Investigations on the origin of the 3,4-dihydroxyphenyllactic acid moiety of rosmarinic acid in *Anthoceros agrestis*

Dissertation zur Erlangung des
Doktorgrades der Naturwissenschaften
(Dr. rer. nat.)

dem Fachbereich der Pharmazie
der Philipps-Universität Marburg

vorgelegt von
Tobias Busch
aus Bad Mergentheim

Marburg/Lahn 2022

Dem Fachbereich Pharmazie der Philipps-Universität Marburg als Dissertation

eingereicht am 31.10.2022

Erstgutachterin: Prof. Dr. Maike Petersen

Zweitgutachter: Prof. Dr. Lars Voll

Tag der mündlichen Prüfung: 14.12.2022

Hochschulkennziffer: 1180

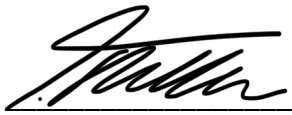
Eidesstattliche Erklärung

Ich versichere, dass ich meine Dissertation

Investigations on the origin of the 3,4-dihydroxyphenyllactic acid moiety of rosmarinic acid in *Anthoceros agrestis*

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 ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.



Tobias Busch

Marburg, den 31.10.2022

Acknowledgements

First and foremost, I would like to thank my supervisor Prof. Dr. Maike Petersen for her encouraging and inspiring nature and her individual and intensive support. You allowed me to plan my project freely and gave me the opportunity to investigate in depth what I considered interesting. I would like to thank you for the marvellous years, which were characterised by great atmosphere, pleasant conversations, and your always open ear.

Of course, I also owe a big thank you to my many great colleagues, which made work so much fun and introduced me into new methods: Soheil Pezeshki, Lennart Poppe, Julia Wohl, Victoria Werner, Jennifer Robinson, Sandra Ditzler, Lucien Ernst, Anne Jahn, Paul Bömeke, Christoph Schwarze, Christoph Kentrath, Janik Marks, Maximilian Ufland, Simona Riedel, Kim Röder, Thanh Son Ta, Elke Bauerbach, Olga Haag, Hanna Lütjens and Marie-Claire Badouin.

The same applies to all the other colleagues from the Institute of Pharmaceutical Biology and Biotechnology, including Prof. Dr. Li and all former and current members of his research group, as well as Dr. Dieter Kreusch and Ronald Bloch for the organisation and realisation of the student lab and Sabine Burgers and Sonja Hiemenz for the smooth management of all organisational matters.

Last, I owe my family and friends and especially Moni a small apology for having to bear my whims while I was completing this work. Without your support this work would not have been possible.

Publications and Presentations

Publications	Busch, T., Petersen M. At the crossroads of primary and secondary metabolism: tyrosine aminotransferase in the hornwort *Anthoceros agrestis*. *Planta* 253, 98 (2021). https://doi.org/10.1007/s00425-021-03623-2
	Pezeshki, S., Warmbier, I., Busch, T., Bauerbach, E., Szövenyi, P., Petersen, M. The first step into phenolic metabolism in the hornwort *Anthoceros agrestis*: molecular and biochemical characterization of two phenylalanine ammonia-lyase isoforms. *Planta* 256, 33 (2022)
Scientific lecture	Two enzymes involved in rosmarinic acid biosynthesis: tyrosine aminotransferase and hydroxyphenylpyruvate reductase in *Anthoceros agrestis*. Section Natural Products of the German Botanical Society (DBG), Warberg, 2019
Poster presentations	Hydroxyphenylpyruvate reductase as an enzyme of rosmarinic acid biosynthesis in *Anthoceros agrestis*. Tagung der Deutschen Botanischen Gesellschaft, Kiel, 2017
	First insights into rosmarinic acid biosynthesis in the hornwort *Anthoceros agrestis* – tyrosine aminotransferase and hydroxyphenylpyruvate reductase. Tagung der Deutschen Botanischen Gesellschaft, Rostock, 2019

Table of contents

I Introduction

1 Early plant records of the conquest of land

The giant sequoia, spined cacti, colourful orchids, the Venus flytrap and dense fern forests - it is hardly imaginable, but the whole plethora of terrestrial plants evolved from a common ancestor, emerged from freshwater streptophyte green algal pioneers. While the classification of land plants as a monophylum seems to be resolved, the exact position of bryophytes as well as their phylogenetic relationship to the vascular plants, was a matter of controversy (Gensel 2008; de Vries and Archibald 2018). The bryophytes comprise three divisions, namely mosses (Bryophyta), liverworts (Marchantiophyta) and hornworts (Anthocerotophyta). Since these plants are inconspicuous and related species are difficult to distinguish from each other, which leads to insufficient classification and redundancy, there is a widely differing information about the actual number of representatives, with conservative estimates of 12,000 to about 20,000 (Hallingbäck 2000; Shaw et al. 2011; Tomescu et al. 2018) and up to 23,000 species (European Redlist), enough to make it the second largest group after the flowering plants. Of these, the mosses form the largest group, followed by the liverworts, and at a great distance by the hornworts, which have a low abundance of about 100-250 species (Hallingbäck 2000; Shaw and Renzaglia 2004; Villarreal et al. 2010; Söderström et al. 2016). However, which of them can be considered most closely related to the vascular plants is not easily resolved and is fiercely contested. Similarly, the algal predecessor is highly debated (Timme et al. 2012; Bowman 2013), especially since the sister taxon to the land plants was narrowed down to one of the three higher charophyte lineages. The Charophyceae and Coleochaetophyceae share several characteristics with Embryophyta, like oogamous sexual reproduction, gametophytic plasmodesmata and a more complex morphology, i.e. apical growth with branching. These observations led, in the context of phylogenetic analyses, to strong support especially for the Charophyceae as sister taxon to the land plants (Graham 1996; Kenrick and Crane 1997b; Karol et al. 2001). Zygnematophyceae, in turn, although structurally simple, are supported by recent extensive phylogenetic analyses (Turmel et al. 2006; Timme et al. 2012; Ruhfel et al. 2014; Wickett et al. 2014; Cheng et al. 2019; One Thousand Plant Transcriptomes Initiative 2019), possibly forming a clade with Coleochaetophyceae (Wodniok et al. 2011). The lack of

advanced morphological traits rather implies an advantageous reductive evolution adapting to semi-terrestrial environment. This is supported by the findings of rudimentary phragmoplasts, consequence of a possible reduction, in some Zygnematophyceae. All three of the aforementioned potential algal ancestors exhibit this characteristic and thus, together with terrestrial plants, form the genus Phragmoplastophyta (Lecointre and Le Guyader 2006; Timme et al. 2012; Wickett et al. 2014; Delwiche and Cooper 2015). The idea that Zygnematophyceae evolved from a structurally more complex ancestor and that reduced filamentous or unicellular representatives (e.g. Desmidiales), adapted to life in drier environments through innovations, i.e. mucilage, lack of flagella, seems plausible (Delwiche and Cooper 2015).

The following terrestrial conquest through bryophyte-like species can be dated to the mid-Ordovician period. Findings of macrofossils like *Aglaophyton* and *Horneophyton* (Devonian Rhynie Chert), both showing characteristics of bryophytes and vascular plants, and *Cooksonia* spec., some of which can be seen as first true vascular plants, provide evidence of the late Silurian-Devonian radiation that caused species diversity to increase rapidly (Kenrick and Crane 1997b; Bateman et al. 1998; Gensel 2008; Kenrick et al. 2012).

Unfortunately, the pivotal phase in the evolution from algae to bryophytes as well as the timely emergence of liverworts, hornworts and mosses is not fully covered by the fossil record, especially since structured features in form of macrofossils which can provide information about their organisation and distinct characteristics have not been discovered. Whether this is due to the lack of resistant structures and compounds that delay decay and thus poor fossilisation in early plant pioneers is debatable (Gensel 2008). Although bryophytes do not possess true lignified plant parts, phenolic, lignin-like compounds or respective pathways are in fact known, at least for recent bryophytes and even algal representatives (Delwiche et al. 1989; Kroken et al. 1996; Ligrone et al. 2008; Espiñeira et al. 2011; Weng and Chapple 2010; Renault et al. 2019). Accordingly, Tomescu et al. (2018) showed that the actual record of bryophyte fossils is much better than stated and fossilisation potential indeed is on par with nonlignified tissue of tracheophytes. Yet, assured bryophyte macrofossils only appear in younger strata (mid-Devonian) and postdate findings of verified early vascular plants (i.e. *Cooksonia* spec.; (Edwards et al. 1992). Other explanations considered include external factors like the absence of fossiliferous rocks or the spatial distribution of early plants without access

to depositional environments in the Ordovician or early Silurian (Wellman and Gray 2000; Gensel 2008; Steemans et al. 2009) or direct influences such as the small size of such plants or shed parts available for fossilisation as well as human bias (Kenrick et al. 2012; Tomescu et al. 2018). The actual composition of the early terrestrial plant flora is therefore best represented by microfossils which date back to the mid-Ordovician, represented through putative liverwort cryptospores (Wellman et al. 2003; Steemans et al. 2009; Rubinstein et al. 2010; Tomescu et al. 2018; Edwards et al. 1995). The fact that the spore record goes back much further can be attributed to the high quantity produced compared to plant individuals, effective dispersal into a wide variety of environments through wind and water and a robust sporopollenin accretion that prevents breakdown (Kenrick et al. 2012). The liverwort affinities of the oldest specimens found, led to the assumption that these might be the most basal form of land plants (Gensel 2008; Tomescu et al. 2018). There are morphological properties that speak for this assumption, like the absence of stomata in liverworts, which occur in sporophytes of hornworts as well as mosses. However, it has been found that these structures are functionally not homologous to the stomata of tracheophytes, i.e. are not suitable to regulate gas exchange, instead contributing to the desiccation of the sporophyte (Lucas and Renzaglia 2002; Pressel et al. 2018; Sussmilch et al. 2019; Pressel et al. 2014). Obviously, stomata were not essential, as evidenced by the occasional loss in both, hornwort and moss representatives. That liverworts have also merely lost this capacity is reasonable and was further substantiated by Harris et al. (2020) recently (Villarreal and Renzaglia 2015; Renzaglia et al. 2020). Individual features arguing for the basal placement of one bryophyte or another are not sufficient for these reasons. This is also evidenced by pyrenoids, bodies mainly consisting of RuBisCO, which are present in algae and hornworts but not in other bryophyte clades. Pyrenoids were likewise affected by losses (Smith and Griffiths 1996; Villarreal and Renner 2012). Therefore, early ambitious cladistic approaches focused on a variety of morphological characters (Mishler and Churchill 1984; Garbary and Renzaglia 1998), on ultrastructural features such as male gametogenesis (Garbary et al. 1993) or combined morphological classification with ribosomal RNA gene sequences (Mishler et al. 1994). The ones mentioned represent only an exemplary sample of a large number of approaches, which include molecular studies on the presence or absence of mitochondrial introns (Qiu et al. 1998) or genetic analyses with respect to single or few nuclear, mitochondrial or chloroplast genes (e.g. Hedderson et al. 1996; Lewis et al. 1997; Duff and Nickrent 1999) emphasising

different phylogenies, mainly focusing on bryophyte paraphyly. Nickrent et al. (2000) thereby provide an useful sketch of the establishment of two competing models, that of successive bryophyte lineages with basal liverworts and that of basal hornworts with monophyletic mosses and liverworts.

Advances in computational phylogenomics and transcriptomics and availability of rapidly growing databases led to multigene analyses (e.g. Qiu et al. 2006; Karol et al. 2010; Chang and Graham 2011; Ruhfel et al. 2014), some of which eventually advocated the monophyly of the bryophytes and endorsed the moss+liverwort clade (Wickett et al. 2014; Nishiyama et al. 2004; Goremykin and Hellwig 2005; Cox et al. 2014). A fourth model comprising the latter together with the hornworts as sister to the vascular plants completes the richness of models. The hypothesised tree topologies of bryophyte evolution and relationship to vascular plants are schematically outlined by Wohl (2020) including supporting literature. Renzaglia et al. (2000) published a comprehensive study based on a multi dataset approach comprising spermatogenesis, morphological as well as nuclear and mitochondrial rRNA data (c.f. Duff and Nickrent 1999; Garbary and Renzaglia 1998; Garbary et al. 1993) which supports the moss+liverwort clade which was resolved as setaphyta (c.f. Renzaglia et al. 2018) by Puttick et al. (2018). Accordingly, the 2018 study supports three main topologies differing only in the position of hornworts, therefore rejecting the idea of successive sister lineages, which was strongly supported to date by the comprehensive study of Qiu et al. (2006), and advocating bryophyte monophyly. These current opinion was recently reaffirmed by Sousa et al. (2019), Harris et al. (2020), Li et al. (2020), Zhang et al. (2020), Su et al. (2021) and the One Thousand Plant Transcriptomes Initiative 2019. Some authors (Cox 2018; Sousa et al. 2020) have already discussed the incongruence of phylogeny inferred of molecular data depending on the particular source (nuclear, mitochondrial, plastid, nucleotides, amino acids). Nevertheless, the consensus on the monophyly of the bryophytes and integration of liverworts and mosses as setaphyte clade seems to be consolidating. No matter how the phylogeny issue is ultimately settled, it has far-reaching implications for the notion of the primal life cycle (Kenrick 2018). A complete bryophyte paraphyly would imply a bryophyte-like alternation of generations in the last common ancestor (antithetic or intercalation theory, Bower 1890). The original alternation of the haplontic alga, with only a haploid multicellular generation, is thereby extended by mitoses preceding the meiotic division of the zygote. This diploid intermediate, the matrotrophic sporophyte, will develop and eventually become independent. This theory

fits well with the assumption of charophyte green algae as direct ancestors of land plants and has been widely accepted. In contrast, the monophyly of the bryophytes could also reintroduce the ideas of a pioneer with isomorphic (homologous) life cycle, i.e. independent gametophyte and sporophyte, and following opposite developments towards the bryophytes, developing the dominant gametophyte, or towards the tracheophyte lineage heading the opposite direction. This pioneer lineage would then be extinct, while the extant haplontic charophyte algae are secondarily aquatic according to Stebbins and Hill (1980), see Harholt et al. (2016) and also Shaw et al. (2011). Fossilised protracheophytes (*Horneophyton*, *Aglaophyton*) in particular show similar to even isomorphic, free-living generations, making sporophytes recognisable only by the sporangia and larger form (Kenrick 1994; Blackwell 2003; Haig 2008; Niklas and Kutschera 2010; Ligrone et al. 2012; Kenrick 2018). Whether the isomorphy represents only an intermediate stage towards higher plants or the crossroads between bryophytes and vascular plants is crucial. This explains the great interest in the basal phylogeny of terrestrial plants and will remain open for novelties for quite some time.

2 Classification and habitus of *Anthoceros agrestis*

The hornworts (Anthocerotophyta) are divided in the two classes Leiosporocerotopsida and Anthocerotopsida, with the field hornwort, *Anthoceros agrestis*, family Anthocerotaceae, belonging to the latter. It can be found in Central and Southern Europe, North Africa and North America, on sufficiently moist sites on arable fields, wet rocks and earth embankments (Frahm 2011). A clear differentiation of the species *Anthoceros agrestis*, *A. punctatus* and *A. crispulus* is not apparent in literature. Often these are used synonymously. Accordingly, the callus culture first introduced by Binding and Mordhorst (1991), the line also used in this work, is originally classified as *A. crispulus*, now as *A. agrestis*. The plant body is the flat haploid, dominant gametophyte, as in other bryophytes, but does not form leaves, unlike the mosses and some liverworts. The plant grows in form of a rosette-shaped, curly and slit thallus (Frahm 2011) and lacks real specialised tissue other than unicellular, filamentous, mostly unbranched rhizoids. The gametophytic thallus has embedded antheridia (male) and archegonia (female) and is therefore monoicous. The biflagellate motile male gametes (sperms) depend like other bryophytes on moist environment to swim to the female gamete (egg). The diploid zygote resulting from the fusion develops into the sporophyte (Frangedakis et al. 2021). Similar to mosses and as well limited to the sporophyte, stomata are present, but probably not

functionally homologous to those of vascular plants (Pressel et al. (2014), see also chapter I.1). The eponymous unbranched horn-like structure develops from a basal meristem and always remains anchored to the parent gametophyte and is nourished by it (matrotrophy) (Frangedakis et al. 2021). The high interest in the species *Anthoceros agrestis*, or hornworts in general, results from quite unique characteristics. Among these are symbiotic, endophytic *Nostoc* colonies housed in mucilaginous cavities. Although *Nostoc* also occurs in other species (Dodds et al. 1995), its presence in the hornworts is characteristic due to the formation of a pore defined by cells resembling the stomatal guard cells of the sporophyte. The pore allows the symbiont to enter and is subsequently closed (Renzaglia et al. 2008; Frangedakis et al. 2021). Further, unlike the other land plants (with exceptions, Liu et al. 2020), there is only one large chloroplast per cell with distinct channel thylakoids and a central pyrenoid, the site of RuBisCO accumulation and important as a CO_2-concentrating mechanism, otherwise only known in algae (Vaughn et al. 1992; Smith and Griffiths 1996; Renzaglia et al. 2008). The sporophyte also shows distinctive features. The first division of the zygote takes place longitudinally, in contrast to the transverse division in liverworts and mosses. The developing sporophyte is characterised by the continuous production of spores from the basal persistent meristem, extending the sporangium from the base, which is unique among land plants. Eventually, seta are missing in comparison to other bryophyte clades (Renzaglia et al. 2008; Frangedakis et al. 2021). Other characteristics, such as fungal associations, i.e. mycorrhizal fungi (Glomeromycota, Mucoromycotina), are not exclusive but emphasise the importance of hornworts in understanding basal plants (Read et al. 2000; Desirò et al. 2013). Consequently, in 2015, Szövényi et al. gave the go-ahead for establishing *Anthoceros agrestis* as a model species with the publication of the genome and protocols for cell culture, induction of reproduction, and nucleic acid extraction.

3 Specialised compounds in *Anthoceros agrestis* and related species

Plants have played an important role in the pharmaceutical history of mankind. While medicines from animals or minerals have almost completely disappeared today, phytomedicines are still used or individual compounds are isolated and used as precursors for semi-synthetic drugs or as prototype of new drug groups. Secondary metabolites, compounds not involved in the primary metabolism of the plant, are of greatest importance. With the beginning of the isolation of active ingredients like morphine (Friedrich Sertürner, 1805) and

numerous other alkaloids in subsequent years, phytotherapy finally broke away from the doctrine of signatures, on the basis of which liverworts, for example, due to their visual similarity to the human liver, were used against liver diseases (Salim et al. 2008). Apart from this application, preparations from bryophytes are rather known from Chinese or North American Indian folk medicine. Nevertheless, a surprisingly diverse pool of new compounds, some of with cytostatic or antibiotic activity is known, with liverworts being the most extensively studied, while little work on hornworts has been done (Zinsmeister et al. 1991). This is consistent with reviews by Asakawa and colleagues, who have collected information on more than a thousand bryophytes and 3000 compounds (Martínez-Abaigar and Núñez-Olivera 2021). The largest group is formed by terpenoids, which are more widely distributed among liverworts due to their storage in specific oil bodies than in mosses and hornworts. The latter comprising only a few records, i.e. some mono- (e.g. limonene) and sesquiterpenoids from *Anthoceros caucasicus*, *A. punctatus* and *A. agrestis* (Ludwiczuk and Asakawa 2019 and literature cited therein). The second largest group is formed by phenolic compounds, followed by a few nitrogen- or sulfur-containing ones. In the following, the focus will be on the second and third group of metabolites isolated from hornworts, i.e. *Anthoceros agrestis* as well as the attached synonyms (*A. punctatus*, *A. crispulus*) and other relevant species. One striking compound, the alkaloid anthocerodiazonin (Fig. 1, a), containing a nine-membered ring system, was reported by Trennheuser et al. (1994). In addition, various glutamic acid amides (Fig. 1, b-d) are shown which comprise either a hydroxybenzoyl, derivatives thereof, or an isoferuloyl or a 4-coumaroyl moiety, respectively (Trennheuser et al. 1994). The latter two, since containing the characteristic C6-C3 structure, belong to the phenylpropanoids. This type of compounds plays an important role in terrestrialisation. To colonise land habitats, the plant pioneers had to cope with novel environmental factors, such as stress from UV-B radiation, drought, new pathogens, and herbivores. Incorporation of phenolic compounds into lipid polymers (cuticle) and spore walls (sporopollenin) or formation of lipid-phenolic co-polymers (suberin) are mainly land plant innovations and, like lignin, form mechanical-chemical barriers. Lignin, a biopolymer comprising of monolignol moieties, ubiquitous in vascular plants, is also essential for sessile life. To overcome the constrained habitat and enable vertical growth, formation of rigid structures withstanding gravity and enabling efficient transport systems, i.e. vascular tissue, are necessary (Weng and Chapple 2010; Niklas et al. 2017; Rensing 2018). Although there is no true lignin in algae and bryophytes, there is increasing evidence that

initial elements as well as the enzymatic machinery of phenylpropanoid biosynthesis were adopted early. A CYP98-dependent formation of caffeoylthreonic acid, component of the lipid-phenolic cuticle matrix in *Physcomitrium patens*, could be extant evidence of the ancestor of suberin, cutin and lignin polymers (Renault et al. 2017). In this regard, the moss contains orthologs of all eight key lignin biosynthetic enzymes required for the biosynthesis of 4-coumaryl and coniferyl alcohol (Weng and Chapple 2010). *Marchantia polymorpha* is able to activate a phenylpropanoid pathway in response to oomycete colonisation (Carella et al. 2019). Originally seen as a unique feature of terrestrialisation, de Vries et al. (2021) state that streptophyte algae already possessed candidates for the genetic toolkit of the biosynthesis of phenylpropanoids. This is supported by several publications. Delwiche et al. (1989) reported lignin-like compounds and sporopollenin in Coleochaetophyceae, the former confirmed by Sørensen et al. (2011), whereas flavonoids have been detected for the zygnematophyceaen alga *Penium margaritaceum* (Jiao et al. 2020). Findings in plant-specific metabolism, including phenylpropanoids in basal organisms, and implications on terrestrialisation, were recently comprehensively reviewed by Rieseberg et al. (2022).The structure of phenylpropanoids is based on (hydroxy)cinnamic acid and can be found in many other metabolites, including flavonoids and associated anthocyanins, mediating UV-B tolerance, general abiotic stresses and environmental interactions among other functions (e.g. Stapleton 1992; Taylor and Grotewold 2005; Dong and Lin 2021). The occurrence of flavonoids in algae is debatable and the definite confirmation of specific genes is still pending. In contrast, flavonoids have been confirmed in liverworts and mosses. An exception are once again the hornworts, which, although thought to be specific to land plants, lack flavonoids (Davies et al. 2020; Davies et al. 2021). Instead, other promising phenylpropanoids were found, when Takeda et al. (1990) reported the first isolation of the lignan anthocerotonic acid (Fig. 1, e) from *A. agrestis* along with megacerotonic acid (Fig. 1, f) from *Megaceros flagellaris*, which were the first examples of lignans from nonvascular plants, whereas descriptions for liverworts are available in the meantime (Lewis and Davin 1999). A short time later, hydroxymegacerotonic acid was also reported in *Anthoceros agrestis* (Trennheuser 1992). Lignans are diverse polyphenolic compounds derived from the oxidative coupling of two (hydroxy)cinnamoyl moieties (C6-C3) commonly linked 8-8', although the structural motifs are not limited to this. Different patterns are sometimes called neolignans, although the definition has varied over time (Lewis and Davin 1999). Lignans exhibit a variety of biological activites, e.g. cytotoxic and antiviral (e.g.

podophyllotoxin, Petersen and Alfermann 2001 and literature cited therein), antifungal and antimicrobial (MacRae and Towers 1984; Zálešák et al. 2019), or have dietary relevance as the phyto-estrogenic mammalian lignans enterodiol and enterolactone metabolised from secoisolariciresinol diglucoside or matairesinol through gastrointestinal bacteria (Landete 2012). The biological activity of the hornwort lignans is almost unexplored. However, Adamu et al. (2021) suggest a potential inhibiton of *Trypanosoma brucei* alternative oxidase for megacerotonic acid. In addition to lignans, rosmarinic acid (Fig. 1, g), an ester of caffeic acid and 3,4-dihydroxyphenyllactic acid and common product in higher plants as well as the standalone unit caffeic acid were detected in *A. agrestis* (Takeda et al. 1990). Interestingly, the latter has already been described in the form of its methyl ester by Méndez and Sanz-Cabanilles (1979) for *A. punctatus*. Rosmarinic acid and its 3'-O-β-D-glucoside as well as the putative precursor molecule caffeoyl-4'-hydroxyphenyllactic acid (isorinic acid) was confirmed in *A. agrestis* by the work of Vogelsang et al. (2006). In addition to the (hydroxy)cinnamic acid structure, derived from phenylpropanoid metabolism, some compounds, i.e. anthocerotonic acid, megacerotonic acid and rosmarinic acid, share the hydroxyphenyllactic acid moiety. It can be assumed that the first two are derivatives of rosmarinic acid, formed either enzymatically controlled or non-enzymatic. Since rosmarinic acid is widely distributed among vascular plants and can be seen as a blueprint of phenolic structures from hornworts, the elucidation of the underlying biosynthesis is of great interest to understand the implications of phenylpropanoids or complex esterified phenolics for terrestrialisation.

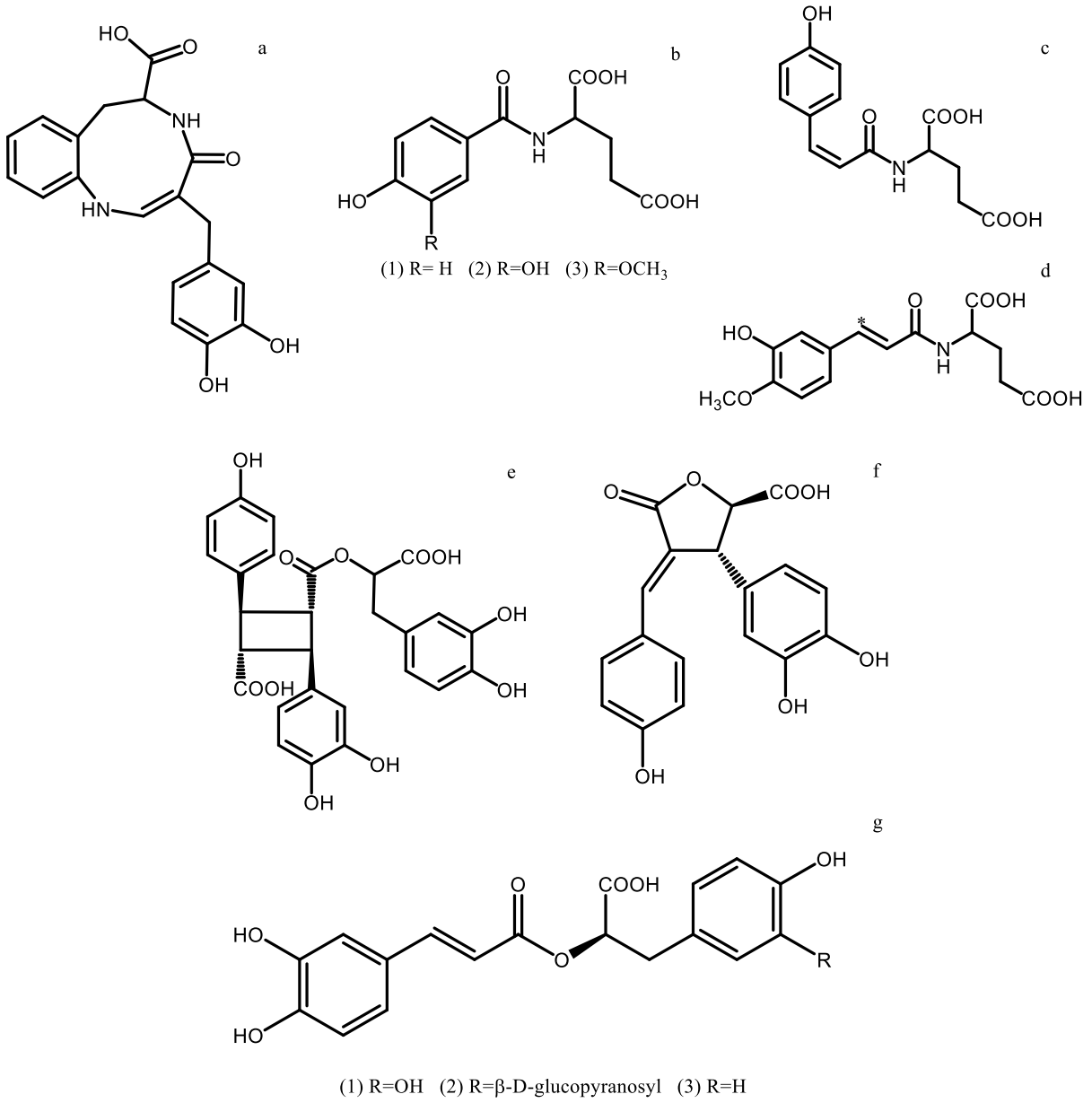


Figure 1 Exemplary secondary metabolites found in *Anthoceros agrestis* (a-e, g) and *Megaceros flagellaris* (f). a Anthozerodiazonin, an alkaloid featuring a characteristic nine-membered ring **b-d** Glutamic acid amides with either hydroxybenzoyl-derivative moiety (b), 4-coumaroyl moiety (c) or isoferuloyl moiety (d) and its *cis*-isomer (indicated through *) **e-f** The lignans anthocerotonic acid and megacerotonic acid, while 3'-hydroxymegacerotonic acid was described for *A. agrestis* **g** rosmarinic acid (R1), its 3'-*O*-glucoside (R2) and isorinic acid (R3)

4 Rosmarinic acid

Rosmarinic acid, as the name implies, was first isolated from rosemary (*Rosmarinus officinalis*, Lamiaceae) and resolved as an ester of caffeic and 3,4-dihydroxyphenyllactic acid by Scarpati and Oriente (1958). The compound together with other phenolic compounds are also called 'Labiatengerbstoffe' due to their tannin properties (see Litvinenko et al. 1975), referring to an

old name for the plant family (Latin: labia = lip). In fact, rosmarinic acid is widely distributed in Lamiaceae, but limited to the subfamily Nepetoideae, as well as in Boraginaceae, where all species studied were found to contain it. Litvinenko et al. (1975) suggested that the distribution of rosmarinic acid and other hydroxycinnamic acid derivatives has some chemotaxonomic significance (see also Harborne 1966). This is not deniable, since he strictly focused on Lamiaceae, however, rosmarinic acid alone proved to be unsuitable for interfamiliary taxonomics in the following years (Petersen and Simmonds 2003). Numerous other families have been identified, mainly from the core eudicots. But even if present, not even all species of a genus need to contain rosmarinic acid, and not all genera of a family or families in an order (Petersen 2013). A special role, especially with regard to the elucidation of rosmarinic acid biosynthesis, is assigned to *Coleus blumei* (syn. *Solenostemon scutellarioides*, *Plectranthus scutellarioides*, Lamiaceae) or painted nettle, see chapter I.5. Away from this, the Chlorantaceae (Zhu et al. 2008; Petersen et al. 2009) from the basal dicots, several families of monocots, e.g. Araceae (Aquino et al. 2001), Zosteraceae (Ravn et al. 1994), and Poaceae (Sterbová et al. 2006), as well as ferns, e.g. Blechnaceae (Harborne 1966; Bohm 1968), Dennstaedtiaceae (Tsumbu et al. 2012) contain rosmarinic acid, whereas it has up to now not been found in gymnosperms (Petersen 2013). The most basal occurrence is in Anthocerotaceae, i.e. *Anthoceros punctatus* (Takeda et al. 1990) and *Anthoceros agrestis* (Vogelsang et al. 2006), with up to 9% per dry weight in suspension cells of the latter (Petersen 2013). The erratic distribution raises the question whether this capacity evolved once early or several times independently in plant evolution (Petersen and Simmonds 2003). Investigations in charophyte algae, i.e. *Chara* spec., did not reveal rosmarinic acid (Petersen 2013). For more information on its occurrence, as well as relevant sources, see Petersen et al. (2009) and Petersen (2013). In addition, there are numerous species that contain corresponding derivatives, putatively derived from rosmarinic acid, among others e.g. lithospermic acid from *Lycopus europaeus* or *Lithospermum ruderale* or lithospermic acid B (syn. salvianolic acid B) from *Salvia milthiorrhiza* (Danshen) (Bulgakov et al. 2012; Petersen 2013), well-known from Traditional Chinese Medicine. For a full review of associated compounds see Pezeshki and Petersen (2018). Since the middle of the 20th century, there was great interest in explaining the nature of 'Labiatengerbstoffe' and defining the active compounds. Numerous studies were devoted to the description of the tannin type, e.g. as a hydrolysable tannin, but based on caffeic acid instead of gallic acid. Antidiarrheal, anti-inflammatory and antiviral effects of

the components were assumed and related compounds, like chlorogenic acid, ester of caffeic acid and quinic acid, were also described (Litvinenko et al. 1975). Meanwhile, numerous additional activities have been demonstrated for rosmarinic acid, e.g. anti-inflammatory, antimicrobial, antiviral, antihyperglycemic, neuroprotective, antitumor and general antioxidative effects (Amoah et al. 2016; Hitl et al. 2021). In addition, there is a nutritional potential as antioxidant (Khojasteh et al. 2020). From an ecological point of view, results indicate that rosmarinic acid can repel herbivores or prevent fungal infections as Simmonds et al. (2019) proved the activity as feeding deterrent against the tobacco hornworm (*Manduca secta*), whereas Szabo et al. (1999) demonstrated that a preparation of the oomycete *Pythium aphanidermatum* can significantly increase the formation of rosmarinic acid and the specific activity of the biosynthetic enzymes in suspension cultures of *Coleus blumei*. A somewhat similiar effect can also be achieved by the addition of the elicitor methyl jasmonate in *Coleus blumei* (Szabo et al. 1999) or *Salvia miltiorrhiza* (Xiao et al. 2009). Potential fields of application for rosmarinic acid are thus given, both medical, nutritional and agronomic, making the biotechnological production a vibrant field of research (see Bulgakov et al. 2012; Khojasteh et al. 2020).

5 Rosmarinic acid biosynthesis in higher plants

Formally, rosmarinic acid is an ester of caffeic acid and 3,4-dihydroxyphenyllactic acid (Fig. 2). Ellis and Towers (1970) performed an early ^{14}C-labelling study in *Mentha* spec. and resolved unequivocally that the phenylpropanoid moiety is derived from L-phenylalanine and the 3,4-dihydroxyphenyllactic moiety from L-tyrosine (see also Razzaque and Ellis 1977). While cinnamic acid and 4-coumaric acid were already correctly suggested as central intermediates of the caffeic acid part, the initial assumption was that tyrosine would be converted to DOPA. This was ultimately challenged by Ellis et al. (1979), who observed that there is no tyrosine 3-hydroxylase activity in *Coleus blumei* and that inhibition of PAL led to accumulation of labelled 4-hydroxyphenyllactic acid. This was confirmed by De-Eknamkul and Ellis (1987a), who proposed the transamination of L-tyrosine through tyrosine aminotransferase (TAT) as initial step. The widely accepted biosynthetic pathway was proposed by Petersen et al. (1993) for *Coleus blumei* and takes place as follows. TAT catalyses the transamination of L-tyrosine, leading to the formation of 4-hydroxyphenylpyruvate (pHPP). This is necessary because pHPP is not a primary intermediate in the shikimate pathway of plants. While bacteria and fungi can

produce the aromatic amino acids L-phenylalanine and L-tyrosine via transamination of pHPP and phenylpyruvate, which is originating from the precursor prephenate derived from chorismate from the shikimate pathway, plants, with exception of some legumes, which possess a cytosolic route to pHPP synthesis (Schenck et al. 2015), favour a divergent pathway. Plastid prephenate aminotransferase catalyses the formation of L-arogenate from prephenate, which is then converted by arogenate dehydrogenase or arogenate dehydratase to the corresponding aromatic amino acids (Maeda and Dudareva 2012; Parthasarathy et al. 2018). Said aromatic aminotransferases, e.g. TAT, are primarily located in the cytosol and involved in the degradation of amino acids to 2-oxoacids (for exceptions: Yoo et al. 2013; Wang et al. 2016) which leads back to the resulting pHPP. This is subsequently reduced to 4-hydroxyphenyllactate, catalysed by hydroxyphenylpyruvate reductase (HPPR; Petersen and Alfermann 1988; Kim et al. 2004). The involvement is supported by RNAi studies (Hücherig and Petersen 2013). Alternatively, pHPP can be transformed to homogentisate through hydroxyphenylpyruvate dioxygenase (HPPD), which is then directed to plastoquinone and tocopherol synthesis, respectively (Whistance and Threlfall 1970; Garcia et al. 1997). The second part of the reaction follows the general phenylpropanoid pathway starting with the deamination of L-phenylalanine by phenylalanine ammonia-lyase (PAL), as indicated by Razzaque and Ellis (1977), resulting in *t*-cinnamic acid and following selective *para*-hydroxylation by cinnamic acid 4-hydroxylase (C4H; Karwatzki et al. 1989) to 4-coumaric acid. The final step before linkage, is the activation as CoA-ester through hydroxycinnamate:coenzyme A ligase (4CL; Vetter 1988; Karwatzki et al. 1989). In addition to rosmarinic acid biosynthesis, 4-coumaric acid and its CoA-ester also serve as precursors for other important secondary metabolites, e.g. lignin, lignans, flavonoids, coumarins and others (Petersen 2013). Eventually, the formation of the rosmarinic acid precursor 4-coumaroyl-4'-hydroxyphenyllactic is catalysed by rosmarinic acid synthase (RAS; Petersen and Alfermann 1988; Petersen 1991), with involvement supported by RNAi data (Hücherig and Petersen 2013). The post-formational introduction of aromatic hydroxy groups through cytochrome P450 enzymes was suggested by Petersen et al. (1993) and resolved later (Petersen 1997). An alternative hypothesis (Di et al. 2013), based on ^{13}C-phenylalanine tracing experiments in *Salvia miltiorrhiza*, assumes the coupling of 4-coumaroyl-CoA and 3,4-dihydroxyphenyllactic acid, since only the 4-coumaroyl-3,4-dihydroxyphenyllactic acid intermediate was detectable in addition to the labelled rosmarinic acid. Initially, however, this only provides evidence for

the assumption that the 4-coumaroyl moiety is derived from L-phenylalanine, whereas the 3,4-dihydroxyphenyllactic acid moiety is not. Free 3,4-dihydroxyphenyllactic acid could not be detected due to the experimental design. Zhou et al. (2018) observed the accumulation of 3,4-dihydroxyphenyllactic acid and lower rosmarinic acid formation in CRISPR/Cas9-mediated mutants. However, complete knockout was not achieved and complementary genes led to substantial rosmarinic acid accumulation even in the homozygous mutant. In contrast, other authors involved in the research of biosynthesis in *Salvia miltiorrhiza* do not contest the current biosynthetic pathway. To date, proof of the involvement of all enzymes and selective characterisation, with exception of RAS, in *S. miltiorrhiza* have been achieved (Zhao et al. 2006; Huang et al. 2008a; 2008b; Song and Wang 2011; 2012; 2015). The characterisation of the *Salvia miltiorrhiza* RAS or tracing experiments with L-tyrosine could provide conclusive information. In general, it can be observed that existing research on rosmarinic acid often tend to focus on the phenylpropanoid moiety (e.g. Tuan et al. 2012; Kim et al. 2014a).

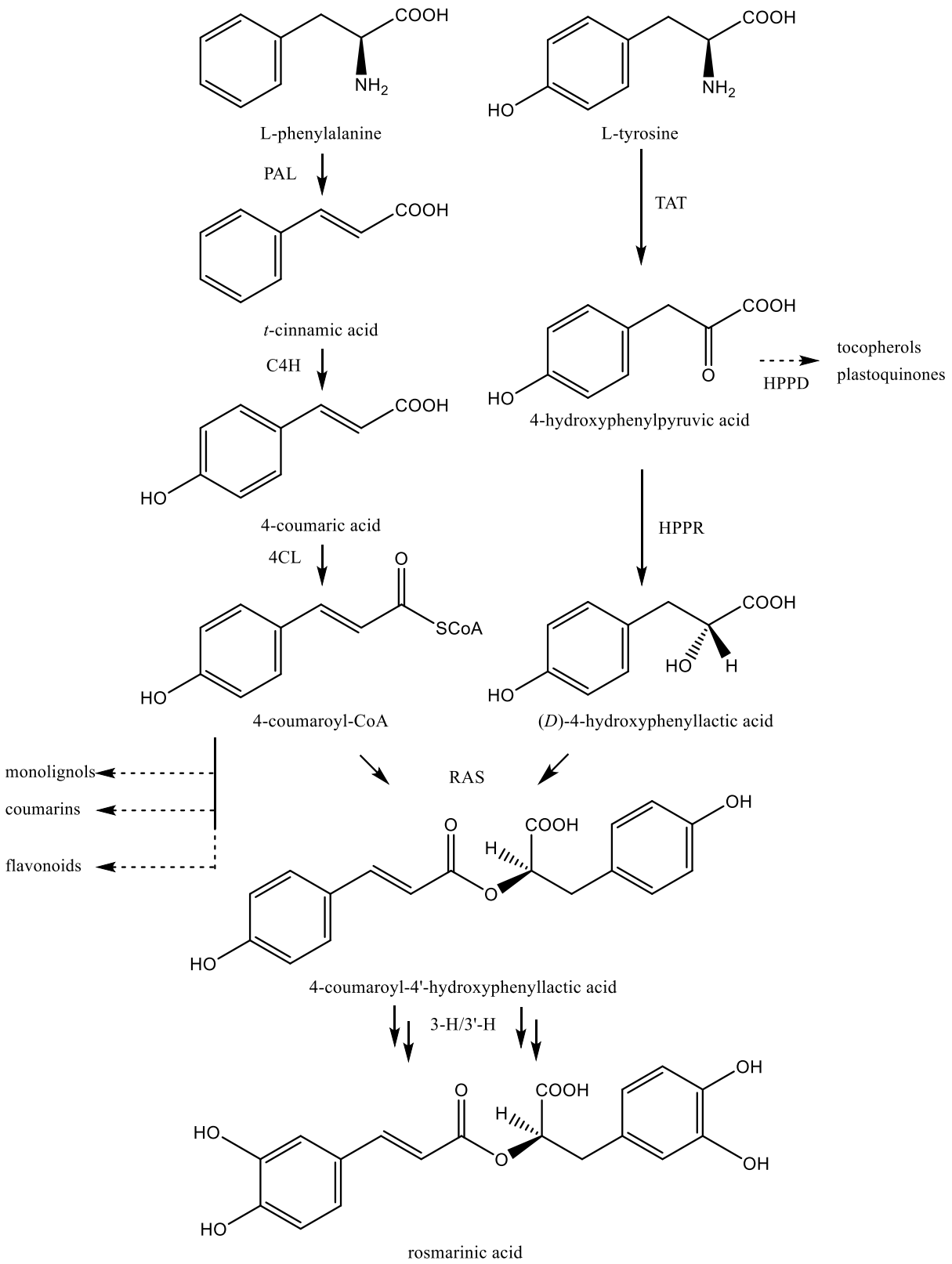


Figure 2 Biosynthethic pathway of rosmarinic acid in *Coleus blumei* according to Petersen et al. (1993), adapted. PAL = phenylalanine ammonia-lyase; C4H = cinnamic acid 4-hydroxylase; 4CL = hydroxycinnamate:coenzyme A ligase; TAT = tyrosine aminotransferase; HPPR = hydroxyphenylpyruvate reductase; RAS = rosmarinic acid synthase; 3-H/3'-H = 3-hydroxylase and 3'-hydroxylase; HPPD = 4-hydroxyphenylpyruvate dioxygenase. Additional metabolites are indicated.

6 Rosmarinic acid biosynthesis in *Anthoceros agrestis*

Whether rosmarinic acid biosynthesis in *Anthoceros agrestis* follows the model from *Coleus blumei* and this pathway possibly originated early in land plant development, or whether it is an independent evolution, can only be clarified by elucidating the enzymes involved. Pezeshki et al. (2022) identified and characterised two PAL isoforms from *Anthoceros agrestis* that demonstrated comparable substrate affinity to typical PALs of spermatophytes. Additionally, a 4-hydroxylase activity leading to 4-coumaric acid was already detected and identified as cytochrome P450-dependent enzyme by Petersen (2003). The enzyme was functionally expressed in *Physcomitrium patens* and biochemically characterised by Wohl and Petersen (2020a). The phenylpropanoid-derived part is completed by the successful expression and characterization of two 4CL variants that showed comparable affinities for the two substrates 4-coumaric acid and isoferulic acid (Wohl and Petersen 2020b). The search for an enzyme corresponding to RAS from *Coleus blumei* has so far been unsuccessful. RAS is classified as hydroxycinnamoyltransferase (HCT) from the BAHD superfamily, a family able to transfer a variety of acyl groups, resulting in a large number of potential donor/acceptor combinations and complicating functional predictions based on sequence information (D'Auria 2006; Berger et al. 2006). Most recently, Ernst et al. (2022) reported about a HCT isoform, which accepted 4-coumaroyl-CoA and caffeoyl-CoA as donor substrates but did not utilise 4-hydroxyphenyllactic acid as acceptor. Instead, shikimic acid was the preferred substrate while amide formation with an anthranilic acid derivative was also possible. This publication also reported on a 3-hydroxylase (CYP98) that accepted 4-coumaroyl-4'-hydroxyphenyllactic acid, but resulting in rather low product formation, while favouring hydroxylation of amides. Work on the L-tyrosine-derived precursor 4-hydroxyphenyllactic acid and the associated enzymes HPPR and TAT will be discussed as part of this work. Even though this fills further gaps in rosmarinic acid biosynthesis, the analogue to RAS from *Coleus blumei* remains the great unknown. Other classes of enzymes, away from the use of CoA-activated 4-coumaric acid, e.g. serine carboxypeptidase-like acyltransferases, which utilise glucose-activated precursors, or lipase-like hydroxycinnamoyltransferases should be taken into consideration (Petersen 2016).

7 Relevant enzymes of L-tyrosine derived metabolism

7.1 Tyrosine aminotransferase (TAT)

TAT (EC 2.6.1.5) is an aromatic aminotransferase which catalyses the reversible transfer of the amino group from L-tyrosine to 2-oxoglutarate generating 4-hydroxyphenylpyruvate and L-glutamate. Aromatic aminotransferases are representatives of pyridoxal 5'-phosphate (PLP)-dependent enzymes. They play an important role in both biosynthesis, i.e. aromatic amino acids in bacteria and fungi, as well as degradative processes, with the latter being important and predominant for plant secondary metabolism (Schenck and Maeda 2018; Xu et al. 2019, chapter I.5). The reaction requires an amino acid (donor) and a 2-oxoacid (acceptor) and is therefore characterised by dual substrate recognition. Aminotransferases generally exhibit a promiscuous character, as demonstrated by Wang and Maeda (2018), which discuss a wide variety of aromatic aminotransferases from different species, making it difficult to draw conclusions about the accepted substrates based only on sequence comparisons. A biochemical analysis of the enzyme is essential. The two subunits of the homodimeric enzyme each consist of two domains (large and small). The catalytic sites are located at the domain interface, resulting in the formation of the active site pocket through incorporation of the subunits (Hirotsu et al. 2005). There are four subgroups of aminotransferases (ATs), with aromatic ATs belonging to subgroup 1 together with aspartate or alanine aminotransferases (Mehta et al. 1993). The mechanism is fundamentally dependent on a central Lys-residue whose ε-amino group forms an internal aldimine with PLP (Kirsch et al. 1984). The detailed mechanism is illustrated by the example of aspartate aminotransferase (Fig. 3), an enzyme that can serve as blueprint for this enzyme family. In contrast to other two-substrate reactions (bi bi), which require the simultaneous binding of the substrates involved (sequential reaction), the reaction is divided into two successive phases. Starting from the internal aldimine, the Lys-residue is initially displaced by the incoming L-aspartate. The external aldimine is formed, which is converted into the ketimine via a quinoid intermediate and subsequently hydrolysed with oxaloacetate leaving the reaction. The cofactor is now present as pyridoxamine phosphate (PMP), acting as caretaker of the amino group. A new ketimine is formed by a condensation reaction with the incoming 2-oxoglutarate, which is converted back

to an external aldimine via the quinoid intermediate. With dissociation of L-glutamate, the initial internal aldimine is formed, regenerating PLP. The whole process is mechanistically called ping pong bi bi (Frey and Hegeman 2007; Berg et al. 2013). The involvement of TAT in rosmarinic acid biosynthesis has been demonstrated for *Coleus blumei* (chapter I.5; De-Eknamkul and Ellis 1987a) and *Perilla frutescens* (Lamiaceae) (Lu et al. 2013a). The presence of a TAT capable of transaminating L-tyrosine in *Anthoceros agrestis* was recently demonstrated (Busch and Petersen 2021).

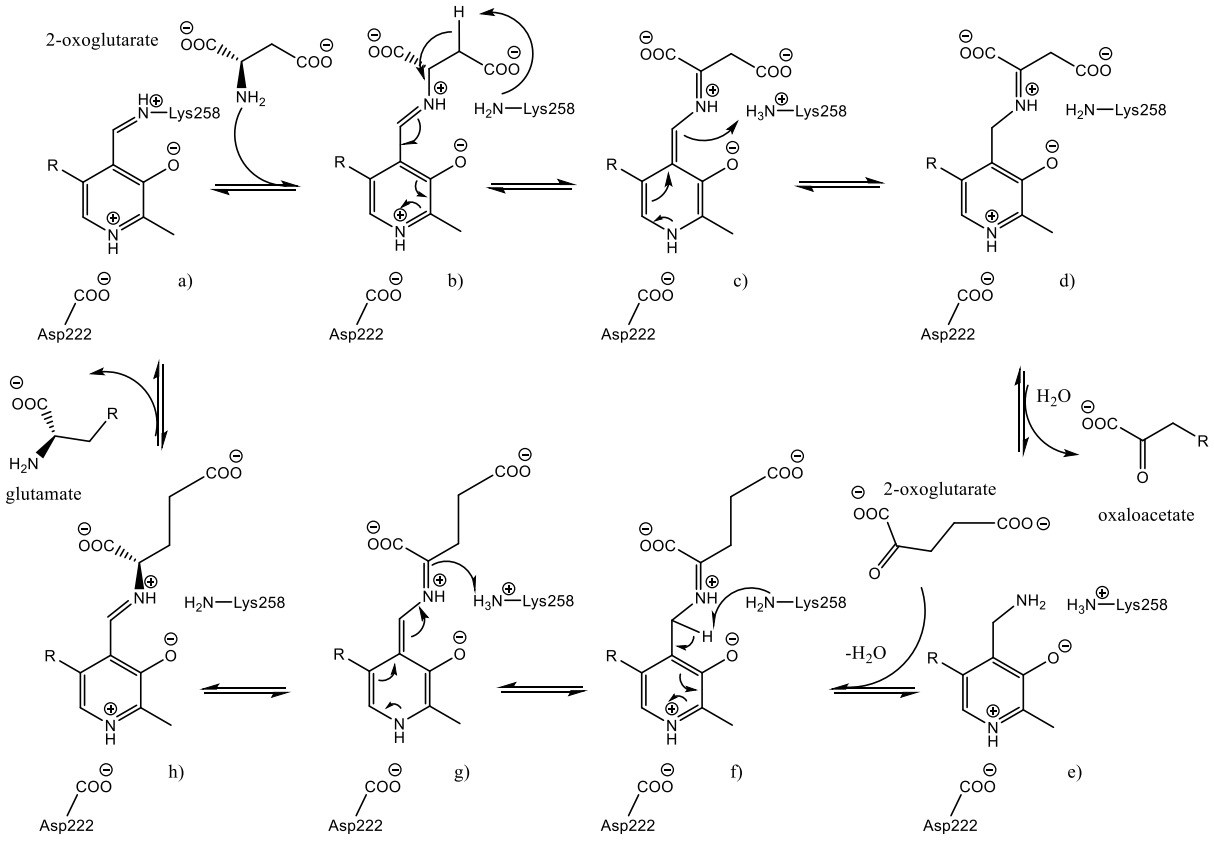


Figure 3 Reaction mechanism of aspartate aminotransferase according to Frey and Hegeman (2007), adapted. PLP is covalently bound to the Lys-residue forming an internal aldimine (a). 2-oxoglutarate (amino donor) replaces the Lys-residue and forms the external aldimine (b). A ketimine (d) is formed via a quinoid intermediate (c). Oxaloacetate dissociates and pyridoxamine remains (e). 2-oxoglutarate attaches and a new ketimine is formed (f), which is converted to the external aldimine (h) via the quinoid intermediate (g). Glutamate dissociates and the internal aldimine regenerates (a).

7.2 Hydroxyphenylpyruvate reductase (HPPR)

HPPR (EC 1.1.1.237) catalyses the NAD(P)H+H^+-dependent reduction of 4-hydroxyphenylpyruvic acid (pHPP) to 4-hydroxyphenyllactic acid (Fig. 4) but also transforms

related derivatives such as 3,4-dihydroxyphenylpyruvic acid and 4-hydroxy-3-methoxyphenylpyruvic acid (Petersen and Alfermann 1988; Häusler et al. 1991). Since the initial description, NADPH has been established as the primary co-substrate, although NADH shows decent affinity and activity as well, and the enzyme has been classified as a D-isomer-specific 2-hydroxyacid dehydrogenase (Meinhard 1991; Häusler et al. 1991; Kim et al. 2004). The heterologously expressed putative HPPR from *Coleus blumei* also revealed additional non-aromatic substrates matching this family, such as hydroxypyruvate and glyoxylate, each significantly exceeding the affinity of the aromatic substrates. In addition, D-isomer specific product formation was demonstrated and HPPR of *Coleus blumei* was resolved as dimeric enzyme through crystallisation and gel permeation chromatography (Janiak 2007). The D-isomer-specific 2-hydroxyacid dehydrogenase family was first introduced by Grant (1989) and comprises a rather diverse group of prokaryotic (e.g. D-3-phosphoglycerate dehydrogenase, D-2-hydroxyisocaproate dehydrogenase, D-lactate dehydrogenase; Taguchi and Ohta 1991), animal (glyoxylate/hydroxypyruvate reductase; Booth et al. 2006), fungal (phenylpyruvate reductase; Fujii et al. 2011) and plant (e.g. D-glycerate dehydrogenase, syn. hydroxypyruvate reductase; Greenler et al. 1989) enzymes, interestingly not closely related to the corresponding L-isomer-specific enzymes, which constitute a family on their own (Kochhar et al. 1992 and sources cited therein). The family comprises dimeric, sometimes tetrameric, enzymes, with each monomer consisting of two domains, a co-substrate and a substrate-binding domain, with a deep catalytic cleft at the interdomain region (Holton et al. 2013; Goldberg et al. 1994). Janiak et al. (2010) was able to determine the crystal structure of H(P)PR from *Coleus blumei* both as apoenzyme and in complex with $NADP^+$. The monomer with a mass of 34.11 kDa and active site located within the said deep cleft, separating the two domains, confirmed the typical topology of a D-isomer-specific 2-hydroxyacid dehydrogenase. The larger domain carries the conserved motif -G-X-G-X-X-G- serving as NAD(P)/NAD(P)H-binding domain, while the two amino acids Arg232 and His279, which are also conserved in other dehydrogenases, form the active site, supported by Glu261, forming a hydrogen bond with the latter. How exactly HPPR fits into this family of mainly primary metabolic enzymes, remains an interesting issue. While aromatic aminotransferases play a weighty role in many metabolic processes, i.e. biosynthesis of primary and specialised compounds, and are therefore studied for multiple backgrounds (Wang and Maeda 2018), HPPR has always been related to rosmarinic acid in the few reports existing besides the comprehensive data on

Coleus blumei (e.g. Mizukami et al. 1993; Xiao et al. 2009; Barberini et al. 2013; Ahmad et al. 2016; Lu et al. 2013b; Wang et al. 2017), with the exception of *Arabidopsis thaliana*, in which, since no rosmarinic acid has yet been detected, HPPR isoforms of a possibly more general tyrosine-derived metabolism are present (Xu et al. 2018). HPPR competes with hydroxyphenylpyruvate dioxygenase (HPPD; Fig. 2) for its substrate and probably is the first specific enzyme in rosmarinic acid biosynthesis, thus forming the junction between primary and secondary metabolism. This inevitably raises questions about the origin of HPPR. Whether HPPR originates in hydroxypyruvate reductase (HPR), a mandatory photorespiratory enzyme, or whether HPPR capacity is attributable to a secondary reaction or rededication of HPR activity remains to be clarified.

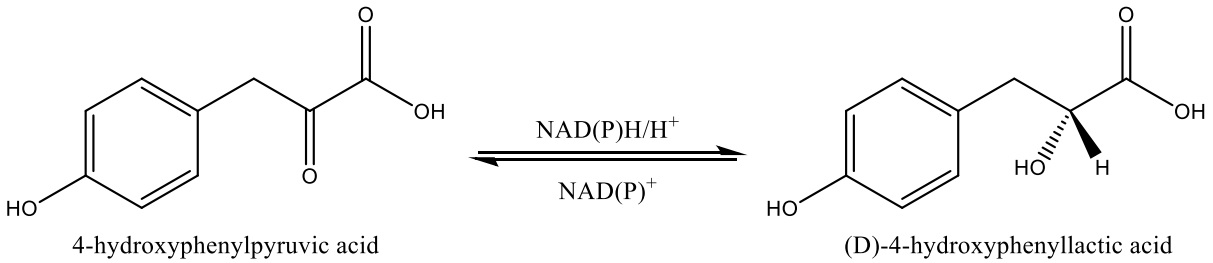


Figure 4 Stereospecific HPPR reaction. Reduction of 4-hydroxyphenylpyruvic acid to (D/R)-4-hydroxyphenyllactic acid and simultaneous oxidation of NAD(P)H/H^+.

8 Photorespiration

Photorespiration is a process that occurs in all oxygenic photosynthetic organisms. While there are adaptations of certain plants (C4 and CAM plants), the initial carbon fixation in C3 plants occurs during the Calvin cycle through the enzyme ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO) which catalyses the carboxylation of ribulose-1,5-bisphosphate, followed by formation of two molecules 3-phosphoglycerate (Reumann and Weber 2006). Those are, after further intermediate steps (e.g. reduction to the carbohydrate), proportionally directed into biosynthesis of carbon skeletons, i.e. starch, sucrose (Raines 2003), while 5/6 of the resulting C3 bodies are channelled into regenerating ribulose-1,5-bisphosphate (Reumann and Weber 2006). However, the enzyme also accepts O_2, resulting in the formation of one molecule 3-phophoglycerate and one molecule toxic 2-phosphoglycolate. The rescue of one molecule 3-phosphoglycerate by the use of two molecules 2-phosphoglycolate at the loss of one molecule CO_2 is achieved by the C2

photorespiratory mechanism spanning three compartments, i.e. chloroplast, peroxisome and mitochondrion (Reumann and Weber 2006). In this process (Fig. 5), peroxisomal hydroxypyruvate reductase (HPR; EC 1.1.1.29) catalyses the NADH-dependent reduction (Tolbert et al. 1970) of hydroxypyruvate to D-glycerate (reviews: Reumann and Weber 2006; Bauwe et al. 2010). Surprisingly, mutants lacking activity of peroxisomal HPR or malate dehydrogenase, which provides peroxisomal NADH (Yu and Huang 1986), do not lead to a collapse in photorespiratory metabolism or photosynthetic rates, as seen in other photorespiration-related mutants that grow only in high CO_2 environments (Murray et al. 1989; Cousins et al. 2008; Timm et al. 2008). This is due to the presence of a cytosolic NAD(P)H-dependent HPR (EC 1.1.1.81) providing an extra-peroxisomal bypass and even a third, chloroplast enzyme, as indicated and eventually suggested for *Arabidopsis thaliana* (Kleczkowski et al. 1988; Kleczkowski and Randall 1988; Timm et al. 2011). The three enzymes, plus a fourth for which no activity has yet been demonstrated, were recently located (GFP fusion) and re-evaluated by Xu et al. (2018). HPR1, favouring NADH, was localised in peroxisomes, while HPPR2, HPPR3 (HPR2 & HPR3 according to Timm et al. 2011) and HPPR4 were localised in the cytosol. Whereas HPPR4 activity was not detectable, HPPR2 and HPPR3 favoured NADPH and accepted hydroxypyruvate besides pHPP. A phylogenetic analysis confirmed HPR1 and the HPPRs as two subfamilies of D-isomer-specific 2-hydroxyacid dehydrogenases and HPPR2 sharing the highest identity with *Coleus blumei* HPPR. They suggest, that HPPRs likely evolved from the HPR clade in ancestral land plants (Xu et al. 2018).

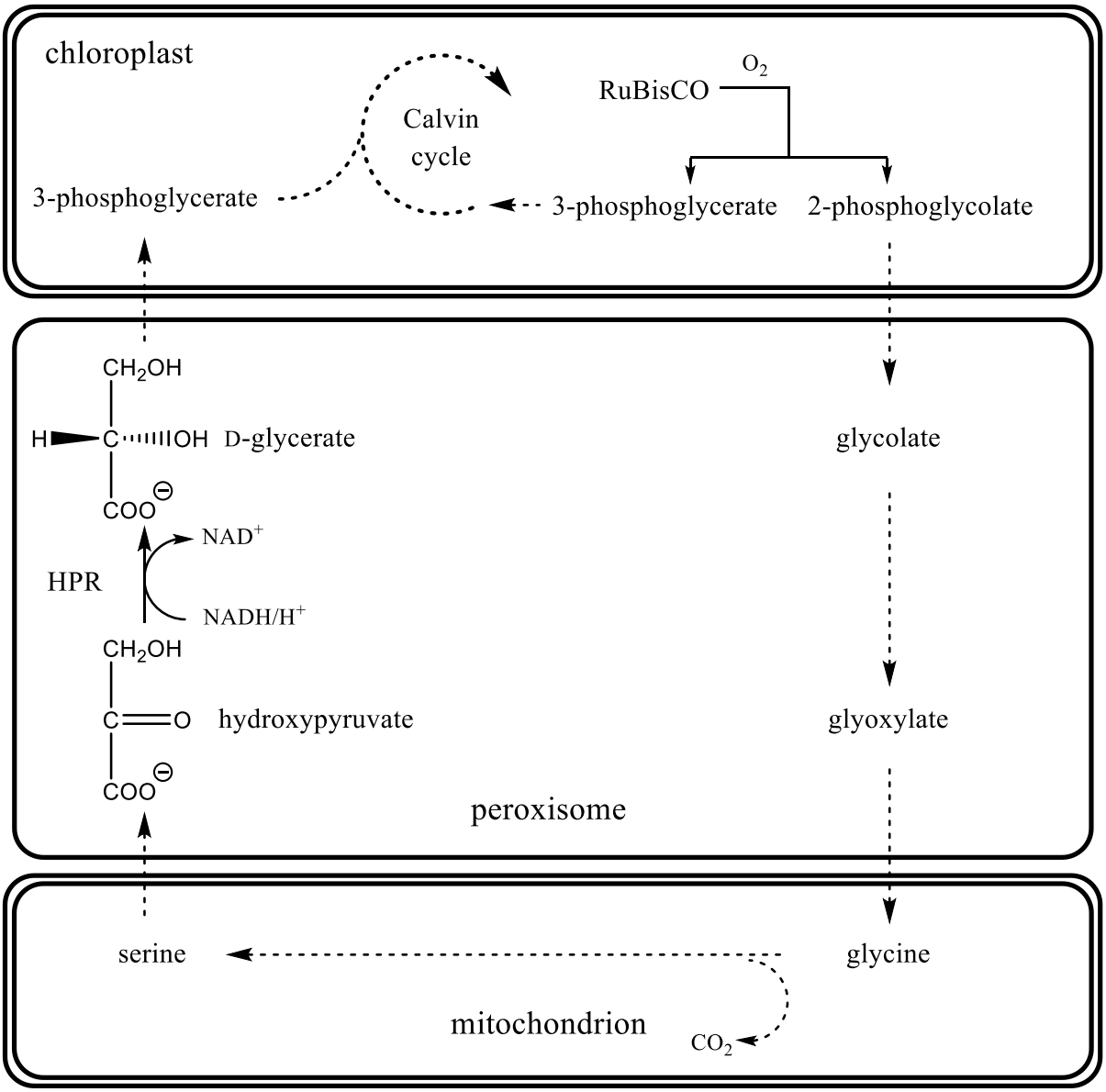


Figure 5 Schematic representation of the photorespiratory mechanism based on Bauwe et al. (2010). The 2-phosphoglycolate formed by O_2 exposure of RuBisCO is rescued after passing through several intermediates and compartments. The peroxisomal HPR catalyses the reduction of hydroxypyruvate to D-glycerate. RuBisCO = ribulose-1,5-bisphosphate carboxylase-oxygenase; HPR = hydroxypyruvate reductase. Non-relevant reaction steps are indicated by dashed arrows and do not reflect physiological processes and actual stoichiometry.

9 Aim of this work

Anthoceros agrestis and related species contain several interesting secondary metabolites derived from hydroxycinnamic acid and 4-hydroxyphenyllactic acid moieties, such as anthocerotonic acid, megacerotonic acid and rosmarinic acid. In this regard, rosmarinic acid, as an ester of the two moieties, provides an interesting starting point to study the biosynthetic pathways of phenolic components in *Anthoceros agrestis*, especially since the elucidation has already been achieved for *Coleus blumei* (Petersen et al. 1993). While the enzymes of the general phenylpropanoid pathway, with the exception of a HCT that catalyses ester formation (= RAS), have already been identified and heterologously expressed (Pezeshki et al. 2022; Petersen 2003; Wohl and Petersen 2020a, 2020b; sorting according to biosynthesis), little progress has been made on the L-tyrosine-derived pathway. Identifying and characterising the enzymes involved, is the objective of this work.

Sequence alignments of promising *Anthoceros* scaffolds and known TAT, HPPR and HPR sequences were to be found. Potential genes had to be amplified and introduced into suitable vectors. Heterologous expression in *E. coli* had to be used to gain active enzyme for biochemical characterisation.

The determined parameters, e.g. substrate acceptance, kinetics and specificity, should allow conclusions about the comparability of rosmarinic acid biosynthesis of *Anthoceros agrestis* and higher plants, i.e. *Coleus blumei*, and reveal new information about the biosynthesis of secondary metabolites in hornworts.

Whether the idea of monophyly or polyphyly of bryophytes will prevail in the long run, in either case, insights into the phenolic compounds of bryophytes, which were either essential in colonising the land as predecessors of vascular plants or were necessary to inhabit a niche in parallel with vascular plants ascension for millions of years, will be valuable.

II Material and Methods

1 General material and methods

1.1 Plant material

The plant material required over the course of this research was collected from *Anthoceros agrestis* (syn. *A. crispulus*) suspension cultures. Two lines in hormone-free CB medium (CBM, see II.1.2.4) were established by Professor Petersen (Petersen 2003), originally deriving from a callus culture kindly provided by Professor Binding (Binding and Mordhorst 1991). A third line was sustained in CB medium with increased sugar content and addition of hormones and N-Z-Amine A (CB2, see I.1.2.4). Viability and growth of the suspension cultures was achieved by weekly transfer of 5.5 g (cell wet weight) cells to 50 ml of fresh medium and cultivation under permanent daylight conditions on a gyratory shaker (100 rpm) at 25 °C.

1.2 Media

1.2.1 Lysogeny broth (LB)

LB was used as standard medium for *E. coli*, either for selection on agar or for plasmid accumulation. It was prepared according to Bertani (1951) without glucose addition. For liquid medium, 100 ml aliquots of 1% (w/v) tryptone/peptone, 0.5% (w/v) yeast extract and 1% (w/v) NaCl, adjusted to pH 7.0 with NaOH were autoclaved and stored under dark conditions until use. For solid medium 1% (w/v) agar was added prior to autoclaving and thoroughly dissolved when still hot. To obtain LB agar plates, solid medium was heated using a microwave and was poured into Petri dishes. If selection through antibiotics or blue-white screening was needed, the medium was supplemented according to chapter II.2.6.5.

1.2.2 Terrific broth (TB)

TB medium was prepared according to Tartoff and Hobbs (1987). This rich medium allows for higher cell density than LB medium and was used for heterologous expression, after an initial attempt with LB medium brought insufficient yields. One flask contains 90 ml of aqueous nutrient solution (1.2% (w/v) tryptone/peptone, 2.4% (w/v) yeast extract and 0.4% (v/v)

glycerol, autoclaved) supplemented with 10 ml of 10x TB salts (0.17 M KH_2PO_4 and 0.72 M K_2HPO_4, autoclaved) prior to use. The separate solutions can be stored under dark conditions, but ideally were prepared fresh before each expression.

1.2.3 Super Optimal broth with Catabolite repression (SOC)

SOC is a nutrient rich medium especially formulated to increase plasmid transformation efficiency through enhancing recovery of competent *E. coli*. It is an adapted version of SOB, supplemented by glucose. The formula is customised according to Hanahan (1983). It contains 2% (w/v) tryptone/peptone, 0.5% (w/v) yeast extract, 8.6 mM NaCl and 2.5 mM KCl dissolved in water, adjusted to pH 7.0 with NaOH and autoclaved. Filter-sterilised 1 M $MgSO_4$ solution and 1 M glucose solution were then used to adjust the concentration to 20 mM each.

1.2.4 Plant cell media

Anthoceros agrestis suspension cells were cultivated using CBM and CB2 medium, as described in II.1.1. CB2 medium was prepared according to Petersen and Alfermann (1988), whereas CBM is a hormone-free, N-Z-Amine A-free variant with halved sugar content; details are given below.

Classification	**Substance**	**CB2 medium**	**CBM medium**
Macro elements	KNO_3	2500.0 mg/l	2500.0 mg/l
	$MgSO_4 \cdot 7\ H_2O$	250.0 mg/l	250.0 mg/l
	$NaH_2PO_4 \cdot 2\ H_2O$	194.4 mg/l	194.4 mg/l
	$CaCl_2 \cdot 2\ H_2O$	150.0 mg/l	150.0 mg/l
	$(NH_4)_2SO_4$	134.0 mg/l	134.0 mg/l
	$FeSO4 \cdot 7\ H_2O$	25.6 mg/l	25.6 mg/l
	$EDTA\text{-}Na_2 \cdot 2\ H_2O$	34.3 mg/l	34.3 mg/l
Micro elements	H_3BO_3	3.0 mg/l	3.0 mg/l
	$ZnSO_4 \cdot 7\ H_2O$	3.0 mg/l	3.0 mg/l
	$MnSO_4 \cdot 1\ H_2O$	1.0 mg/l	1.0 mg/l
	KJ	0.75 mg/l	0.75 mg/l
	$Na_2MoO_4 \cdot 2\ H_2O$	0.25 mg/l	0.25 mg/l
	$CuSO_4 \cdot 5\ H_2O$	0.25 mg/l	0.25 mg/l

	$CoCl_2 \cdot 6\ H_2O$	0.25 mg/l	0.25 mg/l
Vitamins	thiamine-HCl	10.0 mg/l	10.0 mg/l
	pyridoxine-HCl	1.0 mg/l	1.0 mg/l
	nicotinic acid	1.0 mg/l	1.0 mg/l
Hormones	indole-3-acetic acid (IAA)	0.5 mg/l	
	2,4-dichlorophenoxyacetic acid (2,4-D)	2.0 mg/l	
	kinetin	0.2 mg/l	
	1-naphthaleneacetic acid (NAA)	0.5 mg/l	
	myo-inositol	100.0 mg/l	100.0 mg/l
	sucrose	20.0 g/l	10.0 g/l
	N-Z-Amine® A	2.0 g/l	
pH value	After dissolving, the medium was adjusted to pH 5.5 with 0.5 M KOH (CMB) or 0.5 M HCl (CB2).		

Aliquots of 50 ml medium were transferred to 250 ml flasks, sealed with a cellulose stopper, and autoclaved. Medium was stored under dark conditions until use.

1.3 Competent cells

Two strains of competent bacteria were used in this work, namely *E. coli* EZ cells for plasmid multiplication, particularly in preparation of Sanger sequencing, and *E. coli* SoluBL21 for protein expression. A test tube containing 2 ml liquid LB medium and 25 µg tetracycline (12.5 mg/ml stock solution in ethanol) was inoculated with a sample of cells scraped of a permanent *E. coli* EZ glycerol stock (stored at -80 °C) using a sterile toothpick, for *E. coli* SoluBL21, 2 ml liquid LB medium without antibiotic was used. The tube was loosely covered with an aluminium cap and incubated for 14 to 18 h at 220 rpm and 37 °C. Afterwards the sample was poured into a 250 ml Erlenmeyer flask containing 90 ml liquid LB medium and 90 µl 12.5 mg/ml tetracycline (only EZ) and incubated (220 rpm, 37 °C). Optical density (OD_{600}) was measured several times against a medium reference until a value of 0.4-0.6 was reached. The cell suspension was divided between two sterile 50 ml tubes and centrifuged (3000 *g*, 4 °C, 10 min). After discarding the supernatant, the residue was resuspended in 10 ml cold sterile 100 mM calcium chloride solution and samples were combined. The supernatant was

discarded after centrifugation (2500 *g*, 4 °C, 12 min) and resuspended in 10 ml cold sterile 100 mM calcium chloride solution followed by incubation on ice for 20 min. The supernatant was discarded after centrifugation (2500 *g*, 4 °C, 12 min) and the residue was resuspended in 2 ml calcium chloride/15% glycerol. Aliquots of 150 µl each were prepared using sterile 1.5 ml reaction tubes, subsequently frozen with liquid N_2 and stored at -80 °C.

1.4 Agarose gel electrophoresis

Agarose gels were freshly prepared with the addition of 10-20 µg ethidium bromide. For general purposes 0.7% agarose in 1x TAE buffer (20 mM acetic acid, 1 mM EDTA, 40 mM Tris, pH 8) was sufficient. Regarding the demands of some targets, concentrations of up to 1.5% agarose were used. Samples were mixed with 6x loading dye (0.03% (m/v) bromophenol blue, 0.03% (m/v) xylene cyanol, 60 mM EDTA in 60% (w/v) glycerol) and the gel pockets were charged. To visualise nucleic acid sizes, 4-6 µl GeneRuler DNA Ladder Mix or GeneRuler 1 kb DNA Ladder (Thermo Scientific) were used. Migration was achieved with an applied potential of 120 V using TAE buffer. If electrophoresis was used to check RNA integrity, the entire apparatus was thoroughly cleaned to ensure protection from unwanted degradation throughout the process. Then, fresh buffer and, where possible, double-autoclaved equipment was used. The gel was analysed under blue/green LED light, either using an amber filter or captured through photography.

1.5 Gel extraction and purification of PCR amplicons

Gel extration of separated PCR products was performed using the NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel). Bands which corresponded the expected target size were cut out from the agarose gel using a scalpel and were liquefied utilising 200 µl binding buffer NT per 100 mg withdrawn material at 50 °C for ~10 min. The fluid was transferred to a NucleoSpin silica column, centrifuged (11000 *g*, 1 min) and the flow-through discarded. The silica membrane containing target DNA was washed through addition of 700 µl buffer NT3 (10 mM Tris/HCl pH 7.5, 80% ethanol), followed by centrifugation (11000 *g*, 1 min) and discarding of flow-through once again. The membrane was dried (11000 *g*, 2 min) and 15-50 µl, commonly 30 µl, sterile H_2O were added. After 5 min, the spin column was attached to a fresh reaction tube and the elution was collected by centrifugation (11000 *g*, 1 min). Samples were stored at -20 °C.

2 Molecular Biology

2.1 gDNA extraction

The extraction of gDNA was performed according to Rogers and Bendich (1985). Plant material, separated from medium through vacuum filtration, was ground in a mortar with addition of liquid nitrogen. Approximately 200 mg of the ground material was collected in a reaction tube and 300 µl preheated (65 °C) 2x CTAB buffer (2% (w/v) cetyltrimethylammonium bromide (CTAB), 100 mM Tris/HCl pH 8, EDTA pH 8, 1.4 M NaCl, 1% (w/v) polyvinylpyrrolidone) were added. The mixture was stalled on ice for 15 min after incubation (65 °C for 30 min) with occasional shaking. 300 µl chloroform was added followed by shaking and centrifugation (16000 *g*, 5 min). The supernatant was transferred into a fresh reaction tube prepared with 30 µl preheated (65 °C) 10x CTAB buffer (10% (w/v) CTAB, 0.7 M NaCl) and 300 µl chloroform was added. Following shaking and centrifugation (15700 g, 5 min) the supernatant was transferred to a fresh reaction tube containing 250 µl CTAB precipitation buffer (1% (w/v) CTAB, 50 mM Tris/HCl pH 8, 10 mM EDTA pH 8). The supernatant was discarded after shaking and centrifugation (15700 *g*, 10 min) and the pellet was subsequently redissolved in 200 µl high salt TE buffer (10 mM Tris/HCl, 1 mM EDTA pH 8, 1 M NaCl). 300 µl 100% ethanol was added to the sample followed by mixing, precipitation (-20 °C, 15 min), centrifugation (15700 *g*, 5 min) and subsequent discarding of the supernatant. The obtained pellet was purified with 400 µl 70% ethanol followed by centrifugation (15700 *g*, 5 min). The supernatant was gently removed and the pellet was dried (37 °C, 5 min). The pellet was dissolved in 30 µl 1/10 TE buffer containing 100 µg/ml RNase A. Samples were stored at -20 °C.

2.2 RNA extraction

The extraction of total RNA was performed using the acid guanidinium thiocyanate-phenol-chloroform extraction according to Chomczynski and Sacchi (1987). To guarantee preservation of the RNA throughout the whole process, a separated area, double-autoclaved equipment and solutions or, if non autoclavable, heat sterilised (300 °C, 2 h) equipment was used. Gloves had to be worn throughout the whole process. Plant material, separated from medium through vacuum filtration, was ground in a precooled mortar while frequently adding liquid nitrogen. To approximately 50 mg of ground plant material 500 µl solution D (4 M guanidinium thiocyanate, 25 mM citrate buffer pH 7, 0.5% (w/v) sodium lauroyl sarcosinate) was added

followed by shaking and incubation at room temperature. The preparation was gently mixed with 50 µl 2 M sodium acetate pH 4 and 500 µl phenol (equilibrated with 0.1 M citrate buffer pH 4.3) successively. After addition of 100 µl precooled (stored at -20 °C) chloroform and vigorous shaking for 10 s and incubation (on ice, 15 min), each sample was centrifuged (12000 *g*, 4 °C, 15 min). For precipitation, 400 µl of the supernatant was transferred into a fresh reaction tube and 400 µl precooled (stored at -20 °C) isopropanol was added, followed by incubation (-20 °C, 15 min) and centrifugation (12000 *g*, 4 °C, 10 min). Afterwards, the supernatant was discarded and the pellet was step-by-step purified with 400 µl 70% ethanol followed by centrifugation (7500 *g*, 4 °C, 5 min) and discarding the supernatant and 100% ethanol followed by centrifugation (7500 *g*, 4 °C, 5 min) and discarding the supernatant. The pellet was dried (37 °C, 5-10 min) and redissolved in 20 µl H_2O at 50 °C. Finally, RNA concentration and purity was determined photometrically. Additionally, RNA was analysed for integrity, in particular with regard to 28S, 18S and 5S rRNA, and possible DNA contamination electrophoretically (II.1.4). Samples were stored at -80 °C (if designated to qPCR) or at -20 °C.

2.3 Digestion with DNase I

DNase digestion was mandatory in cases where RNA was prepared for qPCR applications. 5 µg total RNA was mixed with 5 µl 10x reaction buffer with $MgCl_2$ and 5 µl DNase I (1 U/µl, RNase free) and adjusted to 50 µl with H_2O. The preparation was incubated (37 °C, 30 min) and extracted by phenol-chloroform extraction. The protocol corresponds to the extraction described in chapter II.2.2 with one tenth of the volumes omitting the addition of solution D. As before, extracted RNA was redissolved in 20 µl H_2O and RNA concentration and purity was determined photometrically.

2.4 cDNA synthesis

2.4.1 cDNA synthesis for PCR

cDNA for standard PCR reactions was synthesised using the RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific). For that, 1 or 2 µg of total template RNA was supplemented with 1 µl Oligo$(dT)_{18}$ primer and adjusted to 12 µl total volume with RNase-free H_2O. After mixing and brief centrifugation (as required), the sample was incubated at 65 °C for 5 min and cooled on ice afterwards. 4 µl 5x reaction buffer, 1 µl RiboLock RNase Inhibitor (20 U/µl), 2 µl

dNTPs (10 mM), 1 µl RevertAid RT (200 U/µl) were added in the indicated order, mixed and centrifuged briefly. Eventually, the mixture was incubated at 42 °C for 60 min and the reaction was finally stopped at 70 °C for 5 min. Obtained cDNA was stored at -20 °C.

2.4.2 cDNA synthesis for PCR and qPCR

RNA for qPCR was transcribed using the qScript cDNA Supermix kit. It was also used to obtain cDNA for standard PCR experiments in addition to the procedure described in chapter II.2.4.1. Total template RNA, optionally digested with DNase I, was mixed with 4 µl 5x qScript cDNA Supermix and adjusted to 20 µl with RNase-free H_2O in micro reaction tubes. For cDNA synthesis intended to be used in general PCR applications, 1 µg RNA was used. If dedicated to qPCR, 0.5 µg was suitable. The mixture was subjected to a three-step temperature program (25 °C for 5 min, then 42 °C for 30 min, finally 85 °C for 5 min) in a thermocycler and the cDNA was stored at -20 °C (PCR applications) or -80 °C (qPCR). Samples to be used in qPCR were diluted with 80 µl H_2O before use or further dilution.

2.4.3 cDNA synthesis for RACE PCR

cDNA for RACE PCR was transcribed using the SMARTer® RACE 5'/3' Kit (Clontech). The resulting cDNA is referred to as 5'- and 3'-RACE-Ready cDNA. A buffer mix containing 4 µl 5x First-Strand Buffer, 0.5 µl DTT (100 mM) and 1 µl dNTPs (20 mM) was prepared in excess for later use. Since RACE PCR requires two separate amplifications with two adapted cDNAs, viz. 5'- and 3'-RACE-Ready cDNA, two reactions must be prepared. The former containing 1 to 10 µl total RNA, 1 µl 5' CDS Primer A, adjusted to 11 µl total volume with sterile H_2O. The latter with 1 to 11 µl total RNA, 1 µl 3' CDS Primer A in a total volume of 12 µl. The contents were mixed and briefly centrifuged, then incubated at 72 °C for 3 min and at 42 °C for 2 min subsequently, followed by centrifugation (14000 *g*, 10 s). Additionally, the 5'-related mixture was complemented by addition of 1 µl SMARTer II A oligonucleotide. In the following, a master mix containing 5.5 µl buffer mix (see above), 0.5 µl RNase Inhibitor (40 U/µl) and 2 µl SMARTScribe Reverse Transcriptase (100 U) was added to the respective 5' and 3' reaction mixtures, followed by mixing and slight centrifugation. The reactions were incubated at 42 °C for 90 min followed by 70 °C for 10 min. Eventually, transcribed cDNA can be diluted with Tricine-EDTA buffer depending on the total RNA concentration used as described in the manual.

2.5 Polymerase chain reaction (PCR)

2.5.1 Standard PCR

The standard PCR protocol was used in reactions aimed to amplify partial, if only fragmentary data of the putative gene was available, as well as full-length sequences. Although sharing the underlying method, some aspects are different. Partial sequences were amplified using GoTaq Polymerase (Promega). Since amplifications of full-length sequences were carried out in preparation of an expression system, a proof-reading polymerase in order to reduce errors as well as gene specific primers containing a restriction site matching the given expression vector were used.

Protocol adapted for partial gene sequences:

1 µl template (cDNA)
5 µl GoTaq reaction buffer (5X)
3 µl $MgCl_2$ (25 mM)
0.5 µl dNTP mix (10 mM)
0.5 µl primer forward (10 µM)
0.5 µl primer reverse (10 µM)
0.1 µl GoTaq Polymerase (5 U/µl)
Adjusted to a final volume of 25 µl with sterile H_2O.

PCR program:

Step / **Cycle**	**Denaturation**	**Annealing**	**Elongation**
1st	94 °C, 120 s	Tm - 5 °C*, 60 s	70 °C, 90 s
2nd-39th	94 °C, 30 s	Tm - 5 °C*, 60 s	70 °C, 90 s
40th	94 °C, 60 s	Tm - 5 °C*, 60 s	70 °C, 600 s
After completion kept at 6-10 °C			

*usually, or gradient of up to T_m – 10 °C.

Protocol adapted for full length sequences:

1 µl template (commonly cDNA or gDNA)
5 µl Phusion HF buffer with $MgCl_2$ (5X)
0.5 µl dNTP mix (10 mM)
0.5 µl primer forward (10 µM)
0.5 µl primer reverse (10 µM)

1 µl Phusion High-Fidelity DNA Polymerase (2 U/µl)
Adjusted to a final volume of 25 µl with sterile H_2O.

PCR program:

Step / Cycle	Denaturation	Annealing	Elongation
1st	94 °C, 120 s	T_m - 5 °C*, 60 s	72 °C, 90 s
2nd-39th	94 °C, 30 s	T_m - 5 °C*, 60 s	72 °C, 90 s
40th	94 °C, 60 s	T_m - 5 °C*, 60 s	72 °C, 300 s
After completion kept at 6-10 °C			

*usually, or gradient of up to T_m – 10 °C.

2.5.2 Rapid Amplification of cDNA ends (RACE) PCR

RACE PCR was used to obtain missing 5’ and 3’-directional regions of the underlying gene and is performed based on a known partial sequence. RACE PCR was performed using the SMARTer® RACE 5’/3’ Kit (Clontech). This kit provides a particular cloning method, called In-Fusion cloning (see II.2.6.2). This method was used occasionally in exchange for ligation and requires a customised primer design with an In-Fusion cloning overhang (see Kit manual and comment in primer list). RACE PCR is performed in separate reactions, in the following referred to as 5’-RACE and 3’-RACE. For both a shared master mix is used.

Master Mix:

15.5 µl PCR-grade H_2O
25.0 µl SeqAmp Buffer (2X)
1.0 µl SeqAmp DNA Polymerase

5’-RACE reaction:

2.5 µl 5’-RACE-ready cDNA (see II.2.4.3)
5 µl 10x Universal Primer Mix (10X)
1 µl 5’ gene specific primer (GSP, 10 µM)
41.5 µl master mix (from above)

3'-RACE reaction:

2.5 µl 3'-RACE-ready cDNA (see II.2.4.3)
5 µl 10x Universal Primer Mix
1 µl 3' gene specific primer (GSP, 10 µM)
41.5 µl master mix (from above)

Depending on the project, two different amplification programs were used. AaHPPR was amplified using program no. 1 and AaTAT using program no. 2.

RACE PCR program no. 1:

Step / **Cycle**	**Denaturation**	**Annealing**	**Elongation**
1st-35th	95 °C, 30 s	68 °C, 30 s	72 °C, 180 s
After completion kept at 6-10 °C			

RACE PCR program no. 2 (touchdown):

Step / **Cycle**	**Denaturation**	**Annealing**	**Elongation**
1st-5th	94 °C, 30 s	Combined: 72 °C, 180 s	
6th-10th	94 °C, 30 s	70 °C, 30 s	72 °C, 180 s
11th-37th	94 °C, 30 s	68 °C, 30 s	72 °C, 180 s
After completion kept at 6-10 °C			

2.5.3 Real-time PCR

The method, often also referred to as quantitative PCR (qPCR), enables quantification of mRNA abundance and was therefore used in gene expression studies. The experiment was performed to determine the relative expression rates of *HPPR*, *TAT*, *HPR1*, *HPPR2* in comparison to the housekeeping genes *actine* and *serine threonine protein phosphatase 2a regulatory subunit* (*St-P 2a*) as reference genes for standardisation. qPCR was performed using cDNA transcribed from DNase-digested RNA from *A. agrestis* (see II.2.2 and following) collected on scheduled days of cultivation (see II.4.4.1). Primers designed for qPCR must meet stringent requirements to produce reproducible and viable results. Therefore, primers with a maximum melting point deviation of 2 °C, a length of 20 ± 3 bases and resulting amplicons of 100-250 bp were designed (see primer list). To reliably exclude non-specific by-products, each

reaction was subjected to PCR and checked by agarose gel electrophoresis. Bands which corresponded the expected target size were cut out from the agarose gel using a scalpel and were extracted (II.1.5), ligated (II.2.6.1), transformed (II.2.6.5) and eventually sequenced. The main measurement was accomplished in a 96 well plate qPCR system (PikoReal Real-Time PCR System). The single reaction consists of:

Actin, *St-P 2a*, *TAT*, *HPPR*

5 µl PerfeCTa SYBR Green Supermix (2x)
4 µl template (cDNA)
0.2 µl primer forward (10 µM)
0.2 µl primer forward (10 µM)
Adjusted to a final volume of 10 µl with sterile H_2O.

HPPR2, *HPR1*

5 µl PerfeCTa SYBR Green Supermix (2x)
4 µl template (cDNA)
0.15 µl primer forward (10 µM)
0.15 µl primer forward (10 µM)
Adjusted to a final volume of 10 µl with sterile H_2O.

The PCR program used was:

Step / **Cycle**	**Denaturation**	**Annealing**	**Elongation**
1st	95 °C, 180 s	-	-
1th-50th	95 °C, 15 s	52 °C, 45 s	68 °C, 60 s, followed by fluorescence measurement
Melting curve 50-95 °C After completion kept at 6 °C			

To be able to make any statement about relative expression rates, it is necessary to calculate the efficiency of amplification. For this purpose, cDNAs were mixed proportionally and serial dilutions (undiluted, 1:4, 1:16, 1:64, and 1:256) were prepared. Those were measured three times for each primer pair in separate reactions. The resulting Cq-values are plotted against the decimal logarithm of the underlying dilution. Assuming optimal amplification, each cycle

leads to the duplication of product. Thus a 1:4 dilution demands two cycles to reach the template amount of the next lower dilution, or:

$$2^n = 4 \qquad n = log_2(4) = \frac{\lg(4)}{\lg(2)} = 2$$

Conversely, this means that the linear regression of a perfect plot with 100% efficiency leads to a slope of 3.32, which corresponds to the number of necessary cycles to increase tenfold. Slope can then be used to calculate efficiency, where the factor 2 corresponds 100%. For this setup, an efficiency of 90-110% or slope of 3.10-3.59 is sufficient.

$$E = 10^{(-\frac{1}{slope})} = 10^{(-\frac{1}{3{,}32})} = 2 \qquad \text{or as a percentage:} \quad E\% = (E-1) * 100 = 100$$

For the main experiment, the cDNAs of the 15 culture days were diluted 1:16. Reactions of all primer pairs for each of the cDNAs were then performed on a 96-well plate. A total of 4 replicates of this scheme were performed.

To check whether a reaction occurs in the absence of the template, no-template controls (NTC) were carried out for the main experiment and efficiency determination, supplemented with an equivalent amount of water instead of template.

2.6 Molecular cloning

Several methods were necessary to further multiply the PCR products of interest (UA ligation, In-Fusion cloning) or to prepare for subsequent recombinant protein expression (restriction site ligation, homologous recombination).

2.6.1 UA ligation

The underlying method of the Qiagen Cloning Kit used for that purpose is based on the ability of unpaired bases to form hydrogen bonds with complementary ones. A DNA ligase links the adjacent ends through the formation of a covalent phosphodiester bond. The kit provides a linearised pDrive vector containing an unpaired uridine nucleotide whereas PCR products amplified with GoTaq polymerase naturally contain adenosine overhangs. With lower success, ligation of blunt ends (when using the proofreading polymerase Phusion) was also possible. The single reaction consists of:

0.5 µl pDrive Cloning Vector
2.5 µl Ligation MasterMix (2x)
2 µl purified PCR product
The samples were incubated at 6 °C for at least 16 h and ligation was stopped at 65 °C for 10 min followed by transformation of *E. coli* EZ.

2.6.2 In-Fusion Cloning

In-Fusion Cloning, a ligation-independent and restriction-free cloning method, was used as part of the SMARTer® RACE 5'/3' Kit. For that, primers extended by specific sequences complementary to the pRACE vector were used. Insert and vector were then cotransformed into *E. coli* EZ. The single reaction consists of:

1 µl linearized pRACE vector
7 µl purified PCR product
2 µl In-Fusion HD Master Mix

The samples were incubated at 50 °C for 15 min and cooled on ice until transformation of *E. coli* EZ (II.2.6.5). RACE PCR products lacking the specific complementary sequences were ligated with the UA-ligation method into pDrive (II.2.6.1).

2.6.3 Homologous recombination

Homologous recombination, another ligation-independent cloning method largely resembles In-Fusion cloning. The method was performed according to Jacobus and Gross (2015). First, the target was amplified using gene-specific primers, each of which was extended with a complementary region to pET-15b (Novagen). The PCR product checked by agarose gel electrophoresis (II.1.4) and bands of the expected size were extracted (II.1.5). The vector was double-digested with NdeI (NEB) and XhoI (Thermo Scientific) and purified (II.1.5), followed by dephosphorylation with Calf Intestinal Alkaline Phosphatase (CIAP). For this, 20 µl of digested pET-15b were mixed with 2.34 µl 10x Buffer for CIAP along with 1 µl CIAP (1 U/µl, Fermentas) and incubated at 37 °C for several hours, followed by inactivation at 65 °C for 15 min. Vector and insert were cotransformed at a twofold insert excess (mass concentration) into 50 µl of *E. coli* EZ (II.1.3).

2.6.4 Ligation into restriction sites

It was the mainly used method to introduce an ORF of interest into the expression vector pET-15b. To provide stable alignment of insert and adjacent vector, both were previously cut with two restriction endonucleases (II.2.7), which allows unidirectional ligation. After digestion, both were checked by agarose gel electrophoresis and were purified (II.1.4 and II.1.5). A three- or five-times excess of insert (volume) was mixed with the linearised pET-15b plasmid followed by addition of water, 0.5 µl 10x T4 DNA Ligase Buffer and 0.5 µl T4 DNA Ligase (5 U/µl, Thermo Scientific) for a total volume of 10 µl. The reaction was incubated at 6 °C for at least 16 h and ligation was stopped at 65 °C for 10 min followed by transformation of *E. coli* EZ .

2.6.5 Heat shock transformation of E. coli

Transformation of chemically competent *E. coli* was used to either further multiply the amplicon and facilitate DNA sequencing (*E. coli* EZ) or to prepare for heterologous protein expression (*E. coli* SoluBL21), i.e. (re-)transformation of plasmid. For the former, if not specified elsewhere, 70 µl of competent *E. coli* EZ was added to the entire ligation assay. After gently mixing and cooling on ice for 30 min, tubes were incubated at 42 °C for 90 sec in a thermomixer and cooled on ice for 5 min. Afterwards 150 µl SOC medium was added and the tubes were incubated at 37 °C for 30 min. The mixture was plated on a LB agar plate containing 100 µg/ml ampicillin, and additionally, allowing for blue-white screening, 50 µM isopropyl β-D-1-thiogalactopyranoside (IPTG) and 80 µg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal). The agar plate was sealed after a short period of drying and incubated at 37 °C overnight (~16 h). For *E. coli* SoluBL21, 70 µl competent cells were added to 1-2 µl plasmid and subsequently treated as just described for *E. coli* EZ with exception of the LB agar plates which contained only 100 µl/ml ampicillin and therefore were not eligible for blue-white screening.

2.6.6 Overnight cultures

After successful transformation and colony formation on LB agar plates, separation and cultivation in individual tubes can be performed. Medium containing X-gal in combination with IPTG allowed for blue-white screening. Accordingly, white colonies indicating the successful ligation of a DNA-fragment into the plasmid can be picked selectively. For transformations

lacking this option, a larger number of colonies were picked for compensation. A tube containing 4 ml LB medium with 100 µg/ml ampicillin was prepared. An imprint of the colony was transferred using a sterile toothpick. The tube was loosely capped and incubated at 220 rpm, 37 °C overnight (~16 h).

2.6.7 Plasmid isolation

1.5 ml cell suspension was transferred to a sterile tube and centrifuged at 3000 *g* for 3 min. Pipetting was performed under laminar air flow conditions if overnight cultures were still needed further (II.2.8). The supernatant was discarded and cell collection was repeated. The isolation of plasmid was carried out using the QIAprep Spin Miniprep Kit (Qiagen). Accordingly, the pellet was resuspended with 250 µl buffer P1 (50 mM Tris/HCl pH 8, 10 mM EDTA, 100 µg/ml RNase A), supplemented with 250 µl buffer P2 (200 mM NaOH, 1% (m/v) SDS) and inverted six times. 350 µl buffer N3 (4.2 M guanidine hydrochloride, 0.9 M potassium acetate, pH 4.8) were added and inverted again, immediately followed by centrifugation at 16000 *g* for 10 min. The supernatant was transferred to a QIAprep Spin column and centrifuged at 16000 *g* for 60 s. Two washing steps, once with 0.5 ml buffer PB (4.69 M guanidine hydrochloride, 30% 2-propanol) and once with 0.75 ml buffer PE (10 mM Tris/HCl pH 7.5, 80% ethanol) and subsequent centrifugation (previous conditions) followed. The column was dried at 16000 *g* for 60 s. The flow-through was removed in each case. The column was placed on a fresh tube and a volume of up to 50 µl sterile H_2O was added. Finally, after incubating at room temperature for 5 min, the tube was centrifuged at 16000 *g* for 60 s. The collected plasmid solution was stored at -20 °C.

2.7 Restriction endonuclease digestion of DNA

Commonly used for confirmation of isolated plasmid, the method is also suitable for targeted digestion of other DNA sequences (e.g. amplicons, vectors in preparation of ligation).

The single reaction for digestion of pDrive plasmid consists of:

3 µl plasmid
1.5 µl EcoRI Buffer or Buffer O (each 10x, Thermo Scientific)
0.5 µl EcoRI (10 U/µl, Thermo Scientific)
Adjusted to a final volume of 15 µl with sterile H_2O.

The single reaction for double digestion of pET-15b plasmid consists of:

3 µl plasmid
1 µl Buffer R (10x, Thermo Scientific)
0.5 µl XhoI (10 U/µl, Thermo Scientific)
0.5 µl NdeI (20 U/µl, NEB)
Adjusted to a final volume of 10 µl with sterile H_2O.

The reaction was incubated at 37 °C overnight (~16 h) if complete digestion was demanded. To qualify for successful ligation, incubation at 37 °C for at least 1.5 h or/and microwaving for up to 20 s (only EcoRI) was sufficient. Digestions were examined via agarose gel electrophoresis (II.1.4)

2.8 Bacterial glycerol stocks

To allow for long-term storage of plasmid or to conserve expression-ready strains (pET-15b in *E. coli* SoluBL21) glycerol stocks were prepared. 425 µl of bacterial overnight culture (II.2.6.6) was thoroughly mixed with 75 µl sterile glycerol, prefrozen with liquid nitrogen and stored at -80 °C. Fresh overnight cultures can easily be generated from those bacterial stocks by inoculation of LB media using sterile toothpicks.

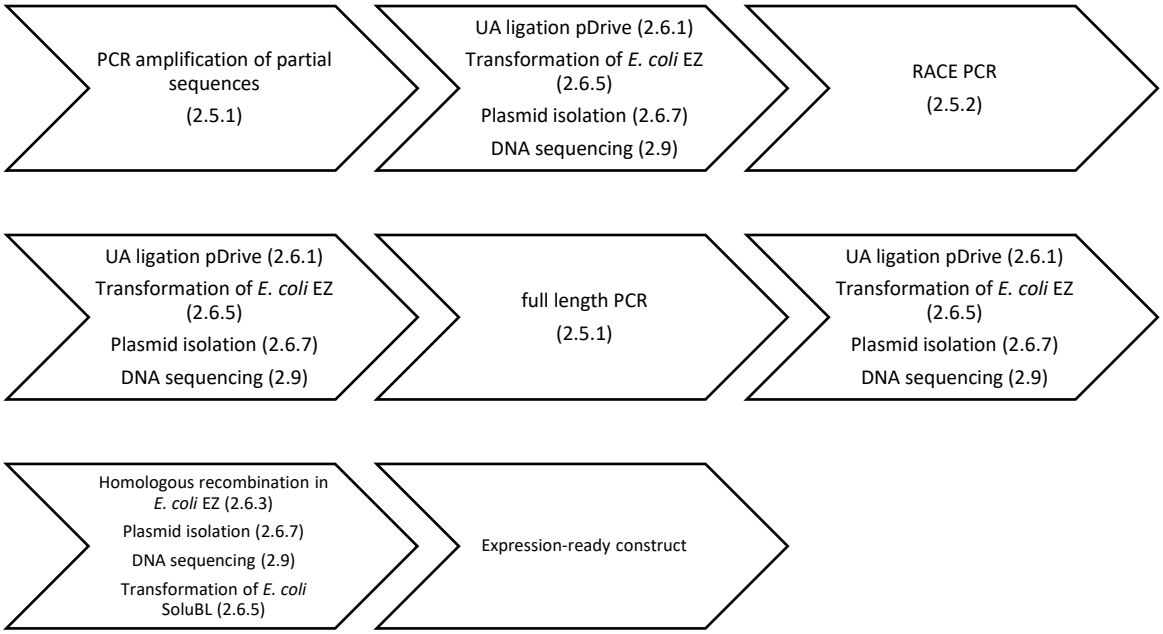
2.9 DNA sequencing

Plasmids, preselected through agarose gel electrophoresis, were measured photometrically and sequenced commercially by Microsynth Seqlab. The plasmids were submitted according to the specifications (plasmid concentration: 40-120 ng/µl), using the provided standard primers for pDrive (M13, M13r) or pET-15b (T7, T7term).

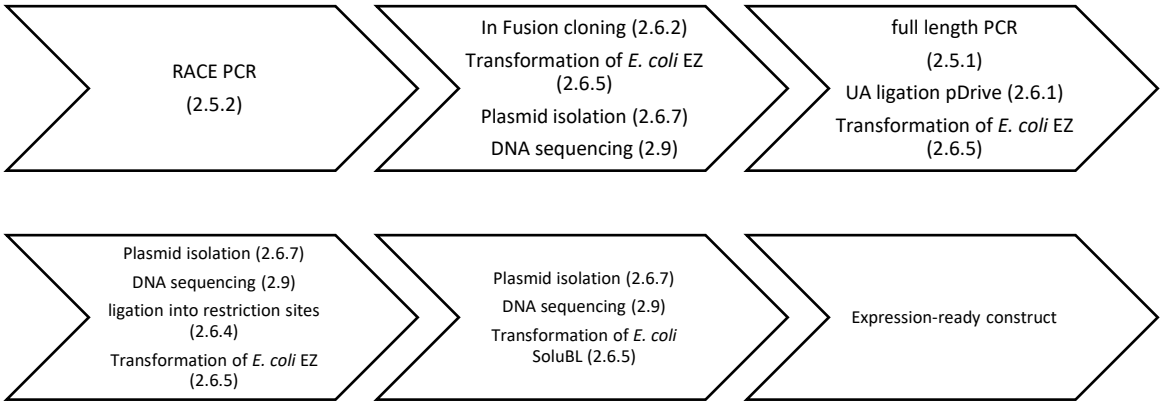
2.10 Flow Chart Molecular Biology

For best understanding of the molecular biology approach, flow charts for all genes are shown in the following.

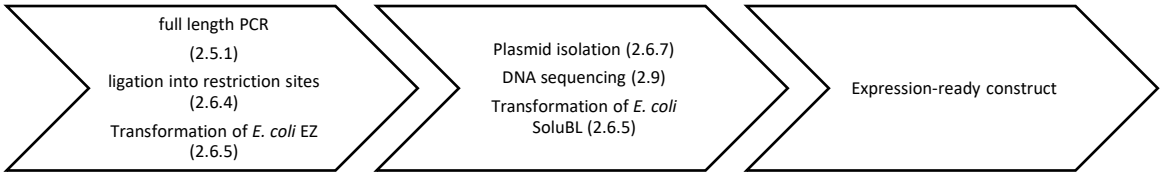
2.10.1 Tyrosine aminotransferase (TAT)



2.10.2 Hydroxyphenylpyruvate reductase (HPPR)



2.10.3 Hydroxyphenylpyruvate reductase 2 (HPPR2) and hydroxypyruvate reductase (HPR1)



3 Protein biochemistry

3.1 Heterologous protein expression

Once *E. coli* SoluBL21 strains containing the qualified target ORF in pET-15b were available, expression was prepared. Overnight cultures were started using the designated strain as described in II.2.6.6 and II.2.8. The next day, up to 2 ml of this culture were used to inoculate a 100 ml LB or TB main culture (250 ml flask or 500 ml baffled flask) containing 100 µg/ml ampicillin. Just before, a reference of said medium was taken. The flask was incubated at 220 rpm (180 rpm for baffled flask), 37 °C. Cell growth was monitored by measuring several times against the medium reference until OD_{600} of 0.4-0.6 was reached. Then, the culture was induced through adjusting to 1 mM IPTG and incubated at 180 rpm, 25 °C for at least 16 h. The cells were pelleted in 50 ml tubes at 3000 *g*, 4 °C for 10 min, the supernatant was removed. The residual pellet was frozen with liquid nitrogen. For cell lysis, 4 ml 50 mM potassium phosphate buffer (KPi) pH 8/ml pellet were added and cells were resuspended. The suspension was incubated on ice for 30 min after addition of ~50 mg lysozyme. The cells were disrupted four times for 30 s on ice through ultrasonication (0.5 cycles, amplitude 100%) with intermittent cooling and were eventually centrifuged at 10000 *g*, 4 °C for 10 min. The cell-free crude extract was collected and usually subjected to purification (II.3.2) immediately, otherwise the crude extract was frozen with liquid nitrogen and stored at -80 °C. *E. coli* SoluBL21 containing the pET-15b plasmid but without insert gene was used as empty vector control and was if necessary subjected to purification as well.

3.2 Purification of His-tagged protein

Since the pET-15b vector already carries a *N*-terminal His-tag sequence (His_6), purification of the respective protein was feasible using metal chelate affinity chromatography with Ni-NTA (nitrilotriacetic acid coupled Ni^{2+} ion, Roth). The crude extract (II.3.1) was adjusted to 10 mM imidazole, 300 mM NaCl and transferred to a column containing 1 ml of Ni-NTA resin which was equilibrated with 3x 1 ml binding buffer (50 mM KPi buffer pH 8, 10 mM imidazole, 300 mM NaCl). The column was sealed and incubated on ice for 1 h. It was then reopened and the flow-through discarded. The column was washed twice with 2 ml wash buffer (50 mM KPi buffer pH 8, 20 mM imidazole, 300 mM NaCl). Eventually, protein was collected by addition of

three times 1 ml elution buffer (50 mM KPi buffer pH 8, 250 mM imidazole, 300 mM NaCl) and kept on ice.

3.3 Buffer exchange and desalting by gel filtration

This step was necessary to remove high imidazole concentrations and impurities while transferring the protein to the designated buffer. The separation is based on size exclusion of the molecules as they flow through the gel matrix Sephadex™ G-25. The PD-10 Desalting column (GE Healthcare) was rinsed with 25 ml H_2O and equilibrated with 25 ml 0.1 M KPi buffer pH 7. 2.5 ml purified protein (II.3.2) was added and the flow-through discarded. The column was placed on a fresh tube and 3.5 ml 0.1 M KPi buffer pH 7 were added, eventually the sample was collected. The protein was kept on ice and the protein concentration was determined according to Bradford (1976) as described below. After completion, aliquots of 50-100 µl were prepared, frozen with liquid nitrogen and stored at -80 °C. The PD-10 columns were rinsed with H_2O and stored at 6 °C for further use.

3.4 Determination of protein concentrations according to Bradford (1976)

The concentration of purified protein was determined based on Bradford (1976). Bradford reagent was prepared with 100 mg/l Coomassie Brilliant Blue G-250, 50 ml/l 95% ethanol, 100 ml 85% *o*-phosphoric acid, filtered twice and stored at 6 °C. For each protein sample of unknown concentration, as well as for the blank reference (buffer) and 1 mg/ml bovine serum albumin (BSA) standard, 1 ml of Bradford reagent was provided in semi microcuvettes. At 30 s intervals, 10 µl of the corresponding sample, i.e. blank, standard, protein sample was added and mixed. The samples were incubated at room temperature for 20 min, then absorption was measured photometrically at 595 nm against the blank reference (buffer in which the protein is stored). For protein designated for kinetics this was performed thrice. Concentration can be calculated using the known BSA-standard. To ensure that concentration is linear to the measured absorption over extensive ranges, increasing BSA concentrations from 0.375 – 3 mg/ml were measured once (n=3).

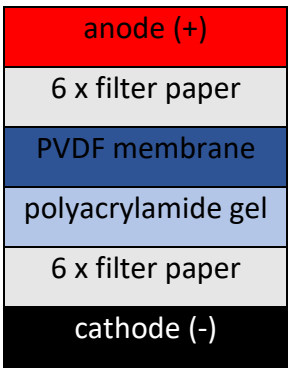
3.5 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)

To verify the purified protein samples for composition (presence of the target protein, impurities), SDS-PAGE was used. It allows discrimination of proteins solely based on their molecular weight through SDS, which masks intrinsic charges of residues and coats the protein with a constant charge to mass ratio. This causes the then negatively charged molecules to migrate toward the positive pole when exposed to an electric field. Depending on the molecular mass, the samples migrate faster or are retained more strongly by the gel polymer. Discontinuous SDS-PAGE was basically performed according to Laemmli (1970). First, the separation gel (per gel: 1.25 ml 1.5 M Tris/HCl pH 8.8, 1.45 ml H_2O, 2.05 ml ROTIPHORESE®Gel 30 (37.5:1), 200 µl 10% (w/v) SDS, 8 µl tetramethylethylenediamine (TEMED), 35 µl 10% (w/v) ammonium persulfate (APS)) was prepared and filled into the empty space between two glass plates included with the apparatus. It was overlaid with 1-butanol or isopropanol for smoothing the surface. After polymerisation, the overlay was carefully removed and the stacking gel (per gel: 625 µl 0.5 M Tris/HCl pH 6.8, 1.40 ml H_2O, 375 µl ROTIPHORESE®Gel 30 (37.5:1), 100 µl 10% (w/v) SDS, 5 µl TEMED, 20 µl 10% (w/v) APS) was prepared and poured on top. The comb was quickly mounted and potentially leaking acrylamide was removed. Meanwhile, the electrophoresis chamber was filled with 1x electrophoresis buffer (for 10x electrophoresis buffer: 250 mM Tris, 1.92 M glycine, 1% (m/v) SDS, pH 8.5). After placement of the polymerised gel, the comb was removed and the pockets were rinsed with buffer several times. The samples were proportionally mixed with 4x Laemmli buffer (2.4 ml 1M Tris/HCl pH 6.8, 4 ml glycerol, 2.5 ml H_2O, 1 ml 2-mercaptoethanol, 0.01% (m/v) bromophenol blue, 8% (m/v) SDS) and heated at 95 °C for 10 min before pipetted into the pockets. 5 µl ROTI®Mark TRICOLOR was added as marker. Eventually, migration was performed at 150 V for 1.5 h or rather until the bromophenol band was to run out at the bottom of the gel. After completion, the stacking gel was clipped and the separation gel was either directed to Western blot (II.3.6) or Coomassie staining. For the latter, it was soaked in a sufficient amount of Coomassie staining solution (0.25% (w/v) Coomassie Brilliant Blue R250, 45% (v/v) methanol, 10% (v/v) acetic acid) until it turned a deep blue. In addition, this can be accelerated by brief heating in the microwave in a covered vessel. The gel was then destained by incubation on a rocking shaker in 45% (v/v) methanol, 10% (v/v) acetic acid; a sponge can be added to speed up the process.

3.6 Western blot

After successful separation of the proteins by size (II.3.5), Western blotting can be used to selectively detect a specific protein. To do so, protein will be quantitively transferred to a membrane and then treated sequentially with two antibodies, the first specifically binding the proteins His_6-tag epitope and the second in turn binding the first antibody. The latter is conjugated with alkaline phosphatase which induces the formation of two insoluble pigments, i.e. dibromodichloroindigo and a formazan derivative, resulting in a purple staining (Sambrook and Russell 2001). Western blot was performed according to Mahmood and Yang (2012) using a polyvinylidene fluoride (PVDF) membrane and semi-dry blotting technique. To avoid carryover and contamination, gloves were worn.

First, the unstained polyacrylamide gel (II.3.5) was equilibrated for 10-30 min in transfer buffer (25 mM Tris, 192 mM glycine, 20% (v/v) methanol, pH 8.3, as described in Towbin et al. (1979)). The transfer membrane (ROTI®PVDF 0.2, Roth) was successively soaked in methanol for 15 s, water for 2 min and transfer buffer for at least 5 min. Besides, at least 12 pieces filter paper of the same size were moistened with transfer buffer. The blot was assembled as follows:



After applying each layer, air patches were removed by applying gentle pressure. Migration was performed with a constant current (2 mA/cm^2 per blot) for 1.5 h.

For the specific detection of proteins (based on www.sysy.com/protocols/westernblot-ap-detection), the membrane was washed three times for 5 min with TBS buffer (10 mM Tris, 0.9% (m/v) NaCl, pH 7.4, Towbin et al. 1979) and was blocked with 5% (m/v) skimmed milk powder in TBS-T buffer (10 mM Tris, 0.9% (m/v) NaCl, 0.05% (m/v) Tween 20, pH 7.4 for at least 1.5 h or overnight. After another two washing steps for 5 min with TBS-T, the membrane

was then incubated upside down on a rocking shaker at 4 °C with the first antibody (mouse anti-His_6-tag antibody, 1:10000 in TBS-T with 1% (m/v) skimmed milk powder) for at least 1 h or overnight. The process was stopped by washing eight times for 5 min with TBS-T. The second antibody (goat anti-mouse IgG, alkaline phosphatase conjugate 1:20000 in TBS-T with 1% (m/v) BSA) was added and incubated at room temperature for at least 1 h. Again, the process was stopped by washing with TBS-T (5 times, 5 min). In preparation for staining, the membrane was equilibrated for 5 min in substrate buffer (100 mM Tris, 100 mM NaCl, 5 mM $MgCl_2$, pH 9.5). Eventually, it was incubated in freshly prepared dye solution with 80 µl 5-bromo-4-chloro-3-indolyl phosphate (BCIP) substrate solution (20 mg/ml BCIP in 100% dimethylformamide) and 60 µl nitro blue tetrazolium (NBT) substrate solution (50 mg/ml NBT in 70% (v/v) dimethylformamide) added to 10 ml substrate buffer until sufficient purple coloration of expected spots has set in, but not longer than 30 min, as the membrane still darkens further on. The dyeing is stopped by washing with H_2O three times and the membrane was dried and photographed.

The blotted polyacrylamide gel can additionally be stained with Coomassie staining solution as described in II.3.5, this can also be used to check whether the transfer was quantitative or incomplete.

4 Determination of enzyme activities

4.1 Tyrosine aminotransferase (TAT)

TAT assays were performed in their entirety as endpoint determinations according to Busch and Petersen (2021). Qualitative studies on TAT enzyme activity were performed through high-performance liquid chromatography (HPLC), optionally coupled with mass spectrometry (LC-MS). The underlying methodology of any quantitative studies analysing pHPP including pH and temperature optima as well as most of kinetics is based on Diamondstone (1966). If phenylpyruvate was the analyte, i.e., for measurements with L-phenylalanine as the amino group donor, measurements were performed according to Ru et al. (2017). Unless otherwise indicated, substrates for enzyme assays were dissolved in water or the related buffer.

4.1.1 Assays for HPLC and LC-MS analyses

To detect the activity of the heterologously expressed enzyme and compare purified enzyme and crude extracts, assays were performed as follows. Each assay contained 4.2 mM L-Tyr, 32 mM 2-oxoglutarate, 0.08 mM pyridoxal 5'-phosphate (PLP), 10 µl purified enzyme or crude extract and 1 M Tris/HCl pH 9 in a total volume of 250 µl. The high buffer concentration was necessary since L-Tyr had to be prepared as a 150 mM solution dissolved in 0.5 M HCl, due to poor solubility in neutral aqueous solutions. Assays were pre-incubated (30 °C) in a thermomixer or water bath and reaction was initiated by addition of enzyme (purified or crude extract) and further incubated at 30 °C. Incubation time was up to 30 min, accordingly the focus here was on qualitative analysis. The reaction was stopped with 100 µl 6 M HCl and extracted with 500 µl ethyl acetate twice. Empty vector control reactions (crude extract and purified) were carried out in the same way. Ethyl acetate phases were combined in each case and evaporated. The residue was redissolved in 100 µl of the HPLC eluent and centrifuged at 16000 *g* for 5 min, 20 µl thereof were analysed by HPLC (4.3.3, a). Where necessary, empty vector controls were prepared by replacing the enzyme/crude extract. To identify the reaction product of L-Tyr, samples were redissolved in 45% (v/v) methanol and were analysed by LC-MS (II.4.3.6, b).

Assays to examine the keto-enol tautomerism and degradation of pHPP as well as its accessibility to chromatographic methods, were performed similar with the difference that instead of HCl, the reaction was occasionally stopped with 50 µl KOH. After an optional rest period of 40 min, the sample was acidified with 100 µl HCl and extracted. The samples were either subjected to HPLC (II.4.3.3, a) or LC-MS (II.4.3.6, b). The pH transition after addition of KOH as well as after HCl addition was controlled by using pH paper (~14-~1).

4.1.2 Conversion of 4-hydroxyphenylpyruvate to 4-hydroxybenzaldehyde

To test different influences on the conversion of 4-hydroxyphenylpyruvate (pHPP) to 4-hydroxybenzaldehyde (HBA), samples with 249 µl 1 M Tris/HCl pH 9, 50 µl KOH were prepared and incubations of 15 min, 30 min and 1 h were started adding 1 µl 10 mM 4-hydroxyphenylpyruvate (pHPP). One series was incubated at room temperature with open lid, the second one at room temperature under vigorous shaking and the third at 45 °C (water bath). The samples were stopped with 100 µl 6 M HCl, references were stopped before adding

pHPP. Each sample was extracted twice with 0.5 ml ethyl acetate and was evaporated. Samples were dissolved in 90% MeOH, 0.01% H_3PO_4 directly before measurement and 20 µl were analysed by HPLC (II.4.3.3, a).

4.1.3 Method development for photometric TAT assays

4.1.3.1 Determination of absorption maximum

The determination of the absorption maximum of HBA was performed through dummy assays. For that, 203 µl 1 M Tris/HCl pH 9, 10 µl 0.1 M KPi pH 7 (enzyme storage buffer), 30 µl H_2O (solvent for PLP and 2-oxoglutarate in real assays), 5 µl 10 mM HBA in 0.5 M HCl, 2 µl 0.5 M HCl and 50 µl 6 N KOH were prepared as well as a corresponding reference in which HBA was exchanged for another 5 µl 0.5 M HCl. This reference was in addition used to dilute samples adequately. The photometric spectrum was recorded against the reference (chapter II.4.3.2).

In addition, background measurements were performed for all kinetic studies to ensure that only the analyte was detected. It is important to emphasize that the measurements must be carried out in a suitable medium that reflects the actual pH value of the test sample, since changes in pH value can have a considerable influence on the measurement. Corresponding adjustments are listed in the respective section.

4.1.3.2 Linear relationship of pHPP, HBA concentrations and absorption

Dummies with 203 µl 1 M Tris/HCl pH 9, 10 µl purified enzyme, 20 µl 400 mM 2-oxoglutarate, 10 µl 2 mM PLP, 50 µl 6 N KOH and increasing HBA or pHPP concentrations in a volume of 7 µl were prepared. After incubation at 45 °C for 30 min, absorption was determined against references lacking HBA at 330 nm using acrylic cuvettes (chapter II.4.3.2). Photometer compensation solution I (II.4.1.4) was used to dilute the samples. This experimental setup was performed in the same way with NaOH as alkaline agent.

4.1.4 Photometric standard method for characterisation of TAT

The underlying principle of the various enzyme assays was based on the combination of amino donor (e.g. L-Tyr), amino acceptor (2-oxoglutarate), the prosthetic group PLP, purified enzyme and a suitable buffer system in a total volume of 250 µl. In general, variants of 1 M Tris/HCl

were used, whose high buffer capacity sufficiently compensated for the influence of some acidic co-substrates and L-Tyr dissolved in 0.5 M HCl. Assays were pre-incubated at the indicated temperature in a thermomixer or water bath, then generally started by addition of the enzyme (a variation would be start by substrate addition) and incubated for the intended interval and stopped by addition of KOH. For quantitative experiments, the incubation period and enzyme dilution was determined by preceding linearity tests so that initial velocities are captured. The determination of PLP dependence and all substrate-saturation kinetics was performed using at least three independent protein expressions to include biological variation. References were stopped before addition of the starter. Standards for the determination of specific activities were performed in the form of dummy tests, i.e. these contained all test components, but instead of the substrate a defined amount of pHPP was supplemented following KOH addition. All samples were then further incubated at 45 °C for 30 min and centrifuged at 16000 *g* for 5 min. Assays, standards as well as references were diluted with a photometer compensation solution adapted to the respective compositions (see below) and measured against the respective reference at 330 nm using acrylic cuvettes (chapter II.4.3.2).

Corresponding modifications or reasonable adjustments are listed in the respective sections.

Compensation solution no.	**Contents**	**Method**
I	1 M Tris/HCl pH 9.0 + H_2O + 6 N KOH (21+4+5)	linear relationship of pHPP pH optimum
II	1 M Tris/HCl pH 8.5 + H_2O + 6 N KOH (22+3+5)	temperature optimum PLP kinetics L-tyrosine kinetics L-phenylalanine kinetics acceptance of L-/D-tyrosine
III	1 M Tris/HCl pH 8.5 + H_2O + 6 N KOH (21+4+5)	oxaloacetate kinetics oxoglutarate kinetics
IV	1 M Tris/HCl pH 7 for pH adjustment	phenylpyruvate kinetics pyruvate

4.1.5 pH optimum

The pH optimum of TAT was determined using three buffer systems. Britton-Robinson buffer (0.04 M phosphoric acid, 0.04 M acetic acid, 0.04 M boric acid, titrated with 0.2 M NaOH in a range of pH 4.0-10.0, Britton and Robinson (1931)), 0.5 M CHES buffer (*N*-cyclohexyl-2-aminomethanesulfonic acid, adjusted with 10 M NaOH to a range of 8.6-10.0) and 0.5 M CAPS buffer (*N*-cyclohexyl-3-aminopropanesulfonic acid, adjusted with 10 M NaOH in a range of 9.7-11.0). Each assay contained 32 mM 2-oxoglutarate, 0.08 mM PLP and purified enzyme. Assays were pre-incubated at 30 °C, then the reaction was initiated by addition of L-Tyr for a final concentration of 4.2 mM and further incubated at 30 °C for 1 min. The remaining conditions are based on II.4.1.4. The actual pH-value was measured for each assay.

4.1.6 Temperature optimum

Assays to determine the temperature optimum in the range of 5.5 °C to 75 °C consisted of 1 M Tris/HCl pH 8.5, 4.2 mM L-Tyr and 32 mM 2-oxoglutarate. Assays were pre-incubated at the respective temperature, then the reaction was initiated by addition of the purified enzyme diluted in an adjusted PLP solution for a final concentration of 0.08 mM (PLP) and further incubated for 1 min. The remaining conditions are based on II.4.1.4. The actual temperature was measured for each assay.

4.1.7 Determination of PLP dependence

To quantify the influence of the prosthetic group PLP on enzyme activity, assays consisting of 1 M Tris-HCl pH 8.5, 80 mM 2-oxoglutarate, 6 mM L-tyrosine and varying PLP-concentrations of up to 1.92 mM were prepared. Solubility was increased by brief heating to 45 °C. Assays were pre-incubated at 40 °C, then the reaction was initiated by addition of the purified enzyme and further incubated at 40 °C for 1 min. The remaining conditions are based on II.4.1.4.

4.1.8 Substrate-saturation kinetics – L-tyrosine

To determine substrate-saturation curves for L-Tyr, assays consisting of 1 M Tris/HCl pH 8.5, 80 mM 2-oxoglutarate and varying L-Tyr concentrations of up to 6 mM were prepared. Solubility was increased by brief heating to 45 °C. Assays were pre-incubated at 40 °C, then

the reaction was initiated by addition of the purified enzyme diluted in an adjusted PLP solution for a final concentration of 0.48 mM (PLP) and further incubated at 40 °C for 2 min. The remaining conditions are based on II.4.1.4.

4.1.9 Substrate-saturation kinetics – L-phenylalanine

To determine substrate-saturation curves for L-Phe, assays consisting of 1 M Tris/HCl pH 8.5, 80 mM 2-oxoglutarate and varying L-Phe concentrations of up to 132 mM were prepared. Assays were pre-incubated at 40 °C, then the reaction was initiated by addition of the purified enzyme diluted in an adjusted PLP solution for a final concentration of 0.48 mM (PLP) and further incubated at 40 °C for 1 min. The remaining conditions are based on II.4.1.4.

4.1.10 Substrate-saturation kinetics – 2-oxoglutarate

To determine substrate-saturation curves for 2-oxoglutarate, assays consisting of 1 M Tris/HCl pH 8.5, 6 mM L-Tyr and 2-oxoglutarate concentrations of up to 120 mM were prepared. Solubility was increased by brief heating to 45 °C. Assays were pre-incubated at 40 °C, then the reaction was initiated by addition of the purified enzyme diluted in an adjusted PLP solution for a final concentration of 0.48 mM (PLP) and further incubated at 40 °C for 1 min. The remaining conditions are based on II.4.1.4.

For direct comparison with phenylpyruvate, 28 mM assays were treated as known, but a pH adjustment with 1 M Tris/HCl pH 7 (II.4.1.12) was applied to the 2-oxoglutarate samples.

4.1.11 Substrate-saturation kinetics – oxaloacetate

To determine substrate-saturation curves for oxaloacetate, assays consisting of 1 M Tris/HCl pH 8.5, 6 mM L-Tyr and oxaloacetate concentrations of up to 120 mM were prepared. Solubility was increased by brief heating to 45 °C. Assays were pre-incubated at 40 °C, then the reaction was initiated by addition of the purified enzyme diluted in an adjusted PLP solution for a final concentration of 0.48 mM (PLP) and further incubated at 40 °C for 2 min. The remaining conditions are based on II.4.1.4.

For direct comparison with pyruvate, 20 mM assays were treated as known, but a pH adjustment with 1 M Tris/HCl pH 7 (II.4.1.13) was applied to the oxaloacetate samples. The oxaloacetate solution was prepared directly before use.

4.1.12 Substrate-saturation kinetics – phenylpyruvate

To determine substrate-saturation curves for phenylpyruvate, assays consisting of 1 M Tris/HCl pH 8.5, 6 mM L-Tyr and phenylpyruvate concentrations of up to 36 mM were prepared. Solubility was increased by brief heating to 45 °C. Assays were pre-incubated at 40 °C, then the reaction was initiated by addition of the purified enzyme diluted in an adjusted PLP solution for a final concentration of 0.48 mM (PLP) and further incubated at 40 °C for 1 min. The remaining conditions are based on II.4.1.4. The pH value of the samples was optimised directly before measurement with 1 M Tris/HCl pH 7. Lowering of the pH value was necessary due to the high absorption of phenylpyruvate in alkaline conditions.

4.1.13 Substrate-saturation kinetics – pyruvate

To determine substrate-saturation curves for pyruvate, assays consisting of 1 M Tris/HCl pH 8.5, 6 mM L-Tyr and pyruvate concentrations of up to 260 mM were prepared. Solubility was increased by brief heating to 45 °C. Assays were pre-incubated at 40 °C, then the reaction was initiated by addition of the purified enzyme diluted in an adjusted PLP solution for a final concentration of 0.48 mM (PLP) and further incubated at 40 °C for 30 s. The remaining conditions are based on II.4.1.4. The pH value of the samples was optimised directly before measurement with 1 M Tris/HCl pH 7.

4.1.14 Acceptance of L-tyrosine and D-tyrosine

To examine a potential substrate specificity for L-Tyr, assays consisting of 1 M Tris/HCl pH 8.5, 80 mM 2-oxoglutarate and 6 mM L-Tyr or D-Tyr, respectively, were prepared. Solubility was increased by brief heating to 45 °C. Assays were pre-incubated at 40 °C, then the reaction was initiated by addition of the purified enzyme diluted in an adjusted PLP solution for a final concentration of 0.48 mM (PLP) and further incubated at 40 °C for 1 min. The remaining conditions are based on II.4.1.4.

4.1.15 Assays for TAT substrate search

The easy-to-implement substrate search was performed by appropriation and further development of the method presented by Matheron and Moore (1973). Assays consisting of 0.1 M KPi pH 8.5, 32 mM 2-oxoglutarate and 2 mM of the substrates to be tested, i.e. L-Tyr,

L-Phe, L-DOPA, L-Asp, L-Trp, L-Ala, L-Ser and prephenate, were prepared (L-Tyr, L-DOPA, L-Asp were dissolved in 0.5 M HCl). For prephenate, an amino acceptor, 2-oxoglutarate was exchanged for L-Glu (in 0.5 M HCl). Assays were pre-incubated at 40 °C, then the reaction was initiated by addition of the purified enzyme diluted in an adjusted PLP solution for a final concentration of 0.48 mM (PLP) and further incubated at 40 °C for up to 45 min. The reaction was stopped by heating at 95 °C for 10 min. To accumulate high amounts of arogenate, the assumed reaction product of prephenate, the reaction time was prolonged to 1 h (cellulose TLC analysis) or up to 20 h (silica TLC analysis). The samples were centrifuged (16000 *g*, 5 min) and further analysed by TLC (II.4.3.1, a, b). By increasing the prephenate concentration to 4 mM and incubating with 16 mM L-glutamate for up to 2 h, direct detection by LC-MS was also possible after OPA derivatisation, without separation by TLC and successive extraction (II.4.3.7).

Assays containing L-Tyr, L-Phe, L-DOPA, L-Asp, L-Trp, L-Ala or L-Ser were additionally subjected to OPA derivatisation and HPLC (II.4.3.3, b).

4.1.16 Identification of the product using oxaloacetate in the reaction

Assays consisting of 0.1 M KPi pH 8.5, 32 mM oxaloacetate and 1 mM or 6 mM L-Tyr (in 0.5 M HCl) were prepared and treated as described in II.4.1.15 with incubation for 2 or 1 h respectively. The stopped and centrifuged assays were analysed by TLC (II.4.3.1, a). If designated for identification via LC-MS, 6 mM L-Tyr was used and the incubation was extended to 90 min and analysed by TLC (II.4.3.1, c).

4.2 Hydroxyphenylpyruvate reductase (HPPR)

HPPR assays were either performed as endpoint determinations mainly analysed by HPLC adapted from Petersen and Alfermann (1988) or photometric interval measurement (substrate-saturation kinetics). LC-MS was used to identify product formation. Unless otherwise indicated, substrates for enzyme assays were dissolved in water or the related buffer.

4.2.1 Assays for HPLC and LC-MS analyses

a) To detect the activity of crude enzyme extracts and compare to empty vector control or to detect activity of purified enzyme, assays were performed as follows. Each assay contained 1 mM 4-hydroxyphenylpyruvate (pHPP, dissolved in 20% ethanol), 2 mM NADPH, 4 mM DTT, 0.04 mM ascorbate and 50 µl crude extract or empty vector control or 10 or 25 µl purified enzyme and 0.1 M KPi pH 7 in a total volume of 250 µl. Assays were pre-incubated (30 °C) in a thermomixer or water bath and the reaction was initiated by addition of enzyme and further incubated at 30 °C. Incubation time was up to 20 min, accordingly the focus here was on qualitative analysis. The reaction was stopped with 25 µl 6 M HCl and extracted twice with 500 µl ethyl acetate. Ethyl acetate phases were combined in each case and evaporated. The residue was redissolved in 150 µl of the HPLC eluent and centrifuged at 16000 *g* for 5 min, 20 µl thereof were analysed by HPLC (II.4.3.3, c).

b) Enzyme assays for pH optimum were performed without ascorbate and DTT addition, as follows. Each assay contained 1 mM pHPP (dissolved in 30% ethanol), 2 mM NADPH, 10 µl purified enzyme and 0.5 M KPi pH 5.0-9.0 (increments of pH 0.5) in a total volume of 250 µl. Assays were pre-incubated (30 °C) in a thermomixer or water bath and the reaction was initiated by addition of enzyme and further incubated at 30 °C for 5 min. The reaction was stopped with 50 µl 6 M HCl and extracted twice with 500 µl ethyl acetate. Actual pH values were measured in a separate batch. Ethyl acetate phases were combined in each case and evaporated. The residue was redissolved in 100 µl of the HPLC eluent and centrifuged at 16000 *g* for 5 min, 20 µl thereof was analysed by HPLC (II.4.3.3, c).

c) Enzyme assays for temperature optimum contained 1 mM pHPP (dissolved in 30% ethanol), 2 mM NADPH, 10 µl purified enzyme and 0.1 M KPi pH 7.0 in a total volume of 250 µl. Assays were pre-incubated at the appropriate temperature (5 to 60 °C, 5 °C increments, plus 70 and 80 °C) in a thermomixer until temperature stability was achieved. The actual temperature was measured in a reference tube with buffer. The reaction was initiated by the addition of enzyme and further incubated for 5 min. The samples were further treated as described in b).

d) Enzyme assays in preparation for stereospecificity verification contained 400 µl 25 mM pHPP (dissolved in 30% ethanol), 400 µl 50 mM NADPH, 200 µl purified enzyme and 0.1 M KPi pH 7.0 in a total volume of 5200 µl. Assays were pre-incubated (30 °C) in a water bath and the reaction was initiated by addition of enzyme and further incubated at 30 °C for 2 h. The

reaction was stopped with 1 ml 6 M HCl and extracted thrice with 5 ml ethyl acetate. The combined ethyl acetate phases were evaporated and the residue was dissolved in 100 µl methanol and subjected to TLC (chapter II.4.3.1, d).

4.2.2 Method development for photometric HPPR assays

4.2.2.1 Relevant spectra

Spectra of relevant substrates or corresponding products, i.e. 4-hydroxyphenylpyruvate (pHPP), 4-hydroxyphenyllactate (pHPL), 3,4-dihydroxyphenylpyruvate (DHPP), 3,4-dihydroxyphenyllactate (DHPL), phenylpyruvate (PP), phenyllactate (PL), 4-hydroxy-3-methoxyphenylpyruvate (4H3MPP) and nicotinamide adenine dinucleotide phosphate (NADPH) were recorded by adding the dissolved compound to 0.1 M KPi pH 7 for a concentration of 0.125 mM (additionally NADPH in 1 mM) and a total volume of 800 µl and measuring at 380 nm against a reference containing just solvent in buffer (chapter II.4.3.2).

4.2.2.2 Protein concentration dependence

To verify whether an approximately linear relationship existed between the change of absorption per time (dA) and the enzyme concentrations used, assays were prepared as follows. Each assay contained 1 mM NADPH and 10, 20 or 40 µl purified enzyme with 0.1 M KPi pH 7 in a total volume of 980 µl. For each enzyme concentration, an equivalent reference with denatured enzyme (95 °C, 10 min, then centrifuged for sedimentation) was prepared. It was pre-incubated at 30 °C for 5 min, then started by adding 20 µl 25 mM pHPP (in 30% ethanol) and incubated for 20 min. An interval measurement at 380 nm was performed throughout the incubation period (chapter II.4.3.2). To test for reactivity apart from heterologously expressed HPPR, a second identical series of measurements was performed using an empty vector control and denatured empty vector control as a reference.

4.2.2.3 Absorption behaviour of NADPH

To verify whether NADPH in the presence of HPPR but without pHPP addition causes a change in absorption, assays with 1 mM NADPH and 20 purified enzyme with 0.1 M KPi pH7 in a total volume of 1000 µl were measured at 380 nm against buffer as reference. A second assay with

1 mM NADPH and 20 purified enzyme with 0.1 M KPi pH7 in a total volume of 1000 µl was measured against an identical reference with denatured enzyme.

4.2.2.4 Determination of NADPH and NADH molar extinction coefficient

The molar extinction coefficient for the co-substrates NADPH and NADH was determined by linear regression of four measurement points. The concentrations 0.10, 0.25, 0.75 and 1 mM were prepared for NADPH and concentrations of 0.10, 0.25, 0.75 and 1.50 mM for NADH with 0.1 M KPi pH 7 in a total volume of 1000 µl. The absorption of the sample was measured against the respective identical reference without co-substrate at 30 °C and 380 nm (chapter II.4.3.2).

4.2.3 Photometrical standard method for characterisation of HPPR

The underlying principle of the various enzyme assays was based on the combination of substrate (pHPP, DHPP, 4H3MPP, PP, pyruvate (P), hydroxypyruvate (β-HP)), the co-substrate NADPH or NADH, purified enzyme and 0.1 M KPi pH 7 in a total volume of 1000 µl. Since this was a live measurement, i.e. photometrical interval measurement, the assays were prepared directly in an acrylic cuvette and pre-incubated at 30 °C in the photometer (with water circuit or Peltier element). In addition to the samples, references were prepared for the corresponding reference channel containing all test components except the primary substrate. The reaction was started by addition of the enzyme and incubated at 30 °C for the intended interval, with simultaneous measurement at 380 nm (chapter II.4.3.2). Slow reactions, with an extensive linear phase, could be started and measured in parallel. For fast reactions, only one test/reference pair was measured per run and very high enzyme dilutions (up to 1:200) were used.

4.2.4 Substrate acceptance and product verification

To examine the substrate acceptance profile of HPPR, assays of several substrates, i.e. pHPP, DHPP, 4H3MPP, PP, P and β-HP consisting of 0.1 M KPi pH 7, 1 mM NADPH and 10-12.5 mM of the substrate were prepared. Assays were pre-incubated at 30 °C. The reaction was initiated by addition of the purified enzyme and further incubated at 30 °C for 5 min. A second comparative assay was performed using empty vector control instead of the purified enzyme. After completion, 500 µl was taken from both the sample with enzyme and the sample with

empty vector, transferred to a fresh reaction tube and stored at -20 °C. The remaining conditions are based on II.4.2.3. After thawing, the stored assays were centrifuged (16000 *g*, 10 min), 450 µl was transferred to a fresh tube and eventually evaporated. The residue was dissolved in 200 µl 5% acetonitrile. 170 µl thereof was extracted twice with 500 µl ethyl acetate after addition of 30 µl 6 M HCl. After evaporation, dissolving in 100 µl 5% acetonitrile and centrifugation at 16000 *g* for 5 min, samples were subjected to LC-MS (4.3.6, c). In addition, 100 µM standards of pHPL, DHPL, 4H3MPL, PL, lactate and glycerate in 5% acetonitrile were analysed. Since no commercial standard for DHPL was available, it was prepared by hydrolysis of rosmarinic acid (chapter II.4.3.8).

4.2.5 Substrate-saturation kinetics – 4-hydroxyphenylpyruvate (pHPP)

To determine substrate-saturation curves for pHPP, assays consisting of 0.1 M KPi pH 7, 1 mM NADPH and varying pHPP concentrations of up to 25 mM were prepared. Assays were pre-incubated at 30 °C. The reaction was initiated by addition of the purified enzyme and further incubated at 30 °C for 20 min. The remaining conditions are based on II.4.2.3.

4.2.6 Substrate-saturation kinetics – hydroxypyruvate (β-HP)

To determine substrate-saturation curves for β-HP, assays consisting of 0.1 M KPi pH 7, 1 mM NADPH and varying β-HP concentrations of up to 6 mM were prepared. Assays were pre-incubated at 30 °C. The reaction was initiated by addition of the purified enzyme and further incubated at 30 °C for 5 min. In contrast to the other measurements, references were prepared with all test components but denatured (95 °C, 10 min) enzyme. The remaining conditions are based on II.4.2.3.

4.2.7 Substrate-saturation kinetics – NADPH

To determine substrate-saturation curves for NADPH, assays consisting of 0.1 M KPi pH 7, 25 mM pHPP and varying NADPH concentrations of up to 1 mM were prepared. Assays were pre-incubated at 30 °C. The reaction was initiated by addition of the purified enzyme and further incubated at 30 °C for 5 min. The remaining conditions are based on II.4.2.3.

4.2.8 Substrate-saturation kinetics – NADH

To determine substrate-saturation curves for NADH, assays consisting of 0.1 M KPi pH 7, 25 mM pHPP and varying NADH concentrations of up to 1.50 mM were prepared. Assays were pre-incubated at 30 °C. The reaction was initiated by addition of the purified enzyme and further incubated at 30 °C for 5 min. The remaining conditions are based on II.4.2.3.

4.2.9 Preliminary HPR1 and HPPR2 assays

To evaluate acceptance of pHPP and β-HP, assays consisting of 0.1 M KPi pH 7, 1 mM NADPH and 1 mM pHPP or β-HP were prepared. To evaluate the favoured co-substrate of AaHPR1, assays consisting of 0.1 M KPi pH 7, 1 mM NADPH or NADH and 0.5 mM β-HP were prepared. Assays were pre-incubated at 30 °C. The reaction was initiated by addition of the purified enzyme and further incubated at 30 °C for 20 min. The remaining conditions are based on II.4.2.3.

4.3 Analytical and chemical methods

4.3.1 Thin layer chromatography (TLC)

a) For qualitative analyses of several TAT substrates as described by Busch and Petersen (2021) and identification of the unknown oxaloacetate product, samples were centrifuged (16000 *g*, 5 min) and 2 µl of each was applied to a silica gel 60 F_{254} TLC plate (Merck). Additionally, references of commonly 20 mmol L-Tyr, L-Phe, L-DOPA, L-Asp, L-Trp, L-Ala, L-Ser were included. To purify the transamination product of prephenate, the full volume was applied along a horizontal line. Plates were developed in 1-butanol/acetic acid/water (4:1:1) according to Matheron and Moore (1973). Then, the plates were sprayed with ninhydrin solution (0.3% (m/v) ninhydrin, 3% acetic acid in 1-butanol) and heated at 90 °C after drying and photographed. One exception was the plates with the applied prephenate assays, those were treated as described below (b).

b) For further analysis of arogenate as described by Busch and Petersen (2021), the TLC plates (cellulose F or silica gel 60 F_{254}, Merck) were first cut into a large and a small segment, then the smaller one was stained and re-attached. The stationary phase was scraped off from the unstained part at the same R_f-value as a spot was visible on the stained part. A second sample

with significantly lower R_f-value but within the TLC-track was scraped off from the unstained part as well, serving as negative control. The adsorbent was extracted with 250 µl H_2O, intermitted by ultrasonication for 5 min, followed by shaking (1500 rpm, 5 min), centrifugation (16000 *g*, 5 min) and filter-sterilisation (sterile 0.2 µm PA, ⌀ 15 mm). It was extracted a second time with 250 µl H_2O, omitting ultrasonication, and the filter was eventually rinsed with 500 µl H_2O. The samples were evaporated in a vacuum centrifuge and the residues were redissolved in up to 40 µl H_2O, centrifuged at 16000 *g* for 5 min and analysed by LC-MS (II.4.3.7, a).

c) To identify the unknown oxaloacetate product as alanine, the amount of two concentrated assays was applied to a TLC plate in a straight line (silica gel 60 F_{254} TLC plate, Merck) as well as 2 µl spots near the right and left edge of the plate. In addition, up to 2 µl 25 mM L-Ala standard was included. Chromatography was performed in 1-butanol/acetic acid/water (40:10:10). The two parts containing the assay spots were cut off and stained with ninhydrin solution (0.3% (m/v) ninhydrin, 3% acetic acid in 1-butanol) after drying and were eventually reattached. The corresponding area on the unstained part was scraped off. In addition, an area with higher R_f-value but within the TLC-track and an area outside the TLC-track as well as the L-Ala standard spot were scraped off. 250 µl water were added to each sample, mixed, and ultrasonicated for 5 min. After incubation at 1500 rpm, 21 °C for 5 min and following centrifugation (16000 *g*, 5 min) the supernatant was filtered (sterile 0.2 µm PA, ⌀ 15 mm). Another 250 µl was added to the material, mixed, centrifuged and passed through the same filter. This was finally rinsed with 500 µl. The samples were evaporated. After dissolving in 40 µl H_2O each, the samples were analysed as described in II.4.3.7.

d) The amount of two redissolved scaled-up HPPR assays was applied to a TLC plate in a straight line (silica gel 60 F_{254} TLC plate, Merck Supelco). In addition, up to 20 µl 10 mM pHPL standard was included. Chromatography was performed in 1-butanol/acetic acid/water (63:10:27). The spot, as indicated by fluorescence quenching, was scraped off, eluted in three runs with a total of 2.5 ml methanol (vigorous shaking, 2 min ultrasonication) and then filtered (PVDF 0.22 µm, Roth). 1 ml aliquots were evaporated and treated as described in II.4.3.4.

e) The product of rosmarinic acid hydrolysis was applied to the TLC plate in a straight line (Alugram Xtra SIL G / UV254, Macherey Nagel). In addition, a sufficient amount of 10 mM DHPL standard was included. Chromatography was performed in 1-butanol/acetic acid/water

(63:10:27) overnight. The spot, as indicated by fluorescence quenching, was scraped off, eluted with 1-2 ml H_2O in 15 fractions and filtered (Büchner funnel with filter paper).

4.3.2 Photometric measurement

Two Specord 200 Plus (Analytik Jena) photometers were used, one temperature-controlled via a water bath, the other via a Peltier element. Acrylic cuvettes were used for photometric measurements in the wavelength range >300 nm. For the range below or for recording spectra, quartz cuvettes were used.

4.3.3 High pressure liquid chromatography (HPLC)

a) Samples were analysed by HPLC using an Equisil ODS RP-18 column (250 x 4 mm with 20 x 4 mm pre column, Dr. Maisch) with 45% methanol, 0.01% H_3PO_4 (1 ml/min) at 283 nm. pHPP, commonly 100 µM, dissolved in eluent or methanol, was used as standard for identification.

b) To visualize glutamate, the reaction product of TAT assays, 50 µl OPA solution (0.067% *o*-phthalaldehyde (OPA), 2.5% methanol, 0.0125% 2-mercaptoethanol and borate buffer pH 8.8 (0.078 M H_3BO_3, 0.03 M $Na_2B_4O_7 \times 10\ H_20$, 0.02 M NaCl)) was added to 25 µl of each sample. After incubation at room temperature for 2 min, 20 µl was analysed by HPLC using a Luna C18(2) column (5 µm, 150 x 4.6 mm with pre column, Phenomenex) with 0.1 M Na_2HPO_4 in 25% methanol pH 6.75 (1.2 ml/min) and detection at 340 nm. The experimental design is based on ESA Application note 70-0160P (Busch and Petersen 2021).

c) Samples were analysed by HPLC using an Equisil ODS RP-18 column (250 x 4 mm with 20 x 4 mm pre column, Dr. Maisch) with 30% methanol, 0.01% H_3PO_4 (1 ml/min) at 280 nm. pHPL, commonly 100 µM, dissolved in eluent or methanol, was used as standard for identification and quantification.

d) To determine rosmarinic acid concentration 20 µl from each of the extracted samples was analysed by HPLC using a Equisil ODS RP-18 column (250 x 4 mm with 20 x 4 mm pre column, Dr. Maisch) with 50% methanol, 0.01% H_3PO_4 (1 ml/min) at 280 nm. Rosmarinic acid, 25 µM and 100 µM, dissolved in eluent or methanol, was used as standard for identification and quantification.

4.3.4 Preparative HPLC

Each sample of the unknown pHPL enantiomer was dissolved in 100 µl eluent and analysed by HPLC using a Nucleodur Sphinx RP column (5 µm, 125 x 4 mm with pre column, Macherey-Nagel) with 30% methanol, 0.01% H3PO4 (1 ml/min) and detection at 280 nm. The pHPL fractions were collected and the combined entity was extracted thrice with 10 ml ethyl acetate. The resulting ten aliquots were evaporated and eventually analysed as described in II.4.3.5.

4.3.5 Chiral HPLC

20 µl of the isolated unknown pHPL enantiomer dissolved in 100 µl isopropanol, 0.1% trifluoroacetic acid (TFA) were analysed by HPLC using a Chiralcel OD column (25 x 4.46 cm with pre column, Daicel) with 92.5% *n*-hexane, 7.5% isopropanol, 0.1% TFA (1 ml/min) and detection at 280 nm. 20 µl of 10 mM DL-pHPL as well as 5 mM L-pHPL were used as references. Additionally, 5 mM L-pHPL were added to the unknown pHPL enantiomer and analysed as described.

4.3.6 Liquid chromatography – mass spectrometry (LC-MS)

LC-MS analyses were thankfully handled by Rixa Kraut, Lena-Christin Ludwig-Radtke and Johanna Schäfer with LC-MS instruments kindly provided by Prof. Dr. Shu-Ming Li.

Analysis was performed using an Agilent 1260 HPLC system with detection in a diode array detector (190-400 nm) and affiliated Bruker Daltonics micrOTOF-Q III MS with ESI source, negative mode. Three LC methods (a, b, c) using 0.1% (v/v) aqueous formic acid (solvent A) and 0.1% (v/v) formic acid in acetonitrile (solvent B) were used, details are shown below:

a) 0–40 min 5% B → 100% (linear gradient), 40–45 min 100% B, 45.10–55 min 5% B, flow rate 0.25 ml/min at 20 °C

b) 0–10 min 5% B → 100% (linear gradient), 10–15 min 100% B, 15.10–20 min 5% B, flow rate 0.50 ml/min at 25 °C

c) 0–30 min 5% B → 100% (linear gradient), 30–35 min 100% B, 35–40 min 5% B, flow rate 0.30 ml/min at 20 °C

a&b) For each run, a sample volume of 5-10 µl was injected and separated using the Multospher 120 RP18 column (250 × 2 mm, 5 µm; CS-Chromatographie Service).

c) For each run, a sample volume of 5-10 µl was injected and separated using the VDspher PUR 100 C18-M-SE (150 × 2 mm, 3 µm; VDS optilab).

4.3.7 LC-MS – OPA derivatisation

To qualitatively identify nascent arogenate, TAT assays with prephenate and L-Glu, either heat inactivated (II.4.1.15) or purified using TLC, were subjected to OPA derivatisation and subsequent analysis by LC-MS. In addition, this method was used to identify putative alanine. For that, 30 µl of each sample was supplemented with 10 µl acetonitrile and derivatised with 10 µl OPA solution (0.27% OPA, 10% methanol, 0.05% 2-mercaptoethanol and borate buffer pH 8.8 (0.078 M H_3BO_3, 0.03 M $Na_2B_4O_7$ × 10 H_20, 0.02 M NaCl)). The prepared samples were quickly analysed by LC-MS using method a (II.4.3.6). Additionally, a solvent reference with 20% (v/v) acetonitrile was analysed. In the case of alanine, a L-Ala standard was derivatised and analysed similiarily.

4.3.8 Rosmarinic acid hydrolysis

20 mg rosmarinic acid and 200 mg Rhozyme HP-150 were dissolved in 1 mM ascorbic acid 0.1 M ammonium acetate. The mixture was flushed with nitrogen for 10 min, sealed with aluminium foil and incubated at 30 °C for 5 h. The reaction was stopped with 1 ml 6 M HCl. 3 ml saturated ammonium sulfate were added, followed by extraction with 4 x 10 ml ethyl acetate, interrupted by centrifugation (5000 *g*, 25 °C, 5 min). The ethyl acetate phases were combined and evaporated in pear-shaped flasks. The residue was dissolved in 1,5 ml ethyl acetate, transferred into reaction tubes and evaporated again. The residue was dissolved in 0.5 ml ethanol and applied onto silica TLC plates (chapter 4.3.1, e). After centrifugation (16000 *g*, 2 min) of each extracted sample, an absorption spectrum (200-400 nm) was recorded in a photometer, matching samples with clear maximum at 280 nm were combined, filtered (PVDF syringe filter 0.45 µm) and freeze-dried. The residue was dissolved in H_2O, the

absorption was measured at 280 nm and used to calculate the concentration. The DHPL samples were adjusted to 10 mM.

$$c = \frac{A}{\varepsilon_{DHPL} \cdot \mathrm{d}}$$

c: molar concentration [mol/l]
A: absorption
ε: molar extinction coefficient [l mol^{-1} cm^{-1}] or [1000 cm^2 mol^{-1}]
d: path length [cm]

4.4 Culture characterisation

4.4.1 Cultivation and harvesting

An *Anthoceros agrestis* culture was propagated for several weeks, pooled at the reference day (day 0), and 5 g of suspension cells were distributed on each of a series of flasks containing 50 ml CBM medium. The flasks were further incubated under permanent daylight conditions on a gyratory shaker (100 rpm) at 25 °C. Cells were harvested by suction filtration on each of the following 14 days and immediately after inoculation. Since fresh weight was low at start, two flasks were needed for day 0, 1, 2 and 3. Fresh weight was determined directly in a pre-weighed Büchner funnel and after transfer to a Petri dish. Samples were then collected for both RNA extraction (II.2.2) and rosmarinic acid determination (II.4.4.3), with the exact weight noted for the latter and were stored at -80 °C after freezing with liquid N_2. The medium was collected and analysed immediately afterwards. After completion of the collection, samples dedicated to rosmarinic acid determination were freeze-dried.

4.4.2 Sugar content, pH value and conductivity of the culture medium

The sugar content of the culture medium was measured using a calibrated hand refractometer. The pH value was measured using a pH electrode at room temperature ($\Delta T \leq 0.5$). For conductivity measurement, the medium was diluted 1:4 with ultrapure water (conductivity < 0.056 µS/cm; $\Delta T \leq 0.8$) and measured with a calibrated conductometer.

4.4.3 Rosmarinic acid extraction from Anthoceros agrestis

Samples were each suspended in 2 ml 70% ethanol and ultrasonicated at 70 °C for 10 min. After vigorous mixing, this procedure was repeated once, followed by centrifugation (3000 *g*, 10 min). From this, 1 ml was taken and stored at -20 °C until measurement (chapter II.4.3.3, d).

5 Bioinformatics

The following table comprises the underlying amino acid sequences of several aminotransferase used for multiple sequence alignment (Clustal Omega, Madeira et al. 2019) and phylogenetics. The term "described by" does not automatically refer to the first mention of the said enzyme or sequence, but to the work which provided the relevant information for inclusion into the comparison. Accordingly, the functional description is, if necessary, strongly broken down and does not reflect the actual research background of the authors. In most cases, an enzymatic-biochemical characterisation or expression profiling took place there, which suggests the specific function. Four sequences are listed as putative since no applicable information was available.

The phylogenetic tree was created with Mega 7 maximum likelihood analysis (Kumar et al. 2016) and reliability was assessed by modelling 1000 bootstrap replicates.

Description	Function/ Description	Acc. No. (NCBI, Uniprot)	Described by
Arabidopsis thaliana	Tyrosine aminotransferase 1 (tyrosine metabolism and degradation)	NP_200208	Wang et al. 2019
Arabidopsis thaliana	Tyrosine aminotransferase 2 (tyrosine metabolism and degradation)	NP_198465	Wang et al. 2019
Scutellaria baicalensis	Tyrosine aminotransferase 1 (RA biosynthesis)	AIV98132	Kim et al. 2014b
Scutellaria baicalensis	Tyrosine aminotransferase 2 (RA biosynthesis)	AIV98133	Kim et al. 2014b
Prunella vulgaris	Tyrosine aminotransferase (RA biosynthesis)	AJW87632	Ru et al. 2017
Salvia milthiorrhiza	Tyrosine aminotransferase (RA biosynthesis)	ABC60050	Huang et al. 2008b Xiao et al. 2011
Perilla frutescens	Tyrosine aminotransferase (RA biosynthesis)	ADO17550	Lu et al. 2013a
Coleus blumei	Tyrosine aminotransferase (RA biosynthesis)	CAD30341	Biastoff 2003

Papaver somniferum	Tyrosine aminotransferase (Alkaloid biosynthesis)	ADC33123	Lee and Facchini 2011
Atropa belladonna	Phenylalanine aminotransferase (tropane alkaloid biosynthesis)	AHN10104	Bedewitz et al. 2014
Rosa Yves Paiget	Phenylalanine aminotransferase (2-phenylethanol biosynthesis)	Supplementary information	Hirata et al. 2012
Physcomitrium patens	Probable aminotransferase TAT2	XP_024361307	putative
Klebsormidium nitens	Tyrosine aminotransferase	GAQ85880	
Marchantia polymorpha subsp. *ruderalis*	Hypothetical protein AXG93_1976s1350	OAE18929	
Selaginella moellendorffii	Probable aminotransferase TAT2	XP_024536669	
Arabidopsis thaliana	Alanine aminotransferase (alanine catabolism, osmotic stress response)	Q9SR86 (At3g08860)	Parthasarathy et al. 2019
Nicotiana tabacum	Histidinol phosphate aminotransferase	CAA70403	El Malki et al. 1998
Arabidopsis thaliana	Methionine:3-IPA aminotransferase (auxin/ethylene biosynthesis, tryptophan homeostasis)	AAP68293 (At1g80360)	Zheng et al. 2013; Pieck et al. 2015
Arabidopsis thaliana	glutamate/aspartate:prephenate aminotransferase	Q9SIE1 (At2g22250)	Graindorge et al. 2010
Arabidopsis thaliana	Tryptophan aminotransferase (auxin biosynthesis)	AAO63403 (At1g70560)	Tao et al. 2008
Oryza sativa	Alanine aminotransferase	BAA77261	Kikuchi et al. 1999
Arabidopsis thaliana	Alanine/glutamate:glyoxylate aminotransferase	AAN62333	Liepman and Olsen 2003
Glycine max	Aspartate aminotransferase (cytoplasmic)	AAC50015	Gebhardt et al. 1998
Arabidopsis thaliana	Aspartate aminotransferase (mitochondrial)	P46643	Wilkie and Warren 1998; Schultz and Coruzzi 1995
Arabidopsis thaliana	Aspartate aminotransferase (cytoplasmic)	P46645	Wilkie and Warren 1998; Schultz and Coruzzi 1995
Arabidopsis thaliana	Aspartate aminotransferase (plastidic)	P46248	Wilkie et al. 1995; Wilkie and Warren 1998

6 Primer list

Partial-length primer

Name	Sequence (5‘-3‘)	Tm [°C]	Target	Comment
Ap6389TAT1f	GAGCACCTGCCGTTCGAG	60.5	TAT	
Ap6389TAT2r	CAAATGAGGATCCATCCCAACC	60.3	TAT	

RACE primer

Name	Sequence (5‘-3‘)	Tm [°C]	Target	Comment
AaTAT_RACE3	TCGATCGCGTGCCTCGCGGCGGAAGG	74.3	TAT	
AaTAT_RACE5	CAGCTCCGCAACCTGCTCGAGGTCCACC	73.9	TAT	
AaHPPR-3RACE	GATTACGCCAAGCTTGGGCCCGTGGTGGACGAGC	67.5	HPPR	
5-AaHPPR-2	GATTACGCCAAGCTTTGCGAGTCTCCACAGTCGCGCTGGCCAC	73.9	HPPR	In-Fusion cloning
5-AaHPPR-3	GATTACGCCAAGCTTCGAGGGGCTTTCCTGCGAAATGCGCG	71.1	HPPR	In-Fusion cloning

Full-length primer

Name	Sequence (5‘-3‘)	Tm [°C]	Target	Comment
AaTAT-VL-f	TATATCATATGGCCACAGTCCTGAACCAGAAGCG	68.0	TAT	NdeI restriction site
AaTAT-VL-r	ATAATCATATGCTAGTTTTGGACGGACGCATGTCGAAGAC	68.1	TAT	NdeI restriction site
P-TAT7	GGTGCCGCGCGGCAGCCATATGGCCACAGTCCTGAACCAGAA	62.4	TAT	Homologous recombination
P-TAT8	GGCTTTGTTAGCAGCCGGATCCTCGAGCTAGTTTTGGACGGACG CATGTC	62.4	TAT	Homologous recombination
AaHPPR-VL-f	ATAATCATATGGCGGCGGTGGAGGGGAAG	67.6	HPPR	NdeI restriction site
AaHPPR-VL-r	TTTCTCGAGCTAGAAATGGAGAGGTGTGACGAGG	64.6	HPPR	XhoI restriction site
AaHPR2-f	ATAATCATATGTGCGCCTGCTTCTACG	58.8	HPPR2	NdeI restriction site
AaHPR2-r	TTTCTCGAGTCACATGTAGTCAGCAGTGAAG	58.4	HPPR2	XhoI restriction site
AaHPR1-f	ATAATCATATGCAGGCACTGAGGTTGTTG	59.8	HPR1	NdeI restriction site
AaHPR1-r	TTTCTCGAGTCAGAGCTTAGCAGACGTTCC	59.8	HPR1	XhoI restriction site

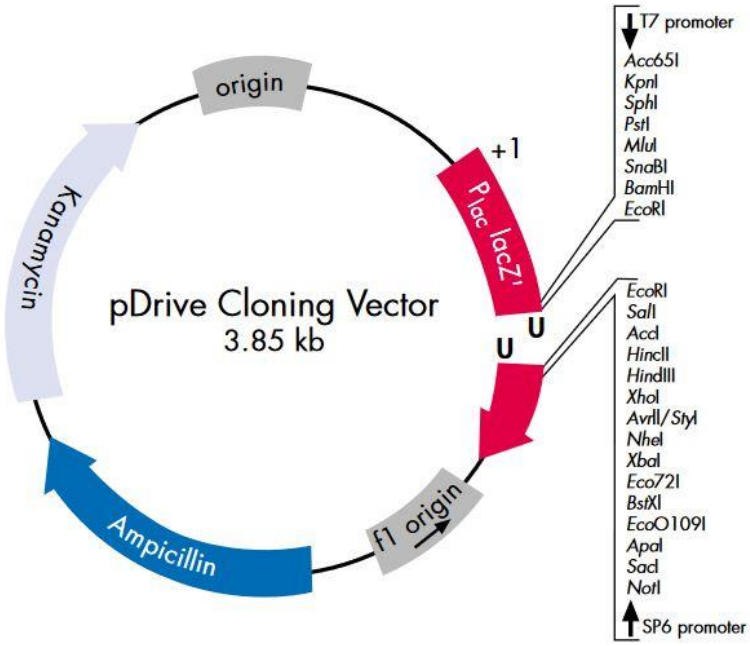
qPCR primer

Name	Sequence (5‘-3‘)	Tm [°C]	Target	Comment
qPCR-P1-Actin-f	TTTTGAGCAGGAACTGGATAC	55.9	Actin	
qPCR-P2-Actin-r	ATACCAGCAGCCTCCATAC	56.7	Actin	

qPCR-P1-StP2a-f	GTGTGTGTCCATCTATCACC	57.3	StP 2a	
qPCR-P2-StP2a-r	AAGACCTCCTCCAACTCC	56.0	StP 2a	
qPCR-P3-TAT-f	GGGACATTGTGGAAACGAAG	57.3	TAT	
qPCR-P4-TAT-r	TCCAAAATCCCAGCCGAGT	56.7	TAT	
qPCR-P5-HPPR-f	CGCGGGTTTGGATGTTTAC	56.7	HPPR	
qPCR-P2-HPPR-r	AATGGAGAGGTGTGACGAG	56.7	HPPR	
qPCR-P3-HPPRv-f	CAAGACTCCCTACACCTAC	56.7	HPPR2	
qPCR-P4-HPPRv-r	CGATAATCTCCCTCCCAAC	56.7	HPPR2	
qPCR-P1-HPR1-f	GCCCAATTCCAACGATGTG	56.7	HPR1	
qPCR-P2-HPR1-r	AGAGCTTAGCAGACGTTCC	56.7	HPR1	

7 Vector maps

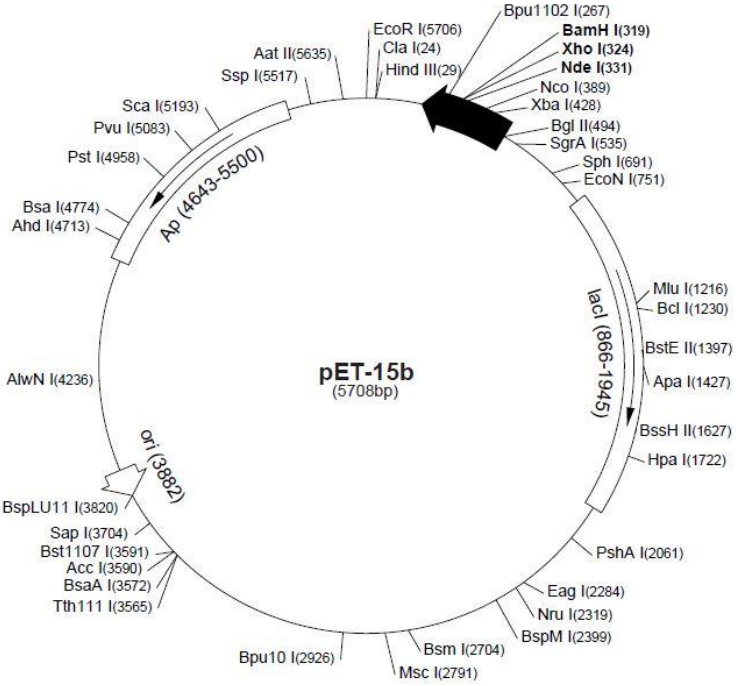
pDrive (Qiagen)

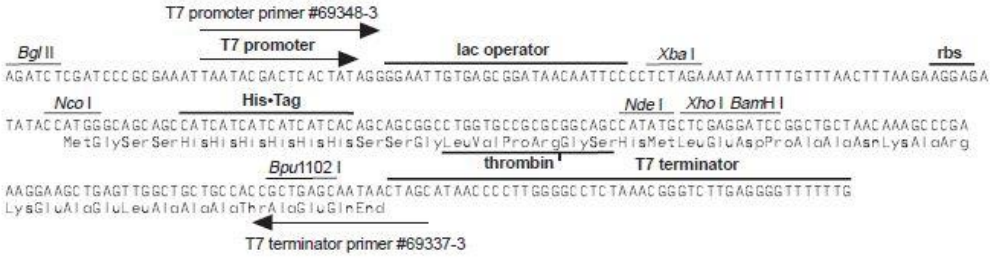



```
M13 reverse primer
binding site                                          T7 promoter
5' GAA ACA GCT ATG ACC ATG ATT ACG CCA AGC TCT AAT ACG ACT CAC TAT AGG GAA   260
3' CTT TGT CGA TAC TGG TAC TAA TGC GGT TCG AGA TTA TGC TGA GTG ATA TCC CTT
             M  T  M  I  T  P  S  S  N  T  T  H  Y  R  E
           LacZ α-petide coding sequence

          Acc65I
             |    KpnI      SphI  PstI|  MluI          SnaBI   BamHI    EcoRI
             |     |          |     |  |               |       |        |
AGC TCG GTA CCA CGC ATG CTG CAG ACG CGT TAC GTA TCG GAT CCA GAA TTC GTG   314
TCG AGC CAT GGT GCG TAC GAC GTC TGC GCA ATG CAT AGC CTA GGT CTT AAG CAC
 S   S   V   P   R   M   L   Q   T   R   Y   V   S   D   P   E   F   V

PCR product
                 SalI
                 |AccI   HindIII          AvrII
                 ||HincII    |     XhoI    |StyI  NheI    XbaI
                 |||         |      |      |      |       |
ATU A TCT GAA TTC GTC GAC AAG CTT CTC GAG CCT AGG CTA GCT CTA GAC CAC   365
TA  UT AGA CTT AAG CAG CTG TTC GAA GAG CTC GGA TCC GAT CGA GAT CTG GTG
        S   E   F   V   D   K   L   L   E   P   R   L   A   L   D   H

Eco72I      EcoO109I      SacI
  |  BstXI      |  ApaI      |  NotI                    <-- SP6 promoter
  |   |         |   |        |   |                      <--
ACG TGT GGG GGC CCG AGC TCG CGG CCG CTG TAT TCT ATA GTG TCA CCT AAA TGG   419
TGC ACA CCC CCG GGC TCG AGC GCC GGC GAC ATA AGA TAT CAC AGT GGA TTT ACC
 T   C   G   G   P   S   S   R   P   L   Y   S   I   V   S   P   K   W

          M13 forward (-20) primer      M13 forward (-40) primer
          binding site                  binding site
          <--                           <--
CCG CAC AAT TCA CTG GCC GTC GTT TTA CAA CGT CGT GAC TGG GAA AAC 3'         467
GGC GTG TTA AGT GAC CGG CAG CAA AAT GTT GCA GCA CTG ACC CTT TTG 5'
 P   H   N   S   L   A   V   V   L   Q   R   R   D   W   E   N
```

pET-15b (Novagen)

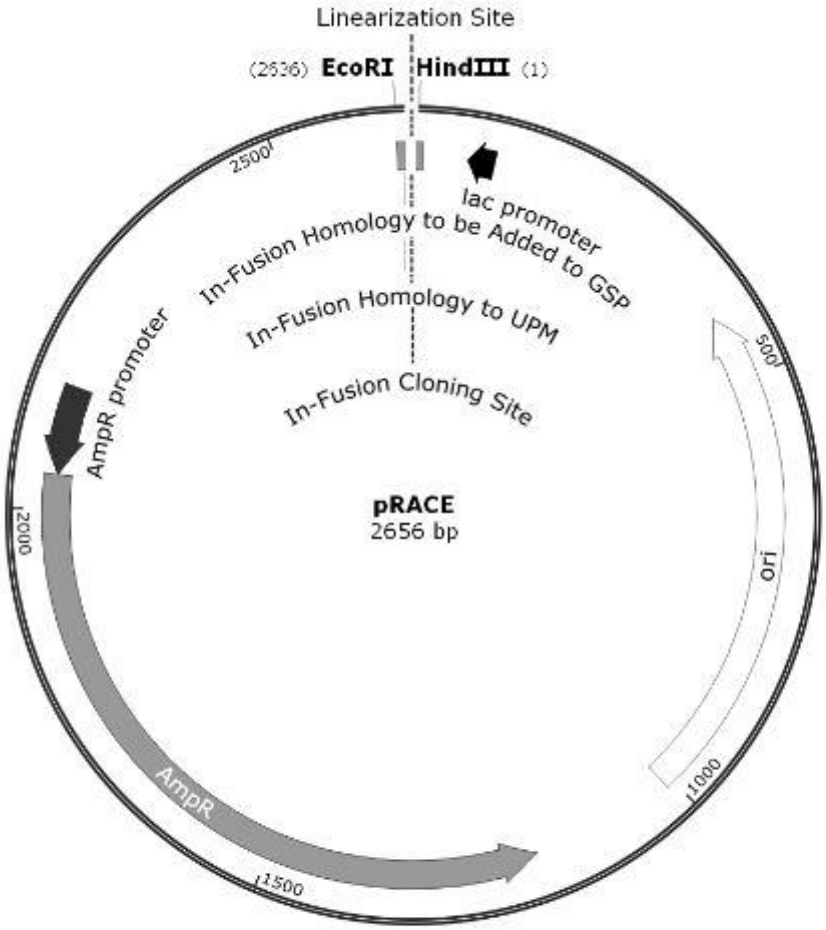
pET-15b sequence landmarks	
T7 promoter	463-479
T7 transcription start	452
His•Tag coding sequence	362-380
Multiple cloning sites	
(*Nde* I - *Bam*H I)	319-335
T7 terminator	213-259
lacI coding sequence	(866-1945)
pBR322 origin	3882
bla coding sequence	4643-5500

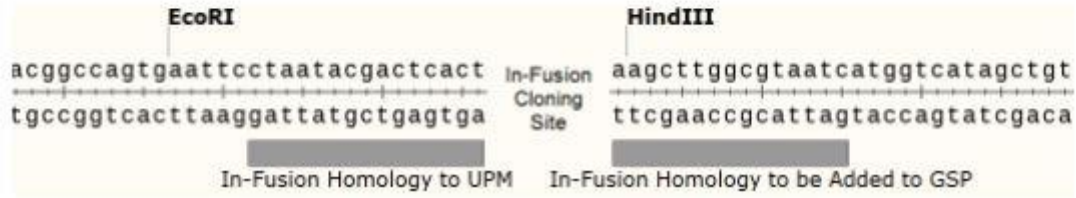




pET-15b cloning/expression region

pRACE (Clontech/Takarabio)





8 Genotypes of bacterial strains

E. coli EZ (Qiagen)

F' Tn10(Tc^r) *lacI^q lacZ* delta-M15 *recA1 end A1 hsdR17 lac glnV44 thi-1 gyrA96 relA1*

Low-copy-number F plasmid [F'::Tn10(Tc^r) *proA$^+$ B$^+$ lacI^q* Z delta-M15]

E. coli SoluBL21 (amsbio)

F$^-$ *ompT hsd*S$_B$ (r_B- m_B-) *gal dcm* (DE3)†

† The SoluBL21 strain contains uncharacterized mutations obtained through special selection criteria. These mutations make the strain able to express insoluble proteins in soluble form, fully or partially, in most tests conducted (ambsio SoluBL21 datasheet, 02.08.2022).

9 List of reagents, enzymes and kits

Product	**Specifications**	**Producer/ Distributor**
Calf Intestine Alkaline Phosphatase (CIAP)	1 U/µl	Fermentas
10X Buffer for CIAP		
DNAse I	RNAse-free, 1 U/µl	Thermo Scientific
10X Reaction buffer for DNAse I	with $MgCl_2$	Thermo Scientific
EcoRI	10 U/µl	Thermo Scientific
10X EcoRI buffer		
GeneRuler™ 1 kb DNA Ladder Mix	0.5 µg/µl	Thermo Scientific
GeneRuler™ DNA Ladder Mix	0.5 µg/µl	Thermo Scientific
Goat anti-mouse secondary antibody	Conjugated alkaline phosphatase	Life Technologies (A16087)
GoTaq® Flexi DNA Polymerase Kit	5 U/µl	Promega
Mouse anti-6xHis-tag monoclonal antibody		Invitrogen (MA1-21315)
NdeI	20 U/µl	New England Biolabs
CutSmart Buffer		
PerfeCTa SYBR Green SuperMix		Quanta Biosciences
NucleoSpin® Extract II Kit		Macherey-Nagel
NucleoSpin® Gel and PCR Clean-up		Macherey-Nagel
Phusion® High-Fidelity DNA Polymerase	2 U/µl	New England Biolabs
Qiagen® PCR Cloning Kit		Qiagen
QIAprep Spin Miniprep Kit		Qiagen
qScript™ cDNA SuperMix		Quanta Biosciences
RevertAid First Strand cDNA Synthesis Kit		Thermo Scientific
RNase A	10 mg/ml	Fermentas
ROTIPHORESE®Gel 30 (37.5:1)		Roth
ROTI®Mark TRICOLOR		Roth
ROTI®Garose-His/Ni Beads		Roth
SMARTer®RACE 5’/3’ Kit		Clontech
T4 DNA Ligase	5 U/µl	Fermentas, Thermo Scientific
XhoI	20 U/µl	Thermo Scientific
10X Buffer R		

10 List of chemicals

Chemical	**Producer/ distributor**
1-butanol	Roth
1-naphthaleneacetic acid (NAA)	Duchefa
2,4-dichlorophenoxyacetic acid (2,4-D)	Duchefa
3,4-dihydroxyphenylpyruvate (DHPP)	Advanced Chemblocks, BLDpharm
3,4-dihydroxyphenyllactate (DHPL)	rosmarinic acid hydrolysis (chapter 4.3.8)
4-hydroxybenzaldehyde (HBA)	Alfa Aesar
4-hydroxy-3-methoxyphenyllactic acid	Sigma
4-hydroxy-3-methoxyphenylpyruvic acid	TCI
DL-4-hydroxyphenyllactic acid (pHPL)	Sigma Aldrich
L-4-hydroxyphenyllactic acid (L-pHPL)	Interchim
4-hydroxyphenylpyruvic acid (pHPP)	Sigma Aldrich, Serva
5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal)	Roth
5-bromo-4-chloro-3-indolyl phosphate (BCIP)	Roth
acetic acid (glacial)	Roth
acetonitrile	Fisher Scientific, Roth
acrylamide 30%/bisacrylamide 0.8% (37.5:1)	Roth
agar	Cero
agarose (NEEO Ultra-Qualität)	Roth
L-alanine	Merck
ammonium acetate	Roth
ammonium persulfate (APS)	Sigma
ammonium sulphate	Roth
ampicillin, sodium salt	Roth
L-ascorbic acid, sodium salt	Roth
L-aspartic acid	Roth, Duchefa Biochemie
boric acid	Roth
bovine serum albumin (BSA)	Roth
bromophenol blue	Merck
buffer solutions pH 4.0, 7.0, 10.0 (calibration of ph-electrode)	Roth, Chemsolute (Th. Geyer)
calcium chloride dihydrate	Sigma-Aldrich
cetyltrimethylammonium bromide (CTAB)	Roth
chloroform	Roth
citric acid	Merck
cobalt(II) chloride, hexahydrate	Merck
Coomassie® Brilliant Blue G250	Serva, Fluka
Coomassie® Brilliant Blue R250	Serva
copper(II) sulphate, pentahydrate	Fluka
dNTPs (dATP, dCTP, dGTP, dTTP)	Fermentas
dimethylformamide	Merck
dipotassium phosphate	Roth

disodium 2-oxoglutarate dihydrate	Merck
disodium ethylenediaminetetraacetate dihydrate (disodium EDTA)	Roth
dithiothreitol (DTT)	biomol
dNTPs (dATP, dCTP, dGTP, dTTP)	Fermentas
L-DOPA	Sigma
ethylenediaminetetraacetic acid (EDTA)	Merck
ethanol	Roth
ethidium bromide	AppliChem
ethidium bromide solution (0.025%)	Roth
ethyl acetate	Roth, Labochem International
L-glutamic acid	Serva
D-(+)-glucose	Roth
glycerol	Roth
glycine	Merck
guanidine hydrochloride solution 6.7 M	Roth
guanidine hydrochloride solution 8 M	PanReac AppliChem
guanidine thiocyanate	Roth
H_2O for injection	Serumwerk Bernburg
hydrochloric acid (37%)	Roth
hydrochloric acid (32%)	Fisher Chemical
imidazole	Roth
indole-3-acetic acid (IAA)	Duchefa
iron(II) sulphate heptahydrate	Fluka
isopropanol	Roth
isopropyl β-D-1-thiogalactopyranoside (IPTG)	Roth
kinetin	Duchefa
lithium β-hydroxypyruvate hydrate	Sigma-Aldrich
lysozyme	Roth
magnesium sulphate heptahydrate	Merck
magnesium chloride hexahydrate	Roth
manganese(II) sulphate monohydrate	Duchefa
β-mercaptoethanol	Merck
methanol	Fisher Scientific
milk powder	Spinnrad
myo-inositol	Roth
NADH, disodium salt	Roth
NADPH, tetrasodium salt	Roth, PanReac Applichem
N-cyclohexyl-2-aminomethanesulfonic acid (CHES)	Roth
N-cyclohexyl-3-aminopropanesulfonic acid (CAPS)	Roth
nicotinic acid	Duchefa
ninhydrin	Merck
nitrogen, liquid	Linde
nitro blue tetrazolium (NBT)	Roth

N-lauroyl sarcosine	Sigma
N-Z-Amine® A	Fluka Analytical
oxaloacetate	Fluka
phenol (citrate buffer saturated)	Sigma Aldrich
L-β-phenyllactate	Sigma
L-phenylalanine	Roth
o-phthalaldehyde (OPA)	TCI
o-phosphoric acid (85%)	Roth
polysorbate 20 (Tween 20)	Serva
polyvinylpyrrolidone (PVP)	Sigma
potassium acetate	Chemsolute
potassium chloride	Roth
potassium dihydrogen phosphate	Roth
potassium hydroxide	Merck
potassium iodide	Merck
potassium nitrate	Roth
pyridoxal phosphate (PLP), monohydrate	Roth
pyridoxine hydrochloride	Duchefa
prephenic acid, barium salt	Sigma Aldrich
Rhozyme HP-150	Pollock and Pool
rosmarinic acid	Isolated from *Melissa officinalis*
L-serine	Roth
sodium acetate trihydrate	Merck
sodium chloride	Roth
sodium citrate	Roth
sodium dihydrogen phosphate dihydrate	Roth
sodium dodecyl sulphate (SDS)	Roth
sodium hydroxide	Merck, Roth
sodium molybdate dihydrate	Fluka
sodium phenylpyruvate	Fluka, Roth
sodium pyruvate	Fluka
sodium tetraborate decahydrate	Roth
sucrose	food quality
tetracycline hydrochloride	Sigma
tetramethylethylenediamine (TEMED)	Roth
thiamine hydrochloride	Roth
trifluoroacetic acid (TFA)	Roth
tris(hydroxymethyl)aminomethane (TRIS)	Roth
tryptone/peptone ex casein	Roth
L-tyrosine	Merck
D-tyrosine	Roth
tris(hydroxymethyl)aminomethane (TRIS)	Roth
L-tryptophan	Roth
tryptone/peptone	Roth
xylene cyanol	Fluka
yeast extract	Roth

zinc sulphate heptahydrate	Merck

11 List of common and analytical consumables

Product	**specifications**	**Producer/ distributor**
cuvettes, acrylic	semi-micro	Sarstedt
cuvettes, polystyrene	micro	Sarstedt
PD-10 columns Sephadex G-25M		GE Healthcare
reaction tubes		Sarstedt
Rotilabo®-syringe filter, PA, sterile	pore size 0.20 µm, ⌀ 15 mm	Roth
Rotilabo®-syringe filter, PVDF, sterile	pore size 0.22 µm, ⌀ 33 mm	Roth
pipette tips		Sarstedt
TLC Cellulose F	20 x 20 cm	Merck
TLC Silica gel 60 F_{254}	20 x 20 cm	Merck, Supelco
Transfer membrane ROTI®PVDF 0.2	300 x 30 cm	Roth

12 List of instruments

Type	**Instrument**	**Manufacturer/ Distributor**
-80 °C deep freezer	C585 innova	New Brunswick Scientific
autoclaves	AL02-02-100	Advantage-Lab
	VX-95	Systec
	VX-150	Systec
blotting chamber	Semi-Dry-Blotter PROfessional	Roth
Bunsen burner	Flammy S	Schütt
centrifuge	Heraeus Fresco 17	Thermo Scientific
	Heraeus Pico 17	Thermo Scientific
	Centrifuge 5415 D	Eppendorf
	3-30KS	Sigma
	Micro Centrifuge SD	Roth
conductometer	Eutech Expert CTS	Thermo Scientific
freeze drier	Alpha 1-2 LDplus	Christ
gel documentation system	FAS-Digi	Nippon Genetics
	Doc-Print VX2 (110-26M)	Vilber
gyratory shaker	Certomat® SII	B. Braun Biotech International
	Ecotron	Infors HT
	10X 400	Gallenkamp

electrophoresis chamber	multiSUB Midi	Cleaver Scientific
	Agagel Mini Biometra	Biome. Analytik GmbH
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
	L-6200 Intelligent Pump	Merck-Hitachi
	655A-40 Auto Sampler	Merck-Hitachi
	SpectraSystem SCM1000	Thermo Electron Corporation
	SpectraSystem UV1000	Thermo Electron Corporation
	SpectraSystem P4000	Thermo Electron Corporation
HPLC columns	Equisil ODS RP-18 column	Dr. Maisch
	Chiralcel OD	Daicel
ice machine	AF 80	Scotsman
	RF-0244A	Manitowoc
incubator	Jouan EB280	Jouan
laminar air flow	Scanlaf Mars	Labogene
	Laminar Flow Workstation	Microflow
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
	VDSpher PUR 100 C18-M-SE	VDS optilab
magnetic stirrer with hotplate	MR 3001	Heidolph
mixer	Vortex-Genie2	Scientific Industries
	Vortex Mixer VM-300	neolab
oven	U 40	Memmert
pH meter	FiveEasy + electrode	Mettler Toledo
	Accumet Basic	Fisher Scientific
	Glass body, combination pH electrode	Hanna instruments
photometer	UV-mini-1240	Shimadzu
	Specord 200 PLUS (blue)	Analytik Jena
	Specord 200 PLUS (red)	Analytik Jena
pipettes	Pipetman 20, 200, 1 ml	Gilson
protein electrophoresis chamber	Mini-Protean 3 Cell	Bio-Rad
refractometer	Optronic HR18	A. Krüss
rocking platform	Rocker 35 EZ	Labnet International

	Duomax 1030	Heidolph Instruments
rotary evaporator	Rotavapor RE 120	Büchi
scale	EG300-3M	Kern
	440-35A	Kern
	440-47	Kern
	Explorer EX225D	Ohaus
	H64	Mettler
thermocycler	Mastercycler gradient	Eppendorf
	MyCycler	BioRad
	PikoReal 96	Thermo Scientific
thermometer	Checktemp 1	Hanna
	Amadigit ad 14 th	Amarell
thermomixer	uniTHERMIX 2	LLG Labware
	Thermomixer comfort	eppendorf
ultrapure water	OmniaPure	stakpure
ultrasonic bath	Sonorex Super RK 510 H	Bandelin
ultrasonic processor	UP 200 s	dr.hielscher GmbH
UV lamp	HL-6-KM	Bachofer
vacuum centrifuge	RVC 2-18 CDplus	Christ
vacuum pump	MZ 2C NT	Vacuubrand
	Drehschieberpumpe P 4 Z	Ilmvac GmbH
Voltage controller	EV3020	Consort
	EV2310	Consort
	E835	Consort
water bath	E-PA/KUE	Julabo Labortechnik GmbH
	Thermomix ME	B. Braun Biotech International
	B 465	Büchi

13 Software and web applications

Software/Application	**Manufacturer/ Distributor**
AspectUV 1.5	Analytik Jena
Blastn	NCBI
Blastp	NCBI
Bruker Data Analysis	Bruker Corporation
ChemDraw 20	PerkinElmer Informatics
CLC Sequence Viewer 8	CLC bio (Qiagen)
Conserved Domain Research (CDD)	NCBI
EMBOSS Needle	EMBL-EBI
GraphPad Prism 5	GraphPad Software
HPLC System Manager 4.0	Hitachi
Inkscape 1.1	Inkscape community, GNU GPL

Mega 7	Pennsylvania State University
Office 365 (Word, Excel), Office 2019	Microsoft
PikoReal Software 2.2.250.602	Thermo Scientific
Universal Protein Resource (UniProt)	UniProt consortium
WinASPECT PLUS 4.2.9.0	Analytik Jena
www.hornworts.uzh.ch/en.html Anthoceros genomes, blast	University of Zurich

III Results

1 Tyrosine aminotransferase (TAT)

1.1 RNA extraction and cDNA synthesis

RNA from *Anthoceros agrestis* suspension-cultured cells was extracted using the acid guanidinium thiocyanate-phenol–chloroform extraction according to Chomczynski and Sacchi (1987). After reconstitution of dried RNA with water, integrity was checked through agarose gel electrophoresis and concentrations at 260 nm as well as the 260/280 ratio were measured photometrically. Fig. 6 and Table 1 show the results of an exemplary extraction of *Anthoceros agrestis* cells after a four-day growth phase in CBM medium. Samples 1 and 2 showed higher RNA concentrations and more distinct ribosomal subunit bands on the agarose gel, i.e. distinct 28S and 18S and blurry 5S rRNA, the latter missing in samples 3 and 4. The 260/280 ratio was comparable for all four samples. Ideally, it should be ~2.0, indicating pure RNA. Since sample 3 and 4 showed unexpectedly low concentrations and it cannot be ruled out that the presence of RNases led to degradation of RNA, samples 1 and 2 were used for cDNA synthesis. RACE-ready cDNA was used from chapter III.2.1.

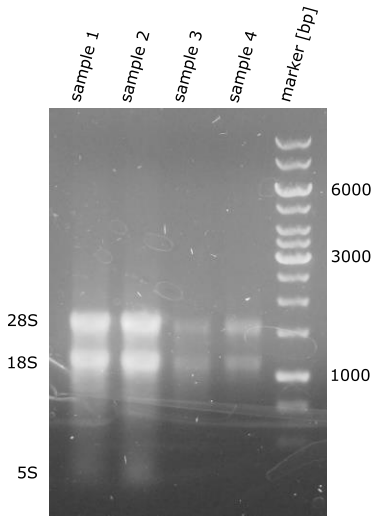


Figure 6 Extracted RNA samples from *Anthoceros agrestis*. Cells were extracted from a four day old suspension cell culture (CBM). Agarose gel electrophoresis with ethidium bromide staining**.** Distinct ribosomal subunit bands indicate RNA integrity. The significantly lower RNA concentration of samples 3 and 4 is clearly visible, the 5 S rRNA band is missing.

Table 1 Concentration and purity of *Anthoceros agrestis* RNA extraction

RNA sample	Concentration [ng/µl]	260/280 ratio
Sample 1	887.8	1.65
Sample 2	957.1	1.68
Sample 3	252.5	1.58
Sample 4	161.5	1.67

1.2 Amplification of cDNA encoding TAT from *Anthoceros agrestis*

Primers targeting an internal sequence of putative AaTAT were designed based on conserved regions obtained through sequence alignments of *Plectranthus scutellarioides* TAT (later referred to syn. *Coleus blumei*, acc. no. AJ458993), *A. punctatus* (scaffold Ap6389, personal communication with Steven Kelly, Department of Plant Sciences, University of Oxford) and *A. agrestis* (Aa34867; Szövényi et al. (2015), personal communication). The gradient PCR amplification of Aa cDNA revealed an internal fragment of 508 bp (Fig. 7 a, appendix 1.1) which enabled new primers to be derived. 5'- and 3'-RACE PCR (Fig. 7 b) was used to amplify the missing upstream and downstream cDNA ends. This revealed an open reading frame of 1395 bp encoding 464 aa residues, where due to a second ATG codon in the more direct environment, a shorter variant of the CDS with 1302 bp or 433 aa would be conceivable (appendix 1.2). The 1395 bp sequence was successfully targeted using full-length gradient PCR (Fig. 7 c) and was deposited in Genbank (appendix 1.3, accession number MN922307; Busch and Petersen 2021).

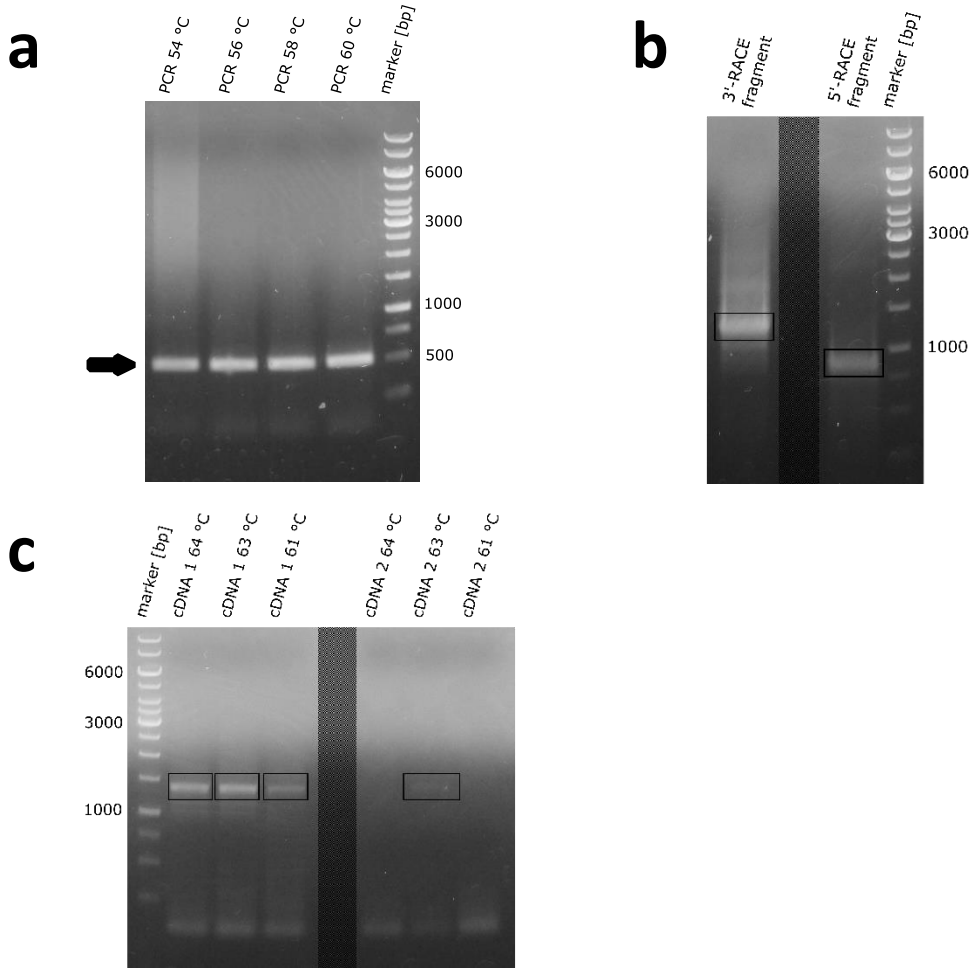


Figure 7 PCR Amplification of *AaTAT*. a Partial gradient PCR **b** 3'- and 5'-RACE PCR **c** Full-length gradient PCR targeting 1395 bp CDS using two Aa cDNAs; agarose gels stained with ethidium bromide.

The second start codon is retrievable in the databases of *Anthoceros agrestis* Oxford strain and in *Anthoceros punctatus*, but not in the sequence of the *Anthoceros agrestis* Bonn strain (Szövényi et al. 2015; https://www.hornworts.uzh.ch/en.html). It could also be found in the draft genome of *Anthoceros angustus* (Zhang et al. 2020). The match with the Oxford strain was highest at 100% (464/464 aa), compared to the Bonn strain with 99.8% (463/464 aa) and *A. punctatus* with 98.9% (459/464 aa). One peculiarity, however, was that the ORF of both *Anthoceros agrestis* sequences featured additional 83 *N*-terminal amino acids in comparison to AaTAT. Only a fragment was retrievable in *Anthoceros punctatus* but limited to genomic data. Part of this sequence was accurately recovered in the 5' RACE product of the AaTAT, but there was, even with consideration of several individual sequencings, hardly any agreement in the region of the additional start codon. To obtain further information about this, the draft genome of *Anthoceros angustus* (Zhang et al. 2020) was searched. Here, too, the section could

be found. Again, not in the mRNA information but in the general scaffold information. In order to be able to make a qualitative statement about the actual CDS, the surrounding areas of the potential start codons were examined for Kozak motifs. Sequence information from *Anthoceros agrestis* Bonn (Szövényi et al. 2015; AagrOXF_evm.model.utg000017l.62.1 cDNA scaffold; https://www.hornworts.uzh.ch/en.html) was examined after reconciling with sequencing data from RACE-PCR in the region of the potential three *N*-termini MPLLH (largest possible CDS; ATAT|ATG|CC), MATVL (amplified CDS; TGAA|ATG|GC) and MSKAV (short CDS; GATT|ATG|TC). The criteria are given in Table 2. The general consensus for green plants (Hernández et al. 2019) is only fulfilled by MATVL, which, except for T at position -4 and A at position 3 respectively, also fulfils the criteria for non-angiosperm plants and early diverged (Joshi et al. 1997). In addition, the occurrence of the universally absent U/C in -3 and U/G at -2 as well as the missing G conserved in all green plants in +4 position also speaks against the competing initiation points (Hernández et al. 2019).

Table 2 Criteria for evaluating the translational origin. Consensus is given for green plants, non-angiosperms, comprising bryophytes, pteridophytes and gymnosperms and early diverged plants, mostly green algae.

Group	**literature**	**Sequence / Position**								
		-4	**-3**	**-2**	**-1**	**1**	**2**	**3**	**4**	**5**
Green Plants	Hernández et al. 2019		A/G	A/C		A	U	G	G	
Non-angiosperms	Joshi et al. 1997	a	A/G	a	a/c	A	U	G	G	C
Lower plants		A/C	A	A/C	a	A	U	G	G	C

Blastp analysis of the translated aa sequence showed overall percent identities ranging from 45% to a high of 63% (100 sequences output), with the highest identities in comparison to the mosses *Physcomitrium patens, Ceratodon purpureus*, *Sphagnum fallax* and the liverwort *Marchantia polymorpha*, although among these only the sequence of the model organism *Physcomitrium patens* is annotated as TAT. The sequences from the lycophyte model organism *Selaginella moellendorffii*, the green alga *Klebsormidium nitens* as well as the useful plants

Salvia miltiorrhiza, *Camellia sinensis*, *Vitis vinifera* and *Gossypium arboreum* then consistently show annotation as TAT (Table 3).

Table 3 Blastp results (16/03/2022). Comparison of AaTAT aa sequence with several species of interest.

Organism	Description	Accession	Identity	E value
Physcomitrium patens	probable aminotransferase TAT2	XP_024361307	62.20%	0.0
Ceratodon purpureus	hypothetical protein M758_8G057100	KAG0607815	62.53%	0.0
Sphagnum fallax	hypothetical protein BDL97_04G130600	KAH8965713	59.51%	0.0
Marchantia polymorpha subsp. ruderalis	hypothetical protein AXG93_1976s1350	OAE18929	52.70%	7E-158
Selaginella moellendorffii	probable aminotransferase TAT2	XP_024536669	51.07%	3E-151
Klebsormidium nitens	tyrosine aminotransferase	GAQ85880	46.32%	3E-130
Salvia miltiorrhiza	tyrosine aminotransferase	ABC60050	45.97%	2E-126
Camellia sinensis	tyrosine aminotransferase-like	XP_028058388	45.60%	2E-125
Vitis vinifera	putative aminotransferase TAT2	RVW14048	44.95%	1E-124
Gossypium arboreum	predicted: tyrosine aminotransferase-like	XP_017632058	45.33%	3E-126

The Conserved Domain Search (CDD) revealed the expected enzyme properties, such as the putative homodimer interface, PLP binding site and catalytic Lys296 residue. Additionally, highly conserved amino acid residues essential for aminotransferase activity were found. As Arg435, which fixes the substrate's deprotonated α-carboxy group of amino acid or 2-oxoacid and Asp263, which stabilises protonation of nitrogen through a hydrogen bond to PLP's N1 (Kirsch et al. (1984), Mehta et al. (1989), cf. Busch and Petersen 2021). The conserved regions and invariant amino acids are visualised in Figure 8 (Busch and Petersen 2021), a sequence alignment of various plant tyrosine aminotransferases with research focusing on RA biosynthesis in comparison to AaTAT. The discussed motifs and conserved amino acids are retrievable in all representatives. Additionally, Pairwise Sequence Alignment (EMBOSS Needle, protein) states a 36.7% identity and 53.3% similarity (Busch and Petersen 2021) of

AaTAT with *Coleus blumei* TAT (CAD30341) as well as a 37.2% identity and 54.8% similarity with *Prunella vulgaris* TAT (AJW87632), both enzymes already enzymatically characterised with proven TAT activity (Biastoff 2003; Ru et al. 2017).

The presence of the stated conserved amino acids does not allow a definitive statement about the subgroup of an aminotransferase. The promiscuous character of these enzymes has already been introduced and the mentioned mandatory amino acids are invariant in aminotransferases of subgroups I-IV (Mehta et al. 1993). Cytosolic pig aspartate aminotransferase for instance, which can be seen as the blueprint for this enzyme family, shows the characteristic Lys296 (258), Arg435 (368) and Asp263 (222) amino acids (Ovchinnikov et al. 1973; Mehta et al. 1989). The numbers in brackets are according to the aa sequence Ovchinnikov et al. suggested. Furthermore, CCD of this pig aspartate aminotransferase (AAA53531; Nagashima et al. 1989) also reveals the same pattern of the known conserved regions for homodimer interface and PLP bindings site.

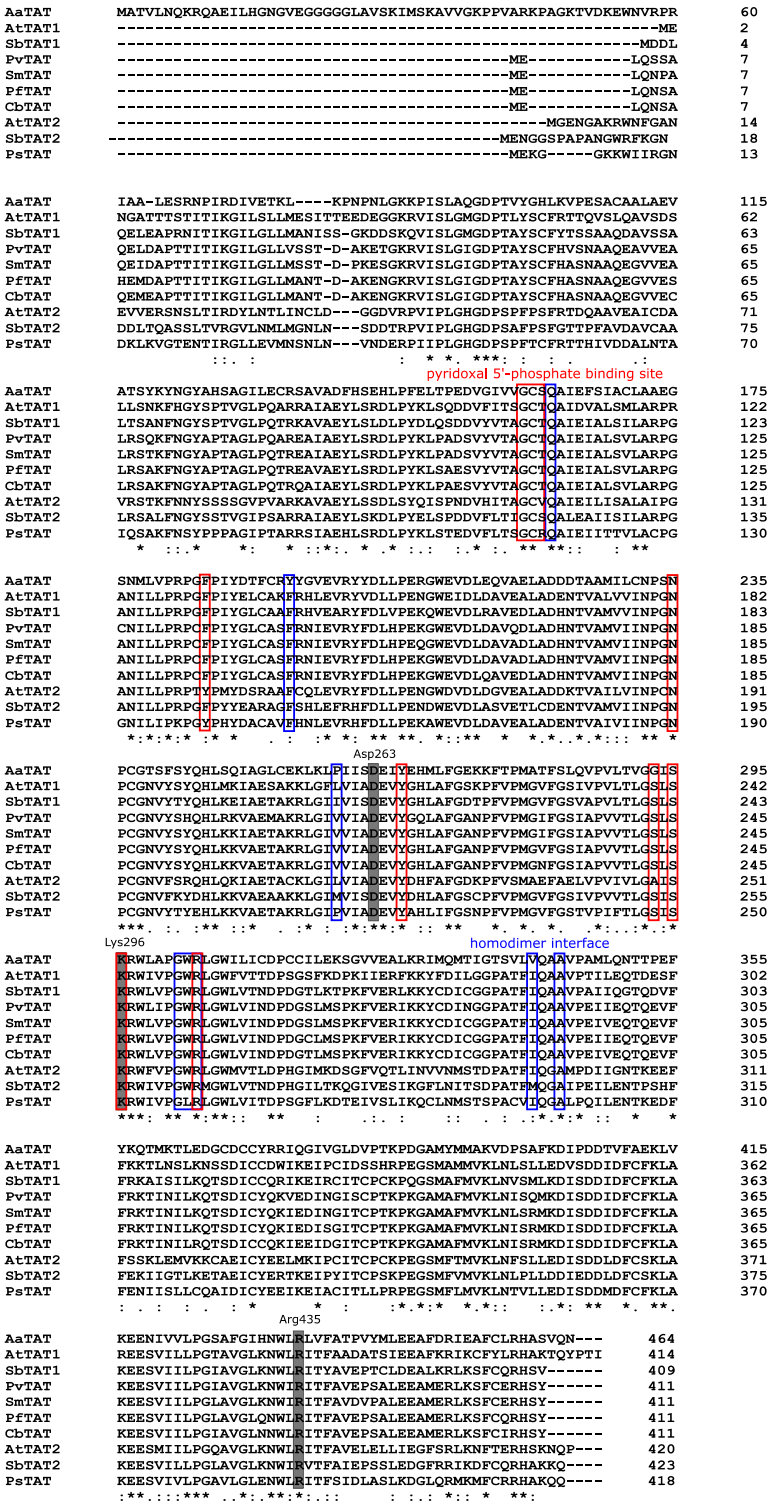


Figure 8 Alignment of some characterised plant TATs and AaTAT (QKX95019). *Arabidopsis thaliana* (NP_200208 for AaTAT1 and NP_198465 for AaTAT2), *Scutellaria baicalensis* (AIV98132 for SbTAT1 and AIV98133 for SbTAT2), *Prunella vulgaris* (AJW87632), *Salvia milthiorrhiza* (ABC60050), *Perilla frutescens* (ADO17550), *Coleus blumei* (CAD30341), *Papaver somniferum* (ADC33123). Motifs are colour coded, important amino acids are annotated (Busch and Petersen 2021).

The direct comparison of AaTAT with characterised TATs revealed that the latter are considerably shorter: 409 aa residues for SbTAT1, 411 aa for PvTAT, SmTAT, PfTAT, CbTAT, 414 aa for AtTAT1, 418 aa for PsTAT, and 420 and 423 aa for AtTAT2 and SbTAT2, respectively. The excess length of AaTAT of 53 aa residues compared to representatives of Nepetoideae (Busch and Petersen 2021) is reflected at the *N*-terminus (Fig. 8). Since two other families, *Arabidopsis thaliana* (Brassicaceae) and *Papaver somniferum* (Papaveraceae), are also represented beside the very similar Nepetoideae sequences, it cannot be assumed that this is a subjective characteristic due to the selective choice (characterised TATs, no putative enzymes) in the alignment. Investigation of plant TATs of the Fabaceae *Glycine max* and *Medicago truncatula* (not shown), which were considered by some authors in their respective alignments (Huang et al. 2008b; Lu et al. 2013a), showed that there was also no *N*-terminal excess. When including non-characterised TATs of more basal species (Fig. 9), i.e. the bryophytes *Physcomitrium patens* (Funariaceae) and *Marchantia polymorpha* (Marchantiaceae), the charophyte green alga *Klebsormidium nitens* (Klebsormidiaceae), the lycophyte *Selaginella moellendorffii* (Selganiellaceae), it becomes apparent that these indeed show a comparable excess in comparison to *Coleus blumei* TAT as example of higher plants. This is most obvious with the moss *Physcomitrium patens*. The C-terminus is largely comparable and shows only minor variances.

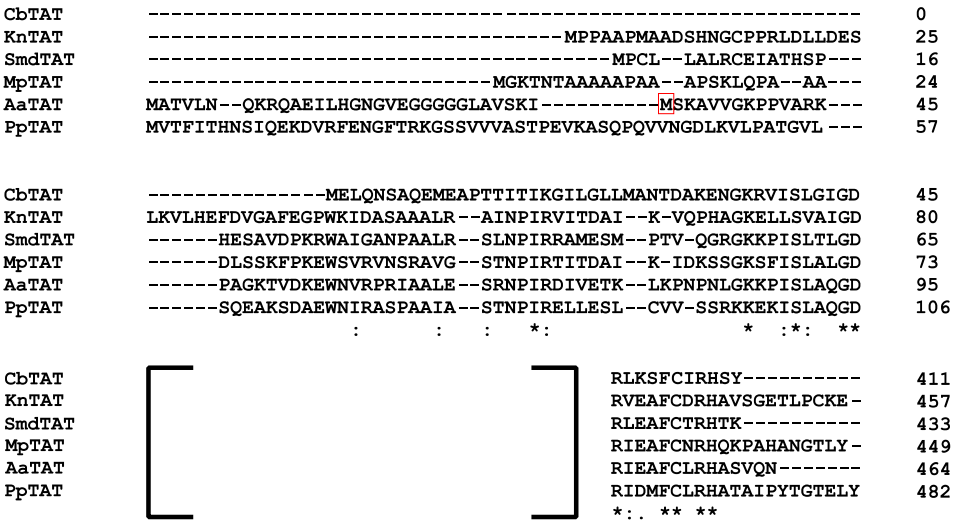


Figure 9 Alignment of TATs from basal plant species and *Coleus blumei*. The detail is on *N*- and *C*-terminal regions, the mid-section has been excluded. The second methionine is indicated in red. *Coleus blumei* (CAD30341), *Klebsormidium nitens* (GAQ85880), *Selaginella moellendorffi* (XP_024536669), *Marchantia polymorpha* (OAE18929), *Anthoceros agrestis* (MN922307, this work), *Physcomitrium patens* (XP_024361307).

The closer relationship of AaTAT to the TATs of the basal species, already indicated by the blastp analysis (Table 3) and the excess length in the region of the *N*-terminus (Fig. 8), is also demonstrated in the phylogenetic tree (Fig. 10). The tree comprises TAT sequences of said basal plants (diamonds), which form the sister branch to the TATs from the higher plant families. The latter to be distinguished as Lamiaceae (circles), all of which contain RA and have proven TAT activity and other dicotyledonous plants (triangles) with proven TAT activity. Since no biochemical or other information on the aminotransferases of the basal species is known, except for *Anthoceros agrestis* in this current work, it cannot be ruled out that those enzymes are in fact biochemically different subgroup I aminotransferases. For this reason, more distantly related characterised aminotransferases (open square) were included in the phylogenetic tree. It shows that only the phenylalanine aminotransferases from *Atropa belladonna* and *Rosa spec*. cannot be clearly distinguished from TATs of related higher plants. However, the clear separation of other distantly related aminotransferases, such as subgroup I representatives (Mehta et al. 1993) , i.e. aspartate, alanine and histidinol aminotransferases, as well as non-classified representatives, i.e. tryptophan and prephenate aminotransferases, suggest that the assignment as TATs or at least as aromatic aminotransferases is correct.

Figure 10 Phylogenetic tree (maximum likelihood analysis) of TAT amino acid sequences of several members of the Lamiaceae (circles) and representatives of other higher plant families (triangles) and lower plants (diamonds). Analyses were conducted in Mega 7. The tree with the highest log likelihood (−13,478.93) is shown. Evolutionary distance is indicated by branch distance and bootstrap values are shown next to the node. Underlying TAT amino acid sequences: *Salvia miltiorrhiza* (ABC60050), *Prunella vulgaris* (AJW87632), *Perilla frutescens* (ADO17550), *Coleus blumei* (CAD30341), *Scutellaria baicalensis* (AIV98132 for SbTAT1 and AIV98133 for SbTAT2), *Arabidopsis thaliana* (NP_200208 for AtTAT1 and NP_198465 for AtTAT2), *Papaver somniferum* (ADC33123), *Selaginella moellendorffii* (XP_024536669), *Klebsormidium nitens* (GAQ85880), *Marchantia polymorpha* (OAE18929), *Physcomitrium patens* (XP_024361307), *Anthoceros agrestis* (MN922307, this paper). The two stars following the name indicate the presence of RA in the plant (first star) and an experimentally (enzyme assay or expression profiling) proven TAT activity (second star), if filled in. Underlying amino acid sequences of distantly related aminotransferases (open squares): *Arabidopsis thaliana* TrpAT (AAO63403), *A. thaliana* IPAAT (AAP68293), *A. thaliana* PrephenAT (Q9SIE1), *Oryza sativa* AlaAT (BAA77261), *A. thaliana* AlaAT (Q9SR86), *A. thaliana* Ala/GluAT (AAN62333), *Glycine max* AspAT (AAC50015), *A. thaliana* (mitochondrial) AspAT1 (P46643), *A. thaliana* (cytoplasmic) AspAT2 (P46645), *A. thaliana* (plastidic) AspAT5 (P46248), *Nicotiana tabacum* HistidinolAT (CAA70403), *Atropa belladonna* PheAT (AHN10104), *Rosa* 'Yves Piaget' PheAT (Hirata et al. 2012). Original figure from Busch and Petersen (2021).

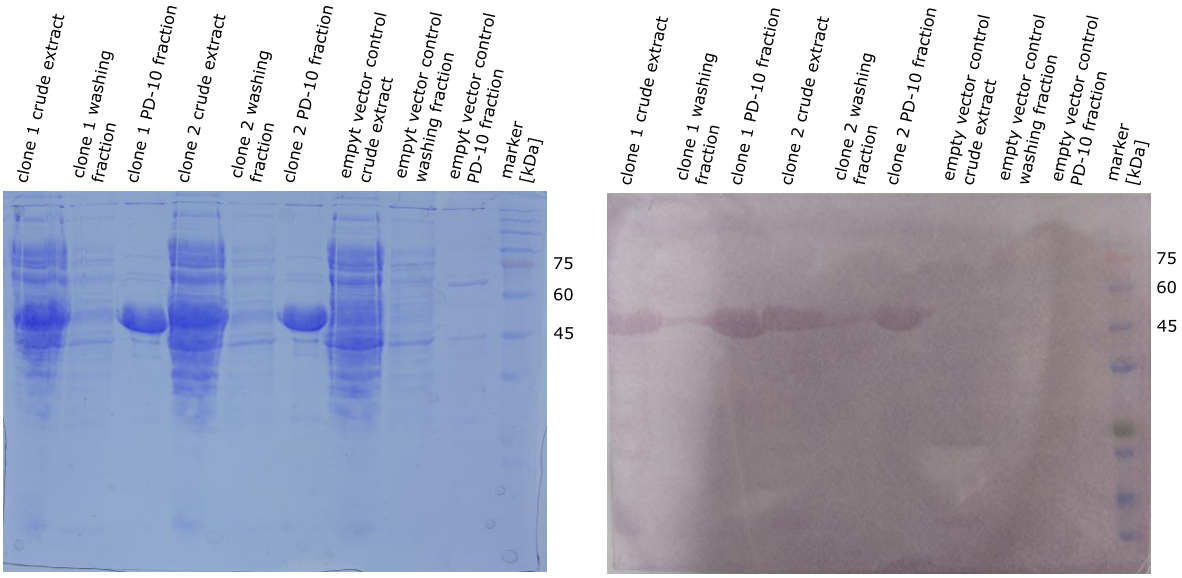


Figure 11 Expression of AaTAT in *E. coli* SoluBL21 (Busch and Petersen 2021). SDS-PAGE (left) and Western blot (right) of AaTAT with crude extract, pooled washing fractions and PD-10 fraction. The two individual clones show distinct bands at the corresponding estimated molecular weight (53.09 kDa). The empty vector control does not show any matching protein bands.

1.3 Expression of TAT from *Anthoceros agrestis*

To prepare for enzyme expression, the CDS was amplified with a second full-length primer set with overlapping regions to pET-15b (Novagen) and introduced through homologous recombination in *E. coli* EZ (Qiagen). After verifying the correct sequence and insertion, pET-15b harbouring full-length *AaTAT* was introduced into *E. coli* SoluBL21 (Amsbio) as described by Busch and Petersen (2021). The protein mass, including the *N*-terminal His-tag sequence (His_6), was predicted to be 53.09 kDa.

The *E. coli* SoluBL21 cells containing pET-15b with *AaTAT* were incubated for at least 16 h at 25 °C in TB medium after induction with 1 mM IPTG. A negative control with cells containing pET-15b without *AaTAT* was performed in parallel. The isolated crude extracts were further subjected to metal chelate chromatography and gel filtration and were analysed by SDS-PAGE and Western blotting, which demonstrated successful protein formation (Fig. 11).

An initial assay for enzyme activity was successful. Crude extracts and purified enzyme of both TAT and empty vector control were analysed. As expected, the crude extract of the empty vector also showed TAT activity, but not the His-Tag purified sample (Fig. 12). Since a reaction

was detectable, but the product did not appear to fit the pHPP standard, TAT enzyme assays were analysed by LC-MS. Two peaks were found to correspond to pHPP (Fig. 13). This is possible, since pHPP is subject to keto-enol tautomerism (Painter and Zilva 1947; Bücher and Kirberger 1952).

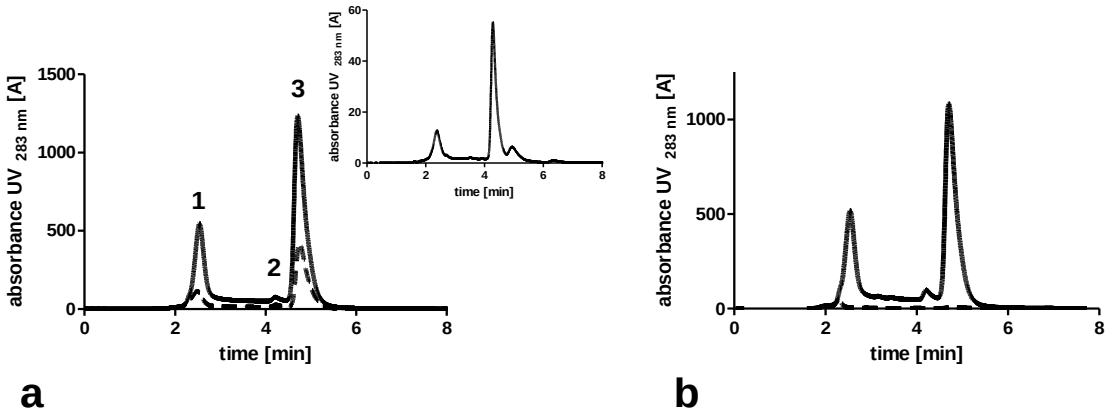


Figure 12 Activity of AaTAT and empty vector control (dashed) with L-Tyr after 30 min. Assays with crude extract (a) and purified enzyme (b) were analysed by HPLC at 283 nm, 45% methanol, 0.01% H_3PO_4. The respective empty vector control is indicated by a dashed line. The chromatogram revealed 3 peaks, with peak 2 corresponding to the main compound of the pHPP reference (inset).

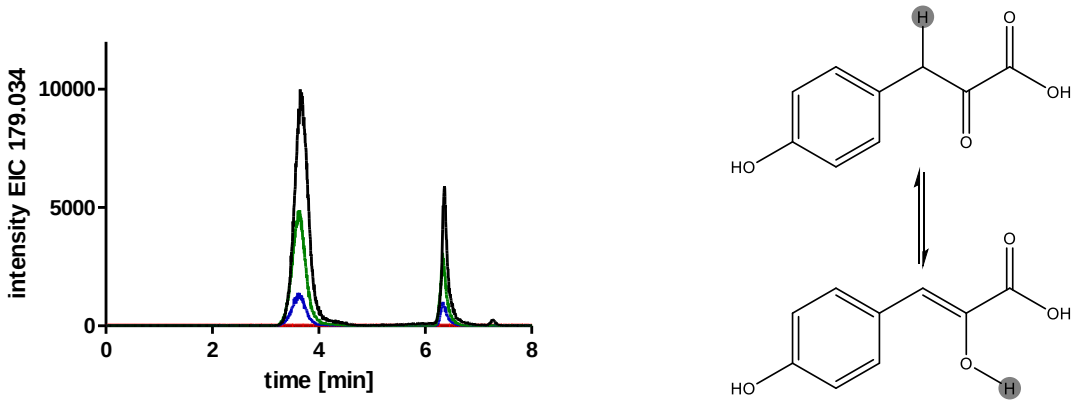


Figure 13 The reaction of AaTAT with L-tyrosine characteristically forms two peaks. Assays were analysed by LC-MS after incubation for 0, 1, 5 and 15 (red, blue, green, black) minutes. The extracted ion chromatogram of pHPP is shown. The keto-enol tautomerism causal for this effect is shown on the right.

In addition, pHPP spontaneously decomposes to 4-hydroxybenzaldehyde (HBA), enhanced in alkaline medium (Doy 1960; Pitt 1962, and literature cited therein). To clarify the assumption

that the main compound in the pHPP standard used as well as the unknown product peak is HBA, further experiments were performed. Fig. 14 demonstrates the change in the chromatogram of enzyme assays that were either extracted directly and were measured instantly after dissolving or were subjected to alkaline treatment. HBA was identifiable by LC-MS which clearly displays the relevant mass in addition to the pHPP peaks and a new unknown peak (Fig. 15).

The complexity of the analytes resulting from pHPP led to the development of a photometric measurement method based on Diamondstone (1966).

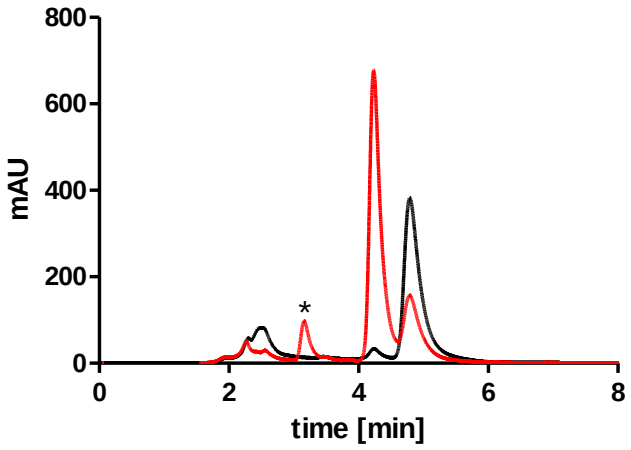


Figure 14 TAT enzyme assays without KOH and with KOH treatment after 10 min. Assays were analysed by HPLC at 283 nm, 45% methanol, 0.01% H_3PO_4. Black indicates the assay measured immediately after redissolving with a dominating enol peak. Red after KOH treatment with emerging probable HBA signal and a new unknown peak (asterisk).

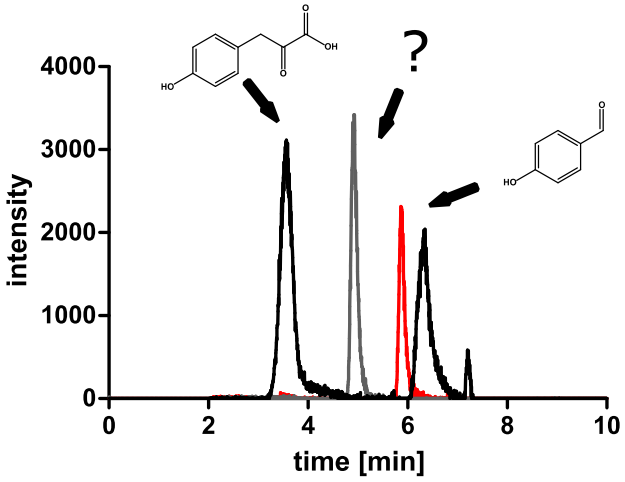


Figure 15 Reaction of AaTAT after 10 min incubation and KOH treatment. Assays were analysed by LC-MS. Two pHPP peaks, resulting from keto-enol tautomerism (black, m/z 179.034 ± 0.01), the degradation product of pHPP, identified as HBA (red, m/z 121.028 ± 0.01) and an unknown peak with a mass corresponding to 4-hydroxyphenylacetate or 3,4-dihydroxyphenylacetaldehyde (grey, m/z 151.039 ± 0.01).

1.4 Implementation of the photometric method

The absorption maximum of HBA, the degradation product of pHPP, was determined to be 330 nm for the present experimental setup (Fig. 16). The measurement of phenylpyruvate, the reaction product of L-phenylalanine, was performed at 320 nm according to Ru et al. (2017) after confirmation. The deviation of 1 nm of the former can be explained by individual deviations in the experimental setup. The absorption maximum depends on the pH value, the effect is exemplarily demonstrated in Fig. 52 (appendix). Background measurements were carried out for all kinetic investigations and, if necessary, the test solution was adjusted or buffered.

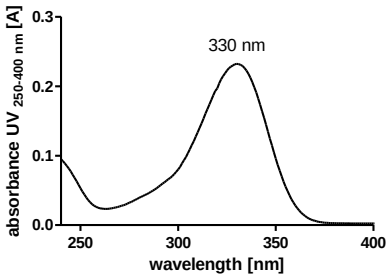


Figure 16 Absorption maximum of 4-HBA at 330 nm. The absorption maximum determined by a recording of a spectrum of 4-HBA (n=3) matching the experimental setup.

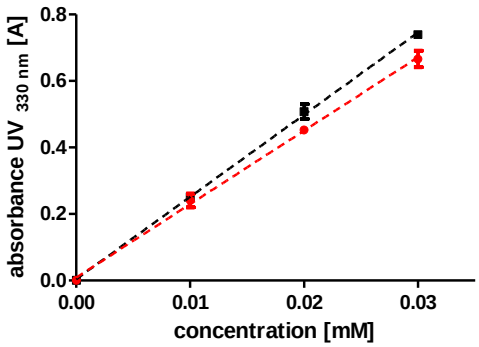


Figure 17 Rate of conversion of HBA (black) and pHPP (red) shown for increasing concentrations.

The rate of conversion of the formation of HBA from pHPP was determined to be at least 92% (Fig. 17). This is in contrast to the 80% that Diamondstone (1966) was able to determine for the conversion with NaOH. A preliminary experiment with NaOH was able to confirm this percentage. More important for the experimental setup, however, is the linear relationship between the concentration of pHPP used and the resulting absorption. This could be confirmed and indicates a uniform conversion. The absolute conversion ratio is therefore negligible, since pHPP was used for the standards in enzyme kinetic studies instead of HBA, which would require a correction factor. In addition, the 30 min incubation for HBA formation given by Diamondstone was tested and extended for the experimental setup presented here. Post-incubation at 45 °C for 15, 30 and 45 min produced consistent results. In each case, no enol peak was detectable, whereas incubation at room temperature with open lid or optionally vigorous shaking seemed to produce no reliable results at 30 min (appendix Fig. 53).

1.5 pH optimum and temperature optimum

Assays performed with purified AaTAT protein and Britton-Robinson (BR) buffer suggested a pH optimum in the range of 8-8.5 (Fig. 18 a). Although BR up to pH 10 was used, the actual pH values did not reflect the range above 8.5. Therefore, CHES as well as CAPS buffers were additionally used. The pH optimum could be determined to be at pH 7.9-8.4 (Busch and Petersen 2021). It should be noted that the standard deviation shown in Figure 18 for pH 7.4 would also allow inclusion in the optimum. However, when considering the individual series of measurements, the maximum is twice attributable to 7.9 and once to 8.4. Including further

measurements, the 7.4 value could thus be excluded as a maximum with high probability. For the study of temperature dependence, the buffer was adjusted to pH 8.5 accordingly. Using a 1 M Tris/HCl buffer, no significant deviation of the actual pH value could be detected. The temperature optimum (Fig. 18 b) was 60 °C (Busch and Petersen 2021). Nevertheless, all subsequent experiments, including substrate-saturation kinetics, were performed at 40 °C to avoid enzyme damage during prolonged pre-incubation and to avoid negative effects on substrates.

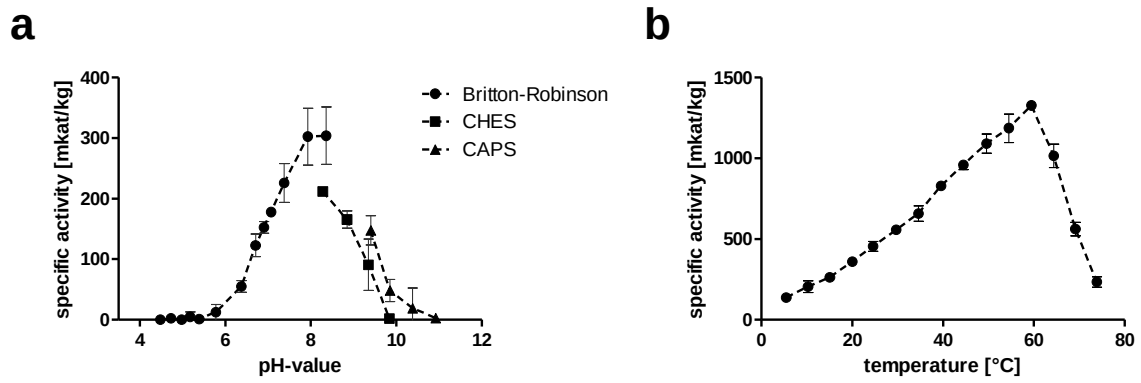


Figure 18 pH optimum (a) and temperature optimum (b). The pH optimum was determined using the three indicated buffer systems. The temperature optimum was determined at pH 8.5. Means of 3 determinations ± SD.

1.6 Dependence on the external pyridoxal phosphate (PLP) concentration

In preparation for substrate-saturation kinetics, the effect of PLP addition on AaTAT was determined. A basal activity, equal to the activity without addition, could be measured in the range up to 0.0024 mM PLP. It increases noticeably starting from 0.024 mM PLP forming a plateau at concentrations between 0.24 mM and 0.96 mM PLP, after which it decreases again (Fig. 19). As a consequence of the results, the PLP concentration to be used was set to 0.48 mM (Busch and Petersen 2021).

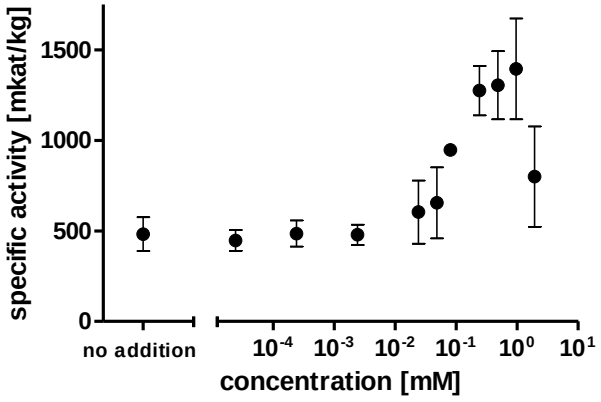


Figure 19 Impact of increasing PLP concentrations on AaTAT activity. A plateau phase from 0.24 to 0.96 mM and basal activity at concentrations from 0 mM to 0.0024 mM as well as a probable decrease in the positive effect at concentrations exceeding the plateau. Means of 3 determinations ± SD. Results according to Busch and Petersen (2021), modified.

1.7 Substrate-saturation kinetics

Kinetic data were evaluated by plotting the specific activities determined at increasing substrate concentrations according to Michaelis-Menten with additional linearisation according to Hanes-Woolf (Fig. 20 and 21). The results (Table 4) were first published in Busch and Petersen (2021). To obtain comparable results, linearity was ensured for the lowest and highest substrate concentration. To determine the amino donors (L-Tyr, L-Phe), the most efficient acceptor, i.e. 2-oxoglutarate, was added at saturation (80 mM). L-Tyr was tested with concentrations of up to 6 mM and was accepted with the highest affinity (K_m 0.53 ± 0.05 mM), with V_{max} at 921.5 ± 142.5 mkat/kg, resulting in the highest kinetic efficiency with 91.7 ± 9.8 $mM^{-1}s^{-1}$. Since the saturating addition of L-Phe was not possible, a direct linear plot according to Cornish-Bowden and Eisenthal (1978) was used to estimate kinetic data. The affinity of AaTAT was 133-fold lower (K_m 70.47 ± 4.18 mM) towards L-Phe, but with a higher V_{max} at 2240.3 ± 505.7 mkat/kg. To be sure that this was not a lack of comparability between the two plots, a replicate of L-Tyr kinetics adapted to the number of measuring points of L-Phe was re-evaluated via Cornish-Bowden and Eisenthal. The resulting K_m (0.65 ± 0.13 mM) and V_{max} (998.9 ± 254.0 mkat/kg) are apparently close to the already shown parameters. It can therefore be assumed that Cornish-Bowden is an adequate substitute for the determination of L-Phe parameters. Due to substrate concentrations of up to 132 mM of L-Phe, and the

resulting high K_m value, the kinetic efficiency is only 1.7 ± 0.4 $mM^{-1}s^{-1}$, despite the high V_{max}, thus L-Tyr exceeds the catalytic efficiency with L-Phe almost 54-fold.

Table 4 Kinetic parameters for AaTAT. Means of 9 determinations ± SD, except L-Phe (n = 3x3), using three independent enzyme expressions according to Busch and Petersen (2021). The K_m- and V_{max}-values for L-tyrosine and all amino acceptors are calculated from Michaelis-Menten equations, parameters for L-phenylalanine were determined by Cornish-Bowden linear plots. k_{cat} was calculated assuming a molecular mass of 53.09 kDa for AaTAT including the His_6-tag.

Varied substrate	Constant substrate	K_m [mM]	V_{max} [mkat/kg]	k_{cat} [s^{-1}]	k_{cat}/K_m [$mM^{-1}s^{-1}$]
L-tyrosine	2-oxoglutarate	0.53 ± 0.05	921.5 ± 142.5	48.9 ± 7.6	91.7 ± 9.8
L-phenylalanine	2-oxoglutarate	70.47 ± 4.18	2240.3 ± 505.7	118.9 ± 26.9	1.7 ± 0.4
2-oxoglutarate	L-tyrosine	15.13 ± 1.34	1376.1 ± 97.6	73.1 ± 5.2	4.9 ± 0.3
oxaloacetate	L-tyrosine	4.05 ± 1.44	12.6 ± 0.8	0.7 ± 0.04	0.2 ± 0.09
phenylpyruvate	L-tyrosine	7.92 ± 2.35	530.1 ± 117.1	28.2 ± 6.2	3.9 ± 1.4
pyruvate	L-tyrosine	81.14 ± 15.12	223.8 ± 30.0	11.9 ± 1.6	0.15 ± 0.02

Furthermore, four amino acceptors were studied, 2-oxoglutarate and oxaloacetate with concentrations of up to 120 mM, phenylpyruvate with up to 36 mM, and pyruvate with up to 260 mM. In each case, L-Tyr, the favoured amino donor, was added in 6 mM concentration. Although 2-oxoglutarate with K_m 15.13 ± 1.34 mM had a lower affinity than oxaloacetate with K_m at 4.05 ± 1.44 mM, the maximum activity exceeded about 100-fold. Overall, 2-oxoglutarate had the highest kinetic efficiency of the examined amino acceptors. Oxaloacetate was the substrate with the lowest V_{max} (12.6 ± 0.8 mkat/kg), trailing by a wide margin, raising questions about the reason for this result (see discussion). Phenylpyruvate, on the other hand, showed ratios comparable to 2-oxoglutarate, with K_m at 7.92 ± 2.35 mM and V_{max} of 530.1 ± 117.1 mkat/kg, resulting in a similar kinetic efficiency. Due to the similar measured values for V_{max} for 2-oxoglutarate and phenylpyruvate, a direct comparison was realised. A fixed concentration of 28 mM was compared for both substrates. The results confirmed the higher conversion of 2-oxoglutarate in line with the kinetic results (Fig. 22 a). Pyruvate had the highest K_m-value (81.14 ± 15.12 mM) and mediocre turnover (V_{max} 223.8 ± 30.0 mkat/kg) (Busch and Petersen 2021). Based on assumptions that pyruvate formed by decarboxylation is responsible for the low oxaloacetate reaction, a direct comparison was also made here. In

each case, 20 mM of the substrate were compared against each other. The measured pyruvate turnover was higher, and both activities correspond to the kinetic results (Fig. 22 b).

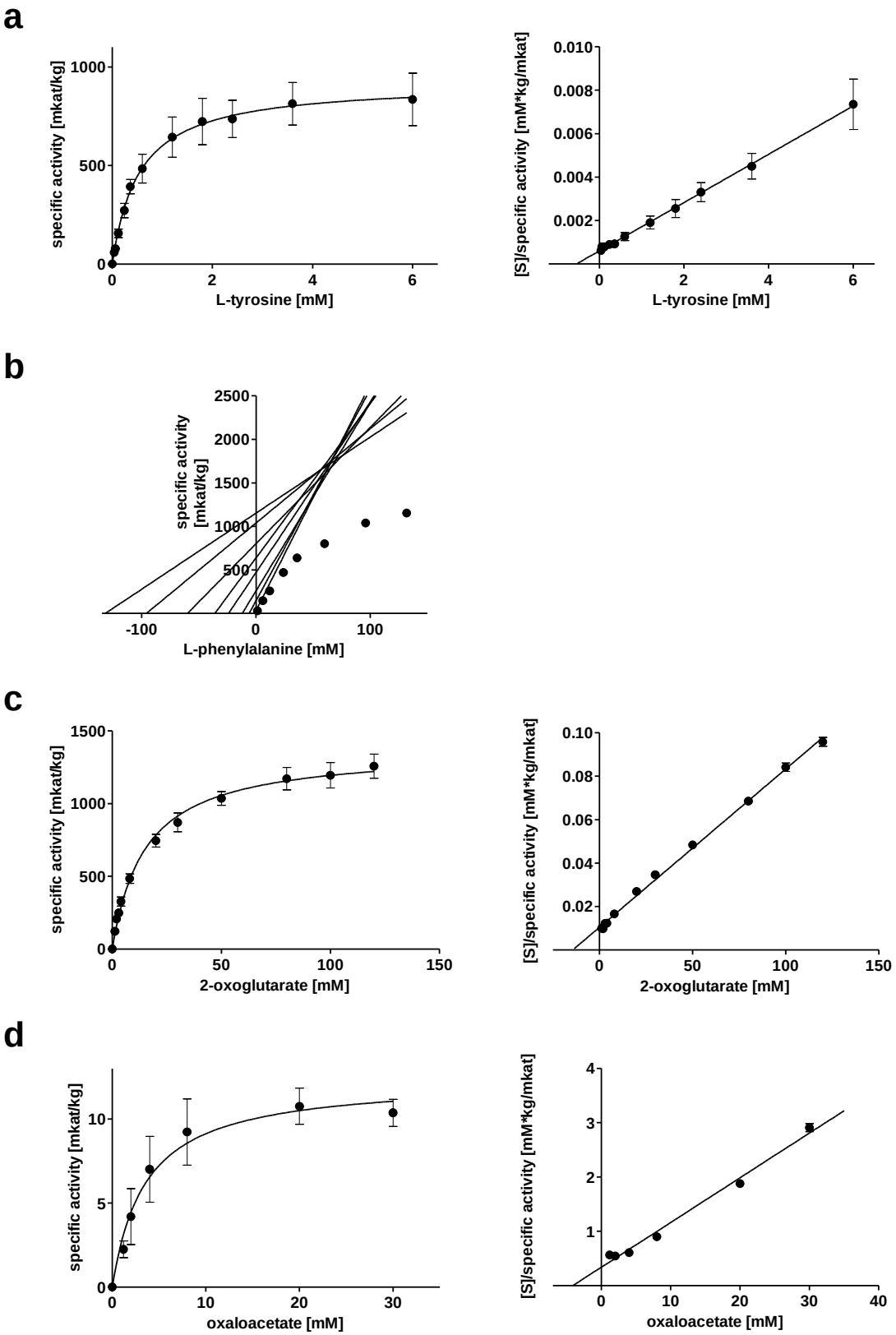


Figure 20 Plots of AaTAT activities for different substrate combination pairs. a L-Tyr with 2-oxoglutarate **b** L-Phe with 2-oxoglutarate **c** 2-oxoglutarate with L-Tyr **d** oxaloacetate with L-Tyr. Means of 9 determinations ± SD, except L-Phe (n = 3x3), using three independent enzyme expressions. Results according to Busch and Petersen (2021).

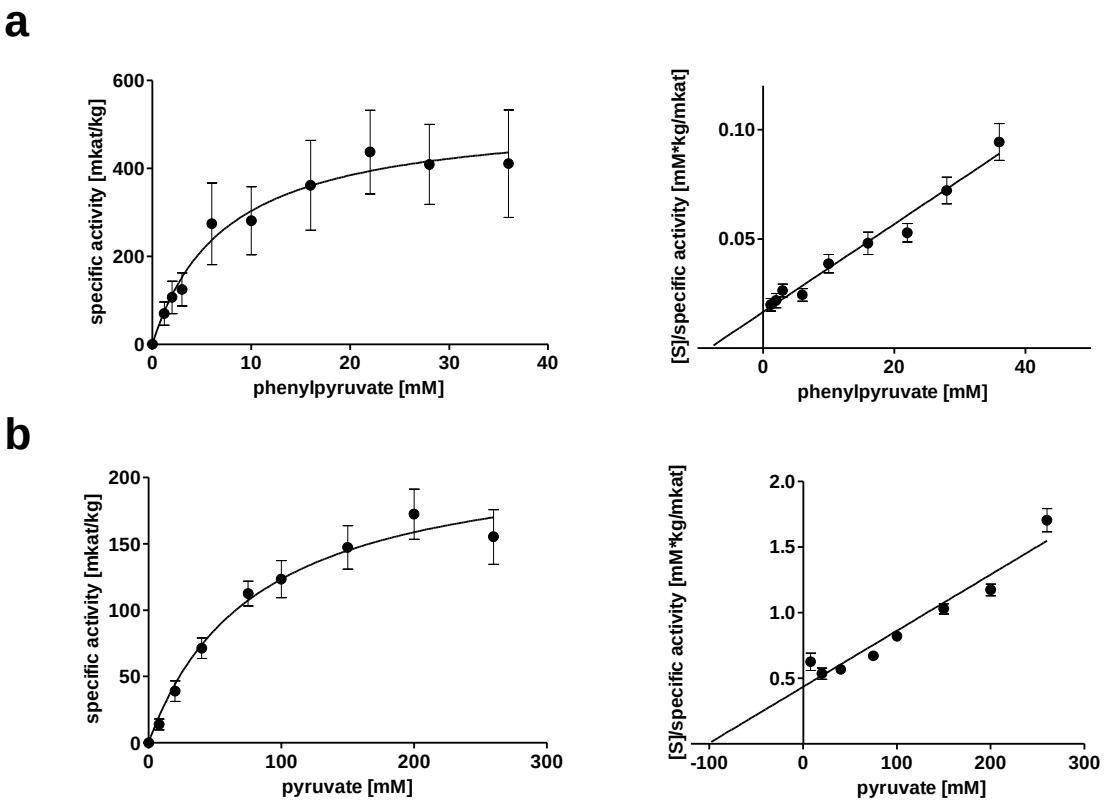


Figure 21 Plots of AaTAT activities for different substrate combination pairs. a phenylpyruvate with L-Tyr **b** pyruvate with L-Tyr. Means of 9 determinations ± SD, using three independent enzyme expressions. Results according to Busch and Petersen (2021).

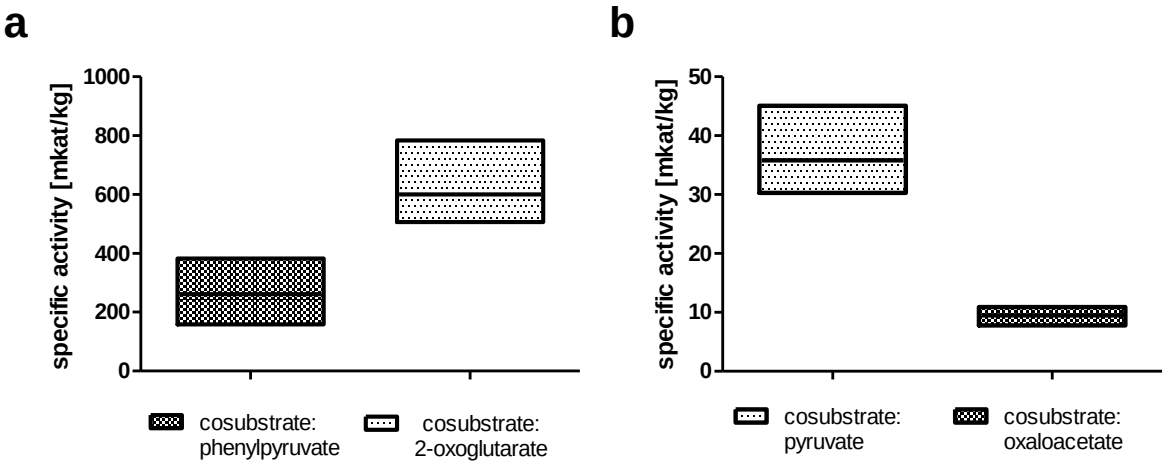


Figure 22 AaTAT activities of 2-oxoglutarate versus phenylpyruvate (a) and oxaloacetate versus pyruvate (b). 2-oxoglutarate and phenylpyruvate 28 mM and oxaloacetate and pyruvate 20 mM. Means (line) and floating bars min to max of 3 determinations, using three independent enzyme expressions.

1.8 Acceptance of D-tyrosine

To evaluate AaTAT for potential acceptance of D-tyrosine, assays were performed with 6 mM D-tyrosine and L-tyrosine, respectively. No activity was detected for D-tyrosine (Fig. 23). The result was confirmed in a linearity test with 10-fold higher concentrated enzyme. In addition, the assay analysed by TLC, with incubation time of up to 45 min, also indicated non-acceptance.

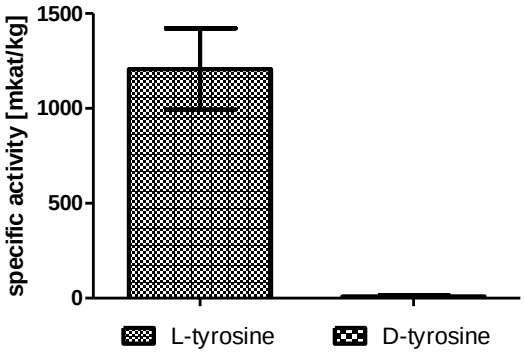


Figure 23 AaTAT activities of L-tyrosine and D-tyrosine. Means of 9 determinations ± SD, using three independent enzyme expressions.

1.9 Search for additional TAT substrates

To investigate the putative promiscuity of AaTAT with respect to additional amino acids, corresponding enzyme assays were incubated for up to 45 min and the product formation was detected via TLC followed by visualisation with ninhydrin solution. The analyses showed that L-tyrosine, L -phenylalanine, L-DOPA, L-tryptophane and L-alanine are accepted as substrates (Fig. 24), whereas L-aspartate, L-serine and D-tyrosine are not accepted (Fig. 25). The increase in glutamate formed from 2-oxoglutarate, indicated by colour depth (roughly correlating with product formation) allows semiquantitative evidence that L-tryptophane and L-alanine are less suitable substrates. Furthermore, the acceptance of the amino acceptor prephenate could be demonstrated. The suspected product arogenate, had to be investigated via extraction and analysis by LC-MS due to the lack of a commercial standard.

To verify substrate acceptance, the simplicity of the TLC method, which is characterised by the consistent formation of glutamate for all substrates used (except prephenate), was transferred to HPLC. Samples were derivatised with OPA solution and analysed by HPLC. A

glutamate peak is apparent for L-tyrosine, L-phenylalanine, L-DOPA, L-tryptophane and L-alanine, but missing for L-aspartate and L-serine. The results confirm the acceptance profile determined by TLC (Fig. 26).

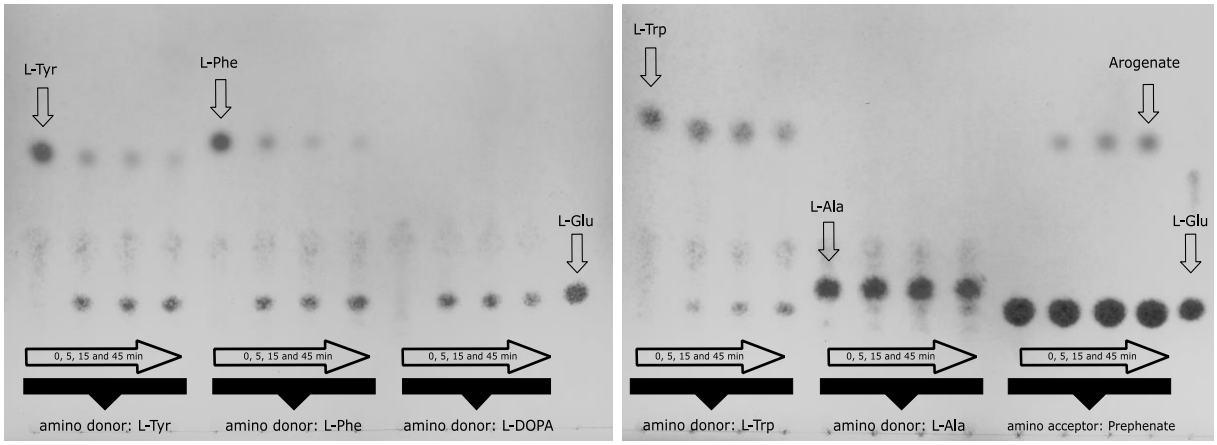


Figure 24 TLC of assays using various putative substrates accepted by AaTAT. Each amino donor (L-Tyr, L-Phe, L-DOPA, L-Trp, L-Ala) was incubated for 0, 5, 15 and 45 min with 2-oxoglutarate as cosubstrate. The TLC plates were stained with 0.3% ninhydrin solution. Glutamate is apparently not detectable in 0 min assays and increases with given reaction time, whereas amino donor spots correspondingly decrease in colour depth. This allows semiquantitative estimation of turnover rates. Prephenate assays are structured differently, since prephenate, the amino acceptor, required a L-glutamate excess. Results according to Busch and Petersen (2021). Amino acids were identified in preliminary tests using commercial standards, except for arogenate.

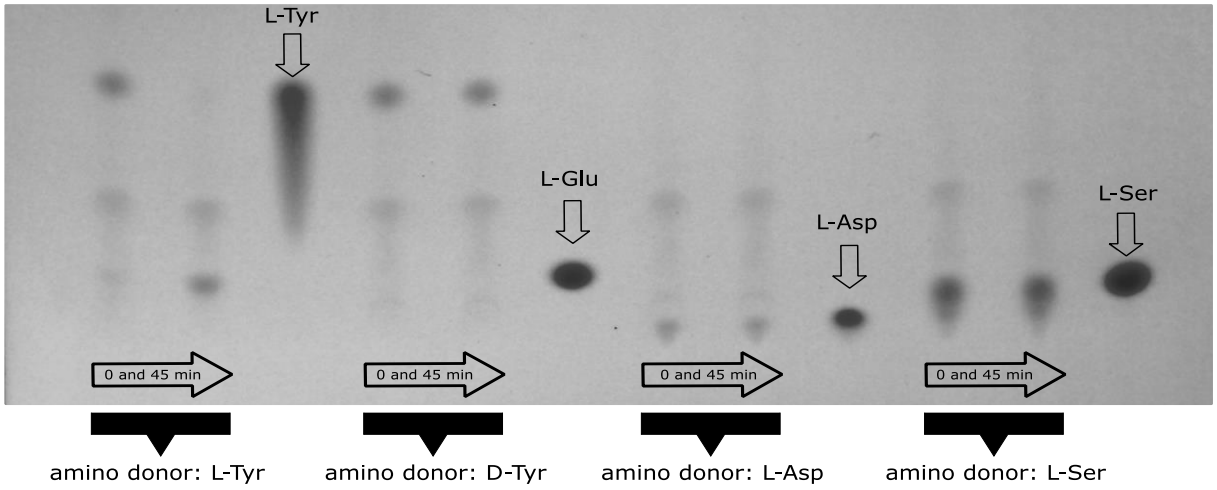


Figure 25 TLC of assays using various putative substrates not accepted by AaTAT. Each amino donor (L-Tyr, D-Tyr, L-Asp, L Ser) was incubated for 0 and 45 min with 2-oxoglutarate as cosubstrate. The TLC plates were stained with 0.3% ninhydrin solution. While L-Tyr shows a glutamate spot at 45 min that is not yet visible at 0 min, the other amino acids show no product formation even after incubation. Results according to Busch and Petersen (2021). Commercial standards are indicated.

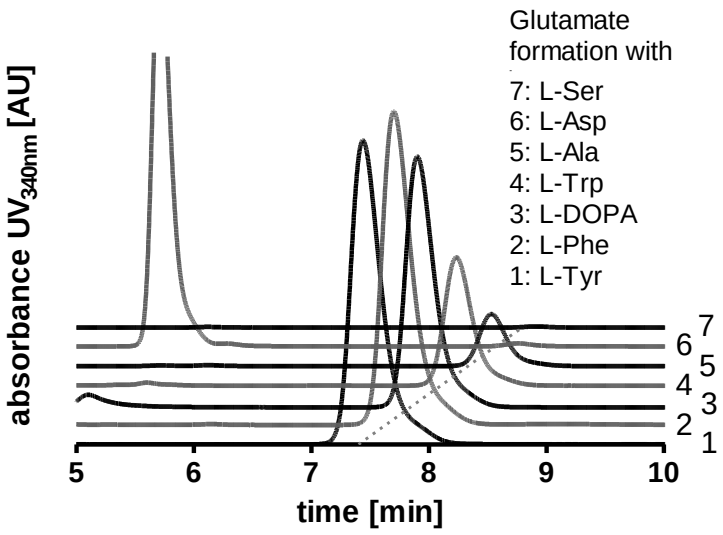


Figure 26 Verification of AaTAT amino donor substrate acceptance through OPA derivatisation and HPLC. Each curve represents the resulting glutamate formation when using the indicated amino donor. No formation is visible for L-Asp (6) and L-Ser (7). The given AUC does not reflect the specific activity of the substrate. Results according to Busch and Petersen (2021).

1.10 Identification of arogenate

AaTAT assays containing prephenate and L-glutamate were used to accumulate high amounts of putative arogenate and were subjected to TLC. After separation, identification was possible using a stained ninhydrin section of the plate, leaving the remaining material unstained. The affected area was scraped off and extracted. Through OPA derivatization and subsequent measurement via LC-MS, arogenate was detected by mass (Fig. 27 a and b). An internal control within the TLC track showed no arogenate. Additionally, arogenate was detectable after doubling prephenate concentration to 4 mM per assay and subsequent analysis by LC-MS (data not shown). However, the results after subjection to TLC are much more significant.

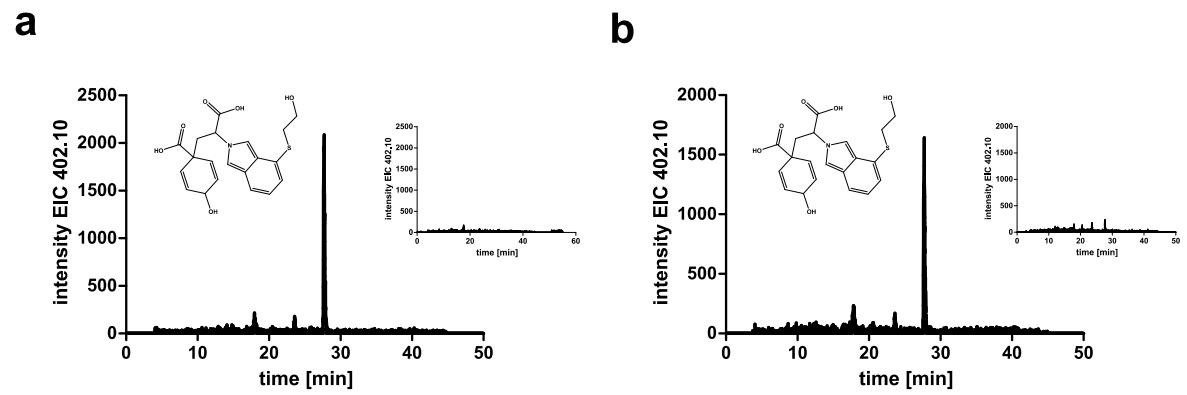


Figure 27 Identification of arogenate as OPA derivative through TLC extraction and LC-MS. Shown is the extracted ion chromatogram for m/z 402.10 [M-H], the structure of the OPA derivative is illustrated. **a** Extraction of arogenate after cellulose TLC vs. control (inset) according to Busch and Petersen (2021) **b** Extraction of arogenate after silica TLC vs. control (inset). The structure of the OPA derivative is shown in each case.

1.11 Identification of a second oxaloacetate product

Since TLC as well as detection via HPLC after OPA derivatisation showed that L-aspartate is not accepted, but in the reverse reaction when using oxaloacetate this was accepted with high affinity and low turnover (III.1.7), oxaloacetate was also tested on TLC. After staining with ninhydrin, it was found that another product was formed in addition to aspartate, which had an R_f value corresponding to that of alanine. Tsai (1967) reports the spontaneous decarboxylation of oxaloacetate to pyruvate, which would explain the alanine formation. An instruction given in the Sigma product information to prepare fresh stock solutions daily was taken as an impetus to compare old and freshly prepared oxaloacetate. The putative alanine peak occurs in both. Due to the weak appearance, a comparison of the colour depth was not possible (Fig. 28).

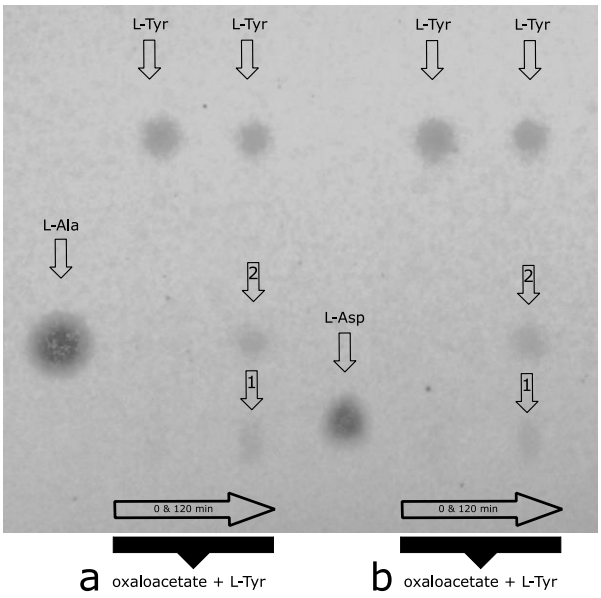


Figure 28 TLC analysis of two potential reaction products of AaTAT with oxaloacetate. Shown are the two resulting ninhydrin stainable products of oxaloacetate as amino donor with L-Tyr (1 mM) as acceptor. **a** Use of an older stock solution of oxaloacetate. **b** Use of an oxaloacetate solution prepared freshly immediately prior to use. 1 represents the expected L-Asp (the slight shift in comparison to the reference results from other assay components and optionally caused by dissolving the L-Asp reference in 0.5 M HCl, compare appendix Fig. 54) and 2 represents L-Ala. According to Busch and Petersen (2021), modified.

To uniquely verify the resulting putative alanine, the amount of two assays containing 32 mM oxaloacetate and L-tyrosine as amino donor was separated by TLC. As for the identification of arogenate, part of the TLC plate was stained with ninhydrin solution. The corresponding band to the alanine spot was scraped out, as well as a band of the same size within the TLC track, a band above the solvent front, and the spot of the alanine reference. The samples were extracted and derivatised with OPA solution with subsequent analysis by LC-MS (Fig. 29).

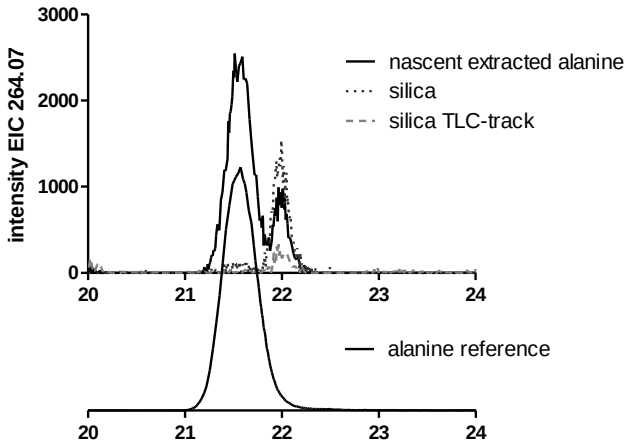


Figure 29 Identification of putative alanine spot through extraction and LC-MS. Shown in the upper half is the putative alanine extracted after separation via TLC as well as two controls, one taken from the TLC track and one outside the solvent front. Shown below is an L-alanine reference that was also extracted. The peak at 22 min is thus shown to originate from the silica material. In the extracted alanine reference (below), this is masked due to the high alanine concentration ($intensity_{t=22}$ = 15594.7).

2 Hydroxyphenylpyruvate reductase (HPPR)

2.1 RNA extraction and cDNA synthesis and gDNA extraction

RNA from *Anthoceros agrestis* suspension-cultured cells was extracted using the acid guanidinium thiocyanate-phenol–chloroform extraction according to Chomczynski and Sacchi (1987). After reconstitution of dried RNA with water, integrity was checked through agarose gel electrophoresis and concentrations at 260 nm as well as the 260/280 ratio were measured photometrically. Fig. 30 and Table 5 show the results of an extraction of *Anthoceros agrestis* cells after a five-day growth phase in CBM medium (cells kindly provided by Julia Wohl). All five samples showed distinct ribosomal subunit bands on the agarose gel, i.e. 28S and 18S and faint 5S rRNA. Based on the best 260/280 ratios (see III.1.1), sample 2 was used for synthesis of 3'-RACE-Ready cDNA, whereas sample 4 was used for synthesis of 5'-RACE-Ready cDNA.

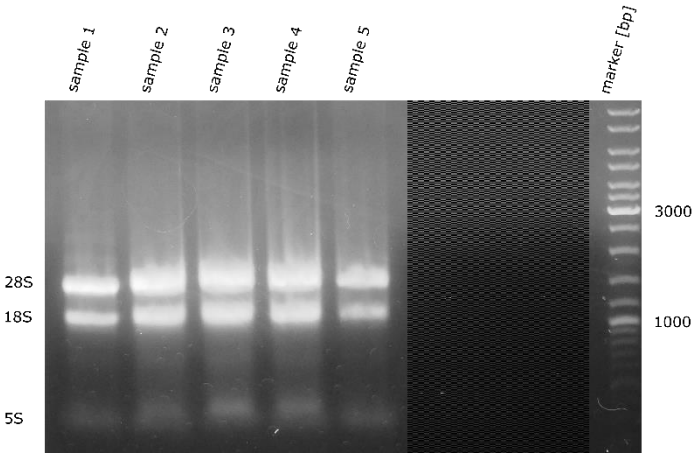


Figure 30 Extracted RNA samples from *Anthoceros agrestis*. Cells were extracted from a five days old suspension cell culture (CBM). Agarose gel electrophoresis with ethidium bromide staining. Distinct ribosomal subunit bands indicate RNA integrity.

Table 5 Concentration and purity of *Anthoceros agrestis* RNA extraction.

RNA sample	Concentration [ng/µl]	260/280 ratio
Sample 1	845	1.69
Sample 2	1255	1.82
Sample 3	915	1.64
Sample 4	835	1.94
Sample 5	915	1.73

The gDNA from *Anthoceros agrestis* suspension culture cells was extracted using the method according to Rogers and Bendich (1985) from pooled CBM and CB2 cultures. After reconstitution of dried gDNA and RNase digestion, concentrations at 260 nm as well as the 260/280 ratio were measured photometrically (Table 6). All three samples had high DNA concentrations and comparable purity, represented through the 260/280 ratio, ideally at ~1.8, indicating pure DNA.

Table 6 Concentration and purity of *Anthoceros agrestis* gDNA extraction.

gDNA sample	Concentration [ng/µl]	260/280 ratio
Sample 1	931.3	1.70
Sample 2	1169.6	1.70
Sample 3	1473.7	1.72

2.2 Amplification of cDNA and gDNA encoding HPPR from *Anthoceros agrestis*

The 3'- and 5'-RACE PCR was used to amplify the missing upstream and downstream cDNA ends and was based on an internal consensus sequence derived from two fragments kindly provided by Maike Petersen. Amplification was successful with cDNA (chapter III.2.1.) used for the 5'-RACE-PCR and with cDNA kindly provided by Julia Wohl for the 3'-RACE experiment, eventually revealing an open reading frame of 987 bp encoding 328 aa residues (Fig. 31 a and b, and appendix 1.4), although two larger variants of the CDS with 1023 bp or 1113 bp would be conceivable in two of four results. However, the additional areas were not homologous and the consensus was in favour of the shorter HPPR. This could be confirmed based on the hornwort sequence information (Szövényi et al. 2015) that had become available in the meantime (https://www.hornworts.uzh.ch/en.html). There were matches of 99% identity (327/328) with *Anthoceros agrestis* Oxford strain, 99% (327/328) for Bonn strain and 98% (319/327) for *Anthoceros punctatus*. The one amino acid difference (serine vs asparagine in AaHPPR) found in the first two was checked over several sequences available from individual sequencing of AaHPPR. Asparagine (AAC) was found in all sequencings that led to the expressible construct. An alternative candidate featured serine (AGC). All RACE sequencings examined also contained serine. It can therefore be assumed that it is indeed a wrong amino acid that appeared during amplification with the proof-reading polymerase. Nevertheless, it cannot be ruled out that this is a deviation specific to the *Anthoceros agrestis* strain used. Overall, the sequence transcribed with this enzyme was of considerably better quality than the variants transcribed without proof reading activity.

After several attempts to generate the full-length sequence via cDNA were unsuccessful, the attempt with gDNA succeeded (appendix 1.5). Apparently, the sequence length on the gel (evident in Fig. 31 d) appears slightly larger than 987 bp. A second PCR experiment showed a

comparable pattern. The full-length seemed also slightly larger in comparison to the 1000 bp marker band (appendix Fig. 55). Since no proofreading polymerase was used here, the approach was not followed further. However, despite the offset, no introns were found in any of the sequencing runs, neither in the regular full-length amplification nor in additional experiment using the non-proofreading enzyme.

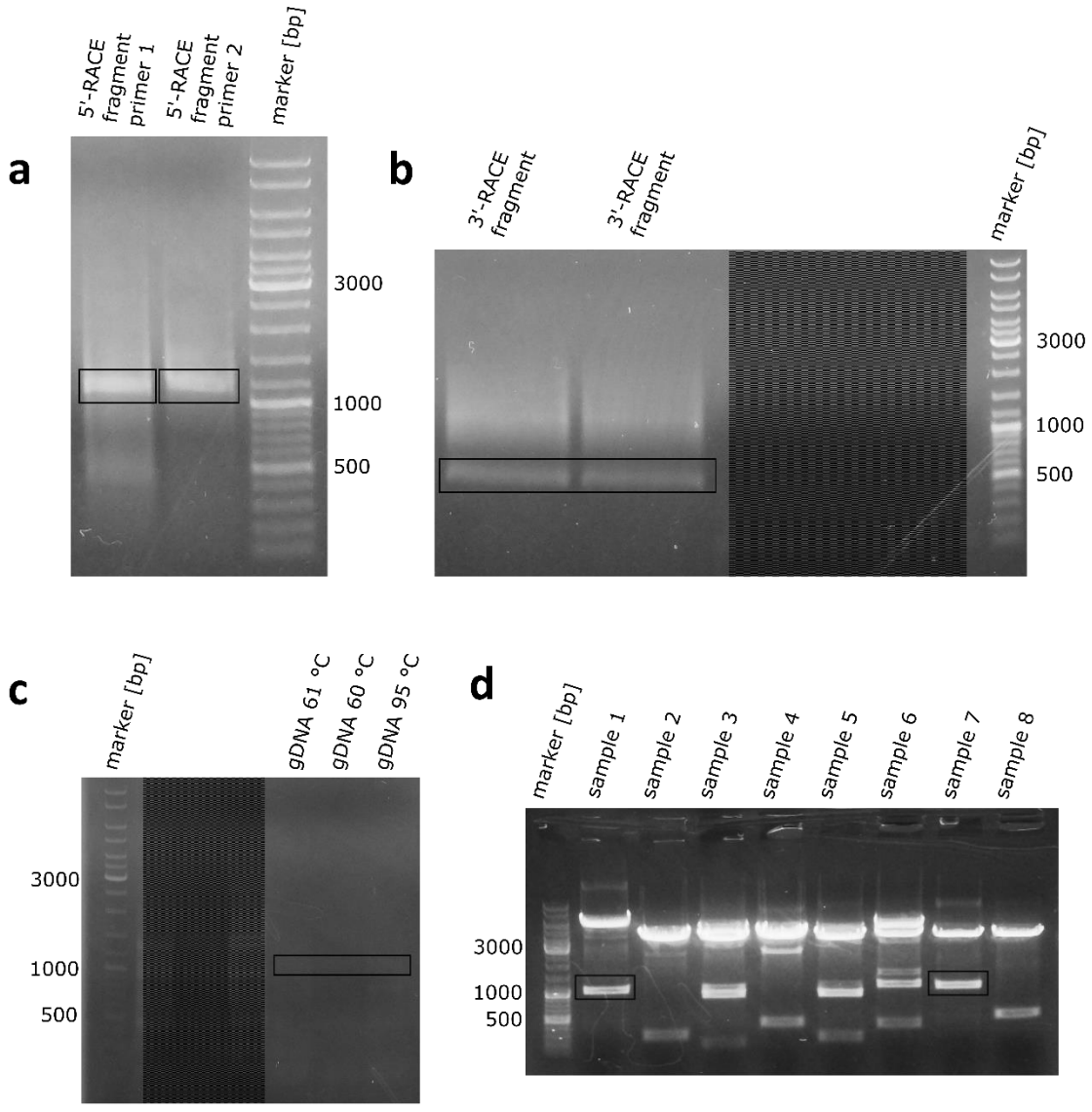


Figure 31 PCR amplification of *AaHPPR*. a 5'-RACE PCR using two different gene-specific primers, primer 2 containing a second ATG codon **b** 3'-RACE PCR **c** Full-length gradient PCR targeting the 987 bp CDS using Aa gDNA **d** Restriction digest of plasmid DNA harbouring *AaHPPR*, exemplary shown to visualise the band size, as corresponding PCR amplification was only visible by eye and could not be captured photographically (indicated through frame). Agarose gel electrophoresis with ethidium bromide staining.

Blastp analysis (Table 7) of the translated aa sequence showed high identity with sequences from the mosses *Sphagnum magellanicum and Physcomitrium patens*, the latter annotated as HPPR. In contrast to blast-searches for *AaTAT*, HPPR-like sequences from *Klebsormidium nitens* and *Selaginella moellendorffii* are not present in the evaluation. Further annotated as HPPR are *Mentha x piperita* and *Prunella vulgaris*, both Lamiaceae, and the grass *Hordeum vulgare*. Other vascular plants, i.e. *Helianthus annuus*, *Rosa chinensis*, are annotated as glyoxylate/hydroxypyruvate reductase or less specific as D-isomer specific 2-hydroxyacid dehydrogenase (*Actinidia rufa*), the respective enzyme family of H(P)PRs. Overall (inclusion of 100 sequences), all percent identities ranged around 55-59%.

Table 7 Blastp results (18/07/2022). Comparison of AaHPPR aa sequence with several selected species of interest.

Organism	Description	Accession	Identity	E value
Sphagnum magellanicum	hypothetical protein CY35_10G030100	KAH9549638.1	57.68%	6E-106
Camellia sinensis	hydroxyphenylpyruvate reductase-like	XP_028127415.1	59.49%	1E-103
Physcomitrium patens	hydroxyphenylpyruvate reductase-like	XP_024392873.1	54.60%	1E-102
Helianthus annuus	glyoxylate/hydroxypyruvate reductase A HPR2 isoform X1	XP_021997110.1	56.65%	5E-100
Actinidia rufa	D-isomer specific 2-hydroxyacid dehydrogenase family protein	GFZ04162.1	56.96%	3E-101
Fragaria vesca subsp. *vesca*	PREDICTED: glyoxylate/hydroxypyruvate reductase A HPR2-like	XP_004302898.1	57.59%	6E-100
Hordeum vulgare subsp. *vulgare*	hydroxyphenylpyruvate reductase isoform X1	XP_044975494.1	57.59%	1E-99
Rosa chinensis	glyoxylate/hydroxypyruvate reductase A HPR2	XP_024180912.1	56.96%	3E-99
Mentha x piperita	hydroxyphenylpyruvate reductase	AVZ47166.1	58.54%	3E-99
Prunella vulgaris	hydroxyphenylpyruvate reductase	AJW87633.1	57.91%	6E-99

The Conserved Domain Search (CDD) revealed the expected enzyme properties, such as the NAD(P)H+H^+/$NAD(P)^+$ binding site and the catalytic site consistent with the aa residues (Arg, His and Glu) detected by Janiak et al. (2010) for the crystallised *Coleus blumei* H(P)PR. For a

detailed elaboration on the motifs, amino acids involved, and special features, including comparison to several plant HPPRs, see chapter III.3.3. Additionally, Pairwise Sequence Alignment (EMBOSS Needle, protein) states 56.1% identity and 70.7% similarity of AaHPPR with *Coleus blumei* H(P)PR (CAD47810.2), an enzyme already enzymatically characterised with proven HPPR activity but with comparable specific activity and even higher affinity for non-aromatic substrates like glyoxylate and hydroxypyruvate.

To prepare for enzyme expression, the pDrive plasmid harbouring the HPPR CDS was cut with two restriction endonucleases to allow unidirectional ligation and was introduced into pET-15b (Novagen). After verifying the correct sequence, the pET-15b harbouring the full-length *AaHPPR* was introduced into *E. coli* SoluBL21 (Amsbio). The protein mass, including the *N*-terminal His-tag sequence (His_6), was predicted to be 36.88 kDa.

2.3 Expression of HPPR from *Anthoceros agrestis*

The *E. coli* SoluBL21 cells containing pET-15b harbouring *AaHPPR* were incubated for at least 16 h at 25 °C in LB medium or TB medium after induction with 1 mM IPTG. All experiments concerning preliminary tests including pH and temperature optimum were carried out with enzyme expressed in LB medium, substrate saturation kinetics and remaining experiments in TB medium. A negative control with cells containing pET-15b without *AaHPPR* was performed in parallel. The isolated crude extracts were further subjected to metal chelate chromatography and buffer exchange and were analysed by SDS-PAGE (Fig. 32), which demonstrated successful protein formation and a molecular mass of just above 35 kDa comparable to the calculated 36.88 kDa (including His-tag) value.

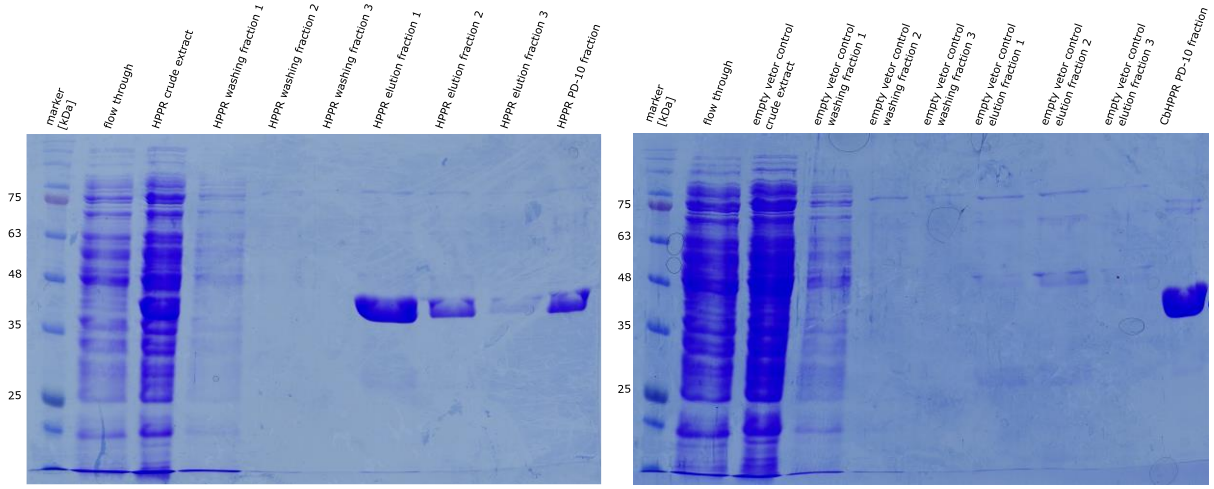


Figure 32 Exemplary expression of AaHPPR in *E. coli* SoluBL21. SDS-PAGE of AaHPPR with crude extract, washing fractions, elution fractions and PD-10 fraction (left). SDS-PAGE empty vector control and the corresponding fractions. In place of the PD-10 fraction a positive control of CbH(P)PR is shown (right). HPPR samples reveal distinct bands at the corresponding estimated molecular weight (36.88 kDa).

An initial assay for enzyme activity with crude extract was successful (Fig. 33). The reaction showed an increasing pHPL peak for 5, 10 and 20 min, dwarfing the reaction of the empty vector control. A CbHPPR assay was also run for 20 min and measured, which could confirm the reaction product in addition to a pHPL reference. As shown in Figure 34, the purified enzyme was also active. However, it became apparent early on, that a side peak or impurity appearing in the chromatogram at the site of the analyte, made further measurements difficult, especially linearity tests with short incubation times and especially with small substrate concentrations used, which resulted in analyte peaks that were not fully pronounced. Attempts to adjust the solvent, the column or to omit test components, i.e. DTT, ascorbate, did not lead to any significant advantage. The results listed in the following chapters on pH and temperature optimum could still be carried out using the HPLC method but without DTT and ascorbate addition. The following kinetics were then performed photometrically.

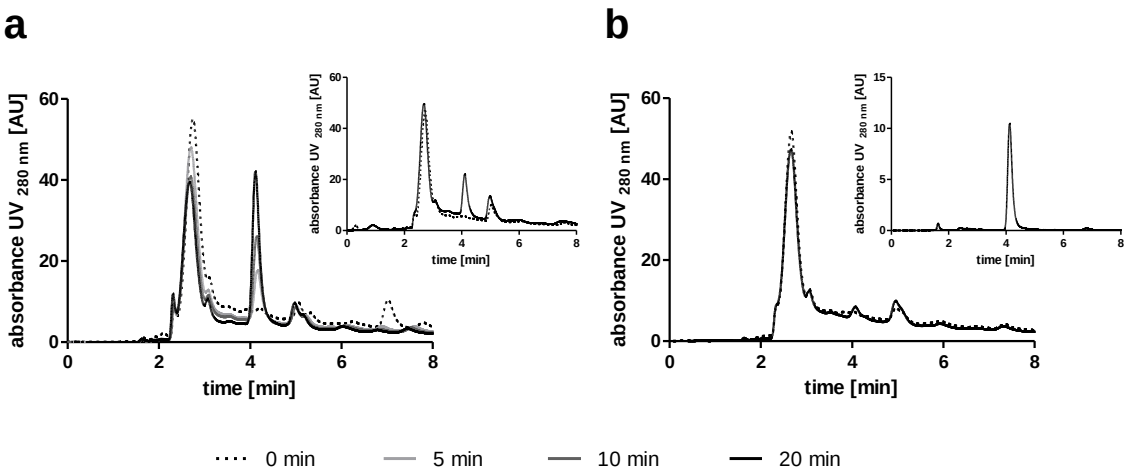


Figure 33 Activity of AaHPPR crude extract and empty vector control with pHPP after up to 20 min and CbHPPR positive control. Assays were analysed by HPLC at 280 nm, 30% methanol, 0.01% H_3PO_4. **a** Assays of AaHPPR crude extract after 0, 5, 10 and 20 min with significant activity and CbHPPR assay (inset) with purified enzyme for 0 and 20 min as reference for the HPPR reaction **b** Empty vector control after 0 and 20 min with slight activity and 100 µM pHPL reference (inset).

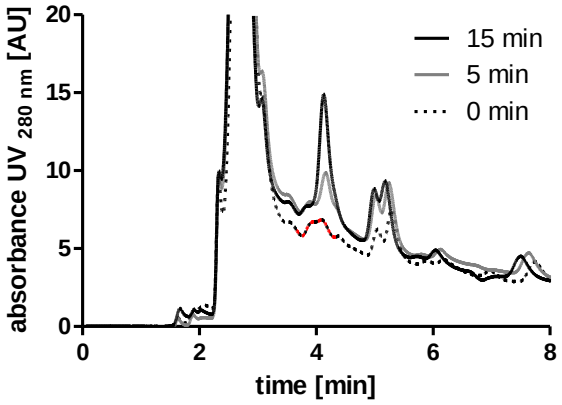


Figure 34 Activity of AaHPPR purified enzyme with pHPP after up to 15 min. Assays were analysed by HPLC at 280 nm, 30% methanol, 0.01% H_3PO_4. The purified enzyme showed distinct activity, however, the chromatogram reveals an emerging problem of the HPLC analyses (shown in red). The 0 min control reveals a peak at the position of the analyte, which interferes with the evaluation of measurements with small resulting peaks.

2.4 pH optimum and temperature optimum of AaHPPR

Assays with purified AaHPPR and KPi buffer revealed a pH optimum of 7.0-7.5 (Fig. 35 a). Accordingly, the buffer for determining the temperature optimum was kept at pH 7. Due to the standard deviation, the resulting temperature optimum is set in the large range of 29.7-49.5 °C, but with an absolute maximum at 44.4 °C (Fig. 35 b). All subsequent experiments were

carried out at 30 °C to avoid enzyme damage during prolonged pre-incubation and to avoid negative effects on substrates.

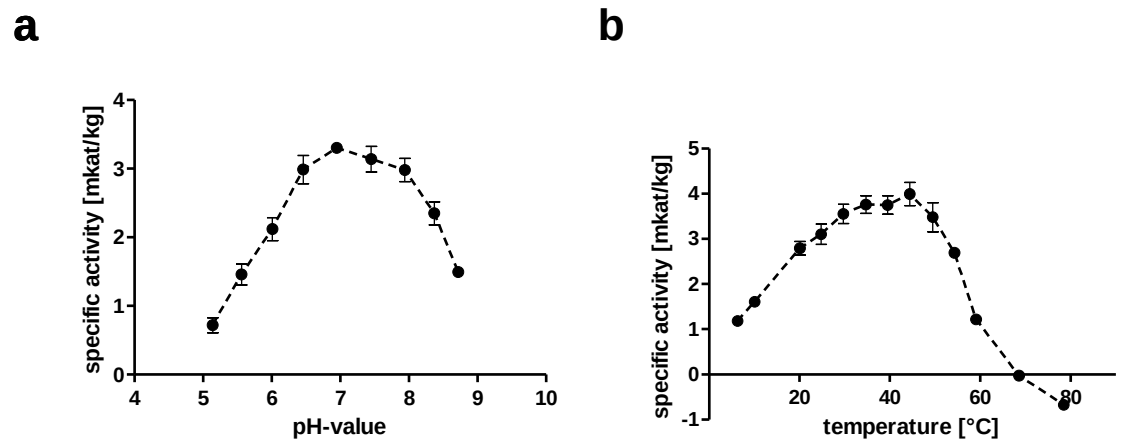


Figure 35 pH optimum (a) and temperature optimum (b) of AaHPPR. pH optimum was determined using KPi buffer system. Temperature optimum was determined at pH 7. Means of 4 determinations ± SD

2.5 Implementation of a photometric method for HPPR activity determination

The photometric characterisation of AaHPPR is based on the quantification of NAD(P)H. In its reduced form, NAD(P)H exhibits an additional absorption maximum at 340 nm compared to the oxidised form. With continuous measurement, an NADPH-consuming reaction can be detected over time through the corresponding decrease in absorption. Based on a direct linear relationship between NAD(P)H concentration and measured absorption, the quantification is then performed by a suitable NAD(P)H standard. The characteristic feature in the determination of aromatic 2-oxoacids such as pHPP is the high inherent absorption of the substrates. One solution is to shift the measurement wavelength to 380 nm. A useful side effect is that the absorption of NAD(P)H in saturation is correspondingly lower. Otherwise, the linear correlation between concentration and absorption would not be given with saturated NAD(P)H (Fig. 36). Said adjustment was adopted from Janiak (2007) .

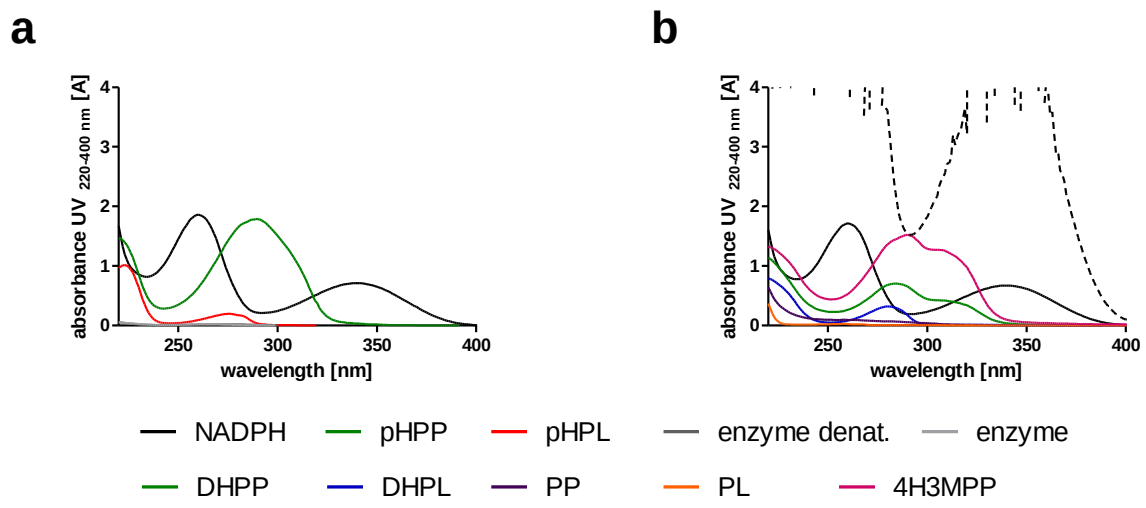


Figure 36 Photometric spectra of several assay components. a Components of the standard assay, i.e. NADPH, pHPP and (denatured) enzyme, and the reaction product pHPL. **b** Additional substrates (DHPP, PP, 4H3MPP) and products (DHPL, PP). The 340 nm NADPH absorption maximum is evident, but the multi-substituted aromatic 2-oxo acids (DHPP, 4H3MPP) still exhibit high absorption in this area. Dashed (b) is the 1 mM NADPH concentration. It is apparent that a continuous measurement at 340 nm would not be possible. The shift to 380 nm combines the advantages of lower NADPH absorption as well as exclusion of the influence of other test components.

In preparation for the determination of the substrate saturation kinetics, the appropriate enzyme concentration to be used was determined and the linear relationship between concentration and measured absorption was confirmed. At the same time, it was demonstrated that no significant absorption decrease was measurable when the enzyme was replaced with the empty vector control (Fig. 37). Furthermore, it was tested whether NADPH alone or with enzyme but without the presence of the substrate, incubated over a longer period of time, leads to a change in absorption. It could be shown that this is indeed the case (Fig. 38). Measuring NADPH in the presence of AaHPPR against a reference containing only the buffer, detected a decrease over 60 min. This was also the case when measuring NADPH without enzyme addition. Thus, the change in absorption is probably due to the NADPH itself. This is shown by the fact that when comparing NADPH plus AaHPPR and a corresponding reference with denatured AaHPPR, no further decrease in absorption is observed. Thus, care had to be taken for all future assays that equivalent NADPH is present in the references.

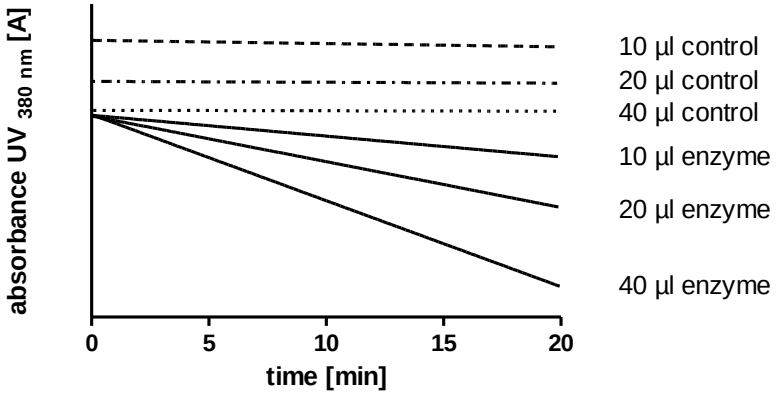


Figure 37 Photometric interval measurement of several AaHPPR additions vs. empty vector control. The assays with purified AaHPPR show an approximately linear decrease in absorption, the slope correlates with the added enzyme concentration. The empty vector controls show no significant activity and there is no correlation between applied concentration and slope.

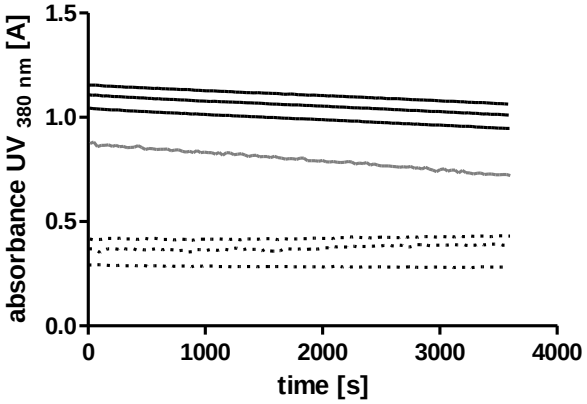


Figure 38 Change of absorption and behaviour of NADPH in different test compositions. The black lines represent replicates of NADPH in combination with AaHPPR measured against buffer as reference. The grey line represents NADPH measured against buffer. The dashed lines represent replicates of NADPH in combination with AaHPPR against a corresponding reference with denatured enzyme.

In order to determine specific activities from the measured absorption decrease, standards of different concentrations were measured for both NADPH and NADH. The molar extinction coefficient can be inferred directly from the slope of the linear regression line. The molar extinction coefficient for NADPH (Fig. 39 a) at 380 nm was 1154.95 l mol^{-1} cm^{-1} and 1099.60 l mol^{-1} cm^{-1} for NADH (Fig. 39 b).

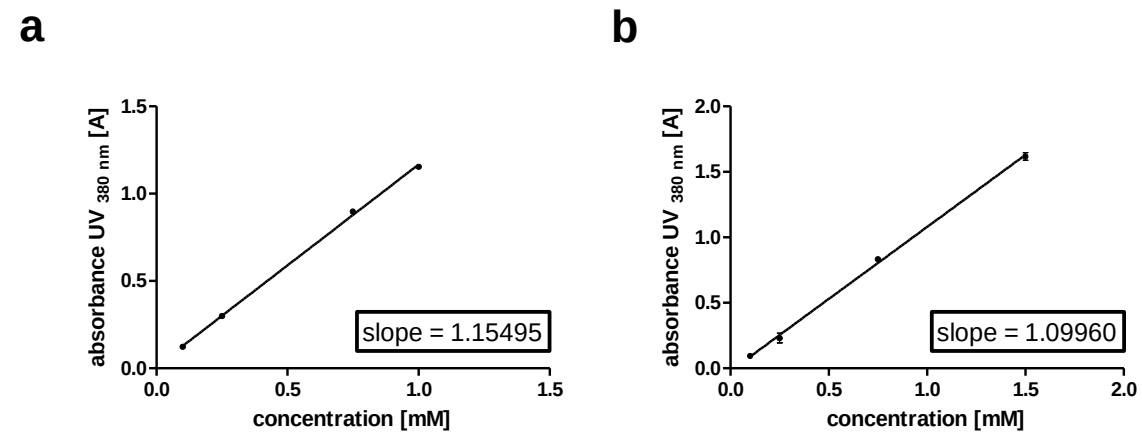


Figure 39 Molar extinction coefficient for NADPH (a) and NADH (b) at 380 nm. Means of 3 determinations ± SD

2.6 Substrate-saturation kinetics

Kinetic data (Table 8) were evaluated by plotting the specific activities determined at increasing substrate concentrations according to Michaelis-Menten with additional linearisation according to Hanes-Woolf (Fig. 40). To obtain comparable results, a linear decrease in absorption was sought for all substrates by adjusting the incubation time, proximity of the measurement points, and enzyme dilution. For pHPP, linearity was easily possible over a period of up to 20 min. For NADPH, NADH and β-HP, the method reached its limitation, especially for small substrate concentrations. Thus, long linear phases could not always be achieved for the latter and poor slopes led to distortions. Therefore, suitable enzyme concentrations were determined and the measurement was consistently performed for 5 min as a compromise. Due to the susceptibility of the small concentrations to leave the linear range more quickly, it can be assumed that the actual K_m values will be somewhat lower and V_{max} values somewhat higher. By doing so, the arbitrary selection of the evaluated time points was omitted in order not to jeopardize the comparability by bias of the evaluator.

Table 8 Kinetic parameters for AaHPPR. Means of 9 determinations ± SD, using three independent enzyme expressions. The K_m- and V_{max}-value for all substrates is calculated from Michaelis-Menten equations. For β-HP, only the range up to 2 mM was evaluated due to substrate inhibition.

Varied substrate	Constant substrate 2	K_m [mM]	V_{max} [mkat/kg]	k_{cat} [s^{-1}]	k_{cat}/K_m [$mM^{-1}s^{-1}$]
pHPP	NADPH	1.79 ± 0.46	7.80 ± 0.67	0.29 ± 0.03	0.18 ± 0.04
NADPH	pHPP	0.090 ± 0.03	15.84 ± 2.33	0.58 ± 0.09	6.86 ± 1.63
NADH	pHPP	0.39 ± 0.35	4.02 ± 1.41	0.15 ± 0.05	0.55 ± 0.33
β HP	NADPH	0.43 ± 0.16	3351.0 ± 1001.3	123.59 ± 36.93	305.86 ± 95.22

To determine kinetic data for pHPP and β-HP, the co-substrate NADPH was added at saturation (1 mM), for determination of the values for NADPH and NADH 25 mM pHPP was added. pHPP was tested at concentrations of up to 25 mM, resulting in a K_m-value of 1.79 ± 0.46 mM and V_{max} of 7.80 ± 0.67 mkat/kg, resulting in a catalytic efficiency of 0.18 ± 0.04 $mM^{-1}s^{-1}$. Hydroxypyruvate, tested in concentrations up to 6 mM, was accepted with much higher affinity (K_m 0.43 ± 0.16 mM) and was rapidly converted (V_{max} 4.02 ± 1.41 mkat/kg), resulting in a catalytic efficiency of 305.86 ± 95.22 $mM^{-1}s^{-1}$. However, it should be noted that at concentrations exceeding 2 mM a steady decrease in reaction rate is observed, probably due to substrate inhibition. This would imply that the actual K_m value should be considered higher. NADPH exhibited both higher affinity (K_m 0.090 ± 0.03 vs. K_m 0.39 ± 0.35 mM) and higher V_{max} (15.84 ± 2.33 vs. 4.02 ± 1.41 mkat/kg) compared to NADH, which was reflected in an approximately 12-fold higher catalytic efficiency (Table 8). As mentioned above, the photometric method reached its limits when measuring NADPH, NADH and β-HP. To test whether the evaluation of the complete incubation time leads to large deviations, an evaluation for the first 100 s was performed for NADPH. The results, K_m 0.06 ± 0.01 mM and V_{max} 18.31 ± 0.56 mkat/kg, are comparable to the values already mentioned for NADPH and are consistent with the expectation that affinity and maximum reaction rates are likely to be somewhat higher. The decision for this mode of evaluation should therefore not have a significant influence on the validity of the results.

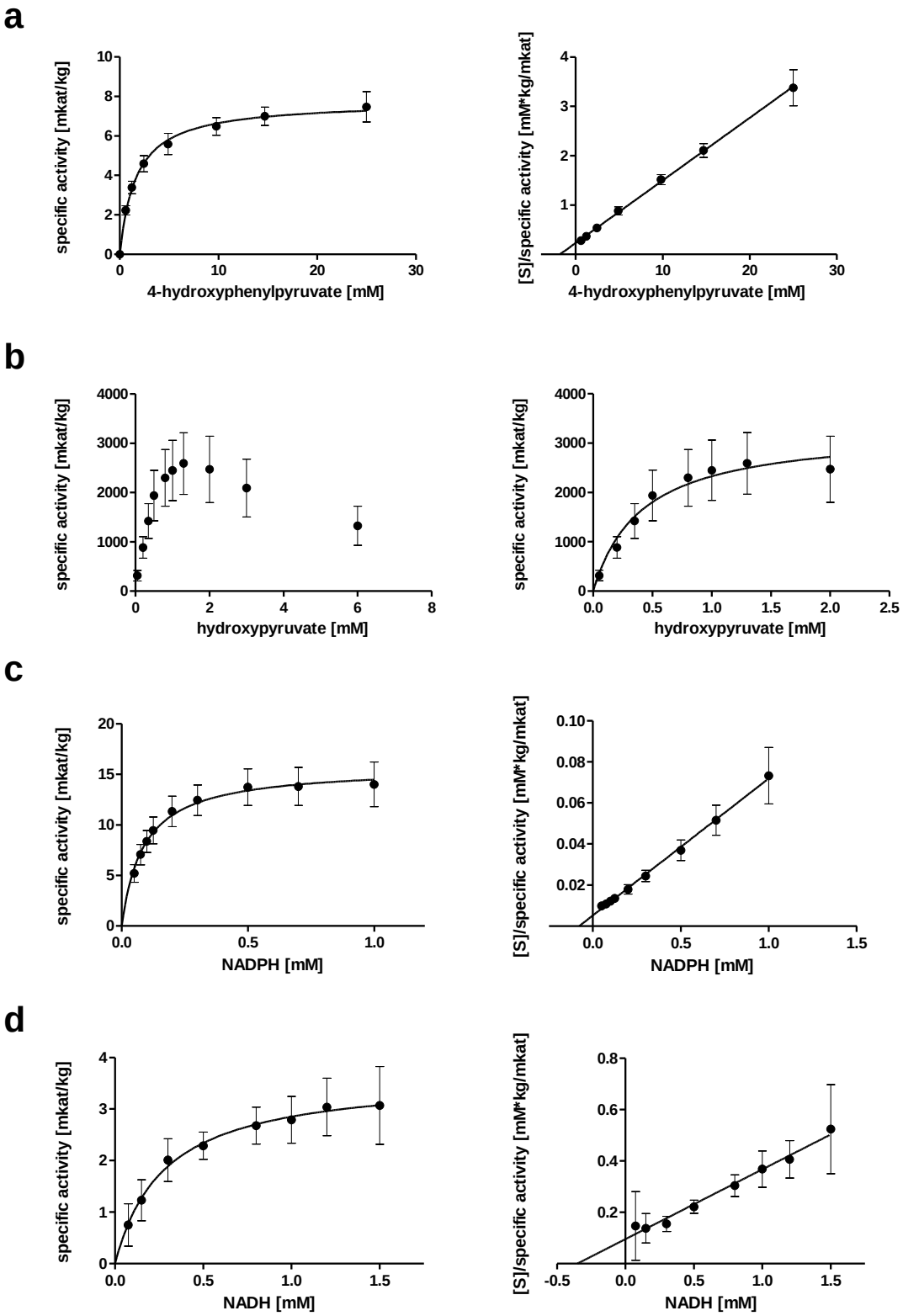


Figure 40 Plots of AaHPPR activities for different (co-)substrates. a 4-hydroxyphenylpyruvate with NADPH **b** hydroxypyruvate with NADPH **c** NADPH with 4-hydroxyphenylpyruvate **d** NADH with 4-hydroxyphenylpyruvate. For **a**, **c** and **d** Michaelis-Menten plot and Hanes-Woolf plot are shown. **b** only depicts the Michaelis-Menten plot, once shown for concentrations of up to 6 mM (left) and evaluated for a peak concentration of 2 mM. Means of 9 determinations ± SD, using three independent enzyme preparations.

2.7 Verification of the stereospecificity of AaHPPR-dependent pHPP reduction

Since HPPR is characterised as D-isomer-specific 2-hydroxyacid dehydrogenase, the stereospecificity of the reaction had yet to be verified. For this purpose, a completed enzyme assay was separated via TLC, the areas corresponding to the pHPL formed were scraped out and extracted. After a pre-purification on a non-chiral column, a comparison was made on the chiral column against the commercially available L-isomer and the racemate. The results (Fig. 41) demonstrated that only the D-isomer is formed by AaHPPR.

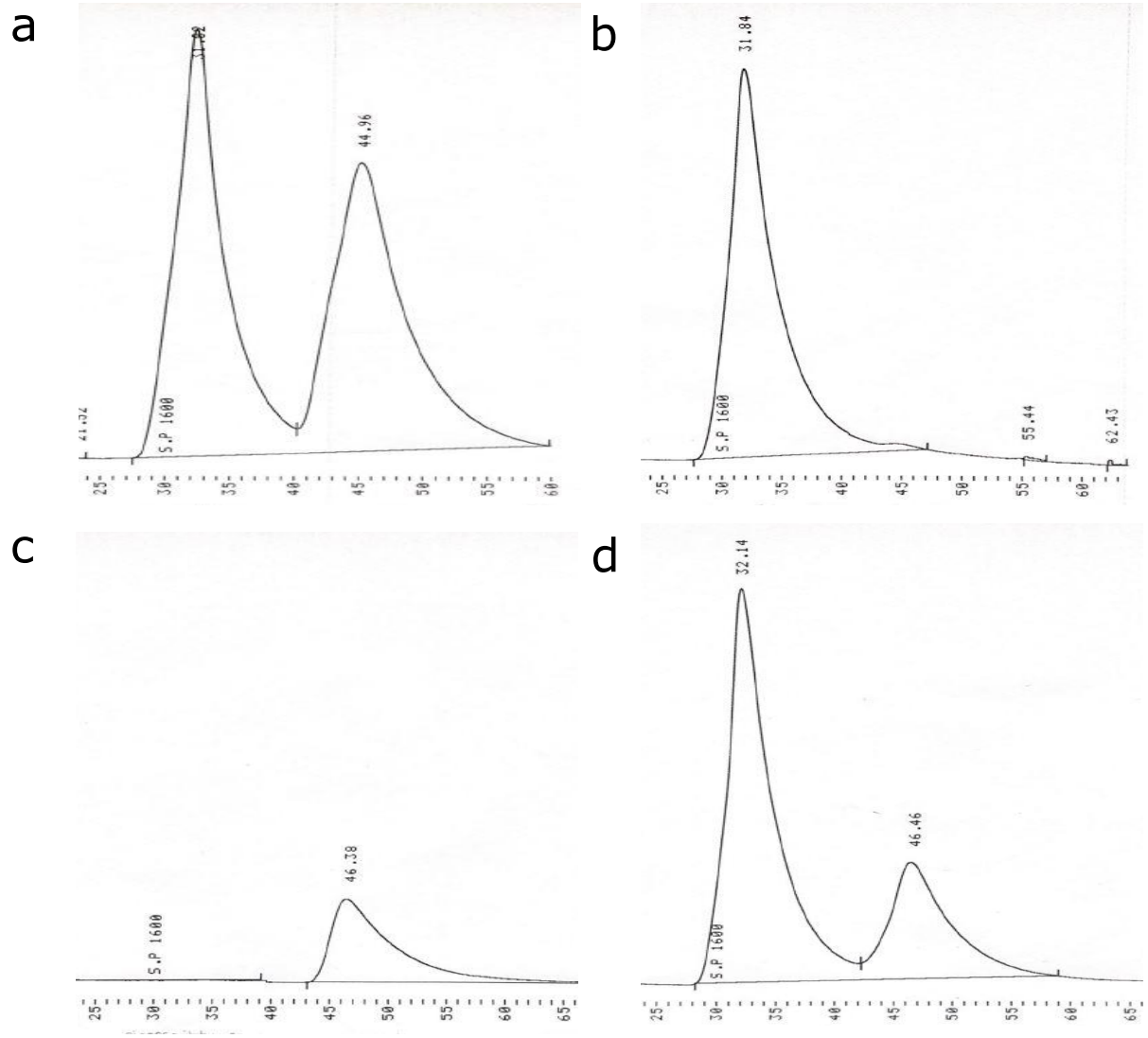


Figure 41 Determination of HPPR stereospecificity through chiral HPLC. Chromatograms of DL-pHPL (a), L-pHPL (b), pHPL product of heterologously expressed AaHPPR (c) and blend of b and c (d). No commercial D-pHPL standard was available, thus the D-enantiomer specificity can be confirmed by exclusion of the L-form.

2.8 Substrate acceptance and product verification

The disadvantage of photometric measurements in contrast to chromatographic methods is the invisibility of the analyte. In particular, because in this case the measurement is indirect, recording the decrease in NADPH concentration. Whether this decrease takes place via other influences and no product at all or some different product is formed, is not visible during the measurement. To identify the reaction products, photometric measurements were performed for 4-hydroxyphenylpyruvate (pHPP), 3,4-dihydroxyphenylpyruvate (DHPP), phenylpyruvate (PP), pyruvate (P) and hydroxypyruvate (β-HP) which were then processed and analysed by LC-MS. All substrates were thereby tested a second time using empty vector to prove the involvement of AaHPPR. The acceptance of all substrates investigated could be demonstrated. While pHPP, 4H3MPP, PP and P show an approximately linear decrease of absorption, this is not the case for DHPP and β-HP (Fig. 42). LC-MS analysis detected the formation of the corresponding products of the four aromatic substrates, whereas lactate (product of pyruvate) and glycerate (product of hydroxypyruvate) could not be detected (Fig. 43). Commercial standards were used for comparison, with the exception of 3,4-dihydroxyphenyllactate, which was prepared by hydrolysis of RA.

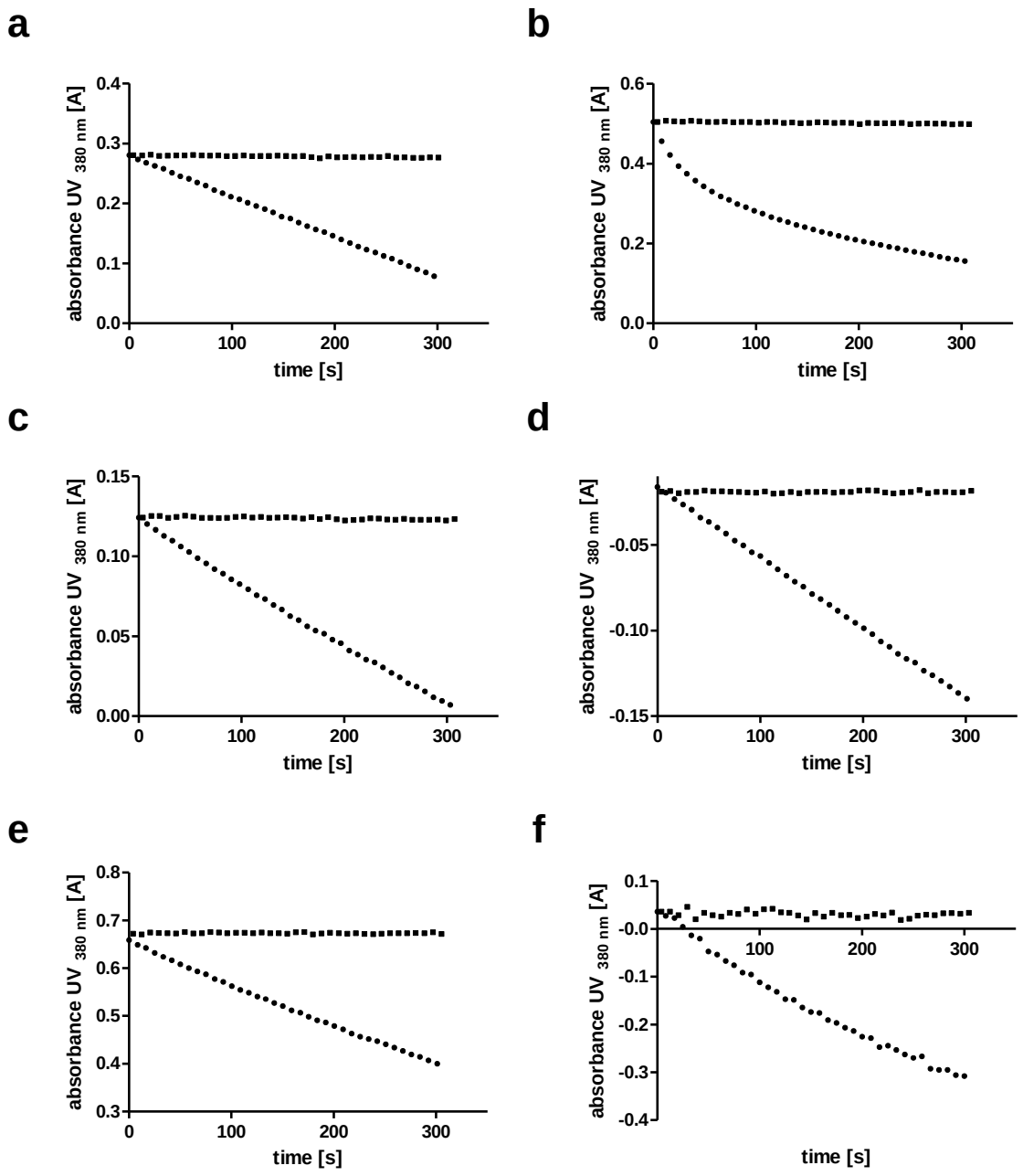


Figure 42 Exemplary reactions of various tested potential substrates (12.5 mM) with purified AaHPPR (circles) and empty vector control (squares). a 4-hydroxyphenylpyruvate **b** 3,4-dihydroxyphenylpyruvate **c** phenylpyruvate **d** pyruvate **e** 4-hydroxy-3-methoxyphenylpyruvate **f** hydroxypyruvate**.** This is a qualitative analysis. The curves of 3,4-dihydroxyphenylpyruvate and hydroxypyruvate are no indicator for the reaction rate, since linearity is not given.

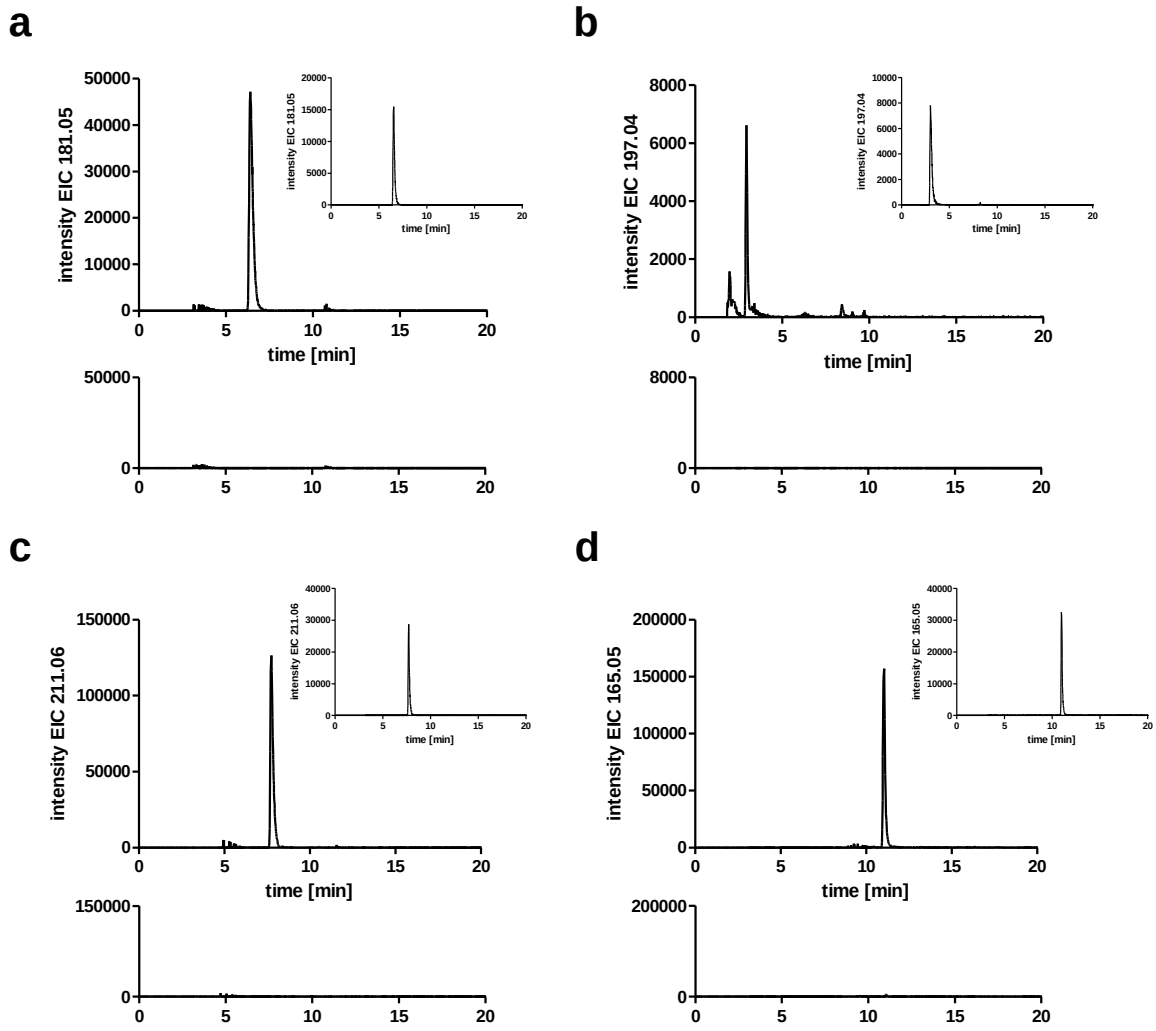


Figure 43 Detection of the product formed by LC-MS after reaction of AaHPPR with various substrates. a Reaction of pHPP results in detectable 4-hydroxyphenyllactate (m/z 181.05) **b** Reaction of DHPP results in detectable 3,4-dihydroxyphenyllactate (m/z 197.04) **c** 4H3MPP results in detectable 4-hydroxy-3-methoxyphenyllactate (m/z 211.06) **d** PP results in detectable phenyllactate (m/z 165.05). The respective comparative analysis with empty vector is shown below the chromatogram and a standard is shown (inset).

3 Hydroxyphenylpyruvate reductase 2 (HPPR2) and hydroxypyruvate reductase (HPR1)

3.1 RNA extraction and cDNA synthesis

RNA from *Anthoceros agrestis* suspension-cultured cells was extracted using acid guanidinium thiocyanate-phenol–chloroform extraction according to Chomczynski and Sacchi (1987). After reconstitution of dried RNA with water, integrity was checked through agarose gel

electrophoresis and concentrations at 260 nm as well as the 260/280 ratio were measured photometrically. Fig. 44 and Table 9 show the results of an extraction of *Anthoceros agrestis* cells after a three-day growth phase in CBM medium. Only sample 1 showed distinct ribosomal subunit bands on the agarose gel, i.e. 28S and 18S and faint 5S rRNA, and contained a sufficient concentration of RNA with good 260/280 ratio (see chapter III.1.1). Accordingly, this sample was used for cDNA synthesis.

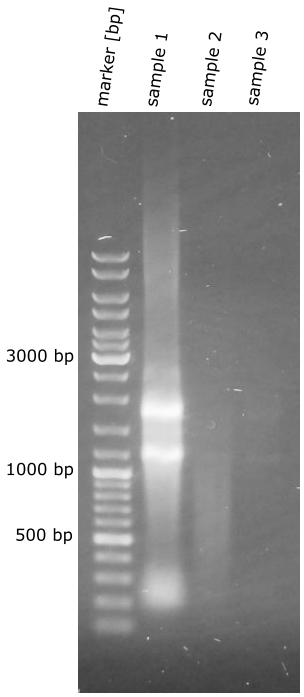


Figure 44 Extracted RNA samples from *Anthoceros agrestis*. Cells were extracted from a three-day old suspension cell culture (from CBM medium). Agarose gel electrophoresis with ethidium bromide staining. Distinct ribosomal subunit bands indicate RNA integrity only for sample 1.

Table 9 Concentration and purity of *Anthoceros agrestis* RNA extraction

RNA sample	Concentration [ng/µl]	260/280 ratio
Sample 1	308.2	1.89
Sample 2	66.6	1.66
Sample 3	8.0	- - -

3.2 Amplification of cDNA encoding HPPR2 and HPR1 from *Anthoceros agrestis*

Based on the now available blast interface (www.hornworts.uzh.ch/en/Blast.html) for the three hornwort genomes *A. agrestis* Bonn, *A. agrestis* Oxford and *A. punctatus* (Szövényi et al. 2015), it was checked whether there are further variants of a H(P)PR (Table 10). For that, the putative sequences (NM_001198420.1, NM_106636.3, BT011735.1) derived from *Arabidopsis thaliana* gene loci annotated by Timm et al. (2011) as HPR1, HPR2, and HPR3 (At1g68010, At1g79870, At1g12550) were translated into the corresponding amino acid sequence and the *Anthoceros* genomes were searched for resemblances. A counterpart with 76% identity was found for HPR1 and a sequence with 58.2% identity for HPR2. Another scaffold turned out to be the already amplified and characterised HPPR (two differing nucleotides). A unique sequence with high identity for HPR3 could not be identified. Based on the nucleotide sequence, primers were designed for HPR1 and HPR2 targeting the full-length sequence. The latter was renamed as HPPR2, due to its much higher similarity to HPPR (compare Fig. 46).

Table 10 Putative H(P)PR scaffolds identified in *A. agrestis* through protein blast of HPR1 and HPR2 sequences.

	Scaffold	Database	Resemblance	Information
1	AagrOXF_evm.model.utg000091l.603.2 Identity: 75.9% E-value: 0	*A. agrestis* Oxford	HPR1	
2	AagrOXF_evm.model.utg000017l.72.1 Identity: 58.1% E-value: 3.20E-127	*A. agrestis* Oxford	HPR2	
3	AagrOXF_evm.model.utg000061l.135.1 Identity: 55.4% E-value: 4.27E-109	*A. agrestis* Oxford	HPR2	Identical to HPPR
4	AagrOXF_evm.model.utg000009l.178.1 Identity: 63.2% E-value: 2.26E-46	*A. agrestis* Oxford	HPR2	Putative fragment of 2

Products of the expected CDS size of 1215 bp (appendix 1.6) for HPPR2 or 404 aa and 1284 bp (appendix 1.7) or 427 aa for HPR1 were successfully amplified (Fig. 45). However, ligation in pDrive was not successful in several attempts. Therefore, it was introduced directly into pET-15b. Since this does not allow preselection via blue-white screening, the success rate was very low. One out of 21 colonies (HPR1) or one out of 24 colonies (HPPR2) contained the respective target gene sequence.

For HPR1, there were matches of 100% identity (427/427 aa) with *Anthoceros agrestis* Oxford strain, and 98% (422/428 aa) for *Anthoceros punctatus*, including a 1 aa gap. The comparable

Anthoceros agrestis Bonn sequence did not fully cover AaHPR1. AaHPPR2 matched 100% identity (404/404 aa) with *Anthoceros agrestis* Oxford and Bonn and 98% identity (394/404 aa) with *Anthoceros punctatus* (https://www.hornworts.uzh.ch/en.html, Szövényi et al. 2015).

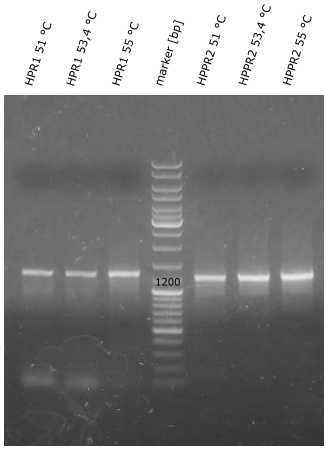


Figure 45 PCR Amplification of AaHPR1 and AaHPPR2. Full-length gradient PCR targeting the 1284 bp CDS of *AaHPR1* and 1215 bp CDS of *AaHPPR2*. Agarose gel electrophoresis with ethidium bromide staining.

Blastp analysis of the translated aa sequence of AaHPR1 showed high overall similarity to other HPR sequences with percent identity consistently above 75 for all 100 significant alignments found (Table 11). The highest identities were found for the mosses *Ceratodon purpureus* and *Physcomitrium patens*, the latter annotated as glycerate dehydrogenase-like, synonymous for hydroxypyruvate reductase due to the reversibility of the reaction, and the liverwort *Marchantia polymorpha*. The sequence of the vascular plant *Handroanthus impetiginosus* is annotated as glyoxylate/hydroxypyruvate reductase, while others, especially in useful plants as *Capsicum annuum*, *Camellia sinensis*, *Macadamia integrifolia* and *Ananas comosus* the sequences are annotated as glycerate dehydrogenase. The sequence from *Chenopodium chinoa* and *Sesamum indicum* is referred to as hydroxypyruvate reductase with indication of being peroxisomal. Additionally, Pairwise Sequence Alignment (EMBOSS Needle, protein) shows that AaHPR1 has 68.4% identity and 79.9% similarity to HPR from *Coleus blumei* (ABL10359.1).

Table 11 Blastp results (27/07/2022). Comparison of AaHPR1 aa sequence with several selected species of interest.

Organism	Description	Accession	Identity	E value
Ceratodon purpureus	hypothetical protein M758_8G110400	KAG0608497.1	83.95%	0.0
Physcomitrium patens	glycerate dehydrogenase-like	XP_024356601.1	82.78%	0.0
Marchantia polymorpha	hypothetical protein MARPO_0015s0044	PTQ45233.1	83.85%	0.0
Capsicum annuum	glycerate dehydrogenase	XP_016539968.1	77.89%	0.0
Handroanthus impetiginosus	glyoxylate/hydroxypyruvate reductase	PIN04558.1	76.84%	0.0
Camellia sinensis	glycerate dehydrogenase	XP_028056183.1	77.11%	0.0
Macadamia integrifolia	glycerate dehydrogenase	XP_042506798.1	76.50%	0.0
Ananas comosus	glycerate dehydrogenase	XP_020085142.1	77.84%	0.0
Chenopodium chinoa	glycerate dehydrogenase HPR, peroxisomal-like	XP_021743047.1	76.58%	0.0
Sesamum indicum	glycerate dehydrogenase HPR, peroxisomal	XP_011092968.1	76.05%	0.0

Blastp analysis of the translated aa sequence of AaHPPR2 (Table 12) showed high identity with a sequence from the moss *Sphagnum magellanicum*, the same protein with highest identity to AaHPPR, followed by the lycophyte *Selaginella moellendorffii* which is annotated as hydroxyphenylpyruvate reductase. A sequence from the moss *Physcomitrium patens* was also issued by the search, but with a lower identity. Seed plants include *Mentha x piperita*, a Lamiaceae, and other useful plants such as *Chenopodium chinoa*, *Capsicum annuum*, *Cucumis sativus*, *Solanum tuberosum* and *Olea europaea*, all of which are listed as hydroxyphenylpyruvate reductases except for *Capsicum annuum*, annotated as glyoxylate/hydroxypyruvate reductase. Overall (100 sequences), all percent identities ranged around 58-63%. Additionally, Pairwise Sequence Alignment (EMBOSS Needle, protein) shows that AaHPPR2 has 46.9% identity and 59.5% similarity to HP(P)R from *Coleus blumei* (CAD47810.2).

Table 12 Blastp results (27/07/2022). Comparison of AaHPPR2 aa sequence with several selected species of interest.

Organism	Description	Accession	Identity	E value
Sphagnum magellanicum	hypothetical protein CY35_10G030100	KAH9549638.1	63.46%	7E-133
Selaginella moellendorffii	hydroxyphenylpyruvate reductase	XP_002978650.1	63.43%	2E-127
Mentha x piperita	hydroxyphenylpyruvate reductase	AVZ47166.1	62.01%	4E-121
Chenopodium chinoa	hydroxyphenylpyruvate reductase	XP_021731988.1	62.66%	1E-120
Capsicum annuum	putative glyoxylate/hydroxypyruvate reductase A HPR2-like	KAF3649663.1	61.36%	5E-120
Cucumis sativus	hydroxyphenylpyruvate reductase	XP_004141766.3	61.69%	4E-119
Solanum tuberosum	PREDICTED: hydroxyphenylpyruvate reductase-like	XP_006343076.1	60.71%	9E-119
Olea europaea var. sylvestris	hydroxyphenylpyruvate reductase	XP_022848529.1	61.69%	9E-119
Papaver somniferum	hydroxyphenylpyruvate reductase-like	XP_026448790.1	60.39%	1E-117
Physcomitrium patens	hydroxyphenylpyruvate reductase-like	XP_024392873.1	58.33%	8E-116

To prepare for enzyme expression, the pET-15b plasmids harbouring the AaHPR1 and AaHPPR2 CDS were introduced into *E. coli* SoluBL21 (Amsbio). The protein mass, including the *N*-terminal His-tag sequence (His_6), was predicted to be 47.79 kDa for AaHPR1 and 45.78 kDa for AaHPPR2.

3.3 Comparison of AaHPPR, AaHPPR2 and AaHPR1

Although all three enzymes, i.e. AaHPR1, AaHPPR and AaHPPR2, are D-isomer-specific 2-hydroxyacid dehydrogenases, one can expect significant differences in aa sequence especially when comparing the HPR (EC 1.1.1.29) and H(P)PR (EC 1.1.1.81) type. Assuming that the peroxisomal enzymes are NADH-preferring and the cytosolic H(P)PRs are NADPH-preferring enzymes with an extended substrate spectrum (Kleczkowski et al. 1988; Julliard and Breton-Gilet 1997; Janiak 2007; Xu et al. 2018), this is quite conceivable and is supported by the alignment shown in Fig. 46. Nevertheless, HPRs and HPPRs share important family-specific

characteristics. One recognisable common feature is the NAD(P)H+H^+/$NAD(P)^+$ binding site, which is indicated by the GxGxxG(x_{17-18})D template, with the aspartate (D) residue primarily reported for NADH-dependent members of the D-isomer-specific 2-hydroxyacid dehydrogenases (Popov and Lamzin 1994; Stoll et al. 1996; Tishkov and Popov 2004), which is consistent with the HPRs shown in the alignment. Janiak et al. (2010) reports of Ser174 present at this position for *Coleus blumei* H(P)PR which is supported by all HPPRs shown here, with exception of AaHPPR, which is solely featuring glycine. In addition, Janiak et al. describe the active site of CbH(P)PR, which features three amino-acid residues, Arg232, His279, and Glu261, the latter stabilising His279 via a hydrogen bond. These three amino acids are known from other dehydrogenases in connection with enzymatic function and are present in all representatives in the alignment shown (Janiak et al. 2010, and literature cited therein). The overlength of the AaHPPR2 and the HPR1 in comparison to the enzymes of higher plants as well as to AaHPPR is also recognisable. The ORF of both enzymes contains a start codon that would also allow a shorter CDS in the size range of the comparative enzymes. As with AaTAT, the potential starting points of the translation were examined for Kozak motifs (see III.1.2). The criteria are shown in Table 2. For AaHPR1, the corresponding scaffold from *Anthoceros agrestis* Oxford (AagrOXF_evm.model.utg000091l.603.2 cDNA scaffold; https://www.hornworts.uzh.ch/en.html) was used. Of the *N*-termini MQALR (TTCC|ATG|CA), MAGYR (AGCT|ATG|GC) and MPGSR (GCCG|ATG|CC) examined, only number two meets the consensus for green plants (Hernández et al. 2019). The other selective requirements (lower plants and non-angiosperms; Joshi et al. 1997) were only partially met. The other two were eliminated due to the absence of G in position 4 and the presence of T/C in position -3. For AaHPPR2, the corresponding scaffold from *Anthoceros agrestis* Oxford (AagrOXF_evm.model.utg000017l.72.1 cDNA scaffold) was examined. Of the *N*-termini MCACF (ATTG|ATG|TG), MVATG (AGCC|ATG|GT) und MLADL (GCCC|ATG|CT) only number two met the general criteria for green plants. Additional properties were met in parts, with AaHPPR2 deviating from the C in position 5 compared to the previous candidates. Since AaHPPR offers only one reasonable possibility and therefore can serve as positive example, the region around the *N*-terminus MAAVE (GGCG|ATG|GC) was examined after reconciling with the 5'-RACE result using the corresponding sequence from *Anthoceros agrestis* Oxford (AagrOXF_evm.model.utg000061l.135.1 genome). As expected, it met the criteria for green plants and some of the additional properties.

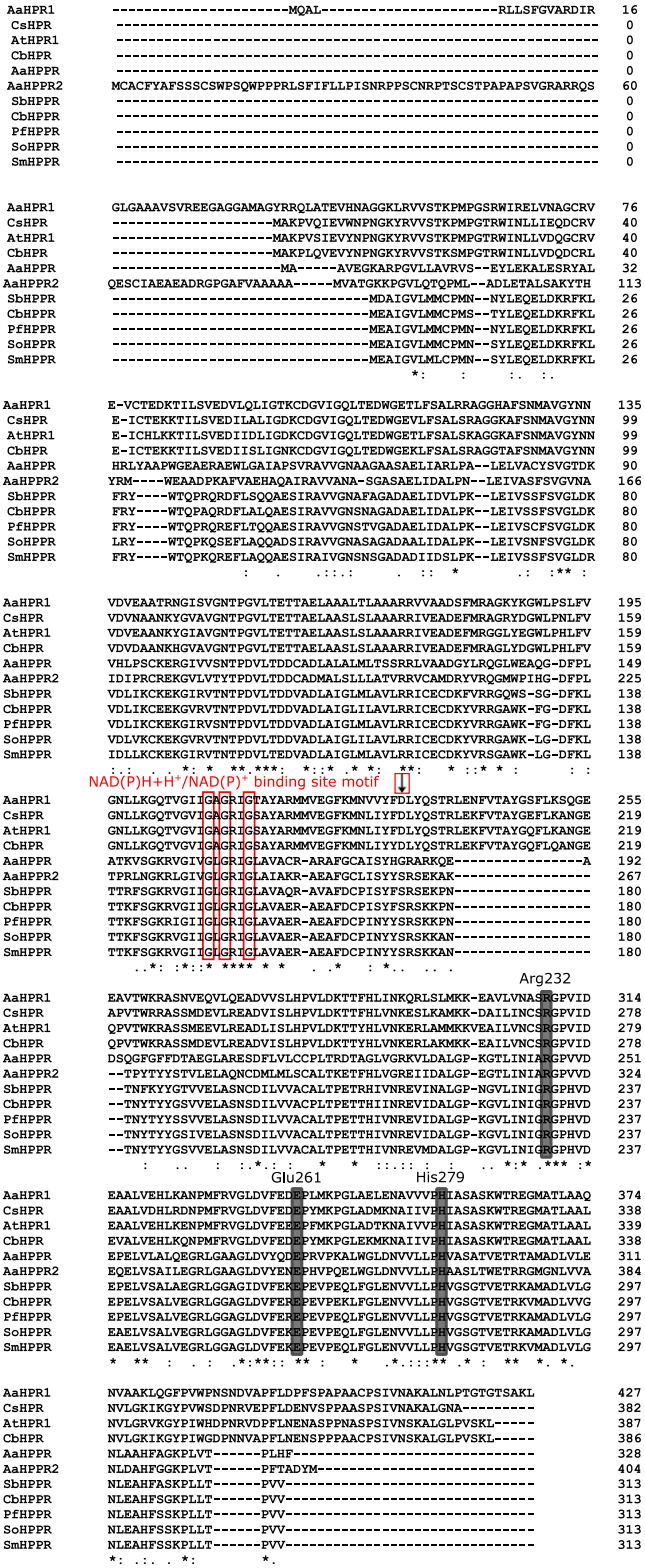


Figure 46 Alignment of some plant hydroxypyruvate and hydroxyphenylpyruvate reductases together with AaHPR1, AaHPPR and AaHPPR2. *Cucumis sativus* (CAA32764.1), *Arabidopsis thaliana* (NP_001185349.1), *Coleus blumei* (ABL10359.1 for HPR1 and CAD47810.2 for H(P)PR), *Scutellaria baicalensis* (AHB63232.1), *Perilla frutescens* (ADM48810.1), *Saliva officinalis* (AGE44245.1), *Salvia miltiorrhiza* (AAZ67354.1). Motifs are colour-coded, important amino acids are annotated according to CbH(P)PR.

3.4 Heterologous expression of HPPR2 and HPR1 from *Anthoceros agrestis*

The *E. coli* SoluBL21 cells containing pET-15b harbouring AaHPPR2 or AaHPR1 were incubated for at least 16 h at 25 °C in TB medium after induction with 1 mM IPTG. The isolated crude extracts were further subjected to metal chelate chromatography and buffer exchange and were analysed by SDS-PAGE and Western blotting (Fig. 47).

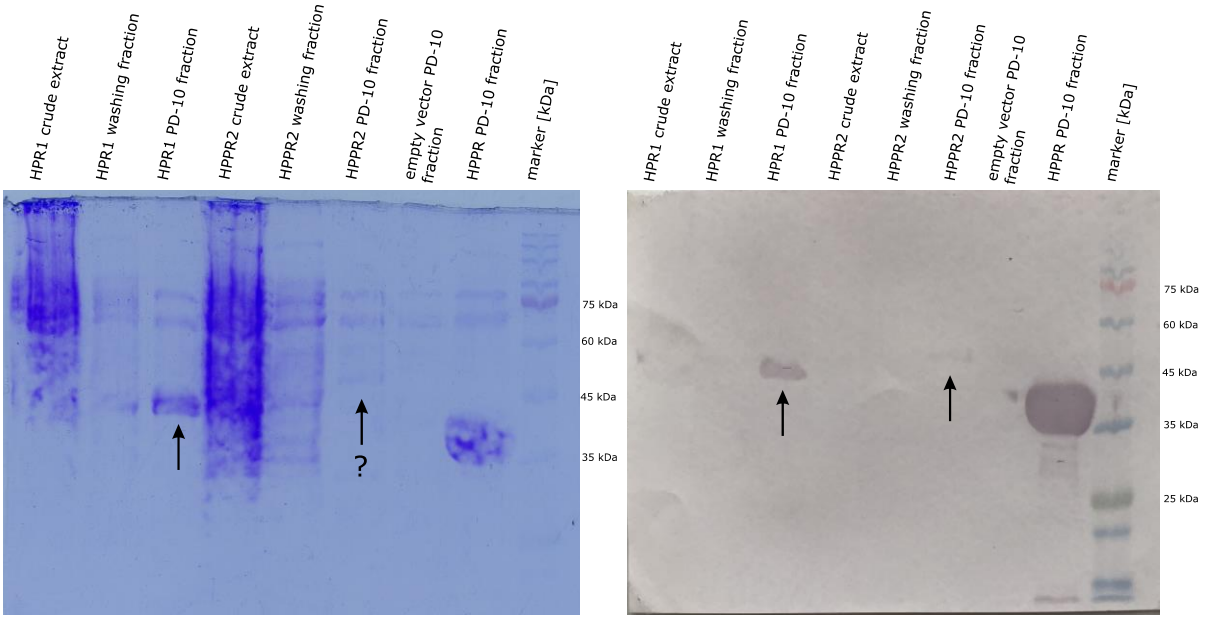


Figure 47 Purification and expression analysis of AaHPR1 and AaHPPR2. SDS-PAGE (left) and Western blot (right) with crude extract, washing fraction and PD-10 fraction. HPR1 shows a distinct band, whereas HPPR2 is only weakly expressed and can only be reliably identified by Western blot. AaHPPR is included as positive and empty vector as negative control.

AaHPPR2 and AaHPR1 were evaluated for activity in initial enzyme assays. AaHPPR2 showed distinct activity using 1 mM pHPP and 1 mM NADPH and weaker activity using the same concentration of β-HP. AaHPR1 exhibited high activity with 1 mM β-HP but activity with 1 mM pHPP was not exceeding empty vector control. It must be noted that this first activity test for AaHPR1 was carried out using 1 mM NADPH as co-substrate. AaHPR1 assays with 0.5 mM hydroxypyruvate and 1 mM NADPH or NADH, respectively, detected a significantly higher enzyme activity for the latter.

4 Culture characterisation and determination of TAT, HPPR, HPPR2 and HPR1 transcript abundances

The objective of the experiments was to determine the rosmarinic acid content *in Anthoceros agrestis* over a cultivation period of 14 days and to carry out an expression analysis of the enzymes molecularly or biochemically investigated and potentially involved in rosmarinic acid biosynthesis. Although not considered as involved in RA biosynthesis, HPR1 was also included. Furthermore, the medium parameters such as pH value, conductivity, sugar content and biomass growth were also recorded.

4.1 Medium parameter, growth and rosmarinic acid content

The pH value was about pH 6 immediately after inoculation and dropped to pH 4 by day one. It then rose steeply to reach pH 8 at day 3 to day 4 and stabilised for the remaining days. The fresh weight at start was 2.8 g per flask and continuously increased until day 12 to 11.5 g, where it remained stable for the following two days. The sugar content, 1.3% after inoculation, dropped sharply until day seven, where it remained at 0.2% for two further days, then dropped to 0.1%, where it remained stable for the remainder of the characterisation (Fig. 48). Conductivity was 4.5 mS/cm at day 0 and dropped to 2.2 mS/cm at day 11 and increased the remaining days to 2.6 mS/cm. In addition to the general medium parameters, the RA accumulation in the cells was determined. It increased steadily after inoculation, with the maximum reached on day 3 (4.8% of the cell dry weight). Over the next 3 days, the content dropped continuously to a level of about 3%, where it remained more or less stable for the rest of the observation period. The maximum reached during this period (3.6%) was on day 13, while the minimum (2.6%) was on day 10 (Fig. 48, for further information see Pezeshki et al. (2022)).

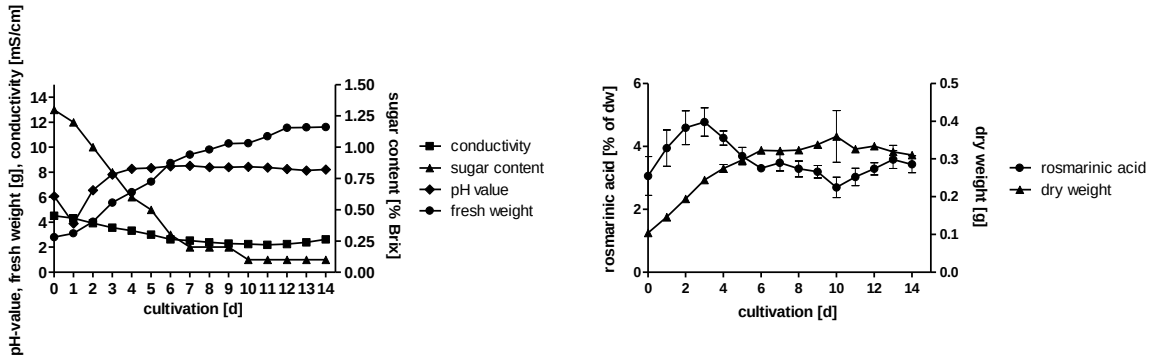


Figure 48 Culture characterisation of *Anthoceros agrestis* in CBM medium over 14 days. a Conductivity (squares), sugar content (triangles), pH value (diamonds) and fresh weight (circles) **b** Rosmarinic acid content (circles, n=6 ± SD) and dry weight (squares, n=3 ± SD) according to Pezeshki et al. (2022).

4.2 Expression analysis

RNA was extracted from 2x 50 mg fresh cell material per culture day. The integrity of the sample was checked by agarose gel electrophoresis and concentrations were determined photometrically (Fig. 49 and Table 13). After digestion with DNase of the better sample based on the 260/280 ratio, a phenol-chloroform extraction was performed and 0.5 µg RNA was reverse-transcribed into cDNA.

Table 13 Concentration and purity of extracted *Anthoceros agrestis* RNA

Day	Concentration [ng/µl]	A260/280 ratio / A260/230 ratio
0	926.7	1.87 / 2.11
1	1138.2	1.92 / 2.37
2	1599.8	1.98 / 2.46
3	1577.6	1.93 / 2.46
4	1289.6	1.91 / 2.51
5	859.5	1.89 / 2.10
6	1300.9	1.92 / 2.31
7	730.4	2.00 / 2.14
8	825.2	1.94 / 2.08
9	905.7	1.91 / 2.27
10	460.3	1.92 / 2.06

Table 13 Concentration and purity of extracted *Anthoceros agrestis* RNA after DNase digestion

Day	Concentration [ng/µl]	A260/280 ratio / A260/230 ratio
0	132.5	2.06 / 2.14
1	141.9	1.98 / 2.18
2	133.7	2.09 / 2.07
3	149.7	2.01 / 2.16
4	145.4	2.06 / 2.04
5	133.8	1.96 / 2.39
6	137.9	2.05 / 2.25
7	164.1	1.96 / 2.12
8	121.5	2.03 / 2.11
9	153.4	1.83 / 1.95
10	144.8	2.02 / 2.33

11	667.3	1.86 / 2.05		11	152.5	2.00 / 2.00
12	531.7	1.95 / 2.18		12	151.7	2.06 / 2.29
13	483.7	1.90 / 2.11		13	144.0	1.90 / 2.15
14	483.0	1.97 / 1.93		14	155.2	1.75 / 1.99

Before performing the actual qPCR experiment, all primer pairs, i.e. for TAT, HPPR, HPPR2 and HPR1 as well as those of the reference genes used, actin and serine threonine protein phosphatase 2a regulatory subunit (St-P 2a), were tested in standard PCR experiments. If distinct bands and no by-products were recognisable, the product was additionally sequenced. Otherwise, the primers used were adapted (appendix Fig. 56).

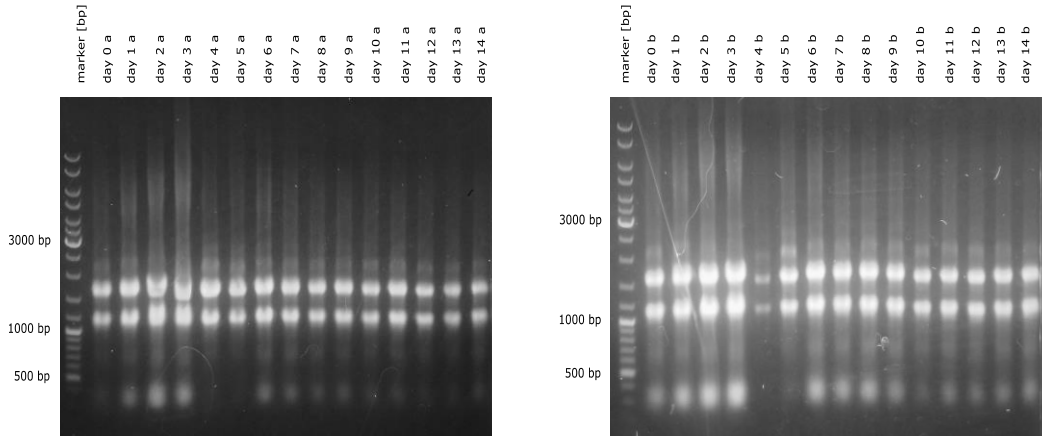


Figure 49 Extracted *Anthoceros agrestis* RNA before DNase treatment. For each cultivation day two samples were prepared and tested.

To enable relative quantification according to Pfaffl (2001), amplification efficiencies were determined for all primer pairs. *TAT*, *HPPR*, *HPR1* and *StP 2a* achieved efficiencies that were within the target range of 90-110%, i.e. 92.46%, 94.73%, 90.39% and 103.94%, respectively (Fig. 50). *HPPR2* and *Actin* missed the range with 86.58% and 82.29%.

Based on the calculated efficiencies, relative expression was determined normalised with both *actin* and *StP 2a* in relation to day 0 for all target genes (Fig. 51). Transcript abundance of *TAT*, normalised with *Actin*, increased until day 4, with a maximum spanning day 4 and 5. Thereafter, the expression decreases, forming a plateau from day 7 to 11 and finally falling back to the expression level of day 1. The transcript abundance of *TAT*, normalised with *StP 2a*,

largely corresponds to the mentioned observations. The first maximum is already reached on day 3 and the plateau is formed at the transcript level of day 1. The relative expression of *HPPR* (A*ctin*) and *HPPR* (*StP 2a*), however, is more different. The former shows an increasing relative expression up to day 7, where it forms a plateau and only drops significantly on day 14. The latter fluctuates from day 1 over the entire observation period with a downward trend. *HPPR2* (*Actin*) and *HPPR2* (*StP 2a*) resemble the curve of *TAT*. The former reaches its maximum from day 3 to day 6 and then declines continuously. This is also shown in *HPPR2* (*StP 2a*), but the measuring points in the range of day 2 to 6 do not show a distinct maximum. *HPR1* (A*ctin*) and *HPR1* (*StP2a*) show stable values over the first 3 days of culture, with a strong jump on day 4. Afterwards, a plateau forms for both normalisations, with a strong drop on day 14.

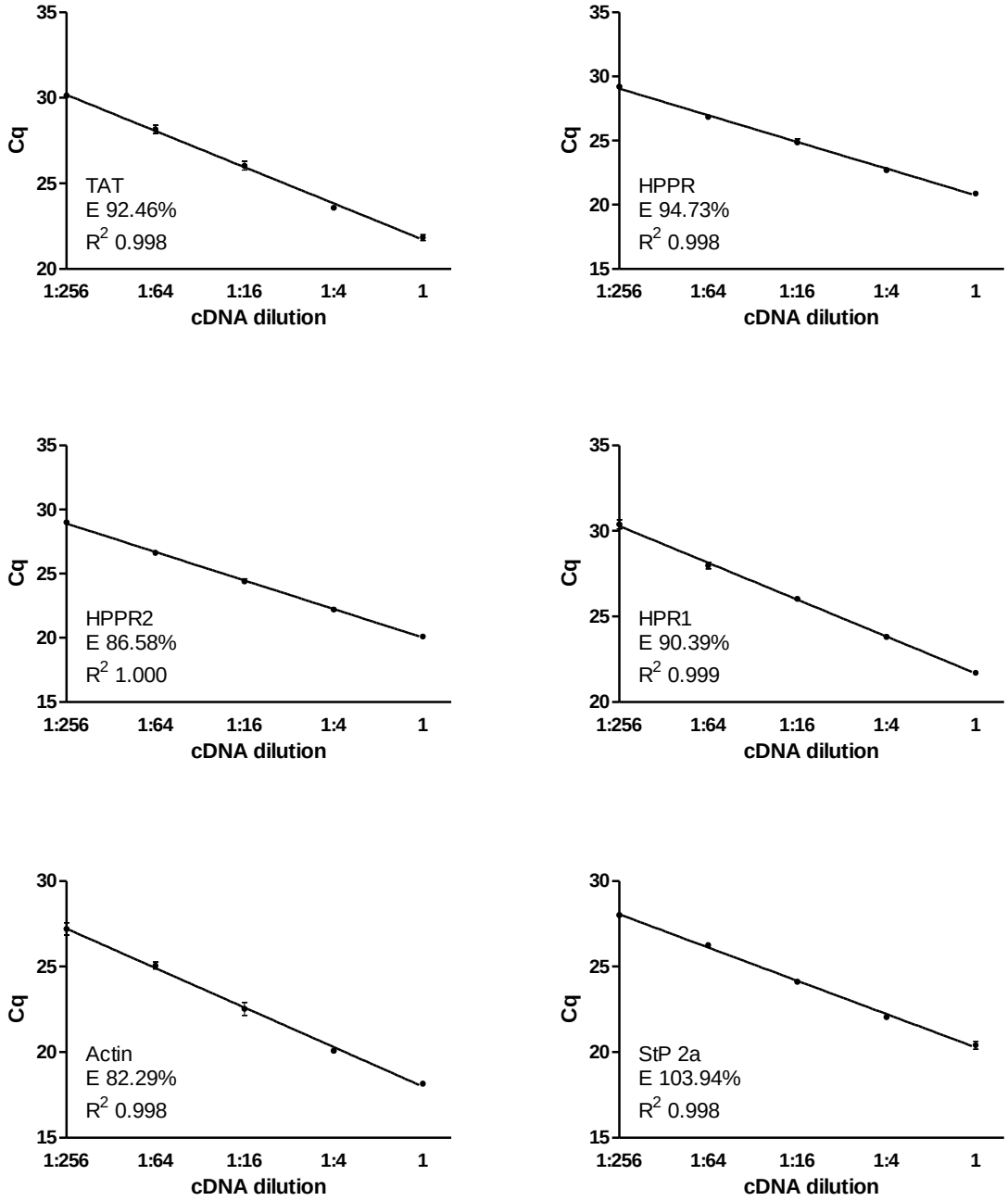


Figure 50 Determination of real-time PCR efficiencies of *TAT*, *HPPR*, *HPPR2*, *HPR1* and the reference genes *Actin* and *StP 2a*. For each primer pair dilutions of up to 1:256 were measured. The concentration for each primer used was 200 nM, with exception of HPPR2 and HPR1, where 150 nM was added. Means of 3 determinations ± SD. For the determination of *HPPR* and *HPPR2* efficiency a single measuring point was excluded due to formation of by-products.

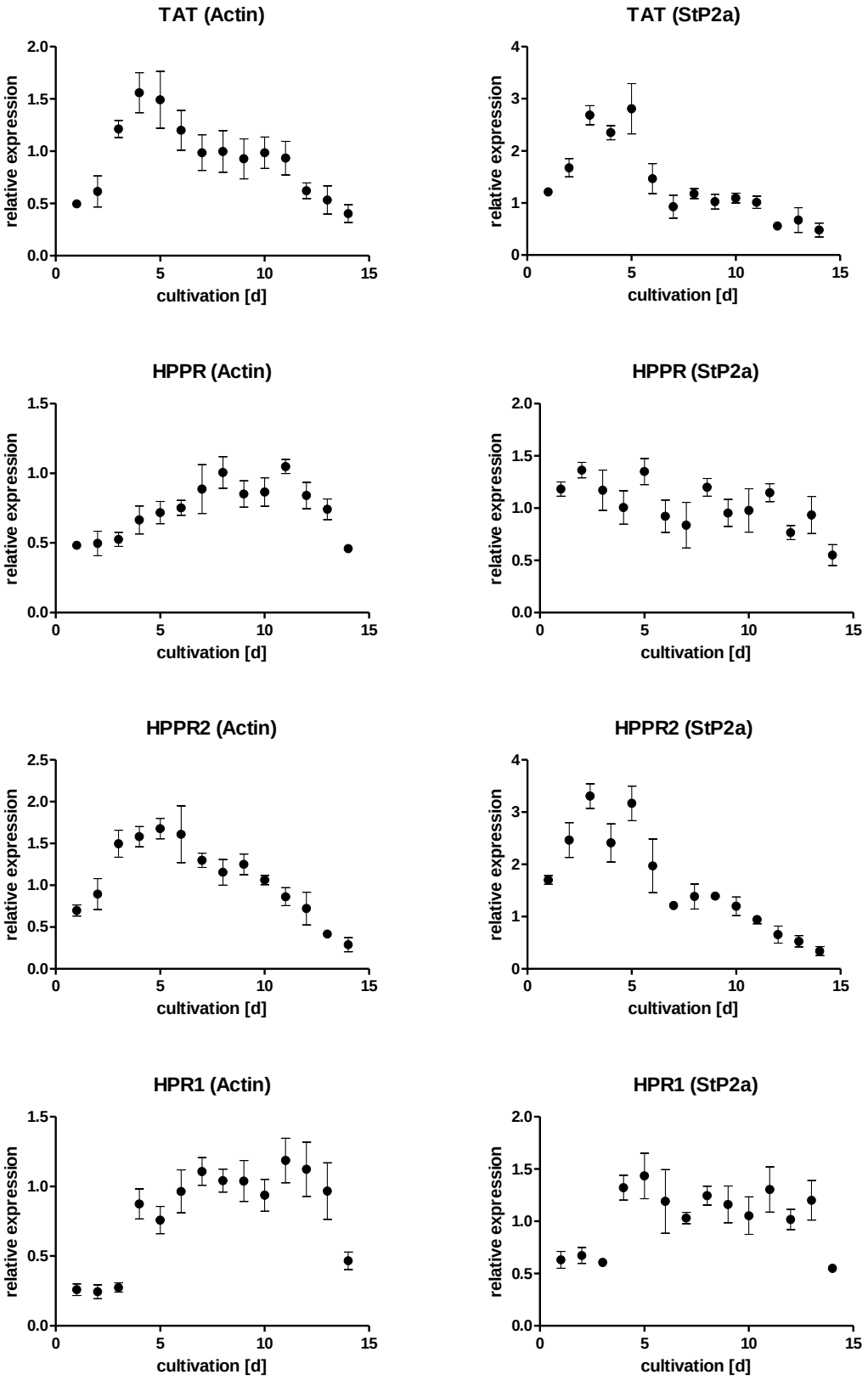


Figure 51 Quantitative real-time PCR analysis of *AaTAT*, *AaHPPR*, *AaHPPR2* and *HPR1*. For each target normalisation with the reference gene *actin* and *StP 2a* is shown. The fold-change (relative expression) was calculated according to Pfaffl (2001). Means of 4 determinations ± SD. In one case, an unwanted by-product was found, the respective value was excluded from calculation, see Appendix, Table 14.

IV Discussion

Rosmarinic acid (RA), formally an ester of caffeic acid and 3,4-dihydroxyphenyllactic acid, is widely distributed in the plant kingdom. While it has not yet been detected in streptophyte algae, it is common in terrestrial plants even in phylogenetically distant species, with *Anthoceros agrestis* being the most basal finding (Takeda et al. 1990; Petersen et al. 2009; Petersen 2013). The biosynthesis of rosmarinic acid as postulated for *Coleus blumei* - a dichotomous reaction with the caffeic acid moiety derived from L-phenylalanine with the intermediates cinnamic acid, 4-coumaric acid and 4-coumaroyl-CoA and the 3,4-dihydroxyphenyllactic acid moiety derived from L-tyrosine with the intermediates 4-hydroxyphenylpyruvate and 4-hydroxyphenyllactate (Petersen et al. 1993) - is generally accepted. To date, *Coleus blumei* is the only species in which all enzymes involved have been elucidated. The comprehensive work on *Melissa officinalis* (Weitzel 2009) and the important Chinese medical plant danshen or *Salvia miltiorrhiza* (Zhao et al. 2006; Huang et al. 2008a; 2008b; Song and Wang 2011; 2012; 2015) suggest that the biosynthetic pathway has a generality at least for Lamiaceae although the introduction of 4-hydroxyphenyllactic acid rather than the dihydroxy-derivative into the RAS reaction is sometimes doubted (Di et al. 2013; Zhou et al. 2018). So far, the interest has been almost exclusively in the Lamiaceae (Nepetoideae, see chapter I.5) and Boraginaceae (e.g. De-Eknamkul and Ellis 1987a, 1987b; Mizukami and Ellis 1991; Mizukami et al. 1993; Matsuno et al. 2002; Levsh et al. 2019). Research on the hornwort *Anthoceros agrestis* could shed light on the extent to which the current idea of RA biosynthesis is valid for genera earlier diverged than *Coleus blumei* (e.g. magnoliids, monocots, other eudicots, see Petersen 2013). Important advances in the biosynthesis of the central intermediate of phenylpropanoid compounds, 4-coumaroyl-CoA, have been achieved for *Anthoceros agrestis* in recent years with the characterisation of two PAL isoforms comparable to PALs from spermatophytes (Pezeshki et al. 2022), cinnamic acid 4-hydroxylase (Petersen 2003; Wohl and Petersen 2020a) and 4-coumarate CoA ligase (Wohl and Petersen 2020b). The results suggest that the formation of said intermediate could thus be homologous to vascular plants. The characterisation of enzymes involved in the formation of the 3,4-dihydroxyphenyllactic moiety starting from L-tyrosine have so far been limited to the use of the crude enzyme extract from plant material (Pezeshki 2016). Recent results on

the enzymes TAT and HPPR, including kinetic and qualitative parameters, as well as early results on HPPR2 and HPR1 from *Anthoceros agrestis* will be discussed below.

1 Sequence, alignment and phylogeny

The investigated aminotransferase, which was assumed to be an aromatic, more precisely a tyrosine aminotransferase (TAT) due to similarities in the sequence to *Coleus blumei* TAT, was successfully amplified. The 1395 bp or 464 aa ORF was targeted for the subsequent expression, although a shorter variant (1302 bp) would also be conceivable due to a second start codon (Busch and Petersen 2021). The aa sequence was clearly identified as TAT, featuring the invariant properties, i.e. the residues Lys296, Arg435 and Asp263 (Mehta et al. 1993, numbering according to AaTAT) as well as the PLP binding site and the homodimer interface, illustrated by the alignment of AaTAT with additional characterised plant TATs (Fig. 8). An interesting observation is the presence of Asn68 (N68), which seems associated with the preference for Tyr in *Trypanosoma cruzi* TAT (see Asn17; Blankenfeldt et al. 1999). In fact, the conserved amino acid sequence NPIR was not only found in *T. cruzi* TAT and AaTAT, but also in all representatives of basal TATs (*P. patens* TAT, *M. polymorpha* TAT, *S. moellendorffii* TAT, *K. nitens* TAT), but not in the characterised TATs from higher plants (Figs. 8 and 9), an observation for which there is currently no explanation. The alignment also displayed a *N*-terminal excess of AaTAT in comparison to TATs of vascular plants. Of the nine TAT sequences, seven (including two isoforms) are listed in the literature in connection to RA biosynthesis. These representatives, *Prunella vulgaris*, *Salvia miltiorrhiza*, *Perilla frutescens*, *Coleus blumei* are Lamiaceae, subfamily Nepetoideae and show strong similarity among each other, while TAT1 and TAT2 from *Scutellaria baicalensis*, subfamily Scutellarioideae, deviate in some areas and length, with the former better matching the enzymes of the Nepetoideae (Busch and Petersen 2021). Since two other families, Brassicaceae (*Arabidopsis thaliana*) and Papaveraceae (*Papaver somniferum*), families lacking RA accumulation, were also included, it cannot be assumed that the excess length of AaTAT in comparison to those enzymes is a subjective characteristic due to the selective choice in the alignment, i.e. focus on characterised TATs and exclusion of putative enzymes. When comparing the sequences of the Fabaceae *Glycine max* and *Medicago truncatula*, which were considered by some authors (Huang et al. 2008b; Lu et al. 2013a) for their respective alignments, there is also no *N*-terminal excess visible. This raises the question of whether the decision to select the larger CDS for

amplification and expression was the right one. The hornwort database with sequence information on *Anthoceros agrestis* (Oxford and Bonn strain) and *Anthoceros punctatus* (https://www.hornworts.uzh.ch/en.html; Szövényi et al. 2015), which became available shortly after enzyme characterisation, showed that the second start codon is retrievable in *Anthoceros agrestis* Oxford strain and *Anthoceros punctatus*, but not in the sequence of the Bonn strain. Additional confirmation in the draft genome of *Anthoceros angustus* (Zhang et al. 2020) supports its existence. Evidence of a *N*-terminal plastid signal peptide using SignalP-5.0 (Almagro Armenteros et al. 2019) was not found according to Busch and Petersen (2021) and SignalP-6.0 (Teufel et al. 2022). The sequence information in the hornwort database, however, brought a second aspect to light. For both *Anthoceros agrestis* strains, the TAT analogue was deposited with an ORF that included additional 83 *N*-terminal amino acids. This excess could be retrieved in the sequence information of *Anthoceros angustus* and fragmentarily in *Anthoceros punctatus*. Part of this sequence was accurately recovered in the 5' RACE product of the AaTAT, but there was hardly any agreement in the region of the upstream start codon. Again, no evidence of a signal peptide could be found using SignalP-6.0 (Teufel et al. 2022). The question of what corresponds to native CDS is not clearly resolvable from today's point of view. An alignment of the AaTAT (Fig. 9) with basal species showed a comparable length, most prominently represented by sequences of the moss *Physcomitrium patens* (Busch and Petersen 2021). Investigation of the underlying sequences (XM_024505539.1; XM_024505540.1) of the latter, revealed neither evidence of a potential excess nor a downstream start codon that would allow a suitable shorter variant. The fact that both the basal and the characterised TATs of the higher plant species and AaTAT are in a comparable order of magnitude, but no matching results are known for the huge excess found in the database, indicates the correct choice of the CDS. Screening of the nucleotide sequence for Kozak motifs confirmed this view. According to the current knowledge, no information on amplified and possibly characterised TATs from other basal species is yet known. The confirmation and characterisation can help to confirm or revise the understanding of AaTAT.

To get a more detailed picture of the classification of AaTAT, a phylogenetic tree (Fig. 10) was established (Busch and Petersen 2021). It demonstrated that more distantly related aminotransferases, i.e. other subgroup I as well as non-classified aminotransferases (tryptophan and prephenate aminotransferases), are not represented in the main branch. The bootstrap values of the base branches of this group were very poor, expectedly due to few

representatives, but the general separation from TATs was still recognisable. The phenylalanine aminotransferases, which can be found among the TATs are one exception. This can be attributed to the similarity of the two substrates. Several examples of double substrate acceptance are known in literature (compare chapter IV.2.1). As expected, AaTAT was found grouped with the basal putative TATs, which are, although uncharacterised and thus not clearly identifiable, almost certainly aromatic aminotransferases of the tyrosine/phenylalanine type. A reliable statement on the relationship of bryophytes and higher plants cannot be given based on the phylogenetic results, since the resolution in the section of the basal representatives is not strongly supported, indicated by low bootstrap values. This is particularly evident in the example of the charophyte alga *Klebsormidium nitens*, which fits in between the lycophyte *Selaginella moellendorffii* and the liverwort *Marchantia polymorpha*. *Anthoceros agrestis* TAT, in contrast, is resolved as sister to TAT from the moss *Physcomitrium patens*, which, according to current consensus (e.g. Puttick et al. 2018, compare chapter I.1), should, in association with the liverwort, constitute the setaphytes (Renzaglia and Garbary 2001). For assessment of the phylogeny, a more comprehensive alignment would be needed. However, this was not the focus of the analysis, i.e. the addition of further putative examples was omitted in favour of a good data set of classified TAT enzymes (corresponding publications are listed in chapter II.5). In some species, isoforms of TAT are already known. The alignment and phylogenetic tree comprise two isoforms for the species *Arabidopsis thaliana*, involved in biosynthesis of tocopherols studied in *Arabidopsis* single and double mutants, with TAT2 having a less important role (Riewe et al. 2012; Wang et al. 2019). In addition, a broader substrate spectrum and the probable favouring of the reverse reaction were reported for TAT2 (Prabhu and Hudson 2010; Wang et al. 2016). Also included are two isoforms of *Scutellaria baicalensis* TAT correlated with RA formation but with a pending proof of the distinct functions (Kim et al. 2014b). In addition, six further putative sequences in *Arabidopsis thaliana* may encode TAT, whereas only TAT1 and TAT2 have been investigated (Riewe et al. 2012). De-Eknamkul and Ellis (1987b) reported three TAT isoforms in *Anchusa officinalis*, where TAT1 is thought to be involved in RA biosynthesis, TAT2 may be involved in this as well, alternatively in plastoquinone, or even in L-tyrosine biosynthesis through transamination of pHPP, a metabolic pathway indicated so far only for legumes (Rubin and Jensen 1979; Schenck et al. 2015). TAT3, on the other hand, proved to be a general aspartate aminotransferase. Due to the occurrence in those exemplary species and partial

specialisation of the isoforms, the hornwort database was screened for potential candidates. For *Anthoceros agrestis*, no evidence of further TAT isoforms was found in each of the three databases (*Anthoceros agrestis* Bonn and Oxford strain and *Anthoceros punctatus*; Szövényi et al. 2015). The search produced only one result in match with the investigated AaTAT as well as hits with low similarity, i.e. less than 30%. Thus, assuming that the formation of the 3,4-dihydroxyphenyllactic acid moiety is analogous to RA biosynthesis in Lamiaceae, the involvement of AaTAT seems likely. However, since the biosynthesis of L-tyrosine and L-phenylalanine has not been elucidated for bryophytes, its involvement would also be conceivable (Busch and Petersen 2021).

HPPR was successfully amplified based on partial sequences provided by Maike Petersen. Although only gDNA was found to work as template, no introns were detected in the sequence obtained. Besides the CDS with 987 bp encoding 328 aa, two longer variants with 1032 bp or 1113 bp were indicated by some of the RACE results, but could be eliminated because of lacking consensus and through comparison with the hornwort sequence information (Szövényi et al. 2015). The highest match for AaHPPR was found in *Anthoceros agrestis* Oxford and Bonn strains with 99% identity. The difference in a single amino acid (asparagine for AaHPPR instead of serine) has been traced back to full-length amplification with a proofreading polymerase. Overall, however, the sequence transcribed with this enzyme was of considerably better quality than the variants transcribed without proofreading activity. An influence on the enzyme activity due to the deviating amino acid cannot be assumed, as this is located *N*-terminally far away from characterised motifs, i.e. NADPH binding site, or substantial amino acids. In the alignment, which includes HPRs and HPPRs, it can also be seen that asparagine (AaHPPR, pos. 62) is conserved in all HPPRs from Lamiaceae. The putative enzyme could be assigned to the D-isomer specific 2-hydroxyacid dehydrogenase family by blastp, with putative hydroxypyruvate/glyoxylate and hydroxyphenylpyruvate reductases as high identity hits.

Since hits appeared in the *Anthoceros* database for two sequences from *Arabidopsis thaliana* annotated by Timm et al. (2011) as HPR1 (peroxisomal) and HPR2 (cytosolic), these were also amplified, while a HPR3 analogue could not be identified. HPR2 was renamed as AaHPPR2 because of its resemblance to AaHPPR. In contrast to AaHPPR, amplification via cDNA was possible. AaHPR1 comprised an ORF of 1284 bp or 427 aa and AaHPPR2 revealed 1215 bp or

404 aa, the latter thus 228 bp larger than AaHPPR. This feature of AaHPPR2, materialising as significant *N*-terminal overlength in direct comparison to AaHPPR, is especially interesting since a second start codon would also allow a shorter sequence roughly corresponding to the length of the latter. Since no reliable information is available about HPPRs in basal plants, the full length was amplified after exclusion of potential signal peptides using SignalP-6.0 (Teufel et al. 2022). AaHPR1 demonstrated the highest identity at 100% with *Anthoceros agrestis* Oxford strain, while AaHPPR2 reached 100% with both strains. The putative AaHPR1 could be assigned to the D-isomer specific 2-hydroxyacid dehydrogenase family by blastp, with putative glycerate dehydrogenase and glyoxylate/hydroxypyruvate reductase as high identity hits, some of them annotated as peroxisomal enzymes. The terms glycerate dehydrogenase and hydroxypyruvate reductase are used synonymously as they refer to the perspective of the bidirectional reaction. The putative AaHPPR2 produced high identity hits with putative hydroxyphenylpyruvate reductases and hydroxypyruvate/glyoxylate reductases. Pairwise sequence alignment with *Coleus blumei* HPPR showed lower identity and similarity for AaHPPR2 than for AaHPPR. However, this can be attributed to the *N*-terminal excess. Truncating AaHPPR2 to the next downstream methionine which leads to a comparable sequence length leads to a result, where AaHPPR2 slightly exceeds AaHPPR in both properties. The common features and differences of AaHPR1, AaHPPR and AaHPPR2 could be emphasised by aligning them with three HPR sequences, i.e. from *Cucumis sativus*, *Arabidopsis thaliana* and *Coleus blumei*, as well as five HPPR sequences, i.e. from *Scutellaria baicalensis*, *Coleus blumei*, *Perilla frutescens*, *Salvia officinalis* and *Saliva miltiorrhiza*. It showed that enzymes of the HPR and HPPR type of the higher plants are clearly distinguishable, earlier demonstrated by Xu et al. (2018) in a phylogenetic tree, but share important family-specific characteristics like the NAD(P)H+H^+/$NAD(P)^+$ binding site and the conserved amino acid residues Arg232, His279 and Glu261, described for *Coleus blumei* H(P)PR by Janiak et al. (2010). In direct comparison, enzymes of the HPR type had a greater average length than HPPRs. The expected *N*-terminal excess of AaHPPR2 compared to the peer group was clearly visible, but AaHPR1 also revealed a significant excess, which was not detectable in putative basal HPRs. Since both enzymes comprise a second start codon, the targeting of which would lead to a more suitable sequence length, it can be suggested that the overlength observed in both sequences might not be part of the actual translated sequence. This conclusion is supported by adjacent

nucleotides in the ATG codon region that conformed to known plant Kozak consensus sequences (Joshi et al. 1997; Hernández et al. 2019).

2 Heterologous expression and properties

2.1 Tyrosine aminotransferase

After amplification of the *AaTAT* full length sequence, *AaTAT* was to be ligated into the expression vector pET-15b. Since, in addition to the NdeI restriction sites of the primers another site was located centrally in the sequence, it had to be partially digested, which was successful. The subsequent ligation and transformation failed in several approaches, with no or only a few colonies growing, which proved to be non-harbouring. The problem was solved by changing the method. *AaTAT* was amplified again using specific primers, extended by a sequence complementary to pET-15b. The product was co-transformed with pET-15b and successfully introduced through homologous recombination in *E. coli*.

AaTAT could be expressed in high yields, visualised by SDS-PAGE and Western blot at 50 kDa and well recognisable in both, the crude extract and the PD-10 fraction (Busch and Petersen 2021). However, some non-specific bands of the washing-fraction remained in the purified enzyme. The 50 kDa indicated for the monomer corresponds to the calculated molecular weight of 53.09 kDa including His-tag. This fits to the range of known results, such as 43 kDa for TAT1 and 56 kDa for TAT2 (purified enzyme) from *Anchusa officinalis* (De-Eknamkul and Ellis 1987b), TAT from *Arabidopsis thaliana* (Prabhu and Hudson 2010) with 52 kDa (including His- and S-tag), TAT from *Prunella vulgaris* (Ru et al. 2017) with 65 kDa (including His-tag) and TAT from *Papaver somniferum* (Lee and Facchini 2011) with 46 kDa (including His-tag). First L-tyrosine assays analysed by HPLC revealed product formation for AaTAT with both, the crude extract and the purified enzyme, but only with crude extract of empty vector control. This activity was expected due to bacterial aromatic aminotransferases, essential for the biosynthesis of aromatic amino acids, which were only removed during His-tag purification (e.g. Gelfand and Steinberg 1977). The proteins in the non-specific bands detected via SDS-PAGE in the purified enzyme sample thus showed no TAT activity. The analysis of the reaction products, however, revealed three peaks, which initially complicated the evaluation. Two correlated with the product formation and were expected due to the keto-enol tautomerism,

i.e. pHPP is present in the enol form both in the solid state as well as in organic media (Kawai et al. 1991) and is gradually converted to the keto form in aqueous medium (Painter and Zilva 1947; Bücher and Kirberger 1952). An unknown third peak was particularly present in the pHPP reference. In the results of the assays, this was only visible in very small amounts. The peak could finally be identified as 4-hydroxybenzaldehyde (HBA), the degradation product of pHPP, whose conversion is significantly accelerated in alkaline environment (Doy 1960; Pitt 1962). The identification was realised by LC-MS, which allowed the analysis of the time-dependent increase of the keto and enol peaks during the reaction of AaTAT with L-tyrosine as well as the spontaneous formation of the HBA peak after alkaline treatment. Although Kawai et al. (1991) reported a reliable method for separation and analysis of pHPP on HPLC, the behaviour of the analyte raised several concerns. Firstly, assays of any kind would have to be subject to a strict time regime in order to have comparable conditions, i.e. by the time the last sample is measured, a partial change may have already occurred. Secondly, a considerable influence of the pH value and, presumably, also of the temperature would have to be assumed, which would have made the determination of the respective optima impossible. The photometric method introduced by Diamondstone (1966) offered a practical alternative, in which the pHPP formed is quantitatively converted to HBA by alkaline treatment, followed by measurement of the latter as auxiliary analyte. Based on own observations, the assay was in parts adapted. Diamondstone reports stable absorption after 30 min at room temperature, which lasted for up to 90 min. Own observations via HPLC showed that when heated to 45 °C, a quantitative conversion was already evident after 15 min. For the assay set-up, 30 min were implemented. It is common practice to use the extinction coefficient to quantify analytes in photometry. But depending on the source, there are different values for HBA (Ewald and Hübener 1961; Pitt 1962; Diamondstone 1966). Diamondstone assumes that this is due to the pH value of the corresponding setup, an observation I share. This places high demands on the standards used. In addition, this is complicated by the maximum conversion of pHPP to HBA of about 80% (Diamondstone 1966) or 92% (this work), which would require a correction factor depending on the experiment. To solve this problem, an internal standard was used that contained pHPP and was carried throughout the experiment up to the analysis. As of today, it can only be speculated why HBA is not 100% recoverable. However, in the HPLC as well as in the LC-MS analyses, a further previously non-existent peak was found after alkaline treatment. The mass could fit e.g. to 4-hydroxyphenylacetate or 3,4-

dihydroxyphenylacetaldehyde. While the formation of the aldehyde by decarboxylation seems conceivable, an additional hydroxy group would have to be introduced. A hitherto unknown alkaline catalysed side reaction could possibly be the cause. Finally, the measurement wavelength was determined to be 330 nm versus the 331 nm of the underlying method, but in the range of what has been reported so far (Pitt 1962; and literature cited therein). The alkaline agent NaOH was replaced by KOH, an implementation already known from Ewald and Hübener (1961).

For AaTAT, a pH optimum of 7.9-8.4 was determined, which corresponds to characterised TATs from *Papaver somniferum* (Lee and Facchini 2011) and *Arabidopsis thaliana* (Prabhu and Hudson 2010; Wang et al. 2016). *Anchusa officinalis* (De-Eknamkul and Ellis 1987b) exhibits a slightly more alkaline optimum with ~9.0 for TAT1 and 9.6 for TAT2. In contrast, AaTAT showed a higher temperature optimum at 60 °C and temperature stability than comparable plant TATs. For *Anchusa officinalis*, a 50% loss of activity is reported at 65 °C, 55 °C, 50 °C for the respective isoforms (De-Eknamkul and Ellis 1987b), while such loss was shown at 60 °C for *Arabidopsis thaliana* TAT with optimum at 30-40 °C (Prabhu and Hudson 2010). However, there are also somewhat more distantly related enzymes such as prephenate aminotransferase from *Anchusa officinalis* (De-Eknamkul and Ellis 1988), which also accepts pHPP, with an optimum between 60 and 70 °C, or prephenate aminotransferase from *Nicotiana sylvestris* with 70 °C (Bonner and Jensen 1985).

(Aromatic) aminotransferases often exhibit broad substrate acceptance (Wang and Maeda 2018), which can be more or less pronounced. For *Arabidopsis thaliana* for example, Wang et al. (2016) report TAT1 with significant activity for L-Tyr, L-Phe and L-Trp, L-Leu, L-Met and L-His, while TAT2 accepts a bouquet of amino acids, in addition to the mentioned there are L-Asp, L-Ala, L-Asn, L-Gln, L-Ser, Leu, L-Cys and L-Arg. Extensive investigations were therefore carried out for AaTAT, with kinetic data being collected for L-Tyr and L-Phe as well as the amino group acceptors 2-oxoglutarate, oxaloacetate, phenylpyruvate and pyruvate. Further amino acids were tested via TLC in an easily established procedure and checked via HPLC by derivatisation with OPA. Of eight amino acids tested, six were accepted by AaTAT (Busch and Petersen 2021). All aromatic amino acids, i.e. L-Tyr, L-Phe, L-DOPA and L-Trp were accepted, the latter two being assessed qualitatively via TLC and HPLC. Similarly, the aliphatic amino acids L-Ala and L-Glu were identified as accepted substrates, whereas L-Ser and L-Asp were not. It is interesting

that L-Asp is not accepted in contrast to L-Glu, a more space-demanding dicarboxylic amino acid with one more CH_2-group (Busch and Petersen 2021). There are reports of active discrimination between aromatic and dicarboxylic amino acids, e.g. complex hydrophobic interactions or charged and neutral pockets (Oue et al. 1997; Hirotsu et al. 2005). However, the discrimination between two such similar amino acids remained surprising, especially since the keto acid counterpart oxaloacetate is converted to L-Asp. In fact, two reaction products, L-Asp and L-Ala, were found in the TLC investigations of the oxaloacetate reaction. This is due to the tendency of the compound to decarboxylate (Tsai 1967). It is conceivable that the very low turnover detected in the kinetics with oxaloacetate is rather due to the turnover of nascent pyruvate and that oxaloacetate is only converted in traces and becomes only visible on the TLC due to the very long incubation. The discrimination between the small non-polar side chain of L-Ala and the polar, more demanding one of L-Ser, on the other hand, seems reasonable (Busch and Petersen 2021). It should be noted that such a simplified conclusion on substrate acceptance falls short. TAT2 from *Arabidopsis thaliana* (Wang et al. 2016), the isoform with the broader acceptance spectrum, showed activity with complex amino acids with a demanding and partly polar or charged side chain such as L-His, L-Arg, L-Asn, L-Gln, L-Leu, while representatives with a small or modified side chain such as L-Val, L-Ile and even L-Gly were not accepted. Rather, it seems that specific interactions play a role. This is supported by some publications. The aromatic aminotransferase from *Paracoccus denitrificans* shows an overall structural homology to aspartate aminotransferase, yet the former also accepts aromatic amino acids in addition to the acidic substrate. This is due to a substrate-dependent rearrangement of the hydrogen bond network. The flexibility of an arginine side chain, responsible for the necessary space for aromatic substrates, alone cannot explain the substrate specificity (Okamoto et al. 1998). To my knowledge, no crystal structures of plant aromatic aminotransferases exist yet, while success has already been achieved for aminotransferases from other organisms, e.g. *Escherichia coli* (Ko et al. 1999), *Trypanosoma cruzi* (Blankenfeldt et al. 1999) and mouse (Mehere et al. 2010). For aspartate aminotransferase, often referred to as the blueprint of subgroup I aminotransferases, no crystal information exists in plants either. On the other hand, for alanine aminotransferase, another representative of said subgroup, a crystal structure from barley (*Hordeum vulgare*) exists, for which a similar fold to that of the human and *Pyrococcus furiosus* enzymes, among others, is named (Duff et al. 2012). Due to the differences in function and reaction direction,

i.e. mainly degradation in plants, amino acid biosynthesis in bacteria and fungi and participation in catabolic processes in animal organisms (Parthasarathy et al. 2018; Wang and Maeda 2018), the availability of a plant representative would be significant in the future. The term promiscuous used by me in this work should be reconsidered, since the sources mentioned regarding substrate specificity suggest that there is more to it.

The acceptances determined so far already allow a comparison with other plant TATs. *Anchusa officinalis* TAT1 and TAT2 primarily accept L-Tyr and, at a lower rate, L-Phe. In addition, L-Glu and L-Asp are also utilised. This also applies to TAT3, whereby the enzyme was suggested as a general aspartate aminotransferase due to its strong preference for L-Asp. As already noted, primarily TAT1 and possibly also TAT2 are associated with RA biosynthesis (De-Eknamkul and Ellis 1987b). *Arabidopsis thaliana* TAT1 and TAT2 show a clearly differentiated substrate profile, as already noted, whereby TAT1 seems to show a focus on aromatic substrates with a distinct preference for L-Tyr, while TAT2 shows a very diverse profile (Wang et al. 2016). Since transamination is a reversible reaction with actually two substrates, the consideration of keto acids is also important. AaTAT accepted all tested compounds, i.e. 2-oxoglutarate, oxaloacetate, phenylpyruvate, pyruvate and even prephenate. For comparison, TAT3 from *Anchusa officinalis* preferred 2-oxoglutarate, while TAT1 additionally accepted 2-oxoadipate and oxaloacetate comparably well and pyruvate somewhat less. TAT2 preferred oxaloacetate with a comparable profile. Glyoxylate was of minor importance for all three enzymes (De-Eknamkul and Ellis 1987b). *Arabidopsis thaliana* TAT1 and TAT2 could utilise 2-oxoglutarate, oxaloacetate and glyoxylate. In addition, aromatic amino acceptors were also tested. While TAT1 mainly accepted phenylpyruvate, TAT2 also accepted 4-hydroxyphenylpyruvate, whose conversion to tyrosine was favoured, and prephenate (Wang et al. 2016). The utilisation of prephenate has already been demonstrated by Prabhu and Hudson (2010) concerning the same enzyme. This is particularly interesting because, as already mentioned, TAT2 favours the reverse reaction, i.e. from keto acid to amino acid, indicated by the k_{cat}/K_m ratio. While prephenate is not the main acceptor, its structure is quite unique. To assume that it is a randomly accepted keto acid is unlikely. Indeed, a specific prephenate aminotransferase also exists for the biosynthesis of the precursor of Tyr and Phe in plants (Graindorge et al. 2010). The prephenate acceptance in connection with the preferred reaction direction to tyrosine could, however, indicate a role in the biosynthesis of

aromatic amino acids. Accordingly, it would be important in future experiments to also determine the favoured reaction direction of AaTAT.

While some of the acceptance profiles identified are qualitative in nature, AaTAT was also characterised kinetically (see also Busch and Petersen 2021). The enzyme demonstrated high substrate specificity for L-Tyr with lower acceptance of L-Phe, represented by a 54-fold higher catalytic efficiency. The higher V_{max} of L-Phe is negated by a 133-fold higher K_m value. A similar picture was shown for TAT1 from *Arabidopsis thaliana*. Here, V_{max} for L-Phe reaches about 50% of that of L-Tyr, but is also counteracted by the high K_m value. Comparable high k_{cat}/K_m values are otherwise not found in additional plant species and also TAT2 from *Arabidopsis thaliana* exhibits rather low values (Prabhu and Hudson 2010; Lee and Facchini 2011; Riewe et al. 2012; Wang et al. 2016; Ru et al. 2017). Since velocity parameters are subject to many influences, e.g. purification method, buffer system, it is also worth looking at the substrate affinities. According to Busch and Petersen (2021), these are comparable with values from the literature: 1.82 ± 0.09 mM in *Papaver somniferum* (Lee and Facchini 2011), 0.19 ± 0.16 mM and 2.9 ± 0.4 mM, respectively, forTAT2 in *Arabidopsis thaliana* (Prabhu and Hudson 2010; Wang et al. 2016), 0.18 ± 0.02 mM and 0.204 ± 0.001 mM, respectively, for TAT1 in *Arabidopsis thaliana* (Riewe et al. 2012; Wang et al. 2016) and 0.40 ± 0.05 mM in *Prunella vulgaris* (Ru et al. 2017). All the enzymes mentioned have in common that L-Tyr has the lowest K_m value of all aromatic amino donors and especially in comparison with L-Phe. This is most evident for TAT from *Prunella vulgaris* (Ru et al. 2017) and *Anthoceros agrestis*. Attempts were also made to determine Michaelis-Menten kinetics for PLP. As expected, due to the covalent binding of PLP in the enzyme, a basic activity was already present without external addition. This could be increased to a maximum, after which it decreased again. The work of Hayashi et al. (1967), however, shows that positive results are possible using the apoenzyme. For AaTAT, it can be assumed that the most likely substrate/co-substrate pair is L-tyrosine/2-oxoglutarate (Busch and Petersen 2021). Phenylpyruvate was also converted efficiently, so that *in vivo* substrate concentrations are likely to be crucial in the end. It can be assumed that 2-oxoglutarate is sufficiently present as an important metabolite (Krebs cycle). A role in RA biosynthesis is very conceivable based on the available data. Pending would be the aforementioned determination of the favoured reaction direction, whereby the calculated k_{cat}/K_m for the reaction of L-Tyr was already rather high. Even if catalytic efficiency of the reverse reaction would be even higher, the reaction *in vivo* could be rather dependent on the prevalent

substrate concentrations. Whether the acceptance of the additional amino acids can be attributed any significance still needs to be clarified by collecting further kinetic data, as the analyses on them to date have been primarily qualitative in nature.

2.2 Hydroxyphenylpyruvate reductase

After successfull ligation of *AaHPPR* into the expression vector pET-15b it was introduced into *E. coli* SoluBL21. AaHPPR could be expressed at high yields. SDS-PAGE revealed a protein of slightly larger than 35 kDa comparable to the estimated molecular weight (36.88 kDa). *Coleus blumei* HPPR in comparison, calculated as 34.11 kDa was detected in the same range. The active CbHPPR later was analysed via gel filtration, capturing the homodimer, to be 45-68 kDa (Kim et al. 2004; Janiak et al. 2010). Other plant HPPRs are also reported in this range, e.g. HPPR from *Salvia miltiorrhiza* with 35 kDa (including His-tag; Wang et al. 2017) and the predicted *Scutellaria baicalensis* enzyme with 34.27 kDa (Kim et al. 2014b). HPPR2 and HPPR3 from *Arabidopsis thaliana* show comparable sizes on Western blot, while HPR1 is characteristically larger. Interestingly, the non-active HPPR4 as well (Xu et al. 2018). AaHPPR showed activity both in the crude extract and as a purified enzyme. In the crude extract, a minimal level was also observed for the empty vector control, which was attributed to bacterial reductase activity of the host, as was also noted by Kim et al. (2004) for *Coleus blumei* HPPR and suggested to originate from lactate dehydrogenase activity of *E. coli*.

The pH optimum was determined to be 7.0-7.5 and the temperature optimum to be in the very wide range 29.7-49.5 °C, with the maximum determined at 44.4 °C. This is comparable to the values for *Coleus blumei* H(P)PR, for which a pH optimum of 7.0 and a temperature optimum in the range 40.0-55.0 °C were reported (Janiak 2007). The native enzyme (Häusler et al. 1991), however, had a comparable pH optimum (pH 6.5-7.0) but a lower temperature optimum (35-37 °C). Janiak (2007) attributed this discrepancy to potential proteases in the isolated protein preparation. Both heterologous enzymes thus show a very broad temperature optimum with a steep decrease of activity after the plateau phase, presumably due to the onset of denaturation. No optimum is mentioned for HPPR from *Salvia miltiorrhiza* (Wang et al. 2017), but the assays reported were also carried out at pH 7 and slightly lower temperature (25 °C). The early results indicated that at small analyte concentrations a side peak, either due to small substrate concentrations or short measurement times for linearity, would cause

problems in the kinetic measurements. Therefore, all other experiments were carried out in the form of a photometric measurement with NADPH as analyte. AaHPPR showed activity with all substrates tested, i.e. 4-hydroxyphenylpyruvate, 3,4-dihydroxyphenylpyruvate, phenylpyruvate, 4-hydroxy-3-methoxyphenylpyruvate, pyruvate as well as hydroxypyruvate. An equally treated empty vector control demonstrated that it was an AaHPPR-specific reaction. Since the continuous absorption measurement used does not allow identification of the reaction product, these were identified via LC-MS, by which the products lactate and glycerate could not be detected. However, this was due to the technical limitations and was expected, i.e. low masses (LC-MS calibrated for masses m/z 100-400) and high hydrophilicity of analytes (unsuitable eluent and column), since a commercial analyte reference was not detectable as well. The result shows that AaHPPR thus has as comparable substrate spectrum to H(P)PR from *Coleus blumei*, although glyoxylate and 2-oxoisocaproate were not tested. For 4-hydroxyphenylpyruvate, the D-enantiomer specificity of the reaction was demonstrated. The results correspond to those of Janiak (2007), fortunately this time a pure L-4-hydroxyphenyllactate reference was available, and Janiak et al. (2010). Kinetic information can only be presented for 4-hydroxyphenylpyruvate and hydroxypyruvate in the course of this work. However, an educated guess based on the patterns of the substrate acceptance assays suggests that pyruvate, phenylpyruvate, 4-hydroxy-3-methoxyphenylpyruvate are also accepted with a decent turnover and that 3,4-dihydroxyphenyllactate has the highest turnover of the aromatic substrates. This is consistent with the results of CbH(P)PR, although only a minor activity for phenylpyruvate was found (Janiak 2007). Consistent with this, no activity at all was detected with phenylpyruvate for the native enzyme (Häusler et al. 1991). HPPR from *Salvia miltiorrhiza*, on the other hand, shows even higher activity for phenylpyruvate than for 4-hydroxyphenylpyruvate (Wang et al. 2017). *Arabidopsis thaliana* revealed isoform-specific results. While HPPR2 shows only low activity with phenylpyruvate, this is significantly present for HPPR3. This work, in particular, is interesting because it looks at the substrate acceptance profiles especially in relation to pHPP and β-HP and thus the differences of HPR and HPPR. Here, HPPR2 exhibits both activities, whereby the HPR activity is clearly more pronounced. HPPR3, on the other hand, revealed a preference for pHPP and shows only minimal HPR activity. HPPR4 could not be characterised due to an almost complete lack of activity (Xu et al. 2018). However, the results show that the HPPRs and their isoforms can have different preference profiles. AaHPPR resembles HPPR2 from *Arabidopsis* or H(P)PR

from *Coleus blumei*, as the HPR activity is clearly more pronounced. In contrast to the aforementioned comparative enzymes, however, phenylpyruvate seems to be better accepted. In future, however, the acceptability and the effect of the non-, single- or double hydroxylation or methoxylation of the phenyl group should also be investigated kinetically.

So far, only the expected primary substrates could be investigated in this respect. AaHPPR revealed a K_m value of 1.79 ± 0.46 mM for pHPP in combination with NADPH, comparable to the known values of 0.80 ± 0.16 mM for HPPR2 and 0.37 ± 0.06 mM for HPPR3 from *Arabidopsis thaliana* and 1.20 ± 0.14 mM for HPPR from *Salvia miltiorrhiza*. H(P)PR from *Coleus blumei*, on the other hand, had a higher value of 16.6 mM (Janiak 2007). However, much lower K_m values of 10 µM (Häusler et al. 1991) and 80 µM (Meinhard et al. 1992) were determined for the native enzyme. The deviations, which can be attributed to the enzyme preparation used which may be blended with other enzyme properties or attached purification tags, have already been discussed by Janiak (2007). An additional feature of the native HPPR investigated by Häusler et al. (1991) is the competitive inhibition by pyruvate. This has not been analysed for the heterologously expressed H(P)PR. In fact, pyruvate is accepted, although with only slight substrate turnover compared to 4-hydroxyphenylpyruvate and a comparably high K_m value (Janiak 2007). Whether this is sufficient for effective competitive inhibition, at least for the reaction with 4-hydroxyphenylpyruvate, cannot be answered without further testing. However, based on the kinetic data known from H(P)PR (Janiak 2007), this is quite conceivable. AaHPPR, which shares many characteristics with H(P)PR from *Coleus blumei*, accepted pyruvate as well but with significant turnover rate. Unfortunately, there is only little information from the literature. For HPR2 from *Spinacia oleracea* (Kleczkowski et al. 1991) it is reported that pyruvate is not accepted. Based on the description as cytosolic, NADPH-preferring and slightly smaller isoform (38 kDa vs. 41 kDa for the peroxisomal, NADH-preferring HPR) it must be the HPPR homologue (Kleczkowski and Randall 1988). At the time of publication, interest in rosmarinic acid was developing in parallel, although 4-hydroxyphenylpyruvate was not tested as a substrate for the enzyme. Why this effect has not yet been observed in heterologously expressed HPPRs or is not to be expected based on the data to data is still unclear. An important feature of *Coleus* HPPR, which also led Janiak (2007) to name it H(P)PR, is the acceptance of hydroxypyruvate and glyoxylate, with the former having the lowest K_m value (1 mM) of all substrates tested. However, it was closely followed by glyoxylate (2 mM). It should be noted that the K_m value of hydroxypyruvate is

positively influenced by the observed substrate inhibition and would be higher without this effect (Janiak 2007). A characteristic that was also observed for AaHPPR and which is known for various H(P)PRs (Kleczkowski and Randall 1988; Kleczkowski and Edwards 1989; Bamforth and Quayle 1977; Husic and Tolbert 1987). The substrates hydroxypyruvate, glyoxylate and pHPP, though the latter could not be added to saturation, showed comparable reaction rates in the enzyme from *Coleus blumei* (Janiak 2007). AaHPPR, on the other hand, revealed a significantly higher conversion of hydroxypyruvate than 4-hydroxyphenylpyruvate. In conjunction with the lower K_m value, this results in a significantly higher catalytic efficiency. It would also be appropriate to designate AaHPPR as H(P)PR since hydroxypyruvate can be seen as primary substrate.

2.3 Hydroxyphenylpyruvate reductase 2 and hydroxypyruvate reductase

After successful ligation of *AaHPPR2* into the expression vector pET-15b, it was introduced into *E. coli* SoluBL21. AaHPPR could be expressed at much lower yields than earlier expressions of AaTAT and AaHPPR. This was illustrated by SDS-PAGE and Western blot, where the suspected band was barely visible and revealed notable impurities. The protein was somewhat larger than 45 kDa but comparable to the estimated molecular weight (45.78 kDa).

AaHPR1 was as well ligated into the expression vector pET-15b and introduced into *E. coli* SoluBL21. It appeared somewhat smaller on the gel with a calculated 47.79 kDa, while on the Western blot, the size differences were smaller. Nevertheless, such behaviour is conceivable based on the individual amino acid differences. The results for AaHPPR2 and AaHPR1 should nevertheless be seen as preliminary, since it has become apparent, as I now suspect, that the native CDS is shorter and is initiated by the second start codon in the same ORF. Both enzymes have an excess length compared to other sequences, which is also reflected in the higher molecular mass. The common molecular mass for HPPRs is about 35 kDa (Wang et al. 2017; Xu et al. 2018) and about 42 kDa for HPRs (peroxisomal), where the native enzymes of both appear as dimers of two subunits (Kleczkowski and Randall 1988).

Initial enzyme tests with AaHPPR2 and AaHPR1 showed that the classification made based on the sequence was confirmed. AaHPPR2 accepted 4-hydroxyphenylpyruvate with a comparable turnover (at a substrate concentration of 1 mM) as AaHPPR. Again, the turnover with hydroxypyruvate was higher, but the discrepancy was smaller and more similar to the

ratios of H(P)PR from *Coleus blumei*. It can be assumed that the enzyme has significant specific activity with 4-hydroxyphenylpyruvate, with part of the activity masked by the low purity of the enzyme preparation and thus higher assumed AaHPPR2 concentration.

AaHPR1 did not accept 4-hydroxyphenylpyruvate and converted hydroxypyruvate with a decent turnover although with lower activity than AaHPPR. It should be noted, however, that these are preliminary experiments. Fine-tuning the reaction, i.e. choosing the suitable enzyme concentration to obtain linear reactions and optimize reaction time, proved already very difficult for AaHPPR with the substrate hydroxypyruvate. In addition, for this comparison NADPH was used as co-substrate. An indication that NADH is better accepted by AaHPR1, at least regarding the turnover rate, was already found. Finally, the pH optimum can also exert a corresponding influence. Givan and Kleczkowski (1992) reported comparable pH optima for the peroxisomal and cytosolic HPRs *in vitro*, which contrasts with the investigation on *Arabidopsis thaliana*, where slightly alkaline pH values (pH 8.4) were measured for the peroxisomal matrix in contrast to the neutral (pH 7.3) cytosol (Shen et al. 2013). It can certainly be said, however, that the direct comparison of enzymes from different organelles should be treated with caution, since other factors, e.g. redox state (DeLoache et al. 2016), are also of importance. Nevertheless, the overall results show that in *Anthoceros agrestis* might be a system homologous to vascular plants consisting of peroxisomal NADH-preferring enzyme and cytosolic NADPH-preferring enzyme(s) with a substrate spectrum extended by 4-hydroxyphenylpyruvate (Givan and Kleczkowski 1992; Xu et al. 2018).

2.4 Culture characterisation and expression analysis

The suspension cell culture of *Anthoceros agrestis* showed continuous increase in fresh weight over 14 days. This was accompanied by a decrease in sugar content in the medium. Overall nutrient uptake was indicated by a decrease in conductivity, which stabilised from the middle of the monitoring period. The pH approached a value of 8, the steep drop on day 1 being due to the uptake of free phosphate. This drop was not evident in the earlier characterisations by Vogelsang et al. (2006), whereas conductivity and sugar content showed a comparable development in CB2 medium but with a correspondingly higher initial value for the latter. It should be noted that the measured sugar content does not correspond exactly to the medium concentration, as other dissolved substances also have an influence on the value. The steep

drop of the pH value at the beginning of the characterisation could be hidden due to the measurement intervals of two days. In comparison, all characterisations of Pezeshki (2016) include day 1, accordingly the characteristic drop is recognisable. The remaining medium parameters correspond to the culture studied here.

The study showed that RA accumulation reaches its maximum at 4.8% early on day 3 and stabilises at about 3% after a declining phase (Pezeshki et al. 2022). Vogelsang et al. (2006), on the other hand, reports a maximum on day 8 (CB2) or later on day 12 (CB4). The CB4 culture, with the highest sugar content, showed lower RA accumulation than CB2. On the other hand, it contained more rosmarinic acid 3'-*O*-β-D-glucoside. In both characterisations, the latter decreased after reaching a maximum at day 4. Thus, if transformation processes take place in parallel to RA accumulation, comparability to the CBM culture (this work) is not easily accessible. This is also well illustrated by the characterisations in CB2 medium by Pezeshki (2016), with high glucoside levels recorded early after inoculation and a corresponding negative influence on the RA content. The RA content is thus strongly dependent on the medium used, which is why CBM culture is also more suitable for elucidating rosmarinic acid biosynthesis (Pezeshki 2016). In contrast to the characterisation of this present work however, the cultures in CBM medium did not show a comparable early maximum, but a more continuous increase to 9% in the first experiment and changing levels of 5-12% in the second experiment. Overall, the achieved RA accumulation is higher (Pezeshki 2016). The discrepancy by almost a factor of 2 can have multiple reasons, e.g. age of the cell culture, location and light conditions and exact temperature.

In the expression analysis of genes encoding enzymes of RA biosynthesis via quantitative real-time PCR, two general patterns could be determined. On the one hand, there was a strong fold-change in the first 5 days with subsequent decreases, i.e. TAT, HPPR2. On the other hand, a more or less continuous increase and late decrease as for HPPR and HPR1. The first pattern demonstrated a sound correlation between transcript abundance and RA content. The data are related to the results on PAL from *Anthoceros agrestis*, which shows a comparable pattern (Pezeshki et al. 2022). However, both PAL and TAT also have general physiological functions, i.e. homogentisic acid biosynthesis (TAT), general phenylpropanoid metabolism (PAL) and therefore cannot be seen as RA-specific enzymes. Thus, the high transcript abundance already a few days after inoculation can be attributed to the rich nutrient supply and thus the ramp-

up of the metabolism. Overall, the pattern of RA content in this study is not particularly characteristic. The drop on culture day 10 is due to variability in dry weight. Thus, a strong correlation as demonstrated for *Anchusa officinalis*, although based on evaluation of TAT activity and not on gene expression (De-Eknamkul and Ellis 1987a; Mizukami and Ellis 1991), cannot be detected here. A more unlikely hypothesis related to TAT, would be the potential involvement of the enzyme in the biosynthesis of L-Tyr, either through the direct biosynthesis of the latter from 4-hydroxyphenylpyruvate or through activity with prephenate. However, this is mainly based on the observation of the expression analysis and the comparable high temperature optimum of prephenate aminotransferase from *Anchusa officinalis*, for which De-Eknamkul and Ellis (1988) found a high activity in the exponential growth phase and before the onset of RA production. Since it is still unknown how the biosynthesis of aromatic amino acids takes place in basal plants, only further investigations in this direction can provide information. The existence of only one isoform, however, makes the involvement of AaTAT at several junctions seem conceivable. The most important finding in the expression analysis was that HPPR and HPPR2 are differentially expressed. Although the present studies suggest that HPPR is not an enzyme exclusively responsible for 4-hydroxyphenyllactic acid synthesis, also taking into account the presence of analogues in species without RA (Wang et al. 2016), HPPR and HPPR2 would also be an interesting target for future studies. It can be assumed that the two enzymes play different roles in *Anthoceros agrestis*, at least one of them possibly in RA biosynthesis.

3 Summary and conclusion

AaTAT was comprehensively characterised, with L-tyrosine/2-oxoglutarate identified as the most likely substrate/co-substrate pair based on kinetic data. The enzyme showed high catalytic efficiency for the substrate combination and the affinity was comparable to the literature. However, the determination of the favoured reaction direction by characterising the substrate 4-hydroxyphenylpyruvate will be important in the future. When dealing with pHPP, attention should be paid to the substance characteristics, so possibly determination should be done at pH value lowered to neutral or slightly acidic setting in order to slow down the degradation of the substrate to 4-hydroxybenzaldehyde (Kawai et al. 1991). Besides the main substrate L-tyrosine, a broad substrate acceptance was identified. For that, two methods for rapid and reliable substrate search could be established. Firstly, the separation of assays

using TLC followed by ninhydrin staining, with the advantage that both the amino acid used and the amino acid produced were detected. In the present case, the method was primarily used qualitatively. However, with the use of controlled staining and an advanced photography system, semi-qualitative evaluation would also be possible. And secondly, a method was established for verifying the results based on separation via HPLC of OPA-derivatised samples. As a result, all tested aromatic amino acids (L-Tyr, L-Phe, L-Trp and L-DOPA) and two aliphatic amino acids (L-Glu and L-Ala) were accepted. The non-acceptance of L-Ser and L-Asp in comparison with the complex acceptance patterns of other plant TATs suggested that the term promiscuous falls somewhat short and that there is probably more to it than that. This is also reflected in the amino acceptors utilised, with acceptance of prephenate as a distinctive feature (Busch and Petersen 2021). The amino acid sequence showed the characteristic features of an aminotransferase, such as PLP binding site, homodimer interface and the conserved amino acids Lys296, Arg435 and Asp263, with an *N*-terminal overlength compared to representatives of vascular plants. Based on a sequence comparison with basal species and further considerations (e.g. Kozak motifs), this could be assumed to belong to the CDS. Since no other isoform could be found in the hornwort database, a high probability of AaTAT involvement in RA biosynthesis in *Anthoceros agrestis* can be concluded.

AaHPPR accepted all tested substrates, either aromatic as 4-hydroxyphenylpyruvate, 3,4-dihydroxyphenylpyruvate, 4-hydroxy-3-methoxyphenylpyruvate and phenylpyruvate, or aliphatic as pyruvate and hydroxypyruvate. The enzyme exhibited typical features of D-isomer specific 2-hydroxyacid dehydrogenases, such as the NAD(P)H+H^+/$NAD(P)^+$ binding site as well as the conserved amino acids Arg232, His279, Glu261, already known from *Coleus blumei* H(P)PR (Janiak et al. 2010). The preference for NADPH as well as the enantiomer specificity could be confirmed experimentally. The acceptance of 4-hydroxyphenylpyruvate confirms, in addition to the results from the alignment, the affiliation to the HPPR type. In contrast to HPPR from *Coleus blumei*, AaHPPR showed significantly higher conversion of hydroxypyruvate. It is therefore proposed to follow the nomenclature of Janiak (2007) and refer to the enzyme as AaH(P)PR. The hornwort sequence information led to the discovery of the HPPR2 isoform as well as HPR1, both of which were successfully amplified and heterologously expressed. Only preliminary characterisation results are available for both enzymes. However, these confirm the classification into the types HPPR and HPR based on the alignment. AaHPR1 favoured NADH over NADPH and did not accept 4-hydroxyphenylpyruvate. AaHPPR2 is to be considered

as an isoform of AaH(P)PR, whereby in initial tests the acceptance of hydroxypyruvate did not exceed that of 4-hydroxyphenylpyruvate as strongly as it was the case for AaH(P)PR. On the basis of more precise sequence analysis, it could be assumed that a too long CDS was selected for expression and that the native enzyme may be shorter. Whether this could have an influence on the activity could not yet be clarified. A dual function of HPPR-type enzymes as providers of phenolic components and as an efficient bypass to peroxisomal hydroxypyruvate reduction (Timm et al. 2008) is the most logical conclusion at present.

To examine a potential correlation of the identified and characterised enzymes TAT, H(P)PR, HPPR2 and HPR1 to RA biosynthesis, a culture characterisation as well as expression analysis was performed. A certain correlation of transcript abundance of TAT and HPPR2 with RA accumulation was found. However, as already discussed, the result is not reliable enough for a definitive statement. It would make sense to carry out supplementary experiments that make the correlation clearer. For this, it should be considered to integrate activity measurements of the investigated enzymes and to influence RA accumulation and therefore expression of the enzymes involved by elicitation. A more characteristic progression of the RA accumulation, especially in the later culture period, would facilitate the interpretation and increase the significance. This can be achieved either chemically, e.g. addition of methyl jasmonate (Mizukami et al. 1993; Szabo et al. 1999; Xiao et al. 2009), or biologically, e.g. preparation of *Pythium aphanidermatum* (Szabo et al. 1999), which have already shown promising effects in increasing RA accumulation in vascular plants including suspension cell cultures. The influence of UV exposure could also be tested. The results could be complemented by additional *in vivo* experiments. Relationships between RA accumulation and enzymes involved could be found, for example, for *Salvia miltiorrhiza* hairy root cultures by overexpression (Xiao et al. 2011) or RNA_i suppression in *Agrobacterium*-mediated *Coleus blumei* hairy root cultures (Hücherig and Petersen 2013). It is conceivable that the suppression or even knockout of AaTAT due to the lack of isoforms has a major impact on the phenotype and is possibly lethal for the cells due to its assumed involvement in plastoquinone and tocopherol biosynthesis (Busch and Petersen 2021). However, further evidence could also be found through *in vitro* inhibitor studies by investigating the influence of involved metabolites such as RA or (di)hydroxyphenyllactate (De-Eknamkul and Ellis 1987b), a method that is of course applicable to HPPR as well (Häusler et al. 1991).

In recent years, the biosynthetic pathway of the caffeic acid moiety of RA via the enzymes phenylalanine ammonia lyase, cinnamic acid 4-hydroxylase and hydroxycinnamate:coenzyme A ligase (Petersen 2003; Wohl and Petersen 2020a, 2020b; Pezeshki et al. 2022) have already been demonstrated for *Anthoceros agrestis*. In this work it was shown, that all the necessary enzymes for the formation of the hydroxyphenyllactic moiety are present. The actual involvement of the enzymes, as suggested, still has to be proven. The big unknown is still the hydroxycinnamoyltransferase, an analogue of the RAS from *Coleus blumei*, which has not yet been found (see e.g. Ernst et al. 2022). An indication that this central step of RA biosynthesis may well be family-specific is provided by Levsh et al. (2019), who were able to show that the RAS in Boraginaceae is paralogous to spermidine hydroxycinnamoyltransferase. So far, no evidence for such activity could be found in *Anthoceros agrestis*, but it is therefore quite conceivable that this specific reaction step might be catalysed by another enzyme, e.g. a serine carboxypeptidase-like acyltransferase or lipase-like hydroxycinnamoyltransferase (Petersen 2016). However, it is reasonable to assume that, given the presence of all the enzymes that catalyse sub-steps, it is unlikely that the vascular plants and *Anthoceros agrestis* have evolved an entirely different biosynthetic pathway. Certainty at least about the incorporated precursors could come from a ^{14}C- or ^{13}C-labelling studies inspired by those of Ellis and Towers (1970), which originally identified L-phenylalanine and L-tyrosine as starting points of RA biosynthesis.

While evidence is growing that streptophyte algae already had the tools to biosynthesise phenylpropanoids (Rieseberg et al. 2022), *Anthoceros agrestis* was shown to additionally possess the necessary enzymes for the biosynthesis of the hydroxyphenyllactic acid moiety of RA. The prevailing view of bryophytes as monophyletic thus presupposes that, if we expect analogy to the biosynthesis in higher plants, the capacity appeared early and may have evolved in whole or in part in the common ancestor of vascular plants and bryophytes. RA and related compounds may therefore have been central in the adaptation of early plants to land environments.

V Summary

Plant terrestrialisation is closely linked to the evolution of phenolic secondary metabolites, which conferred a survival advantage in the presence of new environmental factors, e.g. UV-B radiation, drought, pathogens and herbivores. Besides the incorporation of phenolic components into structural elements (cuticle, spore wall, cell wall), the esters of caffeic acid are of importance (Weng and Chapple 2010). Rosmarinic acid, formally an ester of caffeic acid and 3,4-dihydroxyphenyllactic acid, has been shown to have anti-inflammatory, antimicrobial, antitumour and antioxidant effects, among others (Amoah et al. 2016; Hitl et al. 2021). Besides being abundant in Lamiaceae (Nepetoideae) and Boraginaceae, the component is widespread in the plant kingdom, with hornworts, e.g. *Anthoceros agrestis*, as the most basal representatives (Petersen 2013). Within the scope of this work, the two enzymes tyrosine aminotransferase (TAT) and hydroxyphenylpyruvate reductase (H(P)PR) were identified, heterologously expressed in *E. coli* and enzymatically characterised.

TAT revealed a molecular mass of about 50 kDa, with the active enzyme being expected as homodimer. The highest enzyme activity was measured at a pH value of 7.9-8.4 at 60 °C. The most probable substrate/co-substrate pair was determined to be L-tyrosine/2-oxoglutarate. Depending on the actual substrate concentrations, phenylpyruvate could also be considered. Oxaloacetate and pyruvate, on the other hand, were only poorly accepted. In addition, several aromatic and aliphatic amino acids were accepted, as well as the keto acid prephenate. For the identification of these substrates, a rapid and reliable method via thin-layer chromatography with detection by ninhydrin staining verified through HPLC and derivatisation with *o*-phthalaldehyde was introduced.

HPPR revealed a molecular mass of about 35 kDa and is as well expected as homodimer. The highest enzyme activities were measured at a pH of 7.0-7.5 at a temperature of 44 °C. Contrary to early expectations, hydroxypyruvate was determined as the primary substrate with NADPH as favoured co-substrate and not 4-hydroxyphenylpyruvate. According to the enzyme family, only the D-enantiomer of 4-hydroxyphenyllactate was formed during the reaction. The results correspond to observations on heterologously expressed H(P)PR from *Coleus blumei*. In addition, the 3-hydroxylated and methoxylated derivative of 4-hydroxyphenylpyruvate as well as phenylpyruvate and pyruvate were accepted.

While no further isoforms could be found for TAT in *Anthoceros agrestis*, HPPR2 as well as the related HPR1, the photorespiratory isoenzyme, could be amplified and heterologously expressed in *E. coli*. HPPR2 showed comparable activity with hydroxypyruvate and 4-hydroxyphenylpyruvate, while HPR1 did not accept the latter and favoured NADH.

The investigated enzymes TAT, HPPR as well as HPPR2 could be involved in the biosynthesis of the 3,4-dihydroxyphenyllactic acid moiety analogous to *Coleus blumei*. Expression analysis over a 14-day culture period revealed a parallel progression of TAT and HPPR2 transcript abundance and rosmarinic acid content, which was different from previous observations in different suspension cultures of *Anthoceros agrestis*. A definite statement about the involvement of the enzymes mentioned cannot be made based on these data. However, the identification of the potential core enzymes has laid a solid foundation for future work focusing on *in vivo* analyses.

VI Zusammenfassung

Die Terrestrialisierung der Pflanzen ist eng mit der Entwicklung phenolischer Sekundärmetabolite verbunden, die einen Überlebensvorteil gegenüber neuen Umweltfaktoren wie UV-B-Strahlung, Trockenheit, Krankheitserregern und Herbivoren verschaffen. Neben dem Einbau von phenolischen Komponenten in Strukturelemente (Cutikula, Sporenwand, Zellwand) sind Kaffeesäureester von Bedeutung. Rosmarinsäure, formal ein Ester aus Kaffeesäure und 3,4-Dihydroxyphenylmilchsäure, hat nachweislich unter anderem entzündungshemmende, antimikrobielle, antitumorale und antioxidative Wirkungen. Sie kommt nicht nur in Lamiaceae (Nepetoideae) und Boraginaceae vor, sondern ist im Pflanzenreich weit verbreitet, wobei Hornmoose, z.B. *Anthoceros agrestis*, die basalsten Vertreter sind. Im Rahmen dieser Arbeit wurden die beiden Enzyme Tyrosin Aminotransferase (TAT) und Hydroxyphenylpyruvat Reduktase (H(P)PR) identifiziert, heterolog in *E. coli* exprimiert und enzymatisch charakterisiert.

TAT wies eine Molekularmasse von etwa 50 kDa auf, wobei das aktive Enzym als Homodimer auftritt. Die höchste Enzymaktivität wurde bei einem pH-Wert von 7,9-8,4 bei 60 °C gemessen. Als wahrscheinlichstes Substrat/Co-Substrat-Paar wurde L-Tyrosin/2-Oxoglutarat ermittelt. Je nach den tatsächlichen Substratkonzentrationen könnte auch Phenylpyruvat in Betracht gezogen werden. Oxalacetat und Pyruvat hingegen wurden nur schlecht akzeptiert. Darüber hinaus wurden mehrere aromatische und aliphatische Aminosäuren sowie die Ketosäure Prephenat akzeptiert. Für die Identifizierung dieser Substrate wurde eine schnelle und zuverlässige Methode über Dünnschichtchromatographie mit Nachweis durch Ninhydrinfärbung, verifiziert durch HPLC und Derivatisierung mit *o*-Phthalaldehyd, eingeführt.

Die HPPR hat eine Molekülmasse von etwa 35 kDa, auch dieses Enzym tritt aktiv als Homodimer auf. Die höchste Enzymaktivität wurde bei einem pH-Wert von 7,0-7,5 und einer Temperatur von 44 °C gemessen. Entgegen früheren Erwartungen wurde 4-Hydroxypyruvat als bestes Substrat mit NADPH als bevorzugtem Co-Substrat bestimmt und nicht 4-Hydroxyphenylpyruvat. Entsprechend der Enzymfamilie wurde bei der Reaktion nur das D-Enantiomer von 4-Hydroxyphenyllactat gebildet. Die Ergebnisse stimmen weitgehend mit Beobachtungen an heterolog exprimierter H(P)PR aus *Coleus blumei* überein. Darüber hinaus

wurden das 3-hydroxylierte und methoxylierte Derivat von 4-Hydroxyphenylpyruvat sowie Phenylpyruvat und Pyruvat akzeptiert.

Während für TAT in *Anthoceros agrestis* keine weiteren Isoformen gefunden werden konnten, ließen sich HPPR2 sowie das verwandte HPR1, das photorespiratorische Isoenzym, amplifizieren und heterolog in *E. coli* exprimieren. HPPR2 zeigte eine vergleichbare Aktivität mit Hydroxypyruvat und 4-Hydroxyphenylpyruvat, während HPR1 letzteres nicht akzeptierte und NADH bevorzugte.

Die untersuchten Enzyme TAT, HPPR sowie HPPR2 könnten aufgrund ihrer enzymatischen Eigenschaften an der Biosynthese der 3,4-Dihydroxyphenylmilchsäureeinheit analog zu *Coleus blumei* beteiligt sein. Die Expressionsanalyse über eine 14-tägige Kulturperiode zeigte einen parallelen Verlauf der TAT- und HPPR2-Transkriptabundanz und des Rosmarinsäuregehalts, der sich von einer früheren Beobachtung bei anderen Suspensionskulturen von *Anthoceros agrestis* unterschied. Eine eindeutige Aussage über die Beteiligung der genannten Enzyme kann auf Grundlage dieser Daten nicht getroffen werden. Die Identifizierung der potenziellen Schlüsselenzyme hat jedoch eine solide Grundlage für künftige Arbeiten geschaffen, die sich auf *in-vivo*-Analysen konzentrieren sollten.

VII References

Adamu RM, Ibrahim B, Ibrahim MA, Balogun EO (2021) Identification of megacerotonic acid and a quinazoline derivative from universal natural product database as potential inhibitors of *Trypanosoma brucei* brucei alternative oxidase: molecular docking, molecular dynamic simulation and MM/PBSA analysis. J Biomol Struct Dyn:1–10. https://doi.org/10.1080/07391102.2021.2003862

Ahmad Z, Husaini A, Roslan HA (2016) Characterisation and expression analysis of hydroxyphenylpyruvate reductase derived from *Orthosiphon aristatus*. BJRST 5:34–42. https://doi.org/10.33736/bjrst.226.2015

Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol 37:420–423. https://doi.org/10.1038/s41587-019-0036-z

Amoah SKS, Sandjo LP, Kratz JM, Biavatti MW (2016) Rosmarinic acid - pharmaceutical and clinical aspects. Planta Med 82:388–406. https://doi.org/10.1055/s-0035-1568274

Aquino R, Morelli S, Lauro MR, Abdo S, Saija A, Tomaino A (2001) Phenolic constituents and antioxidant activity of an extract of *Anthurium versicolor* leaves. J Nat Prod 64:1019–1023. https://doi.org/10.1021/np0101245

Bamforth CW, Quayle JR (1977) Hydroxypyruvate reductase activity in *Paracoccus denitrificans*. J Gen Microbiol 101:259–267. https://doi.org/10.1099/00221287-101-2-259

Barberini S, Savona M, Raffi D, Leonardi M, Pistelli L, Stochmal A, Vainstein A, Pistelli L, Ruffoni B (2013) Molecular cloning of SoHPPR encoding a hydroxyphenylpyruvate reductase, and its expression in cell suspension cultures of *Salvia officinalis*. Plant Cell Tiss Org 114:131–138. https://doi.org/10.1007/s11240-013-0300-8

Bateman RM, Crane PR, DiMichele WA, Kenrick PR, Rowe NP, Speck T, Stein WE (1998) Early evolution of land plants: Phylogeny, physiology, and ecology of the primary terrestrial radiation. Annu Rev Ecol Syst 29:263–292. https://doi.org/10.1146/annurev.ecolsys.29.1.263

Bauwe H, Hagemann M, Fernie AR (2010) Photorespiration: Players, partners and origin. Trends Plant Sci 15:330–336. https://doi.org/10.1016/j.tplants.2010.03.006

References

Bedewitz MA, Góngora-Castillo E, Uebler JB, Gonzales-Vigil E, Wiegert-Rininger KE, Childs KL, Hamilton JP, Vaillancourt B, Yeo Y-S, Chappell J, DellaPenna D, Jones AD, Buell CR, Barry CS (2014) A root-expressed L-phenylalanine:4-hydroxyphenylpyruvate aminotransferase is required for tropane alkaloid biosynthesis in *Atropa belladonna*. Plant Cell 26:3745–3762. https://doi.org/10.1105/tpc.114.130534

Berg JM, Tymoczko JL, Stryer L, Gatto GJ (2013) Biochemie, 7. Auflage. Lehrbuch. Springer Spektrum, Berlin, Heidelberg

Berger A, Meinhard J, Petersen M (2006) Rosmarinic acid synthase is a new member of the superfamily of BAHD acyltransferases. Planta 224:1503–1510. https://doi.org/10.1007/s00425-006-0393-y

Bertani G (1951) Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. J Bacteriol 62:293–300

Biastoff S (2003) Biochemische Untersuchung der heterolog exprimierten Tyrosin Aminotransferase aus *Coleus blumei* BENTH. Diploma thesis, Martin-Luther-Universität Halle-Wittenberg

Binding H, Mordhorst G (1991) Gametophyte regeneration and apospory from archegoniate protoplasts under conditions devised for higher plants. Bot Acta 104:330–335. https://doi.org/10.1111/j.1438-8677.1991.tb00238.x

Blackwell WH (2003) Two theories of origin of the land-plant sporophyte: Which is left standing? Bot Rev 69:125–148. https://doi.org/10.1663/0006-8101(2003)069[0125:TTOOOT]2.0.CO;2

Blankenfeldt W, Nowicki C, Montemartini-Kalisz M, Kalisz HM, Hecht HJ (1999) Crystal structure of *Trypanosoma cruzi* tyrosine aminotransferase: substrate specificity is influenced by cofactor binding mode. Protein Sci 8:2406–2417. https://doi.org/10.1110/ps.8.11.2406.

Bohm BA (1968) Phenolic compounds in ferns—III. Phytochemistry 7:1825–1830. https://doi.org/10.1016/S0031-9422(00)86654-6

Bonner CA, Jensen RA (1985) Novel features of prephenate aminotransferase from cell cultures of *Nicotiana silvestris*. Arch Biochem Biophys 238:237–246. https://doi.org/10.1016/0003-9861(85)90161-4

References

Booth MPS, Conners R, Rumsby G, Brady RL (2006) Structural basis of substrate specificity in human glyoxylate reductase/hydroxypyruvate reductase. J Mol Biol 360:178–189. https://doi.org/10.1016/j.jmb.2006.05.018

Bower FO (1890) On antithetic as distinct from homologous alternation of generations in plants. Ann Bot os-4:347–370. https://doi.org/10.1093/oxfordjournals.aob.a090569

Bowman JL (2013) Walkabout on the long branches of plant evolution. Curr Opin Plant Biol 16:70–77. https://doi.org/10.1016/j.pbi.2012.10.001

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/10.1006/abio.1976.9999

Britton HTS, Robinson RA (1931) CXCVIII.—Universal buffer solutions and the dissociation constant of veronal. J Chem Soc 0:1456–1462. https://doi.org/10.1039/JR9310001456

Bücher T, Kirberger E (1952) Über die Enol-Keto-Tautomerie der *p*-Oxyphenylbrenztraubensäure. Biochim Biophys Acta 8:401–406. https://doi.org/10.1016/0006-3002(52)90065-6

Bulgakov VP, Inyushkina YV, Fedoreyev SA (2012) Rosmarinic acid and its derivatives: Biotechnology and applications. Crit Rev Biotechnol 32:203–217. https://doi.org/10.3109/07388551.2011.596804

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

Carella P, Gogleva A, Hoey DJ, Bridgen AJ, Stolze SC, Nakagami H, Schornack S (2019) Conserved biochemical defenses underpin host responses to oomycete infection in an early-divergent land plant lineage. Curr Biol 29:2282-2294.e5. https://doi.org/10.1016/j.cub.2019.05.078

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

Cheng S, Xian W, Fu Y, Marin B, Keller J, Wu T, Sun W, Li X, Xu Y, Zhang Y, Wittek S, Reder T, Günther G, Gontcharov A, Wang S, Li L, Liu X, Wang J, Yang H, Xu X, Delaux P-M, Melkonian B, Wong GK-S, Melkonian M (2019) Genomes of subaerial Zygnematophyceae provide

insights into land plant evolution. Cell 179:1057-1067.e14. https://doi.org/10.1016/j.cell.2019.10.019

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

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

Cousins AB, Pracharoenwattana I, Zhou W, Smith SM, Badger MR (2008) Peroxisomal malate dehydrogenase is not essential for photorespiration in *Arabidopsis* but its absence causes an increase in the stoichiometry of photorespiratory CO_2 release. Plant Physiol 148:786–795. https://doi.org/10.1104/pp.108.122622

Cox CJ (2018) Land plant molecular phylogenetics: A review with comments on evaluating incongruence among phylogenies. Crit Rev Plant Sci 37:113–127. https://doi.org/10.1080/07352689.2018.1482443

Cox CJ, Li B, Foster PG, Embley TM, Civán P (2014) Conflicting phylogenies for early land plants are caused by composition biases among synonymous substitutions. Syst Biol 63:272–279. https://doi.org/10.1093/sysbio/syt109

D'Auria JC (2006) Acyltransferases in plants: a good time to be BAHD. Curr Opin Plant Biol 9:331–340. https://doi.org/10.1016/j.pbi.2006.03.016

Davies KM, Jibran R, Zhou Y, Albert NW, Brummell DA, Jordan BR, Bowman JL, Schwinn KE (2020) The evolution of flavonoid biosynthesis: A bryophyte perspective. Front Plant Sci 11:7. https://doi.org/10.3389/fpls.2020.00007

Davies KM, Jibran R, Albert NW, Zhou Y, Schwinn KE (2021) Conservation and divergence between bryophytes and angiosperms in the biosynthesis and regulation of flavonoid production. In: Reed JD, Freitas VAP, Quideau S (eds) Recent advances in polyphenol research: Volume 7. Wiley Blackwell, Hoboken, NJ, pp 227–263

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 S, Fürst-Jansen JMR, Irisarri I, Dhabalia Ashok A, Ischebeck T, Feussner K, Abreu IN, Petersen M, Feussner I, Vries J de (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

De-Eknamkul W, Ellis BE (1987b) Purification and characterization of tyrosine aminotransferase activities from *Anchusa officinalis* cell cultures. Arch Biochem Biophys 257:430–438. https://doi.org/10.1016/0003-9861(87)90587-X

De-Eknamkul W, Ellis BE (1987a) Tyrosine aminotransferase: The entrypoint enzyme of the tyrosine-derived pathway in rosmarinic acid biosynthesis. Phytochemistry 26:1941–1946. https://doi.org/10.1016/S0031-9422(00)81734-3

De-Eknamkul W, Ellis BE (1988) Purification and characterization of prephenate aminotransferase from *Anchusa officinalis* cell cultures. Arch Biochem Biophys 267:87–94. https://doi.org/10.1016/0003-9861(88)90011-2

DeLoache WC, Russ ZN, Dueber JE (2016) Towards repurposing the yeast peroxisome for compartmentalizing heterologous metabolic pathways. Nat Commun 7:11152. https://doi.org/10.1038/ncomms11152

Delwiche CF, Graham LE, Thomson N (1989) Lignin-like compounds and sporopollenin coleochaete, an algal model for land plant ancestry. Science 245:399–401. https://doi.org/10.1126/science.245.4916.399

Delwiche CF, Cooper ED (2015) The evolutionary origin of a terrestrial flora. Curr Biol 25:R899-910. https://doi.org/10.1016/j.cub.2015.08.029

Desirò A, Duckett JG, Pressel S, Villarreal JC, Bidartondo MI (2013) Fungal symbioses in hornworts: a chequered history. Proc Biol Sci 280:20130207. https://doi.org/10.1098/rspb.2013.0207

Di P, Zhang L, Chen J, Tan H, Xiao Y, Dong X, Zhou X, Chen W (2013) ^{13}C tracer reveals phenolic acids biosynthesis in hairy root cultures of *Salvia miltiorrhiza*. ACS Chem Biol 8:1537–1548. https://doi.org/10.1021/cb3006962

Diamondstone TI (1966) Assay of tyrosine transaminase activity by conversion of *p*-hydroxyphenylpyruvate to *p*-hydroxybenzaldehyde. Anal Biochem 16:395–401. https://doi.org/10.1016/0003-2697(66)90220-X

Dodds WK, Gudder DA, Mollenhauer D (1995) The ecology of *Nostoc*. J Phycol 31:2–18. https://doi.org/10.1111/j.0022-3646.1995.00002.x

References

Dong N-Q, Lin H-X (2021) Contribution of phenylpropanoid metabolism to plant development and plant-environment interactions. J Integr Plant Biol 63:180–209. https://doi.org/10.1111/jipb.13054

Doy CH (1960) Alkaline conversion of 4-hydroxyphenylpyruvic acid to 4-hydroxybenzaldehyde. Nature 186:529–531. https://doi.org/10.1038/186529a0

Duff RJ, Nickrent DL (1999) Phylogenetic relationships of land plants using mitochondrial small-subunit rDNA sequences. Am J Bot 86:372–386. https://doi.org/10.2307/2656759

Duff SMG, Rydel TJ, McClerren AL, Zhang W, Li JY, Sturman EJ, Halls C, Chen S, Zeng J, Peng J, Kretzler CN, Evdokimov A (2012) The enzymology of alanine aminotransferase (AlaAT) isoforms from *Hordeum vulgare* and other organisms, and the HvAlaAT crystal structure. Arch Biochem Biophys 528:90–101. https://doi.org/10.1016/j.abb.2012.06.006

Edwards D, Davies KL, Axe L (1992) A vascular conducting strand in the early land plant *Cooksonia*. Nature 357:683–685. https://doi.org/10.1038/357683a0

Edwards D, Duckett JG, Richardson JB (1995) Hepatic characters in the earliest land plants. Nature 374:635–636. https://doi.org/10.1038/374635a0

El Malki F, Frankard V, Jacobs M (1998) Molecular cloning and expression of a cDNA sequence encoding histidinol phosphate aminotransferase from *Nicotiana tabacum*. Plant Mol Biol 37:1013–1022. https://doi.org/10.1023/A:1006007125448

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

Ellis BE, Remmen S, Goeree G (1979) Interactions between parallel pathways during biosynthesis of rosmarinic acid in cell suspension cultures of *Coleus blumei*. Planta 147:163–167. https://doi.org/10.1007/BF00389519

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

ESA Application note 70-0160P Amino Acids - OPA/βME derivatization. ESA Analytical. Brook Farm Dorton Aylesbury, HP18 9NH United Kingdom

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 13:59–68. https://doi.org/10.1111/j.1438-8677.2010.00345.x

European Redlist: Introduction to Bryophytes, 22/03/2022 14:17. https://ec.europa.eu/environment/nature/conservation/species/redlist/bryophytes/introduction.htm

Ewald W, Hübener HJ (1961) Bestimmung der Tyrosin-α-Ketoglutarat-Transaminase. Naturwissenschaften 48:720. https://doi.org/10.1007/BF00620969

Frahm J-P (2011) Illustrierter Schlüssel für die thallösen Lebermoose Deutschlands. Arch Bryol Special Volume 2

Frangedakis E, Shimamura M, Villarreal JC, Li F-W, Tomaselli M, Waller M, Sakakibara K, Renzaglia KS, Szövényi P (2021) The hornworts: morphology, evolution and development. New Phytol 229:735–754. https://doi.org/10.1111/nph.16874

Frey PA, Hegeman AD (2007) Enzymatic reaction mechanisms. Oxford University Press, Oxford

Fujii T, Shimizu M, Doi Y, Fujita T, Ito T, Miura D, Wariishi H, Takaya N (2011) Novel fungal phenylpyruvate reductase belongs to D-isomer-specific 2-hydroxyacid dehydrogenase family. Biochim Biophys Acta 1814:1669–1676. https://doi.org/10.1016/j.bbapap.2011.05.024

Garbary D, Renzaglia K (1998) Bryophyte phylogeny and evolution of land plants: evidence from development and ultrastructure. In: Bates JW, Ashton NW, Duckett JG (eds) Bryology for the Twenty-first Century. Routledge, London pp 45-63

Garbary DJ, Renzaglia KS, Duckett JG (1993) The phylogeny of land plants: A cladistic analysis based on male gametogenesis. Plant Syst Evol 188:237–269. https://doi.org/10.1007/BF00937730

Garcia I, Rodgers M, Lenne C, Rolland A, Sailland A, Matringe M (1997) Subcellular localization and purification of a *p*-hydroxyphenylpyruvate dioxygenase from cultured carrot cells and characterization of the corresponding cDNA. Biochem J 325:761–769. https://doi.org/10.1042/bj3250761

Gebhardt JS, Wadsworth GJ, Matthews BF (1998) Characterization of a single soybean cDNA encoding cytosolic and glyoxysomal isozymes of aspartate aminotransferase. Plant Mol Biol 37:99–108. https://doi.org/10.1023/a:1005973019045

Gelfand DH, Steinberg RA (1977) *Escherichia coli* mutants deficient in the aspartate and aromatic amino acid aminotransferases. J Bacteriol 130:429–440

Gensel PG (2008) The earliest land plants. Annu Rev Ecol Evol Syst 39:459–477. https://doi.org/10.1146/annurev.ecolsys.39.110707.173526

References

Givan CV, Kleczkowski LA (1992) The enzymic reduction of glyoxylate and hydroxypyruvate in leaves of higher plants. Plant Physiol 100:552–556. https://doi.org/10.1104/pp.100.2.552

Goldberg JD, Yoshida T, Brick P (1994) Crystal structure of a NAD-dependent D-glycerate dehydrogenase at 2.4 Å resolution. J Mol Biol 236:1123–1140. https://doi.org/10.1016/0022-2836(94)90016-7

Goremykin VV, Hellwig FH (2005) Evidence for the most basal split in land plants dividing bryophyte and tracheophyte lineages. Plant Syst Evol 254:93–103. https://doi.org/10.1007/s00606-005-0337-1

Graham LE (1996) Green algae to land plants: An evolutionary transition. J Plant Res 109:241–251. https://doi.org/10.1007/BF02344471

Graindorge M, Giustini C, Jacomin AC, Kraut A, Curien G, Matringe M (2010) Identification of a plant gene encoding glutamate/aspartate-prephenate aminotransferase: the last homeless enzyme of aromatic amino acids biosynthesis. FEBS Lett 584:4357–4360. https://doi.org/10.1016/j.febslet.2010.09.037

Grant GA (1989) A new family of 2-hydroxyacid dehydrogenases. Biochem Biophys Res Commun 165:1371–1374. https://doi.org/10.1016/0006-291X(89)92755-1

Greenler JM, Sloan JS, Schwartz BW, Becker WM (1989) Isolation, characterization and sequence analysis of a full-length cDNA clone encoding NADH-dependent hydroxypyruvate reductase from cucumber. Plant Mol Biol 13:139–150. https://doi.org/10.1007/BF00016133

Haig D (2008) Homologous versus antithetic alternation of generations and the origin of sporophytes. Bot Rev 74:395–418. https://doi.org/10.1007/s12229-008-9012-x

Hallingbäck T (ed) (2000) Mosses, liverworts, and hornworts: Status survey and conservation action plan for bryophytes. IUCN in collaboration with the Swedish Threatened Species Unit, Gland, Cambridge

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

Harborne JB (1966) Caffeic acid ester distribution in higher plants. Z Naturforsch 21b:604–605

Harholt J, Moestrup Ø, Ulvskov P (2016) Why plants were terrestrial from the beginning. Trends Plant Sci 21:96–101. https://doi.org/10.1016/j.tplants.2015.11.010

Harris BJ, Harrison CJ, Hetherington AM, Williams TA (2020) Phylogenomic evidence for the monophyly of bryophytes and the reductive evolution of stomata. Curr Biol 30:2001-2012.e2. https://doi.org/10.1016/j.cub.2020.03.048

Häusler E, Petersen M, Alfermann AW (1991) Hydroxyphenylpyruvate reductase from cell suspension cultures of *Coleus blumei* Benth. Zeitschrift für Naturforschung 46c:371–376

Hayashi S, Granner DK, Tomkins GM (1967) Tyrosine aminotransferase. J Biol Chem 242:3998–4006. https://doi.org/10.1016/S0021-9258(18)95769-8

Hedderson TA, Chapman RL, Rootes WL (1996) Phylogenetic relationships of bryophytes inferred from nuclear-encoded rRNA gene sequences. Plant Syst Evol 200:213–224. https://doi.org/10.1007/BF00984936

Hernández G, Osnaya VG, Pérez-Martínez X (2019) Conservation and variability of the AUG initiation codon context in Eukaryotes. Trends Biochem Sci 44:1009–1021. https://doi.org/10.1016/j.tibs.2019.07.001

Hirata H, Ohnishi T, Ishida H, Tomida K, Sakai M, Hara M, Watanabe N (2012) Functional characterization of aromatic amino acid aminotransferase involved in 2-phenylethanol biosynthesis in isolated rose petal protoplasts. J Plant Physiol 169:444–451. https://doi.org/10.1016/j.jplph.2011.12.005

Hirotsu K, Goto M, Okamoto A, Miyahara I (2005) Dual substrate recognition of aminotransferases. Chem Rec 5:160–172. https://doi.org/10.1002/tcr.20042

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

Holton SJ, Anandhakrishnan M, Geerlof A, Wilmanns M (2013) Structural characterization of a D-isomer specific 2-hydroxyacid dehydrogenase from *Lactobacillus delbrueckii* ssp. *bulgaricus*. J Struct Biol 181:179–184. https://doi.org/10.1016/j.jsb.2012.10.009

Huang B, Duan Y, Yi B, Sun L, Lu B, Yu X, Sun H, Zhang H, Chen W (2008a) Characterization and expression profiling of cinnamate 4-hydroxylase gene from *Salvia miltiorrhiza* in rosmarinic acid biosynthesis pathway. Russ J Plant Physiol 55:390–399. https://doi.org/10.1134/S1021443708030163

Huang B, Yi B, Duan Y, Sun L, Yu X, Guo J, Chen W (2008b) Characterization and expression profiling of tyrosine aminotransferase gene from *Salvia miltiorrhiza* (Dan-shen) in rosmarinic acid biosynthesis pathway. Mol Biol Rep 35:601–612. https://doi.org/10.1007/s11033-007-9130-2

Hücherig S, Petersen M (2013) RNAi suppression and overexpression studies of hydroxyphenylpyruvate reductase (HPPR) and rosmarinic acid synthase (RAS) genes related to rosmarinic acid biosynthesis in hairy root cultures of *Coleus blumei*. Plant Cell Tiss Org 113:375–385. https://doi.org/10.1007/s11240-012-0277-8

Husic DW, Tolbert NE (1987) NADH:hydroxypyruvate reductase and NADPH:glyoxylate reductase in algae: Partial purification and characterization from *Chlamydomonas reinhardtii*. Arch Biochem Biophys 252:396–408. https://doi.org/10.1016/0003-9861(87)90046-4

Jacobus AP, Gross J (2015) Optimal cloning of PCR fragments by homologous recombination in *Escherichia coli*. PLoS ONE 10:e0119221. https://doi.org/10.1371/journal.pone.0119221

Janiak V, Petersen M, Zentgraf M, Klebe G, Heine A (2010) Structure and substrate docking of a hydroxy(phenyl)pyruvate reductase from the higher plant *Coleus blumei* Benth. Acta Crystallogr D Biol Crystallogr 66:593–603. https://doi.org/10.1107/S0907444910006360

Janiak VD (2007) Charakterisierung und Struktur einer Hydroxy(phenyl)pyruvat Reduktase aus *Coleus blumei*. PhD thesis, University of Marburg

Jiao C, Sørensen I, Sun X, Sun H, Behar H, Alseekh S, Philippe G, Palacio Lopez K, Sun L, Reed R, Jeon S, Kiyonami R, Zhang S, Fernie AR, Brumer H, Domozych DS, Fei Z, Rose JKC (2020) The *Penium margaritaceum g*enome: Hallmarks of the origins of land plants. Cell 181:1097-1111.e12. https://doi.org/10.1016/j.cell.2020.04.019

Joshi CP, Zhou H, Huang X, Chiang VL (1997) Context sequences of translation initiation codon in plants. Plant Mol Biol 35:993–1001. https://doi.org/10.1023/a:1005816823636

Julliard JH, Breton-Gilet A (1997) Identification of hydroxypyruvate reductase from parsley by peptide sequence comparison after a two-step purification. Protein Expr Purif 9:10–14. https://doi.org/10.1006/prep.1996.0666

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

Karol KG, Arumuganathan K, Boore JL, Duffy AM, Everett KDE, Hall JD, Hansen SK, Kuehl JV, Mandoli DF, Mishler BD, Olmstead RG, Renzaglia KS, Wolf PG (2010) Complete plastome sequences of *Equisetum arvense* and Isoetes flaccida: implications for phylogeny and plastid genome evolution of early land plant lineages. BMC Evol Biol 10:321. https://doi.org/10.1186/1471-2148-10-321

Karwatzki B, Petersen M, Alfermann A (1989) Transient activity of enzymes involved in the biosynthesis of rosmarinic acid in cell suspension cultures of *Coleus blumei*. Planta Med 55:663–664. https://doi.org/10.1055/s-2006-962258

Kawai S, Hanai K, Ito K, Kitahara S, Kuwae A (1991) High-performance liquid chromatographic separation of *p*-hydroxyphenylpyruvic acid. J Chromatogr A 585:318–321. https://doi.org/10.1016/0021-9673(91)85094-V

Kenrick P (1994) Alternation of generations in land plants: New phylogenetic and palaeobotanical evidence. Biol Rev 69:293–330. https://doi.org/10.1111/j.1469-185X.1994.tb01273.x

Kenrick P (2018) Changing expressions: a hypothesis for the origin of the vascular plant life cycle. Philos Trans R Soc Lond B Biol Sci 373. https://doi.org/10.1098/rstb.2017.0149

Kenrick P, Crane PR (1997b) The origin and early evolution of plants on land. Nature 389:33–39. https://doi.org/10.1038/37918

Kenrick P, Wellman CH, Schneider H, Edgecombe GD (2012) A timeline for terrestrialization: consequences for the carbon cycle in the Palaeozoic. Philos Trans R Soc Lond B Biol Sci 367:519–536. https://doi.org/10.1098/rstb.2011.0271

Khojasteh A, Mirjalili MH, Alcalde MA, Cusido RM, Eibl R, Palazon J (2020) Powerful plant antioxidants: A new biosustainable approach to the production of rosmarinic acid. Antioxidants 9. https://doi.org/10.3390/antiox9121273

Kikuchi H, Hirose S, Toki S, Akama K, Takaiwa F (1999) Molecular characterization of a gene for alanine aminotransferase from rice (*Oryza sativa*). Plant Mol Biol 39:149–159. https://doi.org/10.1023/a:1006156214716

Kim YB, Shin Y, Tuan PA, Li X, Park Y, Park N, Park SU (2014a) Molecular cloning and characterization of genes involved in rosmarinic acid biosynthesis from *Prunella vulgaris*. Biol Pharm Bull 37:1221–1227. https://doi.org/10.1248/bpb.b14-00139

Kim YB, Uddin MR, Kim Y, Park CG, Park SU (2014b) Molecular cloning and characterization of tyrosine aminotransferase and hydroxyphenylpyruvate reductase, and rosmarinic acid accumulation in *Scutellaria baicalensis*. Nat Prod Commun 9:1934578X1400900. https://doi.org/10.1177/1934578X1400900923

Kim KH, Janiak V, Petersen M (2004) Purification, cloning and functional expression of hydroxyphenylpyruvate reductase involved in rosmarinic acid biosynthesis in cell cultures

of *Coleus blumei*. Plant Mol Biol 54:311–323. https://doi.org/10.1023/B:PLAN.0000036367.03056.b2

Kirsch JF, Eichele G, Ford GC, Vincent MG, Jansonius JN, Gehring H, Christen P (1984) Mechanism of action of aspartate aminotransferase proposed on the basis of its spatial structure. J Mol Biol 174:497–525. https://doi.org/10.1016/0022-2836(84)90333-4

Kleczkowski LA, Edwards GE (1989) Identification of hydroxypyruvate and glyoxylate reductases in maize leaves. Plant Physiol 91:278–286. https://doi.org/10.1104/pp.91.1.278

Kleczkowski LA, Randall DD (1988) Purification and characterization of a novel NADPH(NADH)-dependent hydroxypyruvate reductase from spinach leaves. Comparison of immunological properties of leaf hydroxypyruvate reductases. Biochem J 250:145–152. https://doi.org/10.1042/bj2500145

Kleczkowski LA, Givan CV, Hodgson JM, Randall DD (1988) Subcellular location of NADPH-dependent hydroxypyruvate reductase activity in leaf protoplasts of *Pisum sativum* L. and its role in photorespiratory metabolism. Plant Physiol 88:1182–1185. https://doi.org/10.1104/pp.88.4.1182

Kleczkowski LA, Randall DD, Edwards GE (1991) Oxalate as a potent and selective inhibitor of spinach (*Spinacia oleracea*) leaf NADPH-dependent hydroxypyruvate reductase. Biochem J 276 (Pt 1):125–127. https://doi.org/10.1042/bj2760125

Ko TP, Wu SP, Yang WZ, Tsai H, Yuan HS (1999) Crystallization and preliminary crystallographic analysis of the *Escherichia coli* tyrosine aminotransferase. Acta Crystallogr D Biol Crystallogr 55:1474–1477. https://doi.org/10.1107/s0907444999006630

Kochhar S, Hunziker PE, Leong-Morgenthaler P, Hottinger H (1992) Evolutionary relationship of NAD+-dependent D-lactate dehydrogenase: Comparison of primary structure of 2-hydroxy acid dehydrogenases. Biochem Biophys Res Commun 184:60–66. https://doi.org/10.1016/0006-291X(92)91157-L

Kroken SB, Graham LE, Cook ME (1996) Occurrence and evolutionary significance of resistant cell walls in charophytes and bryophytes. Am J Bot 83:1241–1254. https://doi.org/10.1002/j.1537-2197.1996.tb13908.x

Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874. https://doi.org/10.1093/molbev/msw054

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

Landete JM (2012) Plant and mammalian lignans: A review of source, intake, metabolism, intestinal bacteria and health. Food Res Internat 46:410–424. https://doi.org/10.1016/j.foodres.2011.12.023

Lecointre G, Le Guyader H (2006) The tree of life: A phylogenetic classification. Belknap Press of Harvard Univ. Pr, Cambridge, Mass.

Lee EJ, Facchini PJ (2011) Tyrosine aminotransferase contributes to benzylisoquinoline alkaloid biosynthesis in opium poppy. Plant Physiol 157:1067–1078. https://doi.org/10.1104/pp.111.185512

Levsh O, Pluskal T, Carballo V, Mitchell AJ, Weng J-K (2019) Independent evolution of rosmarinic acid biosynthesis in two sister families under the Lamiids clade of flowering plants. J Biol Chem 294:15193–15205. https://doi.org/10.1074/jbc.RA119.010454

Lewis LA, Mishler BD, Vilgalys R (1997) Phylogenetic relationships of the liverworts (Hepaticae), a basal embryophyte lineage, inferred from nucleotide sequence data of the chloroplast gene rbcL. Mol Phylogenet Evol 7:377–393. https://doi.org/10.1006/mpev.1996.0395

Lewis NG, Davin LB (1999) Lignans: Biosynthesis and function. In: Comprehensive Natural Products Chemistry. Elsevier, pp 639–712

Li FW, Nishiyama T, Waller M, Frangedakis E, Keller J, Li Z, Fernandez-Pozo N, Barker MS, Bennett T, Blázquez MA, Cheng S, Cuming AC, Vries J de, Vries S de, Delaux PM, Diop IS, Harrison CJ, Hauser D, Hernández-García J, Kirbis A, Meeks JC, Monte I, Mutte SK, Neubauer A, Quandt D, Robison T, Shimamura M, Rensing SA, Villarreal JC, Weijers D, Wicke S, Wong GKS, Sakakibara K, Szövényi P (2020) *Anthoceros* genomes illuminate the origin of land plants and the unique biology of hornworts. Nat Plants 6:259–272. https://doi.org/10.1038/s41477-020-0618-2

Liepman AH, Olsen LJ (2003) Alanine aminotransferase homologs catalyze the glutamate:glyoxylate aminotransferase reaction in peroxisomes of *Arabidopsis*. Plant Physiol 131:215–227. https://doi.org/10.1104/pp.011460

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

Ligrone R, Duckett JG, Renzaglia KS (2012) Major transitions in the evolution of early land plants: a bryological perspective. Ann Bot 109:851–871. https://doi.org/10.1093/aob/mcs017

Litvinenko VI, Popova TP, Simonjan AV, Zoz IG, Sokolov VS (1975) "Gerbstoffe" Und Oxyzimtsäureabkömmlinge in Labiaten ("Tannins" and derivatives of hydroxycinnamic acid in labiatae (author's transl)). Planta Med 27:372–380. https://doi.org/10.1055/s-0028-1097817

Liu J-W, Li S-F, Wu C-T, Valdespino IA, Ho J-F, Wu Y-H, Chang H-M, Guu T-Y, Kao M-F, Chesson C, Das S, Oppenheimer H, Bakutis A, Saenger P, Salazar Allen N, Yong JWH, Adjie B, Kiew R, Nadkarni N, Huang C-L, Chesson P, Sheue C-R (2020) Gigantic chloroplasts, including bizonoplasts, are common in shade-adapted species of the ancient vascular plant family Selaginellaceae. Am J Bot 107:562–576. https://doi.org/10.1002/ajb2.1455

Lu X, Hao L, Wang F, Huang C, Wu S (2013a) Molecular cloning and overexpression of the tyrosine aminotransferase (TAT) gene leads to increased rosmarinic acid yield in *Perilla frutescens*. Plant Cell Tiss Org 115:69–83. https://doi.org/10.1007/s11240-013-0341-z

Lu X, Hao L, Wang F, Huang C (2013b) Molecular cloning, characterization and expression analysis of a gene encoding hydroxyphenylpyruvate reductase involved in rosmarinic acid biosynthesis pathway from *Perilla frutescens*. In: Zhang T-C, Ouyang P, Kaplan S, Skarnes B (eds) Proceedings of the 2012 International Conference on Applied Biotechnology (ICAB 2012), vol 251. Springer, Dordrecht, pp 1807–1820

Lucas JR, Renzaglia KS (2002) Structure and function of hornwort stomata. Microsc Microanal 8:1090–1091. https://doi.org/10.1017/S143192760210701X

Ludwiczuk A, Asakawa Y (2019) Bryophytes as a source of bioactive volatile terpenoids - A review. Food Chem Toxicol 132:110649. https://doi.org/10.1016/j.fct.2019.110649

MacRae W, Towers G (1984) Biological activities of lignans. Phytochemistry 23:1207–1220. https://doi.org/10.1016/S0031-9422(00)80428-8

Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res 47:W636-W641. https://doi.org/10.1093/nar/gkz268

Maeda H, Dudareva N (2012) The shikimate pathway and aromatic amino acid biosynthesis in plants. Annu Rev Plant Biol 63:73–105. https://doi.org/10.1146/annurev-arplant-042811-105439

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

Martínez-Abaigar J, Núñez-Olivera E (2021) Novel biotechnological substances from bryophytes. In: Natural Bioactive Compounds. Elsevier, pp 233–248

Matheron ME, Moore TC (1973) Properties of an aminotransferase of pea (*Pisum sativum* L.). Plant Physiol 52:63–67. https://doi.org/10.1104/pp.52.1.63

Matsuno M, Nagatsu A, Ogihara Y, Ellis BE, Mizukami H (2002) CYP98A6 from *Lithospermum erythrorhizon* encodes 4-coumaroyl-4'-hydroxyphenyllactic acid 3-hydroxylase involved in rosmarinic acid biosynthesis 1. FEBS Lett 514:219–224. https://doi.org/10.1016/S0014-5793(02)02368-2

Mehere P, Han Q, Lemkul JA, Vavricka CJ, Robinson H, Bevan DR, Li J (2010) Tyrosine aminotransferase: biochemical and structural properties and molecular dynamics simulations. Protein Cell 1:1023–1032. https://doi.org/10.1007/s13238-010-0128-5

Mehta PK, Hale TI, Christen P (1989) Evolutionary relationships among aminotransferases. Tyrosine aminotransferase, histidinol-phosphate aminotransferase, and aspartate aminotransferase are homologous proteins. Eur J Biochem 186:249–253. https://doi.org/10.1111/j.1432-1033.1989.tb15202.x

Mehta PK, Hale TI, Christen P (1993) Aminotransferases: demonstration of homology and division into evolutionary subgroups. Eur J Biochem 214:549–561. https://doi.org/10.1111/j.1432-1033.1993.tb17953.x

Meinhard J, Petersen M, Alfermann AW (1992) Purification of hydroxyphenylpyruvate reductase from cell cultures of *Coleus blumei*. Planta Med 58:598–599

Meinhard EJ (1991) Versuche zur Reinigung der Hydroxyphenylpyruvatreduktase aus Zellkulturen von *Coleus blumei*. Diploma thesis, University of Düsseldorf, German.

Méndez J, Sanz-Cabanilles F (1979) Cinnamic acid esters in *Anthoceros* species. Phytochemistry 18:1409. https://doi.org/10.1016/0031-9422(79)83039-3

Mishler BD, Churchill SP (1984) A Cladistic approach to the phylogeny of the "bryophytes". Brittonia 36:406. https://doi.org/10.2307/2806602

Mishler BD, Lewis LA, Buchheim MA, Renzaglia KS, Garbary DJ, Delwiche CF, Zechman FW, Kantz TS, Chapman RL (1994) Phylogenetic relationships of the "green algae" and "bryophytes". Ann Missouri Bot Garden 81:451. https://doi.org/10.2307/2399900

Mizukami H, Ellis BE (1991) Rosmarinic acid formation and differential expression of tyrosine aminotransferase isoforms in *Anchusa officinalis* cell suspension cultures. Plant Cell Rep 10:321–324. https://doi.org/10.1007/BF00193150

Mizukami H, Tabira Y, Ellis BE (1993) Methyl jasmonate-induced rosmarinic acid biosynthesis in *Lithospermum erythrorhizon* cell suspension cultures. Plant Cell Rep 12:706–709. https://doi.org/10.1007/BF00233424

Murray AJ, Blackwell RD, Lea PJ (1989) Metabolism of hydroxypyruvate in a mutant of barley lacking NADH-dependent hydroxypyruvate reductase, an important photorespiratory enzyme activity. Plant Physiol 91:395–400. https://doi.org/10.1104/pp.91.1.395

Nagashima F, Tanase S, Fukumoto Y, Joh T, Nomiyama H, Tsuzuki T, Shimada K, Kuramitsu S, Kagamiyama H, Morino Y (1989) cDNA cloning and expression of pig cytosolic aspartate aminotransferase in *Escherichia coli*: amino-terminal heterogeneity of expressed products and lack of its correlation with enzyme function. Biochem J 28:1153–1160. https://doi.org/10.1021/bi00429a033

Nickrent DL, Parkinson CL, Palmer JD, Duff RJ (2000) Multigene phylogeny of land plants with special reference to bryophytes and the earliest land plants. Mol Biol Evol 17:1885–1895. https://doi.org/10.1093/oxfordjournals.molbev.a026290

Niklas KJ, Kutschera U (2010) The evolution of the land plant life cycle. New Phytol 185:27–41. https://doi.org/10.1111/j.1469-8137.2009.03054.x

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, Wolf PG, Kugita M, Sinclair RB, Sugita M, Sugiura C, Wakasugi T, Yamada K, Yoshinaga K, Yamaguchi K, Ueda K, Hasebe M (2004) Chloroplast phylogeny indicates that bryophytes are monophyletic. Mol Biol Evol 21:1813–1819. https://doi.org/10.1093/molbev/msh203

Okamoto A, Nakai Y, Hayashi H, Hirotsu K, Kagamiyama H (1998) Crystal structures of *Paracoccus denitrificans* aromatic amino acid aminotransferase: a substrate recognition site constructed by rearrangement of hydrogen bond network. J Mol Biol 280:443–461. https://doi.org/10.1006/jmbi.1998.1869

One Thousand Plant Transcriptomes Initiative 2019: One thousand plant transcriptomes and the phylogenomics of green plants 2019. Nature 574:679–685. https://doi.org/10.1038/s41586-019-1693-2

References

Oue S, Okamoto A, Nakai Y, Nakahira M, Shibatani T, Hayashi H, Kagamiyama H (1997) *Paracoccus denitrificans* aromatic amino acid aminotransferase: a model enzyme for the study of dual substrate recognition mechanism. J Biochem 121:161–171. https://doi.org/10.1093/oxfordjournals.jbchem.a021561

Ovchinnikov Y, Egorov CA, Aldanova NA, Feigina M, Lipkin VM, Abdulaev NG, Grishin EV, Kiselev AP, Modyanov NN, Braunstein AE, Polyanovsky OL, Nosikov VV (1973) The complete amino acid sequence of cytoplasmic aspartate aminotransferase from pig heart. FEBS Lett 29:31–34. https://doi.org/10.1016/0014-5793(73)80008-0

Painter HA, Zilva SS (1947) The tautomeric conversion of *p*-hydroxyphenylpyruvic acid. Biochem J 41:520–522. https://doi.org/10.1042/bj0410520

Parthasarathy A, Cross PJ, Dobson RCJ, Adams LE, Savka MA, Hudson AO (2018) A three-ring circus: Metabolism of the three proteogenic aromatic amino acids and their role in the health of plants and animals. Front Mol Biosci 5:29. https://doi.org/10.3389/fmolb.2018.00029

Parthasarathy A, Adams LE, Savka FC, Hudson AO (2019) The *Arabidopsis thaliana* gene annotated by the locus tag At3g08860 encodes alanine aminotransferase. Plant Direct 3:e00171. https://doi.org/10.1002/pld3.171

Petersen M (1997) Cytochrome P450-dependent hydroxylation in the biosynthesis of rosmarinic acid in *Coleus*. Phytochemistry 45:1165–1172. https://doi.org/10.1016/S0031-9422(97)00135-0

Petersen M, Alfermann AW (1988) Two New Enzymes of Rosmarinic Acid Biosynthesis from Cell Cultures of *Coleus blumei*: Hydroxyphenylpyruvate Reductase and Rosmarinic Acid Synthase. Zeitschrift für Naturforschung 43c:501–504. https://doi.org/10.1515/znc-1988-7-804

Petersen M, Alfermann AW (2001) The production of cytotoxic lignans by plant cell cultures. Appl Microbiol Biotechnol 55:135–142. https://doi.org/10.1007/s002530000510

Petersen M, Häusler E, Karwatzki B, Meinhard J (1993) Proposed biosynthetic pathway for rosmarinic acid in cell cultures of *Coleus blumei* Benth. Planta 189. https://doi.org/10.1007/BF00201337

Petersen M (2003) Cinnamic acid 4-hydroxylase from cell cultures of the hornwort *Anthoceros agrestis*. Planta 217:96–101. https://doi.org/10.1007/s00425-002-0960-9

References

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

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

Petersen M (2016) Hydroxycinnamoyltransferases in plant metabolism. Phytochem Rev 15:699–727. https://doi.org/10.1007/s11101-015-9417-1

Petersen M, Abdullah Y, Benner J, Eberle D, Gehlen K, Hücherig S, Janiak V, Kim KH, Sander M, Weitzel C, Wolters S (2009) Evolution of rosmarinic acid biosynthesis. Phytochemistry 70:1663–1679. https://doi.org/10.1016/j.phytochem.2009.05.010

Petersen MS (1991) Characterization of rosmarinic acid synthase from cell cultures of *Coleus blumei*. Phytochemistry 30:2877–2881. https://doi.org/10.1016/S0031-9422(00)98217-7

Pezeshki S (2016) Biosynthese von Kaffeesäuremetaboliten im Ackerhornmoos *Anthoceros agrestis* und im Kleinen Blasenmützenmoos *Physcomitrella patens*. Dissertation, Philipps University of Marburg

Pezeshki S, Petersen M (2018) Rosmarinic acid and related metabolites. In: Schwab W, Lange BM, Wüst M (eds) Biotechnology of natural products. Springer International Publishing, Cham, pp 25–60

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

Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29:e45. https://doi.org/10.1093/nar/29.9.e45

Pieck M, Yuan Y, Godfrey J, Fisher C, Zolj S, Vaughan D, Thomas N, Wu C, Ramos J, Lee N, Normanly J, Celenza JL (2015) Auxin and tryptophan homeostasis are facilitated by the ISS1/VAS1 aromatic aminotransferase in *Arabidopsis*. Genetics 201:185–199. https://doi.org/10.1534/genetics.115.180356

Pitt BM (1962) Oxidation of phenylpyruvates to aromatic aldehydes and oxalate. Nature 196:272–273. https://doi.org/10.1038/196272a0

Popov VO, Lamzin VS (1994) NAD(+)-dependent formate dehydrogenase. Biochem J 301 (Pt 3):625–643. https://doi.org/10.1042/bj3010625

Prabhu PR, Hudson AO (2010) Identification and partial characterization of an L-tyrosine aminotransferase (TAT) from *Arabidopsis thaliana*. Biochem Res Int 2010:549572. https://doi.org/10.1155/2010/549572

Pressel S, Goral T, Duckett JG (2014) Stomatal differentiation and abnormal stomata in hornworts. J Bryol 36:87–103. https://doi.org/10.1179/1743282014Y.0000000103

Pressel S, Renzaglia KS, Dicky Clymo RS, Duckett JG (2018) Hornwort stomata do not respond actively to exogenous and environmental cues. Ann Bot 122:45–57. https://doi.org/10.1093/aob/mcy045

Puttick MN, Morris JL, Williams TA, Cox CJ, Edwards D, Kenrick P, Pressel S, Wellman CH, Schneider H, Pisani D, Donoghue PCJ (2018) The interrelationships of land plants and the nature of the ancestral embryophyte. Curr Biol 28:733-745.e2. https://doi.org/10.1016/j.cub.2018.01.063

Qiu Y-L, Li L, Wang B, Chen Z, Knoop V, Groth-Malonek M, Dombrovska O, Lee J, Kent L, Rest J, Estabrook GF, Hendry TA, Taylor DW, Testa CM, Ambros M, Crandall-Stotler B, Duff RJ, Stech M, Frey W, Quandt D, Davis CC (2006) The deepest divergences in land plants inferred from phylogenomic evidence. Proc Natl Acad Sci USA 103:15511–15516. https://doi.org/10.1073/pnas.0603335103

Qiu YL, Cho Y, Cox JC, Palmer JD (1998) The gain of three mitochondrial introns identifies liverworts as the earliest land plants. Nature 394:671–674. https://doi.org/10.1038/29286

Raines CA (2003) The Calvin cycle revisited. Photosynth Res 75:1–10. https://doi.org/10.1023/A:1022421515027

Ravn H, Pedersen MF, Borum J, Andary C, Anthoni U, Christophersen C, Nielsen PH (1994) Seasonal variation and distribution of two phenolic compounds, rosmarinic acid and caffeic acid, in leaves and roots-rhizomes of eelgrass (*Zostera marina* L.). Ophelia 40:51–61. https://doi.org/10.1080/00785326.1994.10429550

Razzaque A, Ellis BE (1977) Rosmarinic acid production in *Coleus cell* cultures. Planta 137:287–291. https://doi.org/10.1007/BF00388164

Read DJ, Ducket JG, Francis R, Ligron R, Russell A (2000) Symbiotic fungal associations in 'lower' land plants. Philos Trans R Soc Lond B Biol Sci 355:815-30; discussion 830-1. https://doi.org/10.1098/rstb.2000.0617

References

Renault H, Alber A, Horst NA, Basilio Lopes A, Fich EA, Kriegshauser L, Wiedemann G, Ullmann P, Herrgott L, Erhardt M, Pineau E, Ehlting J, Schmitt M, Rose JKC, Reski R, Werck-Reichhart D (2017) A phenol-enriched cuticle is ancestral to lignin evolution in land plants. Nat Commun 8:14713. https://doi.org/10.1038/ncomms14713

Renault H, Werck-Reichhart D, Weng 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

Rensing SA (2018) Great moments in evolution: the conquest of land by plants. Curr Opin Plant Biol 42:49–54. https://doi.org/10.1016/j.pbi.2018.02.006

Renzaglia KS, Duff RJ, Nickrent DL, Garbary DJ (2000) Vegetative and reproductive innovations of early land plants: implications for a unified phylogeny. Philos Trans R Soc Lond B Biol Sci 355:769–793. https://doi.org/10.1098/rstb.2000.0615

Renzaglia KS, Garbary DJ (2001) Motile gametes of land plants: diversity, development, and evolution. Crit Rev Plant Sci 20:107–213. https://doi.org/10.1080/20013591099209

Renzaglia KS, Villarreal JC, Duff RJ, Goffinet B (2008) New insights into morphology, anatomy, and systematics of hornworts. In: Shaw AJ, Goffinet B, Shaw AJ (eds) Bryophyte biology. Cambridge University Press, Cambridge, pp 139–172

Renzaglia KS, Villareal JC, Garbary DJ (2018) Morphology supports the setaphyte hypothesis: mosses plus liverworts form a natural group. Bryophyte Diversity and Evolution 40:11. https://doi.org/10.11646/bde.40.2.1

Renzaglia KS, Browning WB, Merced A (2020) With over 60 independent losses, stomata are expendable in mosses. Front Plant Sci 11:567. https://doi.org/10.3389/fpls.2020.00567

Reumann S, Weber APM (2006) Plant peroxisomes respire in the light: some gaps of the photorespiratory C2 cycle have become filled--others remain. Biochim Biophys Acta 1763:1496–1510. https://doi.org/10.1016/j.bbamcr.2006.09.008

Rieseberg TP, Dadras A, Fürst-Jansen JMR, Dhabalia Ashok A, Darienko T, Vries S de, Irisarri I, Vries J de (2022) Crossroads in the evolution of plant specialized metabolism. Semin Cell Dev Biol. https://doi.org/10.1016/j.semcdb.2022.03.004

Riewe D, Koohi M, Lisec J, Pfeiffer M, Lippmann R, Schmeichel J, Willmitzer L, Altmann T (2012) A tyrosine aminotransferase involved in tocopherol synthesis in *Arabidopsis*. Plant J 71:850–859. https://doi.org/10.1111/j.1365-313X.2012.05035.x

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

Ru M, Wang K, Bai Z, Peng L, He S, Wang Y, Liang Z (2017) A tyrosine aminotransferase involved in rosmarinic acid biosynthesis in *Prunella vulgaris* L. Sci Rep 7:4892. https://doi.org/10.1038/s41598-017-05290-4

Rubin JL, Jensen RA (1979) Enzymology of L-Tyrosine Biosynthesis in Mung Bean (*Vigna radiata* L. Wilczek). Plant Physiol 64:727–734. https://doi.org/10.1104/pp.64.5.727

Rubinstein CV, Gerrienne P, La Puente GS de, Astini RA, Steemans P (2010) Early Middle Ordovician evidence for land plants in Argentina (eastern Gondwana). New Phytol 188:365–369. https://doi.org/10.1111/j.1469-8137.2010.03433.x

Ruhfel BR, Gitzendanner MA, Soltis PS, Soltis DE, Burleigh JG (2014) From algae to angiosperms-inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. BMC Evol Biol 14:23. https://doi.org/10.1186/1471-2148-14-23

Salim AA, Chin Y-W, Kinghorn AD (2008) Drug discovery from plants. In: Ramawat K, Merillon J (eds) Bioactive Molecules and Medicinal Plants. Springer, Berlin, Heidelberg

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

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

Schenck CA, Maeda HA (2018) Tyrosine biosynthesis, metabolism, and catabolism in plants. Phytochemistry 149:82–102. https://doi.org/10.1016/j.phytochem.2018.02.003

Schenck CA, Chen S, Siehl DL, Maeda HA (2015) Non-plastidic, tyrosine-insensitive prephenate dehydrogenases from legumes. Nat Chem Biol 11:52–57. https://doi.org/10.1038/nchembio.1693

Schultz CJ, Coruzzi GM (1995) The aspartate aminotransferase gene family of *Arabidopsis* encodes isoenzymes localized to three distinct subcellular compartments. Plant J 7:61–75. https://doi.org/10.1046/j.1365-313x.1995.07010061.x

Shaw J, Renzaglia K (2004) Phylogeny and diversification of bryophytes. Am J Bot 91:1557–1581. https://doi.org/10.3732/ajb.91.10.1557

Shaw AJ, Szövényi P, Shaw B (2011) Bryophyte diversity and evolution: windows into the early evolution of land plants. Am J Bot 98:352–369. https://doi.org/10.3732/ajb.1000316

Shen J, Zeng Y, Zhuang X, Sun L, Yao X, Pimpl P, Jiang L (2013) Organelle pH in the *Arabidopsis* endomembrane system. Mol Plant 6:1419–1437. https://doi.org/10.1093/mp/sst079

Simmonds MSJ, Stevenson PC, Hanson FE (2019) Rosmarinic acid in *Canna generalis* activates the medial deterrent chemosensory neurone and deters feeding in the tobacco hornworm *Manduca sexta*. Physiol Entomol 44:140–147. https://doi.org/10.1111/phen.12284

Smith E, Griffiths H (1996) A pyrenoid-based carbon-concentrating mechanism is present in terrestrial bryophytes of the class Anthocerotae. Planta 200:203–212. https://doi.org/10.1007/BF00208310

Söderström L, Hagborg A, Konrat M von, Bartholomew-Began S, Bell D, Briscoe L, Brown E, Cargill DC, Costa DP, Crandall-Stotler BJ, Cooper ED, Dauphin G, Engel JJ, Feldberg K, Glenny D, Gradstein SR, He X, Heinrichs J, Hentschel J, Ilkiu-Borges AL, Katagiri T, Konstantinova NA, Larraín J, Long DG, Nebel M, Pócs T, Puche F, Reiner-Drehwald E, Renner MAM, Sass-Gyarmati A, Schäfer-Verwimp A, Moragues JGS, Stotler RE, Sukkharak P, Thiers BM, Uribe J, Váňa J, Villarreal JC, Wigginton M, Zhang L, Zhu R-L (2016) World checklist of hornworts and liverworts. PhytoKeys:1–828. https://doi.org/10.3897/phytokeys.59.6261

Song J, Wang Z (2011) RNAi-mediated suppression of the phenylalanine ammonia-lyase gene in *Salvia miltiorrhiza* causes abnormal phenotypes and a reduction in rosmarinic acid biosynthesis. J Plant Res 124:183–192. https://doi.org/10.1007/s10265-010-0350-5

Song J, Ji Y, Xu K, Wang Z (2012) An integrated analysis of the rosmarinic acid-biosynthetic genes to uncover the regulation of rosmarinic acid pathway in *Salvia miltiorrhiza*. Acta Physiol Plant 34:1501–1511. https://doi.org/10.1007/s11738-012-0948-4

Song Z, Li X (2015) Expression profiles of rosmarinic acid biosynthesis genes in two *Salvia miltiorrhiza* lines with differing water-soluble phenolic contents. Ind Crop Prod 71:24–30. https://doi.org/10.1016/j.indcrop.2015.03.081

Sørensen I, Pettolino FA, Bacic A, Ralph J, Lu F, O'Neill MA, Fei Z, Rose JKC, Domozych DS, Willats WGT (2011) The charophycean green algae provide insights into the early origins of plant cell walls. The Plant Journal 68:201–211. https://doi.org/10.1111/j.1365-313X.2011.04686.x

Sousa F de, Foster PG, Donoghue PCJ, Schneider H, Cox CJ (2019) Nuclear protein phylogenies support the monophyly of the three bryophyte groups (Bryophyta Schimp.). New Phytol 222:565–575. https://doi.org/10.1111/nph.15587

Sousa F, Civáň P, Brazão J, Foster PG, Cox CJ (2020) The mitochondrial phylogeny of land plants shows support for Setaphyta under composition-heterogeneous substitution models. PeerJ 8:e8995. https://doi.org/10.7717/peerj.8995

Stapleton AE (1992) Ultraviolet radiation and plants: Burning questions. Plant Cell 4:1353–1358. https://doi.org/10.1105/tpc.4.11.1353

Stebbins GL, Hill GJC (1980) Did multicellular plants invade the land? Am Nat 115:342–353. https://doi.org/10.1086/283565

Steemans P, Le Hérissé A, Melvin J, Miller MA, Paris F, Verniers J, Wellman CH (2009) Origin and radiation of the earliest vascular land plants. Science 324:353. https://doi.org/10.1126/science.1169659

Sterbová D, Vlcek J, Kubán V (2006) Capillary zone electrophoretic determination of phenolic compounds in chess (*Bromus inermis* L.) plant extracts. J Sep Sci 29:308–313. https://doi.org/10.1002/jssc.200500297

Stoll VS, Kimber MS, Pai EF (1996) Insights into substrate binding by D-2-ketoacid dehydrogenases from the structure of *Lactobacillus pentosus* D-lactate dehydrogenase. Structure 4:437–447. https://doi.org/10.1016/S0969-2126(96)00049-4

Su D, Yang L, Shi X, Ma X, Zhou X, Hedges SB, Zhong B (2021) Large-scale phylogenomic analyses reveal the monophyly of bryophytes and neoproterozoic origin of land plants. Mol Biol Evol 38:3332–3344. https://doi.org/10.1093/molbev/msab106

Sussmilch FC, Roelfsema MRG, Hedrich R (2019) On the origins of osmotically driven stomatal movements. New Phytol 222:84–90. https://doi.org/10.1111/nph.15593

Szabo E, Thelen A, Petersen M (1999) Fungal elicitor preparations and methyl jasmonate enhance rosmarinic acid accumulation in suspension cultures of *Coleus blumei*. Plant Cell Rep 18:485–489. https://doi.org/10.1007/s002990050608

Szövényi P, Frangedakis E, Ricca M, Quandt D, Wicke S, Langdale JA (2015) Establishment of *Anthoceros agrestis* as a model species for studying the biology of hornworts. BMC Plant Biol 15:98. https://doi.org/10.1186/s12870-015-0481-x

Taguchi H, Ohta T (1991) D-lactate dehydrogenase is a member of the D-isomer-specific 2-hydroxyacid dehydrogenase family. Cloning, sequencing, and expression in *Escherichia coli* of the D-lactate dehydrogenase gene of *Lactobacillus plantarum*. J Biol Chem 266:12588–12594. https://doi.org/10.1016/S0021-9258(18)98939-8

References

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

Tao Y, Ferrer JL, Ljung K, Pojer F, Hong F, Long JA, Li L, Moreno JE, Bowman ME, Ivans LJ, Cheng Y, Lim J, Zhao Y, Ballaré CL, Sandberg G, Noel JP, Chory J (2008) Rapid synthesis of auxin via a new tryptophan-dependent pathway is required for shade avoidance in plants. Cell 133:164–176. https://doi.org/10.1016/j.cell.2008.01.049

Tartoff KD, Hobbs CA (1987) Improved media for growing plasmid and cosmid clones. Bethesda Res Lab Focus 9:12

Taylor LP, Grotewold E (2005) Flavonoids as developmental regulators. Curr Opin Plant Biol 8:317–323. https://doi.org/10.1016/j.pbi.2005.03.005

Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, Winther O, Brunak S, Heijne G von, Nielsen H (2022) SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol 40:1023–1025. https://doi.org/10.1038/s41587-021-01156-3

Timm S, Nunes-Nesi A, Pärnik T, Morgenthal K, Wienkoop S, Keerberg O, Weckwerth W, Kleczkowski LA, Fernie AR, Bauwe H (2008) A cytosolic pathway for the conversion of hydroxypyruvate to glycerate during photorespiration in *Arabidopsis*. Plant Cell 20:2848–2859. https://doi.org/10.1105/tpc.108.062265

Timm S, Florian A, Jahnke K, Nunes-Nesi A, Fernie AR, Bauwe H (2011) The hydroxypyruvate-reducing system in *Arabidopsis*: multiple enzymes for the same end. Plant Physiol 155:694–705. https://doi.org/10.1104/pp.110.166538

Timme RE, Bachvaroff TR, Delwiche CF (2012) Broad phylogenomic sampling and the sister lineage of land plants. PLoS ONE 7:e29696. https://doi.org/10.1371/journal.pone.0029696

Tishkov VI, Popov VO (2004) Catalytic mechanism and application of formate dehydrogenase. Biochemistry (Mosc) 69:1252–1267. https://doi.org/10.1007/s10541-005-0071-x

Tolbert NE, Yamazaki RK, Oeser A (1970) Localization and properties of hydroxypyruvate and glyoxylate reductases in spinach leaf particles. Journal of Biological Chemistry 245:5129–5136. https://doi.org/10.1016/S0021-9258(18)62827-3

Tomescu AM, Bomfleur B, Bippus AC, Savoretti A (2018) Why are bryophytes so rare in the fossil record? A spotlight on taphonomy and fossil preservation. In: Transformative Paleobotany. Elsevier, pp 375–416

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

Trennheuser F (1992) Phytochemische Untersuchung und *in vitro* Kultur ausgewählter Vertreter der Anthocerotopsida. Dissertation, University of the Saarland

Trennheuser F, Burkhard G, Becker H (1994) Anthocerodiazonin an alkaloid from *Anthoceros agrestis*. Phytochemistry 37:899–903. https://doi.org/10.1016/S0031-9422(00)90380-7

Tsai CS (1967) Spontaneous decarboxylation of oxalacetic acid. Can J Chem 45:873–880. https://doi.org/10.1139/v67-145

Tsumbu CN, Deby-Dupont G, Tits M, Angenot L, Frederich M, Kohnen S, Mouithys-Mickalad A, Serteyn D, Franck T (2012) Polyphenol content and modulatory activities of some tropical dietary plant extracts on the oxidant activities of neutrophils and myeloperoxidase. Int J Mol Sci 13:628–650. https://doi.org/10.3390/ijms13010628

Tuan PA, Park WT, Xu H, Park NI, Park SU (2012) Accumulation of tilianin and rosmarinic acid and expression of phenylpropanoid biosynthetic genes in *Agastache rugosa*. J Agric Food Chem 60:5945–5951. https://doi.org/10.1021/jf300833m

Turmel M, Otis C, Lemieux C (2006) The chloroplast genome sequence of *Chara vulgaris* sheds new light into the closest green algal relatives of land plants. Mol Biol Evol 23:1324–1338. https://doi.org/10.1093/molbev/msk018

Vaughn KC, Ligrone R, Owen HA, Hasegawa J, Campbell EO, Renzaglia KS, Monge-Najera J (1992) The anthocerote chloroplast: A review. New Phytol 120:169–190. https://doi.org/10.1111/j.1469-8137.1992.tb05653.x

Vetter E (1988) Biosynthese der Rosmarinsäure - Enzymatische Untersuchung zur Biosynthese der "Zimtsäure"-CoA-Ester bei Suspensionskulturen von *Coleus blumei.* Diplomarbeit, Tübingen, FRG

Villarreal JC, Cargill DC, Hagborg A, Soderstrom L, Renzaglia KS (2010) A synthesis of hornwort diversity: Patterns, causes and future work. Phytotaxa 9:150. https://doi.org/10.11646/phytotaxa.9.1.8

Villarreal JC, Renner SS (2012) Hornwort pyrenoids, carbon-concentrating structures, evolved and were lost at least five times during the last 100 million years. Proc Natl Acad Sci USA 109:18873–18878. https://doi.org/10.1073/pnas.1213498109

Villarreal JC, Renzaglia KS (2015) The hornworts: important advancements in early land plant evolution. J Bryol 37:157–170. https://doi.org/10.1179/1743282015Y.0000000016

Vogelsang K, Schneider B, Petersen M (2006) Production of rosmarinic acid and a new rosmarinic acid 3'-O-β-D-glucoside in suspension cultures of the hornwort *Anthoceros agrestis* Paton. Planta 223:369–373. https://doi.org/10.1007/s00425-005-0089-8

Wang M, Toda K, Maeda HA (2016) Biochemical properties and subcellular localization of tyrosine aminotransferases in *Arabidopsis thaliana*. Phytochemistry 132:16–25. https://doi.org/10.1016/j.phytochem.2016.09.007

Wang G-Q, Chen J-F, Yi B, Tan H-X, Zhang L, Chen W-S (2017) HPPR encodes the hydroxyphenylpyruvate reductase required for the biosynthesis of hydrophilic phenolic acids in *Salvia miltiorrhiza*. Chin J Nat Med 15:917–927. https://doi.org/10.1016/S1875-5364(18)30008-6

Wang M, Maeda HA (2018) Aromatic amino acid aminotransferases in plants. Phytochem Rev 17:131–159. https://doi.org/10.1007/s11101-017-9520-6

Wang M, Toda K, Block A, Maeda HA (2019) TAT1 and TAT2 tyrosine aminotransferases have both distinct and shared functions in tyrosine metabolism and degradation in *Arabidopsis thaliana*. J Biol Chem 294:3563–3576. https://doi.org/10.1074/jbc.RA118.006539

Wellman CH, Gray J (2000) The microfossil record of early land plants. Philos Trans R Soc Lond B Biol Sci 355:717-31; discussion 731-2. https://doi.org/10.1098/rstb.2000.0612

Wellman CH, Osterloff PL, Mohiuddin U (2003) Fragments of the earliest land plants. Nature 425:282–285. https://doi.org/10.1038/nature01884

Weng J-K, Chapple C (2010) The origin and evolution of lignin biosynthesis. New Phytol 187:273–285. https://doi.org/10.1111/j.1469-8137.2010.03327.x

Whistance GR, Threlfall DR (1970) Biosynthesis of phytoquinones. Homogentisic acid: a precursor of plastoquinones, tocopherols and alpha-tocopherolquinone in higher plants, green algae and blue-green algae. Biochem J 117:593–600. https://doi.org/10.1042/bj1170593

Wickett NJ, Mirarab S, Nguyen N, Warnow T, Carpenter E, Matasci N, Ayyampalayam S, Barker MS, Burleigh JG, Gitzendanner MA, Ruhfel BR, Wafula E, Der JP, Graham SW, Mathews S,

Melkonian M, Soltis DE, Soltis PS, Miles NW, Rothfels CJ, Pokorny L, Shaw AJ, DeGironimo L, Stevenson DW, Surek B, Villarreal JC, Roure B, Philippe H, dePamphilis CW, Chen T, Deyholos MK, Baucom RS, Kutchan TM, Augustin MM, Wang J, Zhang Y, Tian Z, Yan Z, Wu X, Sun X, Wong GK-S, Leebens-Mack J (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

Wilkie SE, Warren MJ (1998) Recombinant expression, purification, and characterization of three isoenzymes of aspartate aminotransferase from *Arabidopsis thaliana*. Protein Expr Purif 12:381–389. https://doi.org/10.1006/prep.1997.0845

Wilkie SE, Roper JM, Smith AG, Warren MJ (1995) Isolation, characterisation and expression of a cDNA clone encoding plastid aspartate aminotransferase from *Arabidopsis thaliana*. Plant Mol Biol 27:1227–1233. https://doi.org/10.1007/BF00020897

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 (2020) Molecular and biochemical investigations of genes and enzymes involved in the phenolic metabolism of the hornwort *Anthoceros agrestis.* Dissertation, Philipps University of Marburg

Wohl J, Petersen M (2020a) Functional expression and characterization of cinnamic acid 4-hydroxylase from the hornwort *Anthoceros agrestis* in *Physcomitrella patens*. Plant Cell Rep 39:597–607. https://doi.org/10.1007/s00299-020-02517-z

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

Xiao Y, Gao S, Di P, Chen J, Chen W, Zhang L (2009) Methyl jasmonate dramatically enhances the accumulation of phenolic acids in *Salvia miltiorrhiza* hairy root cultures. Physiol Plant 137:1–9. https://doi.org/10.1111/j.1399-3054.2009.01257.x

Xiao Y, Zhang L, Gao S, Saechao S, Di P, Chen J, Chen W (2011) The c4h, tat, hppr and hppd genes prompted engineering of rosmarinic acid biosynthetic pathway in *Salvia miltiorrhiza* hairy root cultures. PLoS ONE 6:e29713. https://doi.org/10.1371/journal.pone.0029713

References

Xu J-J, Fang X, Li C-Y, Zhao Q, Martin C, Chen X-Y, Yang L (2018) Characterization of *Arabidopsis thaliana* hydroxyphenylpyruvate reductases in the tyrosine conversion pathway. Front Plant Sci 9:1305. https://doi.org/10.3389/fpls.2018.01305

Xu JJ, Fang X, Li CY, Yang L, Chen XY (2019) General and specialized tyrosine metabolism pathways in plants. aBIOTECH. https://doi.org/10.1007/s42994-019-00006-w

Yoo H, Widhalm JR, Qian Y, Maeda H, Cooper BR, Jannasch AS, Gonda I, Lewinsohn E, Rhodes D, Dudareva N (2013) An alternative pathway contributes to phenylalanine biosynthesis in plants via a cytosolic tyrosine:phenylpyruvate aminotransferase. Nat Commun 4:2833. https://doi.org/10.1038/ncomms3833

Yu C, Huang AH (1986) Conversion of serine to glycerate in intact spinach leaf peroxisomes: Role of malate dehydrogenase. Arch Biochem Biophys 245:125–133. https://doi.org/10.1016/0003-9861(86)90196-7

Zálešák F, Bon DJ-YD, Pospíšil J (2019) Lignans and neolignans: Plant secondary metabolites as a reservoir of biologically active substances. Pharmacol Res 146:104284. https://doi.org/10.1016/j.phrs.2019.104284

Zhang J, Fu XX, Li RQ, Zhao X, Liu Y, Li MH, Zwaenepoel A, Ma H, Goffinet B, Guan YL, Xue JY, Liao YY, Wang QF, Wang QH, Wang JY, Zhang GQ, Wang ZW, Jia Y, Wang MZ, Dong SS, Yang JF, Jiao YN, Guo YL, Kong HZ, Lu AM, Yang HM, Zhang SZ, van de Peer Y, Liu ZJ, Chen ZD (2020) The hornwort genome and early land plant evolution. Nat Plants 6:107–118. https://doi.org/10.1038/s41477-019-0588-4

Zhao S-J, Hu Z-B, Di Liu, Leung FCC (2006) Two divergent members of 4-coumarate:coenzyme A ligase from *Salvia miltiorrhiza* Bunge: cDNA cloning and functional study. J Integr Plant Biol 48:1355–1364. https://doi.org/10.1111/j.1744-7909.2006.00302.x

Zheng Z, Guo Y, Novák O, Dai X, Zhao Y, Ljung K, Noel JP, Chory J (2013) Coordination of auxin and ethylene biosynthesis by the aminotransferase VAS1. Nat Chem Biol 9:244–246. https://doi.org/10.1038/nchembio.1178.

Zhou Z, Tan H, Li Q, Chen J, Gao S, Wang Y, Chen W, Zhang L (2018) CRISPR/Cas9-mediated efficient targeted mutagenesis of RAS in *Salvia miltiorrhiza*. Phytochemistry 148:63–70. https://doi.org/10.1016/j.phytochem.2018.01.015

Zhu L, Li Y, Yang J, Zuo L, Zhang D (2008) Studies on chemical constituents of *Sarcandra glabra*. Zhongguo Zhong Yao Za Zhi:33:155–157

References

Zinsmeister HD, Becker H, Eicher T (1991) Bryophytes, a source of biologically active, naturally occurring material? Angew Chem Int Ed 30:130–147. https://doi.org/10.1002/anie.199101301

VIII Appendix

1 Sequences

In the following, the sequences of the heterologously expressed proteins are shown. All genes were introduced into pET-15b, adding an *N*-terminal His6 tag to the protein formed.

1.1 *AaTAT* partial fragment

TGAGCACCTGCCGTTCGAGCTTACCCCTGAGGATGTCGGCATTGTGGTGGGATGCTCCCAAGCCATTGAGTTCTCGATCGC
GTGCCTCGCGGCGGAAGGGAGCAACATGCTGGTCCCCAGGCCCGGGTTCCCGATTTACGATACCTTCTGTAGATATTACG
GTGTCGAAGTGAGATATTACGACCTACTTCCCGAGAGGGGCTGGGAGGTGGACCTCGAGCAGGTTGCGGAGCTGGCCGA
CGATGGTACGGCTGCCATGATCCTGTGCAACCCGAGCAACCCGTGTGGCACCTCCTTCAGTTACCAGCACCTGTCTCAGAT
TGCTGGTTTGTGTGAGAAGCTGAAACTGCCCATAATTTCCGACGAGATCTACGAGCATATGCTTTTCGGAGAGAAAAAGTT
CACTCCGATGGCCACTTTTTCCTTGCAAGTGCCGGTGTTGACTGTCGGAGGAATTTCCAAAAGGTGGTTGGCCCCAGGGTG
GCGGTTGGGATGGATCCTCATTTGA

1.2 Constructed putative full-length *AaTAT* based on RACE-PCR

ATGGCCACAGTCCTGAACCAGAAGCGGCAGGCAGAGATCCTTCACGGCAACGGAGTGGAAGGCGGCGGTGGCGGATTG
GCTGTCAGTAAGATTATGTCCAAGGCTGTGGTTGGGAAGCCTCCGGTGGCCAGGAAGCCGGCCGGGAAGACTGTGGACA
AGGAGTGGAACGTGCGGCCCCGGATCGCGGCTTTGGAGAGCCGGAATCCGATTCGGGACATTGTGGAAACGAAGCTTAA
GCCGAACCCGAATTTGGGGAAGAAGCCGATTTCTCTGGCGCAAGGTGATCCCACGGTGTACGGACACTTGAAGGTGCCTG
AATCCGCTTGCGCTGCTCTGGCGGAGGTGGCTACCAGCTACAAGTACAATGGCTACGCTCACTCGGCTGGGATTTTGGAG
TGCCGGAGTGCTGTGGCGGACTTCCACTCGGAGCACCTTCCGTTCGAGCTTACCCCTGAGGATGTCGGCATTGTGGTGGG
ATGCTCCCAAGCCATCGAGTTCTCGATCGCGTGCCTCGCGGCGGAAGGGAGCAACATGCTGGTCCCCAGGCCCGGGTTCC
CGATTTACGATACCTTCTGTAGATATTACGGTGTCGAAGTGAGATATTACGACCTACTTCCCGAGAGGGGCTGGGAGGTG
GACCTCGAGCAGGTTGCGGAGCTGGCCGACGATGGTACGGCTGCCATGATCCTGTGCAACCCGAGCAACCCGTGTGGCA
CCTCCTTCAGTTACCAGCACCTGTCTCAGATTGCTGGTTTGTGTGAGAAGCTGAAACTGCCCATAATTTCCGACGAGATCTA
CGAGCATATGCTTTTCGGAGAGAAAAAGTTCACTCCGATGGCCACTTTTTCCTTGCAAGTGCCGGTGTTGACTGTCGGAGG
AATTTCCAAAAGGTGGTTGGCCCCAGGGTGGCGGTTGGGATGGATCCTCATTTGTGACCCATGCTGCATTCTCGAGAAGA
GCGGGGTTGTTGAGGCCCTGAAGAGGATTATGCAGATGACCATCGGGACGTCTGTTCTTGTACAGGCTGCTGTCCCGGCC
ATGCTACAGAACACCACTCCTGAATTCTACAAGCAAACCATGAAGACTCTTGAGGACGGTTGCGACTGCTGTTACAGACGC
ATCCAAGGAATTGTAGGGTTGGATGTGCCGACTAAGCCAGATGGTGCGATGTACATGATGGCTAAGGTGGACCCAAGTG
CATTCAAGGACATCCCGGATGATACCGTGTTTGCCGAAAAGCTTGTGAAGGAGGAGAATATCGTTGTGCTGCCAGGCTCT
GCATTCGGTATTCACAACTGGTTACGACTGGTGTTTGCAACACCTGTATATATGCTGGAGGAGGCATTTGACAGGATAGAG
GCGTTCTGTCTTCGACATGCGTCCGTCCAAAACTAG

1.3 Full-length *AaTAT* (Genbank MN922307)

ATGGCCACAGTCCTGAACCAGAAGCGGCAGGCAGAGATCCTTCACGGCAACGGAGTGGAAGGCGGCGGTGGCGGATTG
GCTGTCAGTAAGATTATGTCCAAGGCTGTGGTTGGGAAGCCTCCGGTGGCCAGGAAGCCGGCCGGGAAGACTGTGGACA
AGGAGTGGAACGTGCGGCCCCGGATCGCGGCTTTGGAGAGCCGGAATCCGATTCGGGACATTGTGGAAACGAAGCTTAA
GCCGAACCCGAATTTGGGGAAGAAGCCGATTTCTCTGGCGCAAGGTGATCCCACGGTGTACGGACACTTGAAGGTGCCTG
AATCCGCTTGCGCTGCTCTGGCGGAGGTGGCTACCAGCTACAAGTACAATGGCTACGCTCACTCGGCTGGGATTTTGGAG
TGCCGGAGTGCTGTGGCGGACTTCCACTCGGAGCACCTTCCGTTCGAGCTTACCCCTGAGGATGTCGGCATTGTGGTGGG
ATGCTCCCAAGCCATCGAGTTCTCGATCGCGTGCCTCGCGGCGGAAGGGAGCAACATGCTGGTCCCCAGGCCCGGGTTCC
CGATTTACGATACCTTCTGTAGATATTACGGTGTCGAAGTGAGATATTACGACCTACTTCCCGAGAGGGGCTGGGAGGTG
GACCTCGAGCAGGTTGCGGAGCTGGCCGACGATGATACGGCTGCCATGATCCTGTGCAACCCGAGCAACCCGTGTGGCAC
CTCCTTCAGTTACCAGCACCTGTCTCAGATTGCTGGTTTGTGTGAGAAGCTGAAACTGCCCATAATTTCCGACGAGATCTAC
GAGCATATGCTTTTCGGAGAGAAAAAGTTCACTCCGATGGCCACTTTTTCCTTGCAAGTGCCGGTGTTGACTGTCGGAGGA
ATTTCCAAAAGGTGGTTGGCCCCAGGGTGGCGGTTGGGATGGATCCTCATTTGTGACCCATGCTGCATTCTCGAGAAGAG
CGGGGTTGTTGAGGCCCTGAAGAGGATTATGCAGATGACCATCGGGACGTCTGTTCTTGTACAGGCTGCTGTCCCGGCCA
TGCTACAGAACACCACTCCTGAATTCTACAAGCAAACCATGAAGACTCTTGAGGACGGTTGCGACTGCTGTTACAGACGCA
TCCAAGGAATTGTAGGGTTGGATGTGCCGACTAAGCCAGATGGTGCGATGTACATGATGGCTAAGGTGGACCCAAGTGC
ATTCAAGGACATCCCGGATGATACCGTGTTTGCCGAAAAGCTTGTGAAGGAGGAGAATATCGTTGTGCTGCCAGGCTCTG
CATTCGGTATTCACAACTGGTTACGACTGGTGTTTGCAACACCTGTATATATGCTGGAGGAGGCATTTGACAGGATAGAGG
CGTTCTGTCTTCGACATGCGTCCGTCCAAAACTAG

1.4 Constructed putative full-length *AaHPPR* based on RACE-PCR

ATGGGGGAAGCACCAAAGGGAAAGGTTTCTTCCTCCGCCTCCGCCTCCGCCTCCGCCTCCTTTTCCTTCGCGTGCTTCTTCT
TCTTCGTCCTCGTCGGGTTTCCGAGCTTTTYTCAGACGCTGGCGATGGCGGCGGTGGAGGGGAAGGCGCGGCCGGGGGT
GCTGCTGGCGGTGCGGGTGAGCGAGTACCTGGAGAAGGCGCTGGAGAGCAGGTACGCGCTGCACCGGCTGTACGCGGC
GCCATGGGGGGAGGCAGAGCGGGCGGAGTGGCTTGGCGCCATCGCTCCGTCGGTGCGCGCGGTGGTCGGGAGCGCGGC
GGGCGCGGCCTCCGCGGAGCTCATCGCGCGGCTGCCGGCGCTGGAGCTGGTGGCGTGCTACAGCGTGGGAACGGACAA
GGTGCATCTTCCGAGCTGCAAGGAGCGGGGGATCGTGGTGAGCAACACGCCGGACGTGCTCACGGACGACTGCGCGGAT
CTGGCTTTGGCTCTGATGCTCACTTCCTCCCGCCGCCTGGTGGCCGCGGACGGCTACCTCCGGCAGGGGCTGTGGGAGGC
GCAGGGGGACTTCCCGCTGGCGACCAAGGTGTCAGGGAAGCGGGTAGGGATCGTAGGGCTGGGCCGCATCGGGTTGGC
GGTGGCCTGCAGGGCGAGGGCCTTCGGGTGCGCCATCTCCTACCACGGCCGGGCTCGCAAACAGGAGGCGGACTCGCAG
GGTTTTGGCTTCTTCGACACCGCGGAGGGCCTGGCTCGGGAGTCGGACTTCCTCGTCCTCTGCTGCCCGCTCACCCGGGAC
ACCGCGGGGCTGGTGGGCAGGAAGGTGCTTGACGCGCTTGGCCCCAAGGGGACGCTCATCAACATCGCGCGCGGGCCCG
TGGTGGACGAGCCGGAGCTGGTGCTGGCCCTGCAGGAGGGCCGCCTCGGCGCCGCGGGTTTGGATGTTTACCAGGACGA
GCCCCGCGTGCCCAAGGCGCTGTGGGGGTTGGACAATGTTGTGCTGCTCCCGCATGTGGCCAGCGCGACTGTGGAGACTC
GCACAGCCATGGCGGACCTTGTGCTTGAGAACTTGGCCGCGCATTTCGCAGGAAAGCCCCTCGTCACACCTCTCCATTTCT
AG

Note: One sequencing product featured a second ATG codon in the reading frame, enabling a larger CDS (shown in grey).

1.5 Full-length *AaHPPR*

ATGGCGGCGGTGGAGGGGAAGGCGCGGCCGGGGGTGCTGCTGGCGGTGCGGGTGAGCGAGTACCTGGAGAAGGCGCT
GGAGAGCAGGTACGCGCTGCACCGGCTGTACGCGGCGCCATGGGGGGAGGCAGAGCGGGCGGAGTGGCTTGGCGCCAT
CGCTCCGTCGGTGCGCGCGGTGGTCGGGAACGCGGCGGGCGCGGCCTCCGCGGAGCTCATCGCGCGGCTGCCGGCACTG
GAGCTGGTGGCGTGCTACAGCGTGGGAACGGACAAGGTGCATCTTCCGAGCTGCAAGGAGCGGGGGATCGTGGTGAGC
AACACGCCGGACGTGCTCACGGACGACTGCGCGGATCTGGCTTTGGCTCTGATGCTCACTTCCTCCCGCCGCCTGGTGGCC
GCGGACGGCTACCTCCGGCAGGGGCTGTGGGAGGCGCAGGGGGACTTCCCGCTGGCGACCAAGGTGTCAGGGAAGCGG
GTAGGGATCGTAGGGCTGGGCCGCATCGGGTTGGCGGTGGCCTGCAGGGCGAGGGCCTTCGGGTGCGCCATCTCCTACC
ACGGCCGGGCTCGCAAACAGGAGGCGGACTCGCAGGGTTTTGGCTTCTTCGACACCGCGGAGGGCCTGGCTCGGGAGTC
GGACTTCCTCGTCCTCTGCTGCCCGCTCACCCGGGACACCGCGGGGCTGGTGGGCAGGAAGGTGCTTGACGCGCTTGGCC
CCAAGGGGACGCTCATCAACATCGCGCGCGGGCCCGTGGTGGACGAGCCGGAGCTGGTGCTGGCCCTGCAGGAGGGCC
GCCTCGGCGCCGCGGGTTTGGATGTTTACCAGGACGAGCCCCGCGTGCCCAAGGCGCTGTGGGGGTTGGACAATGTTGT
GCTGCTCCCGCATGTGGCCAGCGCGACTGTGGAGACTCGCACAGCCATGGCGGACCTTGTGCTTGAGAACTTGGCCGCGC
ATTTCGCAGGAAAGCCCCTCGTCACACCTCTCCATTTCTAG

1.6 Full-length *AaHPPR2*

ATGTGCGCCTGCTTCTACGCCTTTTCCTCCTCCTGCTCCTGGCCCAGCCAGTGGCCGCCTCCCCGCCTCTCTTTCATCTTCCTC
CTTCCCATCTCCAATCGCCCACCGAGTTGCAATCGCCCAACAAGTTGCTCTACCCCTGCCCCTGCCCCCAGTGTGGGCCGAG
CCCGGCGCCAGAGCCAGGAGAGCTGCATTGCGGAGGCGGAGGCGGATCGGGGACCGGGGGCGTTTGTTGCCGCTGCTG
CAGCCATGGTGGCCACGGGGAAGAAGCCCGGGGTGCTGCAGACGCAGCCCATGCTGGCGGACCTGGAGACGGCGCTCA
GCGCCAAGTACACGCACTACAGGATGTGGGAGGCGGCTGATCCCAAGGCGTTCGTGGCGGAGCACGCGCAGGCGATCCG
GGCTGTGGTGGCCAACGCGTCGGGAGCCAGCGCTGAGCTCATCGACGCGCTGCCCAACTTGGAGATCGTGGCGTCCTTCA
GCGTGGGCGTGAACGCCATCGACATCCCCCGGTGCCGGGAGAAGGGCGTGCTGGTGACCTACACTCCGGACGTGCTCAC
GGATGACTGCGCGGACATGGCGCTGTCGCTGCTGCTCGCCACCGTGCGGAGAGTGTGCGCGATGGACAGATACGTGCGG
CAGGGCATGTGGCCCATTCATGGCGATTTCCCCCTCACTCCCAGGCTGAATGGGAAGCGGCTGGGGATTGTGGGGCTCGG
GAGGATCGGGCTGGCCATTGCCAAGAGGGCGGAGGCTTTCGGGTGCCTCATCTCCTACTATTCCAGATCCGAGAAGGCCA
AGACTCCCTACACCTACTATTCCACGGTGCTGGAGCTGGCGCAGAATTGTGACATGCTGATGCTGTCTTGTGCCCTGACCA
AGGAGACTTTCCACCTGGTTGGGAGGGAGATTATCGATGCGTTGGGGCCCGAGGGCACGCTCATCAACATTGCCAGGGG
GCCTGTGGTGGACGAGCAGGAGCTGGTGAGCGCTATTCTGGAAGGCCGATTGGGAGCAGCTGGGCTGGATGTATATGAG
AACGAGCCTCATGTGCCGCAGGAGCTCTGGGGGTTGGACAATGTGGTGTTGCTGCCGCATGCTGCCAGCCTGACTTGGGA
AACCAGGCGGGGGATGGGCAACTTGGTGGTTGCTAATCTGGATGCCCACTTTGGAGGGAAGCCGCTTGTCACGCCCTTCA
CTGCTGACTACATGTGA

1.7 Full-length *AaHPR1*

ATGCAGGCACTGAGGTTGTTGAGCTTTGGTGTGGCGAGGGACATTAGAGGCCTGGGAGCTGCAGCGGTGTCAGTGCGAG
AGGAGGGGGCCGGGGGAGCTATGGCGGGTTACAGGCGGCAGCTTGCGACGGAGGTGCACAATGCCGGGGGGAAGCTG
AGGGTGGTGAGCACGAAGCCGATGCCGGGGAGCCGGTGGATTCGGGAGCTGGTGAATGCGGGATGCCGGGTGGAGGT
TTGCACGGAGGACAAGACGATTCTGAGCGTGGAGGACGTGCTGCAGCTGATAGGCACGAAGTGTGACGGCGTGATCGG
GCAATTGACGGAGGACTGGGGGGAGACCCTCTTCTCCGCGCTGAGGAGGGCGGGAGGCCACGCCTTCAGCAACATGGCG
GTCGGGTACAACAATGTGGATGTGGAGGCCGCAACCAGGAACGGGATCTCGGTGGGCAACACCCCGGGCGTGCTGACCG
AGACGACGGCGGAGCTGGCGGCAGCGCTCACTCTGGCCGCGGCCCGGCGGGTGGTGGCGGCAGACAGCTTCATGCGCG
CGGGGAAGTACAAGGGATGGCTGCCATCGCTCTTCGTCGGAAACCTGCTCAAGGGCCAGACTGTGGGGATCATCGGGGC
CGGGAGGATCGGGACCGCGTATGCCAGGATGATGGTGGAGGGATTCAAGATGAACGTGGTCTACTTCGACCTCTACCAG
TCCACCCGGCTGGAGAACTTCGTGACAGCCTACGGGAGCTTCCTCAAGAGCCAGGGGGAGGAGGCCGTGACATGGAAGC
GCGCCAGCAATGTGGAGCAGGTGCTGCAGGAGGCGGATGTGGTGAGCCTGCACCCGGTTCTGGACAAGACCACCTTCCA
TCTCATCAACAAGCAGCGGCTGAGCCTGATGAAGAAGGAGGCTGTGTTGGTGAACGCCAGCCGGGGGCCTGTCATAGAC
GAGGCGGCTTTGGTGGAGCACCTGAAGGCTAACCCCATGTTCCGTGTTGGGCTGGACGTGTTTGAGGACGAGCCTCTGAT
GAAGCCGGGGCTGGCGGAGCTGGAGAATGCAGTGGTGGTGCCGCATATTGCTTCTGCTTCGAAGTGGACCAGGGAAGG
AATGGCCACTCTCGCTGCGCAGAACGTGGCCGCCAAGCTCCAGGGGTTTCCGGTTTGGCCCAATTCCAACGATGTGGCTCC
CTTCTTGGACCCATTTTCCCCCGCGCCTGCTGCCTGTCCCAGCATTGTAAATGCCAAGGCACTCAACCTTCCCACCGGCACC
GGAACGTCTGCTAAGCTCTGA

1.8 qPCR fragment AaTAT

GGGACATTGTGGAAACGAAGCTTAAGCCGAACCCGAATTTGGGGAAGAAGCCGATTTCTCTGGCGCAAGGTGATCCCAC
GGTGTACGGACACTTGAAAGTGCCTGAATCCGCTTGCGCTGCTCTGGCGGAGGTGGCTACCAGCTACAAGTACAATGGCT
ACGCCCACTCGGCTGGGATTTTGG

1.9 qPCR fragment AaHPPR

CGCGGGTTTGGATGTTTACCAGGACGAGCCCCGCGTGCCCAAGGCGCTGTGGGGGTTGGACAATGTTGTGCTGCTCCCGC
ATGTGGCCAGCGCGACTGTGGAGACTCGCACAGCCATGGCGGACCTTGTGCTTGAGAACTTGGCCGCGCATTTCGCAGGA
AAGCCCCTCGTCACACCTCTCCATT

1.10 qPCR fragment AaHPPR2

AATTCAGATTCAAGACTCCCTACACCTACTATTCCACGGTGCTGGAGCTGGCGCAGAATTGTGACATGCTGATACTGTCTTG
TGCCCTGACCAAGGAGACTTTCCACCTGGTTGGGAGGGAGATTATCG

1.11 qPCR fragment AaHPR1

GCCCAATTCCAACGATGTGGCTCCCTTCTTGGACCCATTTTCCCCCGCGCCTGCTGCCTGTCCCAGCATTGTAAATGCCAAG
GCACTCAACCTTCCCACCGGCACCGGAACGTCTGCTAAGCTCT

1.12 qPCR fragment Actin

TTTTGAGCAGGAACTGGATACTGCTCGTGCCAGCTCCTCGCTGGAGAAGAGCTTCGAGCTTCCTGATGGACAGGTCATTAC
CATCGGCTCAGAGCGGTTCAGGTGCCCTGAGGTCTTGTTCCAGCCATCGCTCATCGGTATGGAGGCTGCTGGTAT

1.13 qPCR fragment StP 2a

GTGTGTGTCCATCTATCACCAACAGTTGTCCTACTGCATAACACAGTTTGTGGAGAAAGACTTCAAGCTTGCAGACACGGT
TGTAAGGGGTCTCCTGAAGTACTGGCCTGTCACGAATAGTCAGAAGGAGGTCCTTTTCCTTGGGGAGTTGGAGGAGGTCT
T

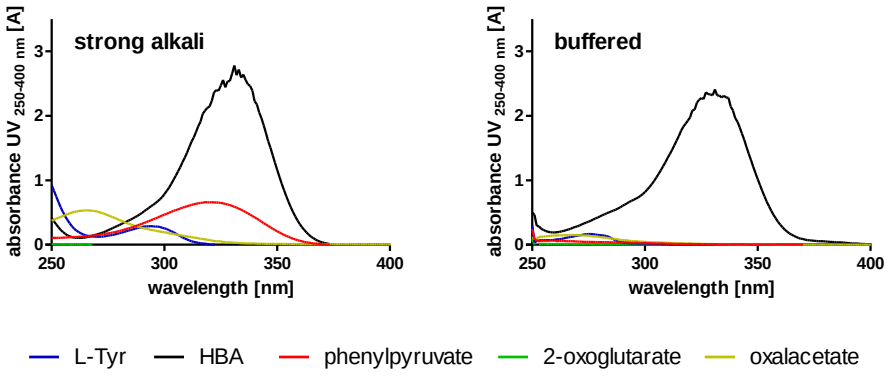


Figure 52 Example of the influence of pH value on the absorption of several components. 4-HBA with maximum at 330 nm in strong alkali and buffered system (≈8.4). Phenylpyruvate with maximum at 320 nm in alkali and no significant absorption in the buffered system. The other components do not show any significant absorption here. This effect was exploited to make phenylpyruvate measurable as amino acceptor.

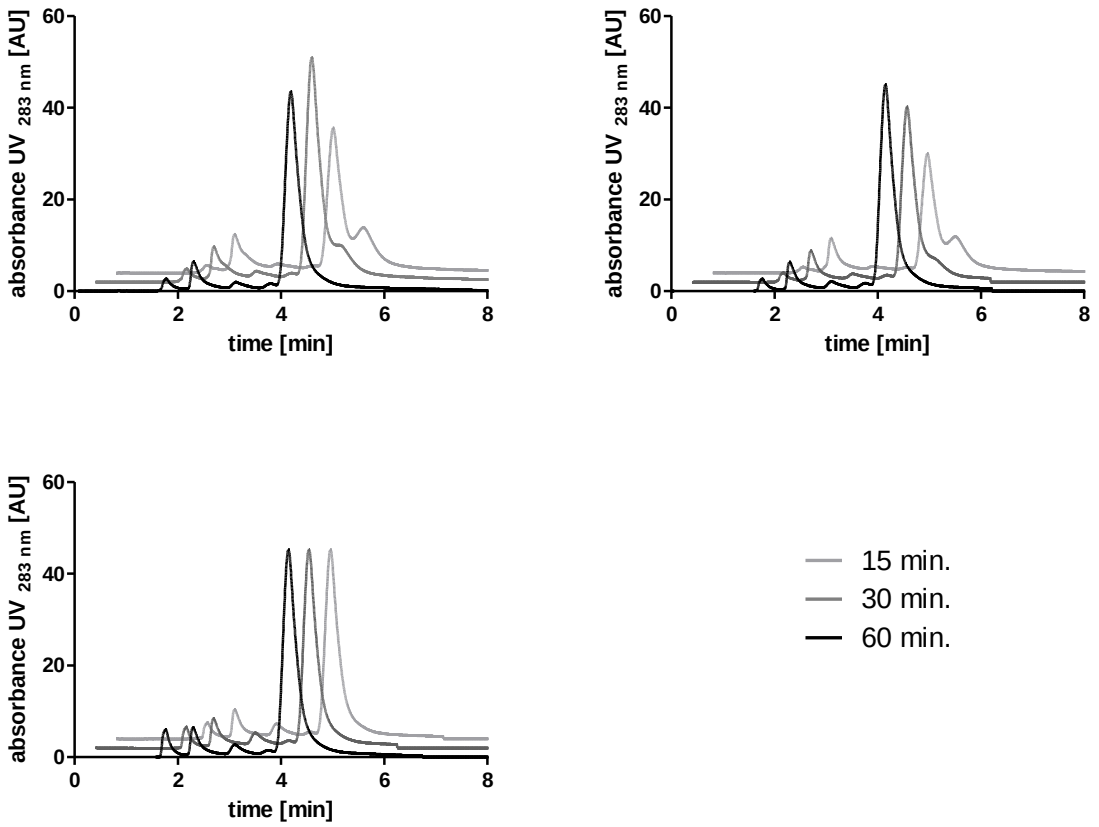


Figure 53 HPLC analyses of HBA formation after 15, 30 and 60 min incubation under several conditions. a pHPP sample treated with KOH and analysed after incubation at room temperature with open lid **b** pHPP sample treated with KOH and analysed after incubation at room temperature and vigorous shaking **c** pHPP sample treated with KOH and analysed after incubation at 45 °C. Each sample was prepared with 1 µl 10 mM pHPP and 50 µl 6 M KOH in a total volume of 300 µl.

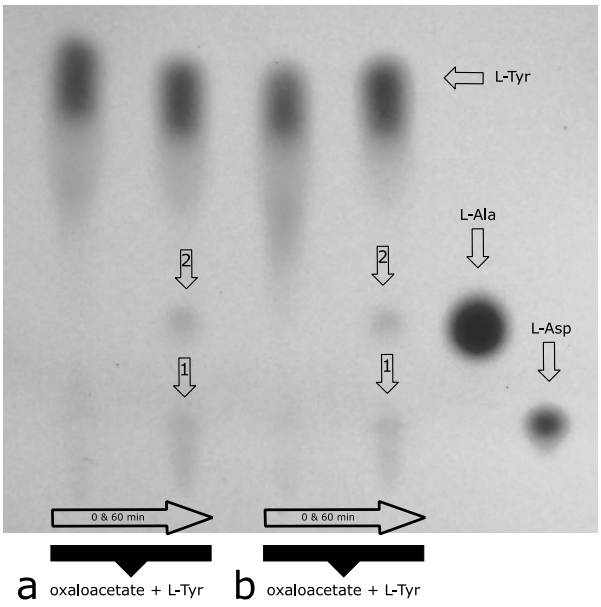


Figure 54 TLC analysis of two potential reaction products of AaTAT with oxaloacetate. Shown are the two resulting ninhydrin stainable products of oxaloacetate as amino donor with L-Tyr (6 mM) as acceptor. **a** Use of an older stock solution of oxaloacetate. **b** Use of a freshly prepared oxaloacetate solution immediately prior to use. 1 represents the expected L-Asp (the L-Asp reference spot here does not show a shift as seen in Fig. 28, probably due to usage of H_2O as solvent) and 2 represents L-Ala.

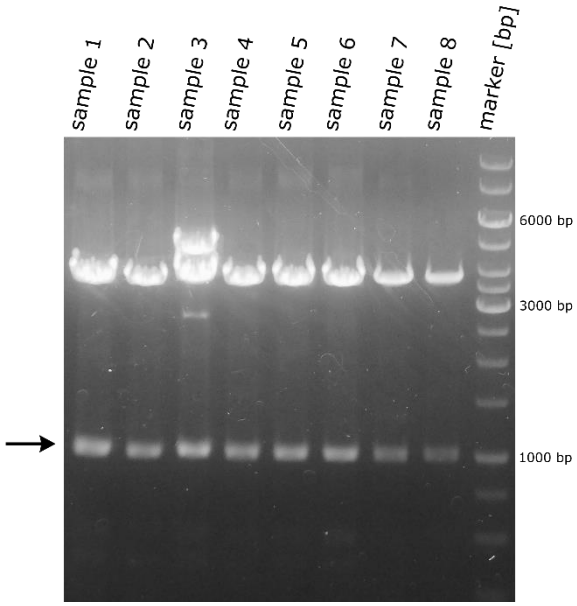


Figure 55 Restriction digest of plasmid DNA harbouring *AaHPPR*. An additional example of the *AaHPPR* full-length (here with non-proofreading enzyme) on gel. The band size appears slightly too large but no intron could be found. Agarose gel electrophoresis with ethidium bromide staining.

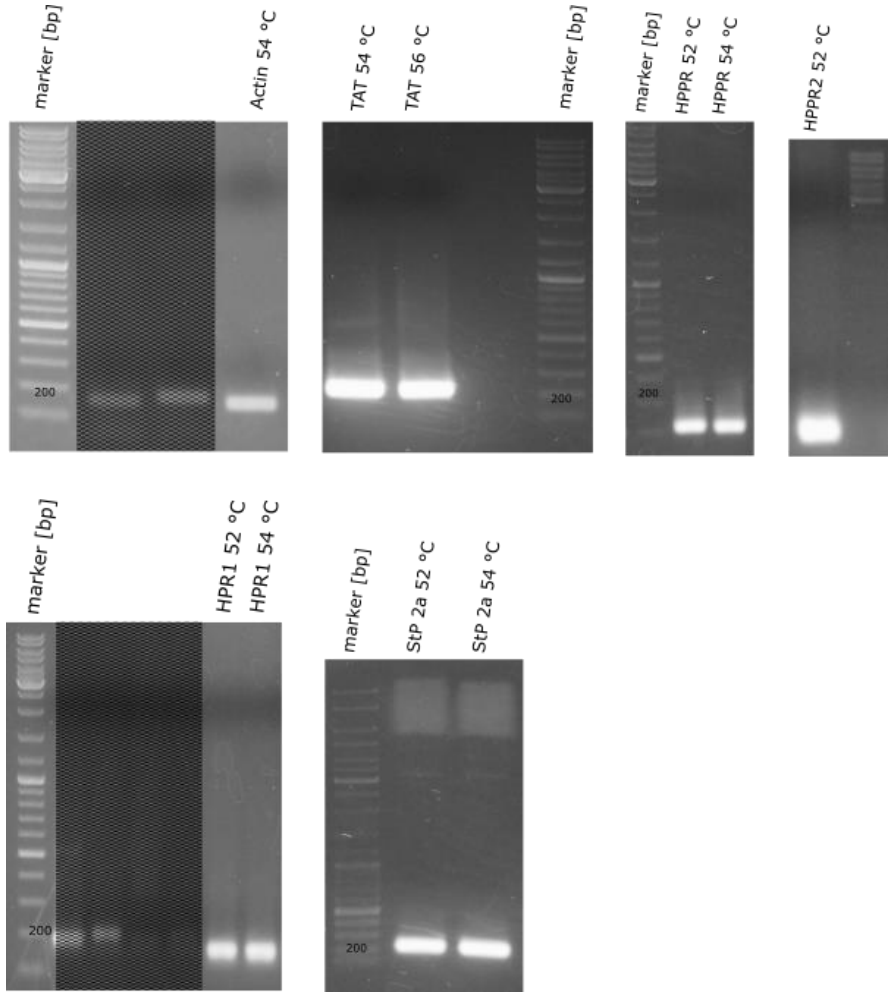


Figure 56 qPCR products of AaTAT, AaHPPR, AaHPPR2, AaHPR1 and the housekeeping genes Actin and StP2a. In some cases, the size marker was not well visible. The gels were additionally stained after completion for identification (not shown here as bands are therefore blurred). The corresponding sequences can be found in Appendix 1.8-1.13.

Table 14 Cq values used for calculation of relative expression

Actin															
Day	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Replicate 1	23,18	20,67	20,48	21,35	21,57	21,97	22,16	22,91	23,07	23,07	23,12	23,49	23,5	24,16	23,95
Replicate 2	24	21,46	21,28	22,09	22,3	22,82	22,94	23,56	23,96	23,83	23,83	24,2	24,23	25	24,59
Replicate 3	23,99	21,41	20,68	22,24	22,17	22,66	22,91	23,91	23,8	23,9	23,84	24,12	24,1	24,76	24,41
Replicate 4	24,17	21,53	21,54	22,11	22,44	22,92	22,93	24,16	24,03	23,82	24,07	24,25	24,31	25	24,65
StP 2a															
Day	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Replicate 1	24,36	23,5	23,38	24,02	23,43	24,08	23,6	24,18	24,53	24,32	24,34	24,62	24,53	25,21	25,03
Replicate 2	25,15	24,15	24,1	24,51	24,15	25,03	24,31	24,65	25,11	25,04	25,08	25,32	25,05	26,13	25,7
Replicate 3	24,35	23,53	23,51	23,97	23,59	24,18	24,01	24,18	24,54	24,4	24,57	24,61	24,42	25,49	25,25
Replicate 4	24,82	23,87	23,86	24,28	24,06	24,76	24,16	24,61	25,1	24,93	24,91	25,24	24,84	26,11	25,55
TAT															
Day	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Replicate 1	26,57	25,31	24,88	24,56	24,37	24,85	25,35	26,07	26,43	26,42	26,48	26,8	-	28,32	28,4

Replicate 2	27,5	26,12	25,37	25,33	25,05	25,47	26,01	27	27,1	27,14	27,1	27,54	28,24	29,37	29,47
Replicate 3	26,68	25,46	24,95	24,87	24,57	25,17	25,35	26,97	26,81	27,01	26,85	27,22	27,62	28,99	28,28
Replicate 4	27,22	25,95	25,47	25,12	25,01	25,55	26,17	27,35	27,29	27,3	27,24	27,56	28,19	28,59	29,56
HPPR															
Day	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Replicate 1	25,6	24,41	24,09	24,73	24,66	24,97	25,01	25,22	25,6	25,86	25,73	25,75	26	26,91	27,44
Replicate 2	26,47	25,21	24,97	25,88	25,42	25,7	25,92	26,17	26,2	26,56	26,68	26,52	26,81	27,8	28,17
Replicate 3	25,31	24,07	23,84	24,73	24,19	24,72	24,74	25,82	25,26	25,25	25,17	25,39	25,86	26,3	26,84
Replicate 4	25,53	24,34	24,06	24,65	24,97	25,06	25	25,79	25,41	25,58	25,79	25,62	26,03	26,94	27,16
HPPR2															
Day	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Replicate 1	25,48	23,71	23,06	23,33	23,3	23,57	23,76	24,92	25,44	25,02	25,43	26,08	26,18	27,79	27,98
Replicate 2	26,44	24,41	24,04	23,82	24,01	24,33	24,37	25,66	26,02	25,78	26,21	26,62	26,99	28,87	28,63
Replicate 3	25,76	24,04	23,19	23,32	23,38	23,62	23,79	25,19	25,42	25,24	25,45	26,13	27,21	27,95	28,18
Replicate 4	26,29	24,26	23,56	23,67	23,75	24,36	24,88	25,76	25,75	25,83	26,03	26,82	26,67	28,44	29,55
HPR1															
Day	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Replicate 1	25,79	25,62	25,59	26,25	24,55	25,26	25,06	25,49	25,72	25,9	25,96	26,09	26	26,83	27,54
Replicate 2	26,91	26,34	26,27	26,94	25,28	25,99	25,76	26,4	26,64	26,45	26,65	26,66	26,59	27,48	28,47
Replicate 3	26,35	26,04	25,89	26,62	25,04	25,5	25,22	26,16	26,13	26,13	26,22	26,05	26,39	27,29	28,2
Replicate 4	26,71	26,64	26,27	26,97	25,36	26,13	25,9	26,35	26,58	26,42	26,92	26,61	26,87	27,8	28,44

1.14 Units and abbreviations

Specific units

% (w/v)	mass concentration (1% = 1 g/100 ml)
% (v/v)	volume concentration
Da	Dalton
kat	katal (mol/s)
k_{cat}	turnover number (s^{-1})
k_{cat}/K_m	kinetic efficiency ($M^{-1}*s^{-1}$)
K_m	Michaelis-Menten constant
OD_{600}	optical density at 600 nm
rpm	revolutions per minute
T_m	melting temperature
U	enzyme unit
v_{max}	maximum reaction rate

Appendix

Amino acids

Amino acid	3 letter code	1 letter code	Amino acid	3 letter code	1 letter code
alanine	Ala	A	leucine	Leu	L
arginine	Arg	R	lysine	Lys	K
asparagine	Asn	N	methionine	Met	M
aspartic acid	Asp	D	phenylalanine	Phe	F
cysteine	Cys	C	proline	Pro	P
glutamine	Gln	Q	serine	Ser	S
glutamic acid	Glu	E	threonine	Thr	T
glycine	Gly	G	tryptophan	Trp	W
histidine	His	H	tyrosine	Tyr	Y
isoleucine	Ile	I	valine	Val	V

Nucleotides

Nucleotide	Symbol	Nucleotide	Symbol	Nucleotide	Symbol
Purines					
adenine	A	guanine	G		
Pyrimidines					
cytosine	C	thymine	T	uracil	U

Other abbreviations

Aa	*Anthoceros agrestis*
AaHPPR	hydroxyphenylpyruvate reductase isoform 1 from *Anthoceros agrestis*
AaHPPR2	hydroxyphenylpyruvate reductase isoform 2 from *Anthoceros agrestis*
AaHPR1	hydroxypyruvate reductase from *Anthoceros agrestis*
AaTAT	tyrosine aminotransferase from *Anthoceros agrestis*
aa	amino acid
APS	ammonium persulphate
BAHD	family of acyltransferases, see e.g. D'Auria (2006)

bp	base pairs
BSA	bovine serum albumin
4H3MPL	4-hydroxy-3-methoxyphenyllactate *
4H3MPP	4-hydroxy-3-methoxyphenylpyruvate*
CAPS	*N*-cyclohexyl-3-aminopropanesulfonic acid
CbHPPR	hydroxyphenylpyruvate reductase from *Coleus blumei*, with respect to Janiak (2007) also indicated as CbH(P)PR
cDNA	complementary DNA
CDS	coding sequence
CHES	*N*-cyclohexyl-2-aminomethanesulfonic acid
Cq	cycle quantification
CYP	cytochrome P450 monooxygenase
DHPL	3,4-dihydroxyphenyllactic acid*
DHPP	3,4-dihydroxyphenylpyruvate*
dNTP	nucleoside triphosphates containing deoxyribose
DTT	dithiothreitol
E	amplification efficiency, optionally given as percentage (E%)
E. coli	*Escherichia coli*
EcoRI	restriction endonuclease from *E. coli*
EDTA	ethylenediaminetetraacetic acid
gDNA	genomic DNA
HCl	hydrochloric acid
β-HP	hydroxypyruvate*
HPLC	high pressure liquid chromatography
HPPR	hydroxyphenylpyruvate reductase
HPR	hydroxypyruvate reductase
IPTG	isopropyl β-D-1-thiogalactopyranoside
KOH	potassium hydroxide
LB	lysogeny broth
NAD/NADH	nicotinamide adenine dinucleotide (oxidized/reduced)
NADP/NADPH	Nicotinamide adenine dinucleotide phosphate (oxidized/reduced)

NaOH	sodium hydroxide
NTC	no template control
OPA	*o*-phthalaldehyde
ORF	open reading frame
P	pyruvate
PA	polyamide
PAGE	polyacrylamide gel electrophoresis
PAL	phenylalanine ammonia-lyase
PCR	polymerase chain reaction
pHPL	4-hydroxyphenyllactate*
pHPP	4-hydroxyphenylpyruvate*
PL	phenyllactate*
PLP	pyridoxal 5'-phosphate
PP	phenylpyruvate *
PVDF	polyvinylidene fluoride
RACE	rapid amplification of cDNA ends
RAS	rosmarinic acid synthase
RNA	ribonucleic acid
rRNA	ribosomal ribonucleic acid
RuBisCO	ribulose-1,5-bisphosphate carboxylase-oxygenase
SD	standard deviation
SDS	sodium dodecyl sulphate
SOB	super optimal broth
SOC	super optimal broth with catabolite repression
St-P 2a	serine threonine protein phosphatase 2a regulatory subunit
TAT	tyrosine aminotransferase
TB	terrific broth
TEMED	tetramethylethylenediamine
TLC	thin-layer chromatography
Tris	tris(hydroxymethyl)aminomethane
qPCR	quantitative real-time PCR

UPM	universal primer mix
UV	ultraviolet
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

***Organic acids**, e.g. 4-hydroxyphenylpyruvic acid and others, are sometimes referred to in the form of the conjugated base (4-hydroxyphenylpyruvate). The terms are used synonymously. This is merely for better readability.