFD (2012) The phenylalanine ammonia lyase (PAL) gene family shows a gymnosperm-specific lineage. BMC Genomics 13 Suppl 3:S1. https://doi.org/10.1186/1471-2164-13-S3-S1

Barbehenn RV, Constabel PC (2011) Tannins in plant-herbivore interactions. Phytochemistry 72:1551–1565. https://doi.org/10.1016/j.phytochem.2011.01.040

Barros J, Dixon RA (2020) Plant Phenylalanine/Tyrosine Ammonia-lyases. Trends Plant Sci 25:66–79. https://doi.org/10.1016/j.tplants.2019.09.011

Bartsch S, Bornscheuer UT (2009) Einfluss einer einzelnen Aminosäure auf den Reaktionsmechanismus von Ammonium-Lyasen und -Mutasen. Angew Chem-Ger Edit 121:3412–3415. https://doi.org/10.1002/ange.200900337

Bata Z, Molnár Z, Madaras E, Molnár B, Sánta-Bell E, Varga A, Leveles I, Qian R, Hammerschmidt F, Paizs C, Vértessy BG, Poppe L (2021) Substrate Tunnel Engineering Aided by X-ray Crystallography and Functional Dynamics Swaps the Function of MIO-Enzymes. ACS Catal 11:4538–4549. https://doi.org/10.1021/acscatal.1c00266

Becker B, Feng X, Yin Y, Holzinger A (2020) Desiccation tolerance in streptophyte algae and the algae to land plant transition: evolution of LEA and MIP protein families within the Viridiplantae. J Exp Bot 71:3270–3278. https://doi.org/10.1093/jxb/eraa105

Beike AK, Stackelberg M von, Schallenberg-Rüdinger M, Hanke ST, Follo M, Quandt D, McDaniel SF, Reski R, Tan BC, Rensing SA (2014) Molecular evidence for convergent evolution and

allopolyploid speciation within the *Physcomitrium-Physcomitrella* species complex. BMC Evol Biol 14:158. https://doi.org/10.1186/1471-2148-14-158

Beilby MJ (2019) *Chara braunii* genome: a new resource for plant electrophysiology. Biophys Rev 11:235–239. https://doi.org/10.1007/s12551-019-00512-7

Bengtson S, Sallstedt T, Belivanova V, Whitehouse M (2017) Three-dimensional preservation of cellular and subcellular structures suggests 1.6 billion-year-old crown-group red algae. PLoS Biol 15:e2000735. https://doi.org/10.1371/journal.pbio.2000735

Berner M, Krug D, Bihlmaier C, Vente A, Müller R, Bechthold A (2006) Genes and enzymes involved in caffeic acid biosynthesis in the actinomycete *Saccharothrix espanaensis*. J Bacteriol 188:2666–2673. https://doi.org/10.1128/jb.188.7.2666-2673.2006

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

Bezanson GS, Desaty D, Emes AV, Vining LC (1970) Biosynthesis of cinnamamide and detection of phenylalanine ammonia-lyase in *Streptomyces verticillatus*. Can J Microbiol 16:147–151. https://doi.org/10.1139/m70-026

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. https://doi.org/10.1111/j.1432-1033.1993.tb17794.x

Bierenbroodspot MJ, Darienko T, Vries S de, Fürst-Jansen JMR, Buschmann H, Pröschold T, Irisarri I, Vries J de (2024) Phylogenomic insights into the first multicellular streptophyte. Curr Biol 34:670-681.e7. https://doi.org/10.1016/j.cub.2023.12.070

Bisswanger H (ed) (2015) Enzyme Struktur, Kinetik und Anwendungen. WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Boerjan W, Ralph J, Baucher M (2003) Lignin biosynthesis. Annu Rev Plant Biol 54:519–546. https://doi.org/10.1146/annurev.arplant.54.031902.134938

Bourgaud F, Hehn A, Larbat R, Doerper S, Gontier E, Kellner S, Matern U (2006) Biosynthesis of coumarins in plants: a major pathway still to be unravelled for cytochrome P450 enzymes. Phytochem Rev 5:293–308. https://doi.org/10.1007/s11101-006-9040-2

Bowman JL, Kohchi T, Yamato KT et al (2017) Insights into Land Plant Evolution Garnered from the *Marchantia polymorpha* Genome. Cell 171:287-304.e15. https://doi.org/10.1016/j.cell.2017.09.030

Brack Y, Sun C, Yi D, Bornscheuer UT (2024) Systematic Analysis of the MIO-forming Residues of Aromatic Ammonia Lyases. Chembiochem 25:e202400016. https://doi.org/10.1002/cbic.202400016

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. https://doi.org/10.1016/0003-2697(76)90527-3

Brand LM, Harper AE (1976) Histidine ammonia-lyase from rat liver. Purification, properties, and inhibition by substrate analogues. Biochemistry 15:1814–1821. https://doi.org/10.1021/bi00654a005

Bryksin AV, Matsumura I (2010) Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. Biotechniques 48:463–465. https://doi.org/10.2144/000113418

Busch T, Petersen M (2021) Identification and biochemical characterisation of tyrosine aminotransferase from *Anthoceros agrestis* unveils the conceivable entry point into rosmarinic acid biosynthesis in hornworts. Planta 253:98. https://doi.org/10.1007/s00425-021-03623-2

Calabrese JC, Jordan DB, Boodhoo A, Sariaslani S, Vannelli T (2004) Crystal structure of phenylalanine ammonia lyase: multiple helix dipoles implicated in catalysis. Biochemistry 43:11403–11416. https://doi.org/10.1021/bi049053

Cannon Homaei S, Barone H, Kleppe R, Betari N, Reif A, Haavik J (2022) ADHD symptoms in neurometabolic diseases: Underlying mechanisms and clinical implications. Neurosci Biobehav Rev 132:838–856. https://doi.org/10.1016/j.neubiorev.2021.11.012

Castañeda-Ovando A, Pacheco-Hernández MdL, Páez-Hernández ME, Rodríguez JA, Galán-Vidal CA (2009) Chemical studies of anthocyanins: A review. Food Chem 113:859–871. https://doi.org/10.1016/j.foodchem.2008.09.001

Cenci U, Nitschke F, Steup M, Minassian BA, Colleoni C, Ball SG (2014) Transition from glycogen to starch metabolism in Archaeplastida. Trends Plant Sci 19:18–28. https://doi.org/10.1016/j.tplants.2013.08.004

Cho M-H, Corea ORA, Yang H, Bedgar DL, Laskar DD, Anterola AM, Moog-Anterola FA, Hood RL, Kohalmi SE, Bernards MA, Kang C, Davin LB, Lewis NG (2007) Phenylalanine biosynthesis in *Arabidopsis thaliana*. Identification and characterization of arogenate dehydratases. J Biol Chem 282:30827–30835. https://doi.org/10.1074/jbc.M702662200

Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 162:156–159. https://doi.org/10.1016/0003-2697(87)90021-2

Chung BK-S, Lee D-Y (2012) Computational codon optimization of synthetic gene for protein expression. BMC Syst Biol 6:134. https://doi.org/10.1186/1752-0509-6-134

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

Cochrane FC, Davin LB, Lewis NG (2004) The *Arabidopsis* phenylalanine ammonia lyase gene family: kinetic characterization of the four PAL isoforms. Phytochemistry 65:1557–1564. https://doi.org/10.1016/j.phytochem.2004.05.006

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. https://doi.org/10.1016/S0969-2126(96)00033-0

Cornish-Bowden A, Eisenthal R (1978) Estimation of Michaelis constant and maximum velocity from the direct linear plot. Biochim Biophys Acta 523:268–272. https://doi.org/10.1016/0005-2744(78)90030-X

Corti B (ed) (1774) Osservazioni microscopiche sulla tremella, e sulla circolazione del fuido in una pianta acquajuola. Giuseppe Rocchi, Lucca

Cotton RG, Gibson F (1965) The biosynthesis of phenylalanine and tyrosine; enzymes converting chorismic acid into prephenic acid and their relationships to prephenate dehydratase and prephenate dehydrogenase. Biochim Biophys Acta 100:76–88. https://doi.org/10.1016/0304-4165(65)90429-0

Cove DJ, Perroud P-F, Charron AJ, McDaniel SF, Khandelwal A, Quatrano RS (2009a) Culturing the moss *Physcomitrella patens*. Cold Spring Harb Protoc 2009:pdb.prot5136. https://doi.org/10.1101/pdb.prot5136

Cukovic D, Ehlting J, VanZiffle JA, Douglas CJ (2001) Structure and evolution of 4-coumarate:coenzyme A ligase (4CL) gene families. Biol Chem 382:645–654. https://doi.org/10.1515/BC.2001.076

Dadras A, Rieseberg TP, Zegers JMS, Fürst-Jansen JMR, Irisarri I, Vries J de, Vries S de (2023) Accessible versatility underpins the deep evolution of plant specialized metabolism. Phytochem Rev. https://doi.org/10.1007/s11101-023-09863-2

de Vries J, Gould SB (2018) The monoplastidic bottleneck in algae and plant evolution. J Cell Sci 131. https://doi.org/10.1242/jcs.203414

de Vries J, Archibald JM (2018) Plant evolution: landmarks on the path to terrestrial life. New Phytol 217:1428–1434. https://doi.org/10.1111/nph.14975

de Vries J, Archibald JM (2017) Endosymbiosis: Did Plastids Evolve from a Freshwater Cyanobacterium? Curr Biol 27:R103-R105. https://doi.org/10.1016/j.cub.2016.12.006

de Vries J, de Vries S, Slamovits CH, Rose LE, Archibald JM (2017) How Embryophytic is the Biosynthesis of Phenylpropanoids and their Derivatives in Streptophyte Algae? Plant Cell Physiol 58:934–945. https://doi.org/10.1093/pcp/pcx037

de Vries S, Fürst-Jansen JMR, Irisarri I, Dhabalia Ashok A, Ischebeck T, Feussner K, Abreu IN, Petersen M, Feussner I, de Vries J (2021) The evolution of the phenylpropanoid pathway entailed pronounced radiations and divergences of enzyme families. Plant J 107:975–1002. https://doi.org/10.1111/tpj.15387

Decker EL, Reski R (2008) Current achievements in the production of complex biopharmaceuticals with moss bioreactors. Bioprocess Biosyst Eng 31:3–9. https://doi.org/10.1007/s00449-007-0151-y

DeGabriel JL, Moore BD, Foley WJ, Johnson CN (2009) The effects of plant defensive chemistry on nutrient availability predict reproductive success in a mammal. Ecology 90:711–719. https://doi.org/10.1890/08-0940.1

Dixon RA, Achnine L, Kota P, Liu C-J, Reddy MSS, Wang L (2002) The phenylpropanoid pathway and plant defence-a genomics perspective. Mol Plant Pathol 3:371–390. https://doi.org/10.1046/j.1364-3703.2002.00131.x

Domozych DS, Bagdan K (2022) The cell biology of charophytes: Exploring the past and models for the future. Plant Physiol 190:1588–1608. https://doi.org/10.1093/plphys/kiac390

Edwards D (2003) Xylem in early tracheophytes. Plant Cell Environ 26:57–72. https://doi.org/10.1046/j.1365-3040.2003.00878.x

Edwards D, LI C-S, Raven JA (2006) Tracheids in an early vascular plant: a tale of two branches. Bot J Linn Soc 150:115–130. https://doi.org/10.1111/j.1095-8339.2006.00450.x

Ehlting J, Shin JJ, Douglas CJ (2001) Identification of 4-coumarate:coenzyme A ligase (4CL) substrate recognition domains. Plant J 27:455–465. https://doi.org/10.1046/j.1365-313X.2001.01122.x

Ellis BE, Towers GH (1970) Biogenesis of rosmarinic acid in Mentha. Biochem J 118:291–297. https://doi.org/10.1042/bj1180291

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. https://doi.org/10.1186/1745-6150-4-7

Ernst L, Wohl J, Bauerbach E, Petersen M (2022) Hydroxycinnamoyltransferase and CYP98 in phenolic metabolism in the rosmarinic acid-producing hornwort *Anthoceros agrestis*. Planta 255:75. https://doi.org/10.1007/s00425-022-03856-9

Espiñeira JM, Novo Uzal E, Gómez Ros LV, Carrión JS, Merino F, Ros Barceló A, Pomar F (2011) Distribution of lignin monomers and the evolution of lignification among lower plants. Plant Biol (Stuttg) 13:59–68. https://doi.org/10.1111/j.1438-8677.2010.00345.x

Fu H, Liang Y, Zhong X, Pan Z, Huang L, Zhang H, Xu Y, Zhou W, Liu Z (2020) Codon optimization with deep learning to enhance protein expression. Sci Rep 10:17617. https://doi.org/10.1038/s41598-020-74091-z

Fürst-Jansen JMR, Vries S de, Vries J de (2020) Evo-physio: on stress responses and the earliest land plants. J Exp Bot 71:3254–3269. https://doi.org/10.1093/jxb/eraa007

Gamborg OL, Miller RA, Ojima K (1968) Nutrient requirements of suspension cultures of soybean root cells. Exp Cell Res 50:151–158. https://doi.org/10.1016/0014-4827(68)90403-5

Gao S, Liu X-Y, Ni R, Fu J, Tan H, Cheng A-X, Lou H-X (2024) Molecular cloning and functional analysis of 4-coumarate: CoA ligases from *Marchantia paleacea* and their roles in lignin and flavanone biosynthesis. PLoS One 19:e0296079. https://doi.org/10.1371/journal.pone.0296079

Gao S, Yu H-N, Xu R-X, Cheng A-X, Lou H-X (2015) Cloning and functional characterization of a 4-coumarate CoA ligase from liverwort *Plagiochasma appendiculatum*. Phytochemistry 111:48–58. https://doi.org/10.1016/j.phytochem.2014.12.017

Garcia-Pichel F (2023) The Microbiology of Biological Soil Crusts. Annu Rev Microbiol 77:149–171. https://doi.org/10.1146/annurev-micro-032521-015202

Gocht M, Marahiel MA (1994) Analysis of core sequences in the D-Phe activating domain of the multifunctional peptide synthetase TycA by site-directed mutagenesis. J Bacteriol 176:2654–2662. https://doi.org/10.1128/jb.176.9.2654-2662.1994

Gulick AM (2009) Conformational dynamics in the Acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase. ACS Chem Biol 4:811–827. https://doi.org/10.1021/cb900156h

Guo D-M, Ran J-H, Wang X-Q (2010) Evolution of the Cinnamyl/Sinapyl Alcohol Dehydrogenase (CAD/SAD) gene family: the emergence of real lignin is associated with the origin of Bona Fide CAD. J Mol Evol 71:202–218. https://doi.org/10.1007/s00239-010-9378-3

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

Hart PH, Norval M, Byrne SN, Rhodes LE (2019) Exposure to Ultraviolet Radiation in the Modulation of Human Diseases. Annu Rev Pathol 14:55–81. https://doi.org/10.1146/annurev-pathmechdis-012418-012809

Hedges SB (2002) The origin and evolution of model organisms. Nat Rev Genet 3:838–849. https://doi.org/10.1038/nrg929

Hellio C, Veron B, Le Gal Y (2004) Amino acid utilization by *Chlamydomonas reinhardtii*: specific study of histidine. Plant Physiol Biochem 42:257–264. https://doi.org/10.1016/j.plaphy.2003.12.005

Herrmann KM, Weaver LM (1999) The shikimate pathway. Annu Rev Plant Physiol Plant Mol Biol 50:473–503. https://doi.org/10.1146/annurev.arplant.50.1.473

Heß D, Holzhausen A, Hess WR (2023) Insight into the nodal cells transcriptome of the streptophyte green alga *Chara braunii* S276. Physiol Plant 175:e14025. https://doi.org/10.1111/ppl.14025

Hess V, Vitt S, Müller V (2011) A caffeyl-coenzyme A synthetase initiates caffeate activation prior to caffeate reduction in the acetogenic bacterium *Acetobacterium woodii*. J Bacteriol 193:971–978. https://doi.org/10.1128/jb.01126-10

Hill AM, Thompson BL, Harris JP, Segret R (2003) Investigation of the early stages in soraphen A biosynthesis. Chem Commun (Camb) 1358–1359. https://doi.org/10.1039/B303542P

Hitl M, Kladar N, Gavarić N, Božin B (2021) Rosmarinic Acid-Human Pharmacokinetics and Health Benefits. Planta Med 87:273–282. https://doi.org/10.1055/a-1301-8648

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

Holzhausen A (2024) What we really know about the dormancy, reproduction, germination and cultivation of charophytes (Characeae). OE 9. https://doi.org/10.3897/oneeco.9.e117655

Holzhausen A, Stingl N, Rieth S, Kühn C, Schubert H, Rensing SA (2022) Establishment and optimization of a new model organism to study early land plant evolution: Germination, cultivation and oospore variation of *Chara braunii* Gmelin, 1826. Front Plant Sci 13:987741. https://doi.org/10.3389/fpls.2022.987741

Hu Y, Gai Y, Yin L, Wang X, Feng C, Feng L, Li D, Jiang X-N, Wang D-C (2010) Crystal structures of a *Populus tomentosa* 4-coumarate:CoA ligase shed light on its enzymatic mechanisms. Plant Cell 22:3093–3104. https://doi.org/10.1105/tpc.109.072652

Huang W-J, Wu C-L, Lin C-W, Chi L-L, Chen P-Y, Chiu C-J, Huang C-Y, Chen C-N (2010) Marchantin A, a cyclic bis(bibenzyl ether), isolated from the liverwort *Marchantia emarginata* subsp. *tosana* induces apoptosis in human MCF-7 breast cancer cells. Cancer Lett 291:108–119. https://doi.org/10.1016/j.canlet.2009.10.006

Hydery T, Coppenrath VA (2019) A Comprehensive Review of Pegvaliase, an Enzyme Substitution Therapy for the Treatment of Phenylketonuria. Drug Target Insights 13:1177392819857089. https://doi.org/10.1177/1177392819857089

Jablonski D, Shubin NH (2015) The future of the fossil record: Paleontology in the 21st century. Proc Natl Acad Sci U S A 112:4852–4858. https://doi.org/10.1073/pnas.1505146112

Jackson C, Clayden S, Reyes-Prieto A (2015) The Glaucophyta: the blue-green plants in a nutshell. Acta Soc Bot Pol 84:149–165. https://doi.org/10.5586/ASBP.2015.020

Jakubska-Busse A, Śliwiński M, Kobyłka M (2013) Identification of bioactive components of essential oils in *Heracleum sosnowskyi* and *Heracleum mantegazzianum* (Apiaceae). Arch biol sci (Beogr) 65:877–883. https://doi.org/10.2298/ABS1303877J

Jensen S, Omarsdottir S, Bwalya AG, Nielsen MA, Tasdemir D, Olafsdottir ES (2012) Marchantin A, a macrocyclic bisbibenzyl ether, isolated from the liverwort *Marchantia polymorpha*, inhibits protozoal growth in vitro. Phytomedicine 19:1191–1195. https://doi.org/10.1016/j.phymed.2012.07.011

Jun S-Y, Sattler SA, Cortez GS, Vermerris W, Sattler SE, Kang C (2018) Biochemical and Structural Analysis of Substrate Specificity of a Phenylalanine Ammonia-Lyase. Plant Physiol 176:1452–1468. https://doi.org/10.1104/pp.17.01608

Jung E, Zamir LO, Jensen RA (1986) Chloroplasts of higher plants synthesize L-phenylalanine via L-arogenate. Proc Natl Acad Sci U S A 83:7231–7235. https://doi.org/10.1073/pnas.83.19.7231

Kamel MY, Maksoud SA (1978) Co-Factor Requirements and Factors Affecting L-Histidine Ammonia-Lease Activity in *Vicia faba*. Z Pflanzenphysiol 88:255–262. https://doi.org/10.1016/S0044-328X(78)80247-5

Karol KG, McCourt RM, Cimino MT, Delwiche CF (2001) The closest living relatives of land plants. Science 294:2351–2353. https://doi.org/10.1126/science.1065156

Kato H, Ishizaki K, Kouno M, Shirakawa M, Bowman JL, Nishihama R, Kohchi T (2015) Auxin-Mediated Transcriptional System with a Minimal Set of Components Is Critical for Morphogenesis through the Life Cycle in *Marchantia polymorpha*. PLoS Genet 11:e1005084. https://doi.org/10.1371/journal.pgen.1005084

Klee CB (1972) Metal Activation of Histidine Ammonia-Lyase. J Biol Chem 247:1398–1406. https://doi.org/10.1016/S0021-9258(19)45572-5

Knosp S, Kriegshauser L, Tatsumi K, Malherbe L, Wiedemann G, Bakan B, Kohchi T, Reski R, Renault H (2023) An ancient role for the CYP73 gene family in t -cinnamic acid 4-hydroxylation, phenylpropanoid biosynthesis and embryophyte development. https://doi.org/10.1101/2023.08.20.551634

Koenen M, Rüther U, Müller-Hill B (1982) Immunoenzymatic detection of expressed gene fragments cloned in the lac Z gene of *E. coli*. EMBO J 1:509–512. https://doi.org/10.1002/j.1460-2075.1982.tb01199.x

Koukol J, Conn EE (1961) The Metabolism of Aromatic Compounds in Higher Plants. J Biol Chem 236:2692–2698. https://doi.org/10.1016/S0021-9258(19)61721-7

Kunz CF, Vries S de, Vries J de (2024) Plant terrestrialization: an environmental pull on the evolution of multi-sourced streptophyte phenolics. Phil Trans R Soc Lond B 379:20230358. https://doi.org/10.1098/rstb.2023.0358

Kurtović K, Schmidt V, Nehasilová M, Vosolsobě S, Petrášek J (2024) Rediscovering *Chara* as a model organism for molecular and evo-devo studies. Protoplasma 261:183–196. https://doi.org/10.1007/s00709-023-01900-3

Kyndt JA, Meyer TE, Cusanovich MA, van Beeumen JJ (2002) Characterization of a bacterial tyrosine ammonia lyase, a biosynthetic enzyme for the photoactive yellow protein. FEBS Lett 512:240–244. https://doi.org/10.1016/S0014-5793(02)02272-X

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

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

Lavhale SG, Kalunke RM, Giri AP (2018) Structural, functional and evolutionary diversity of 4-coumarate-CoA ligase in plants. Planta 248:1063–1078. https://doi.org/10.1007/s00425-018-2965-z

Lee HI, León J, Raskin I (1995) Biosynthesis and metabolism of salicylic acid. Proc Natl Acad Sci U S A 92:4076–4079. https://doi.org/10.1073/pnas.92.10.4076

Lefevere H, Bauters L, Gheysen G (2020) Salicylic Acid Biosynthesis in Plants. Front Plant Sci 11:338. https://doi.org/10.3389/fpls.2020.00338

Lessard JC (2013) Growth media for *E. coli*. Methods Enzymol 533:181–189. https://doi.org/10.1016/B978-0-12-420067-8.00011-8

Lewis LA, McCourt RM (2004) Green algae and the origin of land plants. Am J Bot 91:1535–1556. https://doi.org/10.3732/ajb.91.10.1535

Li Z, Nair SK (2015) Structural Basis for Specificity and Flexibility in a Plant 4-Coumarate:CoA Ligase. Structure 23:2032–2042. https://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. https://doi.org/10.1007/s00606-007-0617-z

Litvinenko VI, Popova TP, Simonjan AV, Zoz IG, Sokolov VS (1975) "Gerbstoffe" Und Oxyzimtsäureabkömmlinge in Labiaten. Planta Med 27:372–380. https://doi.org/10.1055/s-0028-1097817

Liu W, Feng Y, Yu S, Fan Z, Li X, Li J, Yin H (2021) The Flavonoid Biosynthesis Network in Plants. Int J Mol Sci 22. https://doi.org/10.3390/ijms222312824

Lovelock SL, Lloyd RC, Turner NJ (2014) Phenylalanine Ammonia Lyase Catalyzed Synthesis of Amino Acids by an MIO-Cofactor Independent Pathway. Angew Chemie 126:4740–4744. https://doi.org/10.1002/ange.201311061

Mahmood T, Yang P-C (2012) Western blot: technique, theory, and trouble shooting. N Am J Med Sci 4:429–434. https://doi.org/10.4103/1947-2714.100998

Martin WF, Garg S, Zimorski V (2015) Endosymbiotic theories for eukaryote origin. Phil Trans R Soc Lond B 370:20140330. https://doi.org/10.1098/rstb.2014.0330

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. https://doi.org/10.1016/j.cub.2008.12.031

Matern U (1991) Coumarins and other phenylpropanoid compounds in the defense response of plant cells. Planta Med 57:S15-20. https://doi.org/10.1055/s-2006-960224

Medina R, Liu Y, Li-Song W, Shuiliang G, Hylander K, Goffinet B (2015) DNA based revised geographic circumscription of species of *Physcomitrella* s.l. (Funariaceae): *P. patens* new to East Asia and *P. magdalenae* new to East Africa. The Bryologist 118:22. https://doi.org/10.1639/0007-2745-118.1.022

Méjean A, Paci G, Gautier V, Ploux O (2014) Biosynthesis of anatoxin-a and analogues (anatoxins) in cyanobacteria. Toxicon 91:15–22. https://doi.org/10.1016/j.toxicon.2014.07.016

Melkonian M (1989) Flagellar apparatus ultrastructure in *Mesostigma viride* (Prasinophyceae). Plant Syst Evol 164:93–122. https://doi.org/10.1007/BF00940432

Mikhailyuk T, Glaser K, Holzinger A, Karsten U (2015) Biodiversity of Klebsormidium (streptophyta) from alpine biological soil crusts (alps, tyrol, Austria, and Italy). J Phycol 51:750–767. https://doi.org/10.1111/jpy.12316

Moffitt MC, Louie GV, Bowman ME, Pence J, Noel JP, Moore BS (2007) Discovery of two cyanobacterial phenylalanine ammonia lyases: kinetic and structural characterization. Biochemistry 46:1004–1012. https://doi.org/10.1021/bi061774g

Morris JL, Puttick MN, Clark JW, Edwards D, Kenrick P, Pressel S, Wellman CH, Yang Z, Schneider H, Donoghue PCJ (2018) The timescale of early land plant evolution. Proc Natl Acad Sci U S A 115:E2274-E2283. https://doi.org/10.1073/pnas.1719588115

Mubarak A, Bondonno CP, Liu AH, Considine MJ, Rich L, Mas E, Croft KD, Hodgson JM (2012) Acute effects of chlorogenic acid on nitric oxide status, endothelial function, and blood pressure in healthy volunteers: a randomized trial. J Agric Food Chem 60:9130–9136. https://doi.org/10.1021/jf303440j

Nabors MW (ed) (2007) Botanik. Pearson Studium, München

Nagy EZA, Tork SD, Lang PA, Filip A, Irimie FD, Poppe L, Toşa MI, Schofield CJ, Brem J, Paizs C, Bencze LC (2019) Mapping the Hydrophobic Substrate Binding Site of Phenylalanine Ammonia-Lyase from *Petroselinum crispum*. ACS Catal. 9:8825–8834. https://doi.org/10.1021/acscatal.9b02108

Naramoto S, Hata Y, Fujita T, Kyozuka J (2022) The bryophytes *Physcomitrium patens* and *Marchantia polymorpha* as model systems for studying evolutionary cell and developmental biology in plants. Plant Cell 34:228–246. https://doi.org/10.1093/plcell/koab218

Niggeweg R, Michael AJ, Martin C (2004) Engineering plants with increased levels of the antioxidant chlorogenic acid. Nat Biotechnol 22:746–754. https://doi.org/10.1038/nbt966

Niklas KJ, Cobb ED, Matas AJ (2017) The evolution of hydrophobic cell wall biopolymers: from algae to angiosperms. J Exp Bot 68:5261–5269. https://doi.org/10.1093/jxb/erx215

Nishiyama T, Sakayama H, Vries J de et al (2018) The *Chara* Genome: Secondary Complexity and Implications for Plant Terrestrialization. Cell 174:448-464.e24. https://doi.org/10.1016/j.cell.2018.06.033

Otero-Blanca A, Pérez-Llano Y, Reboledo-Blanco G, Lira-Ruan V, Padilla-Chacon D, Folch-Mallol JL, Del Sánchez-Carbente MR, Ponce De León I, Batista-García RA (2021) *Physcomitrium patens* Infection by Colletotrichum gloeosporioides: Understanding the Fungal-Bryophyte Interaction by Microscopy, Phenomics and RNA Sequencing. J Fungi (Basel) 7. https://doi.org/10.3390/jof7080677

Peng Y, Yang J, Li X, Zhang Y (2021) Salicylic Acid: Biosynthesis and Signaling. Annu Rev Plant Biol 72:761–791. https://doi.org/10.1146/annurev-arplant-081320-092855

Petersen M (2013) Rosmarinic acid: new aspects. Phytochem Rev 12:207–227. https://doi.org/10.1007/s11101-013-9282-8

Petersen M, Simmonds MSJ (2003) Rosmarinic acid. Phytochemistry 62:121–125. https://doi.org/10.1016/S0031-9422(02)00513-7

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. https://doi.org/10.1515/znc-1988-7-804

Pezeshki S, Warmbier I, Busch T, Bauerbach E, Szövenyi P, Petersen M (2022) The first step into phenolic metabolism in the hornwort *Anthoceros agrestis*: molecular and biochemical characterization of two phenylalanine ammonia-lyase isoforms. Planta 256:33. https://doi.org/10.1007/s00425-022-03944-w

Pfeifer L, Mueller K-K, Utermöhlen J, Erdt F, Zehge JBJ, Schubert H, Classen B (2023) The cell walls of different Chara species are characterized by branched galactans rich in 3-O-methylgalactose and absence of AGPs. Physiol Plant 175:e13989. https://doi.org/10.1111/ppl.13989

Pfeifer L, Mueller K-K, Classen B (2022) The cell wall of hornworts and liverworts: innovations in early land plant evolution? J Exp Bot 73:4454–4472. https://doi.org/10.1093/jxb/erac157

Ponce De León I (2024) Evolution of immunity networks across embryophytes. Curr Opin Plant Biol 77:102450. https://doi.org/10.1016/j.pbi.2023.102450

Ponce-Toledo RI, Deschamps P, López-García P, Zivanovic Y, Benzerara K, Moreira D (2017) An Early-Branching Freshwater Cyanobacterium at the Origin of Plastids. Curr Biol 27:386–391. https://doi.org/10.1016/j.cub.2016.11.056

Poppe L, Rétey J (2005) Friedel-Crafts-type mechanism for the enzymatic elimination of ammonia from histidine and phenylalanine. Angew Chem Int Ed Engl 44:3668–3688. https://doi.org/10.1002/anie.200461377

Prigge MJ, Bezanilla M (2010) Evolutionary crossroads in developmental biology: *Physcomitrella patens*. Development 137:3535–3543. https://doi.org/10.1242/dev.049023

Ralph J, Lundquist K, Brunow G, Lu F, Kim H, Schatz PF, Marita JM, Hatfield RD, Ralph SA, Christensen JH, Boerjan W (2004) Lignins: Natural polymers from oxidative coupling of 4-hydroxyphenyl-propanoids. Phytochem Rev 3:29–60. https://doi.org/10.1023/B:PHYT.0000047809.65444.a4

Rao G, Pan X, Xu F, Zhang Y, Cao S, Jiang X, Lu H (2015) Divergent and Overlapping Function of Five 4-Coumarate/Coenzyme A Ligases from *Populus tomentosa*. Plant Mol Biol Rep 33:841–854. https://doi.org/10.1007/s11105-014-0803-4

Reboledo G, Agorio AD, Vignale L, Batista-García RA, Ponce De León I (2021) Transcriptional profiling reveals conserved and species-specific plant defense responses during the interaction of *Physcomitrium patens* with *Botrytis cinerea*. Plant Mol Biol 107:365–385. https://doi.org/10.1007/s11103-021-01116-0

Rechler MM (1969) The Purification and Characterization of L-Histidine Ammonia-lyase (*Pseudomonas*). J Biol Chem 244:551–559. https://doi.org/10.1016/S0021-9258(18)94392-9

Renault H, Werck-Reichhart D, Weng J-K (2019) Harnessing lignin evolution for biotechnological applications. Curr Opin Biotechnol 56:105–111. https://doi.org/10.1016/j.copbio.2018.10.011

Renault H, Alber A, Horst NA et al (2017) A phenol-enriched cuticle is ancestral to lignin evolution in land plants. Nat Commun 8:14713. https://doi.org/10.1038/ncomms14713

Rencoret J, Gutiérrez A, Marques G, Del Río JC, Tobimatsu Y, Lam PY, Pérez-Boada M, Ruiz-Dueñas FJ, Barrasa JM, Martínez AT (2021) New Insights on Structures Forming the Lignin-Like Fractions of Ancestral Plants. Front Plant Sci 12:740923. https://doi.org/10.3389/fpls.2021.740923

Rensing SA, Goffinet B, Meyberg R, Wu S-Z, Bezanilla M (2020) The Moss *Physcomitrium* (*Physcomitrella*) *patens*: A Model Organism for Non-Seed Plants. Plant Cell 32:1361–1376. https://doi.org/10.1105/tpc.19.00828

Rensing SA, Lang D, Zimmer AD et al (2008) The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. Science 319:64–69. https://doi.org/10.1126/science.1150646

Rensing SA, Ick J, Fawcett JA, Lang D, Zimmer A, van de Peer Y, Reski R (2007) An ancient genome duplication contributed to the abundance of metabolic genes in the moss *Physcomitrella patens*. BMC Evol Biol 7:130. https://doi.org/10.1186/1471-2148-7-130

Reski R, Parsons J, Decker EL (2015) Moss-made pharmaceuticals: from bench to bedside. Plant Biotechnol J13:1191–1198. https://doi.org/10.1111/pbi.12401

Rétey J (1996) Enzymatic catalysis by Friedel-Crafts-type reactions. Naturwissenschaften 83:439–447. https://doi.org/10.1007/bf01144012

Rieseberg TP, Dadras A, Fürst-Jansen JMR, Dhabalia Ashok A, Darienko T, Vries S de, Irisarri I, Vries J de (2023) Crossroads in the evolution of plant specialized metabolism. Semin Cell Dev Biol 134:37–58. https://doi.org/10.1016/j.semcdb.2022.03.004

Ritter H, Schulz GE (2004) Structural basis for the entrance into the phenylpropanoid metabolism catalyzed by phenylalanine ammonia-lyase. Plant Cell 16:3426–3436. https://doi.org/10.1105/tpc.104.025288

Rogers SO, Bendich AJ (1985) Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. Plant Mol Biol 5:69–76. https://doi.org/10.1007/BF00020088

Rösler J, Krekel F, Amrhein N, Schmid J (1997) Maize phenylalanine ammonia-lyase has tyrosine ammonia-lyase activity. Plant Physiol 113:175–179. https://doi.org/10.1104/pp.113.1.175

Rouse B, Matalon R, Koch R, Azen C, Levy H, Hanley W, Trefz F, La Cruz F de (2000) Maternal phenylketonuria syndrome: congenital heart defects, microcephaly, and developmental outcomes. J Pediatr 136:57–61. https://doi.org/10.1016/S0022-3476(00)90050-7

Sambrook J, Russell DW (eds) (2001) Molecular cloning: A laboratory manual, 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Sato M, Sakayama H, Sato M, Ito M, Sekimoto H (2014) Characterization of sexual reproductive processes in *Chara braunii* (Charales, Charophyceae). Phycol Res 62:214–221. https://doi.org/10.1111/pre.12056

Scarpati ML, Oriente G (1958) Isolamento e costituzione dell'acido rosmarinico (dal *rosmarinus off.*). Ric Sci 28:2329–2333

Schneider K, Hövel K, Witzel K, Hamberger B, Schomburg D, Kombrink E, Stuible H-P (2003) The substrate specificity-determining amino acid code of 4-coumarate:CoA ligase. Proc Natl Acad Sci U S A 100:8601–8606. https://doi.org/10.1073/pnas.1430550100

Schubert H, Blindow I, Schories D, Mages M, Tümpling W von, Woelfl S (2018) Biogeography of Chilean Charophytes – determined by climate or by water chemistry? Botany Letters 165:129–145. https://doi.org/10.1080/23818107.2017.1370612

Schwarze CM, Petersen M (2024) Phenylalanine ammonia-lyases and 4-coumaric acid coenzyme A ligases in *Chara braunii*, *Marchantia polymorpha*, and *Physcomitrium patens* as extant model organisms for plant terrestrialization. Plant J. https://doi.org/10.1111/tpj.16950

Schwede TF, Rétey J, Schulz GE (1999) Crystal structure of histidine ammonia-lyase revealing a novel polypeptide modification as the catalytic electrophile. Biochemistry 38:5355–5361. https://doi.org/10.1021/bi982929q

Sgroi M, Paszkowski U (2020) Transcriptional responses to arbuscular mycorrhizal symbiosis development are conserved in the early divergent Marchantia paleacea. bioRxiv 2020.12.14.422721; https://doi.org/10.1101/2020.12.14.422721

Shimamura M (2016) *Marchantia polymorpha*: Taxonomy, Phylogeny and Morphology of a Model System. Plant Cell Physiol 57:230–256. https://doi.org/10.1093/pcp/pcv192

Shockey J, Browse J (2011) Genome-level and biochemical diversity of the acyl-activating enzyme superfamily in plants. Plant J 66:143–160. https://doi.org/10.1111/j.1365-313X.2011.04512.x

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. https://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. https://doi.org/10.1016/j.phytochem.2008.06.014

Soares AR, Marchiosi R, Siqueira-Soares RdC, Barbosa de Lima R, Dantas dos Santos W, Ferrarese-Filho O (2014) The role of L-DOPA in plants. Plant Signal Behav 9:e28275. https://doi.org/10.4161/psb.28275

Steele CL, Chen Y, Dougherty BA, Li W, Hofstead S, Lam KS, Xing Z, Chiang S-J (2005) Purification, cloning, and functional expression of phenylalanine aminomutase: the first committed step in Taxol side-chain biosynthesis. Arch Biochem Biophys 438:1–10. https://doi.org/10.1016/j.abb.2005.04.012

Stöckigt J, Zenk MH (1975) Chemical syntheses and properties of hydroxycinnamoyl-coenzyme A derivatives. Z Naturforsch C Biosci 30:352–358. https://doi.org/10.1515/znc-1975-5-609

Strullu-Derrien C, Selosse M-A, Kenrick P, Martin FM (2018) The origin and evolution of mycorrhizal symbioses: from palaeomycology to phylogenomics. New Phytol 220:1012–1030. https://doi.org/10.1111/nph.15076

Strullu-Derrien C, Kenrick P, Selosse M-A (2016) Origins of the mycorrhizal symbioses. In: Martin F (ed) Molecular Mycorrhizal Symbiosis. Wiley, pp 1–20. https://doi.org/10.1002/9781118951446.ch1

Stuible H, 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. https://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. https://doi.org/10.1074/jbc.M100355200

Suchi M, Sano H, Mizuno H, Wada Y (1995) Molecular cloning and structural characterization of the human histidase gene (HAL). Genomics 29:98–104. https://doi.org/10.1006/geno.1995.1219

Szövényi P (2016) The Genome of the Model Species *Anthoceros agrestis* Genomes and Evolution of Charophytes, Bryophytes, Lycophytes and Ferns. Elsevier, pp 189–211. https://doi.org/10.1016/bs.abr.2015.12.001

Takahashi H, Asakawa Y (2017) Transcriptome Analysis of Marchantin Biosynthesis from the Liverwort *Marchantia polymorpha*. Nat Prod Commun 12:1934578X1701200. https://doi.org/10.1177/1934578X1701200831

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. https://doi.org/10.1016/S0040-4039(00)97569-5

Tanaka M, Esaki T, Kenmoku H, Koeduka T, Kiyoyama Y, Masujima T, Asakawa Y, Matsui K (2016) Direct evidence of specific localization of sesquiterpenes and marchantin A in oil body cells of *Marchantia polymorpha* L. Phytochemistry 130:77–84. https://doi.org/10.1016/j.phytochem.2016.06.008

Tavares S, Vesentini D, Fernandes JC, Ferreira RB, Laureano O, Ricardo-Da-Silva JM, Amâncio S (2013) *Vitis vinifera* secondary metabolism as affected by sulfate depletion: diagnosis through phenylpropanoid pathway genes and metabolites. Plant Physiol Biochem 66:118–126. https://doi.org/10.1016/j.plaphy.2013.01.022

Towbin H, Staehelin T, Gordon J (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc Natl Acad Sci U S A 76:4350–4354. https://doi.org/10.1073/pnas.76.9.4350

Ufer M (2005) Comparative pharmacokinetics of vitamin K antagonists: warfarin, phenprocoumon and acenocoumarol. Clin Pharmacokinet 44:1227–1246. https://doi.org/10.2165/00003088-200544120-00003

Ullmann A, Jacob F, Monod J (1967) Characterization by in vitro complementation of a peptide corresponding to an operator-proximal segment of the beta-galactosidase structural gene of *Escherichia coli*. J Mol Biol 24:339–343. https://doi.org/10.1016/0022-2836(67)90341-5

Vance CP, Nambudiri AM, Wat C-K, Towers G (1975) Isolation and properties of hydroxycinnamate: CoA ligase from *Polyporus hispidus*. Phytochemistry 14:967–969. https://doi.org/10.1016/0031-9422(75)85168-5

Vanholme R, Meester B de, Ralph J, Boerjan W (2019) Lignin biosynthesis and its integration into metabolism. Curr Opin Biotechnol 56:230–239. https://doi.org/10.1016/j.copbio.2019.02.018

Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W (2010) Lignin biosynthesis and structure. Plant Physiol 153:895–905. https://doi.org/10.1104/pp.110.155119

Walton E, Butt VS (1971) The demonstration of cinnamyl-CoA synthetase activity in leaf extracts. Phytochemistry 10:295–304. https://doi.org/10.1016/S0031-9422(00)94043-3

Walton E, Butt VS (1970) The Activation of Cinnamate by an Enzyme from Leaves of Spinach Beet (*Beta vulgaris* L. ssp. *vulgaris*). J Exp Bot 21:887–891. https://doi.org/10.1093/jxb/21.4.887

Wang L, Gamez A, Archer H, Abola EE, Sarkissian CN, Fitzpatrick P, Wendt D, Zhang Y, Vellard M, Bliesath J, Bell SM, Lemontt JF, Scriver CR, Stevens RC (2008) Structural and biochemical characterization of the therapeutic *Anabaena variabilis* phenylalanine ammonia lyase. J Mol Biol 380:623–635. https://doi.org/10.1016/j.jmb.2008.05.025

Weise NJ, Ahmed ST, Parmeggiani F, Galman JL, Dunstan MS, Charnock SJ, Leys D, Turner NJ (2017) Zymophore identification enables the discovery of novel phenylalanine ammonia lyase enzymes. Sci Rep 7:13691. https://doi.org/10.1038/s41598-017-13990-0

Weng J-K, Akiyama T, Bonawitz ND, Li X, Ralph J, Chapple C (2010) Convergent evolution of syringyl lignin biosynthesis via distinct pathways in the lycophyte *Selaginella* and flowering plants. Plant Cell 22:1033–1045. https://doi.org/10.1105/tpc.109.073528

Whetten R, Sederoff R (1995) Lignin Biosynthesis. Plant Cell 7:1001–1013. https://doi.org/10.1105/tpc.7.7.1001

Whetten RW, Sederoff RR (1992) Phenylalanine ammonia-lyase from loblolly pine: purification of the enzyme and isolation of complementary DNA clones. Plant Physiol 98:380–386. https://doi.org/10.1104/pp.98.1.380

Wickett NJ, Mirarab S, Nguyen N et al (2014) Phylotranscriptomic analysis of the origin and early diversification of land plants. Proc Natl Acad Sci USA 111:E4859-68. https://doi.org/10.1073/pnas.1323926111

Williams JS, Thomas M, Clarke DJ (2005) The gene stlA encodes a phenylalanine ammonia-lyase that is involved in the production of a stilbene antibiotic in *Photorhabdus luminescens* TT01. Microbiol 151:2543–2550. https://doi.org/10.1099/mic.0.28136-0

Wodniok S, Brinkmann H, Glöckner G, Heidel AJ, Philippe H, Melkonian M, Becker B (2011) Origin of land plants: do conjugating green algae hold the key? BMC Evol Biol 11:104. https://doi.org/10.1186/1471-2148-11-104

Wohl J, Petersen M (2020) Phenolic metabolism in the hornwort *Anthoceros agrestis*: 4-coumarate CoA ligase and 4-hydroxybenzoate CoA ligase. Plant Cell Rep 39:1129–1141. https://doi.org/10.1007/s00299-020-02552-w

Wolf L, Rizzini L, Stracke R, Ulm R, Rensing SA (2010) The molecular and physiological responses of *Physcomitrella patens* to ultraviolet-B radiation. Plant Physiol 153:1123–1134. https://doi.org/10.1104/pp.110.154658

Wüstenberg A, Pörs Y, Ehwald R (2011) Culturing of stoneworts and submersed angiosperms with phosphate uptake exclusively from an artificial sediment. Freshwater Biol 56:1531–1539. https://doi.org/10.1111/j.1365-2427.2011.02591.x

Xiang L, Moore BS (2005) Biochemical characterization of a prokaryotic phenylalanine ammonia lyase. J Bacteriol 187:4286–4289. https://doi.org/10.1128/jb.187.12.4286-4289.2005

Xie M, Zhang J, Tschaplinski TJ, Tuskan GA, Chen J-G, Muchero W (2018) Regulation of Lignin Biosynthesis and Its Role in Growth-Defense Tradeoffs. Front Plant Sci 9:1427. https://doi.org/10.3389/fpls.2018.01427

Xu J-G, Hu Q-P, Liu Y (2012) Antioxidant and DNA-protective activities of chlorogenic acid isomers. J Agric Food Chem 60:11625–11630. https://doi.org/10.1021/jf303771s

Xue JY, Hind KR, Lemay MA, Mcminigal A, Jourdain E, Chan CX, Martone PT (2022) Transcriptome of the coralline alga *Calliarthron tuberculosum* (Corallinales, Rhodophyta) reveals convergent evolution of a partial lignin biosynthesis pathway. PLoS One 17:e0266892. https://doi.org/10.1371/journal.pone.0266892

Yadav V, Wang Z, Wei C, Amo A, Ahmed B, Yang X, Zhang X (2020) Phenylpropanoid Pathway Engineering: An Emerging Approach towards Plant Defense. Pathogens 9. https://doi.org/10.3390/pathogens9040312

Ye Z-H, Zhong R (2022) Cell wall biology of the moss *Physcomitrium patens*. J Exp Bot 73:4440–4453. https://doi.org/10.1093/jxb/erac122

Yu H-N, Liu X-Y, Gao S, Han X-J, Cheng A-X, Lou H-X (2014) Molecular cloning and functional characterization of a phenylalanine ammonia-lyase from liverwort *Plagiochasma appendiculatum*. Plant Cell Tiss Organ Cult 117:265–277. https://doi.org/10.1007/s11240-014-0438-z

Yuan J-C, Xiong R-L, Zhu T-T, Ni R, Fu J, Lou H-X, Cheng A-X (2021) Cloning and functional characterization of three flavonoid O-glucosyltransferase genes from the liverworts *Marchantia emarginata* and *Marchantia paleacea*. Plant Physiol Biochem 166:495–504. https://doi.org/10.1016/j.plaphy.2021.06.009

Zang Y, Jiang T, Cong Y, Zheng Z, Ouyang J (2015) Molecular Characterization of a Recombinant *Zea mays* Phenylalanine Ammonia-Lyase (ZmPAL2) and Its Application in trans-Cinnamic Acid Production from L-Phenylalanine. Appl Biochem Biotechnol 176:924–937. https://doi.org/10.1007/s12010-015-1620-4

Zenk MH (1979) Recent Work on Cinnamoyl CoA Derivatives. In: Swain T, Harbone JB, van Sumere CF (eds) Biochemistry of Plant Phenolics. Springer US, Boston, MA, pp 139–176. https://doi.org/10.1007/978-1-4684-3372-2_5

Zhou H, Schwartzenberg K von (2020) Zygnematophyceae: from living algae collections to the establishment of future models. J Exp Bot 71:3296–3304. https://doi.org/10.1093/jxb/eraa091

8 Appendix

8.1 Gene and protein sequences

In the following (Table 26), all the nucleotide and protein sequences of the genes characterized in this work are listed. It has to be noticed that the genes that were codon-optimized possess a different nucleotide sequence than the original annotation suggested, yet represent the same protein sequence.

Table 26 Nucleotide and amino acid sequences according to their last sequencing for all investigated enzymes.

CbPAL 1	ATGGCCCTGCTGGAGCCAGCCACGACTCCTCCGCAAAATGGTCACTATCACGACTCGGCAGGAATCTGTTTCGACGGCAGCCATCTGAGGATCGAAGACATC GTCGCCATCTCTCGGCGTTCCATGCCGGTACGGCTGTCGGATGATCCGAAGTTCTGCAGCCGTATCTCTCTGGGAGCCAATTTTCTGGATCGCCTTCTGCGTGA GGAGAAGGCGATCTATGGCGTGACTACCGGTTATGGGGACTCCTGCACCGTGAGCATTCCTGCAGATCTGGTCGAAGAACTGCCGCACCGGCTATACACCTT CCATGGCTGCGGTGCAGGTCGAATGCTCACTGCCGCGGAGACGCGAGCGGTACTCGCCGTCCGACTTGCCTCGCTGGCAAGAGGGGTGTCCGGGGTCAGCC TAGACCTCTTGCGGCAGCTCGAATTACTCCTGGTGCACGACATTCTGCCGCTGATTCCCGCCGAGGGCTCCGTCGGCGCGAGTGGCGACCTCACGCCGCTGTC CTACGTGGCAGCGGGCGCTCTGCGGTGAGCGCAAGGTGCTTTTCCGCGGGGAGACCTGTGACGCGGCGCAGGCGCTGGATGCCATCGGCGCCCGACCTCTGC GGCTCCGCCCCAAGGAAGGTCTAGCAGTCCTGAACGGCACCTCCGTGATGACGGCGCTGGCCTGCCTCGCATGGCGGCGAGCCGACTACTTCAGCCGGCTT GCAACGCGGCTCACAGCGTGCACTGTGGTCGCCACCGCTGGCAATCCGGGGGCACTTCGACGAGGGCTTGCATGCCGTTAAGCCTCATCCTGGCCAGCAGCGA GCGGCGGCGCGCCTGCGAAGCGATCTGCAAACGGACCGTGCACCAGCCAAAGAACGCCGCCTGCAGGATCGCTATTCCCTGCGCTGCGCCCCTCACGTAATC GGCGTGCTCGAAGACGCGCTGCCTTGGCTCAGAAGCATCATTGAGAGCGAGCTCAACAGCGCTAATGACAACCCCATAATTCTCGGCGACGAGGAGCGCGT TCTGCACGGCGGGCATTTCTACGGCGGCCACATCGCCTTCGCGATGGACACACTGAAGATCGCGGTGGCCAACGTGGCCGACCTGCTGGACCGCCAGCTGG CACTGCTGGTCGATCCGCGTTTAACCACGGGCTGCCCTCAATCTGTCGGGCACAGTCGGCCACCGCGCGGCCATTAACCATGGGCTGAAGGCACTGCAGA TCAGCGTGTCGGCATGGACGGCGAAGCGTTGAAGCTCACCATGCCTGCTTCCGTCTTTTCGCGGTCCACGGAGTGCCACAATCAGGACAAGGTGAGCATG GGCACCATCGCAGCCCGCGATTGCCTCCGTGTTCTCGAGCTGACCGAACAGGTCACTGCGGCAATGCTGATTGCGGCGCGGCAGGGCGTGACCTTGCGAGG GCAGGTTGATGGCAGCTTGTCCTTGTCGCCGGTGCTGGATTCCTTCATCGCCGATCTCGGTCAACGCATCAAGCTGGTCGTGGAAGACCGTGCCCTCGACGA GGACCTGCAGGACCTTCTTGCCGCCCTGCGGAAGGAGTCGTGGCAGCTTTATTCATGA	MALLEPATTPPQNGHYHDSAGICFDGSHLRIEDIVAISRRSMPVRLSDDPKFCSR ISLGANFLDRLLREEKAIYGVTTGYGDSCTVSIPADLVEELPHRLYTFHGCGAGR MLTAAETRAVLAVRLASLARGVSGVSLDLLRQLELLLVHDILPLIPAEGSVGASGD LTPLSYVAAALCGERKVLFRGETCDAAQALDAIGARPLRLRPKEGLAVLNGTSV MTALACLAWRRADYFSRLATRLTACTVVATAGNPGHFDEGLHAVKPHPGQQR AAARLRSDLQTDRAPAKERRLQDRYSLRCAPHVIGVLEDALPWLRSIIESELNSA NDNPIILGDEERVLHGGHFVGGHIAFAMDTLKIAVANVADLDRQLALLVDPRF NHGLPSNLSGTVGHRAAINHGLKALQISVSAWTAEALKLTMPASVFSRSTECHN QDKVSMGTIAARDCLRVLELTEQVTAAMLIAARQGVTLRGQVDGSLSLSPVLDS FIADLGQRIKLVVEDRALDEDLQDLLAALRKESWQLYS
CbPAL 2s	ATGCAGAGCGGCGAACAGGCCAGCATGGATATTCGTGAAATTATTGAAAAGGTGCGTCAGATTATTGAAAGTAATGCAACCCTGGCCCTGAGCGAACAGGC AAAACAGAATATTGCCGGCTGCCGCGCATATCTGGATCGTAAAATGGCCAGTCAGAAAGAACCGGTGTATGGCATTAATACCGGCTTCGGCAGTCTGTGTAA TGTTATTATTCCGGATAATGAAATCGAGCAGCTGCAGACCAATCTGGTGATGAGTCATGCATGCGGTATGGGTGATGAAGTGCCGGTTGAAATTGTGCGCCT GATGCTGCTGCTGAAAGTGCAGGCACTGTGTTATGGTAAAAGTGCAGTGCAGCTGGTTACCGTGGAACGCCTGGTGGATCTGTATAATCATCGTATCTATCC GATTGTGTATCAGCAGGGTAGTCTGGGTGCCAGTGGTGATCTGGCCCGGTGCACATCTGAGTCTGCCGCTGCTGGGCATGGGTGAAGTGAATTATCAGG GCAGCAAATATGAAACGCCGTGGTTCTGCGTCAGCTGGGCCTGGCACCGATTGCACTGAAAAGTAAAGAAGGTCTGGCACTGCTGAATGGCACCCAGTTCA TGAATGCATATGGTATCTGGTGTGCCCTGAAAGCCGAACGTCTGAATATGCTGGCCGATCTGGCCGGTGCCATTAGCCTGGATGCTATTCGATGGTCGTATTG AACCGTTCAAAGATAATATTCATACCATTCGTCCGCATAAAGGTCAGCTGCAGACAGCCCAGGGCATTCGCAAAATTCTGGAAGGCAGTGAAATTGCCCGCC AGCATAAAACACAAGTTCAGGATCCGTATAGCTTCCGCTGTATTCCGCAGGTTCATGGCGCCACCAAAGATACCACCCGTTATGTGACCAGCGTGTTCGAAAC CGAAATGAATGCAGTTACCGATAATCCGACCGTGTTCCCGGAAGATGATGAAATTATTAGCGGTGGCAACTTCCATGGTCAGCCGCTGGCACTGGCCCTGGA CTTCCTGGCAATTGCAATTGCCGAACTGGGTAATATTAGCGAACGTCGTACCTATCAGCTGATTAGTGGTGTGCGTCAGCTCCGGCCTTCCTGGTTGCCAAA CCGGGCCTGAATAGCGGCTTCATGATTCCGCAGTATACCGCCGCAAGTATTGTTAGTCAGAATAAACAGCTGTGCACCCCGGCAAGTATTGATAGTATTGTGA GTAGTAATGGTCAGGAAGATCATGTTAGCATGGGCGCAAATGCAGCAACCAAATGTTATAAAGTTATGGAAAACGTGGAGAATATCCTGGCAATTGAACTG CTGAATGCCACCCAGGCACTGCACTTCCGTAAACCGCTGAAAAGCAGCCCGAAAATGAAAAAATTGTGAATCTGTTCCGCACCACCATTCCGTTCATTGAAG AAGATAAAATTATGTACCGCGAACTGGAAGCAACCCTGCAGTTCGTTAAACATCTGAATCTGGAAGGCCTGTATTAA	MQSGEQASMDIREIIEKVRQIIESNATLALSEQAKQNIAGCRAYLDRKMASQKE PVYGINTGFGSLCNVIIPDNEIEQLQTNLVMSHACGMGDEVPVEIVRLMLLLKV QALCYGKSAVQLVTVERLVDLYNHRIYPIVYQQGSLGASGDLAPLAHLSLPLLG MGEVNYQGSKYETAVVLRQLGLAPIALKSKEGLALLNGTQFMNAYGIWCALKA ERLNMLADLAGAISLDAFDGRIEPKFDNIHTIRPHKGQLQTAQGIRKILEGSEVAR QHKTQVQDPYSFRCIPQVHGATKDTTRYVTSVFETEMNNAVTDNPTVFPEDDEII SGGNFHGQPLALALDFLAIAIAELGNISERRTYQLISGVRQLPAFLVAKPGLNSGF MIPQYTAASIVSQNKQLCTPASIDSIVSSNGQEDHVSMGANAATKCYKVMENV ENILAIELLNATQALHFRKPLKSSPKIEKIVNLFRTTIPFIEEDKIMYRELEATLQFVK HLNLEGLY
CbPAL 2l	ATGACCGAACAGACCCATTGTCTGCAGCTGGCAGCACAGCTGCGTGCACATGGCATTAATACCGAAGTGTATAGCGATGCAGTGAAAATGAAAAAACAGTTC GAATATGCCGATAAAAAGCAGATTCCGTATGTGTGCGTTGTGGGCAGCGATGAAATTACCGCCAGCGTTATACCCTGCGCAATATGCAGAGCGGCGAACA GGCCAGCATGGATATTCGCGAAATTATTGAAAAGTGCGCCAGATTATTGAAAGCAATGCAACCCTGGCCCTGAGCGAACAGGCAAAACAGAATATTGCAG GCTGCCGCGCATATCTGGATCGCAAAATGGCCAGTCAGAAAGAACCGGTGTATGGCATTAATACAGGCTTCGGTAGCCTGTGCAATGTTATTATTCCGGATA ATGAAATCGAGCAGCTGCAGACCAATCTGGTTATGAGTCATGCTGCGGTATGGGCGATGAAGTTCCGGTTGAAATTGTGCGCCTGATGCTGCTGCTGAAAG TTCAGGCACTGTGTTATGGCAAAAGTGCCGTTCAGCTGGTGACCGTGGAACGCCTGGTGGATCTGTATAATCATCGTATCTATCCGATTGTGTATCAGCAGGG CAGCCTGGGCGCCAGCGGTGATCTGGCACCGCTGGCTCATCTGAGCCTGCCGCTGCTGGGTATGGGTGAAGTGAATTATCAGGGTAGTAAATATGAAACCG CCGTGGTTCTGCGTCAGCTGGGTCTGGCACCGATTGCCCTGAAAAGCAAAGAAGGCCTGGCACTGCTGAATGGCACCCAGTTCATGAATGCATATGGTATCT GGTGCGCACTGAAAGCCGAACGTCTGAATATGCTGGCAGACTTAGCAGGTGCCATTAGCCTGGATGCCTTCGATGGTCGTATTGAACCGTTCAAAGATAATA TTCATACCATTCGTCCGCATAAAGGTCAGCTGCAGACAGCCCAGGGTATTCGTAAAATTCTGGAAGGCAGCGAAATTGCCCGTCAGCATAAAACACAAGTGC AGGATCCGTATAGCTTCCGTTGTATTCCGCAGGTGCATGGTGCAACCAAAGATACCACCCGCTATGTTACCAGCGTGTTCGAAACCGAAATGAATGCCGTGAC CGATAATCCGGAAGATGATGAAATTATTAGTGGTGGCAACTTCCATGGTCAGCCGCTGGCACTGGCCCTGGACTTCCTGGCCATTGCAATT GCAGAACTGGGCAATATTAGTGAACGTCGCACCTATCAGCTGATTAGTGGTGTGCGCCAGCTGCCGGCCTTCCTGGTGGCAAAACCGGGTCTGAATAGTGGC TTCATGATTCCGCAGTATACCGCCGCAAGTATTGTGAGCCAGAATAAACAGCTGTGTACCCCGGCAAGTATTGATAGCATTGTGAGTAGCAATGGTCAGGAA GATCATGTTAGTATGGGTGCAAATGCAGCAACCAAATGTTATAAAGTTATGGAAAACGTGGAGAATATCCTGGCAATTGAACTGCTGAATGCACCCAGGCA CTGCACTTCCGCAAACCGCTGAAAAGTAGTCCGAAAATTGAAAAAATCGTGAATCTGTTCCGCACCACCATTCCGTTCATTGAAGAAGATAAAATTATGTACC GCGAACTGGAAGCAACCCTGCAGTTCGTTAAACATCTGAATCTGGAAGGCCTGTATTAA	MTEQTHCLQLAAQLRAHGINTEVYSDAVKMKKQFEYADKKQIPYVCVVGSDEI TAQRYTLRNMQSGEQASMDIREIIEKVRQIIESNATLALSEQAKQNIAGCRAYLD RKMASQKEPVYGINTGFGSLCNVIIPDNEIEQLQTNLVMSHACGMGDEVPVEI VRLMLLLKVQALCYGKSAVQLVTVERLVDLYNHRIYPIVYQQGSLGASGDLAPLA HLSLPLLGMGEVNYQGSKYETAVVLRQLGLAPIALKSKEGLALLNGTQFMNAYG IWCALKAERLNMLADLAGASLDAFDGRIEPKFDNIHTIRPHKGQLQTAQGIRKI LEGSEIARQHKTQVQDPYSFRCIPQVHGATKDTTRYVTSVFETEMNAVTDNPT VFPEDDEIISGGNFHGQPLALALDFLAIAIAELGNISERRTYQLISGVRQLPAFLVA KPGLNSGFMIPQYTAASIVSQNKQLCTPASIDSIVSSNGQEDHVSMGANAATK CYKVMENVENILAIELLNATQALHFRKPLKSSPKIEKIVNLFRTTIPFIEEDKIMYRE LEATLQFVKHLNLEGLY
MpPAL 1	ATGAACGAATTCTTCTCGTCCAACTTTGAGCAGATCAAAATGCCCACTGGGAAAACCGGTGATTCTATGCAGTATTGCCTCAGTCCTTCGGCGATGACGAAGC CAGCTTCACAACACTGAGGACCCTCTGCAGTGGGCGAAAGCAGCCCGTGCGATGCATTGCAGTCACTGGGAGGAAGTCGCCGACATGGTTCATGAGTTTCAG GATGCCAAAGAAATACTCATTCAAGGTGCCAATCTCAAAGTGGCCCATGTCACAGCTGTTGCCAGGCGTTCTGATGTCACCGTGGTGCTCGATGCAGCCGTTG CCAGGAACAGAGTTGACGAGAGCTCCGAATGGGTGCTGCAACATATGAGGAAAGGCACTGACACTTACGGTGTCACTACAGGCTTCGGTGCCACATCCCAC CGCCGCACTGACCAGGGAGTTGAGTTACGAAAGAGTTGATTCGTTTCCTCAATGCTGGAGTTCTGGGTAAGGACGCTGCCAACATCCTTCCCGTGTCCACAA CCAGAGCAGCAATGCTGGTAAGAACCAACACTCTGCTGCAAGGCTACTCCGGCATTCGGTGGGCAATCCTCCAAACTCTGCAGAAACTGCTCAACAAGCAGA TCACTCCTCAACTTCCTCTTCGAGGCACCATCACAGCTTCAGGAGATCTCGTGCCACTTTCATACATTGCAGGTCTACTCACCGGACGTCCTAACTCCAAAGCTG TCACCGAATGTGGAGACATTGTCAGCGCACAACAAGCTCTGAAGCTCGTAGGTCTGCTGAAGCCCTTTGAGCTGCAACCCAAAGAAGGACTGGCCCTCGTGA ACGGCACAGCAGTCGGCTCAGCTTGGCTTCCACATGTTGCTACGATGCCAACGTGCTCGCCCTCTTCGCCGAAGTGTCGTCGGCTCTGTTCTGTGAAGTGAT GCAGGGCAAGCCTGAATTCACCGACCACCTCACCCACAAGCTCAAGCATCACCCCGGTCAGATTGAAGCGGCTGCCATCATGGAGCATGTGCTGGAAGGCA GCTCCTACATGAAGGCCGCAGCCAAGCTCCACGAAACGGACCCGCTGCAGAAGCCCAAGCAGGACAGGTACGCTCTGCTCAGCAGCTCCCCAGTGGCTGGGC CCCCAAATCGAGGTGATCCGCGCTGCCACACACTCCATCCAGAGGGAGATCAACTCTGTGAACGACAACCCGCTCATCGATGTGTCCCGAGATATGGCAGTG CACGGTGGAAACTTCCAGGGCACTCCCATCGGAGTGTCCATGGACAACATGCGACTGGCAATTGCAGCCATGGGCAAGCTCATGTTCGCTCAGATGTCGGAG CTCGTCAACGACTTCTATAACGGCGGGCTTCCATCGAATCTGACGGGAGGTCCTAATCCGAGCTTGGATTATGGATTCAAAGGCGCGGAGATAGCCATGGCG TCTTACACTTCAGAGCTCTTGTATTTGGCCGGTCCAGTGACCACCCATGTGCAGAGTGCGGAGCAACACAACCAGGACGTGAACTCTCTTGGGTCTCATCTCAGT CTCGCAAAACAGAGGAAGCTATCGAAATCTTCAAGCTCATGTGCTCCACATATTTGGTGGCGTTGTGCCAGGCTGTGGATCTGCGACACTTGGAGGAGAACA TGCTAGCAAGCGTCAAGCAAGTGGTCTCTCAAGCGGCCAGGAAGAACTTTGTTCACCGATAACGAAGGCTGCCTACTGCATCCGCGCTTCTGTGAGAAGGACC TTCTGAAAGTGGTGGATAATTTGCCTCCTTTCTCGTACATAGACGATGCTACCAGCGCATCCTACCCTCTGATGCAGAAACTGAGGCAATGTCTGGTGAATAG TGCGCTCAGTCCCGCCAATGATGATGAGGGCTGTAGTACATTCAGAAGAATCTCGGTCTTCGAGGAGCAGCTAAAGATCGGGCTGGAGTCTACCATCCCCTC CATTCGAGGCAATTCGACAATGGCGTCACAGCAGTACCAAACAGAATCACTGACTGCAGGAGTTATCCTCTGTACAATTTTGTAAGATCATTAGGCACCAAG TTACTAGTAGGCACGGAGACGAGGTCACCAGGGCAGGATATCGAGCTCGTTTTTGAAGCCATTTCGGAAGGCAAGATGGCCACTCCCTTACTTGAATGTTTA TCAGTGTGGGACGGATGTCCGAGATCCATCAAGCCTTCCAAAACAATAGCGTAA	MNEFFSSNFEQIKMPTGKTGDSMQYCLSPSAMTKPASHTEDPLQWAKAARA MHCSHWEEVADMVHEFQDAKEILIQGANLKVAHVTAVARRSDVTVVLDAAV AKERVDESSEWVLHNMRKGTDTYGVTTGFGATSHRRTDQGVELQKELIRFLNA GVLGKDAANILPVSTTRAAMLVRTNTLLQGYSGIRWAILQTLQKLLNKQITPQLP LRGTITASGDLVPLSYIAGLLTGRPNSKAVTECGDIVSAQQALKLVGLLKPFELQP KEGLALVNGTAVGSALASTCCYDANVLALFAEVSSALFCEVMQGKPEFTDHLTH KLKHHPGQIEAAAIMEHVLEGSSYMKAAAKLHETDPLQKPKQDRYALRTSPQW LGPQIEVIRAATHSIQREINSVNDNPLIDVSRDMAVHGGNFQGTPIGVSMDNM RLAIAAMGKLMFAQMSELVNDFYNGGLPSNLTGGPNPSLDYGFKGAEIAMAS YTSELLYLAGPVTTHVQSAEQHNQDVNSLGLISARKTEEAIEIFKLMCSTYLVALC QAVDLRHLEENMLASVKQVVSQAARRTLFTDNEGCLLPSRFCEKDLLKVVDNLP PFSYIDDATSASYPLMQKLRQCLVNSALSPANDDEGCSTFRRISVFEEQLKIGLES TIPSIRGQFDNGVTAVPNRITDCRSYPLYNFVRSLGTKLLVGTETRSPGQDIELVF EAISEGKMATPLLECLSVWDGCPRSIKPSKTIA
MpPAL 2	ATGATGAACGAATTCTTCTCATCGAATTTTCAGCAGATGAAAATGTCTGATGAGAAGACTGCAGATTCGATGCAGTACTGTTCACACCCTACGGCCCAGGTCG CTCCACGAACAGAAGATCCCCTGAATTGGGCAAAGGCAGCCCGTGCCATGGATTGCAGTCACTGGGAGGAGGTCTTCGACCTGGTGCAGGAATTTCAGAAC GAGAAGGAGGTGGTCATTCAGGGCGTGACGCTCACAGTGGCCCAGGTTGCAGCGGTTGCCAGGCGTCCTGACGTCACTGTGGTCCTCGACGCAGCTGTAGC TAAGGACCGCGTTGACGAGAGCTCGGAATGGGTGCTCCACCACATGAGGAAGGGCACTGACACTTACGGTGTCACTACAGGCTTCGGTGCCACCTCCCACCG CCGCACTGATCAGGGGGTCGACTTACAGAGAGAGTTGATTCGTTTTCTCAACGCTGGAGTTCTCGGCAAAGATGCTTCCAACATCCTTCCGGCCTCCACGACC CGAGCAGCAATGCTGGTGAGAACCAACACTCTGCTGCAAGGGTATTCCGGAATTCGGTGGTCCATCCTCCAAACTCTGCAGAAATTTCTCAACAAGCAGATCA CTCCTCAACTTCCTCTTCGGGGCACCATCACAGCTTCAGGAGATCTCGTCCCCTTTCCTACATTGCAGGTCTGCTTACCGGCCGTCCCAACTCCAAAGCCGTCA CAGAATGCGGAGATGTTGTGACTGCGCACGAAGCTTTGAAGCTCGTCGGATTGCTGAAGCCATTCGAGCTGCAGCCCAAAGAGGGACTCGCACTTGTGAAT GGTACAGCAGTCGGCTCAGCCTTGGCTTCCATATGTTGCTACGATGCCAACGTGCTCGCGCTCTTCGCCGAAGTGTCGTCGGCTCTGTTCTGTGAGGTGATGC AGGGCAAGCCCGAGTTCACCGACCACCTCACCCACAAGCTCAAGCATCACCCGGACAGATTGAAGCCGCTGCCATCATGGAGCATGTGCTAGAGGGTAGCT CCTACATGAAGGCCGCAGCCAAGCTCCACGAAACGGACCCGCTCCAGAAACCCAAGCAGGACAGGTACGCTCTGAGGACGTCTCCCCAGTGGCTGGGCCCC CAAATCGAGGTGATCCGTGCAGCCACACACTCCATCGAGAGGGAGATCAACTCTGTCAACGACAACCCGCTCATCGATGTGTCCCGAGACATGGCAGTGCAC GGTGGAAACTTCCAGGGCACTCCCATCGGAGTGTCCATGGACAACATGCGACTGGCAATTGCAGCCATGGGCAAGCTCATGTTCGCTCAGATGTCGGAGCTC GTCAACGACTTCTATAACGGCGGGCTTCCATCGAATCTGACGGGAGGTCCCAACCCGAGCTTGGACTATGGATTCAAAGGAGCGGAGATAGCCATGGCATC GTACACCTCGGAGCTCTTGTACTTGGCCAACCCAGTGACCACCCACGTGCAGAGTGCAGAGCAACACAACCAGGACGTGAACTCTCTGGGTCTCATCTCAGCT CGTAAAACAGAGGAAGCTATCGAGATCTTCAAGCTCATGTGCTCCACCTATTTGGTGGCGTTGTGCCAGGCAGTGGATCTGCGGCACCTGGAGGAGAACATG CAAGTGAGTGTCAAGAATGTGGTCTCTCAAGCCGCGAAGAAGACTCTCTTCACCGATGCCCAAGGCTGCTTACTGCCATCGCGCTTCTGTGAGAAGGATCTCT TGAAAGTGGTGGACAACCAGCCTCCATTCTCCTACGTCGACGATGCAACCAGCGCATCCTACCCTCTGATGCAGAAGCTGAGGCAATGCCGTGGAGCCATG CTCTCAGTTCCGAGAACGAAGAGGAGAGATGTAGTGTTTTCAGAAGAATCTCAGTCTTCGAGGAGCAGGTCAAGACGGATCTCGAAGCCACGGTGCCCAAA GTCAGAGAGCAATTCGACAATGGCGTCGCTGCCATACCTAACAGAATCAGCGACTGCAGAAGCTATCCCCTCTACGATTTTGTTAGATCTTTGGGTACCAAAT	MMNEFFSSNFQQMKMSDEKTGDSMQYCSHPTAQVAPRTEDPLNWAKAAR AMDCSHWEEVFDLVQEFQNEKEVVIQGVTLTVAQVAAVARRPDVTVVLDAAV AKDRVDESSEWVLHHMRKGTDTYGVTTGFGATSHRRTDQGVDLQRELIRFLN AGVLGKDASNILPASTTRAAMLVRTNTLLQGYSGIRWSILQTLQKFLNKQITPQL PLRGTITASGDLVPLSYIAGLLTGRPNSKAVTECGDVVTAHEALKLVGLLKPFELQ PKEGLALVNGTAVGSALASICCYDANVLALFAEVSSALFCEVMQGKPEFTDHLT HKLKHHPGQIEAAAIMEHVLEGSSYMKAAAKLHETDPLQKPKQDRYALRTSPQ WLGPQIEVIRAATHSIEREINSVNDNPLIDVSRDMAVHGGNFQGTPIGVSMDN MRLAIAAAMGKLMFAQMSELVNDFYNGGLPSNLTGGPNPSLDYGFKGAEIAM ASYTSELLYLANPVTTHVQSAEQHNQDVNSLGLISARKTEEAIEIFKLMCSTYLVA LCQAVDLRHLEENMQVSVKNVVSQAAKKTLFTDAQGCLLPSRFCEKDLLKVVD NQPPFSYVDDATSASYPLMQKLRQCLVSHALSSENEEERCSVFRRISVFEEQVKT DLEATVPKVREQFDNGVAAIPNRISDCRSYPLYDFVRSLGTKLLVGTETRSPGQD IELVYEAISQGKMASPLLQCLAGWNGCPKSIKPCKIVV

	TACTCGTGGGCACGGAGACGAGGTCGCCAGGACAGGATATCGAGCTTGTCTATGAAGCCATCTCGCAAGGCAAGATGGCTTCGCCCCTGATTCAATGTTTGG CAGGCTGGAATGGGTGTCCCAAATCAATCAAGCCTTGCAAAATTGTAGTTTAA	
MpPAL 3	ATGGACTTCCCGCTGTGCCTGGAACCGCGTGATGCACCGATTGAAGAACATGGCGGCGATGCCCTGTATTGGGCACGTGCAGCCCGCCGTGCAGAAGGCAG TCATTGGGAAGAAGTTCGCAGTATGGTTCAGGAATGTAGTGAAGCCAAAGAAGTTGCAATTCAGGGCGTGACCCTGAGCATTGGCCAGGTTACCGCCGTTG CACGCAATGCCGATGTTACCGTTGTGCTGGATGCAGAAAGCGCCAAAGGCCGCGTTGATGAAAGCGCCCGCTGGGTTAATGATAATATGACCAAAGGCACC GATACCTATGGCGTTACCACCGGCTTCGGTGCAACCAGCCATCGTCGCACCAATGAAGGCGTTCTGCTGCAGAAAGAACTGATTCGCTTCCTGAATGCAGGC GTTCTGACCGATGAAGCCAGCATTCTGCCGGTTGCAACCACCCGCGCAGCCATGCTGGTTCGCACCAATACCCTGATGCAGGGTTATAGTGGTATTCGTTGGG AAATTCTGCAGACCATGCAGAAACTGCTGAATAAACAGCTGACCCCGCGCCTGCCGCTGCGCGGTACCATTACCGCCAGTGGCGATCTGGTGCCGCTGAGTT ATATTGCAGGTCTGCTGACCGGTCGCCCGAATAGCCGTTGTGTTGATGCAAGCGGTCAGACCCTGACCGCACATGAAGCACTGAAACAGGTTGGCCTGATTA AACCGTTCGAACTGCTGCCGAAAGAAGGCCTGGCAATGGTTAATGGCACCGCCGTGGGTGCAGCACAGGCAAGCATGGCCTGTTATGATGCAAATGTGCTG GCACTGTTCGCCGAAGTTAGTAGTGCACTGTTCTGTGAAGCCATGCAGGGCAAACCGGAGTTCACCGATCATCTGACCCATAAACTGAAACATCATCCGGGTC AGATTGAAGCCGCCGCCATTATGGAATGGGTTCTGGAAGGCAGCGCCTATATGAAAGCAGCCGCAAAACTGCATGAAACCGATCCGCTGAAAAAACCGAAA CAGGATCGCTATGCCCTGCGTACCAGCCCGCAGTGGCTGGGTCCGCAGATTGAAGTTATTCGTGCAGCCACCCATAGCATTGAACGCGAAATTAATAGTGTT AATGACAATCCGCTGATTGATGTGAGTCGTGATATGGCCGTTCATGGCGGCAACTTCCAGGGTACCCCGATTGGCGTTAGTATGGATAATATGCGTCTGGCA CTGGGTGCAATTGGCAAACTGATGTTCGCCCAGATGAGTGAACTGGTTAATGACTTCTATAATAGTGGTCTGCCGAGTAATCTGACCGGTGGCCCGAATCCG AGTCTGGATTATGGCTTCAAAGGCGCCGAAATTGCAATGGCCAGCTATTGTAGTGAACTGCTGTATCTGGCCAATCCGGTGACCACCCATGTGCAGAGCGCC GAACAGCATAATCAGGATGTGAATAGCCTGGGTCTGATTAGTGCCCGCAAAACCGCAGAAGCCGTTGAAATTCTGAAACTGATGGCAAGTACCTTCCTGGTG GCCCTGTGTCAGGCAGTTGATCTGCGTCATCTGGAAGAAAATATGCAGGCCGCAGTTGCAAATGCCGTTGCAATGGCAGCCAAAAAAACCTTAGCAGTTGAA GCAGCCGCCCCGCTGCTGAAAGTTGTGGCAGATACCCCGCCGTTCACCTATGTTGATGATGCAACCAGCGCAGCCTATCCGCTGATGCAGAAATTACGTGAA GTGCTGGTGGCACGTGCCCTGCAGAGTGGTCAGGATGATGAAGGTTGTACCGCATTCCGTAGTATTAGCGCATTCGAAGCCGAAGTGAAACTGCAGCTGGA AACCGATGTTCCGGCAGTTCGCCAGCAGTTCGATGATGGCGTTAGCAGCGTTCCGAATAAAATTACCGATTGCCGTAGCTTCCCGCTGTATAACTTCGTTCGC AGCCTGGGCACCAAACTGCTGGTGGGTACCGATACCCGCAGCCCGGGTGAAGATATTCAGCTGGTGTATGAAGCCATGACCGAAGGCCGTCTGGCAGCACC GCTGCTGCGTTGTCTGGATGGCTGGAGCGGCACCCCGACCCCTATTACCCCGCTGATTCATTAA	MDFPLCLEPRDAPIEEHGGDALYWARAARRAEGSHWEEVRSMVQECSEAKEV AIQGVTLSIGQVTAVARNADVTVVLDAESAKGRVDESARWVNDNMTKGTDTY GVTTGFGATSHRRTNEGVLLQKELIRFLNAGVLTDEASILPVATTRAAMLVRTNT LMQGYSGIRWEILQTMQKLLNKQLTPRLPLRGTITASGDLVPLSYIAGLLTGRPN SRCVDASGQTLTAHEALKQVGLIKPFELLPKEGLAMVNGTAVGAAQASMACYD ANVLALFAEVSSALFCEAMQGKPEFTDHLTHKLKHHPGQIEAAAIMEWVLEGS AYMKAAAKLHETDPLKKPKQDRYALRTSPQWLGPQIEVIRAATHSIEREINSVN DNPLIDVSRDMAVHGGNFQGTPIGVSMDNMRLALGAIGKLMFAQMSELVND FYNSGLPSNLTGGPNPSLDYGFKGAEIAMASYCSELLYLANPVTTHVQSAEQHN QDVNSLGLISARKTAEAVEILKLMASTFLVALCQAVDLRHLEENMQAAVANAV AMAAKKTLAVEAAAPLLKVVADTPPFTYVDDATSAAYPLMQKLREVLVARALQ SGQDDEGCTAFRSISAFEAEVKLQLETDVPAVRQQFDDGVSSVPNKITDCRSFP LYNFVRSLGTKLLVGTDTRSPGEDIQLVYEAMTEGRLAAPLLRCLDGWSGTPTPI TPLIH
MpPAL 4	ATGGATAGCAGTAAACATCTGGGCCATAGCAGCGCAGGTCCGAGCCCGAGCGCAAATGGTTGGGTGGCCCCGCGTAATCACTTCGTGCGTATTCGCGATAT GATGCTGACCGTGGAAAAAAGTACCGTTGTGTTCGTGGGCCGTGCAGATCCGGCCGCACTGACCATTGAACAGGTTACCTGTGTGAGTCGCAATCTGCATGT GAAAGTGAGTATTCTGGGCATTGATTGCGGTGCCCAGGGTGAAGAAGAACTGCAGGGTGAAGGTCGCGCCACCAATGGCACCAGCGCCGTGGAACCGCTG CAGACCTGTAATCAGGCAAGCAGCGCAACCGAAACCGCCGAAAGTATTCGCTTCCTGACCGCAGGTGTGGCCAGCTGCCGCGGTCTGCGTACCATTGCACCG CTGGATCCGAGTATTACCCGTGCAGCACTGGCAGTTCGTACCAATACCCTGCTGAATGGCTATGGCCGTCATGAAGTGCGTTGGGAAATTCTGCATGCCTTCC TGAAACTGCTGAATCAGCGCATTACCCCGCAGCTGCCGACCCCGCGGTTGTATTAGTGCCAGCGGCGGTGATCTGCTGCAGCTGAGCTATCTGGCACGCCTGC TGAAAGGCCAGCCGAATGTGAAACTGCTGACCGAAAGTGGCGAAGAAATTGAAGCACTGGAAGCACTGCGTAGCATTGGCCTGAGCCAGCCGCTGAATCTG CAGCCGAAAGAAGAAATTGCCATTGTGAATGGCACCTCAGTTAGTAGCGCACTGGCAGCAACCATCTGTTATGATGTGAATGTGCTGGCACTGCTGGCCACC CTGGCAAGCGTTCTGTTCTGCGAAATTATTCAGTGGGATGGCTTCGAACCGCAGTTCCCGGAAGACTTAACCACCAGTGTTACCATGCATCCGCGCCAGCTGG AAACCGTGGACTTCATGGAGTTCGTGCTGCGCCGTCCGAGCCTGAGTCAGCTGCGCAATTGCGCCTTCTGCCATAGTAGTACCAGTGGCTGGCTGCGTAGTA GTAGTAGTATGGATAGCAGCAGCGATTGGGAACTGCATGGTCGCGAAAGCAATGGTGAAACACCTACCCCGAGCCCGCTGCCGAGTCATGATACCCTGCTG AAAGGTGTTGAAAGTGTTAAAGTTGATCTGGATCGCGCCCCGGATCTGTTCCGTTGCTGCGCCAATTGGCTGGGTGAACAGGTTGAAGGCATTCAGGCCGCC ACCGAAGCCGTTCAGATGGAAATTAATAGCGTGAACTTCCAGGGCGCATTCATTGGCCCGATGATGGAAATTCTGCGCAGTGCCATCTATGCACTGGGCAAA CTGATGAATTATCAGCTGATTGAACTGGCCCGTTATGTGGAAAAATGTAATCTGCCGGGCGAAGCAGTTCATGGTGCAAAAATTGTGATGGGCAGTTATGCC TGTGAACTGAGCTATATGGCCCATCCGCTGGGCAAACAGAGTCAGATTAGCGATATGGATCCGGATAATGCATATCAGAGCCTGGGCTTCCTGAGCGGTTGC AAAACCGCCGAAGCAGTGACCACCCTGCAGCTGATGTGTGCAACCAGCCTGGTTGCCCTGTGCCAGGTGTATGATCTGCAGCTGCTGGAAGCAAGTGCATTC GAATATCTGAAATATGGCGTTAGTCGCGCAGCAGTGTGTAATAAATTCGGCAGTGCCCTGGGCATTAAAGAAGAAGATATTGGTCGCTTCAATGGCAGCCGC CCGACCCTGGGTACCGCAGATCATGAAACACCTAAAGTGTTCACCGATCTGCTGGAAGTGCTGCGTGATTATAGCCCGTTCCATAATGATGCCCGTAGTATTA TTCCGAAACTGGAACAGGTTCTGAAAGATCATGGCATTGAAGAACAGCTGATTAAAGGCTTCCGTGAAAGTAAAGAACTGGAAGGCGGCTTCCATGAAGTG GATCGCATGATGAATGACTTCAATAAACATGTGATGCGCAATATTAGCGCCGGTCGCCGTTGTAAAGGCATCTATCCGGTGTATAGCTTCGTTCGTAAAGTGG GTACCAAAATGATGATTGGCAAAGACTTCGAAAATCGTGTTCCGGGTGAAGATGTTGGCCGTGTGTATGAAGCAATCTGTGGTGGTGAAATGGCAGGCAAT CTGCTGGCAGCCTTCCATGATATTCAGAATCATCTGTAA	MDSSKHLGHSSAGPSPSANGWVAPRNHFVRIRDMMLTVEKSTVVFVGRADP AALTIEQVTCVSRNLHVKVSILGIDCGAQGEEELQGEGRATNGTSAVEPLQTCN QASSATETAESIRFLTAGVASCRGLRTIAPLDPSITRAALAVRTNTLLNGYGRHEV RWEILHAFLKLLNQRITPQLPTRGCISASGGDLLQLSYLARLLKGQPNVKLLTESG EEIEALEALRSIGLSQPLNLQPKEEIAIVNGTSVSSALAATICYDVNVLALLATLASV LFCEIIQWDGFEPQFPEDLTTSVTMHPRQLETVDFMEFVLRRPSLSQLRNCAFC HSSTSGWLRSSSSMDSSSDWELHGRESNGETPTPSPLPSHDTLLKGVESVKVDL DRAPDLFRCCANWLGEQVEGIQAATEAVQMEINSVNFQGAFIGPMMEILRSAI YALGKLMNYQLIELARYVEKCNLPGEAVHGAKIVMGSYACELSYMAHPLGKQS QISDMDPDNAYQSLGFLSGCKTAEAVTTLQLMCATSLVALCQVYDLQLLEASAF EYLKYGVSRAAVCNKFGSALGIKEEDIGRFNGSRPTLGTADHETPKVFTDLLEVLR DYSPFHNDARSIIPKLEQVLKDHGIEEQLIKGFRESKELEGGFHEVDRMMNDFN KHVMRNISAGRRCKGIYPVYSFVRKVGTKMMIGKDFENRVPGEDVGRVYEAIC GGEMAGNLLAAFHDIQNHL
PpPAL 1	ATGCCGATGGCAATATCAGGGCATAAGGTTGACAGCCAAGGCACCGACCACTGCAACGGCGTGGACCACAGCAGCGGCCACCACGACCCCATGCTTGTGGA CTCCAGTGCGTTCTGTATCAGTGCGCCCCTGCCTGCTCTGTCCGCGGACCCCTTGAACTGGAAGCGCACCGCCAAAGCCATGGAAGGCAGCCACCTCGACGA GGTGAAGAGCTTCATCAGGACGTTCTCCGAGTCCACTTTCGTTTCCTTGGAGGGCGTCAGCCTCACCATCGCCCACGTTGCGGCGGTGGCTCGCAGGCCCGA GCTCCAGGTGCGCTTGGATGCCGCCATCGCCAAGAAGCTCGTGGACCAGAGCTCCGACTGGGTGCTGAACAAGATCATGCGCGGATCCGACATCTACGGTG TCACCACCGGCTTTGGCGCCACTTCCCATCGCCGGACGCAGCAGGGAGTGGAGCTTCAGCGGGAACTCATCAGGTTCTTGAACGCCGGAGTGGTGAATTCCG GAAACACGCTGCCGGTGGCCACCACGCGTGCTGCGATGCTGGTGCGGACCAACACATTGTTGCAGGGATTCTCCGGCATCAGGTGGGAGATTCTTGAAGCC ATGGAGAAGCTCCTCAATTCCCAAGTCACGCCCAAGTTGCCGCTCCGAGGCACCATCACAGCCTCCGGCGACCTCGTCCCGCTGTCGTACATCGCTGGTCTCC TCACTGCCAGGCCCAATTCCGAGGCGGTCACCGAGGATGGCAGCACGGTGTCGGCCGTGGAAGCATTGCAGATGGCGGGCATCGACAAGCCTTTCGAGCTG CAGCCCAAGGAGGGGCTGGCCATGGTGAATGGCACGGCCGTGGGATCGGCGCTGGCCTCCACCACCTGCTTCGACGCCAACATCCTGGCGGTCATGGCCGA AGTGCTGTCCGCGCTGTTCTGCGAAGTGATGCAGGGGAAGCCAGAGTTCGCGGACCCCTTGACGCACAAGCTGAAGCACCACCCGGGTCAGATGGAGGCCG CGGCCATCATGGAGTACCTCCTCGACGGGAGCTCCTACATGAAGGCCGCTGCCAAGTTGCACGAGACGGACCCGTTGAAGAAGCCCAAGCAGGATCGGTAC GCCCTGCGCACCTCGCCGCAGTGGCTGGGGCCTCAGGTGGAAGCCATCCGCAACGCCACGCATTCCATCCAGCGGGAGATCAACTCCGTCAACGACAACCCC CTCATCGACGCCGCGGGAGACAGAGCACTCCACGGCGGCAACTTCCAAGGCACGCCCATCGGCGTCTCCATGGACAACATGCGCCTCGCCGTGGCCGCCATC GGGAAGCTCATGTTCGCGCAATTCAGCGAGCTCGTGAACGACTTCTACAACAACGGCCTGCCATCCAACTTGAGCGGAGGGCCCAACCCCAGCCTCGACTAC GGCATGAAGGGCGCGGAGATCGCCATGGCTTCCTACCTGTCCGAGCTCAACTACCTCGCCAACCCGGTCACCACCCACGTCCAGAGCGCCGAGCAGCACAAC CAGGACGTCAACTCGCTGGGCCTCGTGTCCGCGCGCAAAACCGAGGAGGCCATCGAGATCCTCAAGCTCATGACGTCCACATTCTTGGTCGGGTTGTGCCAG GCCATCGACCTGCGCCACGTCGAGGAGACCATGCAGGCGGCCGTGAAGAGCATCGTCACGCAGGTGGCGAAGAAGACGCTGTTCATGAACCCCGCGGATG GCACCCTCCTGCCCGCCAGGTTCGCCGAGAAGGACCTCCTCAGAGTGGTGGATCACCAGCCCGTGTTCACGTACATCGACGACGCAGCCAGCGGCACGTATC CTCTCATGGAGAAGCTGCGACAGGCGCTCGTCCAGCACGCCCTGCAGACCAACAAGTCGGACAAGCAGGCGGCCAGCTCGACCTTCAATCTCATCGCGGCGT TCGAGGAGGAACTCAAGGTGAGGCTGCAAACGGAGGTGCCACAGATTCGGGAAGCATACGACAACGAGGGCTACTCGCCGGTGCCCAACAGGATCGAGAA CTGCAGAACGTACCCGCTGTACAAGTTCGTGCGCAACGGATTGAAGACGCAGCTTCTGTCGGGACTGCGCACCATCTCACCCGGCCAGGAAATCGAGAAGGT ATACGACGCCATCTGCGAGGGCAAGCACGTGGCTCCGCTGCTGGAGTGCATCGGAGGATGGAACGGAGCACCGGGACCCTTCTCATGCTGA	MPMAISGHKVDSQGTDHCNGVDHSSGHHDPMLVDSSAFCISAPLPALSADPL NWKRTAKAMEGSHLDEVKSFIRTFSESTFVSLEGVSLTIAHVAAVARRPELQVRL DAAIAKKLVDQSSDWVLNKIMRGSDIYGVTTGFGATSHRRTQQGVELQRELIRF LNAGVVNSGNTLPVATTRAAMLVRTNTLLQGFSGIRWEILEAMEKLLNSQVTP KLPLRGTITASGDLVPLSYIAGLLTARPNSEAVTEDGSTVSAVEALQMAGIDKPFE LQPKEGLAMVNGTAVGSALASTTCFDANILAVMAEVLSALFCEVMQGKPEFAD PLTHKLKHHPGQMEAAAIMEYLLDGSSYMKAAAKLHETDPLKKPKQDRYALRT SPQWLGPQVEAIRNATHSIQREINSVNDNPLIDAAGDRALHGGNFQGTPIGVS MDNMRLAVAAIGKLMFAQFSELVNDFYNNGLPSNLSGGPNPSLDYGMKGAEI AMASYLSELNYLANPVTTHVQSAEQHNQDVNSLGLVSARKTEEAIEILKLMTST FLVGLCQAIDLRHVEETMQAAVKSIVTQVAKKTLFMNPADGTLLPARFAEKDLL RVVDHQPVFTYIDDAASGTYPLMEKLRQALVQHALQTNKSDKQAASSTFNLIA AFEEELKVRLQTEVPQIREAYDNEGYSPVPNRIENCRTYPLYKFVRNGLKTQLLSG LRTISPGQEIEKVYDAICEGKHVAPLLECIGGWNGAPGPFSC
PpPAL2	ATGGCACCACGAGCTGGGGAGCTTAATGTCGTCGGATCAGGTGAGCAAATTGATGCGGCATCTTCAGACATGATGCTTGATGACAGCACATCTCCGTACTGC ATTGGCCTACTCGGTAATGGAGCAGTTCATGGTGCGGATCCTCTCAACTGGGCCAAGACCGCCAAGGCCATGGAATGCAGCCACCTCGAAGAAATCAAGCG GATGGTCGACACTTACCAAAATGCCACCCAGGTCATGATCGAGGGAGCCACCCTCACTGTACCTCAGGTAGCAGCCATTGCTCGCAGACCTGACGTGCATGT CGTGCTCGATGCTGCCAATGCCAAGAGTAGAGTGGACGAGAGCTCCAACTGGGTGCTCGACAGGATCATGGGCGGAGGTGACATCTACGGCGTTACCACCG GGTTCGGAGCGACCTCTCATCGTCGCACGCAGCAAGGAGTGGAGCTCCAGCGCGAACTCATCCGCTTCCTGAATGCGGGAGTCCTCAGCAAGGGTAACTCCC TGCCATCCGAGACTGCCCGAGCGGCGATGCTAGTGCGTACCAACACGCTGATGCAGGGCTATTCTGGCATCCGCTGGGAGATCCTCCAAGCTATGGAGAAGC TACTCAACGCGCACGTCACCCCCAAGTTGCCCTTGCGGGGTACCATCACGGCATCCGGCGACCTCGTCCCCTTGTCGTATATCGCCGGCCTGCTTACCGGTCG CCCCAATTCCAAAGCCGTCACCGAGGACGGACGAGAGGTTTCCGCGCTGGAGGCCTTGCGCATCGCAGGCGTGGAGAAGCCCTTCGAGCTGGCGCCCAAGG AAGGGCTCGCCCTTGTCAACGGCACCGCAGTCGGTTCCGCCCTCGCTTCCACCGTGTGCTACGATGCCAACATCATGGTGTTGCTCGCCGAGGTTCTGTCCGC GCTCTTTTGCGAGGTAATGCAGGGCAAGCCCGAATTTGCCGACCCCCTGACTCACAAGCTGAAGCACCACCCCGGACAGATGGAAGCTGCGGCTGTCATGGA GTGGGTGCTGGACGGGAGTTCCTTCATGAAGGCTGCCGCCAAGTTCAATGAGACCGACCCCCTGCGCAAGCCCAAGCAGGACAGATACGCGCTGCGAACTT CCCCGCAGTGGCTGGGCCCTCAGGTGGAGGTCATCCGCAACGCCACCCACGCCATCGAGCGGGAGATCAACTCTGTGAACGACAATCCCATCATCGACGCTG CCAGGGGCATCGCCCTGCACGGCGGCAACTTCCAGGGCACGCCCATCGGCGTGTCTATGGACAACACGCGCCTGGCGCTGGCCGCCATCGCTAAGCTCATGT TCGCACAATTCTCCGAACTTGTGAACGACTACTACAACAACGGATTACCGTCCAATTTGAGCGGGGGACCCAACCCCAGCCTGGATTACGGCATGAAAGGGG CAGAGATCGCCATGGCCTCGTACTTATCTGAGATTAACTACCTAGCCAACCCGGTCACCACCCATGTCCAAAGCGCGGAGCAACACAACCAAGATGTGAACTC ACTGGGATTGGTGTCAGCCCGCAAAACGGAGGAAGCCATGGAGATTCTCAAGCTCATGAGCGCCACCTTCTTGGTAGGCTTGTGCCAGGCTGTCGACCTGCG TCATGTCGAAGAGACCATGCAGTCGGCCGTGAAGCAAGTAGTGACACAGGTGGCCAAGAAGACGCTCTTCATGGGCAGCGACGGATCACTGCTGCCATCAC GCTTCTGCGAAAAGGAACTACTGATGGTAGTGGATCGTCAGCCGGTGTTCAGCTACATCGACGATTCCACCAGTGATTCATACCCTCTAATGGAGAAACTGCG AGGGGTGCTGGTGAGCCGTGCCTTGAAGAGTGCCGACAAGGAGACCTCCAACGCAGTGTTTAGGCAGATTCCAGTGTTTGAGGCAGAGCTGAAGCTGCAGC TGAGCCGTGTGGTGCCCGCGGTGCGAGAAGCATATGACACAAAGGGACTATCCTCGGTTCCCAACAGGATTCAAGACTGTAGGACGTATCCATTATACAAGC TAGTGCGCGGAGACTTGAAGACGCAGCTCCTGTCCGGGCAGAGGACGGTGTCTCCTGGGCAAGAGATTGAGAAGGTGTTCAATGCCATTTCTGCAGGCCAA TTGGTGGCTCCTTTGTTGGAATGCGTTCAAGGCTGGACCGGCACACCAGGGCCCTTCCCCGCTCGCACCTCTTGCTAG	MAPRAGELNVVGSGEQIDAASSDMMLDDSTSPYCIGLLGNGAVHGADPLNW AKTAKAMECSHLEEIKRMVDTYQNATQVMIEGATLTVPQVAAIARRPDVHVVL DAANAKSRVDESSNWVLDRIMGGGDIYGVTTGFGATSHRRTQQGVELQRELI RFLNAGVLSKGNSLPSETARAAMLVRTNTLMQGYSGIRWEILQAMEKLLNAHV TPKLPLRGTITASGDLVPLSYIAGLLTGRPNSKAVTEDGREVSALEALRIAGVEKPF ELAPKEGLALVNGTAVGSALASTVCYDANIMVLLAEVLSALFCEVMQGKPEFAD PLTHKLKHHPGQMEAAAVMEWVLDGSSFMKAAAKFNETDPLRKPKQDRYAL RTSPQWLGPQVEVIRNATHAIEREINSVNDNPIIDAARGIALHGGNFQGTPIGV SMDNTRLALAAIAKLMFAQFSELVNDYYNNGLPSNLSGGPNPSLDYGMKGAEI AMASYLSEINYLANPVTTHVQSAEQHNQDVNSLGLVSARKTEEAMEILKLMSA TFLVGLCQAVDLRHVEETMQSAVKQVVTQVAKKTLFMGSDGSLLPSRFCEKELL MVVDRQPVFSYIDDSTSDSYPLMEKLRGVLVSRALKSADKETSNAVFRQIPVFE AELKLQLSRVVPAVREAYDTKGLSSVPNRIQDCRTYPLYKLVRGDLKTQLLSGQR TVSPGQEIEKVFNAISAGQLVAPLLECVQGWTGTPGPFPARTSC
PpPAL 3	ATGGCACCACGAGCTGAGGAGCTTAATGTCGTCGGATCAGGTGAGCAAATTGATGCGGCATCTTCAGACATGATGATGCTCGATGATAACACATCTCCGTAC TGCATTGGCCTACTCGGTAATGGAGCAGTTCATGGTGCGGATCCTCTCAACTGGGCCAAGACCGCCAAGGCCATGGAATGCAGCCACCTCGAAGAAATCAAG CGGATGGTCGACACTTACCAAAATGCCACCCAGGTCATGATCGAGGGAGCCACCCTCACCGTACCTCAGGTAGCAGCCGTTGCTCGCAGACCCGACGTGCAT GTCGTGCTCGATGCTGCCAATGCCAAGAGTAGAGTGGACGAGAGCTCCAACTGGGTGCTCGACAGGATCATGGGCGGAGGTGACATCTACGGCGTTACCAC CGGGTTCGGAGCGACCTCTCATCGTCGCACGCAGCAAGGAGTGGAGCTCCAGCGCGAACTCATCCGCTTCCTGAATGCGGGAGTCCTCAGCAAGGGTAACTC CCTGCCATCCGAGACTGCCCGAGCGGCGATGCTAGTGCGTACCAACACGCTGATGCAGGGCTATTCTGGCATCCGCTGGGAGATCCTCCAAGCTATGGAGAA GCTACTCAACGCGCACGTCACCCCCAAGTTGCCCTTGCGGGGTACCATCACGGCATCCGGCGACCTCGTCCCCTTGTCGTATATCGCCGGCCTGCTTACCGGT CGCCCCAATTCCAAAGCCGTCACCGAGGACGGACGAGAGGTTTCCGCGCTGGAGGCCTTGCGCATCGCAGGCGTGGAGAAGCCCTTCGAGCTGGCGCCCAA GGAAGGGCTCGCCCTTGTCAACGGCACCGCAGTCGGTTCCGCCCTCGCTTCCACCGTGTGCTACGATGCCAACATCATGGTGTTGCTCGCCGAGGTTCTGTCC GCGCTCTTTTGCGAGGTAATGCAGGGCAAGCCCGAATTTGCCGACCCCCTGACTCACAAGCTGAAGCACCACCCCGGACAGATGGAAGCTGCGGCTGTCATG GAGTGGGTGCTGGACGGGAGTTCCTTCATGAAGGCTGCCGCCAAGTTCAATGAGACCGACCCCCTGCGCAAGCCCAAGCAGGACAGATACGCGCTGCGAAC TTCCCCGCAGTGGCTGGGCCCTCAGGTGGAGGTCATCCGCAACGCCACCCACGCCATCGAGCGGGAGATCAACTCTGTGAACGACAATCCCATCATCGACGC TGCCAGGGGCATCGCCCTGCACGGCGGCAACTTCCAGGGCACGCCCATCGGCGTGTCTATGGACAACACGCGCCTGGCGCTGGCCGCCATCGCTAAGCTCAT GTTCGCACAATTCTCCGAACTTGTGAACGACTACTACAACAACGGATTACCGTCCAATTTGAGCGGGGGACCCAACCCCAGCCTGGATTACGGCATGAAAGG GGCAGAGATCGCCATGGCCTCGTACTTATCTGAGATTAACTACCTAGCCAACCCGGTCACCACCCATGTCCAAAGCGCGGAGCAACACAACCAAGATGTGAA CTCACTGGGATTGGTGTCAGCCCGCAAAACGGAGGAAGCCATGGAGATTCTCAAGCTCATGAGCGCCACCTTCTTGGTAGGCTTGTGCCAGGCTGTCGACCT GCGTCATGTCGAAGAGACCATGCAGTCGGCCGTGAAGCAAGTAGTGACACAGGTGGCCAAGAAGACGCTCTTCATGGGCAGCGACGGATCACTGCTGCCAT CACGCTTCTGCGAAAAGGAACTACTGATGGTAGTGGATCGTCAGCCGGTGTTCAGCTACATCGACGATTCCACCAGTGATTCATACCCTCTAATGGAGAAACT GCGAGGGGTGCTGGTGAGCCGTGCCTTGAAGAGTGCCGACAAGGAGACCTCCAACGCAGTGTTTAGGCAGATTCCAGTGTTTGAGGCAGAGCTGAAGCTGC AGCTGAGCCGTGTGGTGCCCGCGGTGCGAGAAGCATATGACACAAAGGGACTATCCTCGGTTCCCAACAGGATTCAAGACTGTAGGACGTATCCATTATACA AGCTAGTGCGCGGAGACTTGAAGACGCAGCTCCTGTCCGGGCAGAGGACGGTGTCTCCTGGGCAAGAGATTGAGAAGGTGTTCAATGCCATTTCTGCAGGC CAATTGGTGGCTCCTTTGTTGGAATGCGTTCAAGGCTGGACCGGCACACCAGGGCCCTTCCCCGCTCGCGCCTCTTGCTAG	MAPRAEELNVVGSGEQIDAASSDMMMLDDNTSPYCIGLLGNGAVHGADPLN WAKTAKAMECSHLEEIKRMVDTYQNATQVMIEGATLTVPQVAAVARRPDVH VVLDAANAKSRVDESSNWVLDRIMGGGDIYGVTTGFGATSHRRTQQGVELQR ELIRFLNAGVLSKGNSLPSETARAAMLVRTNTLMQGYSGIRWEILQAMEKLLNA HVTPKLPLRGTITASGDLVPLSYIAGLLTGRPNSKAVTEDGREVSALEALRIAGVE KPFELAPKEGLALVNGTAVGSALASTVCYDANIMVLLAEVLSALFCEVMQGKPE FADPLTHKLKHHPGQMEAAAVMEWVLDGSSFMKAAAKFNETDPLRKPKQDR YALRTSPQWLGPQVEVIRNATHAIEREINSVNDNPIIDAARGIALHGGNFQGTPI GVSMDNTRLALAAIAKLMFAQFSELVNDYYNNGLPSNLSGGPNPSLDYGMKG AEIAMASYLSEINYLANPVTTHVQSAEQHNQDVNSLGLVSARKTEEAMEILKLM SATFLVGLCQAVDLRHVEETMQSAVKQVVTQVAKKTLFMGSDGSLLPSRFCEK ELLMVVDRQPVFSYIDDSTSDSYPLMEKLRGVLVSRALKSADKETSNAVFRQIPV FEAELKLQLSRVVPAVREAYDTKGLSSVPNRIQDCRTYPLYKLVRGDLKTQLLSG QRTVSPGQEIEKVFNAISAGQLVAPLLECVQGWTGTPGPFPARASC
PpPAL 4	ATGGCACCACGAGCTGGGGAGCTTATTGTCGTCGGATCAGGTGAGCAAATTGATGCGGCATCTTCAGACATGATGCTTGATGATGATACATCTCCGTACTGC ATTGGCCTACTCGGTAATGGACCAGTTCATGGTGTGGATCCTCTCAACTGGGCCAAAACCGCCAAGGCCATGGAATGCAGCCACCTCGAAGAAATCAAGCGG ATGGTCGACACTTACCAAAATGCCACCCAGGTCATGATCGAGGGAAACACCCTCACTGTACCTCAGGTAGCAGCCATTGCACGCAGACCTGACGTGCGTGTC ATGCTCGATGCTGCCAATGCCAAGAGTAGAGTGGTCAAGAGCTCCAACTGGGTGCTCGACAGGATCATGGGTGGAGGTGACATCTACGGCGTTACCACCGG GTTCGGAGCGACCTCTCATCGTCGCACGCAGCAAGGAGTGGAGCTCCAGCGCGAACTCATCCGCTTCCTGAATGCGGGAGTCCTCAGCAAGGGTAACTCCCT GCCATCCGAGGCTGCCCGAGCAGCCATGCTAGTGCGTACCAACACGCTGATGCAGGGCTATTCTGGCATCCGCTGGGAGATCCTCCAAGCTATGGAGAAGCT ACTCAACGCGCACGTCACCCCCAAGTTGCCCTTGCGGGGTACCATCACGGCATCCGGCGACCTCGTCCCCTTGTCGTATATCGCCGGCCTACTCACCGGTCGC CCCAATTCCAAAGCCGTCACCGAGGACGGACGAGAGGTTTCCGCGCTGGAGGCCTTGCGCATCGCAGGCGTGGAGAAGCCCTTCGAGCTGGCGCCCAAGGA AGGGCTCGCCCTTGTCAACGGCACCGCAGTCGGTTCTGCCCTCGCTTCTACCGTGTGCTACGATGCCAACATCATGGTATTGCTTGCCGAGGTTTTGTCCGCGC TCTTTTGCGAGGTAATGCAGGGCAAGCCTGAATTTGCCGACCCCCTGACTCACAAGCTGAAGCACCACCCCGGACAGATGGAAGCTGCGGCTGTCATGGAGT GGGTGCTGGACGGGAGTTCCTTCATGAAGGCTGCCGCCAAGTTCAACGAGACCGACCCCCTGCGCAAGCCCAAGCAGGACAGATACGCGCTGCGCACTTCT CCGCAGTGGCTGGGCCCTCAGGTGGAGGTCATCCGCAACGCCACCCACGCCATCGAGCGGGAGATCAACTCTGTGAACGACAATCCCATCATCGACGCTGCC AGGGGCATCGCCCTGCACGGCGGCAACTTCCAGGGCACGCCCATCGGCGTTTCTATGGACAACATGCGCCTGGCGCTGGCCGCCATCGCTAAGCTCATGTTC GCACAATTCTCCGAACTTGTGAACGACTACTACAACAACGGATTACCGTCCAATTTGAGCGGAGGACCCAACCCCAGCCTGGATTACGGCATGAAAGGGGCA GAGATCGCCATGGCCTCGTACTTATCTGAGATTAACTATCTAGCCAACCCGGTCACCACCCATGTCCAAAGCGCGGAGCAGCACAACCAAGATGTGAACTCAC TGGGATTGGTGTCAGCCCGCAAGACGGAGGAAGCCATGGAGATTCTCAAGCTCATGAGCGCCACCTTCTTGGTAGGCTTGTGCCAGGCTATCGACCTGCGTC ACGTCGAAGAGACCATGCAGTCGGCCGTGAAGCAAGTAGTGACACAGGTGGCGAAGAAGACGCTCTTCATGGGCAGCGACGGATCACTGCTGCCATCACGC TTCTGCGAGAAGGAACTACTAATGGTAGTGGATCGTCAGCCGGTGTTCAGCTACATCGACGATTCCACCAGTGATTCATACCCGCTAATGGAGAAACTGCGA GGGGTGCTGGTGAGCCGTGCCTTGAAGAGTGCAGACAAGGAGACCTCCAACGCAGTATTTAGGCAGATTCCAGTGTTTGAGGCAGAGCTGAAGCTGCAGCT GAACCGTGTGGTGCCCGCGGTGCGAGAAGCATATGACACAAAGGGACTATCCTTGGTCTCCAACAGGATTCAAGACTGTAGGACGTATCCATTGTACAAGCT	MAPRAGELIVVGSGEQIDAASSDMMLDDDTSPYCIGLLGNGPVHGVDPLNWA KTAKAMECSHLEEIKRMVDTYQNATQVMIEGNTLTVPQVAAIARRPDVRVML DAANAKSRVVKSSNWVLDRIMGGGDIYGVTTGFGATSHRRTQQGVELQRELIR FLNAGVLSKGNSLPSEAARAAMLVRTNTLMQGYSGIRWEILQAMEKLLNAHVT PKLPLRGTITASGDLVPLSYIAGLLTGRPNSKAVTEDGREVSALEALRIAGVEKPFE LAPKEGLALVNGTAVGSALASTVCYDANIMVLLAEVLSALFCEVMQGKPEFADP LTHKLKHHPGQMEAAAVMEWVLDGSSFMKAAAKFNETDPLRKPKQDRYALR TSPQWLGPQVEVIRNATHAIEREINSVNDNPIIDAARGIALHGGNFQGTPIGVS MDNMRLALAAIAKLMFAQFSELVNDYYNNGLPSNLSGGPNPSLDYGMKGAEI AMASYLSEINYLANPVTTHVQSAEQHNQDVNSLGLVSARKTEEAMEILKLMSA TFLVGLCQAIDLRHVEETMQSAVKQVVTQVAKKTLFMGSDGSLLPSRFCEKELL MVVDRQPVFSYIDDSTSDSYPLMEKLRGVLVSRALKSADKETSNAVFRQIPVFE AELKLQLNRVVPAVREAYDTKGLSLVSNRIQDCRTYPLYKLVRGDLKTQLLSGQR TVSPGQEIEKVFNAI**SAG**QLVAPLLECVQGWTGTSGPFLSRASC

	AGTGCGCGGAGACTTGAAGACGCAGCTCCTGTCTGGGCAGAGGACGGTGTCTCCGGGGCAGGAGATTGAGAAGGTATTTAATGCCATTTCTGCCGGCCAAT TGGTGGCTCCTTTGTTGGAATGCGTTCAAGGCTGGACCGGCACATCAGGGCCCTTCCTCTCTCGCGCCTCTTGCTAG	
PpPAL 5	ATGCTGATGAGCAAACGTCCGAAAATTGCCGCCCTGGATGGTGACATGAGCGCCATTATTAGTAGTAAAGGTCATATTGGTTACGGTACCACCGTGAGCAGT CCGACCTTTCAGCTGGAACAGCCGTATACCAATGGTCTGATTGGTGCAGCCTTTCCGGAAAAATGTATGGTTGGCTTTCGTGATGTTAATGCACAGAGTGAAG ATCCGCTGAATTGGGGCAAAGTTGCAGCAGGCCTGGCCGGTAGCCATCTGCAGGAAGTTCGCCAGATGGTTACCGCATTTTTCGAATGTAGCGTGGTTGTTC TGGAAGGCGCCAGTCTGACCATTGCCCAGGTTGCAGCAATTGCCCGTCGTCCGGAAGTGAAAGTTGTGCTGGATGCAGATACCGCCAAAGAACGCGTTGAT GAAAGCAGTGATTGGGTTCTGAGTTGTGCCCTGAAAGGTACCGATACCTATGGTGTGACCACCGGCTTTGGTGCAACCAGCCATCGCCGCACCAATCAGGTT GCAGAACTGCAGTTTGAACTGATTCGCTTTCTGAATGCCGGCATTGTTGGCAAAGGTGACGGTAATACCCTGCCGAGCCCGACCACCCGTGCCGCAATGCTG GTTCGTACCAATACCCTGATGCAGGGTTATAGCGGCATTCGTTGGGAAATTCTGGTTGCCATTAGTAAACTGCTGAATGCCAATGTTACCCCGAAACTGCCGC TGCGCGGTACCATTACCGCAAGTGGCGATCTGGTGCCGCTGAGTTATATTGCAGGCTGTCTGACCGGTCGCCAGAATAGTCGTGCATATACCGCAGAAGGCA AAGAAGTGACCAGTGAAGAAGCACTGCGCATTGCCGGTGTGGGTAAACCGTTTCAGCTGGAGCCGAAAGAAGGTCTGGCACTGGTTAATGGTACCGCCGTT GGTAGCGCACTGGCAAGTACCGTTTGTTTTGATGCCAATGTTCTGGTTGTTATGGCAGAAGTGCTGAGCGCACTGTTTTGTGAAGTGATGCAGGGTAAACCG GAATTTGCAGATCCGCTGACCCATAAACTGAAACATCATCCGGGTCAGATGGAAGCAGCCGCCATGATGGAATGGGTTCTGGATGGCAGCAGTTATATGAAA GCAGCCGCCGAAGCCCATGCCACCGATCCGCTGAGCAAACCGAAACAGGATCGTTATGCCCTGCGTACCAGTCCGCAGTGGCTGGGCCCGCATATTGAAACC ATTCGTGCCGCCACCCATAGCATTCAGCGCGAAATTAATAGTGTGAATGATAATCCGCTGATTGATGTGAGCCGTGATATGGCACTGCATGGCGGTAATTTTC AGGGTACCCCGATTGGTGTGAGCATGGATAATATGCGTCTGAGTATTGCAGGTATTGGTAAACTGATGTTTGCACAGTTTAGCGAACTGGTGAATGATTTTTA TAACAATGGTCTGCCGAGTAATCTGAGTGGTGGCCCGAATCCGAGCCTGGATTATGGCATGAAAGGTGGCGAAATTGCCATGGCCAGCTATACCAGCGAACT GCAGTATCTGGCAAATCCGGTTACCACCCATGTGCAGAGCGCAGAACAGCATAATCAGGATGTGAATAGCCTGGGTCTGATTAGTGCACGCAAAACCGCAG AAGCAATTGAAATTCTGAAACTGATGACCAGTACCTATCTGGTGGCCCTGTGCCAGGCAATTGATCTGCGTCATCTGGAAGAAACCATGCAGGCAACCGTTA AAACCGTGGTTGCACAGGCCGCAAAAAAGACCCTGTTTCTGGGTACCAATGGCGCCCTGCAGGTGGCACGTTTTTGTGAAAAAGAACTGCTGAAAGTGGTTG AACGCCAGCCGGTTTTTACCTATGTTGATGATCCGGCCAATACCAAATATCCGCTGATGGAAAAACTGCGTCAGGTTCTGGTGGAATATAGTCTGAAATATAT TACCGAAAAGGACGATGGCCTGAGTATTTTTAATAAGATTCCGGCCTTTGAAGAAGAAGTGAAAACCCAGCTGCGCCTGGAAGTTGCCGCAGCACGTGCAA GTTTTGATAATGGCGTGAGTCCGATTCCGAATCGTGTGATGGATTGCCGTAGTTATCCGCTGTATCAGTTTGTTCGCAGTGAACTGGGCACCGCACTGCTGAC CGGTCAGAGTAGCCAGAGCCCGGGCACCGATTTTGAAAAAGTGTATGATGCAATTTGCGAAGGTCGCCATGTTGCACATCTGATGAAAGTGCTGGAAGGTT GGAGCTGCGTGCCGGGCAGCCTGTGCGTG	MLMSKRPKIAALDGDMSAIISSKGHIGYGTTVSSPTFQLEQPYTNGLIGAAFPEK CMVGFRDVNAQSEDPLNWGKVAAGLAGSHLQEVRQMVTAFFECSVVVLEGA SLTIAQVAAIARRPEVKVVLDADTAKERVDESSDWVLSCALKGTDTYGVTTGFG ATSHRRTNQVAELQFELIRFLNAGIVGKGDGNTLPSPTTRAAMLVRTNTLMQG YSGIRWEILVAISKLLNANVTPKLPLRGTITASGDLVPLSYIAGCLTGRQNSRAYTA EGKEVTSEEALRIAGVGKPFQLEPKEGLALVNGTAVGSALASTVCFDANVLVVM AEVLSALFCEVMQGKPEFADPLTHKLKHHPGQMEAAAMMEWVLDGSSYMK AAAEAHATDPLSKPKQDRYALRTSPQWLGPHIETIRAATHSIQREINSVNDNPLI DVSRDMALHGGNFQGTPIGVSMDNMRLSIAGIGKLMFAQFSELVNDFYNNGL PSNLSGGPNPSLDYGMKGGEIAMASYTSELQYLANPVTTHVQSAEQHNQDVN SLGLISARKTAEAIEILKLMTSTYLVALCQAIDLRHLEETMQATVKTVVAQAAKKT LFLGTNGALQVARFCEKELLKVVERQPVFTYVDDPANTKYPLMEKLRQVLVEYS LKYITEKDDGLSIFNKIPAFEEEVKTQLRLEVAAARASFDNGVSPIPNRVMDCRSY PLYQFVRSELGTALLTGQSSQSPGTDFEKVYDAICEGRHVAHLMKVLEGWSCVP GSLCV
Cb4CL 1	ATGGATAAGGTGATGGATAAAGCCATGATTATGGATAAAGCAATGATTATGGACACCCTGCGCGCAGAAAGCAGCCCGAGTTGTCATCGTCATGCAAATGT GGCCCAGGCCCTGTGTGATATTGCAGAACGCCATGGCAGCCGCATGCTGACCCTGTGTGGCAGCCGCCGCCGCACAGGTCAGGAGTTCCATCGCCGCGTGA TGCGTCTGAGCATTGGCCTGCGCCGTCTGGGTCTGCGTCCGGGTCAGCGTGTGGCAATTGCAGCACTGAATAGTGATCGTTATCTGGAATGGATTCTGGCCG TGCCGATGGTTGGTGGCATTATCTGTCCGATTAATTATCGCTGGAATATTGGTGAAGCAACCGATGCAGTGTGTCAGATTACCCCGCGTATTGTTGTTATGGA TGCCCATTGCCAGAATTGGTGGCCGAAACTGGATCAGACCTGCGCAGACTTATGTACCGGTATCTTCATTGATGATGATCCGAGTGCATATCATAATCTGGTG CGCGGTGAAGAACTGATTGAAAGTGTGAAAACCATTGATGGTGCCAGCTTCCTGAGTAGCCCGGATGGCGTTGCCCTGATCTGCTTCACCAGTGGCACCACC GGTCGCCCGAAAGGTGTGGCAATTAGCCATAGCGCAATTATTAGCCAGAGTATGGCAAAACTGGATCGTGTGGGTTATAATAAACAGGATGTGTACTTCCAT GCAGCCCCGCTGTGCCATATTGGTGGTCTGAGCAGTGCATATGCAATGGTGATGGCCGGCGCACAGCATATCTTCGTGAGCCGCTTCAAAGCAGAAGAAGCA ATTGCAGCCATTCAGATGAATCAGGCAACCGCACTGATTGCCGTTCCGGCAATGCTGGTTGATATGGTGCGTGCCGCCGAAAAACTGCAGAAAATTCAGGGC AAAAGTAGTATGGGTGAAGATCGCTTCGTTCTGCGTACCGTTAGTAAACTGCTGATTGGTGGTGGTAGCATGAATCCGAGTCTGCTGAGCCATGTGTTCCAG CTGTTCCCGAGCGCCTGCATTACCAGCGCATATGGCATGACCGAAGCCTGTAGTAGTATGACCTTCCGCATTGTGCGTGAAAGCCCGCTGGCCCGTCCGAGCC CTGCAGTTACCGGTATTCCGAATGAACGCATGGGCGAAAGCGTGGCAGCCCTGATTGTGATTCATAGTGGCTGGGATTGGCAGGGTATTATGGCCCCGACC GAAGATGAACATGGCGATGATGATGAAGATGAAGATGGTGATGATGATAATCCGAATAATCCGTGTGGTGAACAGAGCCAGAAAGAAGAACGCCGTCAGC AGCGTGATGAAGCCAGCATGAGCATTGATCTGAGTGTTACCGCAAGCGGTCAGAATCGCGTTAGCACCCAGGACTTCCGCAATCATTGCCGCGAAATTGGCC TGACCGGCTTCAAAATTCCGGCCGTGTTCGTGGCAAAAAAAGGCAATAGTCTGCCGCAGACCAGCACCGGTAAAATTCAGAAAGATGTTGTTCGCCGCGTTA TTCTGCGTGCCATGCATAAAGCAAATAGCAGTGAAAATCAGCTGCGTATGAAAATTATTAGTGGTCAGCTGCGTCCGCATTGGGAACTGCGTAGCCGTCTGT AA	MDKVMDKAMIMDKAMIMDTLRAESSPSCHRHANVAQALCDIAERHGSRML TLCGSRRRTGQEFHRRVMRLSIGLRRLGLRPGQRVAIAALNSDRYLEWILAVPM VGGIICPINYRWNIGEATDAVCQITPRIVVMDAHCQNWWPKLDQTCADLCTGI FIDDDPSAYHNLVRGEELIESVKTIDGASFLSSPDGVALICFTSGTTGRPKGVAISH SAIISQSMAKLDRVGYNKQDVYFHAAPLCHIGGLSSAYAMVMAGAQHIFVSRF KAEEAIAAIQMNQATALIAVPAMLVDMVRAAEKLQKIQGKSSMGEDRFVLRTV SKLLIGGGSMNPSLLSHVFQLFPSACITSAYGMTEACSSMTFRIVRESPLARPSPA VTGIPNERMGESVAALIVIHSGWDWQGIMAPTEDEHGDDDEDEDGDDDNPN NPCGEQSQKEERRQQRDEASMSIDLSVTASGQNRVSTQDFRNHCREIGLTGFK IPAVFVAKKGNSLPQTSTGKIQKDVVRRVILRAMHKANSSENQLRMKIISGQLR PHWELRSRL
Cb4CL 2	ATGGTGCAGGCAAAAAAATGAAATGCTGAGTACCCTGCCGGCCCTGCTGCGTAGCGTTGGCAAAGCAGATGATGTGGCAGTTATTGTGGCCAGTGGCAGTCC GCAGCTGAGTTATGCAGAACTGCATGCCGAAGTTGATCGCGTTGCAGCAGTTCTGCGTAGTAGCTTCGGCGTGGCACAGGGTAGCGTGGTGAGTATGGCCT TCCCGAATACCATTGAACTGGTGATTCTGTTCCTGGCCGTGACCCGCGCACGCGCTATTGCCGCACCGCTGAATGCCGCCTATAAAGAAGAAGAGTTCCGCTT CTATATGGAAGATGCAGGTGCAGCATGCCTGATTATTCCGGCAAAAGCCGGCGGCGGCAGCGCAGAAAAAGCCGGTACCACCCTGGGTATTCCGGTGGTTG CCGCAGAATGGATGACCAGCAGCGAAAGCGCACAGGCACTGGGCCTGAAACTGGTTAGTGGTGCCGTTAAAGGCGAAAGTAGCATTGATGATGATGATGG TCCGGGCCCGCTGCCGGAACCGGATGATGTGGCGCTGTTCCTGCATACCAGCGGCACCACCAGCCGCCCGAAAGGCGTTCCGCTGACCCATGGCAATCTGAC CGCCAGTATGAAAAATATTATTCGTACCTATGACCTGCAGAGTACCGATCGCACCCTGATTGTGATGCCGCTGTTCCATGTTCATGGTCTGATGGCCGCCCTGC TGGCCACCCTGGCTGCTGGTGGTACCGTTGTGCTGCCGGCCGCCGGTCGCTTCAGCGCAACAACCTTCTGGCATGATATTGCAACCTATGGTGTGACCTGGTA TACCGCCGTGCCGACCATTCATCAGATTCTGCTGAGCCATCATCATCGCAGCCCGCTGGTGGAATATCCGCGCCTGCGCTTCATTCGTAGTTGTAGTAGCAGTC TGCCGCCGCCGGTGCTGCATAGTCTGGAAGAAGAGTTCAAAGCCCCGGTTCTGGAAGCATATGCAATGACCGAAGCAAGTCATCAGATGACCAGTAATCCGC TGCCGGAGCATGGTGGTCATAAACCGGGTAGTGTGGGCCGTCCGACCGGTATTGAACTGGCCGTGATTGATAGTGATGGCAAAATGCTGGAAACCGGTCGC ATTGGTGAAGTGTGTGTGAAAGGCGCCAATGTGACCAAAGGTTATGTGAATAATGAAGAAGCCAATAAAGCAGCATTCACCGGCGGCTGGTTCCATACCGG TGATCGCGGCTTCCTGGATGCCGATGGCTATCTGATGCTGACCGGCCGTATTAAAGAACTGATTAATCGCGGTGGTGAAAAAATTAGCCCGAGCGAAGTTGA TAATGCACTGCTGAGCCACCCGGCCGTTAGCGAAGCCGTTGCCTTCGCCGCCCCGGATGAAAAATATGGTGAAGTTGTTAATGCCGCAGTTGTTCTGCATCCG GGCGCAGTTGCCAGCAGCAAAGATATTATTGAACATTGCAAAACCAAGGTTGCACCGTTCAAAGTGCCGGCACATGTGTTCATTACCGATAGCGTTCCGCGC ACCGGCACCGGTAAAATTCAGCGTCGCTTCGTGGCCGAAAAATTCCTGAAAGAAGCCCAGCAGCAGCATGGTCCGGCCGCCAATCCGGTGGAAGCCAGTGC CAAAAGCGGCAAAGATGCCAGTTAA	MVQAKNEMLSTLPALLRSVGKADDVAVIVASGSPQLSYAELHAEVDRVAAVLR SSFGVAQGSVVSMAFPNTIELVILFLAVTRARAIAAPLNAAYKEEEFRFYMEDAG AACLIIPAKAGGGSAEKAGTTLGIPVVAAEWMTSSESAQALGLKLVSGAVKGES SIDDDDGPGPLPEPDDVALFLHTSGTTSRPKGVPLTHGNLTASMKNIIRTYDLQS TDRTLIVMPLFHVHGLMAALLATLAAGGTVVLPAAGRFSATTFWHDIATYGVT WYTAVPTIHQILLSHHHRSPLVEYPRLRFIRSCSSSLPPPVLHSLEEEFKAPVLEAY AMTEASHQMTSNPLPEHGGHKPGSVGRPTGIELAVIDSDGKMLETGRIGEVCV KGANVTKGYVNNEEANKAAFTGGWFHTGDRGFLDADGYLMLTGRIKELINRG GEKISPSEVDNALLSHPAVSEAVAFAAPDEKYGEVVNAAVVLHPGAVASSKDIIE HCKTKVAPFKVPAHVFITDSVPRTGTGKIQRRFVAEKFLKEAQQQHGPAANPVE ASAKSGKDAS
Mp4CL 1	ATGACCAGTATCTTCAGTCGTCGCGCACTGGCAAGCAGCATTAGTGAACGTCGCAAAAGCGTTATGGCCCGTAGTCTGATGATGGCCGAACGTGAATATATC TTCGAAAGTAAACTGCCGGCCTATCCGATTGCAGATAGCGCCATTGATATGACCCTGCCGGAGTTCGTTCTGGATGGTGCCGAACAGGACTTCGATCGCGTT GCCATTGTGGATAGTAGTGATGGTCGCCAGTATACCTATAGCGAGTTCCGCCGTCTGGTGAAAAATGTGGCCGCAGGCCTGGCCAGTCTGGGCATTCATAAA GGTGATGTGGTGTGCTTCGTGCTGCCGAATGTGGCAGAATACTTCATTCTGGTGTTCGGCGTTATGAGTGTGGGCGCCATCTTCAGTGGCTGTAATCCGTATG CCCATCCGAGCGAAATTCTGAAACAGGTTCTGAGCAGCGATGCCAAACTGCTGATTACCGATCCGCCGACCCTGCGCAAAGTGGCAGATATTGAAGTTCCGA TTGTGATGGTTGGTGGCGAAGCAATGGCCGGCACCATTACCGTGCAGACCCTGCTGACCGCCGATGGCGCCAAAGCACCGGCAGTTAGCATGAGTCCGGAT GATGTGTGTCTGCTGCCGTATAGCAGCGGCACCACCGGCCTGCCGAAAGGTGTGATGATTACCCATCGTAATATTATTGCAAACCTGTGTCAGACCCTGTGTG TGACCCAGTGTAAATTCAGCAGCCAGCGCATGGGTCCGCATGAAACCGTTCTGGGCCTGATGCCGTTCTTCCATATCTATGGCATTAGTGGTCTGGGCTGCGC CACCATGCGCAATCGCGGTACCATTGTTGTGATGGAACGTTATGAAATTCGTCGTCTGCTGCAGGTTCTGCTGGACTTCGAAGTGACCAGCGCACCGCTGGTT CCGCCGATTATTCTGAGTCTGCTGAAAAGCCCGGTGGTGGCCGAGTTCGATCTGACCGCACTGAAACTGAGCGCAATTCTGTGCGCAGCAGCACCGCTGAGT CCGGAACTGCAGCAGGCATTCGAAGATAAATTCCCGGGTGTTCTGATCTGTCAGGCATATGGCCTGACCGAATATAGCTGCGTGACCCTGAGCCATGTGCCG CCGCCGTATGGCCTGGATGTTAAACCGAGTAAAAAAGGCAGTGTGGGCTTCCTGATGCCGGGTACCCAGATTAAATTCGTTGATCTGAGTACCGGTAAAAGT GTTCCGGCCCATGCACATGGTGAACTGTGTGTGCGTGGTCCGACCATTATGAAAGGTTATTATAAAAACGCCAAGGCAACCGAAAGCATTATTGATGCAGAT GGCTGGCTGCATACCGGCGATGTGGGCTATATTGATGATGATGGCGATGTGTATATTGTGGAACGTGTGAAAGAACTGATTAAATATAAGGGCTTCCAGGTT CCGCCGGCAGAACTGGAAGCAATTCTGATTAGCCATAGCGCCGTGACCGATGCCGCCGTGGTTGGTATTCCGGATGATGAAGCAGGTGAAATTCCGGCCGC CTGTATTGTTCTGAGCCCGAATGCATTCGTTACCGAAAAAGAAGTGAAAGATTATGTTAACAGCCAGGTGGCAAGTTATAAAAAAGTGCGCTTCGTTGAACA GGTTCCGAAAATTCCGAAAAGTACCAGTGGTAAAATTCTGCGTCGCCTGCTGAAAGATCAGCTGATTCTGAAATATCGTGCACAGACCCAGCGCCTGTAA	MTSIFSRRALASSISERRKSVMARSLMMAEREYIFESKLPAYPIADSAIDMTLPEF VLDGAEQDFDRVAIVDSSDGRQYTYSEFRRLVKNVAAGLASLGIHKGDVVCFVL PNVAEYFILVFGVMSVGAIFSGCNPYAHPSEILKQVLSSDAKLLITDPPTLRKVADI EVPIVMVGGEAMAGTITVQTLLTADGAKAPAVSMSPDDVCLLPYSSGTTGLPK GVMITHRNIIANLCQTLCVTQCKFSSQRMGPHETVLGLMPFFHIYGISGLGCAT MRNRGTIVVMERYEIRRLLQVLLDFEVTSAPLVPPIILSLLKSPVVAEFDLTALKLS AILCAAAPLSPELQQAFEDKFPGVLICQAYGLTEYSCVTLSHVPPPYGLDVKPSKK GSVGFLMPGTQIKFVDLSTGKSVPAHAHGELCVRGPTIMKGYYKNAKATESIID ADGWLHTGDVGYIDDDGDVYIVERVKELIKYKGFQVPPAELEAILISHSAVTDAA VVGIPDDEAGEIPAACIVLSPNAFVTEKEVKDYVNSQVASYKKVRFVEQVPKIPK STSGKILRRLLKDQLILKYRAQTQRL
Mp4CL 2	ATGGGCCAGCAGAAAGGCTTCTATGATGCCGAACATGCCACCTATCATGGCCCGTGGCCGCAGGTGCATCTGCCGGAACAGGCCAATATGAATCTGGTGACC TATCTGTTCAGTCCGGCATATTATAAACCGGAGTTCGCAAGTAAAAAAATTCTGATTGATAGCGAAACCGAAGAGTTCGTTACCTATGGTGAACTGCATGAAA CCGTGAAACGTGTTGCCAATGGCCTGGCAGACTTAGGTCTGCGTAAAGGTAATGTTGTTGTTATTCTGCTGAATAATACCATCTATTACCCGATTATGTTCCTG GCAATTGTTAGCCTGGGTGCCATTGTTACCACCATGAATCCGCTGAGTAGTAAACATGAAATTGCAAATCAGATGCGTCTGAGCGGCGCCCAGTTCGCAATTA CCAGCGCCCAGTATGTGGAAAAAGTTCGCGATTATGGTATGGTTCTGATTCTGTATGGCGATGTTGGCCGCCAGCGTGAACTGGCAGACTTAAGCTATACCC CGTTCACCATTCTGCTGGCAGGTGATAGCACCGACTTCAGCGAAGTGCAGGTGCTGCAGAGTGATACCGCAGCCCTGATGTTCAGTAGTGGTACCACCGGTG TTACCAAAGGTGTGATTCTGACCCATCGTAACTTCATTGCACTGCAGGCAGCCATGAATAAATTCGAAGAAGATAGTCGCGATCCGGCAAGCGAATGTCTGG CCTATCATGATCTGATTAGCCTGGTTCCGATTCCGCTGTTCCATATCTATGGTCTGTGTACCGTTGCAGGTCTGCAGCTGCGCTGGGGTTGTACCAATGTTGTG CTGAATAAATTCGACTTCGTTGGTATGCTGCATGCAATTGAAAAATTCCGTTGTACCAGCATTCCGGCAGTGCCGCCGATTATTGTGGCACTGGCCAAACAGC CGATTGTGGAAAAATTCGATCTGAGCAGCCTGAAAATGATTGGTTGTGGTGGCGCAAGCCTGAGCAAAGGCGTTATTGAAGACTTCCGTAGCCGCTTCGCAG ATATTAAATTCAGTCAGGGTTATGGCATGACCGAAACCGTTAGTGTTATCTGTGCCAGCAATCATACCCCGAATGTGCCGGTTGCCCCGTATGGTAGTGTTGG CTATCTGATTCCGAATATTGAAGCCAAAATTATTGAGGTGGGTACCGGTAAAAGCCTGCCGCCGCGTGGCGAAGGTGAACTGTGTATTCGTGGCCCGCAGGT TATGAAAGGTTATGTTAATGATCCGGAAGCCACCAGTGCCACCATTGATGCCGATGGTTGGCTGCATACCGGCGATCTGGCCTACTTCACCGAAACCGGCGC CCTGTTCATTCTGGATCGTATTAAAGAACTGATTAAGTATAAGGCATACCAGGTGGCCCCGGCAGAAATTGAAAAAGTTCTGATTCAGCATCCGGATATTAGC GAAGCAGTTGTTGTTCCGGTGCCGGATGAAGAAGCCGGCCAGCTGCCGGTGGCATGCGTGGTTCTGAATAGCAATGGTAGTCTGACCGAAAGTGAAATTAC CAGCTTCGTGGCCAAAGAAGTGGCACCGCATAAAAAAATTCGTCGCGTGTTCTTCCTGGATGCAATTCCGCGCGGTGAAAGCGGTAAAATTCAGCGCCGCCA GCTGGCACGTAAAGCCGCAGAACATATTCGCGGCCGCACCTAA	MGQQKGFYDAEHATYHGPWPQVHLPEQANMNLVTYLFSPAYYKPEFASKKILI DSETEEFVTYGELHETVKRVANGLADLGLRKGNVVVILLNNTIYYPIMFLAIVSLG AIVTTMNPLSSKHEIANQMRLSGAQFAITSAQYVEKVRDYGMVLILYGDVGRQ RELADLSYTPFTILLAGDSTDFSEVQVLQSDTAALMFSSGTTGVTKGVILTHRNFI ALQAAMNKFEEDSRDPASECLAYHDLISLVPIPLFHIYGLCTVAGLQLRWGCTNV VLNKFDFVGMLHAIEKFRCTSIPAVPPIIVALAKQPIVEKFDLSSLKMIGCGGASLS KGVIEDFRSRFADIKFSQGYGMTETVSVICASNHTPNVPVAPYGSVGYLIPNIEA KIIEVGTGKSLPPRGEGELCIRGPQVMKGYVNDPEATSATIDADGWLHTGDLAY FTETGALFILDRIKELIKYKAYQVAPAEIEKVLIQHPDISEAVVVPVPDEEAGQLPV ACVVLNSNGSLTESEITSFVAKEVAPHKKIRRVFFLDAIPRGESGKIQRRQLARKA AEHIRGRT
Mp4CL 3	ATGGCCCCTACCATGGCCCCGCCGGCTGCTTCACCGGTGGTGGTTAGTAGTGTTCCGAGTGATGGTCAGACCCGCGAATATATCTATCGTAGTAAACTGCCG GATATTGAAATTGATAATAACCAGAATCTGGCACGTTATAGTCTGGAAAAAGTGGGCGATCATCTGGATCGCCCGGCCCTGATTGATGGCACCACCGGCCTG GAATATACCTATCGTGAAACCGAACTGCTGACCCGTAAAATTGCCGCAGGCCTGAATAATCTGGGTATTCAGAAAGGTGATGTTCTGGCACTGCTGCTGCCG AATTGCTGCGAGTTCGTTCTGGTGTTCCTGGGTAGTGCAATTCGCGGTAGTATTGTTACCACCGCCAATCCGTTCTATACCCCGGGCGAAATTGCAAAACAGC TGAATGCATGTCGTACCAAACTGATTGTGACCCAGGCAAGCTATGTTGATAAAGTGCGTGATCTGGGCTGCACCATTGTGACCATTGATAGTCCGCCGGAAG GCTGCCTGCCGATGAGTGTTCTGATGGAAGCCGATGAAAATGATTGCCCGCAGGTTGAAGTGGATCCGGATGATGTGGTGTGTCTGCCGTATAGTAGCGGC ACCACCGGTCTGCCGAAAGGCGTGATGCTGACCCATAAAAGTCTGATTAGTAGCGTGGCCCAGCAGGTTGATGGTGAAAATCCGAATCTGAATATTGTTGCA GATGATGTTATGGTGTGTCTGCTGCCGATGTTCCATATCTATAGTCTGAATAGCGTTCTGCTGTGTAGTCTGCGCGTGTGTTGTACCCTGGTTATTATGCATAA ATATGAAATCAGCCAGCTGCTGGAACTGATTCAGAAATATCGTGTTACCGTTCTGCCGGTTGTTCCGCCGATTGTGCTGGGCATTGCAAAAAGCCCGATTGTG GGCAATTATGATGTGAGTAGTGTTAAAGTTGTGCTGAGCGGTGCAGCCCCGCTGGGTAAAGACTTAGAAGATGCCTTCATTGCCAAACTGCCGGGTGCAGTT ATTGGCCAGGGTTATGGTATGACCGAAGCAGGCCCGGTTCTGAGTATGTGTCTGGCCTTCGCCAAAGATCCGTATCCGGTGAAACCGGGTAGTTGTGGTACC GTGGTGCGTAATGCCGAAGTGAAAATTGTTGATCCGGAAACCGGTGATAGCCTGCCGCATGGCCAGGCAGGCGAAATCTGTATTCGCGGCCCGCAGATTAT GAAAGGTTATCTGGATAATCCGGAAGCCACCGCCAATACCGTGGATAAAGAAGGTTGGCTGCATACCGGTGATATTGCACTGATTGATAGCGATGAAGAAG TGTTCATTGTTGATCGTGTGAAAGAAATTATCAAGTTCAAAGGCTTCCAGGTTCCGCCGGCCGAACTGGAAGGTATTCTGGTTAGTCATCCGCTGATTACCGA TGCCGCAGTGGTGCCGAAAAAAGATGAAGGCGCAGGTGAAGTGCCGGTTGCCTTCGTTGTTAAAGCCAAAGATGCAGTGCTGACCGAAGATGATGTTAAAA CCTTCGTGGCAAAACAGGTGGTGTTCTATAAAAAACTGCATAGCGTTATCTTCGTTGATAGTATTCCGAAAAGCCCGAGTGGCAAAATTCTGCGTAAAGACTT ACGCTGCCTGGTGTAA	MAPTMAPPAASPVVVSSVPSDGQTREYIYRSKLPDIEIDNNQNLARYSLEKVGD HLDRPALIDGTTGLEYTYRETELLTRKIAAGLNNLGIQKGDVLALLLPNCCEFVLVF LGSAIRGSIVTTANPFYTPGEIAKQLNACRTKLIVTQASYVDKVRDLGCTIVTIDSP PEGCLPMSVLMEADENDCPQVEVDPDDVVCLPYSSGTTGLPKGVMLTHKSLIS SVAQQVDGENPNLNIVADDVMVCLLPMFHIYSLNSVLLCSLRVCCTLVIMHKYE ISQLLELIQKYRVTVLPVVPPIVLGIAKSPIVGNYDVSSVKVVLSGAAPLGKDLEDA FIAKLPGAVIGQGYGMTEAGPVLSMCLAFAKDPYPVKPGSCGTVVRNAEVKIV DPETGDSLPHGQAGEICIRGPQIMKGYLDNPEATANTVDKEGWLHTGDIALIDS DEEVFIVDRVKEIIKFKGFQVPPAELEGILVSHPLITDAAVVPKKDEGAGEVPVAF VVKAKDAVLTEDDVKTFVAKQVVFYKKLHSVIFVDSIPKSPSGKILRKDLRCLV
Mp4CL 4	ATGATCAACCCGGGTGCAAAAATGGATGTTATTAAAACCGCAGAAGCCCATGGTTATAGCCGTGAAGATGGCATCTGGCGCAGTCAGCGTCCGAGTATTAGT ATTCCGGAAGAAGAAACCGTTGATACCGTTGGTATGTATTATCGCATGGCCCCGGATGATGATCGCCTGGCCCTGGTTGATGTGGTTAGCAAAGAAGAACTG ACCTATAAAGAGTTCAAAATTAAGTTCCGTACCGTTGCAAGCGGCCTGCATGAACTGGGTGTGCGTAAAGGCGATGTTGTGATGATTCTGAGCCCGAATAGT GTGTGGTTCCCGGTGATTATCTTCGGTTGCTGGCTGGTGGGCGCAACCGTGACCACCGTGAATCCGGTTAATACCAGCTATGAAATTGAAAAACAGCTGAAA GATAGCGGTGCCCGTCTGATCTTCACCGTGCCGGAACAGTATACCAAAGTGCAGAACTTCGGTAAACCGGTGGTTCTGATTGGTGGTACCGAACAGCAGCAT AATGTGCCGAAAGGCGCCACCCCGTTCACCAAACTGCTGAGCGCAGATCCGCGTCGTGTGCCGGAAGTGACCCGTCGCCAGAGCGAAATTGCAGCACTGCT GTATAGTAGCGGTACCACCGGCATGAGTAAAGGCGTGGTTCTGACCCATCGCAATATTATGGCCGTTATTCTGCAGCTGCTGTATGATGCCCCGCCGGAAGT GCGTGATCGCCGCTATATGGTTATGCTGCCGATGTATCATGTGTATGGTCTGGCAGTGGTGAGCATTGCCCAGTTCTTCCGCCCGGGTGCACTGATTATTATG CCGCGCTTCGATACCCAGCTGCTGCTGAGCAGCATTGAAAAATATAAAATTACCGACCTGCCGGTTGTGCCGCCGATTATTATTATGCTGGCAAAAAGCAAAG AGGTTGATCGCTATGATATGAGCAGCCTGGTGGAAATTACCACCGGTGCCGCACCGTGCGGCCTGGAACTGACCATGGAACTGCATCGTCGTCTGCCGAATG TTCGCATTCGCCAGGGCTTCGGCATGACCGAAACCAGTGGTATTGTTGGTGTGAATGTTAGCCGCCCGGATAATAGTCACTTCGGTAGTACCGGCACCCTGGT GAGTAATATGGAAGCAAAAATTATTGACGTGACCACCGGCAAAGCACTGCCGCCGTATGATCAGGGTGAAATCTGTGTGCGTGGTCCGAATGTTATGCGCG AATATCTGAATAATCCGCAGGCCACCGCAAGCACCATTAAAGAAGGTTGGCTGCATACCGGCGATCTGGGTTATTATGATGAAACCGGCCAGCTGCATGTTG TGGATCGTCTGAAAGAACTGATTAAATATAAGGGCTTCCAGGTTGCACCGGCAGAACTGGAAGCCCTGCTGCTGACCCATAGTAAAATTCTGGATGCAGCCG TTGTTCCGCTGCCGGATGAAAAAGCAGGCGAACTGCCGGTTGCCTATGTTGTTAAAGTGGCAAATAGCAAACTGACCGAACAGGAAGTGATTGCCTTCGCAG CAGAAAAAGTTAGCCCGTATAAACGTCTGCATAAAGTGACCTTCATTGATGCCATTCCGAAAAGCCCGAGTGGCAAAATTCTGCGTCGTGAACTGACCCTGCT GGCCAAAGCAAAAGCAAAACTGTAA	MINPGAKMDVIKTAEAHGYSREDGIWRSQRPSISIPEEETVDTVGMYYRMAPD DDRLALVDVVSKEELTYKEFKIKFRTVASGLHELGVRKGDVVMILSPNSVWFPVII FGCWLVGATVTTVNPVNTSYEIEKQLKDSGARLIFTVPEQYTKVQNFGKPVVLIG GTEQQHNVPKGATPFTKLLSADPRRVPEVTRRQSEIAALLYSSGTTGMSKGVVL THRNIMAVILQLLYDAPPEVRDRRYMVMLPMYHVYGLAVVSIAQFFRPGALII MPRFDTQLLLSSIEKYKITDLPVVPPIIIMLAKSKEVDRYDMSSLVEITTGAAPCGL ELTMELHRRLPNVRIRQGFGMTETSGIVGVNVSRPDNSHFGSTGTLVSNMEAK IIDVTTGKALPPYDQGEICVRGPNVMREYLNNPQATASTIKEGWLHTGDLGYYD ETGQLHVVDRLKELIKYKGFQVAPAELEALLLTHSKILDAAVVPLPDEKAGELPVA YVVKVANSKLTEQEVIAFAAEKVSPYKRLHKVTFIDAIPKSPSGKILRRELTLLAKA KAKL
Pp4CL 1	ATGAGCGCAGGCATTGCAGAACTGGGCAGTCTGTGTCTGTGTGCAGCCCTGACCGAACGTCCGTTCAGTCGTGTTCGCAGTAGTCGCGTGTACTTCGTTAGT GACTTCCTGGTGAATCAGGAAGAAGGTGCCCGTCGTTGGCTGGATCAGGAAGCACGCGTGAGCGAAATGAGTATTAGTGTTATTGCCGAAAGTCTGATTAG CGAACCGGCCACCGTGAGCATGAAAATGAAAACCGCACAGGCCGAAGTGCGCGAAGCCCGTATTCCGCCGACCGCCGTGAGCGAACCGCTGTTACTGGAAA GCCCGATGGAAACCGAAACCGAACCGGCCCAGTATGGCCAGGTGCAGGAGTTCATCTATCGCAGCAAACTGCCGGATATTGATATTCCGAATCACATGCCGC TGGCAGATTATTGTCTGGAAAAAGCCGCACAGTGGCCGGAAAATGTGTGTCTGATTGATGGTAATACCGGTCGCAAACATACCTATGGCGAAATTGAAGTGA GTATGCGCCGTGTGGCCGCAGGCCTGGCAAATATTGGTGTGAAACAGGGTGATGTGATTGCCCTGCTGCTGCCGAATTGTGCAGAGTTCGTTCAGGTGTTCC TGGGTGCAGCAAAACGCGGCGCCGTGATTACCACCGCCAATCCGTTCTATACCAGTGCCGAACTGCGTAAACAGATTCTGGCAAGCGGTACCACCATGGTTG TGACCCAGAGTAGCTATGTTGAAAAACTGGAAGGTCTGATTGTGCAGATTGTTACCGTGGATCAGCATGTTGATGGTTGTCTGCATATTAGTGCCCTGCTGGA	MSAGIAELGSLCLCAALTERPFSRVRSSRVYFVSDFLVNQEEGARRWLDQEARV SEMSISVIAESLISEPATVSMKMKTAQAEVREARIPPTAVSEPLLLESPMETETEP AQYGQVQEFIYRSKLPDIDIPNHMPLADYCLEKAAQWPENVCLIDGNTGRKHT YGEIEVSMRRVAAGLANIGVKQGDVIALLLPNCAEFVQVFLGAAKRGAVITTAN PFYTSAELRKQILASGTTMVVTQSSYVEKLEGLIVQIVTVDQHVDGCLHISALLEA DEAECPQVEIHPDDVVCLPYSSGTTGLPKGVMLTHKGLVSSVSQQVDGEVPNF NITVEDTMMCVLPMFHIYSLNSILLCGLRVGAALVVMSKFELPKLLDLIQRYKVT VGPFVPPIVLAIAKNPIVDNYDLSSIRMVMSGAAPLGKELEDAFRARLPNAVLG

	AGCCGATGAAGCCGAATGTCCGCAGGTGGAAATTCATCCGGATGATGTGGTGTGTCTGCCGTATAGTAGTGGTACCACCGGCCTGCCGAAAGGTGTTATGCT GACCCATAAAGGCCTGGTTAGCAGCGTTAGCCAGCAGGTTGATGGTGAAGTGCCGAACTTCAATATTACCGTGGAAGATACCATGATGTGCGTGCCGAT GTTCCATATCTATAGTCTGAATAGTATCCTGCTGTGCGGTCTGCGCGTTGGTGCAGCACTGGTTGTGATGAGTAAATTCGAACTGCCGAAACTGCTGGATCTG ATTCAGCGTTATAAAGTTACCGTGGGCCCGTTCGTTCCGCCGATTGTTCTGGCCATTGCAAAAAATCCGATTGTTGATAATTACGATCTGAGTAGCATTCGTAT GGTGATGAGTGGCGCCGCCCCGCTGGGCAAAGAACTGGAAGATGCCTTCCGTGCCCGCCTGCCGAATGCAGTGCTGGGCCAGGGCTATGGTATGACCGAAG CCGGTCCGGTGCTGGCCATGTGTCTGGCATTCGCCAAAACACCGTTCCCGGTGAAACCGGGCAGCTGCGGTACCGTGGTTCGTAATGCAGAAGTGAAAATTG TTGATACCGAAACCGGCATGAGCCTGCCGTATAATCAGCCGGGCGAAATCTGTATTCGTGGCCCGCAGATTATGAAAGGCTATCTGAATAATCCGGAAGCCA CCGCCAATACCATTGATAAAGATGGCTTCCTGCATACCGGTGATGTTGCCTTCATTGATGAAGATGAAGAAATGTTCATCGTTGATCGCGTGAAAGAAATTAT TAAATTCAAGGGCTTCCAGGTGCCGCCAGCAGAACTGGAAGCCCTGCTGTTCAGCCATCCGCTGATTCAGGATGCCGCAGTGGTGAGCCGTAAAGATGATAT TGCAGGCGAAGTTCCGGTTGCCTTCGTGGTGCGTACCCCGAGTAGTATTATTAGCGAAGAAGAAGTGAAAGCATATATTGCAGATCAGGTTGTGTTCTATAA AAAAATTCACAGTGTGTACTTCGTGGATAGCATTCCGAAAAGTGCCAGCGGCAAAATTCTGCGTAAAGACTTACGCAATAAAGTTTAA	QGYGMTEAGPVLAMCLAFAKTPFPVKPGSCGTVVRNAEVKIVDTETGMSLPV NQPGEICIRGPQIMKGYLNNPEATANTIDKDGFLHTGDVAFIDEDEEMFIVDRV KEIIKFKGFQVPPAELEALLFSHPLIQDAAVVSRKDDIAGEVPVAFVVRTPSSIISEE EVKAYIADQVVFYKKIHSVYFVDSIPKSASGKILRKDLRNKV
Pp4CL 2	ATGAGCCCGAGCCTGCTGCCGCAGCCGATTGTTAGCGAAAGTACCGGTGAAAGCGTTATGAAAATGAGTCTGCAGAGCGAAGTGCGCGAAGCAAGCCTGGC AACCGGCGAAAATCCGGAACCGTTCCTGCTGGAAACCGATGCCGAAAGTCAGATTATGGAACCGGTTCATGCCGAAGTTCATGACTTCATCTATCGACGCAA ACTGCCGGATATTGATATTCCGAATCACATGCCGCTGGCCGATTATTGCCTGGAAAAAGCAGTCAGTGGCCGGATAAAGTGTGTCTGATTGATGGTGTGAC CGGTCGTGAACATCGTTATGGCGAAATTGAACTGAGCAGCCGCCGTGTGGCCGCAGGCTTAGATAAAATTGGCGTTAAACAGGGTGATGTGATTGCCCTGCT GCTGCCGAATTGCGCAGAGTTCGTGCTGGTGTTCCTGGGTGCAGCCAACGTGGCGCAGTTGTGACCACCGCAAATCCGTTCTATACCGCCGCAGAACTGGA AAAACAGATTGAAGCAAGTGGCGCAGGCATTGTTATTACCCAGAGCAGTTATATTGAAAACTGGCAGGTCTGAATGTGCAGATTATTACCGTTGATCAGCA TGTTGCAAATTGCATGCATATTAGCGTTCTGCTGAATGCCTGTGAAGATGAATGTCCGCAGGTTCGTATTCATCCGGATGATCTGGTGTGTCTGCCGTATAGC AGCGGCACCACCGGTCTGCCGAAAGGTGTGATGCTGACCCATAAAAGCCTGGTGAGTAGCGTGAGTCAGCAGGTTGATGGCGAAGCACCGAACTTCAATAT TACCGTTGAAGATACCCTGATGTGCGTGCTGCCGATGTTCCATATCTATAGTCTGAATAGCATTCTGCTGTGTGGCCTGCGCGTGGGTGCCACCCTGGTTATTA TGCCGAAATTCGAACTGCCGAAACTGCTGGATCTGATTCAGCGTCATAAAGTTACCATGGGCCCGTTCGTGCCGCCGATTGTTCTGGCAATTGCAAAAAATCC GATTGTGGAAAATTACGATCTGAGCAGTATGCGCATGGTGATGAGTGGCGCAGCCCCGCTGGGTCGCGAACTGGAAGATGCCTTCCGCGCACGCCTGCCGA ATGCAGTTCTGGGCCAGGGTTATGGCATGACCGAAGCCGGCCCGGTTCTGGCCATGTGTCTGGCATTCGCAAAAACACCGTTCAGTGTTAAACCGGGTAGTT GCGGTACCGTTGTGCGTAATGCAGAAGTGAAAATTGTTGATACCGAAACCGGCATGAGCCTGCCGTATAATCAGCCGGGTGAAATCTGTATTCGCGGCCCGC AGATTATGAAAGGTTATCTGAAAAATCCGGAAGCAACCGCCAATACCATTGATAAAGATGGCTTCCTGCATACCGGTGATGTTGCATTCATTGATGAAGATGA AGAAATGTTCATCGTTGATCGCGTTAAAGAAATTATTAAGTTCAAGGGCTTCCAGGTTCCGCCGGCAGAACTGGAGGCCCTGCTGTTAAGCCATAAAGAAATT CAGGATGCCGCAGGTGTTATGCCGCAAAGATGATGTGGCAGGCGAAGTGCCGGTTGCATTCGTGGTTCGCGCAACCAGCAGCACCATTACCGAAGATGAAGT GAAAGATTATATCGCAAAACAGGTTGTGTTCTATAAAAAAATCCACAATGTGTACTTCGTGGATAGCGTTCCGAAAAGCCGAGCGGTAAAATTCTGCGTAA AGACTTACGTAATAAGGTGTAA	MSPSLLPQPIVSESTGESVMKMSLQSEVREASLATGENPEPFLLETDAESQIMEP VHAEVHDFIYRSKLPDIDIPNHMPLADYCLEKSSQWPDKVCLIDGVTGREHRYG EIELSSRRVAAGLDKIGVKQGDVIALLLPNCAEFVLVFLGAAKRGAVVTTANPFYT AAELEKQIEASGAGIVITQSSYIEKLAGLNVQIITVDQHVANCMHISVLLNACEDE CPQVRIHPDDLVCLPYSSGTTGLPKGVMLTHKSLVSSVSQQVDGENAPNFNITVE DTLMCVLPMFHIYSLNSILLCGLRVGATLVIMPKFELPKLLDLIQRHKVTMGPFV PPIVLAIAKNPIVENYDLSSMRMVMSGAAPLGRELEDAFRARLPNAVLGQGYG MTEAGPVLAMCLAFAKTPFSVKPGSCGTVVRNAEVKIVDTETGMSLPYNQPGE ICIRGPQIMKGYLKNPEATANTIDKDGFLHTGDVAFIDEDEEMFIVDRVKEIIKFK GFQVPPAELEALLLSHKEIQDAAVVSRKDDVAGEVPVAFVVRATSSTITEDEVKD YIAKQVVFYKKIHNVYFVDSVPKSPSGKILRKDLRNKV
Pp4CL 3	ATGAGCCCGAGTGTTATGAGCGAACCGATTATGAGCGAACCTGAAAGTCAGACCAAAAGCATGAAACCGGTTCAGAGTGTTATTCGCGAACGCGCAACCGA TAGTATGGTTGTTAAAAGCATGGAAAGCGAAGTGTGTGACTTCATCTATCGTAGCAAACTGCCGGATATTGATATTCCGAATCACATGCCGCTGAGTGATTAT TGCCTGGAAAAAGCAGCCCAGTGGCCGGATAAAGTGTGTCTGATTGATGGTGTTACCGGCCGTGAACATACCTATGGCGAAATTGAACTGAGCACCCGTCGT GTTGCCGCAGGCCTGTTCAAAATTGGCGTTAAACAGGGTGATGTTATTGCCCTGCTGCTGCCGAATTGTGCAGAGTTCGTTCAGGTGTTCCTGGGTGCCGCCA AAATGGGTGCAATTGTGACCACCGCCAATCCGTTCTATACCAGCGCAGAACTGGAAAAACAGACCATTGCAAGTGGTGCCGGCATTGTTGTTACCCATAGCA GTTATATTGAAAAACTGGCCGGTCTGAATGTGCAGATTATTACCGTGGATCAGCATGTGGATAAATGCATGCATATTAGTATGCTGCTGGAACCGAATGAAG CCGAATGCCCGCAGGTGGAAATTCATCCGGATGATGTTGTGTGTCTGCCGTATAGTAGCGGCACCACCGGTCTGCCGAAAGGCGTGATGCTGACCCATAAAA GTCTGGTTAGTAGTGTTAGCCAGCAGGTGGATGGTGATAGCCCGAACTTCAATATTACCGTGGAAGATACCCTGATGTGCGTTCTGCCGATGTTCCATATCTA TAGTCTGAATAGCATTCTGCTGTGCGGCCTGCGTGTGGGTGCCACCCTGGTTATTATGCCGAAATTCGAACTGAGCAAAATGCTGGAACTGATTCAGAAACAT AAAGTGACCATGGGTCCGTTCGTGCCGCCGATTGTTCTGGCCATTGCAAAAAATCCGATTGTGGAAAATTACGATCTGAGCAGCATTAAAATGGTTATGAGT GGTGCCGCACCGCTGGTGAAAGAACTGGAAGATGCCTTCCGTGCACGCTCTGCCGAATGCCGTTCTGGGTCAGGGCTATGGCATGACCGAAGCCGGCCCGGT TCTGGCCATGTGCCTGGCATTCGCCAAAAGTCCGTTCCCGGTGAAACCGGGCAGCTGTGGTACCGTTGTGCGCAATGCAGAAGTTAAAATTGTGGATACCGA AACCGGTATGAGTCTGCCGTATAATCAGCCGGGTGAAATCTGTATTCGTGGTCCGCAGATTATGAAAGGCTATCTGAATAATCCGGAAGCCACCGCCAATAC CATTGATAAAGATGGCTTCCTGCATACCGGCGATGTTGCCTTCATTGATGAAGATGAAGAAATGTTCATCGTTGATCGTGTGAAAGAAATTATTAAGTTCAAG GGCTTCCAGGTTCCGCCGGCAGAACTGGAGGCACTGCTGCTGAGCAATGAAGAAATTCAGCATGCAGCAGTTGTGAGTCGTAAAGATGATGTGGCCGTGA AGTTCCGGTGGCCTTCGTTGTGCGCACCAGGCCGGTAGCACCATTAGCGAAGAAGAAGTTAAAGATTATGTGGCCAAACAGGTTGTGTTCTATAAAAAAATTCA CAACGTGTACTTCGTGGATAGTATTCCGGAAAGCCCGAGTGGTAAAATTCTGCGCAAAGACTTACGTAATAAAGTGTAA	MSPSIISESLMSEPLTYSIMKSLQSDVQELKPSLTESWAETKTESAPSDQVQEFIF RSKLPDIYISNHMPLTDYCLEKVTQWPDKVCLIDGNTGREYSYGEMELTSRRVA AGLAKIGVEQGGVIALLLPNCAEFVQVFLGAAKRGAIVTTANPFYTSTELAKQIIA SGATVVVTQSRYIEKLAGLNIQIIVVDQYVDGYLHVSALLEADEAECPEVDIHPD DVVCLPYSSGTTGLPKGVMLTHKSLVTSVAQQVDGEVPHFNINVEDTLMCVLP MFHIYSLNSILLCGLRAGATLVIMAKFELSKLLEFIQKYKVTMGPFVPPIMLAIAK NPIVENYDLSSIKMIMSGAAPLGKELEDAFRARLPNAILGQGYGMTEAGPVLA MSLAFAKTPFPVKPGSCGTVVRNAEVKIIDTETGMSLSYNQPGEICIRGPQIMK GYLNNPEATAYTIDEDGFLHTGDVAFIDEDEEMFIVDRVKEIIKYKGFQVPPAELE AVLLSHQQIQDAAVVSRKDEVAGEVPVAIVVRSPGCTITEDEVKDHVAKRVVFY KKIHDVYFADSIPKSAYGKILRKDITLKF
Pp4CL 4	ATGAGTCCGAGCATTATTAGTGAAAGCCTGATGAGCGAACCGCTGACCTATAGCATTATGAAAAGTCTGCAGAGCGATGTGCAGGAACTGAAACCGAGTCT GACCGAAAGCTGGGCAGAAACCAAAACCGAAAGCGCACCGAGTGATCAGGTTCAGGAGTTCATCTTCCGCAGCAAACTGCCGGATATCTATATTAGCAATCA CATGCCGCTGACCGATTATTGTCTGGAAAAAGTGACCCAGTGGCCGGATAAAGTGTGTCTGATTGATGGTAATACCGGCCGCGAATATAGTTATGGTGAAAT GGAACTGACCAGTCGTCGTGTTGCCGCAGGCCTGGCCAAAATTGGTGTGGAACAGGGTGGCGTTATTGCCCTGCTGCTGCCGAATTGTGCAGAGTTCGTTCA GGTGTTCCTGGGCGCCGCAAAACGTGGCGCCATTGTGACCACCGCAAATCCGTTCTATACCAGCACCGAACTGGCCAAACAGATTATTGCAAGCGGTGCCAC CGTTGTTGTGACCCAGAGTCGCTATATTGAAAAACTGGCCGGCCTGAATATTCAGATTATTGTTGTTGATCAGTACGTGGATGGCTATCTGCATGTGAGTGCA CTGCTGGAAGCCGATGAAGCAGAATGTCCGGAAGTTGATATTCATCCGGATGATGTGGTGTGTCTGCCGTATAGTAGTGGTACCACCGGTCTGCCGAAAGGT GTTATGCTGACCCATAAAAGTCTGGTGACCAGCGTTGCCCAGCAGGTGGATGGTGAAGTTCCGCACTTCAATATTAATGTTGAAGATACCCTGATGTGTGTGC TGCCGATGTTCCATATCTATAGCCTGAATAGCATTCTGCTGTGCGGTCTGCGTGCCGGTGCCACCCTGGTTATTATGGCAAAATTCGAACTGAGCAAACTGCT GGAGTTCATTCAGAAATATAAAGTGACCATGGGTCCGTTCGTTCCGCCGATTATGCTGGCAATTGCAAAAAATCCGATTGTGGAAAATTACGATCTGAGTAGC ATTAAAATGATCATGAGTGGTGCCGCCCCGCTGGGTAAAGAACTGGAAGATGCATTCCGCGCCCGTCTGCCGAATGCAATTCTGGGCCAGGGCTATGGTATG ACCGAAGCAGGTCCGGTGCTGGCCATGAGCCTGGCATTCGCCAAAACACCGTTCCCGGTTAAACCGGGCAGTTGCGGTACCGTTGTGCGTAATGCAGAAGTT AAAATTATGATACCGAGACCGGCATGAGCCTGAGCTATAATCAGCCGGGTGAAATCTGTATTCGCGGTCCGCAGATTATGAAAGGTTATCTGAATAATCCG GAAGCCACCGCCATACCATTGATGAAGATGGCTTCCTGCATACCGGTGATGTTGCATTCATTGATGAAGACGAAGAAATGTTCATTGTGGATCGCGTTAAAG AAATTATTAAGTATAAGGGCTTCCAGGTGCCGCCGGCAGAACTGGAAGCAGTGCTGCTGAGCCATCAGCAGATTCAGGATGCCGCAGTTGTTAGTCGCAAA GATGAAGTGGCAGGTGAAGTGCCGGTTGCCATTGTGGTTCGCAGCCCGGGCTGTACCATTACCGAAGATGAAGTTAAAGATCATGTTGCAAAACGCGTTGT GTTCTATAAAAAAATTCACGATGTGTACTTCGCAGATAGTATTCCGAAAAGCGCATACGGTAAAATTCTGCGTAAAGATATTACCCTGAAATTCTAA	MSELVKRTAAGLVTLGVKPKDVVVFLSPNSPEFPVVYLAIVYIGAVAALGNSLNT ENDIALQLVQTKAVFVITVPEQFSKIQKCNLPTILIGDDVLASQLKTTTSVSLLSEV CEGDYIEMAPPECHPDDTCSLLFSSGTTGLTKAIQLTHRNLMSSVTAYNTLEPGD STREDDVCVAIIPMFHVFGLGIIMLSTLQRGACVVTMTRYSFPSMLQYIEKYKIT VAIVVPPILVYLVKNQEMLAKYDLSSLRILMTGAAPLREDTMKSIQAIFPKCVTR QGYGMTECPLISYSVVANEHPLPVMTTGEVWVRGPQIMKGYLNNVEQTSATI DSDGWLHTGDLGYMDNNNYLFIIDRLKEMIKYRGHQVAPGDLEAVLLKNPRIL DACVVPCPDDDNCELPMAFVVKRECTDLTELEVMNYVAQLVAPYKKVRKVEFI DAIPKSPTGKILRKQLTLKLRETGCGDRSVGRIQTASNLGQPKSSKSLNLRWTSST KLNIVKSHQVERTATYL

8.2 Unit abbreviations

% (w/v)	mass concentration (1 % = 1 g/100 ml)
% (v/v)	volume concentration
°C	degree Celsius
cm	centimetre
Da	Dalton
g	gram
g	gravitational acceleration [g = 9.81 m/s^2]
h	hour
kat	katal [mol/s]
k_{cat}	turnover number [s^{-1}]
k_{cat}/K_m	catalytic efficiency [s^{-1} mol^{-1} l]
kg	kilogram
K_m	Michaelis-Menten constant [µM]
M	moles per litre [mol/l]
min	minute

mm	millimetre
mM	millimoles per litre [mmol/l]
nm	nanometre
OD_{600}	optical density at 600 nm
rpm	revolutions per minute
s	second
U	enzyme unit
V_{max}	maximum reaction velocity [mkat kg^{-1}]

8.3 List of further abbreviations

4CL	4-coumarate CoA-ligase
A	adenosine
Aa	*Anthoceros agrestis*
Aa4CL	4-coumarate CoA-ligase derived from *Anthoceros agrestis*
AaPAL	phenylalanine ammonia-lyase derived from *Anthoceros agrestis*
appr.	approximately
APS	ammonium persulfate
ATP	adenosine triphosphate
BCIP	5-bromo-4-chloro-3-indolyl phosphate
bp	base pairs
BSA	bovine serum albumin
C	cytosine
C4H	cinnamic acid 4-hydroxylase
CAD	cinnamyl alcohol dehydrogenase
Cb	*Chara braunii*
Cb4CL	4-coumarate CoA-ligase derived from *Chara braunii*
CbPAL	phenylalanine ammonia-lyase derived from *Chara braunii*
cDNA	complementary DNA
CoA	coenzyme A
COMT	caffeic acid *O*-methyltransferase
CCoAOMT	caffeoyl-CoA *O*-methyltransferase
CCR	cinnamoyl-CoA reductase
CSE	caffeoylshikimic acid esterase
dNTP	nucleoside triphosphates containing deoxyribose
L-DOPA	levodopa
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
F5H	ferulate 5-hydroxylase (F5H)
fus	fusion
G	guanine
gDNA	genomic DNA
HAL	histidine ammonia-lyase
HCl	hydrochloric acid

HCT	hydroxycinnamoyltransferase
L-His	L-histidine
His-Tag	protein tag with 6x His
HPLC	high performance liquid chromatography
HW	Hanes-Woolf
IPTG	isopropyl-β- L -1-thiogalactopyranoside
KOH	potassium hydroxide
LB	lysogeny broth
LCMS	liquid chromatography-mass spectrometry
MeOH	methanol
$MgCl_2$	magnesium chloride
$MgSO_4$	magnesium sulfate
MM	Michaelis-Menten
$MnCl_2$	manganese chloride
Mp	*Marchantia polymorpha*
Mp4CL	4-coumarate CoA-ligase derived from *Marchantia polymorpha*
MpPAL	phenylalanine ammonia-lyase derived from *Marchantia polymorpha*
m/z	mass-to-charge ratio
NaOH	sodium hydroxide
NBT	nitro-blue tetrazolium chloride
Pa	*Plagiochasma appendiculatum*
PAGE	polyacrylamide gel electrophoresis
PAL	phenylalanine ammonia-lyase
PaPAL	phenylalanine ammonia-lyase derived from *Plagiochasma appendiculatum*
PCR	polymerase chain reaction
pHPL	4-hydroxyphenyllactate
D-Phe	D-phenylalanine
L-Phe	L-phenylalanine
Pp	*Physcomitrium patens*
Pp4CL	4-coumarate CoA-ligase derived from *Physcomitrium patens*
PpPAL	phenylalanine ammonia-lyase derived from *Physcomitrium patens*
PTAL	phenylalanine/tyrosine ammonia-lyase
PVP	polyvinylpyrrolidone
RAS	rosmarinic acid synthase
RNA	ribonucleic acid
SDS	sodium dodecyl sulfate
SOC	super optimal broth with catabolite repression
T	thymine
TAL	tyrosine ammonia-lyase
TB	terrific broth
TE	Tris-EDTA
TEMED	tetramethylethylenediamine
Tris	tris(hydroxymethyl)aminomethane
L-Tyr	L-tyrosine
UV	ultraviolet
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

8.4 Amino acids

Amino acid	**3 letter code**	**1 letter code**
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