

Biosynthesis of the necic acid moiety of
lycopsamine type pyrrolizidine alkaloids

-

investigations into an
acetoxyhydroxyacid synthase

Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Annika Engelhardt

Kiel, Dezember 2022

Erstgutachter: Prof. Dr. Dietrich Ober

Zweitgutachter: Prof. Dr. Wolfgang Bilger

Tag der mündlichen Prüfung: 13.02.2023

Zum Druck genehmigt: 13.02.2023

Annika Engelhardt

**Biosynthesis of the necic acid moiety of lycopsamine type pyrrolizidine alkaloids -
investigations into an acetohydroxyacid synthase**

Dissertation, Faculty of Mathematics and Natural Sciences
Christian-Albrechts-Universität zu Kiel



~ nature just needs the right change to experience a new function ~

Zusammenfassung

Pyrrolizidinalkaloide (PAs) sind typische pflanzliche Sekundärmetabolite, die an der pflanzlichen Abwehr von Herbivoren beteiligt sind. Mehr als 600 verschiedene Strukturen sind aus Pflanzenarten verschiedenster Pflanzenfamilien beschrieben worden. Das Grundgerüst der PAs besteht aus einer charakteristischen bizyklischen Necinbase, die allen PAs gemeinsam ist und die mit einer oder mehreren Necinsäuren verestert ist. Der erste Schritt der Biosynthese der Necinbase wird durch die Homospermidin Synthase (HSS) katalysiert und wurde in den letzten Jahrzehnten intensiv untersucht. Erst kürzlich wurde durch die Identifizierung der Homospermidin-Oxidase der zweite Schritt der Synthese der Necinbase aufgeklärt. Die umfassende evolutionäre Charakterisierung der HSS ergab, dass sich die PA Biosynthese mehrmals unabhängig voneinander in verschiedenen Pflanzenstämmen entwickelt hat, unter anderem bei den Boraginales, den Apocynaceae und den Eupatorieae (Asteraceae). Bemerkenswert ist, dass die unabhängigen Ursprünge dieser drei Linien zur Bildung von PAs mit identischen Strukturen und einer charakteristischen verzweigt-kettigen C₇-Necinsäure, den sogenannten Lycopsamintyp PAs.

In dieser Arbeit wurde der erste Schritt der Bildung der verzweigt-kettigen C₇-Necinsäuren der Lycopsamintyp PAs aufgeklärt. In *Symphytum officinale* (*S. officinale*, Boraginaceae), *Heliotropium indicum* (*H. indicum*, Heliotropiaceae, Boraginales) und *Eupatorium cannabinum* (*E. cannabinum*, Eupatorieae, Asteraceae) wurde je ein Duplikat einer katalytischen Untereinheit (CSU) der Acetohydroxysäure Synthase (AHAS) identifiziert, die C₇-Hydroxysäuresynthase (engl. C₇-hydroxyacid synthase, C₇HAS) genannt wird und das erste spezifische Enzym für die Biosynthese der Necinsäure ist. Die Charakterisierung der C₇HAS aus *S. officinale* und *H. indicum* zeigte, dass das Enzym die Übertragung eines aus Pyruvat gebildeten aktivierten Acetaldehyds auf 2-Oxoisovalerat katalysiert, wodurch der direkte Vorläufer der C₇-Necinsäuren, die C₇-Pronecinsäure, gebildet wird. Es wurde eine GC-MS-basierte Assaymethode entwickelt um alle Komponenten der AHAS- und C₇HAS-katalysierten Reaktionen, inklusive der C₇-Pronecinsäure, simultan nachweisen zu können.

Die CRISPR/Cas9-vermittelte Genome Editierungstechnik wurde angewandt, um *in planta* die Beteiligung einer HSS an der PA-Biosynthese von *E. cannabinum* und die Beteiligung einer C₇HAS an der PA-Biosynthese von *Symphytum* und *Eupatorium* nachzuweisen. Der Knock-out des Gens, das für HSS in *E. cannabinum* kodiert, führte zu PA-freien hairy roots ohne nachweisbare PA-spezifische Zwischenprodukte. Knock-outs des Gens, das für die C₇HAS kodiert, führten bei *Symphytum* und *Eupatorium* ebenfalls zu PA-freien hairy roots.

Bemerkenswert ist, dass in diesen hairy roots die gesättigte Necinbase Trachelanthamidin akkumuliert und so als Veresterungspartner für die Necinsäureeinheit in beiden Pflanzenarten identifiziert werden konnte. Als regenerierte ganze Pflanzen aus *S. officinale c7has* Knock-out hairy roots Blütenstände zu entwickeln begannen, konnte gezeigt werden, dass die PA-Biosynthese nach der Bildung von Trachelanthamidin auch hier unterbrochen wird und keine PAs vom Lycopsamintyp gebildet werden. Dies lässt den Schluss zu, dass in *Symphytum* nur eine C₇HAS die Bildung der C₇-Pronecinsäure an beiden Biosyntheseorten katalysiert.

Durch den Vergleich der Metabolome von *c7has*-Knockout und Kontrolllinien hairy roots von *S. officinale* konnten wir mögliche Zwischenprodukte der Kernbiosynthese bis hin zu *Symphytums* einfachstem Lycopsamintyp PA identifizieren. Trachelanthamin, nicht unterscheidbar Trachelanthamin *N*-Oxid oder 7-Hydroxytrachelanthamin und 7-Hydroxytrachelanthamin *N*-Oxid sind ausschließlich in Kontrolllinien nachweisbar. Dies deutet auf die folgende Abfolge von enzymatischen Schritten bei der PA-Biosynthese in *S. officinale* hin: Reduktion der C₇-Pronecinsäure, *O*⁹-Veresterung von Trachelanthamidin mit der C₇-Necinsäure, C⁷-Hydroxylierung und *N*-Oxygenierung in ungeklärter Reihenfolge und schließlich Bildung der protoxischen 1,2-ungesättigten PAs durch Desaturierung in der 1,2-Position der Necinbase.

Interessanterweise deuten phylogenetische Analysen der drei in dieser Arbeit untersuchten C₇HAS-kodierenden Gene auf einen unabhängigen Ursprung der C₇-Necinsäurebiosynthese in den Heliotropiaceae, Boraginaceae und Eupatorieae hin, was einen weiteren einzigartigen Aspekt der konvergenten Evolution der PAs darstellt.

Zusammenfassend lässt sich sagen, dass die in dieser Arbeit vorgestellten Daten und die etablierten Methoden ein tieferes Verständnis der Biosynthese von PAs des Lycopsamintyps, ermöglichen. Sie werden eine hilfreiche und wichtige Quelle für die laufende Forschung zur PA-Biosynthese, insbesondere zu der Synthese der Necinsäuren, und den daran beteiligten Enzymen sein.

Abstract

Pyrrolizidine alkaloids (PAs) are typical specialized plant metabolites involved in plant defense. More than 600 different structures have been described from plant species of various lineages. The backbone of PAs consists of a necine base, the characteristic bicyclic structure common to all PAs that is esterified with one or more necic acids. The first step of the biosynthesis of the necine base moiety is catalyzed by homospermidine synthase (HSS) and was intensively studied over the past decades. Most recently, the identification of homospermidine oxidase elucidated the second step in the necine base formation. The comprehensive evolutionary characterization of HSS revealed that the PA biosynthesis evolved several times independently in different plant lineages, among others, in the Boraginales, the Apocynaceae, and in the Eupatorieae (Asteraceae). Remarkably, the independent origins in these three lineages resulted in the formation of PAs with identical structures and a characteristic branched-chain C₇-necic acid, the lycopsamine type PAs.

In this thesis, the first step of the formation of the branched-chain C₇-necic acids of lycopsamine type PAs was elucidated. In *Symphytum officinale* (*S. officinale*, Boraginaceae), *Heliotropium indicum* (*H. indicum*, Heliotropiaceae, Boraginales), and *Eupatorium cannabinum* (*E. cannabinum*, Eupatorieae, Asteraceae) a duplicate of an acetohydroxyacid synthase (AHAS) catalytic subunit (CSU), now named C₇-hydroxyacid synthase (C₇HAS), was identified to be the first specific enzyme in the biosynthesis of the necic acid moiety. Characterization of the C₇HAS from *S. officinale* and *H. indicum* showed that the enzyme catalyzes the transfer of an activated acetaldehyde formed from pyruvate to 2-oxoisovalerate, forming the direct precursor of the C₇-necic acids, C₇-pronecic acid. A GC-MS based assay method was developed to simultaneously detect all components of AHAS- and C₇HAS-catalyzed reactions, including the formation of the C₇-pronecic acid.

The CRISPR/Cas9 mediated genome editing technique was applied to provide *in planta* evidence for the involvement of HSS in the PA biosynthesis of *E. cannabinum* and for the involvement of C₇HAS in the PA biosynthesis of *Symphytum* and *Eupatorium*. Knock-outs of the gene encoding HSS in *E. cannabinum* resulted in PA free hairy roots without any detectable PA-specific intermediates. Knock-outs of the gene encoding C₇HAS also resulted in PA free hairy roots for *Symphytum* and *Eupatorium*. Remarkably, the saturated necine base trachelanthamidine accumulated in these hairy roots and was thus identified as the esterification partner for the necic acid unit in both plant species. When regenerated whole plants from *S. officinale* *c7has* knock-out hairy roots started to develop inflorescences, we were able to show that again, PA biosynthesis is interrupted after trachelanthamidine and no lycopsamine type

PAs are produced. This allows us to conclude that in *Symphytum* only one C₇HAS catalyzes the formation of the C₇-pronecic acid at both biosynthetic sites.

By comparing the metabolomes of *c₇has* knock-out hairy root lines and control lines of *S. officinale*, we identified possible intermediates of core biosynthesis up to intermedine *N*-oxide, *Symphytum*'s simplest lycopsamine type PA. Trachelanthamine, the indistinguishable trachelanthamine *N*-oxide or 7-hydroxytrachelanthamine, and 7-hydroxytrachelanthamine *N*-oxide have been identified and are detectable exclusively in control lines. This suggests the following sequence of enzymatic steps in PA biosynthesis in *S. officinale*: Reduction of C₇-pronecic acid, *O*⁹ esterification of trachelanthamidine with the C₇-necic acid, C⁷ hydroxylation and *N*-oxygenation in unexplained order, and finally formation of the pre-toxic 1,2-unsaturated PAs by desaturation in 1,2-position of the necine base.

Interestingly, phylogenetic analyses of the three C₇HAS encoding genes studied in this thesis indicate an independent origin of C₇-necic acid biosynthesis in the Heliotropiaceae, Boraginaceae, and Eupatorieae, adding another unique aspect to the convergent evolution of PAs.

In conclusion, the data presented and the methods established in this work provide a deeper understanding of the biosynthesis of lycopsamine type PAs. They will be a useful and valuable resource for ongoing research on PA biosynthesis, especially on the synthesis of necic acids, and the enzymes involved in it.

List of Abbreviations

2OB	2-oxobutyrate
2OIV	2-oxoisovalerate
AHAS	Acetohydroxyacid synthase
AHOB	2-acetohydroxy-2-oxobutyrate
AHOIV	2-acetohydroxy-2-oxoisovalerate
AL	Acetolactate
<i>A. rhizogenes</i>	<i>Agrobacterium rhizogenes</i>
BCAA	Branched-chain amino acid
C ₇ HAS	C ₇ -hydroxyacid synthase
C ₇ PNA	C ₇ -pronecic acid
CSU	Catalytic subunit
CRISPR	Clustered regulatory interspaced short palindromic repeats
Cas9	CRISPR associated protein 9
CL	Empty vector control line
ECF	Ethyl chloroformate
<i>E. cannabinum</i>	<i>Eupatorium cannabinum</i>
EIC	Extracted ion chromatogram
ESI-	Negative ion mode electrospray ionization
ESI+	Positive ion mode electrospray ionization
<i>H. indicum</i>	<i>Heliotropium indicum</i>
HCD	Higher energy collision induced dissociation
HSO	Homospermidine oxidase
HSS	Homospermidine synthase
HeThDP	Hydroxyethyl thiamine diphosphate
MCF	Methyl chloroformate
PA	Pyrrolizidine Alkaloid
PYR	Pyruvate
RSU	Regulatory subunit
sgRNA	Short guide RNA
<i>S. officinale</i>	<i>Symphytum officinale</i>
ThDP	Thiamine diphosphate
TMS	Trimethylsilyl ester
TP	Transit peptide

Content

1 Introduction	1
2 A rapid and efficient assay for the characterization of substrates and products of acetohydroxyacid synthases	8
2.1 Abstract	9
2.2 Introduction	10
2.3 Results	13
2.3.1 Enzyme preparation	13
2.3.2 Requirements for an activity assay to characterize AHAS-catalyzed reactions	14
2.3.3 Evaluation of methods for monitoring AHAS-catalyzed reactions	15
2.3.4 Characterization of <i>Ec</i> AHASII reaction assays using MCF derivatization.....	16
2.3.5 Ethylation instead of methylation improves the detectability of keto acids and acetohydroxyacids	20
2.3.6 Characterization of <i>So</i> CSU	23
2.4 Discussion	27
2.4.1 State of the art of AHAS assay methods.....	27
2.4.2 Improvements in AHAS assay detection by ECF	28
2.4.3 Possible applications.....	30
2.4.4 Conclusion	31
2.5 Experimental	32
2.5.1 Material	32
2.5.2 Heterologous expression of <i>E. coli</i> AHAS isoenzyme II	32
2.5.3 Heterologous expression of <i>So</i> CSU	33
2.5.4 AHAS activity assay based on indirect colorimetric detection	33
2.5.5 AHAS activity assay based on alkylation of substrates and products	34
2.5.6 ECF Ethylation of acids	35
2.6 Supplementary.....	36

3 The role of an acetohydroxyacid synthase in the biosynthesis of lycopsamine type pyrrolizidine alkaloids in <i>Symphytum officinale</i> (Boraginaceae).....	42
3.1 Abstract	43
3.2 Introduction	44
3.3 Results	48
3.3.1 Tissue-specific transcript quantification suggests the gene encoding <i>SoC₇HAS</i> to be co-expressed with <i>SoHSS</i>	48
3.3.2 CRISPR/Cas9 knock-outs of <i>soc₇has</i> result in the inability of hairy roots to produce PAs	49
3.3.3 No transport of trachelanthamidine intermediate within regenerated plants	55
3.3.4 Second side – same enzyme.....	55
3.3.5 Heterologous expression and activity assay of <i>SoC₇HAS</i>	57
3.3.6 <i>SoC₇HAS</i> catalyzes the formation of C ₇ pronecic acid.....	58
3.3.7 Kinetic Parameters of <i>SoC₇HAS</i>	59
3.3.8 <i>SoC₇HAS</i> is promiscuous regarding its acceptor substrate.....	62
3.4 Discussion	64
3.4.1 <i>In planta</i> studies of <i>SoC₇HAS</i>	64
3.4.2 Promiscuity of <i>SoC₇HAS</i>	66
3.4.3 Does the <i>SoC₇HAS</i> form a complex with an RSU?.....	67
3.4.4 Conclusion	68
3.5 Experimental	70
3.5.1 Material	70
3.5.2 Plant Material.....	70
3.5.3 RNA extraction and quantitative reverse transcription PCR (RT-qPCR)	70
3.5.4 CRISPR/Cas9 vector construction and generation of transgenic hairy root mutants ...	71
3.5.5 Plant regeneration from hairy roots	71
3.5.6 Genetic identification of transgenic hairy root lines.....	72
3.5.7 Pyrrolizidine alkaloid identification and quantification by GC-MS.....	72
3.5.8 Heterologous expression in <i>E. coli</i>	74

3.5.9 Enzyme activity assays	74
3.5.10 Derivatization with ECF	74
3.6 Supplementary.....	75
4 Core biosynthesis of lycopsamine type pyrrolizidine alkaloids in <i>Symphytum officinale</i> – a metabolomic approach	79
4.1 Abstract	80
4.2 Introduction	81
4.3 Results	84
4.3.1 Hypothetical biosynthetic pathway of PA core biosynthesis.....	84
4.3.2 Metabolic profiles of knock-out lines and control lines show distinct differences	86
4.3.3 Production of PAs represents a contributing factor to differences in the metabolite profiles of CL and KO.....	87
4.3.4 Accumulation of trachelanthamidine in knock-out lines of <i>SoC₇HAS</i>	92
4.3.5 C ₇ -necic acid is esterified with trachelanthamidine.....	93
4.3.6 Identification of potential intermediates	94
4.4 Discussion	97
4.4.1 Metabolomic approach.....	97
4.4.2 Is trachelanthamine first <i>N</i> -oxygenated or C ⁷ hydroxylated?	98
4.4.3 Predicted metabolic pathway of PA biosynthesis in <i>S. officinale</i>	98
4.4.4 Conclusion	100
4.5 Material and methods	101
4.5.1 Plant material	101
4.5.2 Metabolite extraction	101
4.5.3 UHPLC-ESI-Q Exactive MS/MS Analysis of Extracted Metabolites.....	101
4.5.4 Manual identification of metabolites	102
4.5.5 Data conversion	102
4.5.6 Data processing with XCMS™ Online.....	102
4.5.7 Statistical analysis with Metaboanalyst 5.0	103
4.5.8 Data processing with MZmine3	103

4.5.9 Determination of the molecular formulae using SIRIUS 4	104
4.5.10 Structure assignment using CSI:FingerID	104
4.6 Supplementary.....	105
5 Characterization of the necic acid biosynthesis of lycopsamine type pyrrolizidine alkaloids – miscellaneous efforts.....	106
5.1 Abstract	107
5.2 Introduction	108
5.3 Results	110
5.3.1 CRISPR/Cas9 knock-outs of <i>EcanC7HAS</i> and <i>EcanHSS</i>	110
5.3.2 Formation of C ₇ -pronecic acid is exclusively catalyzed by C ₇ HAS.....	114
5.3.3 C ₇ HAS recruitment from AHAS occurred several times independently.....	116
5.4 Discussion	118
5.4.1 C ₇ HAS is involved in the PA biosynthesis of <i>E. cannabinum</i>	118
5.4.2 C ₇ HAS still possesses AHAS CSU activity.....	118
5.4.3 The gene <i>ahas csu</i> was duplicated several times	119
5.4.4 Conclusion	119
5.5 Material and Methods.....	120
5.5.1 Material	120
5.5.2 Plant Material.....	120
5.5.3 CRISPR/Cas9 vector construction and generation of transgenic hairy root mutants .	120
5.5.4 Genetic identification of transgenic hairy root lines.....	121
5.5.5 Pyrrolizidine alkaloid analysis by GC-MS	121
5.5.6 Heterologous expression.....	121
5.5.7 Enzyme activity assays	121
5.5.8 Derivatization with MCF	121
5.5.9 Phylogenetic analysis.....	121
5.6 Supplementary.....	125
6 Conclusion and general perspectives	127
7 Outlook.....	132

8 Contributions.....	136
9 Acknowledgements	137
10 References	139
Erklärung	154

1 Introduction

Previously considered as waste products of plant metabolism, plant secondary metabolites now offer botanists a diverse and interesting field of research. One fascinating characteristic is the immense structural diversity of metabolites that does not come by chance but results from a complex network of biosynthetic pathways. To learn about the origin of this diversity, it is necessary to understand the evolution of the secondary metabolite pathway. The great similarities between many reactions and enzyme properties of plant primary and secondary metabolism can clearly be seen by an inspection of the two (Modolo et al., 2009). In our research group, the evolution of secondary metabolism is studied in more detail using pyrrolizidine alkaloids (PAs) as an example.

Plants produce PAs as a chemical defense

PAs are considered defense compounds against herbivores (Boppré, 1986; Schneider, 1987; Boppré, 1990; Hartmann and Witte, 1995). PAs are synthesized and stored as nontoxic *N*-oxide derivatives. Reduction to the tertiary pre-toxic form occurs spontaneously in the digestive system of the feeding herbivores (Figure 1 E), making the PAs available for bioactivation by cytochrome P450 monooxygenases, which are part of the xenobiotic metabolism in the liver. If the PAs possess a 1,2-double bond, an esterified hydroxy group at position C⁹, and a free or esterified C⁷ hydroxy group, the bioactivation results in toxic pyrrolic compounds (Winter and Segall, 1989; Prakash et al., 1999). These PA pyrroles are highly reactive alkylating agents and are able to cross link DNA and proteins and thus exhibit hepatotoxic, carcinogenic, tumorigenic, and mutagenic properties. PAs missing those structural features are speculated to deter herbivores by bitter taste or might serve different unknown ecological functions. (Mori et al., 1985; Zijlstra and Vogel, 1988a; Zijlstra and Vogel, 1988b; Frei et al., 1992; Fu et al., 2002; Fu et al., 2004).

Necine base moiety is a common trait of all PAs

A common trait all PAs share is the bicyclic nitrogen-containing necine base that is derived from 1-hydroxymethylpyrrolizidine (Figure 1). In its biosynthesis homospermidine synthase (HSS, EC 2.5.1.45), the first PA pathway-specific enzyme, catalyzes the formation of homospermidine by transferring an aminobutyl moiety from spermidine to putrescine in a NAD⁺-dependent reaction (Ober and Hartmann, 1999). HSS evolved by duplication of the gene encoding the enzyme deoxyhypusine synthase of primary metabolism. HSS was recruited several times independently in different angiosperm lineages. Independent gene duplication

events have been described to have occurred in the monocots, the Boraginales, the Convolvulaceae, the Apocynaceae, the Fabaceae, and with the Eupatorieae and the Senecioneae even twice within the Asteraceae (Anke et al., 2004; Reimann et al., 2004; Nurhayati et al., 2009; Kaltenegger et al., 2013; Irmer et al., 2015; Livshultz et al., 2018). As often described for enzymes involved in the biosynthesis of secondary metabolites, the site of biosynthesis of PAs is characterized by a high tissue and cell specificity in the individual lineages, but is quite diverse when different lineages are compared. Immunolabeling studies of HSS have described unique sites of synthesis at the organ, tissue, or cell level for almost every species or family. HSS was localized in tissues ranging from the apical meristem of aerial root tips of *Phalaenopsis* orchids, the lower leaf epidermis of *Heliotropium indicum* (Heliotropiaceae, Boraginales) to the endodermis or cortex parenchyma of primary roots in *Symphytum officinale* (*S. officinale*, Boraginaceae, Boraginales) and *Eupatorium cannabinum* (Eupatorieae, Asteraceae), respectively. In *Senecio vernalis* (Senecioneae, Asteraceae) special cells in the endodermis and cortex parenchyma opposite of the phloem have been identified to express HSS (Moll et al., 2002; Anke et al., 2008; Niemüller et al., 2012). To boost PA levels during flower development, additional production sites in the flower buds of *Phalaenopsis* and in the bundle sheath cells of young leaves subtending the developing inflorescences of *S. officinale* are activated following a developmental program and contribute substantially to the variability of PA biosynthetic sites in plants (Anke et al., 2008; Kruse et al., 2017).

Most recently, the identification of homospermidine oxidase (HSO, EC 1.4.3.22) elucidated the second step in the necine base formation (Zakaria et al., 2022). As HSS, HSO also resulted from a duplication of a gene coding an enzyme of primary metabolism, *i.e.*, a copper-containing amine oxidase (EC 1.4.3.6). Characterization of the HSO of *H. indium* revealed that only one enzyme is sufficient to catalyze the oxidative deamination of the two primary amino groups of homospermidine and to control the formation of the bicyclic pyrrolizidine backbone 1-formylpyrrolizidine (Zakaria et al., 2022). Just an additional reduction has to be catalyzed to result in the formation of the necine base, the 1-hydroxymethylpyrrolizidine.

Structural variety of PAs is attributed to the necic acid moieties

The necine bases only differ in their 1,2-desaturation, the hydroxylation mostly at C⁷ and more rarely at C² or C⁶, and in their stereochemistry as the necine base possesses one or two stereocenters in the unsaturated and saturated necine bases, respectively. However, PAs exhibit an astonishing diversity with more than 600 known structures. This diversity results from esterification of the necine bases with a wide variety of necic acids at O⁹ and/or O⁷ and the

diversification of the resulting structures. The necic acids have been used to chemosystematically classify the PAs into different types as shown in Figure 1. The senecionine type (A) with their twelve-membered macrocyclic diesters are derivatives of senecionine. The triangularine type (B) is characterized by C5 necic acids, more precisely angeloyl-, tigloyl- or seneciroyl mono- or diesters. The monocrotaline type (C) consists of eleven-membered macrocyclic diesters and variable substitution patterns of the acid moiety. In the phalaenopsine type (D), monoesters of saturated necine bases with aryl or aralkyl and less commonly with alkyl necic acids are compiled. Lycopsamine type PAs (E) occur as mono- or diesters and always contain a hydroxylated 2-isopropylbutyric acid residue, a branched C₇-acid that is uniquely found in PAs. PAs with unusual necic acid esters, like thesinine and ipanguline, and necine base derivatives that do not fit into any of these categories are grouped under miscellaneous PAs (Hartmann and Witte, 1995). It is noteworthy that the type of PAs is often lineage specific. For example, within the Asteraceae, which comprise the two PA-producing tribes, the Senecioneae and the Eupatorieae, the PAs within the Senecioneae belong to the senecionine type, whereas the PAs of the Eupatorieae belong to the lycopsamine type (Hartmann and Witte, 1995). The lycopsamine type PAs are not only interesting because of the unique structure of the necic acid moiety, but also from an evolutionary point of view. They occur in species of not closely related plant lineages, *i.e.*, of the Apocynaceae, the Boraginales, and the Eupatorieae. For all these three lineages an independent evolutionary origin is postulated based on phylogenetic data for HSS (Reimann et al., 2004; Livshultz et al., 2018). Nevertheless, lycopsamine type PAs produced by plants of these tribes are identical with respect to the stereochemistry of the necic acid moieties (Hartmann and Witte, 1995; El-Shazly and Wink, 2014).

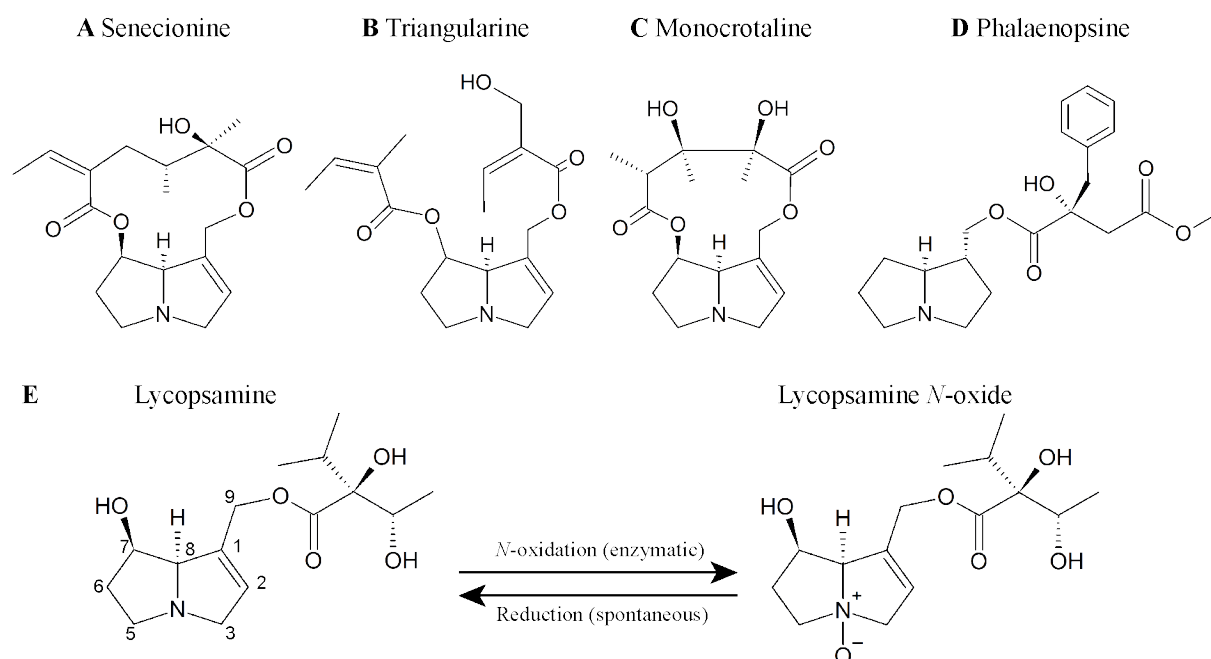


Figure 1 Typical structures of the different PA types as defined by Hartmann and Witte (1995). **A** Senecionine type, **B** Triangularine type, **C** Monocrotaline type, and **D** Phalaenopsine type. **E** Lycopsamine type, PAs are synthesized and stored as nontoxic *N*-oxide derivatives. Reduction to the tertiary pre-toxic form occurs spontaneously in the digestive system of the herbivores.

Necic acid moieties originate in the branched-chain amino acid pathway

Our knowledge about the biosynthesis of necic acids is scanty and could not be clarified on enzyme level so far. Tracer feeding experiments conducted in the 1960s until the early 1980s showed that all studied aliphatic necic acids are derived from the branched-chain amino acids (BCAAs) valine, leucine, and isoleucine and from their precursors. Isoleucine is incorporated into monocrotalic acid (Robins et al., 1974) and triangularine type necic acids (Crout, 1967). Even two molecules of isoleucine are incorporated into the senecionine type necic acids seneciphylllic, senecic and isatinecic acid (Crout et al., 1966; Crout et al., 1970; Crout et al., 1972; Bale et al., 1978). Leucine on the other hand, has been shown to be a precursor of senecioic acid (O'Donovan and Long, 1975). ^{14}C -labelled threonine, isoleucine, valine, and leucine were all incorporated into trichodesmic acid (senecionine type), whereby threonine and isoleucine mainly labelled one half of the C_{10} -acid, while valine and leucine predominantly labelled the other half (Devlin and Robins, 1984). For lycopsamine type PAs labelling studies in *Cynoglossum officinale* (Boraginaceae) suggested that either valine or its biosynthetic precursor 2-oxoisovalerate, is incorporated into the (-)-viridifloric acid with the missing C_2 unit coming from acyloin condensation with an activated acetaldehyde (Crout, 1966).

Prediction of the enzymatic formation of the necic acid moiety of lycopsamine type PAs

Crouts (1966) suggestions on the biosynthesis of the acid moiety of lycopsamine type PAs were confirmed by feeding ^{13}C -labelled glucose to root cultures of *Eupatorium clematideum* (Eupatoriaceae). Glucose is metabolized during glycolysis to ATP and Pyruvate (PYR), in this case ^{13}C -labeled PYR, which is in turn available as a substrate for BCAA biosynthesis. As shown in Figure 2, analyzing the ^{13}C labelling pattern of the 1', 2', 5', 6', and 7' carbon atoms of the (+)-trachelanthic acid moiety of trachelanthamine and the five carbon atoms of valine revealed that they are identical (Weber et al., 1999). This indicates that (+)-trachelanthic acid and valine are biosynthesized in the same plant compartment with 2-oxoisovalerate as a plausible common intermediate. Additionally, Weber et al. (1999) demonstrated that the coupling pattern of carbon atoms 3'/4' of (+)-trachelanthic acid are identical to the carbon atoms 5'/6' and to the carbon atoms 3/4 of valine. This indicates that hydroxyethyl, activated by thiamine diphosphate (HeThDP), is most likely the activated acetaldehyde transferred to 2-oxoisovalerate. The same activated acetaldehyde is transferred in the first step of BCAA biosynthesis by an acetohydroxyacid synthase (AHAS, EC 2.2.1.6). The AHAS with the tightly bound co-factor ThDP activates one donor PYR and transfers the resulting hydroxyethyl residue to either another PYR (C_3) or 2-oxobutyrate (C_4) to form 2-acetolactate or 2-acetohydroxy-2-oxobutyrate, respectively. Assuming an extended substrate specificity, it is quite conceivable that an AHAS might also be able to accept 2-oxoisovalerate (C_5) as a recipient of the hydroxyethyl residue, forming the C_7 backbone of necic acids of lycopsamine type PAs.

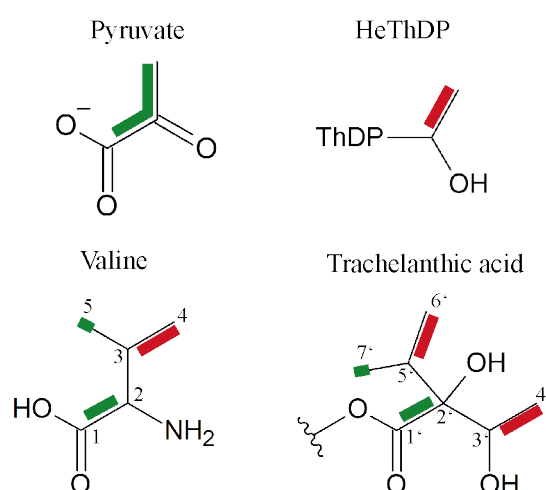


Figure 2 ^{13}C -labelling patterns after feeding of $[\text{U-}^{13}\text{C}_6]$ -glucose to root cultures of *Eupatorium clematideum*. Pyruvate, hydroxyethyl thiamine diphosphate (HeThDP), valine and the trachelanthic acid moiety of trachelanthamine. Carbon-carbon coupling patterns marked in green are derived from pyruvate, carbon-carbon coupling patterns marked in red are derived from HeThDP (according to Weber et al., 1999).

Objectives and outline of this work

In general, PA biosynthesis can be divided into two parts: (1) the synthesis of the bicyclic necine base and (2) the synthesis of the necic acid units. With respect to the first part, the biochemical characterization of HSS, the key enzyme of PA biosynthesis, and of HSO elucidated the steps leading from the polyamine putrescine and spermidine to the formation of 1-formylpyrrolizidine, the direct precursor of the saturated necine bases. Since the necine base is common to all PAs, it seems likely that the biosynthetic steps for its formation are mechanistically more or less identical. In contrast, very little is known about the second part of PA biosynthesis, the formation of the various necic acid units. The main objective of this work is to elucidate the first step of the formation of the unique branched-chain C₇-necic acid of lycopsamine type PAs. The main hypothesis follows the prediction of Weber et al. (1999) that the catalyzed reaction is the transfer of an activated acetaldehyde, formed from PYR, to 2-oxoisovalerate and that the catalyzing enzyme is an AHAS homolog.

The data of this thesis is presented in four chapters:

Chapter 2 describes the development of a rapid and easy assay method by which the keto acid substrates and the unstable acetohydroxyacid products of AHAS-catalyzed reactions can be detected and quantified. For this the reactive groups of the analytes were methylated or ethylated with methyl chloroformate or ethyl chloroformate, respectively, allowing GC-MS analysis of the derivatives. The method was tested with a variety of different keto acids with the well characterized AHAS isoenzyme II of *Escherichia coli* and a previously uncharacterized AHAS catalytic subunit of *S. officinale*.

Chapter 3 focusses on the analysis and characterization of a putatively PA-specific AHAS CSU in *S. officinale*, referred to as C₇-hydroxyacid synthase. *In planta* the gene encoding C₇-hydroxyacid synthase was targeted in a CRISPR/Cas9-mediated genome editing approach using *Agrobacterium rhizogenes* to generate stable mutant hairy roots. The hairy roots resulting from different constructs were tested at genomic level and the PA composition was analyzed by GC-MS. Additionally, the developed method described in Chapter 2 was used to confirm that the C₇-hydroxyacid synthase is able to catalyze the proposed formation of the C₇-necic acid backbone of the lycopsamine type PAs.

In **Chapter 4** different CRISPR/Cas9 C₇-hydroxyacid synthase knock-out and control line hairy roots from Chapter 3 were used to elucidate the following steps of the biosynthetic pathway of PAs in *S. officinale* to the core structure of intermedine in a metabolomic approach. Crude extracts of the different hairy roots were analyzed by LC-MS/MS and the resulting metabolomic data was compared to predict the sequence of the individual biosynthetic steps.

Chapter 5 summarizes miscellaneous efforts that were made to determine the role of an AHAS in PA biosynthesis. Thus, the putative C₇-hydroxyacid synthase from *E. cannabinum* was knocked out in a CRISPR/Cas9 approach and the effects at the genomic and PA phenotypic levels were investigated. As these are the first CRISPR/Cas9 experiments in *E. cannabinum*, the HSS was targeted in a control experiment. In addition, an attempt was made to heterologously overexpress the catalytic subunits of different AHASs from *E. cannabinum* and *H. indicum* to test them for their substrate specificity. This chapter briefly addresses also the evolutionary origins of the C₇-hydroxyacid synthases studied in this thesis.

In the last part of this thesis, a general conclusion is drawn based on the research results presented in the four chapters and follow-up experiments are proposed that might help in the future to answer the remaining open questions.

2 A rapid and efficient assay for the characterization of substrates and products of acetohydroxyacid synthases

A rapid and efficient assay for the characterization of substrates and products of acetohydroxyacid synthases

Annika Engelhardt^a, Elisabeth Kaltenegger^a, Dorothee Langel^a, Britta Muhs^a and Dietrich Ober^a

^a Botanisches Institut und Botanischer Garten, Christian-Albrechts-Universität zu Kiel, D-24098 Kiel, Germany

Detailed author contributions are listed at the end of the thesis.

2.1 Abstract

Acetohydroxyacid synthase (AHAS, EC 2.2.1.6) is the enzyme that catalyzes the first step in the synthesis of the branched-chain amino acids (BCAA) valine, leucine and isoleucine. AHAS catalyzes the decarboxylation of pyruvate and the specific condensation of the resulting thiamin diphosphate-bound (ThDP) two-carbon intermediate, hydroxyethyl-ThDP (HEThDP), with a second ketoacid, to form (2*S*)-2-acetolactate (AL) or (2*S*)-2-acetohydroxy-2-oxobutyrate (AHOB). A range of different methods to analyze the affinity of AHAS to PYR for the formation of AL and one to determine the preference of the AHAS for PYR or 2-oxobutyrate (2OB) as a second keto acid have been established. However, the simultaneous detection and quantification of substrates and products of the AHAS-catalyzed reaction is to our knowledge not possible with any of the methods available in the literature.

Here we present a new gas chromatography and mass spectrometry (GC–MS) method to monitor and characterize the AHAS-catalyzed reactions. Alkylation of the alcohol of the keto acid group, which is present in both substrates and products, stabilizes and volatilizes all test components, making them accessible for analysis by GC-MS. The method has validated over a broad range of different keto acids and acetohydroxy acid reaction products. To establish this method, we used the well-characterized AHAS isoenzyme II from *Escherichia coli* (*E. coli*, EcAHASII). We could show that the EcAHASII accepts 2-oxovalerate and 2-oxooctonate in addition to the substrates of BCAA biosynthesis. EcAHASII is not able to use one of the keto acids with branched side chain, *i.e.* 2-oxoisovalerate, 3-methyl-2-oxopentanoate, 4-methyl-2-oxopentanoate as acceptor of the activated acetaldehyde. In the absence of PYR, the EcAHASII is also able to effectively decarboxylate 2OB and transfer the resulting activated 3-carbon intermediate to another molecule of 2OB to form 2-ethyl-2-hydroxy-3-oxopentanoate. For comparative purposes we included an enzyme of plant origin, *i.e.*, the so far undescribed CSU of an AHAS from *Symphytum officinale* (comfrey, Boraginaceae, SoCSU) that we characterized with respect to its AHAS properties. We could show that the SoCSU catalyzes the specific condensation of HeThDP with the substrates of BCAA biosynthesis and 2-oxovalerate, but not with 2-oxooctonate or any keto acids with a branched side chain. A K_M of 11.1 ± 3.5 mM could be estimated for affinity of SoCSU to PYR.

2.2 Introduction

L-valine, L-leucine, and L-isoleucine form the small group of branched-chain amino acids (BCAAs). They are classified by their branched hydrocarbon residues that are responsible for their aliphatic character. The thiamine diphosphate (ThDP) dependent enzyme acetohydroxyacid synthase (AHAS, formerly called ‘acetolactate synthase’, EC 2.2.1.6) was identified as the first enzyme common to the biosynthetic pathways of all BCAAs. As shown in Figure 3, AHAS catalyzes the decarboxylation of the substrate pyruvate (PYR) to form the activated acetaldehyde hydroxyethyl ThDP (HeThDP) and the condensation of the activated acetaldehyde another molecule of PYR to form (2*S*)-2-acetolactate (AL), the biosynthetic precursor of L-leucine and L-valine. Additionally AHAS catalyzes the formation of (2*S*)-2-acetohydroxy-2-oxobutyrate (AHOB), the biosynthetic precursor of L-isoleucine, from one molecule of 2-oxobutyrate (2OB) and the activated acetaldehyde (Umbarger and Brown, 1958; Radhakrishnana and Snell, 1960).

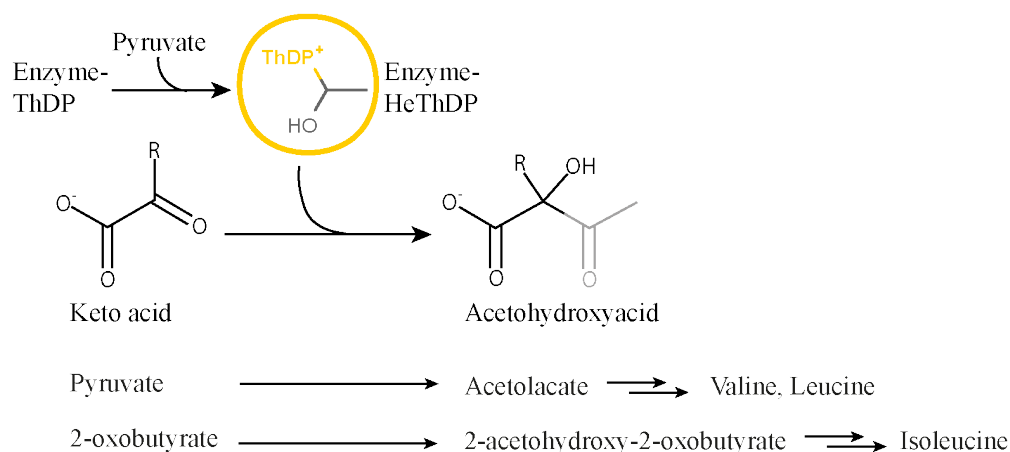


Figure 3 AHAS catalyzed reactions in branched-chain amino acid biosynthesis.

AHAS decarboxylates pyruvate to form an activated acetaldehyde bound to the thiamine diphosphate (HeThDP) and transfers it to one of the keto acids pyruvate or 2-oxobutyrate resulting in the acetohydroxyacid products acetolactate and 2-acetohydroxy-2-oxobutyrate, respectively.

The AHAS is a complex of catalytic (CSU) and regulatory subunits (RSU) (Duggleby and Pang, 2000). In addition to stabilizing the enzyme complex, the RSUs regulate the AHAS by feedback inhibition in response to BCAA levels in the cell, particularly that of valine (Pang et al., 2002; McCourt and Duggleby, 2006). The CSUs are active as dimer and it was shown that already this dimer possesses the entire catalytic machinery of AHAS (Vinogradov et al., 2006). This is particularly advantageous for *in vitro* analyses of AHAS, as the AHAS complex described to be very unstable (Grimminger and Umbarger, 1979; Eoyang and Silverman, 1986; Vinogradov et al., 2006). Therefore, the dimers can serve as a substitute for the holoenzyme when

characterizing AHAS with respect to catalytic activity and substrate specificity. For its catalytic activity AHAS requires one molecule each of three different cofactors: ThDP, a magnesium ion (Mg^{2+}) that allow the ThDP to be tightly bound in the catalytic center, and flavin adenine dinucleotide (FAD). In higher plants AHAS contains no introns, is nuclear encoded, and localized in the chloroplasts (Miflin, 1974; Chaleff and Mauvais, 1984; Jones et al., 1985). The amino acid sequence of the proteins derived from the gene sequence contains an N-terminal extension that is not present in the prokaryotic AHAS proteins and serves as a transit peptide to direct the enzyme post-translationally to the correct subcellular compartment (Jones et al., 1985; Rutledge et al., 1991; Duggleby and Pang, 2000). Further properties of AHAS enzymes have been compiled extensively in recent reviews (Liu et al., 2016). Due to the relevance of AHAS for BCAA biosynthesis in primary metabolism, they are an important target for herbicides of different classes (Duggleby et al., 2008; Li et al., 2018; Xie et al., 2021; Wei et al., 2022). Sulfonylurea derivatives, imidazoles and numerous others compounds inhibit AHAS in bacteria (LaRossa and Schloss, 1984), plants (Chaleff and Mauvais, 1984; Shaner et al., 1984) and fungi (Richie et al., 2013; Garcia et al., 2018). The BCAA pathways are expressed in plastids of the plant cell, the mitochondria of fungi, and are absent in animals. Therefore, inhibitors that are designed to target AHAS are regarded to be harmless to vertebrates and insects. In addition AHAS-inhibiting herbicides are studied as promising drugs for the treatment of microbial (Pue and Guddat, 2014) and fungal pathogenic infections (Garcia et al., 2018; Agnew-Francis et al., 2020; Meng et al., 2022) and are used to study AHAS-like enzymes that are involved in plant secondary metabolism (Weber et al., 1999; Chang et al., 2020). To study AHAS in all respects requires the characterization of enzymatic activity including (1) the affinity of AHAS for a specific substrate defined by the Michaelis-Menten constant and (2) the range of substrates, *i.e.*, the substrate specificity (Punekar, 2018). To analyze the affinity of AHAS to PYR for the formation of AL, various assay methods with discontinuous (Singh et al., 1988; Epelbaum et al., 1990) and continuous detection (Schloss et al., 1985; Vinogradov et al., 2005; Xie et al., 2019) have been established. Their detection based on indirect colorimetric detection (Singh et al., 1988; Epelbaum et al., 1990), the spectrophotometric observation of PYR consumption (Schloss et al., 1985), the detection of the newly formed chiral center of acetohydroxyacid products (Vinogradov et al., 2005), and the detection a fluorescence signal of an artificial acceptor substrate (Xie et al., 2019). For the determination of substrate specificity only one method is available that is based on a time-consuming decarboxylation of the acetohydroxyacids produced in the reaction. Decarboxylation results in the corresponding 2,3-diketones that have been detected by GC (Gollop et al., 1987). Direct detection of the reaction

products that would also provide structural data by fragmentation is not possible. The simultaneous detection and quantification of substrates and products of the AHAS-catalyzed reaction is to our knowledge also not possible with any of the methods available in the literature.

Here we present a new method to monitor and characterize the AHAS-catalyzed reactions. By alkylating the alcohol of the keto acid group that is present in both substrates and products all assay components are stabilized and become volatile so that they become accessible for analyses by GC-MS. To establish this method we used the well-characterized AHAS isoenzyme II from *Escherichia coli* (*E. coli*) (Hill et al., 1997; Engel et al., 2004b; Chipman et al., 2009; Vyazmensky et al., 2009; Steinmetz et al., 2010) for comparative purposes and an enzyme of plant origin, *i.e.*, the so far undescribed CSU of an AHAS from *Symphytum officinale* (comfrey, Boraginaceae) that we characterized with respect to its AHAS properties.

2.3 Results

To establish the assay method described here, the AHAS activity of two enzymes was analyzed. For well-known bacterial AHAS isoenzyme II from *E. coli* (*EcAHASII*), the substrates and products were derivatized with methyl chloroformate (MCF), whereas for the previously undescribed plant CSU of AHAS from *Symphytum officinale* (*S. officinale*, *SoCSU*), the compounds were ethylated with ethyl chloroformate (ECF).

2.3.1 Enzyme preparation

For expression of the *EcAHASII* we used the plasmid described by Hill et al. (1997) that contained the CSU with an *N*-terminal oligohistidine tag and of the RSU in its native form. After expression in the *E. coli* BL21(DE3) strain, the recombinant protein was purified via Ni-NTA affinity chromatography. SDS polyacrylamide gel electrophoresis confirmed the purification of a protein with an apparent molecular mass of 64 kDa (Figure 4 A) corresponding to that of the CSU. While a distinct protein band at about 9.5 kDa is seen in the run-through and wash step fractions, it is not seen in the elution fractions. However, an activity assay by the colorimetric method of Westerfeld showed that the purified CSU exhibited significant AHAS activity. This enzyme preparation was used for all further studies on substrate specificity of *EcAHASII*.

The sequence identification, amplification and transit peptide determination of *SoCSU* was done by Langel (2008). Since the *SoCSU* with N-terminal 6xHis-tag had no activity, an N-terminal Strep-tag®II was used instead. For this, the N-terminal 6xHis-tag of the first multiple cloning site of pCDFDuetTM-1 expression vector was replaced with a Strep-tag®II using oligonucleotides P1 and P2. The resulting vector was named pCDFDuet1_modStrep. The open reading frame of *SoCSU* without transit peptide was amplified and cloned into the SalI/NotI-linearized pCDFDuet1_modStrep vector. After transformation of the resulting construct into *E. coli* DH5a cells for DNA propagation and sequencing, the protein was heterologously expressed in *E. coli* BL21(DE3). The recombinant catalytic subunit with a predicted molecular weight of 64 kDa was purified via StrepTactin® affinity chromatography. In SDS polyacrylamide gel electrophoresis, the resulting protein showed a single band with an apparent molecular mass of approximately 60-70 kDa (Figure 4 B). An activity assay by the colorimetric method of Westerfeld (Singh et al., 1988; Epelbaum et al., 1990) showed that the purified CSU exhibited significant AHAS activity. This enzyme preparation was used for all further studies concerning the substrate specificity and kinetics of the plant enzyme.

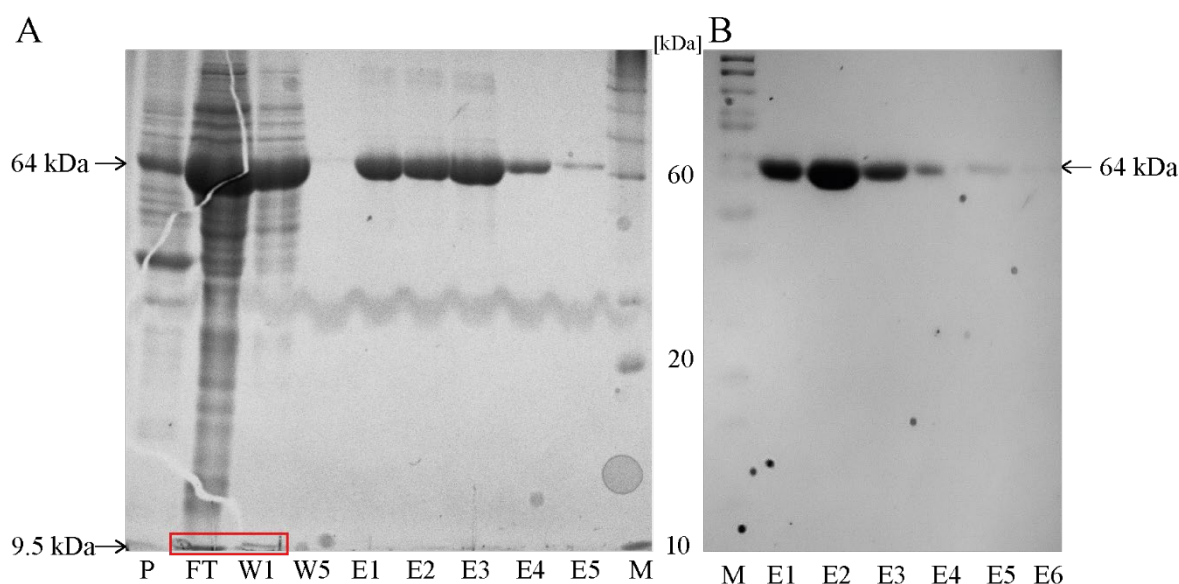


Figure 4 SDS-PAGE of fractions resulting from affinity purification of proteins. (12% gel, comassie blue stained). **A** *E. coli* AHAS II. The 6xHis-tagged catalytic subunit with a size of appx.64 kDa is labelled by an arrow. Co-purified non-tagged RSU of about 9.5 kDa was only detectable in the flow through and washing fractions. **B** *S. officinale* catalytic subunit. The strep-tagged catalytic subunit with a size of appx.64 kDa is labelled by an arrow. P = pellet, FT = flow through, W = washing fractions, E = elution fractions, M = protein size marker (PageRuler™ 10-200 kDa)

2.3.2 Requirements for an activity assay to characterize AHAS-catalyzed reactions

The validity of any measurement of the catalytic properties of an enzyme requires a suitable activity assay. In order to be able to perform activity tests of AHAS variants that allow the quantification of substrates and products, it was decided to develop a new method that would meet the following requirements:

- (i) Observation/identification of substrates
- (ii) Observation/identification of products, including those with unknown structure
- (iii) Simultaneous detection of several different products
- (iv) Rapid, simple and reproducible
- (v) Quantification of products

In addition to these requirements, one aspect has proved to be particularly challenging for method development: acetohydroxyacids are volatile and the solvent cannot be exchanged by evaporation and resolving the residue. Therefore, the development of a method to monitor the AHAS-catalyzed reaction requires the direct detection of substrates and products directly in the assay buffer or a protocol that allows their derivatization in the aqueous environment of the assay buffer. To our knowledge, none of the commonly used methods for characterizing AHAS

meets all of these criteria. Therefore, we tested various methods to implement as many of these requirements as possible. These are briefly described below.

2.3.3 Evaluation of methods for monitoring AHAS-catalyzed reactions

Gollop, Barak, and Chipman (1987) described an assay that allowed the simultaneous detection of the formation of AL and AHOB from PYR and 2OB. The described method involves oxidative decarboxylation of the acetohydroxyacids to the corresponding 2,3-diketones and their transfer to methanol by an elaborate air distillation apparatus to allow gas chromatography (GC) analytics. One strategy to simplify the elaborate assay of Gollop and colleagues to avoid the “air distillation” step by direct application of the volatile diketones to the GC by a headspace injector. This method worked very well for the diketone standards of acetoin and 2,3-butanedione. However, the detection of products formed in AHAS-catalyzed reactions showed very low sensitivity and reproducibility. The insufficient sensitivity might be due to very low concentrations of product formed in the assay or to an insufficient decarboxylation, causing product peaks to be lost in baseline noise.

Another idea was to measure the substrates and products of the assay reactions directly, without further processing in a high-sensitivity, high-resolution MS. Fourier-transform ion cyclotron resonance mass spectrometry is an ion-trap mass spectrometry that provides high mass resolution for product identification. In theory, the acetohydroxyacids and possibly also the keto acids should be measured with very accurate mass (± 2 ppm) and corresponding signal intensity. Unfortunately, no reproducible measurement data could be obtained using this method.

Different derivatization methods in aqueous milieu were tested: 2,4-dinitrophenylhydrazine is a derivatizing reagent that attaches the 2,4-dinitrophenylhydrazine to the carbonyl group, a common characteristic of the keto acids and acetohydroxyacids (Zwiener et al., 2002; Dong and Moldoveanu, 2004). The carbonyl group is also the target for *O*-(2,3,4,5,6-pentafluorobenzyl)-hydroxylamine hydrochloride (Cancilla and Que Hee, 1992). 4,5-dichloro-1,2-diaminobenzol should bind to the oxidative decarboxylation products of the acetohydroxyacids, the 2,3-diketones (Landaud et al., 1998). None of those derivatizations lead to reproducible results.

Methylation of AL with diazomethane (Langel, 2008) was the only method that resulted in clearly distinguishable, distinct peaks in GC-MS but the reproducibility was insufficient. On this basis, further methods for the alkylation of keto acids in the aqueous milieu were tested. First experiments with methyl chloroformate (MCF) (Hušek, 1991; Hušek, 1998; Smart et al., 2010; Madsen et al., 2016) and ethyl chloroformate (ECF) (Hušek, 1991; Hušek, 1998; Qiu et al., 2007) with typical keto acid substrates of the AHAs reaction, *i.e.*, PYR and 2OB resulted

in a rapid and reproducible alkylation. This observation encouraged us to test the derivatization with methyl and ethyl chloroformate as a method for the simultaneous detection of substrates and products of AHAS-catalyzed reactions.

2.3.4 Characterization of *Ec*AHASII reaction assays using MCF derivatization

Before we tested MCF derivatization as a suitable method for characterizing AHAS assays with *Ec*AHASII, we tested whether the substrates of the AHAS assays can be methylated and detected by GC-MS. For this purpose, PYR and 2OB were dissolved in assay buffer and derivatized using MCF and analyzed by GC-MS. For all three keto acids, distinct and well-separated peaks were detected that could be distinguished and identified by mass spectra (RI and mass spectra data are summarized in Supplementary Table 1). Due to the structural similarity and the presence of the same chemical group that serves as the target of methylation by MCF, we postulated that this method should also be suitable for detecting acetohydroxyacids, *i.e.*, the products of the AHAS reaction. To test this, the recombinant *Ec*AHASII was incubated in reaction mixtures containing a total of 10 mM substrate; either 10 mM PYR as donor and acceptor or 2.5 mM PYR as donor and 7.5 mM 2OB as acceptor substrate and stopped after 30 min by the addition of 5% NaOH. After derivatization with MCF the samples were analyzed by GC-MS. Figure 5 A and C show that the *Ec*AHASII exhibited AHAS activity catalyzing the formation of AL from two molecules of PYR and AHOB from PYR and 2OB. Additionally, *Ec*AHASII catalyzed the formation of 2-ethyl-2-hydroxy-3-oxopentanoate from two molecules of 2OB. These product peaks were attributed as such firstly by their clear formation after addition of the enzyme and secondly by their 70-eV-EI mass spectra (Figure 5 C and D). For AL and AHOB, the molecular ion peak, m/z $[M]^+$ 146 and m/z $[M]^+$ 160, correspondingly, could be detected. In addition, it was found that the keto group is preferentially cleaved during fragmentation, resulting in the fragment ion peaks of m/z $[M]^+$ 43 and counterparts m/z $[M]^+$ 103 for AL and m/z $[M]^+$ 117 for AHOB.

In addition to the substrates of BCAA biosynthesis, the substrate specificity of *Ec*AHASII was tested with five other structurally similar keto acids with carbon chains of different lengths and branching: 2-oxovalerate, 2-oxoisovalerate, 3-methyl-2-oxopentanoate, 4-methyl-2-oxopentanoate and 2-oxooctonate. The acceptor substrates were used at a concentration of 7.5 mM with 2.5 mM PYR as donor of the active acetaldehyde. Enzyme reactions were started by the addition of the protein, stopped after 30 min, and remaining substrates and products were methylated with MCF and analyzed by GC-MS. Figure 5 E and G show that *Ec*AHASII was able to convert two of the five substrates tested, *i.e.*, 2-oxovalerate and 2-oxooctonate respectively, into the corresponding acetohydroxyacids. Small amounts of

AL were also synthesized in these two reactions. The large difference in peak size of 2-oxooctonate between the comparative chromatogram of the enzyme reaction, despite only minor product formation, is due to the fact that 10 mM 2-oxooctonate and no PYR was used in the comparative chromatogram. The fragmentation of acetohydroxy-2-oxovalerate in the 70-eV-EI mass spectrum (Figure 5 F) follows those of AL and AHOB. In addition to the molecular ion peak of m/z $[M]^+ 174$, the keto group is cleaved, which is reflected in the characteristic fragment ions m/z $[M]^+ 43$ and m/z $[M]^+ 131$. These characteristic ions are also found in mass spectrum of acetohydroxy-2-oxooctonate (Figure 5 F), but while the ion of the keto group is clearly detectable, its counterpart m/z $[M]^+ 173$ is only detectable in traces. This strong fragmentation is due to the ionization process used, electron ionization (EI), which leads to significant fragmentation, especially of large molecules, already in the ion source. Chromatographic and mass spectrometric properties of all methylated substrates and products are summarized in Supplementary Table 1, 70-eV-EI mass spectra of additional products are shown in Supplementary Figure 2. The *Ec*AHASII is not able to use one of the keto acids with branched side chain, *i.e.* 2-oxoisovalerate, 3-methyl-2-oxopentanoate, 4-methyl-2-oxopentanoate as acceptor of the activated acetaldehyde and instead catalyzes the formation of AL from the available PYR (data not shown). We were thus able to show that the *Ec*AHASII is more promiscuous than previously described (Hill et al., 1997; Engel et al., 2003; Steinmetz et al., 2010). MCF derivatization of AHAS assays is an efficient method by which it is not only possible to follow standard reactions from substrate to product, but that also allows to be attentive to unexpected conversions. Impurities with 2,2-dihydroxypropionate were detectable in all three batches of PYR tested with a stated purity greater than 99%. Crucially, this 2,2-dihydroxypropionate is converted as a substrate, which could potentially lead to measurement inaccuracies with other detection methods. Another contamination with heptanoate was shown for the batch of 2-oxooctonate used here. But even though its structurally similar, heptanoate lacks the keto group and is thus not accepted as substrate by AHAS. However, one observation we made by derivatization with MCF is the fact that the PYR methyl ester elutes from the GC-MS column before the pyridine that is used in the derivatization reaction (chromatograms Figure 5). The huge pyridine peak distorts the chromatogram of the analytes of interest and makes accurate quantification of PYR difficult. This led us to change from methylation of substrates and products with MCF to ethylation with ethyl chloroformate (ECF) for all the following analyses.

Of note: The first step of the derivatization process, the addition of 5% NaOH to the assay mixture, results in an alkaline environment that is necessary for the transfer of an alkyl group

to the acid group but also inactivates enzymes present in the assay mixture. Thus, would allow to process several reaction mixtures in parallel that are stopped after a given time interval by the addition of NaOH and then successively derivatized. To test the stability of the reaction products in the alkaline environment, reaction mixtures were stopped with NaOH and allowed to stand for up to two hours before proceeding the methylation with MCF. No significant differences have been observed in the peak areas after GC-MS analyses of the compounds after these different derivatization protocols (data not shown).

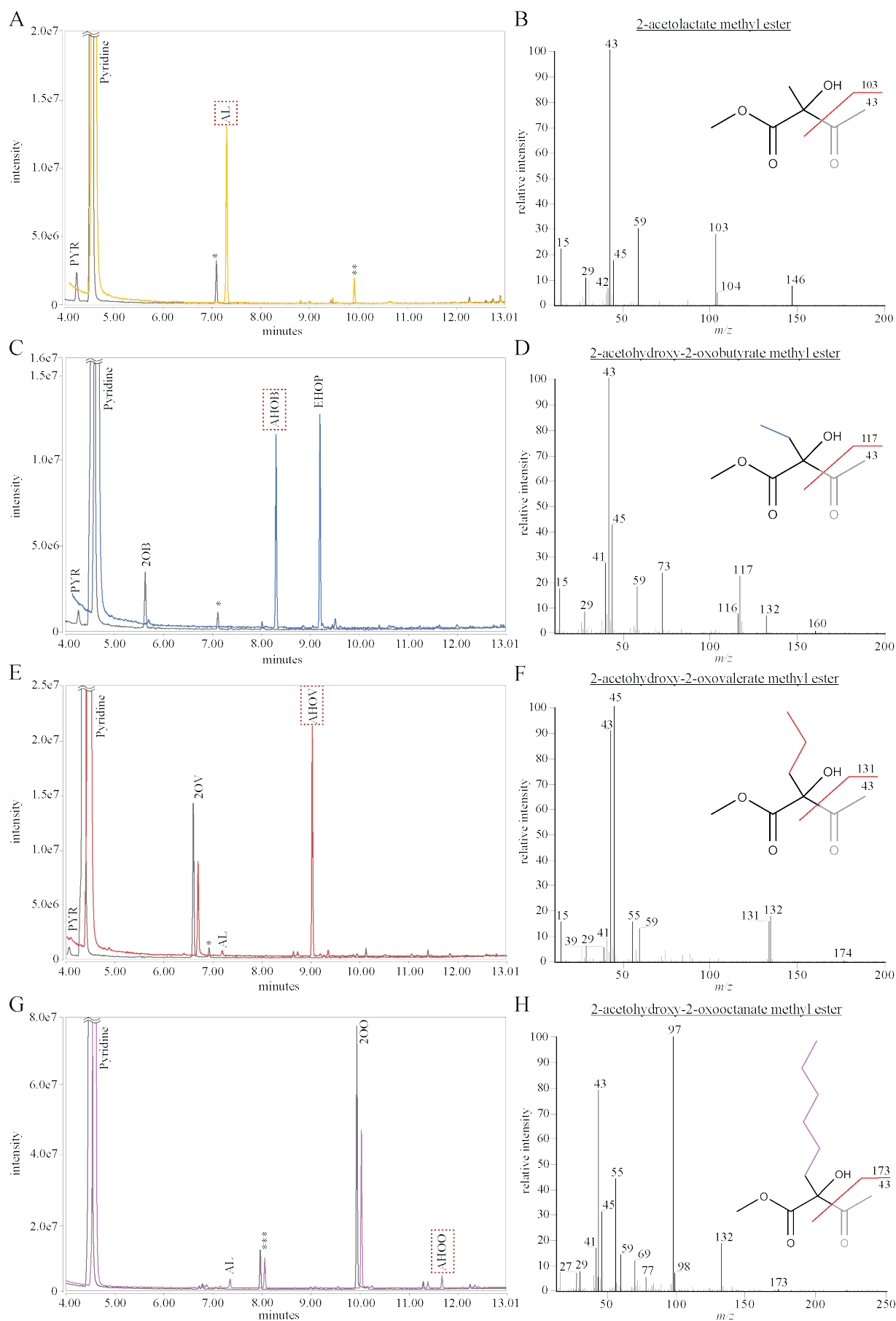


Figure 5 GC-MS total ion chromatograms and corresponding mass spectra recorded after MCF methylation of 30-minute enzymatic reactions of *Ec*AHASII with different substrate combinations. Chromatograms with active enzyme are marked in color, the corresponding chromatograms of the control reaction, with heat-inactivated enzyme, are shown in grey in the background. Red-dashed boxes indicate main products of which the mass spectra are shown. All analytes are present as methyl esters. Red lines indicate possible fragmentations with the m/z of the corresponding fragment ions that are supported by the observed data. **A** Substrate: PYR [10 mM]. **B** 70-eV-EI mass spectrum of methylated AL, m/z [M]⁺ 146. **C** Substrates: PYR [2.5 mM] and 2OB [7.5 mM]. **D** 70-eV-EI mass spectrum of AHOB methyl ester m/z [M]⁺ 160. **E** Substrates: PYR [2.5 mM] and 2OV [7.5 mM]. **F** 70-eV-EI mass spectrum of AHOV methyl ester, m/z [M]⁺ 174. **G** Substrates: PYR [2.5 mM] and 2OO [7.5 mM]; control reaction 2OO [10 mM]. **H** 70-eV-EI spectrum of AHOO methyl ester m/z [M]⁺ 216.

PYR = pyruvate; AL = acetolactate; 2OB = 2-oxobutyrate; AHOB = 2-acetohydroxy-2-oxobutyrate; EHOP = 2-ethyl-2-hydroxy-3-oxopentanoate; 2OV = 2-oxovalerate; AHOV = 2-acetohydroxy-2-oxovalerate; 2OO = 2-oxooctonate; AHOO = 2-acetohydroxy-2-oxooctonate. *threefold methylated 2,2-dihydroxypropionate, an impurity that was found in all tested batches of PYR; **product of AHAS reaction with 2,2-dihydroxypropionate as substrate; ***methylated heptanoate, an impurity of 2-oxooctonate

2.3.5 Ethylation instead of methylation improves the detectability of keto acids and acetohydroxyacids

PYR methyl ester has a retention index (RI) of 663, pyridine of 698, hence pyridine elutes later from the GC capillary column than PYR methyl ester. Due to the chain extension of the derivatized side chain, ECF ethylation of PYR results in a RI of 803, eluting after pyridine and all other solvents and waste products of the derivatization reaction. This allows to start the recording after the pyridine peak eluted from the column to obtain chromatograms that included all substrates and products of the AHAS reaction and allowed their accurate quantification and direct comparison.

Another advantage is that some of the substrates, *i.e.*, PYR, 2OB, and 2OIV, are commercially available as ethylated compounds and can be used for accurate characterization of ethylated derivatives. Figure 6 shows the efficiency of derivatization with ECF. Comparing the peak areas of the substrates ethylated with ECF with their ethylated standards showed that the peak area for all tested substrate concentrations in the range from 0.625 mM to 5.0 mM is about 75% smaller, suggesting that about 25% of the keto acids are derivatized. With higher substrate concentration (10 mM), the derivatization efficiency decreases to about 20%.

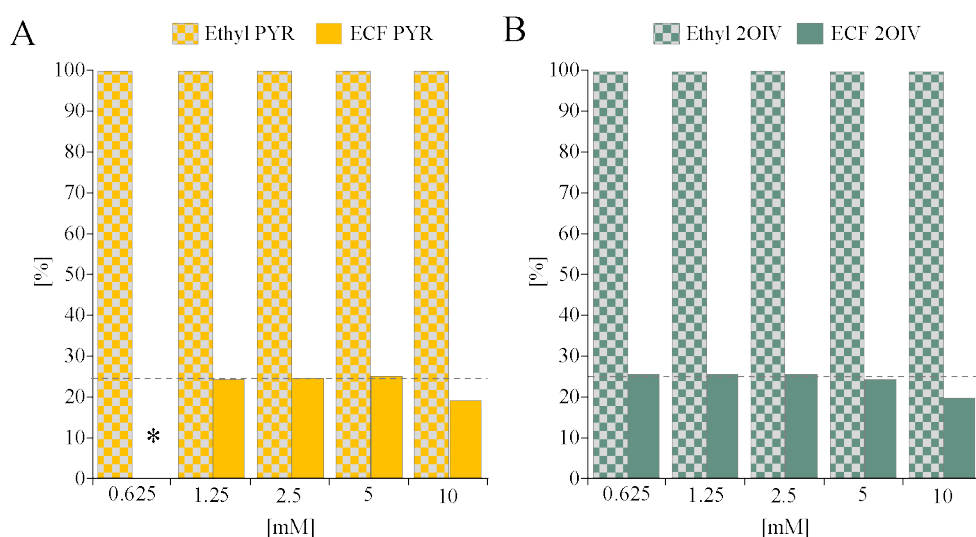


Figure 6 Derivatization efficiency of ECF with keto acids. Different concentrations (0.625, 1.25, 2.5, 5, 10 [mM]) of **A** PYR and **B** 2-oxoisovalerate (2OIV) were ethylated with ECF and the peak area in the total ion chromatogram was compared with that of purchased ethylated PYR and 2OIV, respectively. After ECF derivatization of 0.625 to 5.0 mM substrate, the peak area of the derivatives was shown to account for approximately 25% (grey dashed line) of the area of the purchased standards, while this value drops to only approximately 20% if 10 mM substrate were used for derivatization. * peak to noise ratio too low for clear peak integration.

To test whether the derivatization efficiency is constant over a concentration range of 1.0 to 6 mM keto acids, we were obtained calibration curves for the ethylation of PYR, 2OB and 2-oxoisovalerate, Figure 7 A, B and C respectively. It could be shown that all keto acids are linearly represented in this quantity range. The same can be assumed for the products, the acetohydroxyacids for which we were unable to prepare calibration curves due to the unavailability of reference compounds. Therefore, we decided to limit the total amount of keto acids per assay to 6 mM. Assay reactions with higher amounts of substrate require appropriate dilution before derivatization to guarantee comparable derivatization efficiency and comparability of substrate and product quantification. Of note, the size of the peaks in the total ion chromatogram depends not only on the amount of analyte but also on the size of its side chain. Figure 7 D shows the effect of the molecular size on the peak area of the ethylated keto acids PYR, 2OB, 2-oxoisovalerate, 4-methyl-2-oxopentionate, and 2-oxooctonate that have been derivatized in a mixture with equimolar amounts in a single derivatization reaction with ECF. Relative to the peak area of the internal standard, ethylated 2OB, the peak area of ethylated PYR differs by the factor of 0.7 while that of 2-oxoisovalerate, 4-methyl-2-oxopentionate, and 2-oxooctonate differ by the factor of 1.6, 4.0, and 18.5, respectively. This suggest that with the size of the keto acids also the factor referring the peak area to the internal standard increases. Therefore, only compounds should be used as an internal standard that are similar in size to the

substrates and products under study. For our studies, we have chosen 2OB as internal standard for all measurements except for those where 2OB was used as a substrate for which we have chosen 2-oxoisovalerate. Chromatographic and mass spectrometric properties of all ethylated keto acids are summarized in Supplementary Table 2, their 70-eV-EI mass spectra are shown in Supplementary Figure 3.

One way to determine the activity of the enzymes would be to observe the reduction in the amount of substrate. However, the smaller the conversions at high substrate concentrations, the less accurate the quantification would be. Therefore, we have opted for the quantification of the products. Since there are no acetohydroxyacids commercially available that could be used as standards to generate an external calibration curve for the direct quantification of the products, we used a keto acid as an internal standard. For this purpose, the peak area ratios of the keto acid substrates were applied analogously to the corresponding acetohydroxyacid products. Additionally, an impact of the molecule size on the peak area of the acetohydroxyacids has to be expected based on the observations with the keto acid substrates. To test this hypothesis, we used two commercially available compounds, *i.e.*, PYR methyl ester and the methyl ester of AL. These two compounds represent the substrate (PYR) and the product (AL) of the AHAS-catalyzed reaction after methylation with MCF. The two compounds were dissolved in chloroform at different molar concentrations and analyzed by GC-MS. Figure 7 E shows the normalized peak areas of PYR-methyl ester and AL-methyl ester measured at different molarities. It can be clearly seen that the peak areas of the AL methyl ester are always larger than those of the PYR methyl ester despite the same concentration. This difference averages at 2.7 and since the change in molecular size from substrate to product remains the same, whether methylated or ethylated, assumed the same factor between ethylated keto and acetohydroxyacids. Thus, for the calculation of the specific activity we considered that the molecular sizes of the keto acids among each other and the addition of the activated acetaldehyde to the keto acids have a decisive influence on the peak area and therefore on the quantification of the analyte.

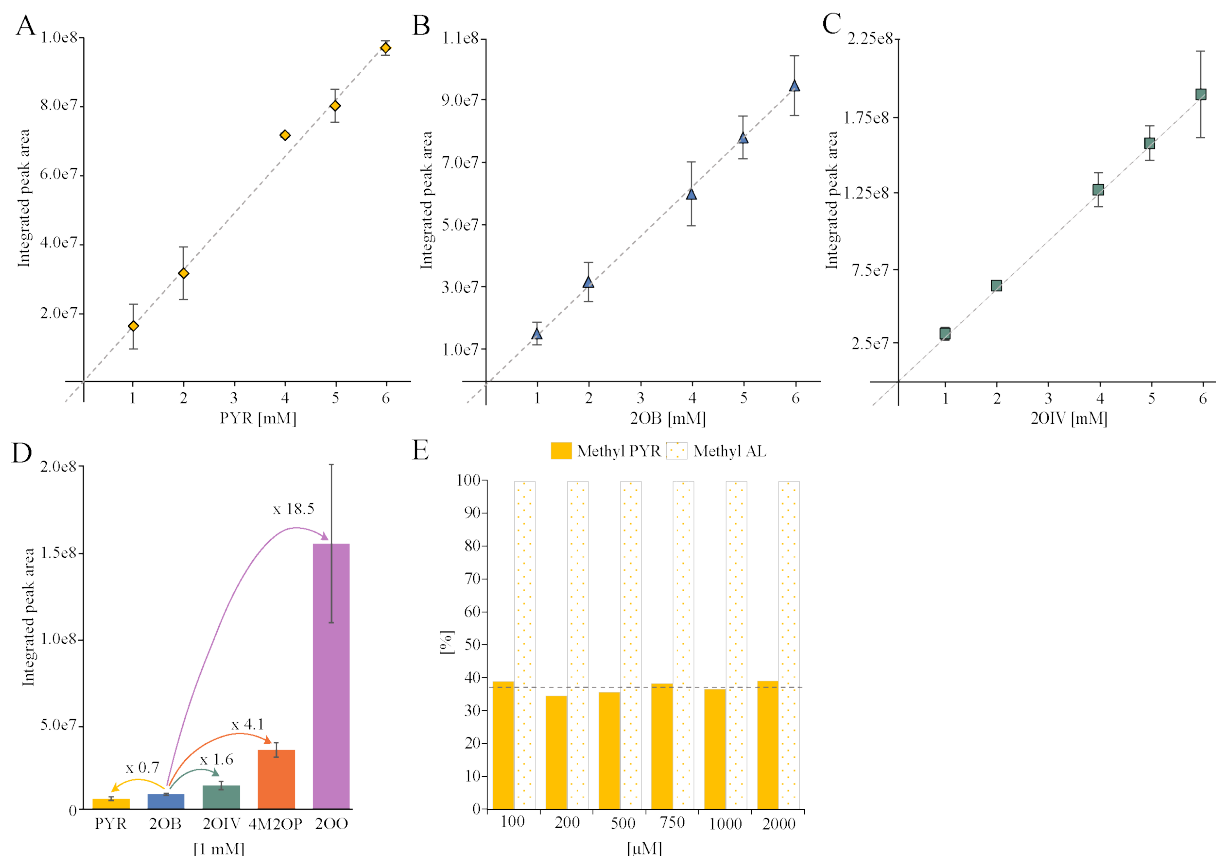


Figure 7 Peak areas of keto acids, ethylated with ECF. **A** Integrated peak areas of [1-6 mM] ethyl PYR. **B** Integrated peak areas of [1-6 mM] ethyl 2OB. **C** Integrated peak areas of [1-6 mM] ethyl 2OIV. **D** Pyruvate (PYR), 2-oxobutyrate (2OB), 2-oxoisovalerate (2OIV), 4-methyl-2-oxopentionate (4M2OP), and 2-oxooctonate (2OO) were mixed (1 mM each) and derivatized with ECF. The integrated peak areas were related to that of 2OB. All data points represent the mean \pm standard error of the mean (s.e.m.) of three technical replicates. **E** Comparison of integrated peak areas of purchased PYR and AL methyl ester at different molar concentrations. 100% values correspond to the integrated peak area of AL methyl ester.

2.3.6 Characterization of *SoCSU*

To test improved method for the detectability of keto acids and acetohydroxyacids by ethylation to monitor the AHAS-catalyzed conversion, we used the affinity-purified *SoCSU* in standard reactions with 6 mM substrate. In the case of the one-substrate reaction, 6 mM PYR as donor and acceptor of the activated acetaldehyde and in the case of two-substrate reactions, 1 mM PYR as donor and 5 mM of one of the various acceptor substrates, namely, 2OB, 2-oxovalerate, 2-oxoisovalerate, 3-methyl-2-oxopentionate, 4-methyl-2-oxopentionate, and 2-oxooctonate were used. After derivatization with ECF the samples were analyzed by GC-MS. Figure 8 A and C show that the *SoCSU* exhibited AHAS activity catalyzing the formation of AL from two molecules of PYR and AHOB from PYR and 2OB. In addition, the *SoCSU* was able to transfer the activated acetaldehyde to 2-oxovalerate as acceptor substrate, a conversion that results in the formation of 2-acetohydroxy-2-oxovalerate (Figure 8 E). None of the other substrates have

been accepted as substrates, including 2OO with a long aliphatic side chain and the branched-chain keto acids. Figure 8 B, D and F show the 70-eV-EI mass spectra of AL, AHOB, and 2-actohydroxy-2-oxovalerate where the molecular ion peaks (m/z $[M]^+$ 160, m/z $[M]^+$ 174, and m/z $[M]^+$ 188, respectively) could be detected, in the case of 2-actohydroxy-2-oxovalerate in traces. With respect to the fragmentation pattern we observed as for the methyl derivatives of the acetohydroxyacids also for the ethyl derivatives that the keto group is reproducibly split off during fragmentation. As consequence, the fragment ion peaks of m/z $[M]^+$ 43 is detectable in the MS spectra of all acetohydroxyacids, in addition, the compound specific fragments with m/z $[M]^+$ 117 for AL, m/z $[M]^+$ 131 for AHOB, and m/z $[M]^+$ 146 for 2-actohydroxy-2-oxovalerate are clearly detectable. 70-eV-EI mass spectra of all keto acids studied are summarized in Supplementary Figure 3. Chromatographic and mass spectrometric properties of all ethylated substrates and products are summarized in Supplementary Table 2. Of note, in the case of the one-substrate reaction with PYR resulting in the formation of AL and the two-substrate reaction with PYR and 2OB resulting in the formation of AHOB, two product peaks with identical mass spectra and almost identical retention times can be distinguished (inlets Figure 8 A and C). For 2-actohydroxy-2-oxovalerate, in contrast, only one product peak could be detected (Figure 8 E inlet). The identity of mass and fragmentation prompted us to speculate that the two peaks represent the different enantiomers of the product in their (S)- and (R)-conformation, even though such enantiomers should not be separated on the column we used. A second hypothesis was that the two peaks represent the two keto-enol tautomers. At least the first assumption was ruled out by derivatizing the structurally related amino acids valine and isoleucine in their (D)- and (L)-conformation. Neither derivatizing each amino acid enantiomer individually nor derivatizing the (D)- and the (L)-enantiomer simultaneously in one reaction provided evidence that the enantiomers are separated by our experimental setup (Supplementary Figure 4). The reason for the detection of almost identical product peaks needs further evaluation. For quantification of the amount of product, both peaks were integrated and the peak areas summed-up.

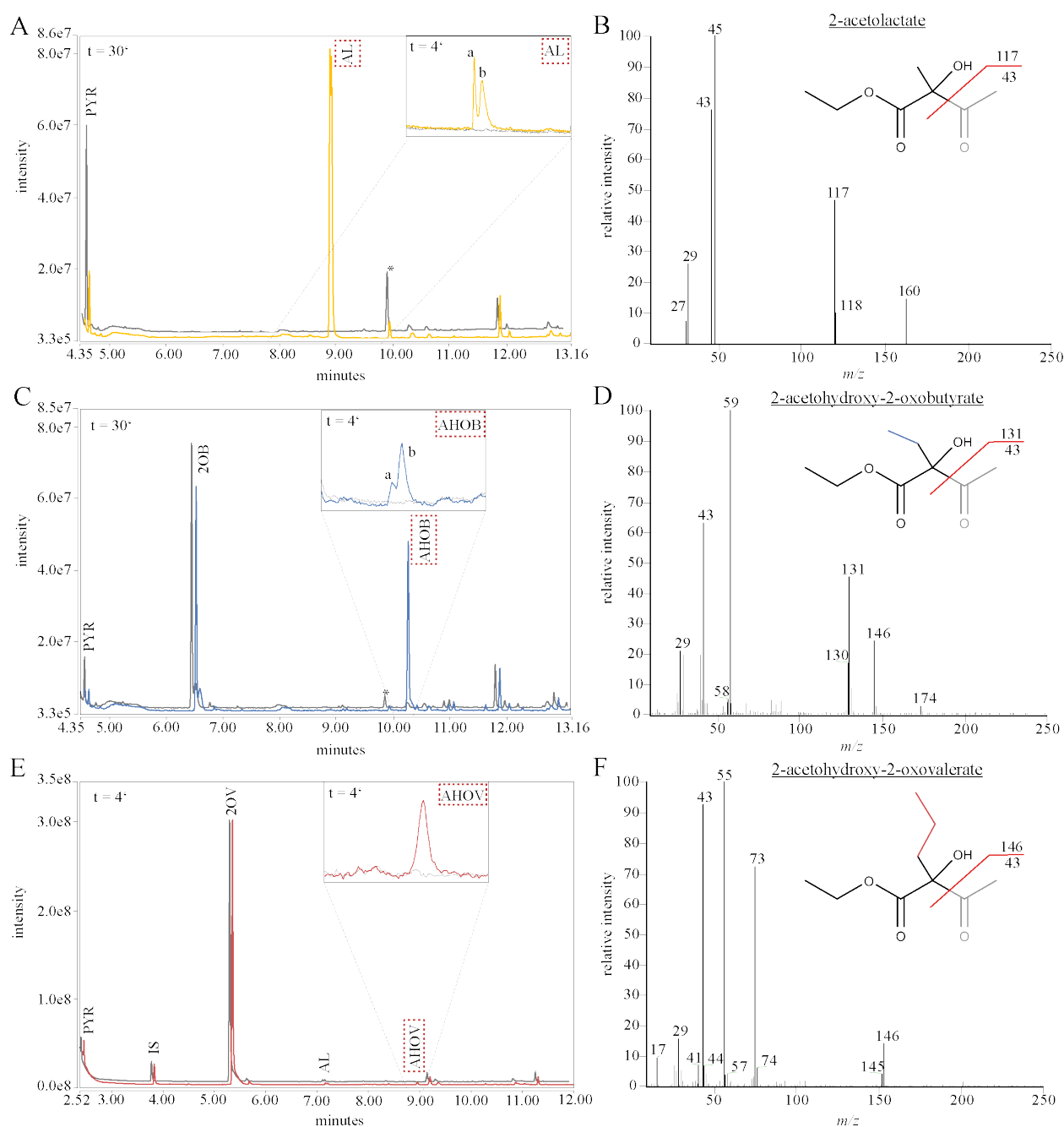


Figure 8 GC-MS total ion chromatograms and corresponding mass spectra recorded after ECF ethylation of enzymatic reactions of *So*CSU with different substrate combinations. Chromatograms resulting from active enzyme are marked in color, the comparative chromatograms resulting from incubations with heat-inactivated enzyme are marked in grey in the background. Red lines in the 70-eV-EI spectra represent possible fragmentations fitting the observed fragment ions. **A** Substrate: PYR [6 mM], 30-minute reaction. The inset shows the same assay procedure with a shorter reaction time of 4 minutes. This results in no signal overlap and thus a better separation of the possible stereoisomers of the products. **B** 70-eV-EI spectrum of AL ethyl ester m/z [M]⁺ 160. **C** Substrates: PYR [1 mM] and 2OB [5 mM]. **D** 70-eV-EI spectrum of AHOB ethyl ester m/z [M]⁺ 174. **E** Substrates: PYR [2.5 mM] and 2OV [7.5 mM]. **F** 70-eV-EI spectrum of the 2-actohydroxy-2-oxovalerate ethyl ester m/z [M]⁺ 188.

To test whether the GC-MS based assay could also be applied to study the substrate affinity of AHAS with enzyme kinetics the affinity of *So*CSU to PYR should be determined. Assays with affinity purified enzyme were performed with different PYR concentrations [0.5 – 20 mM] and stopped after 4 minutes. When plotting the specific activity against the substrate concentration the saturation curve roughly follows a Michaelis-Menten saturation curve (Figure 9). A K_M of 11.1 ± 3.5 mM could be estimated by the double reciprocal plot.

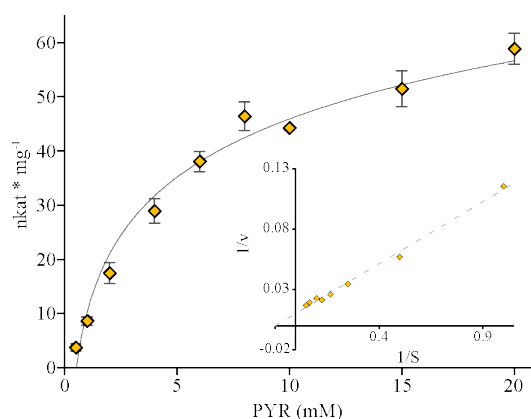


Figure 9 Formation of AL as a function of the pyruvate concentration. Michaelis-Menten plot with an inset Lineweaver-Burk plot for affinity-purified *So*CSU. 10 μ g of *So*CSU were incubated in a standard reaction assay with varying amounts of PYR [0.5-20 mM]. The V_{max} and the K_M values have been calculated to be 58.84 ± 2.85 nkat mg^{-1} and 11.1 ± 3.5 mM, respectively. All data points represent the mean \pm standard error of the mean (s.e.m.) of three technical replicates.

2.4 Discussion

2.4.1 State of the art of AHAS assay methods

The aim of this work was to develop a method that allows the simultaneous detection of substrates and products of a typical AHAS-catalyzed reaction, *i.e.*, keto acids and acetohydroxyacids, respectively. Our motivation was to be able to quantify in two-substrate reactions the amount of AL formed as the product resulting from PYR as sole substrate separately from the amount of acetohydroxyacids that result from the transfer of the activated acetaldehyde to another acceptor molecule. Over the years, various assays have been developed to characterize the enzymatic properties of AHAS. These assays will be discussed in more detail with respect to their applicability for our research question.

A well-established method developed by Singh et al. (1988) and modified by Epelbaum et al. (1990) for the determination of AHAS activity is based on an acid-catalyzed conversion of the chemically unstable acetohydroxyacids. The acetoin resulting from the conversion of AL and 3-hydroxy-2-pentanone resulting from the conversion of AHOB are determined colorimetrically after their reaction with a guanidine group under alkaline conditions (Westerfeld, 1945). This assay is characterized by an excellent sensitivity and offers an easy way to check whether the AHAS preparations show activity. However, this method does not allow to distinguish between the different products in two-substrate reactions. A method described by Schloss et al. (1985) allows a continuous assay, since here the consumption of PYR is observed spectrophotometrically at 333 nm. However, in this method the signal-to-noise ratio is low, since PYR has a rather low molar extinction coefficient at the respective wavelength and the detection of different products is not possible. Vinogradov et al. (2005) have taken advantage of the fact that in the AHAS-catalyzed reaction the achiral substrates are converted into products with a new chiral center, which is formed by the transfer of the activated acetaldehyde. These products display strong circular dichroism that can be detected by circular dichroism spectroscopy. However, again, the two products (*S*)-AL and (*S*)-AHOB, cannot be distinguished as the differences in the observed degrees of ellipticity are too small at 300 nm. For a clear differentiation of the products these must be enzymatically decarboxylated and the wavelength for optimal detection must be determined in each individual case. Another method to continuously monitor the activity of an AHAS-catalyzed conversion is a fluorogenic assay based on an intramolecular charge transfer-based fluorescent probe. This method allows a high throughput of samples and has a much better signal-to-noise ratio than is the case with the observation of pyruvate as described above (Xie et al., 2019). In this assay, the acceptor of the activated acetaldehyde is a fluorescent probe that binds efficiently in the

second substrate binding pocket and competes with the natural substrate PYR. Since the probe is used as an artificial acceptor substrate, this assay is limited to the observation of product formation with the probe and does not allow studies on the substrate specificity with keto acids. Gollop et al. (1987) developed a method that is based on the detection of the products in their decarboxylated form, *i.e.*, as 2,3-diketones, by GC. This makes this method unique as it allows to discriminate between the different products. The method involves the oxidative decarboxylation of the acetohydroxyacids to the respective volatile 2,3-diketones that are separated from the reaction mixture to methanol by using an “air distillation”. Disadvantages of this method are: (I) The products are not detected directly but in a degradation form that does not allow to distinguish between possible stereoisomers formed. (II) The detection of the substrates is not possible. (III) The “air distillation” is laborious and requires a lot of routine to obtain reproducible results.

2.4.2 Improvements in AHAS assay detection by ECF

As described above, the only method that allows the differentiation between different acetohydroxyacids resulting from AHAS-catalyzed reactions is the method by Gollop et al. (1987) that is based on the GC-analysis of 2,3-diketones. The relevance of a method that allows the differentiation between the different products has been emphasized by the authors:

“The outstanding characteristic of this method is its suitability for studying the specificity of purified ALS (acetolactate synthase = formerly used for AHAS) isozymes. Data (...) in which both physiologically significant reactions are followed, are important for understanding the physiological control of branched-chain amino acid biosynthesis. The method can also be used to investigate effects of inhibitors on specificity, and other aspects of the mechanism of action of the enzyme.” (Gollop et al., 1988)

We follow this statement and developed a method that is characterized by easier handling and by the direct alkylation of the substrates and products providing several advantages over the method based on the measurement of the 2,3-diketones:

(1) The substrates can be detected simultaneously with the products.

If the formation of the products can be observed in parallel with the consumption of the substrates, all aspects of the enzyme reaction can be monitored in the measurement, including aspects like substrate limitations or different rates of substrate consumption in two-substrate

reactions. Thus, with this method maximum information can be obtained about the enzymatic properties of AHAS.

(2) A variety of different substrates can be tested.

The proposed method allows a comprehensive characterization of the substrate specificity of AHAS. It also allows the analysis of unusual keto acids as substrates, since the derivatization targets the functional group and the side chain should have no or little influence for derivatization and detectability. In addition to the extensively studied substrates of BCAA biosynthesis, *i.e.*, PYR and 2OB, longer-chain and branched-chain keto acids have been successfully detected in addition to their products.

(3) The purity of the substrates can be verified.

The purity of substrates, buffers and other assay components plays a crucial role in the final outcome of an enzyme assay. We have been able to show that all analyzed batches of pyruvate contained minor amounts of 2,2-dihydroxypropionic acid (Figure 5 and Figure 8). As the AHASs studied in this project are able to convert also this by-product, the detectability of this impurity and of the product formed from this impurity is essential to analyze enzyme activities. Just to regard this impurity as background might have an impact on the kinetic data.

(4) The products of the reaction can be detected directly rather than their degradation products.

In BCAA biosynthesis, not only the formation of AL and AHOB can be observed but also products formed due to a wider substrate spectrum that is accepted by the AHAS. In this study we have been able not only to detect compound 2-ethyl-2-hydroxy-3-oxopentanoate formed by *Ec*AHASII from two molecules of 2OB, but also further products that have been formed from keto acids with longer unbranched side chains.

(5) Stereoisomers of the products could be separated via GC with a stereo column.

Discussing the supposed separation of enantiomers by the method presented here drew the attention already to the fact that the products of the AHAS-catalyzed reaction have stereocenters and should occur as individual enantiomers. As the alkylation of the acid group has no impact on the stereocenter, separation and detection of acetohydroxyacid enantiomers, *e.g.*, 2-*S*-AL and 2-*R*-AL by GC is in principle possible and necessary for a proper interpretation of the stereospecificity of AHAS.

(6) The efficiency of the derivatization of the acetohydroxyacids can be determined.

Gollop et al. (1987) pointed out that it would be desirable to be able to quantify the efficiency of the conversion of the acetohydroxyacids to their corresponding 2,3-diketones. In our study we have been able to show by comparing equimolar amounts of ethylated keto acids obtained

as standards and of keto acids resulting from derivatization with ECF that the derivatization efficiency is approximately 25%. Due to the structural similarity of the analytes, it can be speculated that the derivatization efficiency for acetohydroxyacids should be in the same range of magnitude. Our analyses confirm previous observations where it was shown that the choice of substrate does not have an impact on the derivatization efficiency of the chloroformate derivatizing agents (Hušek, 1998; Smart et al., 2010; Madsen et al., 2016).

2.4.3 Possible applications

Alkylation as a derivatization method allows the conversion of amino acids and organic acids into volatile carbamates and esters that are easily detectable by GC and encompass a group of compounds that represent many intermediates of central carbon metabolism (Hušek, 1998; Smart et al., 2010).

Considering all substrates, intermediates, and products of BCAA biosynthesis (Figure 10), it is evident that we have already been able to derivatize and detect 10 out of 14 of them in the context of our investigations of substrate specificity by GC-MS by using the ethylation with ECF presented in this manuscript. Based on the structural similarity, we assume that the remaining compounds not analyzed in this study, *i.e.*, 2-isopropylmalate, 3-isopropylmalate, 2,3-dihydroxy-3-methylbutyrate, and 2,3-dihydroxy-3-methylpentanate can also be detected by GC-MS after ethylation with ECF. Thus, the method presented here could be applied to the analysis of the entire BCAA biosynthesis, starting from the formation of 2OB from threonine by threonine dehydration/deaminase up to the formation of the BCAAs valine, leucine and isoleucine by the branched-chain aminotransferase.

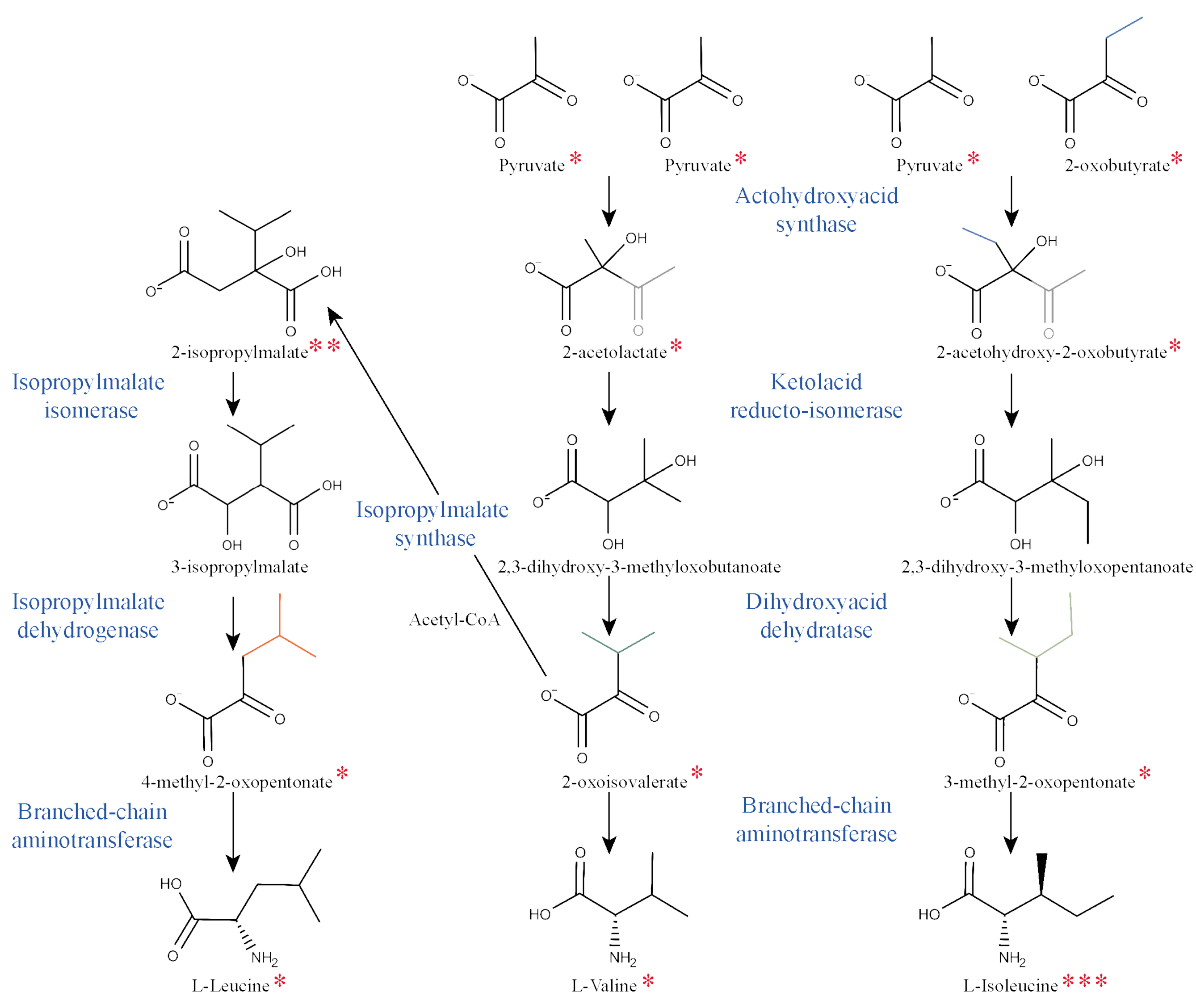


Figure 10 Key enzymes and metabolites of branched-chain amino acid biosynthesis. *derivatized with MCF and/or ECF in this work; **derivatized with MCF by Smart et al; ***derivatized with ECF by Qui et al 2006

2.4.4 Conclusion

We described a convenient, quantitative and precise analytical method using GC–MS with MCF and ECF alkylation for the analysis of AHAS enzymatic reactions. It allows the detection of all substrates and products of the AHAS-catalyzed reaction within branched-chain amino acid biosynthesis and thus allows studies on substrate affinity and substrate preference of PYR and 2OB. In addition, studies of substrate specificity are possible by the analysis of structurally similar keto acids and acetohydroxyacids.

2.5 Experimental

2.5.1 Material

The plasmid pET-GM (Hill et al., 1997) was obtained from Dr. Ronald Duggleby (University of Queensland, Brisbane, Australia). This construct expresses the large subunit (59 kDa) with a 50-residue (5.43 kDa) N-terminal fusion oligohistidine-tag, and the small subunit in its native form (9.5 kDa).

Restriction endonucleases, DNA-depending enzymes, and other molecular biology reagents were purchased from Thermo Fisher Scientific (Waltham, MA, USA), New England Biolabs (Ipswich, MA, USA) or Promega (Fitchburg, WI, USA) and, unless stated otherwise, were used in accordance with the manufacturers' instructions. Other chemicals were from Merck (Darmstadt, Germany), Fluka (Seelze, Germany), Carl Roth (Karlsruhe, Germany) or Sigma-Aldrich (St Louis, MO, USA). Oligonucleotides were synthesized at MWG Eurofins Genomics (Ebersberg, Germany) and are given in Table 1.

Table 1 Oligonucleotides.

No	Sequence 5'→3'	purpose
P01	<u>CATGGCTAGCTGGAGCCACCCG</u> CAGTTCGAAAAAGCCAG	SrepTagII cloning forward, overhang for ligation underlined
P02	GATCCTGGCTTTTTTCGAACGCGGGTGGCTCCAGCTAGC	SrepTagII cloning reverse, overhang for ligation underlined
P03	TATAGTCGACCCACCTTCTTTTCCAGGTACGCCCCGACGAG	SoCSU cloning primer forward with Sall <u>restriction site</u>
P04	TATATAGCGGCCGCTCAGTATTTTGTCTTCCATCGCCTTC	SoCSU cloning primer reverse with NotI <u>restriction site</u>

2.5.2 Heterologous expression of *E. coli* AHAS isoenzyme II

The AHAS isoenzyme II of *E. coli* was expressed and purified as described by Hill et al. (1997). As it was reported that AHAS might be light sensitive (Chang et al., 1997), all steps of protein purification were protected from light whenever possible. Briefly, after transformation of the plasmid pET-GM, containing the CSU with a with an N-terminal oligohistidine tag and of the RSU in its native form, into *E. coli* expression strain BL21(DE3), expression cultures were induced at OD₆₀₀ of 0.6 to 0.8 with 0.5 mM IPTG and cultivated for 20 h at 22°C on a gyratory shaker. Harvested cells were resuspended in 10 ml lysis buffer (50 mM NaH₂PO₄·xH₂O, 300 mM NaCl, 5 mM β-mercaptoethanol, 20 mM imidazole, pH 7.5) and sonicated twice for 7 min using a Branson sonifier (Heinemann, Schwäbisch Gmünd, Germany, cycle 40, duty 4.). The supernatant was added to 200 µl of nickel-nitrilotriacetic agarose (Qiagen, Hilden, Germany) equilibrated with lysis buffer and incubated end-over-end at 4°C for 1 h. Recombinant proteins were purified according to the manufacturer's protocol.

2.5.3 Heterologous expression of *SoCSU*

For heterologous expression of the *SoCSU*, a modified pCDFDuetTM-1 vector (Novagen) was used in which the His-tag was replaced with a Strep-tag®II. For this purpose, the oligonucleotides P01 and P02 (Table 1) were hybridized resulting in NcoI- and BamHI-compatible sticky ends for cloning into the NcoI-BamHI-linearized pCDFDuetTM-1 vector. The resulting vector was named pCDFDuet1_modStrep, the vector map is given in Supplementary Figure 1. The open reading frame of *SoCSU* without transit peptide was amplified using the primer pair P03 and P04 introducing SalI and NotI restriction sites and the proofreading Phusion HF DNA polymerase (Thermo Fisher Scientific). The resulting PCR product was cloned into the SalI/NotI-linearized pCDFDuet1_modStrep vector. After propagation in *E. coli* DH5 α cells (Thermo Fisher Scientific) the constructs were sequenced and transformed into the expression strain *E. coli* BL21(DE3). Expression cultures were induced at OD₆₀₀ of 0.6 to 0.8 with 0.5 mM IPTG and cultivated for 20 h at 22°C. Harvested cells were resuspended in 10 ml buffer W (100 mM Tris pH 8, 150 mM NaCl, 1 μ M DTT) and sonicated twice for 7 min using a Branson sonifier. For purification, a Strep-Tactin® Superflow® column (1.2 ml bed volume, IBA Lifesciences, Göttingen, Germany) was used according to the manufacturer's instructions. Briefly, the protein extract was added to the column, washed five times with buffer W, and eluted with 1x buffer E (buffer W supplemented with 2.5 mM desthiobiotin, pH 8.0). The affinity-purified protein was rebuffed via a PD10 column (GE Healthcare, Chicago, IL, USA) to assay buffer (0.05 M K₂HPO₄, 1 mM MgCl₂, 1 mM Thiamine pyrophosphate, 20 μ M FAD, 1 μ M DTT, pH 7.9) and stored at -80°C after addition of 1% glycerin without major loss of activity for up to one year.

Successful expression and purity were monitored by SDS-PAGE analysis. Protein quantities were estimated according to Bradford (Bradford, 1976) with bovine serum albumin as reference.

2.5.4 AHAS activity assay based on indirect colorimetric detection

The colorimetric assay of AHAS was first described by Westerfeld 1945 and used with PYR as substrate according to an optimized protocol (Singh et al., 1988) to test enzyme preparations for activity. The produced AL is measured spectrophotometrically after decarboxylation to acetoin that forms with creatine and naphthol a color complex. 10 μ g of enzyme was incubated in 250 μ l assay buffer containing 10 mM PYR for 30 min. at 30°C. The reaction was stopped with 25 μ l of 6N H₂SO₄ and incubated for the formation of the diketones for 15 min. at 60°C. This was followed by the addition of 275 μ l of 5% (w/v) 1-naphthol solution in 4 M NaOH and 275 μ l of 0.05% (w/v) creatine solution in H₂O. The mixture was incubated for 15 min. at 60°C

to allow color complex formation. Enzyme activity was indicated by a color change from brownish to pink (Figure 11).

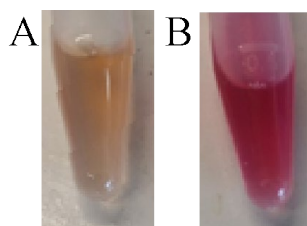


Figure 11 Colorimetric assay of AHAS with 10 mM PYR. **A** Control reaction without enzyme. The reaction has the brownish color of 1-naphtol. **B** Reaction with active AHAS. A pink color complex is formed from acetoin with 1-naphtol and creatine.

2.5.5 AHAS activity assay based on alkylation of substrates and products

Reactions were carried out in a total volume of 200 μ l with varying concentrations of substrates that have been prepared as 0.2 M stock solutions in assay buffer to guarantee that the concentration of all cofactors were identical in all reactions. The reactions were started by addition of the enzyme and incubated at 30°C. Reactions were stopped by addition of 40 μ l 5% (w/v) NaOH. The reactions were conducted without direct light exposure and darkened whenever possible.

For derivatization of substrates and products alkyl chloroformates were used that are highly toxic and require to work under a fume hood during derivatization and to exercise special caution. Methylation of the keto and acetohydroxy acids was performed with methyl chloroformate (MCF, Sigma-Aldrich) in 2 ml reaction tubes. To the enzyme reactions that have been stopped with 40 μ l 5% (w/v) NaOH (total volume 240 μ l). 200 μ l methanol (donor of the methyl group), 50 μ l pyridine (catalyst of the reaction), and 25 μ l MCF were added and the reaction mixture vortexed for 30 s. During the procedure, the tube was opened repeatedly (in intervals of approx. 10 s) to prevent overpressure in the reaction tube due to gas formation. The methylated compounds were extracted into an organic phase by addition of 400 μ l chloroform and subsequent vortexing for 10 s. If several samples were processed in parallel, a pause point was inserted here. By adding 400 μ l NaHCO_3 and 10 s vortexing possible precipitates were removed from the organic phase. After removal of the upper, aqueous phase 100 mg Na_2SO_4 were added followed by vortexing for 10 s and centrifugation (12 000 g) for 30 s to dry the organic phase. During transfer of the sample to a glass μ -vial with a 300 μ l insert (Carl Roth) for GC-MS analytics care was taken not to transfer any crystals from the bottom of the tube as these would be harmful to the GC-MS.

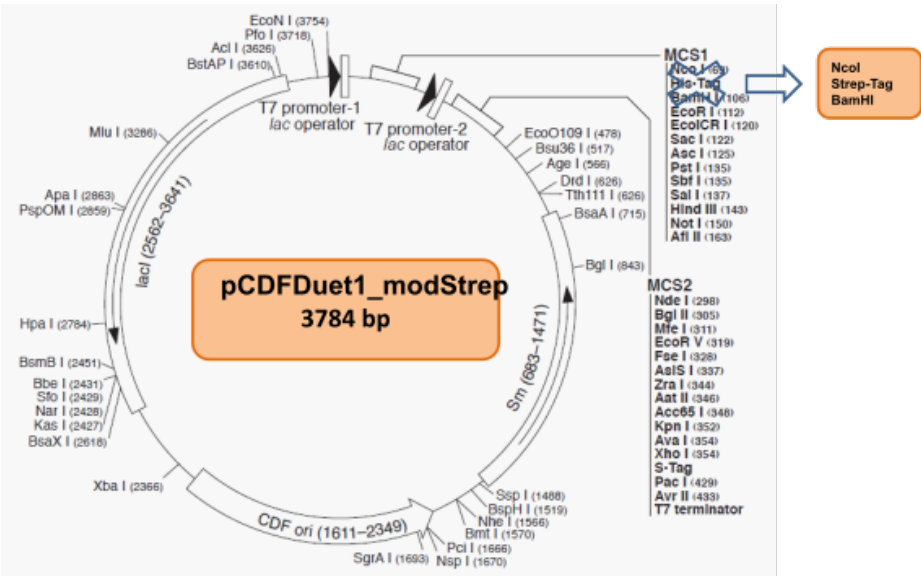
GC-MS data were obtained with a Trace Ultra GC coupled to a Dual Stage Quadrupole Mass Spectrometer (Thermo Fisher Scientific) equipped with a split/splitless (SSL) injector and a TG-5SILMS (30 m × 0.25 mm i.d., 0.25 µm film thickness, Thermo Fisher Scientific). GC conditions were: split ratio 1:10, temperature program 70°C for 3 min, 70–200°C at 10°C/min, 200–280°C at 25°C/min followed by 1 min constant at 280°C, carrier gas helium 1 ml/min, MS transfer line 300°C. EI-mass spectra were recorded at 70 eV (ion source temperature 230°C). The linear retention indices were calculated according to Kováts (Kováts, 1958) with a reference set of co-injected hydrocarbons (Sigma-Aldrich). Identification of analytes was achieved via MS spectra and RI value comparison with NIST database, literature (Smart et al., 2010; Madsen et al., 2016), and comparative analyses of purchased standard compounds.

2.5.6 ECF Ethylation of acids

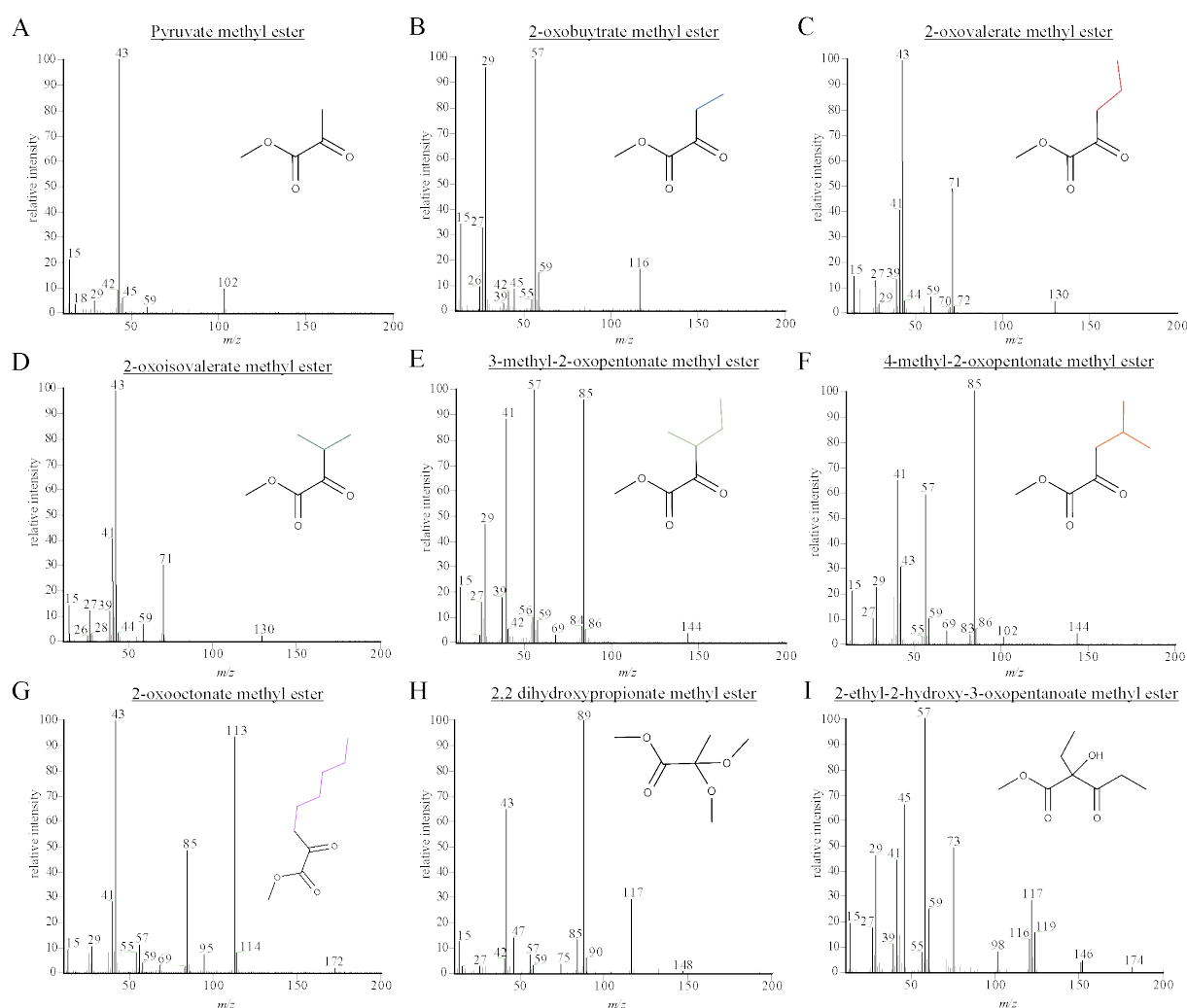
Ethylation of the keto and acetohydroxy acids was performed with ethyl chloroformate (ECF, Sigma-Aldrich) in 2ml reaction vials following the protocol given for derivatization with MCF with following modification: 100% ethanol were used instead of methanol as donor of the ethyl group and ECF instead of MCF for derivatization.

GC-MS data were obtained with a Thermo TSQ TRACE1300/1310 coupled to a Triple Quadrupole Mass Spectrometer (Thermo Fisher Scientific) equipped with a PTV injector and a TG-5SILMS (30 m × 0.25 mm i.d., 0.25 µm film thickness, Thermo Scientific). GC conditions were: PTV injector 280°C, split ratio 1:10, temperature program 50°C for 3 min, 50°C - 200°C at 10°C/min, 200–280°C at 25°C/min followed by 1 min at 280°C, carrier gas helium 1 ml/min, MS transfer line 300°C. EI-mass spectra were recorded at 70 eV (ion source temperature 280°C).

2.6 Supplementary



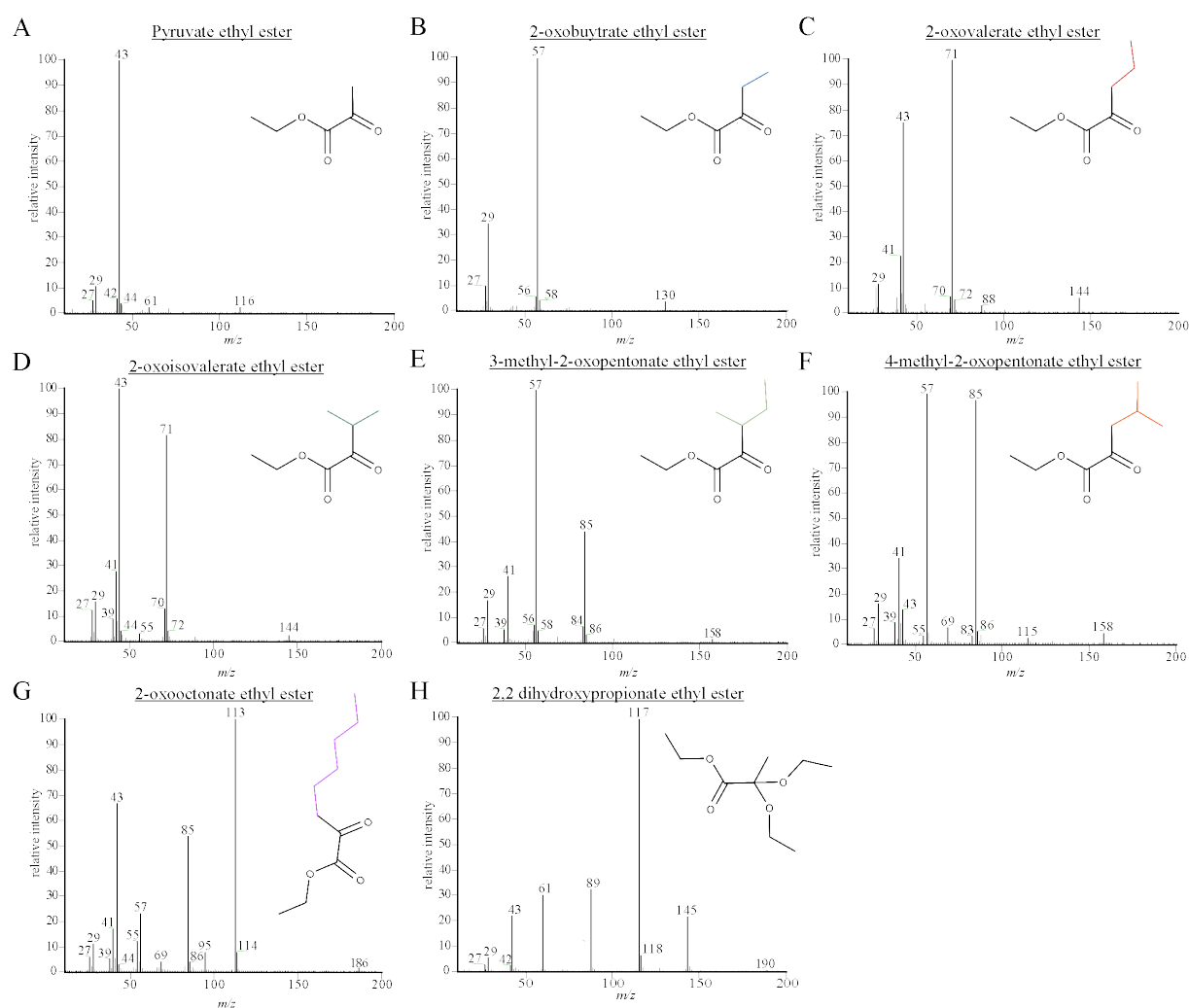
Supplementary Figure 1 Vector map of pCDFDuet-1_modStrep. The N-terminal His-Tag in MCS1 was replaced with a Strep-Tag using restriction sites NcoI and BamHI. The annealed complementary oligonucleotides P01 and P02, contain the Strep-Tag and matching overhangs to the restriction sites.



Supplementary Figure 2 GC 70-eV-EI mass spectra of keto acids and additional acetohydroxyacids after methylation with MCF. **A** Pyruvate methyl ester, m/z [M]⁺ 102. **B** 2-oxobutyrate methyl ester, m/z [M]⁺ 116. **C** 2-oxovalerate methyl ester, m/z [M]⁺ 130. **D** 2-oxoisovalerate methyl ester m/z [M]⁺ 130. **E** 3-methyl-2-oxopentionate methyl ester, m/z [M]⁺ 144. **F** 4-methyl-2-oxopentionate methyl ester, m/z [M]⁺ 144. **G** 2-oxooctonate methyl ester, m/z [M]⁺ 172. **H** 2,2-dihydroxypropionate methyl ester, m/z [M]⁺ 148. **I** 2-ethyl-2-hydroxy-3-oxopentanoate (EHOP) methyl ester, m/z [M]⁺ 174.

Supplementary Table 1 Keto acid and acetohydroxyacid derivatives of the AHAS assay reactions methylated with MCF. Retention indices (RI) on TG5-SILMS capillary column

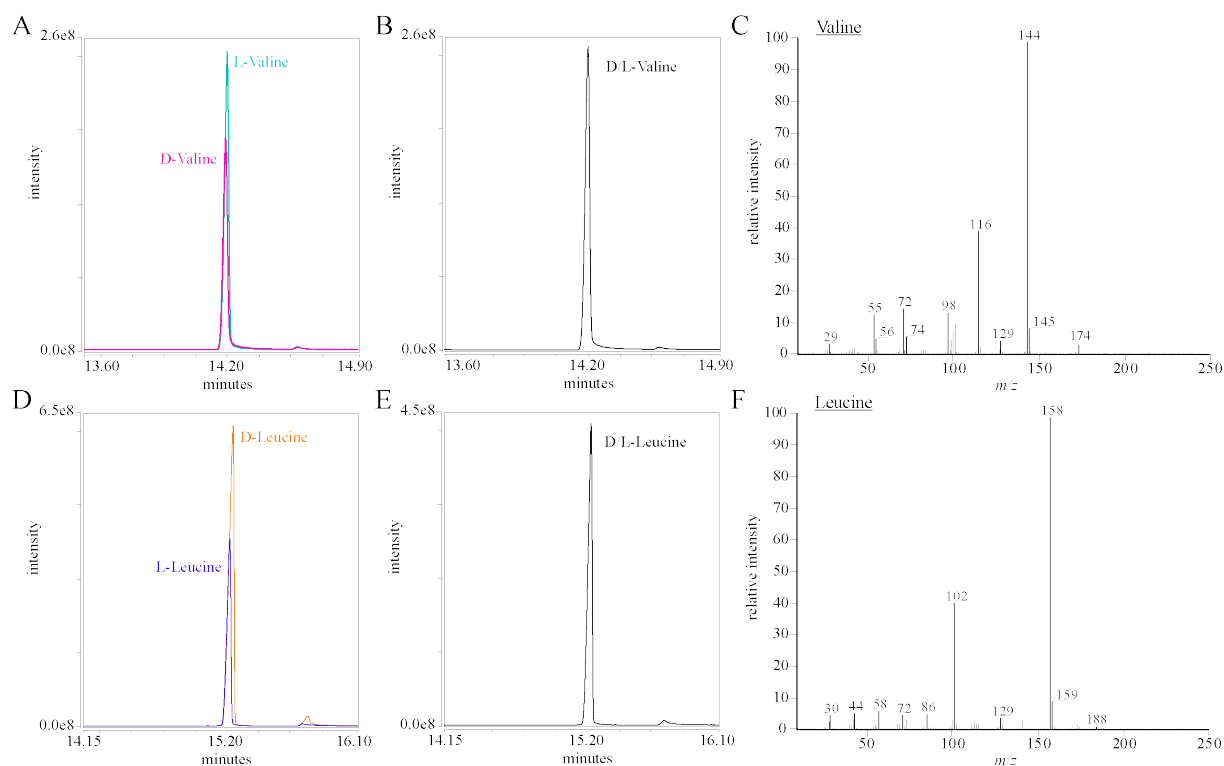
Substance	Formula	[M] ⁺	RI	MS data: characteristic ions m/z (% relative abundance)
Keto acids				
Pyruvate methyl ester	C ₄ H ₆ O ₃	102	663	14(5), 15(25), 16(1), 17(1), 28(3), 29(5), 30(1), 41(2), 42(10), 43(100), 44(3), 45(6), 47(2), 59(1), 73(2), 82(4), 83(2), 85(4), 102(10), 103(2)
2,2-dihydroxypropionate methyl ester	C ₆ H ₁₂ O ₄	148	966	14(1), 15(10), 17(1), 29(4), 31(2), 42(4), 43(59), 44(2), 45(2), 47(12), 57(7), 58(1), 59(2), 75(3), 85(11), 86(1), 89(100), 90(4), 117(23), 133(2)
2-oxobutyrate methyl ester	C ₅ H ₈ O ₃	116	809	26(2), 27(10), 28(4), 29(36), 30(1), 31(1), 39(1), 41(1), 42(1), 43(1), 45(2), 55(1), 56(5), 57(100), 58(4), 73(1), 74(1), 75(2), 87(1), 130(3)
2-oxovalerate methyl ester	C ₆ H ₁₀ O ₃	130	932	15(15), 26(2), 27(16), 28(3), 29(4), 38(2), 39(12), 40(3), 41(43), 42(13), 43(100), 44(4), 45(2), 55(2), 59(6), 69(1), 70(2), 71(49), 72(2), 130(5)
2-oxoisovalerate methyl ester	C ₆ H ₁₀ O ₃	130	871	15(16), 18(4), 27(13), 28(4), 29(3), 39(11), 40(2), 41(39), 42(11), 43(100), 44(4), 59(6), 70(3), 71(25), 130(2)
3-methyl-2-oxopentionate methyl ester	C ₇ H ₁₂ O ₃	144	997	15(22), 26(2), 27(16), 28(9), 29(47), 39(18), 40(3), 41(87), 42(5), 45(2), 55(7), 56(9), 57(100), 58(5), 59(8), 69(3), 84(4), 85(86), 86(4), 144(3)
4-methyl-2-oxopentionate methyl ester	C ₇ H ₁₂ O ₃	144	998	15(20), 27(10), 28(2), 29(23), 39(19), 40(3), 41(68), 42(17), 43(30), 45(2), 55(3), 57(61), 58(3), 59(10), 69(5), 83(3), 85(100), 86(5), 102(2), 144(4)
2-oxooctonate methyl ester	C ₉ H ₁₆ O ₃	172	1145	15(9), 27(8), 28(3), 29(10), 39(7), 40(1), 41(27), 42(9), 43(100), 44(3), 55(7), 56(2), 57(10), 59(4), 69(2), 85(43), 86(3), 95(6), 113(80), 114(6)
Acetohydroxyacids				
Acetolactate methyl ester	C ₆ H ₁₀ O ₄	146	985	14(2), 15(22), 26(1), 27(4), 28(2), 29(11), 31(10), 39(2), 41(5), 42(6), 43(100), 44(4), 45(18), 55(1), 59(30), 71(2), 87(2), 103(28), 104(5), 146(7)
2-acetohydroxy-2-oxobutyrate methyl ester	C ₇ H ₁₂ O ₄	160	1047	15(17), 27(4), 29(8), 31(3), 39(5), 41(27), 42(7), 43(100), 44(5), 45(42), 55(2), 57(3), 59(18), 69(2), 73(24), 84(1), 116(7), 117(22), 118(5), 132(7), 160(1)
2-ethyl-2-hydroxy-3-oxopentanoate methyl ester	C ₈ H ₁₄ O ₄	174	1099	15(21), 27(18), 28(7), 29(46), 39(11), 41(47), 42(6), 43(15), 45(67), 55(7), 57(100), 58(5), 59(25), 69(5), 73(49), 98(8), 116(12), 117(29), 119(16), 146(5), 174(2)
2-acetohydroxy-2-oxovalerate methyl ester	C ₈ H ₁₄ O ₄	174	1099	15(16), 27(6), 28(2), 29(7), 33(2), 39(6), 41(9), 42(4), 43(89), 44(5), 45(100), 46(2); 55(16), 57(5), 59(13), 73(5), 83(3), 87(3), 131(15), 132(17)
2-acetohydroxy-2-oxooctonate methyl ester	C ₁₁ H ₂₀ O ₄	216	1244	15(7), 17(4), 27(5), 29(9), 39(6), 41(21), 42(5), 43(93), 44(5), 45(32), 55(49), 56(4), 59(15), 69(11), 71(4), 77(5), 88(5), 97(100), 98(9), 132(19), 173(1)



Supplementary Figure 3 GC 70-eV-EI mass spectra of keto acids and additional acetohydroxyacids after ethylation with ECF. **A** Pyruvate ethyl ester, m/z $[M]^+ 116$. **B** 2-oxobutyrate ethyl ester, m/z $[M]^+ 130$. **C** 2-oxovalerate ethyl ester, m/z $[M]^+ 144$. **D** 2-oxoisovalerate ethyl ester m/z $[M]^+ 144$. **E** 3-ethyl-2-oxopentionate ethyl ester, m/z $[M]^+ 158$, **F** 4-methyl-2-oxopentionate ethyl ester, m/z $[M]^+ 158$ **G** 2-oxooctonate ethyl ester, m/z $[M]^+ 186$. **H** 2,2-dihydroxypropionate ethyl ester, m/z $[M]^+ 190$.

Supplementary Table 2 Keto acid and acetohydroxyacid derivatives of the AHAS assay reactions ethylated with ECF. Retention indices (RI) on TG5-SILMS capillary column. When two values are given for RI, they refer to two peaks with almost the same retention time and identical ion peaks in the mass spectrum.

Substance	Formula	[M] ⁺	RI	MS data: characteristic ions m/z (% relative abundance)
Keto acids				
Pyruvate ethyl ester	C ₅ H ₈ O ₃	116	800	14(1), 15(1), 25(1), 26(1), 27(5), 28(1), 29(13), 30(1), 41(1), 42(5), 43(100), 44(3), 45(4), 56(1), 57(1), 59(1), 61(2), 73(1), 91(2), 116(2)
2,2-dihydroxypropionate ethyl ester	C ₉ H ₁₈ O ₄	190	1097	27(2), 29(5), 42(1), 43(23), 44(1), 45(1), 47(1), 61(33), 62(1), 88(1), 89(32), 90(2), 91(1), 117(100), 118(6), 119(1), 129(1), 145(21), 146(2), 147(1)
2-oxobutyrate ethyl ester	C ₆ H ₁₀ O ₃	130	890	15(1), 26(1), 27(9), 28(3), 29(35), 30(1), 41(1), 42(1), 55(1), 56(5), 57(100), 58(4), 59(1), 73(1), 74(1), 75(1), 87(1), 130(3)
2-oxovalerate ethyl ester	C ₅ H ₁₂ O ₃	144	989	26(1), 27(11), 28(2), 29(11), 38(1), 39(6), 40(1), 41(23), 42(8), 43(74), 44(3), 45(1), 55(4), 69(2), 70(6), 71(100), 72(5), 73(1), 88(2), 144(6)
2-oxoisovalerate ethyl ester	C ₅ H ₁₂ O ₃	144	936	27(9), 28(1), 29(6), 39(7), 40(2), 41(28), 42(7), 43(100), 44(3), 55(2), 56(1), 57(1), 69(1), 70(14), 71(85), 72(4), 73(1), 88(1), 89(1), 144(2)
3-methyl-2-oxopentionate ethyl ester	C ₈ H ₁₄ O ₃	158	1043	27(5), 28(2), 29(16), 39(5), 40(1), 41(28), 42(2), 43(1), 45(1), 53(1), 55(5), 56(7), 57(100), 58(5), 69(2), 84(6), 85(47), 86(3), 102(2), 158(1)
4-methyl-2-oxopentionate ethyl ester	C ₈ H ₁₄ O ₃	158	1046	27(6), 29(16), 39(9), 40(2), 41(35), 42(9), 43(14), 45(2), 55(3), 56(2), 57(100), 58(4), 69(6), 83(3), 84(2), 85(95), 86(5), 88(3), 115(2), 158(4)
2-oxooctonate ethyl ester	C ₁₀ H ₁₈ O ₃	186	1312	27(6), 28(2), 29(11), 39(5), 41(18), 42(6), 43(70), 44(3), 55(12), 56(4), 57(23), 68(2), 69(4), 83(2), 84(2), 85(51), 86(3), 95(7), 113(100), 114(8), 186(1)
Acetohydroxyacids				
Acetolactate ethyl ester	C ₇ H ₁₂ O ₄	160	1109 1112	18(3), 27(6), 28(4), 29(15), 31(6), 41(3), 42(4), 43(78), 44(6), 45(100), 46(2), 55(3), 56(1), 71(4), 73(5), 75(2), 88(4), 117(17), 118(3), 160(4)
2-acetohydroxy-2-oxobutyrate ethyl ester	C ₈ H ₁₄ O ₄	174	1132 1134	27(5), 28(3), 29(14), 31(9), 39(3), 41(13), 42(4), 43(62), 44(3), 45(3), 55(2), 57(3), 58(4), 59(100), 60(3), 69(3), 87(2), 130(3), 131(7), 146(4)
2-acetohydroxy-2-oxovalerate ethyl ester	C ₉ H ₁₆ O ₄	188	1225	26(1), 27(10), 28(2), 29(12), 38(1), 39(6), 40(1), 41(23), 42(7), 43(72), 44(3), 45(1), 55(4), 69(2), 70(5), 71(100), 72(5), 73(1), 88(2), 144(5)
Branched-chain amino acids				
D/L Valine ethyl ester	C ₇ H ₁₅ O ₂ N	145	1387	28(1), 29(3), 44(2), 55(11), 56(4), 70(2), 72(14), 74(5), 83(1), 98(13), 100(4), 101(7), 102(9), 116(38), 117(2), 128(1), 129(4), 144(100), 145(8)
D/L Leucine ethyl ester	C ₇ H ₁₅ O ₂ N	159	1461	29(2), 30(4), 43(3), 44(5), 58(6), 72(5), 74(3), 86(5), 101(3), 102(43), 103(2), 112(2), 114(3), 115(2), 116(2), 129(3), 130(2), 142(3), 158(100), 159(9)



Supplementary Figure 4 Comparative GC-MS analysis of ECF ethylated D and L valine and leucine on TG-5SILMS capillary column. **A** stacked total ion chromatograms (TICs) of 5 mM of D-valine and L-valine derivatized with ECF. **B** TIC of 2.5 mM D-valine and L-valine derivatized with ECF in the same reaction. **C** 70-eV-EI mass spectrum of ethylated valine, m/z $[M]^+$ 174 **D** stacked TICs of 5 mM of D-leucine and L-leucine derivatized with ECF. **E** TIC of 2.5 mM D-leucine and L-leucine derivatized with ECF in one reaction. **F** 70-eV-EI mass spectrum of ethylated leucine, m/z $[M]^+$ 188.

3 The role of an acetohydroxyacid synthase in the biosynthesis of lycopsamine type pyrrolizidine alkaloids in *Symphytum officinale* (Boraginaceae)

The role of an acetohydroxyacid synthase in the biosynthesis of lycopsamine type pyrrolizidine alkaloids in *Symphytum officinale* (Boraginaceae)

Annika Engelhardt^a, Britta Muhs^a, Sebastian Fritz^a, Dorothee Langel^a, and Dietrich Ober^a

^a Botanisches Institut und Botanischer Garten, Christian-Albrechts-Universität zu Kiel, D-24098 Kiel, Germany

Detailed author contributions are listed at the end of the thesis.

3.1 Abstract

Pyrrolizidine alkaloids (PAs) are typical specialized plant metabolites involved in plant defense. The backbone of PAs consists of the characteristic bicyclic necine base that is esterified with one or more necic acids. The great structural diversity of the PAs with more than 600 different structures has its origin mainly in the diversity of the necic acids. Lycopsamine type PAs are characterized by a branched chain aliphatic C₇-necic acid moiety that is uniquely found in this group of PAs. Feeding experiments suggested that a C₂-moiety is transferred from hydroxyethyl thiamine diphosphate to 2-oxoisovalerate (2OIV) by an acetohydroxyacid synthase (AHAS, EC 2.2.1.6)-like enzyme to generate the C₇-necic acid backbone. AHASs are known to catalyze the transfer of a C₂-moiety to pyruvate (PYR) or 2-oxobutyrate (2OB) early in the branched-chain amino acid biosynthesis. AHASs are heteromers composed of catalytic subunits (CSUs) that are responsible for the biochemical reaction and regulatory subunits (RSUs) that are responsible for feedback inhibition.

Here we present the identification and characterization of the AHAS CSU that was recruited for PA biosynthesis in *Symphytum officinale* (*S. officinale*, Boraginaceae). In this plant, two CSUs and one RSU of AHAS are known. The most likely candidate to be involved in PA biosynthesis was identified by RT-qPCR and named C₇-hydroxyacid synthase (*SoC₇AS*). CRISPR/Cas9-mediated knock-out hairy roots provided *in planta* evidence that this candidate is indeed involved in the formation of the C₇-necic acid moiety. We could show that hairy roots with a genotypic knock-out of *soc₇has* are not able to produce lycopsamine type PAs and instead accumulate the saturated necine base trachelanthamidine. Regeneration of whole plants from the knock-out hairy roots and the analysis of the inflorescences showed that there is only one C₇HAS in both biosynthetic sites of PAs in *S. officinale*. Subsequently, the *SoC₇HAS* was heterologously expressed in *E. coli* and characterized with respect to its biochemical properties using a GC-MS based assay method. We could show that *SoC₇HAS* is able to convert the substrates of primary metabolism, *i.e.*, PYR and 2OB. Additionally, *SoC₇HAS* is accepts branched-chain and larger long chain ketoacids as acceptor molecules for the C₂-moiety. This includes the reaction involved in necic acid biosynthesis, *i.e.*, the formation of acetohydroxy-2-oxoisovaleric acid, or C₇-pronecic acid, the direct precursor of the C₇-necic acids (+)-trachelanthic and (-)-viridifloric acid from 2OIV and the activated acetaldehyde.

3.2 Introduction

Pyrrolizidine alkaloids (PAs) are typical secondary plant metabolites (also called “specialized metabolites”) involved in plant defense. More than 600 different structures have been described from plant species of various lineages. The backbone of PAs consists of a necine base, the characteristic bicyclic structure common to all PAs that is esterified with one or more necic acids. The great structural diversity of the PAs has its origin mainly in the diversity of the necic acids that are derived from various amino acid precursors, *i.e.*; branched-chain amino acids in the case of aliphatic necic acids (Crout, 1966; Crout et al., 1966; Crout, 1967; Crout et al., 1970; Robins et al., 1974; O’Donovan and Long, 1975; Bale et al., 1978; Hartmann and Witte, 1995). The necic acids are further diversified as part of transformations that occur within the plant and that result in the often unique species-specific PA patterns (Hartmann and Dierich, 1998). Comfrey (*Symphytum officinale*, Boraginaceae) does contain PAs of the lycopsamine type, a class of PAs typically found in species belonging to the Boraginaceae, Apocynaceae, and Eupatorieae (Asteraceae). Lycopsamine type PAs are characterized by at least one branched chain aliphatic C₇-necic acid moiety with the carbon skeleton of 2-isopropylbutyric acid that is uniquely found in this group of PAs (Hartmann and Witte, 1995). (+)-Trachelanthic acid and (-)-viridifloric acid are the most abundant necic acids of this type (Hartmann and Witte, 1995). The biosynthesis of the necine base moiety was intensively studied over the past decades and resulted in the comprehensive biochemical and evolutionary characterization of homospermidine synthase (HSS, EC 2.5.1.45, (Ober and Hartmann, 1999; Ober et al., 2003; Reimann et al., 2004; Kaltenegger et al., 2013; Irmer et al., 2015; Livshultz et al., 2018). Immunolabeling analyses of HSS demonstrated a different site of synthesis at the organ, tissue, or cell level in almost every species or family (Moll et al., 2002; Anke et al., 2004; Frölich et al., 2007; Anke et al., 2008; Niemüller et al., 2012; Irmer et al., 2015; Kruse et al., 2017). In *Symphytum officinale*, two biosynthetic sites could be identified. The constitutive biosynthesis in the endodermis of the roots (Niemüller et al., 2012) and an additional biosynthesis in the bundle sheath cells of young leaves subtending the developing inflorescences (Kruse et al., 2017). Most recently, the identification of homospermidine oxidase (HSO, EC 1.4.3.22 (Zakaria et al., 2022)) elucidated the second step in the necine base formation. On the other hand, our knowledge about the biosynthesis of the necic acids is scanty. Early feeding experiments in root cultures of *Cynoglossum officinale* indicated that valine or its biosynthetic precursor, 2-oxoisovalerate (2OIV), are incorporated into PAs, with an additional two carbon moiety integrated as an activated acetaldehyde (Crout, 1966). Later feeding experiments with [¹³C₆] glucose using root cultures of *Eupatorium clematideum* (Asteraceae) that produce the

lycopsamine type PA trachelanthamine, allowed analyses of the ^{13}C ^{13}C coupling within the necic acid moiety of the PA molecule, *i.e.*, (+)-trachelanthic acid. Comparison of this coupling pattern to that of valine revealed that five of the seven carbon atoms of the necic acid show an identical ^{13}C -labelling pattern as that of valine. The remaining two carbon moiety of the necic acid, trachelanthic acid, showed an identical ^{13}C labelling pattern as the two carbon moieties in valine that are derived from hydroxyethyl thiamine diphosphate (HeThDP, 'activated acetaldehyde') that is transferred early in this pathway to pyruvate (PYR) by acetohydroxyacid synthase (AHAS, 'acetolactate synthase', EC 2.2.1.6) (Weber et al., 1999). These observations resulted in the hypothesis that firstly, (+)-trachelanthic acid and valine share the early steps in the biosynthesis until the intermediate 2OIV in the same cell compartment and, secondly, the biosynthesis of this C_7 -necic acid requires a second transfer of a two-carbon moiety from HeThDP to 2OIV (Weber et al., 1999). It has been proposed that an AHAS-like enzyme catalyzes this second transfer of the two-carbon residue to 2OIV resulting in an C_7 -intermediate (the " C_7 -pronecic acid") that is reduced to result in the formation of the necic acid (+)-trachelanthic acid (Weber et al., 1999). AHAS is the first enzyme in the biosynthesis of the branched-chain amino acids (BCAAs) valine, leucine, and isoleucine (Umberger and Brown, 1958). As shown in Figure 12 A and B, this reaction proceeds in two steps: first, the donor PYR is decarboxylated and the reactive intermediate HeThDP is formed. Second, the acceptor substrate, PYR in the biosynthesis of valine and leucine and 2-oxobutyrate (2OB) in the biosynthesis of isoleucine, reacts with the HeThDP to form 2-acetolactate (AL) or 2-acethydroxy-2-oxobutyrate (AHOB), respectively (Duggleby and Pang, 2000). These reactions are catalyzed in the catalytic subunits (CSU) of the holoenzyme and requires the co-factors flavin adenine dinucleotide (FAD), magnesium ions and ThDP (Duggleby and Pang, 2000). To form fully functional active sites, it is essential that the CSUs dimerize (Lee and Duggleby, 2001; Pang et al., 2002; Lonhienne et al., 2020). Thereby, the two active sites of the dimer work synchronized complementary to each other (Belenky et al., 2012; Lonhienne et al., 2017; Lonhienne et al., 2020), while in one active site the decarboxylation of PYR occurs, in the other active site the carbon-carbon bond is formed and vice versa (Lonhienne et al., 2017). In addition to CSUs, the holoenzyme also consists of regulatory subunits (RSUs) that mediate feedback inhibition of the pathway by BCAAs (Mifflin, 1971; Durner and Böger, 1988; Singh et al., 1988; Southan and Copeland, 1996; Pang et al., 2002). Additionally, RSUs stabilize the complex of the holoenzyme and thereby mediate the activity of the AHAS by improving the connection of the active sites of the CSU dimers (Lonhienne et al., 2020). AHAS has two substrate-binding pockets, where pocket 1 is quite specific for PYR but pocket 2 can accept a

variety of keto acids and even unnatural aldehydes (Engel et al., 2003; Tittmann et al., 2005). If an AHAS is involved in the biosynthesis of C₇-necic acids, it should accept not only PYR and 2OB, the substrates for the biosynthesis of valine and leucine, but also 2OIV as acceptor of the activated two carbon unit (Figure 12 C).

Here we present the identification of an AHAS CSU duplicate, the C₇-hydroxyacid synthase (C₇HAS), to be the first-specific enzyme in the biosynthesis of C₇-necic acids of lycopsamine type PAs in *Symphytum officinale* (*S. officinale*, Boraginaceae). This enzyme is co-transcribed with HSS in the roots and young leaves subtending the inflorescences. *In planta* evidence for the involvement in PA biosynthesis is provided by CRISPR/Cas9-mediated genome editing in *S. officinale* targeting the C₇HAS that resulted in hairy roots free of lycopsamine type PAs whilst accumulating the saturated necine base trachelanthamidine. Regeneration of whole plants from knock-out hairy roots demonstrated that trachelanthamidine is not translocated from roots to leaves as is the case with lycopsamine type PAs in the control lines. Analysis of the inflorescences of the knock-out plants showed that one C₇HAS catalyzes the necic acid formation at both biosynthetic sites of PAs in *S. officinale*. Biochemical characterization revealed that the C₇HAS accepts a wider spectrum of substrates than the SoCSU of AHAS, including those of the predicted reaction of 2OIV and PYR to the C₇-pronecic acid, the direct precursor of the necic acids (+)-trachelanthic acid and (-)-viridifloric acid.

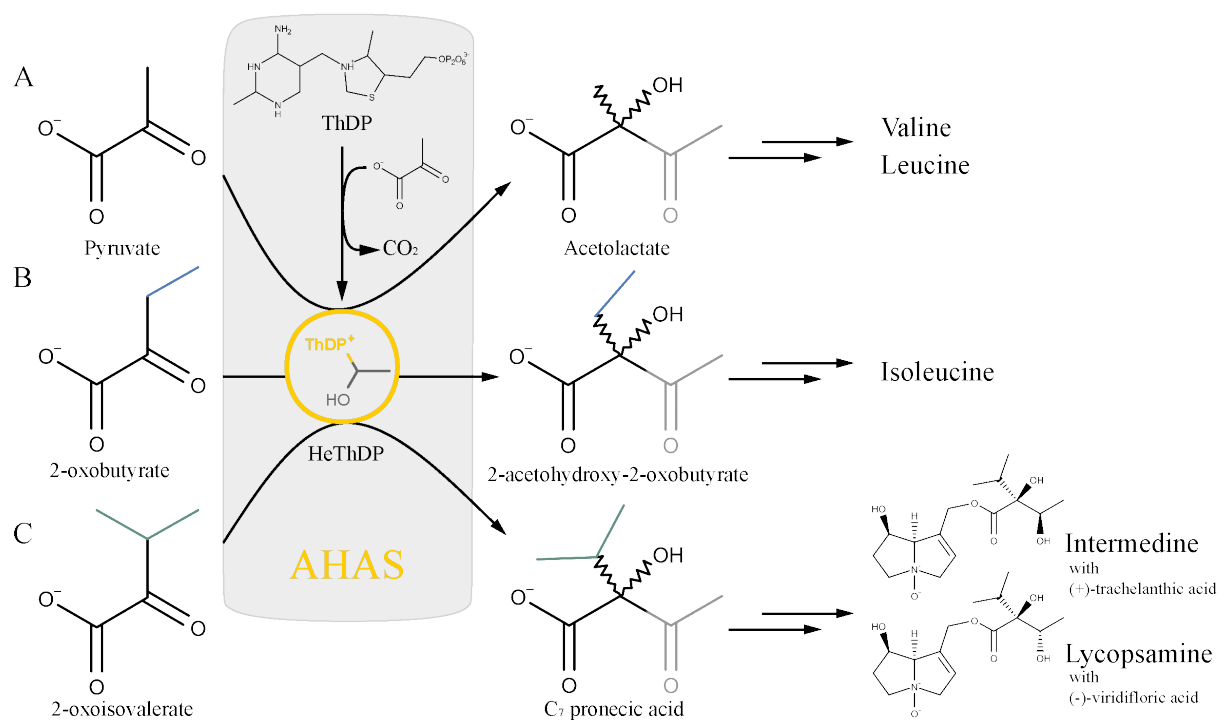


Figure 12 Reactions catalyzed by AHAS. **A** Formation of acetolactate from two molecules of pyruvate in the biosynthesis of valine and leucine. **B** Formation of 2-acetohydroxy-2-oxobutyrate from 2-oxobutyrate and pyruvate in the biosynthesis of isoleucine. **C** Formation of the C_7 moiety (pronecic acid) as a precursor of (+)-trachelanthic and (-)-viridifloric acid as part of lycopsamine type PAs as postulated by Weber (1999). Activation of pyruvate by thiamine diphosphate (ThDP) bound in the catalytic center of the AHAS to form hydroxyethyl ThDP (HeThDP). Transfer of the active C_2 -unit to the various keto acids leads to the formation of the associated acetohydroxyacids. Via intermediate steps the final products valine, leucine and isoleucine in primary metabolism and to the lycopsamine type PAs in secondary metabolism are formed.

3.3 Results

3.3.1 Tissue-specific transcript quantification suggests the gene encoding *SoC₇HAS* to be co-expressed with *SoHSS*

In *S. officinale* two CSU-like sequences (*socsu* and *socsu2*, from here on referred to as C₇-hydroxyacid synthase, *soc₇has*), and one RSU (*sorsu*) of an AHAS have been identified by a reverse genetic approach (Langel, 2008). To identify the subunits involved in PA biosynthesis, we used reverse-transcription quantitative PCR (RT-qPCR) to quantify the transcript level of the genes encoding for the AHAS subunits in various tissues in comparison with that of *SoHSS*. *SoHSS* is expressed in the endodermis of young roots (Niemüller et al., 2012) and in bundle sheath cells of young leaves subtending an inflorescence (Kruse et al., 2017). Therefore, candidates involved in the biosynthesis of the PA backbone should show a transcript profile similar to that of *SoHSS*. Or, if two paralogs are involved in PA biosynthesis, each specific for one of the two PA-producing sites, transcript occurrence should be restricted to one of these tissues.

Total RNA of roots, stems, young leaves, old leaves and the inflorescence of *S. officinale* was extracted and reverse transcribed into cDNA. The transcripts of *sohss*, *socsu*, *soc₇has*, and *sorsu* were quantified by RT-qPCR using the gene encoding elongation factor 1- α (*efl α*) as reference. The relative transcript levels of the genes of interest normalized to that of *efl α* are shown in Figure 13. *sohss* transcripts (A) were detected in roots and young leaves, confirming the earlier immunolabeling and tracer feeding experiments (Niemüller et al., 2012; Kruse et al., 2017). Traces of *sohss* transcripts were also detected in inflorescences. Transcripts of *sorsu* (B) were detected in all analyzed tissues, except the roots. While transcripts of *socsu* (C) could be detected in all tissues examined, *soc₇has* transcript (D) showed the same transcriptional pattern as *sohss* in roots, the young leaf and the inflorescence. Consequently, the *SoC₇HAS* was identified as candidate for the PA biosynthesis. *SoCSU* was included in the further analyses for comparative reasons as a candidate for BCAA biosynthesis. Of note, transcript levels of *socsu*, *soc₇has* and *sorsu* have been remarkably lower than that observed for that of *sohss*.

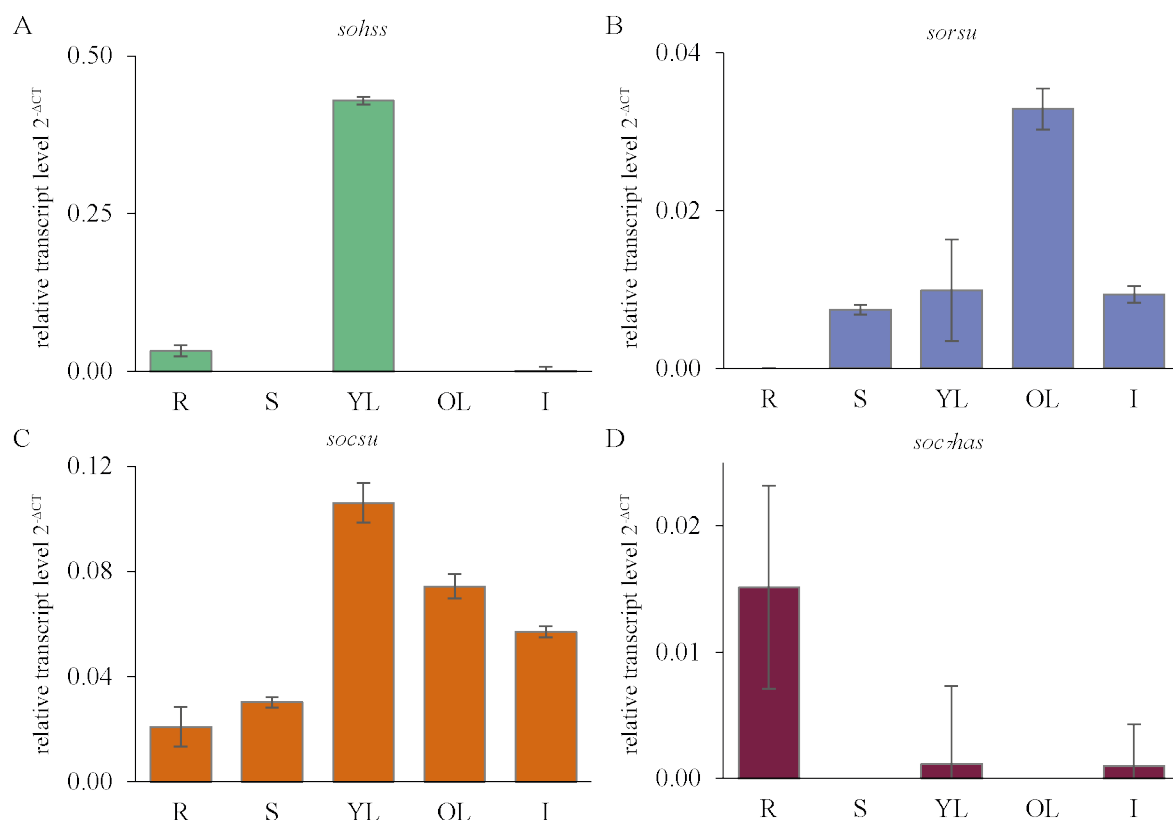


Figure 13 Tissue-specific analysis of relative transcript levels of genes of interest in different tissues of *S. officinale*. A ■ *sohss*, B ■ *sorsu* C ■ *socsu* and D ■ *soc7has* transcripts in roots (R), stems (S), young leaves (YL), old leaves (OL), and inflorescences (I.) were quantified using RT-qPCR. The transcript levels of *efla* served as reference and were set to 1 (= relative transcript level 2^{-ΔCT}).

3.3.2 CRISPR/Cas9 knock-outs of *soc7has* result in the inability of hairy roots to produce PAs

To test if the candidate *SoC₇HAS* is involved in PA biosynthesis, the gene encoding *SoC₇HAS* was inactivated by a genome editing approach using clustered regulatory interspaced short palindromic repeats (CRISPR) with CRISPR associated protein 9 (CRISPR/Cas9). As the roots are a site of PA biosynthesis in *S. officinale* we used hairy roots resulting from transformation with the hairy root-inducing *Agrobacterium rhizogenes* (*A. rhizogenes*). Three expression constructs for the short guide RNA (sgRNA) targeting the *soc7has* gene were created (constructs A, B, and C) and cloned into the destination vector containing the Cas9 and hygromycin resistance encoding genes. After transformation into *A. rhizogenes* these were used for infection of *S. officinale* leaf sections. Transformed lines were selected by hygromycin resistance and analyzed for mutations in the *soc7has* gene, and for an effect on PA levels in comparison to control lines that resulted from infection with the empty destination vector. The

identification and quantification of the PAs synthesized by the hairy roots was performed by GC-MS.

Figure 14 A shows the total ion chromatogram of purified extracts of a selected control line where 17 different PAs could be detected. Their identification is summarized in Figure 14 B and was done by analyzing the extracts on a DB-1 capillary column, calculating the retention indices (RIs) of the compounds and compare them and the mass spectra with literature values. Doubtful cases are marked. All identified PAs could be classified to the lycopsamine type with 7-acetylintermedine and intermedine being the predominant PAs found. In addition, lycopsamine and 7-acetyllycopsamine, 3'-acetyllycopsamine and 7,3'-diacetylintermedine or -lycopsamine were detected in small amounts. The PAs numbered 10-13 are most likely echiumine, symlandine, myoscorpine, and symphytine which could be identified based on their mass spectra. Measurements of (Roeder et al., 1991) allow us to assume that they elute from the column in the order given here. Compounds numbered 14-17 were identified from their mass spectra as the respective 3'-acetyl derivatives of echiumine, symlandine, myoscorpine, and symphytine assuming the same elution order. Detailed MS data with characteristic ions m/z and their relative abundance is given in Supplementary Table 4.

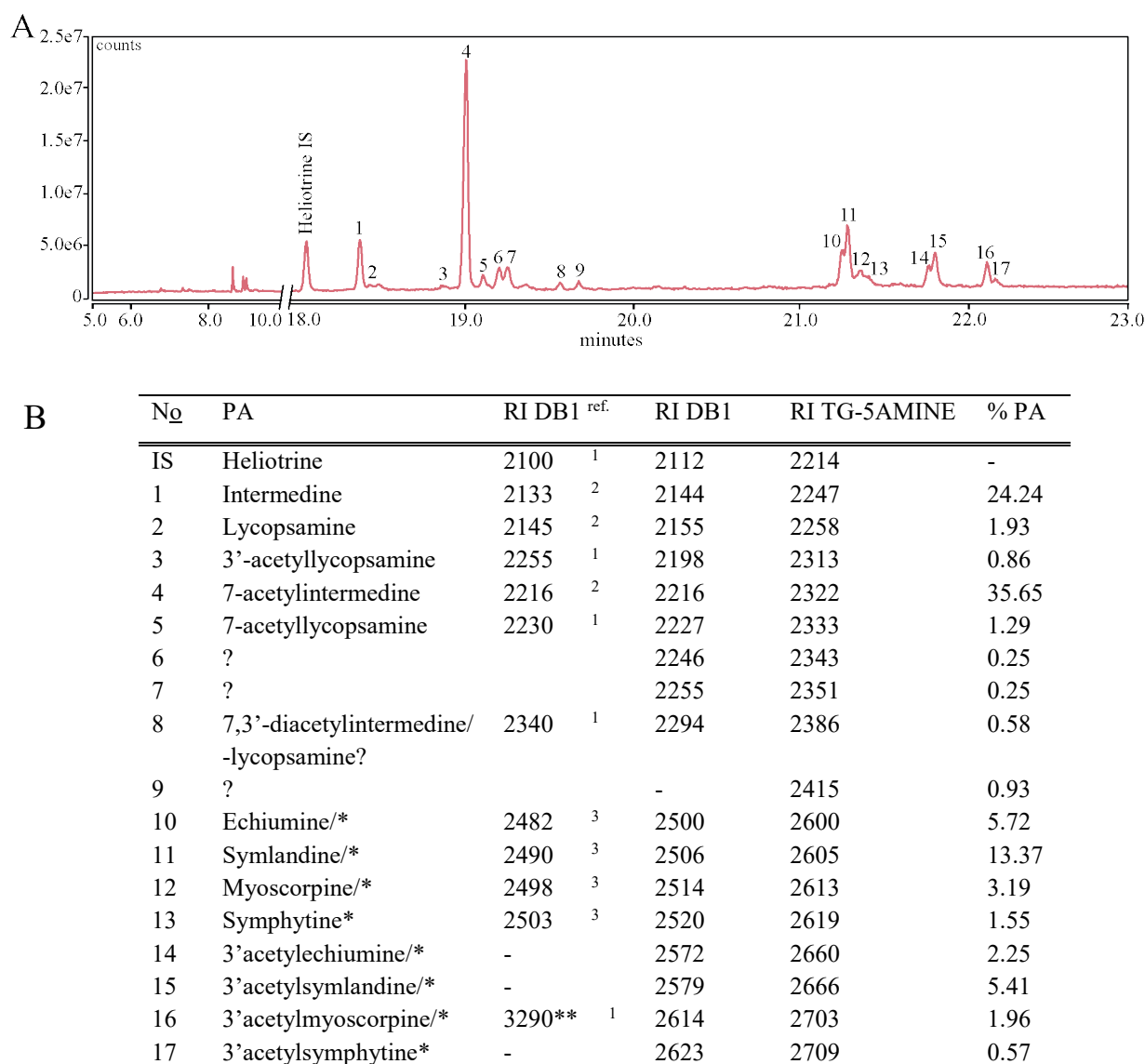


Figure 14 Chemotyping of transgenic HR lines. **A** Total ion chromatogram of purified extract of control line CL5 hairy root on TG-5SILMS capillary column: control line (CL5), **B**: identified PAs, No 1-17 as labelled in the total ion chromatogram. % PA = Percentage of PA in total PA content. ? = unidentified PA; * elution sequence of the stereoisomers is not confirmed; ** RI for DB-1701 capillary column; - no data available. References: ¹ (El-Shazly and Wink, 2014); ²(Fröhlich et al., 2007), ³ (Haberer et al., 2002).

The PA content of 40 hairy roots resulting from transformation with construct A, 31 from construct B, and 48 from construct C were analyzed. While no line resulting from transformation with construct B showed any effects on PA levels (data not shown), 15 lines resulting from construct A and 3 resulting from construct C showed significantly lower PA levels in comparison to the 6 CL or were PA free (Figure 15, red bars). Instead, in knock-out and knock-down lines an accumulating intermediate was identified via its mass spectrum as a saturated necine base. An RI of 1246 indicates (+)-trachelanthamidine as the most likely candidate (Figure 15, blue bars). To demonstrate a correlation between altered PA content,

accumulating trachelanthamidine and CRISPR/Cas9-mediated inactivation of the *soc7has* gene in the analyzed hairy roots, the targeted region in the genomic DNA was amplified, sequenced, and compared to that of the *soc7has* gene of the wilt type plant (WT) used for transformation. In Figure 15 the observed modifications within the genomic DNA are listed for the hairy root lines with modified PA levels. Even though *S. officinale* is a diploid organism, up to five different gene variants could be identified in certain lines, an observation that might be explained by the fact that the CRISPR/Cas9 system can continuously target and cleave genes even after the first division of the initially infected cell. As a consequence, the CRISPR/Cas9-mediated effects on the two alleles in the individual cells might be quite different, resulting in developing root, in which the genotype of the individual cells is not identical, at least as the *soc7has* gene is concerned, a phenomenon referred to as ‘genetic mosaicism’ (Mehravari et al., 2019). Thus, the gene variants identified as different PCR products might not represent all variants present in these lines but are sufficient to confirm a CRISPR/Cas9-mediated effect on the *soc7has* gene. The mutations detected in the *soc7has* gene varied from replacements of single nucleotides at the CRISPR/Cas9 site to insertions and deletions of a few bases up to deletions of more than 400 bp and insertions of up to 171 bp. Almost all of these mutations resulted in frameshift mutations that should result in potentially inactive *SoC₇HAS*. Exceptions are the 3 bp-deletions which, depending on their position in the sequence, should have an effect on a maximum of two amino acids. In several hairy root lines WT alleles have been detected in addition to mutated alleles. In these lines lower PA levels than the CL were expected (expected effect = knock-down, KD). If no WT variants were detected and all mutations resulted in a frame shift, no PAs should be detectable, if *SoC₇HAS* is involved in PA biosynthesis (expected effect = knock-out, KO). A detailed overview of all gene variants found is given in Supplementary Table 5. In the majority of the lines studied, the predicted and observed effect coincided. For lines A29, A37, and A50, the deletion of 3 bp and therefore the alteration of 1-2 amino acids appear to have a greater effect on enzyme activity than predicted, *i.e.* these lines do not produce any lycopsamine type PAs (Figure 15, green arrows). In contrast, in lines A12 and C106, a KO was predicted and a KD was observed (Figure 15, yellow arrows). In these lines, it is likely that not all gene variants were identified and that an unmodified gene variant exists in these lines. Overall, it can be clearly seen that the CRISPR/Cas9 complex with the sgRNA of construct A and C have successfully cleaved the *soc7has* and that the resulting changes in the nucleotide and amino acid sequence have an impact on the production of PAs. That both sgRNAs targeting different regions within the *soc7has* gene have the same effect on PA levels, an off-target effect is most unlikely to be responsible for this phenotype. In

conclusion, the absence of the complex lycopsamine type PAs and the accumulation of trachelanthamidine in the KD and KO hairy roots lines in combination with the mutations on the *soc7has* gene provide convincing evidence that the SoC₇HAS is involved in the biosynthesis of the necic acid.

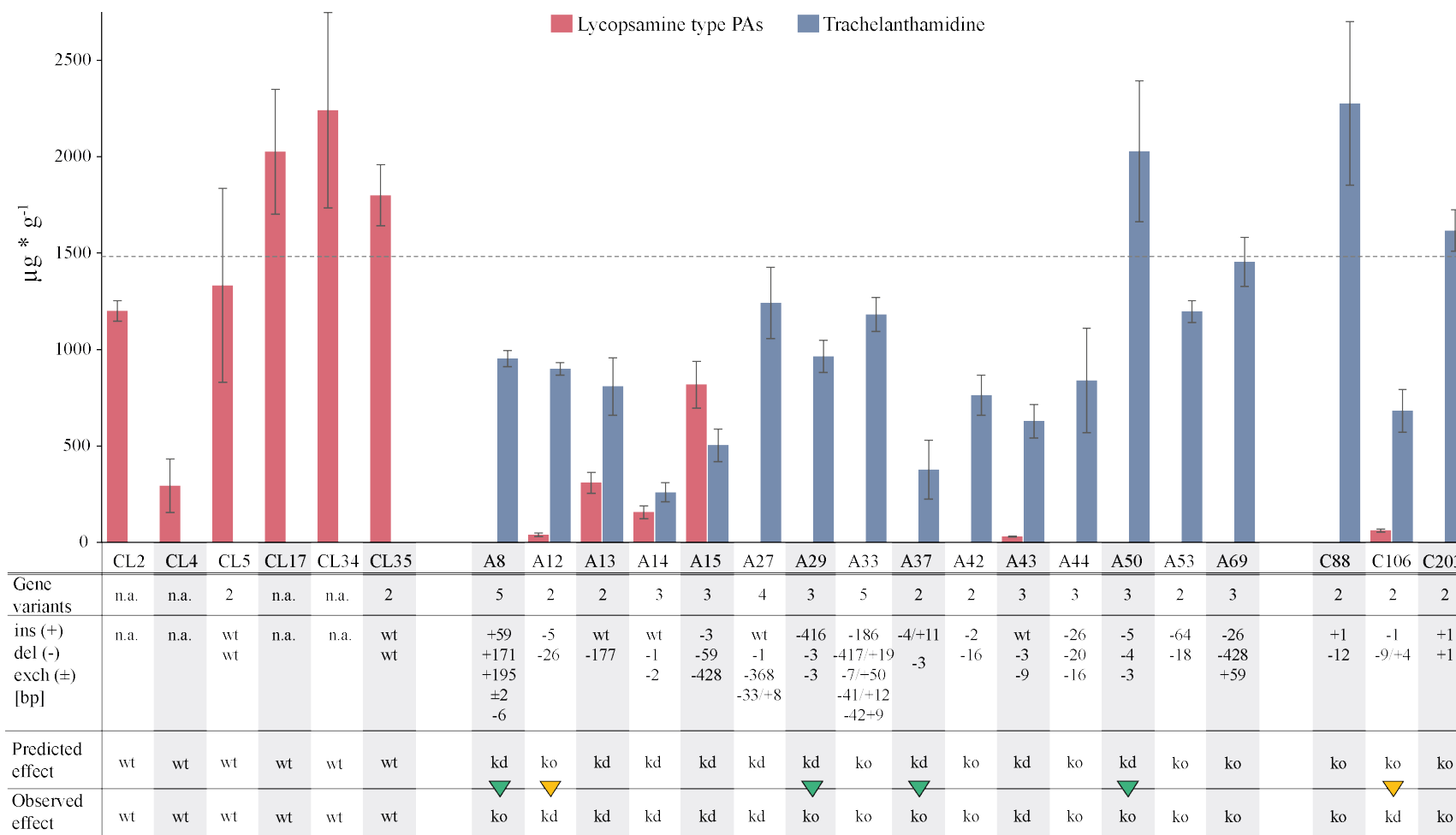


Figure 15 Genotyping, PA content and accumulation of (-)-trachelanthamidine in different hairy root lines. Six control lines (CL) resulting from infection with the empty destination vector, 15 lines resulting from infection with construct A, and 3 lines resulting from infection with construct C, both targeting the *soc7has* gene were analyzed. PA content and amount of accumulated trachelanthamidine are given in µg/g dry weight (n=3) and are calculated as heliotrine equivalents. The dashed line represents the average PA content of the five CLs. The modifications observed within the *soc7has* gene by genotyping are indicated by ‘+’ for insertions, ‘-’ for deletions, and ‘±’ for base substitutions. If different mutations have occurred simultaneously on the different alleles, they are separated by ‘/’. The predicted effect on the PA content is based on the genotyping results. The observed effect is based on the PA analyses. wt = *soc7has* sequence identical to wildtype; kd = knock-down, ko = knock-out. Discrepancies in prediction and observation are marked by colored arrowheads. ▼ indicates that according to PA analyses the changes in the gene resulted in the inability to produce complex PAs to lead to ko after all. ▼ indicates that despite substantial mutations in the gene that should have resulted in lines unable to produce complex PAs, PAs could still be detected.

3.3.3 No transport of trachelanthamidine intermediate within regenerated plants

In *S. officinale*, two sites of PA biosynthesis are known: the endodermis of the primary roots and the bundle sheath cells of the leaves below the inflorescences. Both are known to possess the complete enzyme apparatus for the synthesis of lycopsamine type PAs. The resulting PAs are subsequently distributed throughout the plant organs. Thus, it is assumed that the transporters specifically transport only the complex PAs and no early intermediates of the metabolism (Hartmann and Dierich, 1998). To test this hypothesis, we regenerated plants without inflorescences from one hairy root control line (CL5) and five hairy root *soc7has* gene KO lines (lines 29, 37, and 69 resulting from construct A and lines 88 and 203 from construct C) and screened the roots and leaves for PAs. As shown in Figure 16, in CL5 lycopsamine type PAs were detectable in both roots and leaves, confirming the transport of complex PAs within the plant. No trachelanthamidine was detected in either tissue. In the plants regenerated from the KO lines on the other hand no lycopsamine type PAs were detectable and trachelanthamidine was found exclusively in the roots and not in the leaves. These observations show that trachelanthamidine, as an intermediate of PA biosynthesis, is not transported from the roots into the shoot of the plant.

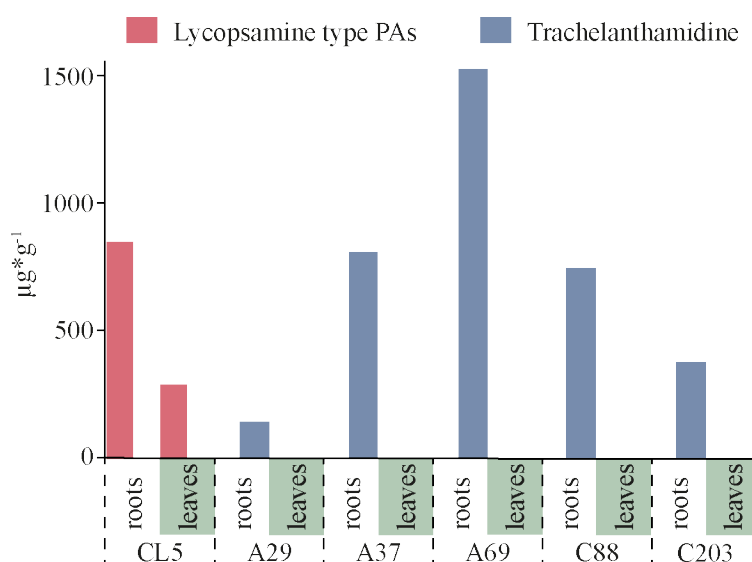


Figure 16 PA analysis of plants regenerated from hairy root control line CL5 and *soc7has* knock-out lines A29, A37, A69, C88, C203. PA content and amount of accumulated trachelanthamidine are given in µg/g dry weight and are calculated as heliotrine equivalents.

3.3.4 Second side – same enzyme

The two biosynthetic sites in *S. officinale* raise another interesting question: (i) is just one C₇HAS responsible for the formation of the C₇-necic acid moiety in both sites or (ii) are two

paralogs involved in PA biosynthesis, each specific for one of the two PA-producing sites. To test this, flower formation was induced in regenerated plants by cold treatment. The inflorescences were collected at different stages of development from flower buds (stage I) until all single flowers were fully bloomed (stage IV) (Figure 17 A). An old leaf of the rosette plant, the young leaves subtending the individual inflorescences, the flowers of each inflorescence, and, as a positive control, the roots of the plant at developmental stage IV were analyzed for their PA content. Without derivatization, the signal to noise ratio of the analytes was too low for peak integration and retention time determination due to the small amounts. Therefore, the H-acidic functional groups of the analytes were silylated with N-methyl-N-(trimethylsilyl)trifluoroacetamide, forming trimethylsilyl (TMS) esters. In the GC-MS analysis, no lycopsamine type PAs could be detected in any organs at any developmental stage. Figure 17 B shows that the roots accumulated trachelanthamidine that is not transported into the leaves of the rosette plant. In development stages I and II, trachelanthamidine was also not detectable in the organs of the inflorescences. However, the young leaves at stage III and the flowers in stages III and IV contained 34, 12, and 13 µg/g of trachelanthamidine respectively. In stage IV there was no trachelanthamidine found in the young leaves subtending the inflorescences, which indicates that at this stage PA biosynthesis was already switched off again and all alkaloids were transported into the flowers. Two conclusions can be drawn from these results. First, at both sites of synthesis of the PAs in *S. officinale* the same C₇HAS initiates the necic acid biosynthesis. And second, unlike in roots, in inflorescences there is a translocation of trachelanthamidine from the site of synthesis in the young leaves into the flowers.

Of note: It should be mentioned that after silylation of the analytes, traces of a compound were detectable in all leaves, which could be assigned to the mono- or di-silylated platynecine (C₇-hydroxylated trachelanthamidine) based on its MS signature. We currently have no explanation for this observation and it would need to be clarified whether it is indeed platynecine or a case of mistaken identity.

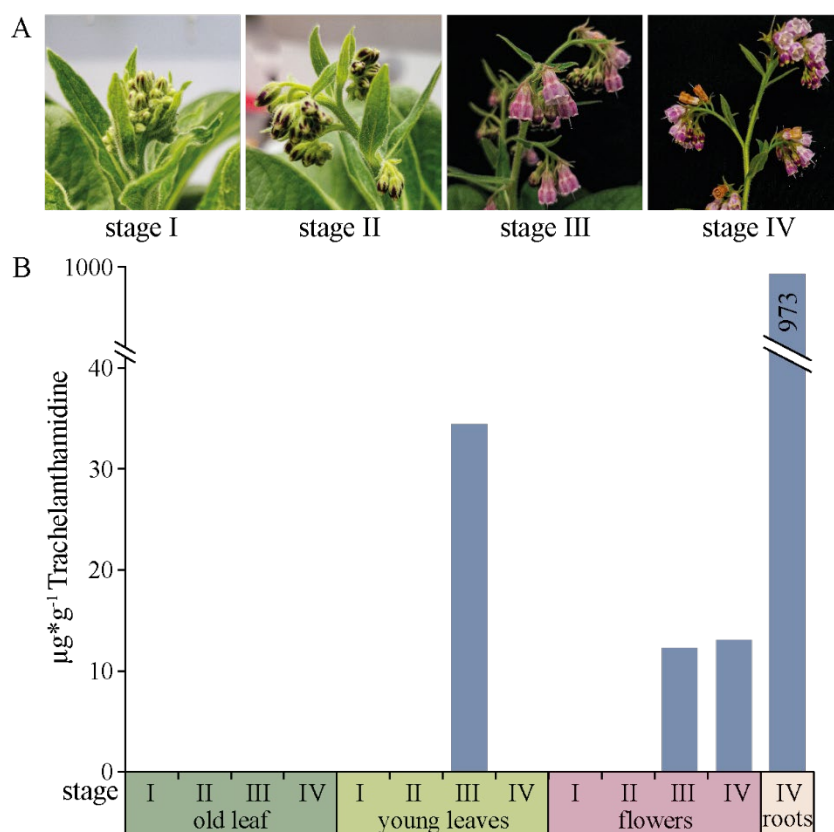


Figure 17 PA analysis of flowering plant regenerated from *soc7has* knock-out line A69. **A** Different developmental stages of the collected inflorescences (I) closed flower buds, (II) opening flower buds, (III) some flower buds open, others still closed, (IV) fully bloomed inflorescence. **B** Extracts of an old leaf of the rosette plant, the young leaves subtending the individual inflorescences, and the flowers of each inflorescence and the roots of the plants at developmental stage IV were trimethyl silylated with N-Methyl-N-(trimethylsilyl)trifluoroacetamid and measured by GC-MS. The trachelanthamidine content is given in $\mu\text{g}\cdot\text{g}^{-1}$ dry weight and is calculated as heliotrine equivalents.

3.3.5 Heterologous expression and activity assay of *SoC7HAS*

Encouraged by the results of the CRISPR/Cas9 knock-out experiments confirming an involvement of *SoC7HAS* in the biosynthesis of the necic acid, *in vitro* experiments were used to establish differences on the biochemical level between the *SoC7HAS* and the previously described *SoCSU* involved in BCAA biosynthesis (Chapter 2, this work). Therefore, the ORF of the *SoC7HAS* without the predicted *N*-terminal chloroplast targeting sequence was amplified using the primer pair 19/20, and cloned into a modified pCDFDuet1TM expression vector where the *N*-terminal 6xHis-tag was replaced by a Strep-tag as described in Chapter 2. After heterologous expression in *E. coli* the SDS-PAGE analyses of affinity-purified *SoC7HAS* confirmed the expression of a protein of about 64 kDa (Figure 18) that showed activity in a colorimetric AHAS assay with PYR as substrate indicating the formation of an AL.

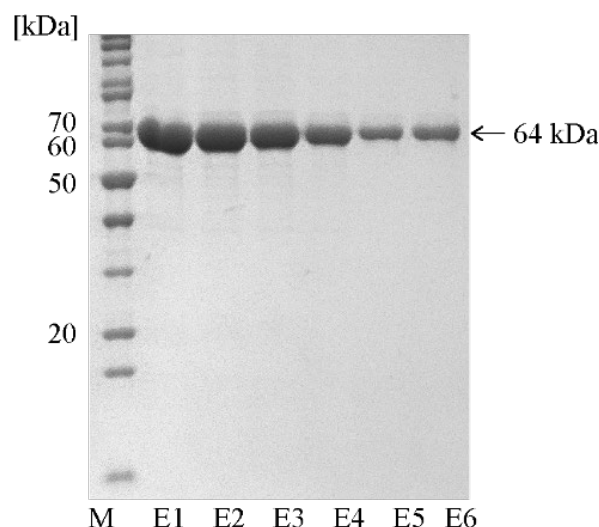


Figure 18 SDS-PAGE of affinity purification of *SoC7HAS* catalytic subunit (12% gel, coomassie blue stained). M= marker (PageRuler™ 10-200 kDa), E= elution fractions. The purified enzyme shows a distinct band between 60 and 70 kDa.

3.3.6 *SoC7HAS* catalyzes the formation of C₇ pronecic acid

SoC7HAS was identified as an AHAS homolog which has been shown to be involved in PA biosynthesis. According to our hypothesis this AHAS should have a modified substrate specificity that allows the conversion of PYR and 2OIV, a substrate with a branched side chain, to C₇-pronecic acid (acetohydroxy-2-oxoisovalerate) (Figure 12C). To test this, the recombinant *SoC7HAS* was incubated in reaction mixtures containing a total of 6 mM substrate; either 6 mM PYR as donor and acceptor or 1 mM PYR as donor and 5 mM acceptor substrate, *i.e.*, 2OB or 2OIV and stopped after 30 min by the addition of 5% NaOH. After derivatization with ECF the samples were analyzed by GC-MS. Figure 19 A and B show that the *SoC7HAS* exhibited AHAS activity catalyzing the formation of AL from two molecules of PYR and AHOB from PYR and 2OB. But, more importantly, the *SoC7HAS* also catalyzes the formation of acetohydroxy-2-oxoisovaleric acid, or C₇-pronecic acid, the direct precursor of the C₇-necic acids (+)-trachelanthic and (-)-viridifloric acid (Figure 19 C and D, detailed MS data with characteristic ions *m/z* and their relative abundance is given in Supplementary Table 6). Because of this enzymatic property and its role in PA biosynthesis, we decided to name this enzyme C₇-hydroxyacid synthase, or C₇HAS for short.

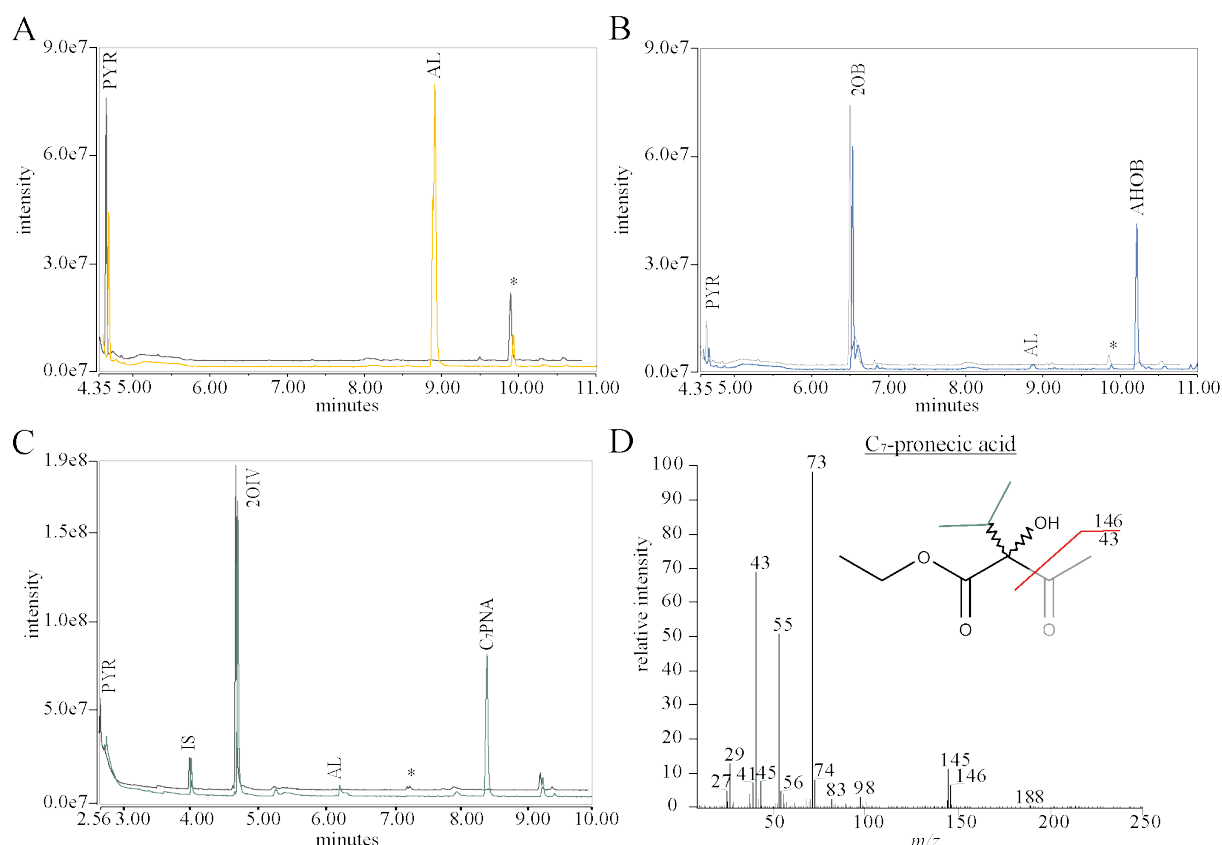


Figure 19 GC-MS total ion chromatograms of reactions catalyzed by *SoC₇HAS* with various substrates after derivatization with ECF. Chromatograms resulting from active enzyme (incubation time 30 min) are marked in color, the comparative chromatogram with heat-inactivated enzyme are marked in grey in the background. All metabolites are present as ethyl esters. **A** Substrate: PYR [6 mM]. **B** Substrates: PYR [1 mM] and 2OB [5 mM]. **C** Substrates: PYR [1 mM] and 2OIV [5 mM]. 2OB was added as internal standard (IS). **D** 70-eV-EI spectrum of ethylated C₇-pronecic acid (C₇PNA). Red lines indicate postulated fragmentation with *m/z* values that agree with the observed fragment ions. * 2,2-dihydroxypropionic acid, impurity of PYR.

3.3.7 Kinetic Parameters of *SoC₇HAS*

To characterize the ability of affinity purified *SoC₇HAS* to convert 2OIV acid in more detail, the kinetics of the conversion of this substrate together with PYR was compared with that of the conversion of PYR as sole substrate. Using only PYR as substrate concentrations from 0.5 to 20 mM have been used in assays with 4 min incubation time. When plotting the specific activity against the substrate concentration the curve roughly follows that of a Michaelis-Menten kinetic (Figure 20 A). A K_M of 1.8 mM was estimated by the double reciprocal plot (Figure 20 B). To characterize the reaction with PYR and 2OIV, two kinetic assays were performed in which either the concentration of PYR or 2OIV was kept constant, while that of the second substrate varied. Figure 20 C shows the enzyme activity with 3 mM PYR and varying amounts of 2OIV (1 to 20 mM). The formation of AL is catalyzed as long as

PYR is available in excess and drops to almost zero when C₇-pronecic acid is formed at 2OIV concentrations higher than 7.5 mM. The specific activity of SoC₇HAS based on C₇-pronecic acid production exhibits a shallow hyperbolic curve when plotted against the 2OIV concentration and results in an estimated K_M-value of 76 mM using the double reciprocal plot (Figure 20 D). Figure 12 E shows the enzyme activity 5 mM 2OIV and varying amounts of PYR (0.25 to 7.5 mM). In this case the amount of C₇-pronecic acid increases only until PYR concentrations of 2 mM and drops as soon as PYR is in excess, while the formation of AL shows a classical hyperbolic curve with a reduced slope resulting from the competitive reaction with 2OIV at lower PYR concentrations. The double reciprocal plot results in a K_M value of 4.9 mM for PYR, indicating that the presence of 2OIV affects the binding of PYR as acceptor to the enzymes active site.

If only the first four values of enzyme activity base on C₇-pronecic acid formation are used for the double reciprocal plot, a K_M of 0.25 mM for PYR is estimated. We interpret the K_M of 0.25 mM as an estimate for the affinity of SoC₇HAS to PYR as substrate for decarboxylation to the active acetaldehyde, *i.e.*, the first step of the classical AHAS reaction mechanism, as PYR is only used as a donor in the formation of C₇-pronecic acid while 2OIV serves as acceptor. The K_M of 4.9 mM, estimated by the AL formation at higher PYR concentrations should be an estimate of the affinity of SoC₇HAS to PYR as acceptor substrate for the active acetaldehyde in the second substrate binding pocket.

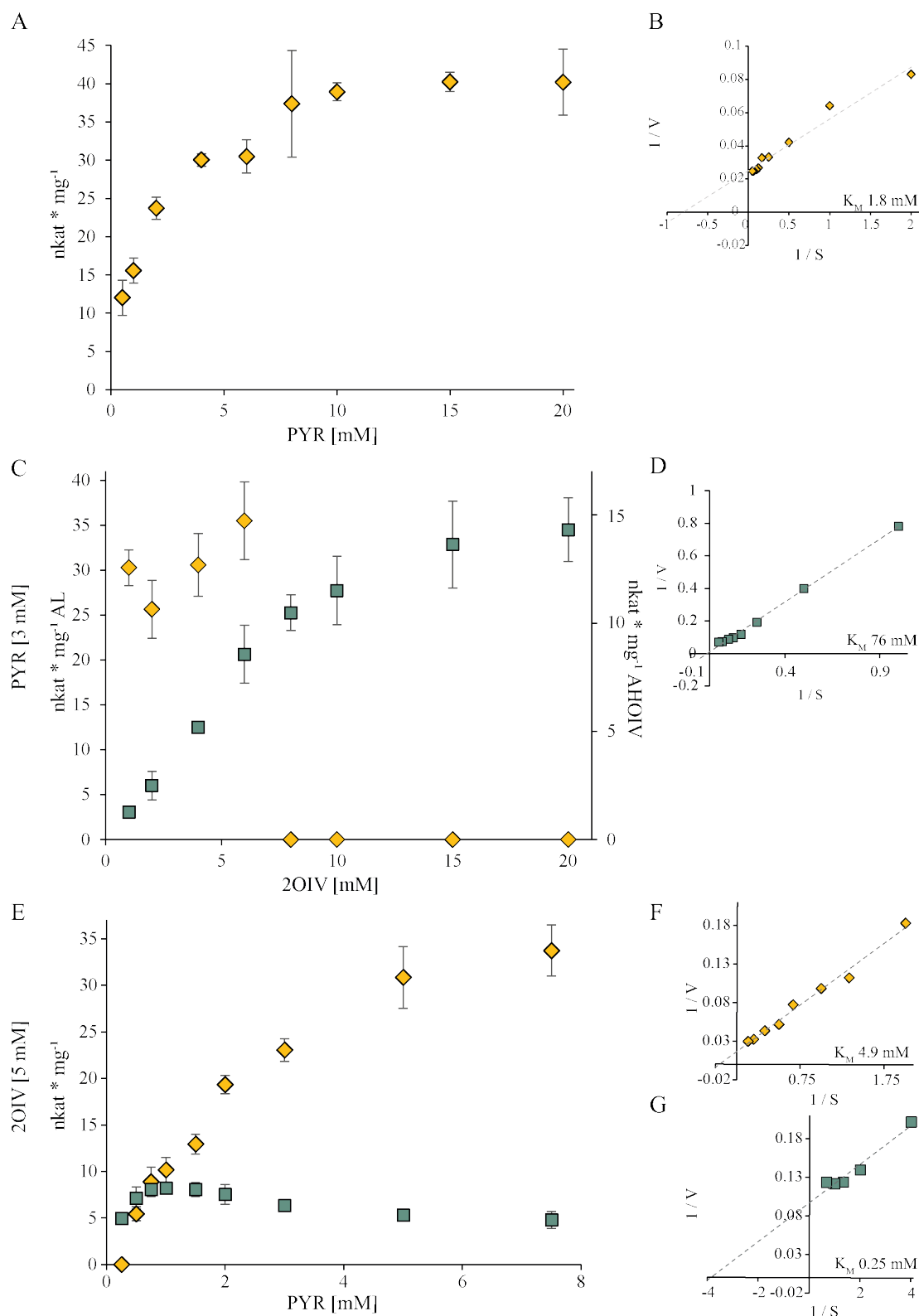


Figure 20 Formation of acetohydroxyacids as a function of the substrate concentrations.

◆ AL formation, ■ C_7 -pronecic acid formation. $10 \mu\text{g}$ of *SoC7HAS* were incubated in *in vitro* assays with different substrate ratios. **A** AL formation in reactions with 0.5-20 mM PYR. **B** Double reciprocal plot (Lineweaver-Burk plot) of the same data. **C** AL and C_7 -pronecic acid formation in reactions with constant amount of PYR (3 mM) and varying amounts of 2OIV (1-20 mM). **D** Double reciprocal plot of the same data for C_7 -pronecic acid formation. **E** AL and C_7 -pronecic acid formation in reactions with constant amount of 2OIV (5 mM) and varying amounts of PYR (0.25 - 7.5 mM). Double reciprocal plot of the same data as in **E** for AL formation **F** and C_7 -pronecic acid formation **G**.

3.3.8 *SoC₇*HAS is promiscuous regarding its acceptor substrate

The observation that *SoC₇*HAS accepts in addition TO the substrates of BCAA biosynthesis also 2OIV as a substrate with a branched side chain, raised the question how far the substrate acceptance differs from that of *SoCSU* that was identified before as a candidate most likely being involved in BCAA biosynthesis (Chapter 2). Therefore, we tested further structural analogues as substrates, *i.e.* the two long-chain keto acids 2-oxovalerate and 2-oxooctonate and the two branched compounds 3-methyl-2-oxopentionate and 4-methyl-2-oxopentionate (Figure 21 A). To favor the two-substrate reaction in the presence of PYR, the reactions contained 1 mM PYR to allow the formation of the activated acetaldehyde and 5 mM of the second keto acid as acceptor of the activated acetaldehyde. After 4 min incubation the reactions were stopped, derivatized with ECF and analyzed by GC-MS. The specific activities of *SoC₇*HAS and of *SoCSU* (Chapter 2) for the various substrates show that *SoC₇*HAS accepts a wider substrate spectrum than *SoCSU* Figure 21 B. While *SoCSU* accepts, in addition to PYR and 2OB, 2-oxovalerate, a substrate with a small unbranched side chain. The *SoC₇*HAS converts all keto acids tested. *SoC₇*HAS shows the highest activities with the substrates of BCAA biosynthesis and smaller compounds without branching, *i.e.*, PYR, 2OB and 2-oxovalerate with 27.1, 12.1 and 17.5 nkat mg⁻¹, respectively. But it also accepts substrates with branched or longer side chains as 2OIV (7.3 nkat mg⁻¹), 4-methyl-2-oxopentionate (2.1 nkat mg⁻¹), and 2-oxooctonate (1.8 nkat mg⁻¹). 3-methyl-2-oxopentionate is only converted in trace amounts. The GC-MS properties of the ethylated products C₇-pronecic acid, 2-acetohydroxy-3-methyl-2-oxopentionate, and 2-acetohydroxy-4-methyl-2-oxopentionate, shown here for the first time, are summarized in Supplementary Table 6. In conclusion, the *SoC₇*HAS is highly promiscuous in the choice of acceptor compounds for the transfer of the active acetaldehyde, a property not observed for CSUs most likely involved in BCAA biosynthesis.

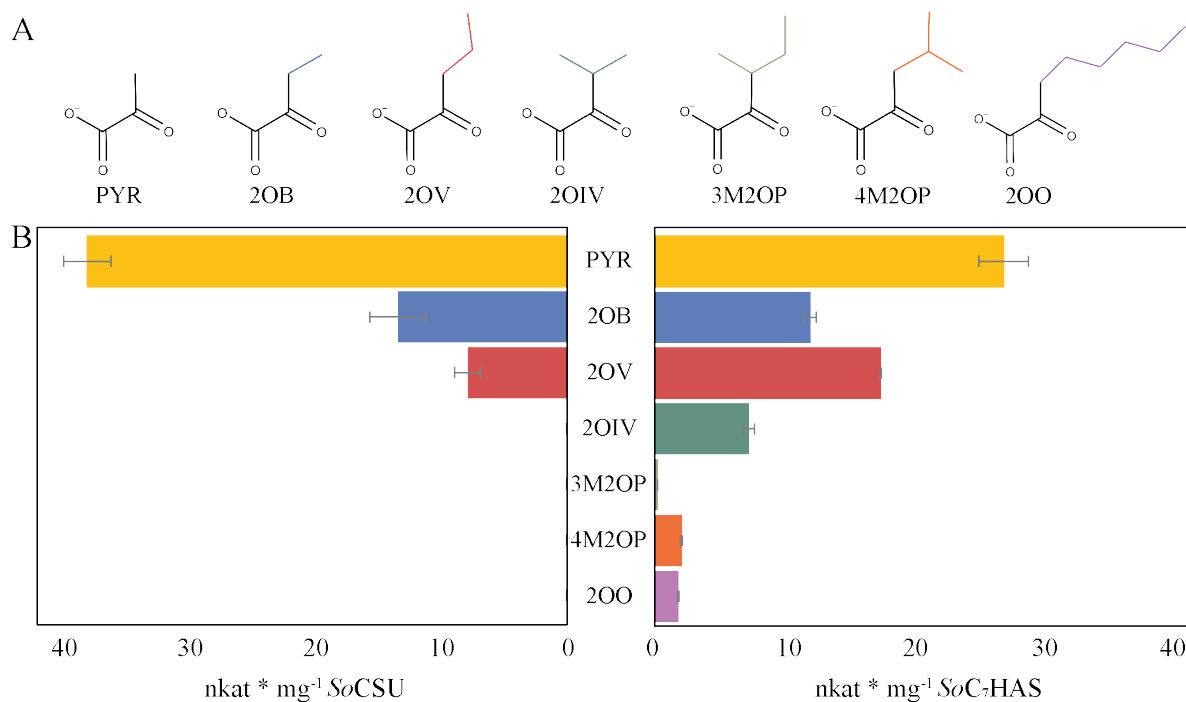


Figure 21 Specific activities of CSU and C₇HAS of *S. officinale* with different α -ketoacids as acceptor substrates for the activated acetaldehyde. **A** Different ketoacids tested as acceptor substrates. Common to tested substrates is the C₃ structural core identical to pyruvate that carries carbon side chains of different length and/or branching (highlighted in color). **B** Specific activities of *SoCSU* (left) and *SoC₇HAS* (right) with different substrates. Substrate concentrations have been 6 mM PYR when given as sole substrate and 1 mM PYR as donor molecule with 5 mM acceptor substrate. Specific activities result from 4 min incubations. The reactions were subsequently ethylated with ECF and analyzed by GC-MS. 2OV = 2-oxovalerate, 3M2OP = 3-methyl-2-oxopentanoate, 4M2OP = 4-methyl-2-oxopentanoate, 2OO = 2-oxooctanoate.

3.4 Discussion

By catalyzing the first metabolic reaction in the biosynthesis of the BCAAs valine, leucine, and isoleucine, AHASs are essential enzymes of primary metabolism in bacteria, fungi, and plants. This fundamental reaction has been well studied and the enzymes have been extensively characterized. Furthermore, since animals do not have AHASs, they provide a target structure for a wide variety of herbicides which in turn brings the enzymes into the focus of agricultural and structural research (Saari et al., 1994; McCourt et al., 2006; McCourt and Duggleby, 2006; Tan et al., 2006; Lonhienne et al., 2018). Additionally there is great interest in AHAS-inhibiting herbicides as potential drugs for the treatment of human fungal pathogenic infections (Garcia et al., 2018; Meng et al., 2022) and as a target for antimicrobial drug discovery (Pue and Guddat, 2014). Despite all this research, little is known about the function of the AHAS gene copies that can be found in several plant species. AHAS homologs are known, e.g. from *Brassica napus* (Rutledge et al., 1991) *Helianthus annuus* (Kolkman et al., 2004), and *Nicotiana tabacum* (Mazur et al., 1987; Lee et al., 1988). Just recently it was shown that one of the AHAS homologs in *Nicotiana tabacum* is involved in secondary metabolism (Chang et al., 2020). *Nicotiana tabacum* synthesizes acyl sugars in glandular trichomes as defense substances against herbivores (Keene and Wagner, 1985; Kandra and Wagner, 1988). In earlier experiments with herbicides targeting the AHAS, it was observed that the plants were no longer able to produce acyl sugars (Kandra and Wagner, 1990). CRISPR/Cas9 introduced loss-of-function mutants of the trichome-specific CSU paralogue resulted in significantly lower acyl sugar content in the mutants, confirming the involvement of an AHAS (Chang et al., 2020). It would be interesting to test whether the catalyzed reaction is the fundamental AHAS reaction of primary metabolism, or whether this AHAS catalyzes acyl sugar pathway specific reactions. Here we show another recruitment of an AHAS subunit for secondary metabolism: in *S. officinale*, a CSU paralogue is involved in the biosynthesis of lycopsamine type PAs, more specifically in the biosynthesis of the unique C₇-necic acids of these PAs.

3.4.1 In planta studies of SoC₇HAS

As shown in numerous cases, CRISPR/Cas9 genome editing is an invaluable tool in the research of non-model plants (Bhowmik et al., 2021; Fizikova et al., 2021; Zhang et al., 2021). In transgenic hairy roots of *S. officinale* it was shown that CRISPR/Cas9 knock-outs of HSS, the key enzyme of PA biosynthesis, lead to PA deficient hairy roots (Zakaria et al., 2021). Mutants with loss-of-function of HSO, the enzyme that catalyzes the second step of necine base formation, also lack PAs while simultaneously showing an accumulation of homospermidine

(Zakaria, 2022). A similar effect can be expected from other enzymes involved in PA biosynthesis, presumably with the accumulation of intermediates preceding the affected catalytic step. We applied CRISPR/Cas9 genome editing to elucidate the role of the C₇HAS in the PA biosynthesis in *S. officinale*. Thereby it was shown that a genotypic knock-out of *soc7has* leads to a phenotypic knock-out of the complex lycopsamine type PAs. Simultaneously, the saturated necine base trachelanthamidine accumulates lacking the C₇-necic acids (+)-trachelanthic and (-)-viridifloric acid (Figure 15), providing *in planta* evidence of an involvement of SoC₇HAS in the biosynthesis of the C₇-necic acid moieties of lycopsamine type PAs. The same effect was observed for two different target sequences, making an off-target effect very unlikely. Knock-out hairy roots have not only provided evidence for the involvement of SoC₇HAS in PA biosynthesis, they have also provided deeper insights into the regulation and organization of PA biosynthesis in *S. officinale*. For example, the regeneration of transgenic plants from hairy roots showed that there is no transport of the saturated necine base from the roots into the plants. This confirms the tissue specificity of the enzymes involved in PA biosynthesis. As known from other alkaloids, the metabolic pathways are characterized by high tissue and even cell specificity (Facchini, 2001; Facchini and St-Pierre, 2005). Even though hairy root cultures of *S. officinale* are known to be capable of synthesizing the entire bouquet of naturally occurring PAs (Frölich et al., 2007; Kruse et al., 2019; Zakaria et al., 2021), the lack root to shoot translocation of trachelanthamidine demonstrated that the enzymatic activity of the enzymes downstream of HSS is restricted to the roots. When regenerated whole plants from *S. officinale* *c7has* knock-out hairy roots started to develop inflorescences, we were able to show that again, PA biosynthesis is interrupted after trachelanthamidine and no lycopsamine type PAs are produced. This allows us to conclude that in *Symphytum* only one C₇HAS catalyzes the formation of the C₇-pronecic acid at both biosynthetic sites. This shows that PA biosynthesis in *S. officinale* is very versatile. Of the HSS as the key enzyme of PA biosynthesis, only one coding gene is known to be transcribed and expressed at both biosynthetic sites (Kruse et al., 2017). The second step of base biosynthesis is catalyzed by two different HSO homologs, one that is root-specific and one that is leaf-specific (Zakaria, 2022). In the first step of C₇-necic acid synthesis, it is again only one enzyme, adding another fascinating facet to the versatile recruitment and integration of enzymes to the PA pathway in *S. officinale*.

Of note: Sequencing of SoC₇HAS PCR products of hairy roots resulting from the CRISPR/Cas9-mediated approach resulted in many cases in the identification of more than two alleles. This phenomenon is attributed to the nature of the CRISPR/Cas9 system: once transformed into an organism it can continuously target and cleave genes with different

efficiency and at different developmental stages of the organism, in our study of the hairy roots. As a result, not all cells have identical mutations of the targeted gene, a phenomenon referred to as mosaicism (Mizuno et al., 2014; Oliver et al., 2015; Mehravar et al., 2019).

3.4.2 Promiscuity of *SoC7HAS*

Weber et al. (1999) predicted that a CSU of an AHAS is involved in the biosynthesis of lycopsamine type PAs and transfers an activated acetaldehyde, originating from PYR, to 2OIV. As the product the direct precursor of the C₇-necic acids (+)-trachelanthic and (-)-viridifloric acid, the so-called C₇-pronecic acid is formed. The application of our recently developed GC-MS based assay to identify and quantify substrates and products of AHAS reactions (Chapter 2) provided *in vitro* evidence that *SoC7HAS* is indeed able to catalyze the predicted reaction. In contrast, the *SoCSU* is not able to accept 2OIV as substrate. In evolutionary terms, the *SoC7HAS* thus followed one of the key evolutionary mechanisms of gene duplicates leading to the formation of secondary metabolites, neofunctionalization (Edger et al., 2015; Freeling et al., 2015; Birchler and Yang, 2022). With a K_M of 76 mM the affinity of *SoC7HAS* towards 2OIV is rather low, further supporting the hypothesis that (+)-trachelanthic acid and valine share the early steps in the biosynthesis until the intermediate 2OIV in the same cell compartment and thereby not causing any limitation of 2OIV as a substrate for PA biosynthesis. In addition to the formation of the pronecic acid as part of PA biosynthesis, *SoC7HAS* is also able to use the branched-chain C₆ keto acids 3-methyl-2-oxopentionate and 4-methyl-2-oxopentionate, and the C₈ unbranched 2-oxooctonate as acceptors of the activated acetaldehyde. This shows that *SoC7HAS* is a fairly promiscuous enzyme. In contrast, *SoCSU* only uses short unbranched acceptor molecules. In few earlier studies non-standard substances were tested as substrates for AHAS. For example, it was shown that the *E. coli* AHAS isoenzyme III is capable of accepting 2-oxovalerate in addition to PYR and 2OB. In the same study, 2OIV was also tested as a substrate, which competed very poorly, if at all, with the normal substrates and the results were not shown (Gollop et al., 1989). Another example is the use of AHASs for the *in vitro* synthesis of pharmacological precursors. The AHASs of *E. coli*, in particular isoenzyme I, are potent enzymes for the synthesis of (*R*)-phenylacetylcarbinol from benzaldehyde and pyruvate, and are also able to convert other derivatives of benzaldehyde (Engel et al., 2003; Engel et al., 2004a). However, most substrate specificity studies of AHAS focus on the preference of the enzymes for the formation of the BCAA precursors AL and AHOB. This focus led to the identification of a number of amino acids that have a crucial influence on the substrate preference of PYR over 2OB and vice versa in several bacterial CSUs (Ibdah et al., 1996). For example, site-directed mutagenesis of the conserved tryptophan W464

in the AHAS isoenzyme II of *E. coli* decreases the preference towards 2OB by 6-20-fold (Engel et al., 2004b). Modeling the amino acid sequence of AHAS isoenzyme II on the tertiary structure of *Arabidopsis thaliana* (PDB 6U9H) (Lonhienne et al., 2020), W464 (W574 in *Arabidopsis thaliana*) is located directly in the α -helix of the substrate access tunnel. Despite the structural similarity of all the keto acids tested, the branched substrates appear to fit only in the substrate access tunnel of the SoC₇HAS and not that of the SoCSU (Chapter 2), suggesting a crucial change in the shape or size of the substrate access tunnel.

3.4.3 Does the SoC₇HAS form a complex with an RSU?

Of note: The question of which role the regulatory subunit has in the SoC₇HAS remains an exciting one. Several possible scenarios come to mind:

(1) The SoC₇HAS uses the RSU of the primary metabolism AHASs.

The two main roles of the RSUs of AHASs are to stabilize the enzyme complex and to inhibit catalysis by a feedback mechanism after binding BCAAs as end products of the pathway. The greatest inhibitory effect is achieved when a combination of different BCAAs is used (Mifflin, 1971; Durner and Böger, 1988; Lee and Duggleby, 2001). An inhibition of PA biosynthesis by BCAAs should not be of biological relevance and is regarded as unlikely.

(2) Both CSU and RSU were duplicated and functionalized for PA biosynthesis.

In this case, co-expression of PA-specific CSU and RSU would be expected. In our view, two points argue against this hypothesis: In *S. officinale* we could only identify one RSU and no transcripts of this subunit could be detected in the roots, the constitutive biosynthetic site of PAs in *S. officinale* (Niemüller et al., 2012). And secondly, modification of the effector site in a PA-specific RSU to allow feedback inhibition by C₇-pronecic acid or lycopsamine type PAs as end product of the pathway would be very complex and we would consider the cost-benefit factor to be insufficient.

(3) The SoC₇HAS no longer requires an RSU.

Another hypothesis is that only the CSU gene duplicate was optimized and recruited to PA biosynthesis. That only single subunits of hetero-multimeric ancestor proteins are recruited for new functions in secondary metabolism is described in the literature for the evolution of the first step in the biosynthesis of benzoxazinoids, compounds produced by various species of the Poaceae to fend off insect herbivores. This step is catalyzed by Bx1 that was shown to originated from a duplication of the gene encoding the alpha subunit of the tryptophan synthase (Frey et al., 2009). Currently, we consider the recruitment of only the CSU for PA biosynthesis to be the most likely hypothesis.

3.4.4 Conclusion

Concluding the *in planta* and *in vitro* studies in this work allows the introduction of an updated biosynthetic pathway to the core structure of lycopsamine type PAs in *S. officinale* (Figure 22). The left part of the pathway represents what is confirmed by previous data: HSS uses spermidine and putrescine to form homospermidine (Ober and Hartmann, 1999). Homospermidine oxidase (HSO) deaminates the homospermidine and forms the double ring backbone 1-formylpyrrolizidin that is reduced by an unidentified reductase to trachelanthamidine (Zakaria, 2022). Highlighted in green is what we can add, following the results of this work:

- (1) *SoC₇HAS* catalyzes the formation of C₇-pronecic acid by forming an activated acetaldehyde from PYR and transferring it to 2OIV.
- (2) Trachelanthamidine is the esterification partner of the acid moiety, although it remains unclear whether C₇-pronecic acid or its reduced forms (+)-trachelanthic or (-)-viridifloric acid are esterified to trachelanthamidine.
- (3) All modifications of the base, more precisely desaturation, C₇-hydroxylation, and *N*-oxygenation, occur only after esterification with the necic acid, overruling the previous assumption of supinidine as esterification partner (Frölich et al., 2007).

The grey part and grey dashed arrows represent uncertain biosynthetic steps whose sequence, substrates, enzymes and/or intermediates have yet to be determined and offer an interesting field for further studies.

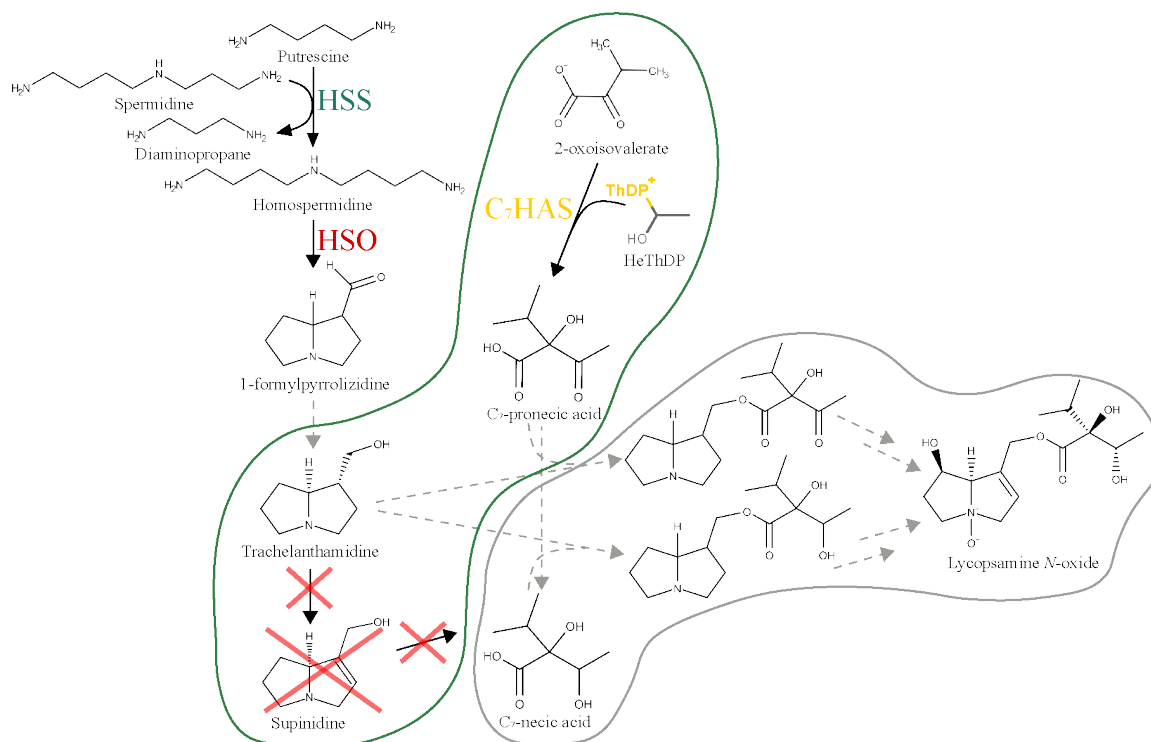


Figure 22 PA biosynthesis extended by the results of this work. The left part represents the previously known steps, results of this work are highlighted in green, uncertain steps are highlighted in grey. Grey, dashed arrows represent reactions for which the sequence and/or enzymes are not yet known. HSS = homospermidine synthase; HSO = homospermidine oxidase; C₇HAS = C₇-hydroxyacid synthase; HeThDP = hydroxyethyl-thiamine diphosphate.

described earlier (Kruse et al., 2017). To calculate the relative transcript levels the $2^{\Delta Ct}$ method was employed (Schmittgen and Livak, 2008). *efl α* served as a reference gene to normalize the expression levels of the genes of interest.

Genome editing approach of *SoC γ HAS*

3.5.4 CRISPR/Cas9 vector construction and generation of transgenic hairy root mutants

For generation of the CRISPR/Cas9 vector constructs, the entry vector pEn-Chimera and the destination vector pDe-Cas9 were kindly gifted from Holger Puchta (Schiml et al., 2016). The pDe-Cas9 plasmid was modified by replacing the cassette mediating phosphinotricin resistance in the T-DNA against that mediating resistance against hygromycin (Zakaria et al. 2021). Potential CRISPR sites were predicted using Geneious Prime software (V2020.1.2, Biomatters, Auckland, New Zealand) based on (Doench et al., 2014). The selected sgRNA oligonucleotides for modification of the gene encoding C γ HAS of *S. officinale* contained the *Streptococcus pyogenes* Cas9 protospacer adjustment motif (PAM) NGG and the specific 20 nucleotides long targeting sequence to guide the Cas9 protein to the cleavage site. The three individual annealed sgRNA oligonucleotide pairs 11/12, 13/14 and 15/16 (Supplementary Table 3) were cloned into pEN-Chimera. Subsequently, the sgRNA cassettes were transferred from the entry vectors into the destination vector pDe-Cas9HYG using Gateway® cloning with Clonase® II (Thermo Fisher Scientific). The constructs were transformed into *A. rhizogenes* strain TR105. Transformation of *S. officinale* leaves was carried out as described before (Zakaria et al., 2021). Briefly, the bacterial suspensions were injected into the sterilized leaf pieces through a syringe. These were then placed on plates with solid MS20 medium (1 % [w/v] agar) for three days. Thereafter they were transferred to plates with 250 μ g/ml of a mixture of ticarcillin and clavulanic acid (mixed at a ratio of 15:1, Duchefa) to prevent overgrowth with *A. rhizogenes*. Emerging hairy roots of approximately 1 cm were cut from the leaf and transferred to fresh MS20 media plates containing the ticarcillin and clavulanic acid (250 μ g/ml), and hygromycin B (25 μ g/ml) for the selection of successfully transformed hairy root lines. Selected hairy root lines were cultivated on MS20 media plates containing the ticarcillin and clavulanic acid (250 μ g/ml).

3.5.5 Plant regeneration from hairy roots

To regenerate plants from hairy roots, shoot growth was induced by adding thidiazuron (1 μ g/ml) to solid MS20 and cultivation in an artificial climate-controlled chamber under a 16 h/8 h light/dark cycle (206 μ mol m⁻¹ sec⁻¹) at a constant temperature of 24°C and relative humidity of 65 %. After seven days the roots were transferred onto solid MS20 plates without

thidiazuron until shoots formed. These were transferred to fresh MS20 plates and subsequently to bigger glass vessels containing solid MS20 media to allow growth. If necessary, ticarcillin and clavulanic acid mix (250 µg/ml) was added to the media. Plants with well-developed shoots and roots were transplanted into soil and grown as described for the wild-type plant material. Inflorescence formation was induced by cold at 4°C and darkness for 3 weeks.

3.5.6 Genetic identification of transgenic hairy root lines

Transgenic hairy root lines were genetically identified by PCR analysis using specific primers. Approximately 100 mg of each transgenic hairy root line were used for the extraction of genomic DNA by the GeneJET Plant Genomic DNA Purification Kit (Thermo Fisher Scientific) according to the manufacturer's instructions with one modification: Approx. 10 mg polyvinylpyrrolidone (Polyclar AT, Serva, Heidelberg, Germany) was added to lysis buffer A to bind the phenolics present in the extract and prevent interaction with nucleic acids. The quality and concentration genomic DNA were analyzed by NanoDrop2000 UV–vis Spectrophotometer (Thermo Fisher Scientific). Primer pair 17/18 (Supplementary Table 3) was used to amplify the regions of genomic DNA targeted by the three CRISPR/Cas9 sgRNA constructs in a PCR using GoTaq G2 DNA polymerase (Promega) or JumpStart™ REDAccuTaq® LA DNA Polymerase (Sigma Aldrich) in reactions of 20 µl with appx. 10 ng genomic DNA as template. PCR amplicons were purified with the GeneJET Gel Extraction Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. DNA was quantified by using a NanoDrop ND2000 UV/VIS spectrometer and amplicons were sent for sequencing directly when only one band was clearly visible in the agarose gel. If PCR amplification resulted in more bands or when sequencing of PCR products resulted in ambiguous sequence signals, DNA fragments were cloned into pGEM T-easy vector (Promega), purified with GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific), and sequenced. Genomic sequences were aligned and analyzed by Geneious software with the “map to reference” and “pairwise alignment” tools in comparison to the respective sequence of the control lines and the wild type.

3.5.7 Pyrrolizidine alkaloid identification and quantification by GC-MS

For the quantification of PAs root material taken from a well-grown hairy root line were transferred into 80 ml liquid MS20 media (250 µg/ml ticarcillin/clavulanic acid mixture) and grown in the dark at 16 °C for 14 d on an orbital shaker at 130 rpm. The hairy root mutants were grown three times independently under the same conditions to give the technical replicates for PA analysis. The hairy roots were patted dry with tissue paper, weighed and freeze-dried for PA extraction. Appx. 100 mg pulverized root material was mixed with 12 ml of 0.05 M

H₂SO₄ and 3 g of zinc powder. After the addition of 20 µg heliotrine as an internal standard, samples were incubated end over end for appx. 16 h. Subsequently, the extracts were centrifuged at 14 000 rpm for 10 min and the supernatant was purified over Strata®-X-C strong cation exchange solid phase extraction columns (30 mg / 1 ml, Phenomenex®). For this purpose, the columns were first washed with 1 ml methanol and conditioned twice with 1 ml 0.05 M H₂SO₄ and then the supernatant from the extraction was applied. The column was washed twice with 1 ml ethyl acetate and twice with 1 ml methanol. The elution of the analytes was performed in three steps, each with 1 ml of 5% ammonia solution in methanol (v/v). Samples were dried under continuous air flow and reconstituted in pure MeOH, transferred to µVials (Carl Roth) and analyzed by GC-MS.

GC-MS measurements were carried out using a Focus GC system coupled to a DuoMass Spectrometer (Thermo Scientific™) equipped with a split/splitless (SSL) injector and a TraceGOLD TG-5SilMS (modified 5 % phenyl methyl polysiloxane for basic compounds; 30 m× 0.25 mm i.d., 0.25 µm film thickness, Thermo Scientific™). The applied GC conditions were as follows: injector temperature: 250 °C; temperature program: 100°C for 5 min, 100°C to 160°C at 25°C/min, 160°C to 300°C at 10°C/min, final hold at 300°C for 5 min; carrier gas: helium with a flow of 1 ml/min; injection mode: split ratio 1:10; MS transfer line: 300°C. EI-MS-spectra were recorded at 70 eV and an ion source temperature of 280°C.

Most retention indices (RI) for PAs refer to GC-MS measurements with a DB-1 column (100 % dimethyl polysiloxane, non-polar). In order to be able to compare the literature values, selected samples were additionally tested on a TG-1MS column (DB-1 equivalent; 30 m× 0.25 mm i.d., 0.25 µm film thickness, Thermo Scientific™). GC-MS data were obtained with a Thermo TSQ TRACE1300/1310 coupled to a Triple Quadrupole Mass Spectrometer (Thermo Scientific) equipped with a PTV injector. GC conditions were: PTV injector 280°C, split ratio 1:10, carrier gas helium with a flow rate of 1 ml/min, temperature program 100°C for 5 min, then increasing by 6°C/min to 300°C followed by 10 min at 300°C, MS transfer line 300°C. EI-mass spectra were recorded at 70 eV (ion source temperature 280°C).

The internal standard (heliotrine) and PAs were identified via comparison with in-house and literature reference data of mass spectra and Kovats indices (RI values). The linear retention indices were calculated with a reference set of co-injected hydrocarbons (Sigma-Aldrich). PAs were quantified via their signal peak area in the total ion chromatogram relating to heliotrine as internal standard (= heliotrine equivalents).

3.5.8 Heterologous expression in *E. coli*

For heterologous expression of the *SoC₇HAS* the gene was cloned into pCDFDuet1_modStrep. The procedure was described before for *SoCSU* (Chapter 2) using the gene-specific primers of *SoC₇HAS*, primer pair 19/20 (Supplementary Table 3) for cloning. The gene was amplified without the N-terminal transit peptide. Overexpression and affinity purification via StrepTag were described previously in Chapter 2.

3.5.9 Enzyme activity assays

The enzyme activity assays have been described previously in Chapter 2.

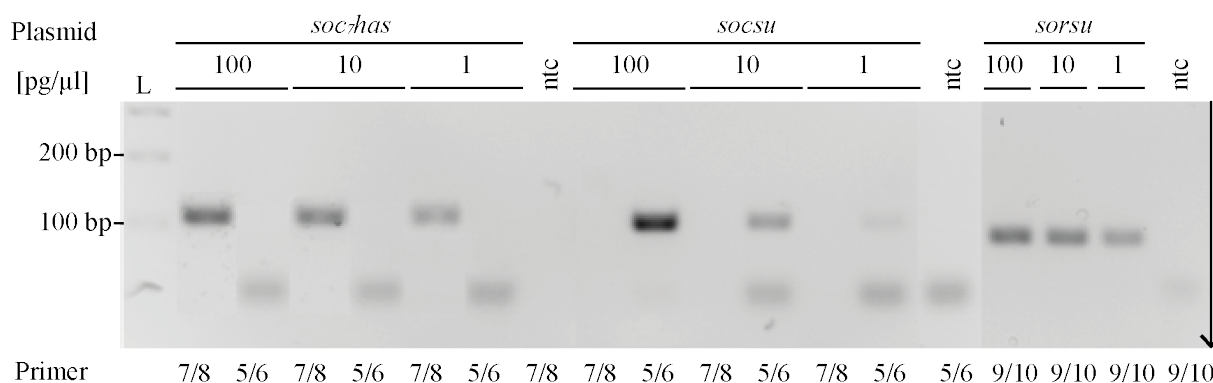
3.5.10 Derivatization with ECF

The derivatization of the enzymatic reaction has been described before in Chapter 2.

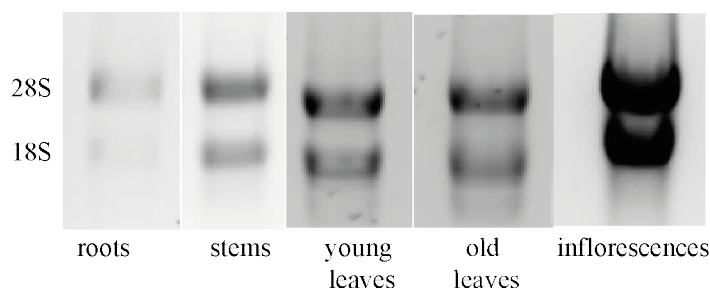
3.6 Supplementary

Supplementary Table 3 Oligonucleotides

No.	5'-sequence-3'	Purpose
1	CCACCACCCCAAAATATTCCAAG	qPCR amplification <i>efla</i>
2	CAGTCCAGGTTGGTGGACCT	qPCR amplification <i>efla</i>
3	AGTGCTATGGACAATGAATCAGTGA	qPCR amplification <i>hss</i>
4	CAGCAAAATCAGCGCCTCCA	qPCR amplification <i>hss</i>
5	TGCAGATGTTGGGGATGCAT	qPCR amplification <i>socsu1</i>
6	CAGCTTCCCAGTCACACGAT	qPCR amplification <i>socsu1</i>
7	CATGAATGTGCAAGAGTTGGCT	qPCR amplification <i>soc7has</i>
8	TGGTCTTCCCATTGTACGGC	qPCR amplification <i>soc7has</i>
9	GCTCTGCAGAGACTGTTGGA	qPCR amplification <i>sorsu</i>
10	TTGGAGTCAACGCCTGATCC	qPCR amplification <i>sorsu</i>
11	<u>ATTGAGTATAATCGACCTAGGCAG</u>	N20 oligo for motif A, overhang for ligation underlined
12	<u>AAACCTGCCTAGGTCGATTATACT</u>	N20 oligo for motif A, overhang for ligation underlined
13	<u>ATTGATAATAGCACCCATAGCTGC</u>	N20 oligo for motif B, overhang for ligation underlined
14	<u>AAACGCAGCTATGGGTGCTATTAT</u>	N20 oligo for motif B, overhang for ligation underlined
15	<u>ATTGTCTTGGTGA</u> CTCGAGCAGCA	N20 oligo for motif C, overhang for ligation underlined
16	<u>AAACTGCTGCTCGAGTCACCAAGA</u>	N20 oligo for motif C, overhang for ligation underlined
17	TCCAGCCTCTGATGAGTTGT	Amplification of <i>SoC7HAS</i> to test for mutations forward
18	CCACTGGGGATCATAGGCAA	Amplification of <i>SoC7HAS</i> to test for mutations reverse
19	TATAGTCGACCCCTTGTTTCATTTCTAGGTACTCCCTTCGGAG	<i>soc7has</i> cloning primer forward with <u>SalI</u> restriction site
20	TATATAGCGGCCGCTAAACTTGTGGGATTTCATTGGCTGAAAC	<i>soc7has</i> cloning primer reverse with <u>NotI</u> restriction site



Supplementary Figure 5 RT-qPCR primer specificity. Primers for the amplification of *soc7has* (primers 7/8) *socsu* (primers 5/6) and *sorsu* (9/10) were tested for their specificity in PCR reactions containing the respective plasmids as templates at different concentrations. PCR was done with GoTaq DNA polymerase (Promega) at an annealing temperature of 60°C. M, DNA ladder (GeneRuler 100 bp Plus DNA Ladder, Thermo Scientific). ntc = no template control. Resulting PCR products were separated on a 3% agarose gel.



Supplementary Figure 6 Quality test of total RNA used for reverse transcription and RT-qPCR. Total RNA was analyzed by agarose gel electrophoresis (1.5 % gel) to exclude degradation.

Supplementary Table 4 Pyrrolizidine alkaloid GC EIMS data. PAs No 1-17 as labelled in the total ion chromatogram of Figure 14 * elution sequence of the stereoisomers is not confirmed.

No	PA	[M] ⁺	MS data: characteristic ions m/z (% relative abundance)
	Trachelanthamidine	141	53(6), 54(7), 55(42), 56(4), 67(5), 68(9), 70(8), 73(3), 80(9), 81(5), 82(66), 83(100), 84(11), 96(3), 108(7), 110(9), 122(4), 124(14), 140(7), 141(18)
	Trachelanthamidine TMS	213	55(18), 59(6), 68(5), 70(4), 73(16), 75(5), 80(6), 82(48), 83(100), 84(17), 96(4), 108(7), 110(17), 122(21), 124(11), 184(5), 185(11), 198(10), 212(4), 213(7)
	Heliotrine	313	53(3), 59(34), 67(10), 80(15), 93(89), 94(36), 95(16), 106(3), 108(5), 111(3), 118(4), 119(3), 120(6), 131(6), 136(19), 137(22), 138(100), 139(40), 140(4), 156(15)
	Heliotrine 3xTMS	515	59(48), 67(21), 71(12), 73(73), 74(7), 75(21), 80(8), 89(10), 93(78), 94(100), 95(23), 118(7), 120(25), 133(16), 136(5), 203(8), 208(9), 210(88), 211(27), 212(6)
1	Intermedine	299	53(6), 54(3), 67(10), 71(5), 73(3), 80(17), 93(100), 94(59), 95(16), 106(3), 108(5), 111(4), 117(5), 120(12), 121(3), 136(14), 137(18), 138(87), 139(26), 156(8)
2	Lycopsamine	299	53(5), 55(3), 65(4), 67(11), 73(5), 80(19), 81(4), 92(6), 93(90), 94(76), 95(15), 106(3), 108(4), 118(4), 120(12), 136(14), 137(14), 138(100), 139(25), 156(8)
3	3'-acetyllycopsamine	341	53(4), 65(3), 66(3), 67(8), 73(3), 80(14), 93(90), 94(62), 95(7), 99(9), 108(3), 117(7), 118(5), 119(3), 120(11), 136(18), 137(18), 138(100), 139(18), 156(3)
4	7-acetylintermedine	341	67(5), 80(6), 92(3), 93(44), 94(23), 95(6), 101(10), 106(3), 117(4), 118(6), 119(12), 120(44), 121(13), 136(13), 178(4), 179(9), 180(100), 181(32), 182(3), 198(8)
5	7-acetyllycopsamine	341	67(6), 80(10), 81(6), 82(9), 93(51), 94(22), 95(12), 101(8), 108(7), 119(18), 120(57), 121(14), 136(22), 179(12), 180(100), 181(10), 182(7), 198(6)
6	?		55(19), 67(9), 68(7), 73(11), 80(12), 83(21), 93(82), 94(54), 106(13), 107(9), 118(13), 119(23), 120(52), 121(10), 136(78), 141(15), 211(8), 220(100), 221(11), 237(17)
7	?		55(34), 83(59), 93(70), 96(20), 104(14), 106(14), 119(43), 120(69), 136(72), 141(17), 145(13), 180(18), 181(13), 191(13), 211(14), 219(20), 220(100), 221(15), 237(13)
8	7,3'-diacetylintermedine/-lycopsamine	383	66(5), 73(6), 78(7), 79(5), 89(5), 93(29), 95(6), 99(11), 101(5), 118(7), 119(19), 120(27), 121(6), 126(5), 136(20), 179(8), 180(100), 181(22), 198(6)
9	?		73(17), 80(7), 81(7), 82(9), 93(44), 94(27), 106(14), 108(8), 118(9), 120(57), 121(14), 122(23), 136(17), 194(100), 195(25), 221(17), 269(9), 281(24)
10	Echiumine/*	381	55(11), 67(8), 80(15), 83(50), 93(68), 94(36), 95(11), 108(8), 117(8), 118(14), 119(22), 120(66), 121(30), 136(63), 137(10), 138(17), 141(11), 192(8), 220(100), 221(22)

11	Symlandine/*	381	52(8), 55(34), 71(9), 80(18), 83(22), 93(85), 94(37), 95(16), 106(11), 117(9), 119(28), 120(83), 121(28), 122(10), 136(39), 137(10), 141(21), 219(11), 220(100), 221(33)
12	Myoscorpine/*	381	55(17), 80(23), 83(68), 93(12), 94(56), 96(12), 108(25), 117(11), 119(20), 120(100), 136(63), 138(15), 141(13), 209(13), 220(86), 221(27)
13	Symphytine*	381	55(61), 78(20), 93(16), 108(23), 115(19), 118(19), 120(59), 122(55), 126(17), 136(67), 146(26), 161(33), 165(18), 172(24), 191(35), 220(100), 221(5)
14	3'acetylechiumine/*	423	55(6), 67(6), 80(6), 83(25), 92(6), 93(42), 94(37), 95(6), 99(10), 117(13), 118(9), 119(14), 119(39), 120(51), 136(67), 137(6), 141(9), 174(6), 220(100), 221(19)
15	3'acetylsymlandine/*	423	55(10), 67(11), 80(10), 82(12), 83(44), 92(10), 93(64), 94(42), 95(12), 99(10), 108(10), 118(19), 119(38), 120(76), 121(23), 136(78), 141(21), 220(100), 221(19)
16	3'acetylmyoscorpine/*	423	55(9), 80(9), 83(50), 93(81), 94(44), 95(9), 99(13), 117(13), 118(10), 119(48), 120(72), 121(13), 135(15), 136(96), 137(10), 141(13), 219(10) 220(100), 221(20)
17	3'acetylsymphytine*	423	55(16), 75(17), 80(16), 83(36), 93(95), 94(23), 96(36), 99(16), 117(21), 119(47), 120(100), 121(60), 122(19), 136(80), 137(25), 147(35), 192(27), 220(97), 221(53)

Supplementary Table 5 Sequence analyses of CRISPR/Cas9- induced mutations within the *soc7has* in transgenic hairy roots of *Symphytum officinale*. For all analyzed lines, 45 bp neighboring the 20 nt-long protospacer sequences with the attached PAM motif are shown in comparison with the sequence of the *soc7has* of the control line resulting from transformation with the empty vector. Two different protospacer constructs (A and C) with attached PAM are highlighted in the sequence, predicted cleavage site implied by scissors. Dashes indicate deletions, colored nucleotides insertions, numbers indicate larger insertions and deletions. PA and trachelanthamidine content are given in $\mu\text{g/g}$ fresh weight and are the mean of 3 independent purifications and measurements, the standard deviation is given in brackets.

hr	Target Sequence with Protospacer construct and PAM	PAs [$\mu\text{g/g}$]	Trach. [$\mu\text{g/g}$]
CL35	CAG TAC TAC AAG TAT AAT CGA CCT AGG CAG TGG TTG ACA TCC TCT	1800 (± 158)	-
A8	CAG TAC TAC AAG TAT AAT CGA CCT +59 AGG CAG TGG TTG ACA TCC CAG TAC TAC AAG TAT AAT CGA CCT +171 AGG CAG TGG TTG ACA TCC CAG TAC TAC AAG TAT AAT +195 CGA CCT AGG CAG TGG TTG ACA TCC CAG TAC TAC AAG TAT AAT CGA CCT AGA CAA TGG TTG ACA TCC TCT CAG TAC TAC AAG TAT AAT CGA CCT AGG --- --- TTG ACA TCC TCT	-	953 (± 42)
A12	CAG TAC TAC AAG TAT AAT CGA CC- --- -G TGG TTG ACA TCC TCT CAG TAC TAC AAG TAT AAT CGA CCT AGG --- --- --- --- ---	40 (± 9)	900 (± 33)
A13	CAG TAC TAC AAG TAT AAT CGA CCT AGG CAG TGG TTG ACA TCC TCT --- --- --- --- --- -177 --- --- --- --- ---	309 (± 54)	809 (± 149)
A14	CAG TAC TAC AAG TAT AAT CGA CCT AGG -AG TGG TTG ACA TCC TCT CAG TAC TAC AAG TAT AAT CGA CCT A-- CAG TGG TTG ACA TCC TCT CAG TAC TAC AAG TAT AAT CGA CCT AGG CAG TGG TTG ACA TCC TCT	157 (± 33)	260 (± 50)
A15	CAG TAC TAC AAG TAT AAT CGA CCT AGG --- TGG TTG ACA TCC TCT --- --- --- --- --- -59 --- --- --- --- --- TCT --- --- --- --- --- -428 --- --- --- --- ---	818 (± 121)	504 (± 85)
A27	CAG TAC TAC AAG TAT AAT CGA CCT AGG CAG TGG TTG ACA TCC TCT CAG TAC TAC AAG TAT AAT CGA CCT AGG -AG TGG TTG ACA TCC TCT --- --- --- --- --- -368 --- --- --- --- --- CTG --- --- --- --- --- --- --- -AG TGG TTG ACA TCC TCT	-	1242 (± 186)
A29	CAG TAC TAC AAG TAT AAT CGA CCT AGG --- TGG TTG ACA TCC TCT CAG TAC TAC AAG TAT AAT CGA CCT --- CAG TGG TTG ACA TCC TCT	-	964 (± 83)

A33	CAG TAC TAC AAG TAT AAT CGA CCT AGG --- +43 ---TG ACA TCC CAG TAC TAC AAG TAT TAG GGG AGG AAC ATT TGG T--- -417 --- --- -186 --- CAG TAC TAC AAG TAT A--- -41 --- CAG TAC TAC AAG TAT AAT CGT CAT CAT TC- -42 ---	-	1182 (±88)
A37	CAG TAC TAC AAG TAT AAT CGA C-- -GG CAG TGG TTG ACA TCC TCT CAG TAC TAC AAG TAT AAT CGA CCT AGG --- -GG CCA TGG ATT TTG	-	377 (±153)
A42	CAG TAC TAC AAG TAT AAT CGA CCT AGG -G TGG TTG ACA TCC TCT CAG TAC TAC AAG TAT AAT --- --- -TG ACA TCC TCT	-	764 (±104)
A43	CAG TAC TAC AAG TAT AAT CGA CCT AGG --- TGG TTG ACA TCC TCT CAG TAC TAC AAG TAT AAT CGA CCT --- --- TTG ACA TCC TCT CAG TAC TAC AAG TAT AAT CGA CCT AGG CAG TGG TTG ACA TCC TCT	31 (±3)	629 (±87)
A44	CAG TAC TAC AAG TAT AAT CGA CCT AGG --- -26 --- CAG TAC TAC AAG TAT AAT CGA CCT AGG --- -20 --- CAG TAC TAC AAG TAT AAT --- --- -TG ACA TCC TCT	-	839 (±271)
A50	CAG TAC TAC AAG TAT AAT CGA CC- --- -AG TGG TTG ACA TCC TCT CAG TAC TAC AAG TAT AAT CGA CCT AG- --- TGG TTG ACA TCC TCT CAG TAC TAC AAG TAT AAT CGA CCT AGG --- TGG TTG ACA TCC TCT	-	2028 (±366)
A53	--- -64 --- CAG TAC TAC AAG T--- --- -GG TTG ACA TCC TCT	-	1198 (±57)
A69	CAG TAC TAC AAG TAT AAT CGA CCT AGG --- -26 --- --- -428 --- CAG TGG TTG ACA TCC TCT CAG TAC TAC AAG TAT AAT CGA CCT AG +59 G CAG TGG TTG ACA TCC	-	1455 (±127)
CL35	GT AAT ATC CCT GCT GCT CGA GTC ACC AAG AAA GAT CAA CTC AGA	1800 (±158)	-
C88	GT AAT ATC CCT GCT GCT TCG AGT CAC CAA GAA AGA TCA ACT CAG A GT AAT ATC CC- --- --- C ACC AAG AAA GAT CAA CTC AGA	-	2276 (±423)
C106	GT AAT ATC CCT GCT -CT CGA GTC ACC AAG AAA GAT CAA CTC AGA GT AAT ATC CCT GC- --- --- -G ATA AAG AAA GAT CAA CTC AGA	62 (±8)	682 (±111)
C203	GT AAT ATC CCT GCT GCT TCG AGT CAC CAA GAA AGA TCA ACT CAG A GT AAT ATC CCT GCT GCT TCG AGT CAC CAA GAA AGA TCA ACT CAG A	-	1617 (±107)

Supplementary Table 6 GC EIMS data of products of the AHAS assay reactions ethylated with ECF. RI on TG-5 AMINE capillary column.

Substance	[M] ⁺	RI	MS data: characteristic ions m/z (% relative abundance)
Ethyl-2-acethydroxy-2-oxoisovalerate (C7-pronecic acid)	188	1185	27(7), 29(15), 31(4), 39(4), 41(7), 42(2), 43(92), 44(6), 45(5), 55(100), 56(6), 57(3), 72(2), 73(69), 74(7), 83(2), 90(3), 101(2), 145(5), 146(15)
Ethyl-2-acetohydroxy-3-methyl-2-oxopentionate	202	1275	17(25), 29(7), 31(9), 39(6), 41(21), 42(4), 43(33), 44(6), 45(3), 47(3), 52(3), 56(9), 57(100), 58(5), 72(6), 84(3), 85(44), 127(3), 177(4)
Ethyl-2-acetohydroxy-4-methyl-2-oxopentionate	202	1245	27(6), 29(14), 32(4), 39(4), 41(17), 42(3), 43(95), 44(7), 45(31), 55(7), 57(5), 59(8), 69(100), 70(7), 71(19), 73(3), 74(14), 87(34), 97(8), 146(16)
Ethyl-2-acetohydroxy-2-oxooctonate	230	1483	27(4), 29(15), 39(3), 41(12), 42(3), 43(64), 44(6), 45(3), 55(95), 56(6), 57(6), 59(4), 69(25), 71(5), 74(7), 91(6), 97(100), 98(8), 115(21), 146(11)

4 Core biosynthesis of lycopsamine type pyrrolizidine alkaloids in *Symphytum officinale* – a metabolomic approach

Core biosynthesis of lycopsamine type pyrrolizidine alkaloids in *Symphytum officinale* – a metabolomic approach

Annika Engelhardt^a, Riya Christina Menezes^b, Aleš Svatoš^b and Dietrich Ober^a

^aBotanisches Institut und Botanischer Garten, Christian-Albrechts-Universität zu Kiel, D-24098 Kiel, Germany

^bMax Planck institute for chemical ecology, D-07745 Jena, Germany

Detailed author contributions are listed at the end of the thesis.

4.1 Abstract

Symphytum officinale (*S. officinale*, Boraginaceae) produces pyrrolizidine alkaloids (PAs) of the lycopsamine type with a unique aliphatic branched-chain C₇-necic acid, 2,3-dihydroxy-2-isopropylbutyric acid. Recently, the enzyme catalyzing the formation of the C₇-backbone, referred to as C₇-pronecic acid, was identified to be an acetohydroxyacid synthase with extended substrate specificity, the C₇-hydroxyacid synthase (C₇HAS). A CRISPR/Cas9-mediated approach resulted in the inactivation of the gene encoding C₇HAS in hairy roots that proved to be unable to produce lycopsamine type PAs. Instead, these hairy roots accumulated trachelanthamidine, the saturated necine base moiety of PAs. These results contradicted previous feeding studies which suggested not trachelanthamidine but rather the unsaturated necine base supinidine as the partner for *O*⁹ esterification with the necic acid moiety. By comparing the metabolomic profiles of transgenic hairy roots of *S. officinale* with inactivated C₇HAS to control lines (empty vector control) we wanted to identify intermediates of core biosynthesis up to intermedine *N*-oxide, the simplest lycopsamine type PA without further chemical diversification in *S. officinale*. More specifically, we wanted to determine the sequence of (1) reduction of the necic acid, (2) *O*⁹ esterification, (3) *N*-oxygenation, (4) 1,2-desaturation, and (5) C⁷ hydroxylation that follow the formation of trachelanthamidine and C₇-pronecic acid.

Extracts of a total of four different knock-out and five control line hairy roots have been used for a non-targeted UHPLC-ESI MS/MS approach with positive and negative ionization mode. Mass features exclusively present or increased in the control lines over the knock-out lines were identified by manually and computationally analyzing the high-resolution mass spectra and tandem mass spectra. The data confirmed a massive accumulation of the saturated necine base trachelanthamidine in the knock-out hairy roots. Additionally, the three possible intermediates of PA biosynthesis trachelanthamine, the indistinguishable trachelanthamine *N*-oxide or 7-hydroxytrachelanthamine, and 7-hydroxytrachelanthamine *N*-oxide have been identified and are detectable exclusively in control lines. The detectability of these intermediates suggests the following sequence of enzymatic steps in PA biosynthesis in *S. officinale*: Reduction of C₇-pronecic acid, *O*⁹ esterification of trachelanthamidine with the C₇-necic acid, C⁷ hydroxylation and *N*-oxygenation in unexplained order, and finally formation of the toxic 1,2-unsaturated PAs by desaturation in 1,2-position of the necine base.

4.2 Introduction

Within the pyrrolizidine alkaloids (PAs) one major group, the lycopsamine type PAs, are represented by monoesters or open-chain diesters with a unique aliphatic branched-chain C₇-necic acid, 2,3-dihydroxy-2-isopropylbutyric acid (Hartmann and Witte, 1995; Frölich et al., 2007; El-Shazly and Wink, 2014). PAs of this type have been detected in representatives of the tribe Eupatorieae (Asteraceae) and in the families Apocynaceae, Heliotropiaceae and Boraginaceae (Hartmann and Witte, 1995; Ober et al., 2000). The PA profile of *Symphytum officinale* (*S. officinale*, Boraginaceae, Boraginales) shows a mixture of retronecine O⁹-monoesters and O⁷-O⁹-diesters of the lycopsamine type (Figure 23). In recent years, many efforts have been made to elucidate the biosynthetic steps that ultimately lead to the formation of the core structure of lycopsamine type PAs, the lycopsamine/intermediate N-oxide (Weber et al., 1999; Kruse et al., 2017; Kruse et al., 2019; Zakaria et al., 2021; Zakaria et al., 2022) (Chapter 3 this work). The present knowledge on PA biosynthesis *S. officinale* is summarized in the green highlighted part of Figure 24. Homospermidine synthase (HSS) was identified as the first and key enzyme in the biosynthesis of necine bases (Böttcher et al., 1993; Ober and Hartmann, 1999; Ober and Kaltenegger, 2009). The fate of the homospermidine has long been speculated on and has only recently been elucidated using *in vitro* enzyme assays and virus-induced gene silencing in *H. indicum* (Heliotropiaceae, Boraginales) (Zakaria et al., 2022) and was confirmed by CRISPR/Cas9 mediated genome editing in *S. officinale* (Zakaria, 2022). A homospermidine oxidase (HSO) deaminates both primary amino groups and cyclizes the molecule to 1-formylpyrrolizidine. An alcohol dehydrogenase is postulated to be involved in the conversion of 1-formylpyrrolizidine to the saturated necine base trachelanthamidine (Robins, 1982; Veverka et al., 1997) but this step has not yet been elucidated on enzyme level. However, the fact that this reduction undoubtedly occurs prior to the O⁹-esterification of the necic acid was confirmed by two observations. Firstly, feeding a HSS knock-out hairy root of *S. officinale* with ¹³C-labeled homospermidine allowed to detect labelled intermediates of the pathway and confirmed the formation of trachelanthamidine, which was successively incorporated into complex lycopsamine type PAs without detectable intermediates (Zakaria et al., 2021). Secondly, trachelanthamidine was detected as an accumulating intermediate in hairy roots, in which C₇-hydroxyacid synthase (C₇HAS), an enzyme involved in the formation of the C₇-necic acid characteristic for lycopsamine type PAs, was knocked-out (Chapter 3). This knock-out provided *in planta* evidence for the involvement of C₇HAS in PA biosynthesis. By transferring an active acetaldehyde from pyruvate to 2-oxoisovalerate, C₇HAS catalyzes the formation of the direct precursor of the C₇-necic acids, C₇-pronecic acid (Chapter 3). The

accumulation of trachelanthamidine contradicted previous feeding studies which suggested the unsaturated necine base supinidine as the most likely partner for O^9 esterification with the necic acid moiety (Frölich et al., 2007). In the same study, Frölich et al. (2007) were also able to show that PA biosynthesis in *S. officinale* proceeds so slowly that 17 days after feeding with radiolabeled putrescine, early intermediates could still be detected, whereas in other analyzed PA-producing plants these were only detectable for the first 2-4 days. This makes *S. officinale* an optimal candidate to identify traces of possible intermediates of PA biosynthesis.

By comparing the metabolomic profiles of transgenic hairy roots of *S. officinale* with inactivated C₇HAS and control lines (empty vector control) we want to identify intermediates of the core biosynthesis up to intermediate *N*-oxide, the simplest lycopsamine type PA without further chemical diversification in *S. officinale* (Figure 23). With this the sequence of the biosynthetic steps that follow the formation of trachelanthamidine and C₇-pronecic acid, *i.e.* (1) reduction of the necic acid, (2) O^9 -esterification, (3) *N*-oxygenation, (4) 1,2-desaturation, and (5) C⁷-hydroxylation should be determined.

Here we present the results of a non-targeted UHPLC-ESI MS/MS approach with positive and negative ionization mode. Mass features of potential intermediates exclusively present or increased in the control lines over the knock-out lines were identified by manually and computationally analyzing the high-resolution mass spectra and tandem mass spectra.

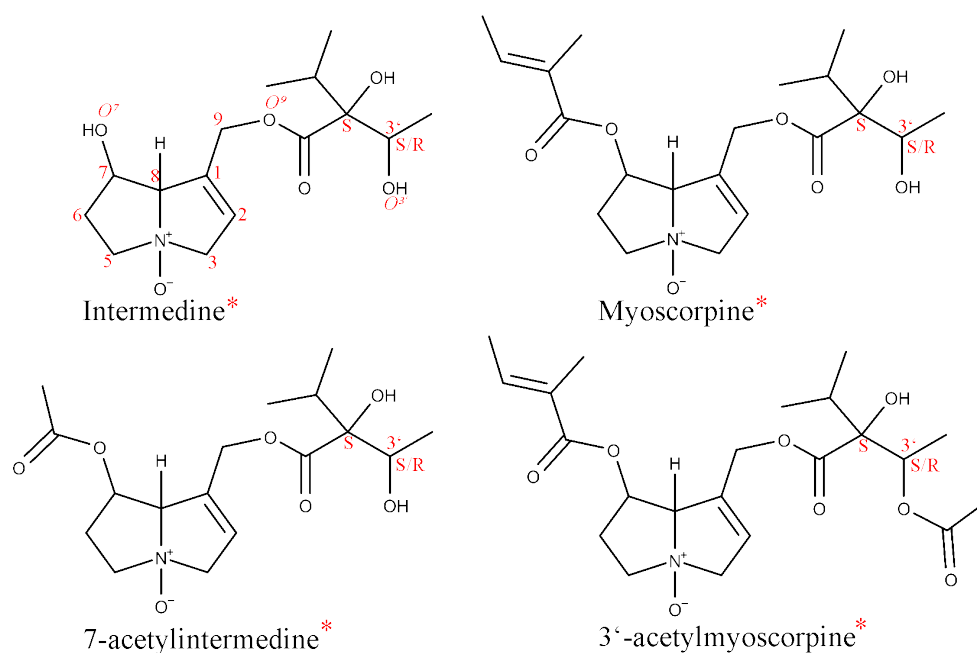
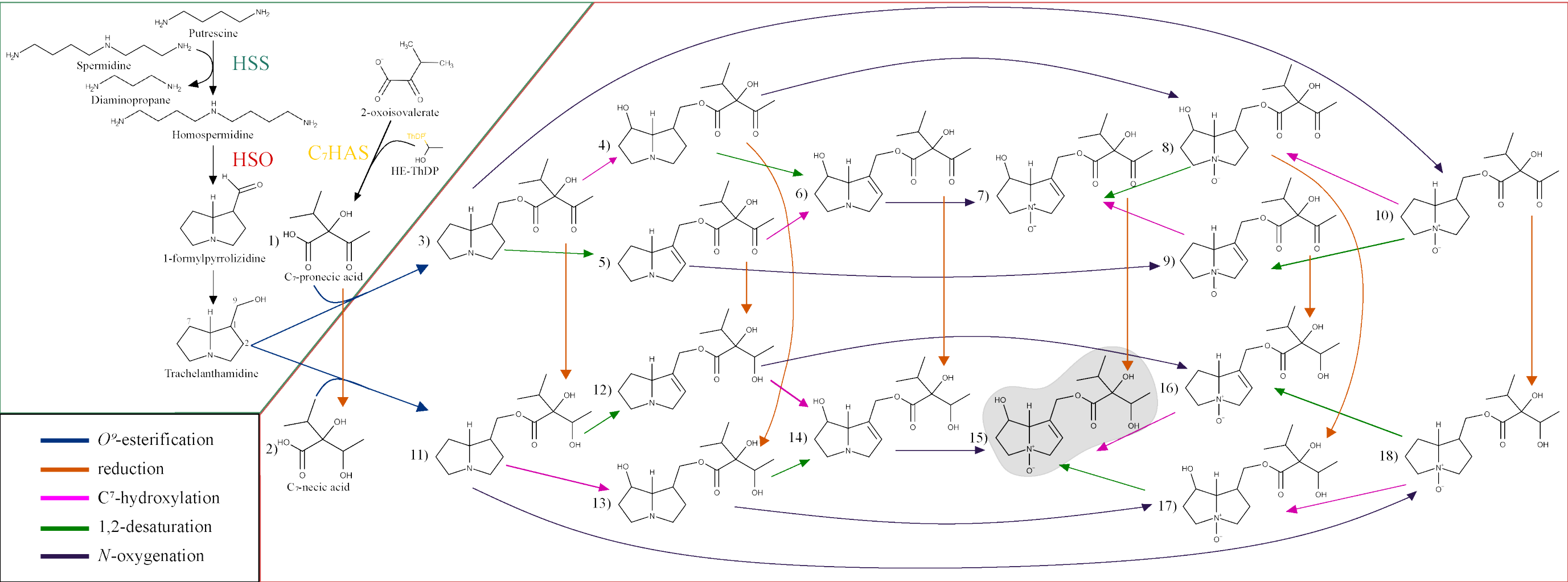


Figure 23 Retronecine O^9 -monoesters and O^7 -, O^9 -diesters of the lycopsamine type described for *Symphytum officinale*. *PAs chosen here could synonymously be one of several stereoisomers, *i.e.*, intermedine or lycopsamine *N*-oxide; 7-acetylintermedine or 7-acetyllycopsamine *N*-oxide; myoscorpine or symmlandine/ echiumine/ symphytine *N*-oxide, and 3'-acetylmyoscorpine or 3'-acetylsymmlandine/ 3'-acetylechiumine/ 3'-acetylsymphytine *N*-oxide.

4.3 Results

4.3.1 Hypothetical biosynthetic pathway of PA core biosynthesis

PA biosynthesis in *S. officinale* is far from being understood; only the first steps are described. C₇HAS catalyzes the biosynthesis of the characteristic C₇-skeleton of the necic acids of lycopsamine type PAs by producing the so-called C₇-pronecic acid that requires further reduction to result in the formation of (-)-viridfloric and (+)-trachelanthic acid ('C₇-necic acids', Chapter 3). HSS and HSO catalyze the formation of the bicyclic intermediate 1-formylpyrrolizidine (Zakaria et al., 2021, 2022) that will be reduced to the necine base trachelanthamidine. Knock out-lines with inactivated C₇HAS accumulated trachelanthamidine, suggesting that the saturated necine base is an intermediate in PA biosynthesis in *S. officinale* and that this reduction occurs before the necine base is esterified with the necic acid (Chapter 3). The sequence of all following steps remains an open question. Figure 2 summarizes all hypothetical pathways for the later steps of the biosynthesis of the lycopsamine type PA core structure, lycopsamine *N*-oxide (Figure 24, compound 15). Marked in green are the known parts of PA biosynthesis with HSS, HSO, and C₇HAS. Red labels that part of the pathway that is not described yet involving the five chemical reactions. Starting with the *O*⁹-esterification of either C₇-pronecic acid or C₇-necic acid with trachelanthamidine, all possible sequences of the five required chemical reactions, C⁷-hydroxylation, reduction of the necic acid moiety, 1,2-desaturation, and *N*-oxygenation. (Figure 24). To simplify the illustration, the stereochemistry of the compounds is not shown and the structures are representatives of all possible stereoisomers. The table in Figure 2 summarizes the molecular formula and the calculated high-resolution mass of the protonated (m/z [M+H]⁺) and deprotonated (m/z [M-H]⁻) putative intermediates resulting from positive ion mode electrospray ionization (ESI+) and negative ion mode electrospray ionization (ESI-), respectively. The protonated necine base fragment ion is a PA defining feature of fragmentation and thus serves to accurately assign mass features as PAs. As such, the expected m/z [M+H]⁺ of the necine base fragment ion of the hypothetical intermediates is also given in the table in Figure 2.



No	1	2	3	4	5	6	7	8	9
Formula	C ₇ H ₁₂ O ₄	C ₇ H ₁₄ O ₄	C ₁₅ H ₂₅ NO ₄	C ₁₅ H ₂₅ NO ₅	C ₁₅ H ₂₃ NO ₄	C ₁₅ H ₂₃ NO ₅	C ₁₅ H ₂₃ [N ⁺]O ₅ [O ⁻]	C ₁₅ H ₂₅ [N ⁺]O ₅ [O ⁻]	C ₁₅ H ₂₃ [N ⁺]O ₄ [O ⁻]
m/z [M+H] ⁺	-	-	284.18563	300.18054	282.16998	298.16489	314.15981	316.17546	298.16489
m/z [M-H] ⁻	159.0652	161.0808	-	-	-	-	-	-	-
Base m/z [M+H] ⁺	-	-	142.1226	158.1175	140.1070	156.1019	172.0968	174.1125	156.1019

No	10	11	12	13	14	15	16	17	18
Formula	C ₁₅ H ₂₅ [N ⁺]O ₄ [O ⁻]	C ₁₅ H ₂₇ NO ₄	C ₁₅ H ₂₅ NO ₄	C ₁₅ H ₂₇ NO ₅	C ₁₅ H ₂₅ NO ₅	C ₁₅ H ₂₅ [N ⁺]O ₃ [O ⁻]	C ₁₅ H ₂₅ [N ⁺]O ₄ [O ⁻]	C ₁₅ H ₂₇ [N ⁺]O ₅ [O ⁻]	C ₁₅ H ₂₇ [N ⁺]O ₄ [O ⁻]
m/z [M+H] ⁺	300.18054	286.20128	284.18563	302.19619	300.18054	316.17546	300.18054	318.19111	302.19619
m/z [M-H] ⁻	-	-	-	-	-	-	-	-	-
Base m/z [M+H] ⁺	158.1175	142.1226	140.1070	158.1175	156.1019	172.0968	156.1019	174.1125	158.1175

Figure 24 Hypothetical biosynthetic routes of the pathway to intermedine *N*-oxide, the core structure of lycopsamine type PAs in *S. officinale*. Characterized steps catalyzed by homospermidine synthase (HSS), homospermidine oxidase (HSO), and C₇-hydroxyacid synthase (C₇HAS) are highlighted in green. The red part represents the hypothetical biosynthetic routes with colored arrows representing one of the five required modifications: *O*⁹-esterification of necine base and necic acid, reduction of the necic acid moiety, hydroxylation at the C⁷, 1,2-desaturation, and *N*-oxygenation. In the table, the molecular formula and the respective high-resolution *m/z* are given protonated [M+H]⁺ for the analysis in positive ion mode and deprotonated [M-H]⁻ for the analysis in negative ion mode. The high-resolution *m/z* of the base fragment ions of the intermediates 3-18 are given as adducts [M+H]⁺. To simplify the illustration, the stereochemistry of the compounds is not shown and the structures are representatives of all possible stereoisomers.

4.3.2 Metabolic profiles of knock-out lines and control lines show distinct differences

To elucidate the sequence of biosynthetic steps in PA biosynthesis in *S. officinale*, the metabolites of four hairy root knock-out lines of *SoC₇HAS* (KO; A29, A37, A50 and A69, Chapter 3) and five empty vector control lines (CL; CL2, CL4, CL11, CL17, CL34, Chapter 3) were extracted. The resulting extracts were subjected to non-targeted analyzes using ultra-high-performance liquid chromatography ESI tandem mass spectrometry (UHPLC-ESI-Q Exactive MS/MS) with ESI+ and ESI-. The result files were converted to mzXML format using ProteoWizard 3.0 X software (Chambers et al., 2012) and subsequently processed with XCMSTM Online (Tautenhahn et al., 2012). In this process, peaks were detected and a nonlinear retention time alignment was performed to obtain metadata tables where peaks with a unique *m/z* ratio and retention time, termed mass features, and associated signal intensities are given. A total of 38439 mass features could be identified in the ESI+ measurements, and a total of 17224 mass features could be identified in the ESI- measurements. To evaluate the differences of metabolite production of the different hairy root lines, unsupervised multivariate principal component analyses (PCA) were conducted by analyzing the mass features for the positively and negatively ionized samples separately using the statistical analysis tool for metadata tables in Metaboanalyst5 (Pang et al., 2021; Pang et al., 2022). The resulting PCA plots showed a clear separation of detected metabolites of KO and CL. Figure 25 A shows the PCA of the protonated metabolites. Here PC1 and PC2 explained 67.5 % and 20.7 % of the total variation, respectively. The metabolic profiles of the CL separated along PC1 but not PC2. The metabolic profiles of the KO lines separated along PC1 and PC2 within the group but still show a distinct cluster separate from that of the CL. The PCA of the deprotonated metabolites (Figure 25 B) shows that PC1 explained 45 % and PC2 17.5 % of the total variation within the metabolic profiles. Again, the profiles of the CL separated along PC1 but hardly along PC2. Remarkably the KO lines do not separate along PC1 but along PC2. In total these results clearly show differences of metabolic profiles of KO and CL hairy roots for both ionization modes, confirming the classification into two separate groups based on the GC-MS analysis (Chapter 3).

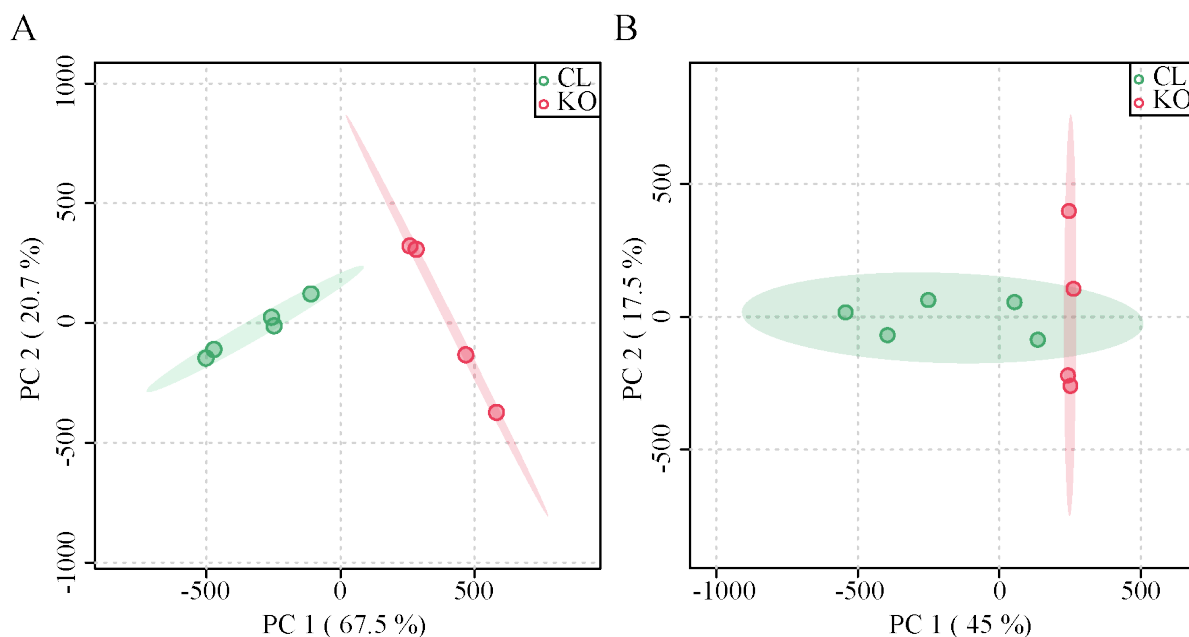


Figure 25 Metabolomic profiles of different *S. officinale* hairy root knock-out lines (KO, n=4) and control lines (CL, n=5). Ellipses show the 95 % confidence regions. **A** PCA scores plot of metabolomic profiles after positive ionization. **B** PCA scores plot of metabolomic profiles after negative ionization.

4.3.3 Production of PAs represents a contributing factor to differences in the metabolite profiles of CL and KO

Using PCA, we have shown that the metabolomic profiles of CL and KO differ and that the lineages of each treatment cluster together. Next, we wanted to show that PAs are a contributing factor to these differences, and that we can distinctively identify PAs with the approach chosen here. Intermedine *N*-oxide, 7-acetylintermedine *N*-oxide, myoscorpine *N*-oxide and 3'-acetylmyoscorpine *N*-oxide, typical PAs in *S. officinale* (Figure 23) (Hartmann and Witte, 1995; El-Shazly and Wink, 2014)(Chapter 3), should serve as a positive control. All these PAs occur in different, indistinguishable stereoisomeric forms. Thus, the PAs chosen here could synonymously be one of several stereoisomers, *i.e.*, intermedine or lycopsamine *N*-oxide; 7-acetylintermedine or 7-acetyllycopsamine *N*-oxide; myoscorpine or symlandine/ echiumine/ symphytine *N*-oxide, and 3'-acetylmyoscorpine or 3'-acetylsymlandine/ 3'-acetylechiumine/ 3'-acetylsymphytine *N*-oxide. For clarity, only one PA is mentioned as a placeholder for all possible stereoisomers.

We used the following workflow to identify analytes:

- (1) manually analyzing the high-resolution mass spectra for accurate masses with Xcalibur™, including molecular formula prediction

- (2) computational data analysis with Sirius4 (Dührkop et al., 2019) to confirm molecular formula (detailed results, including Sirius score, tree score, explained peaks and total explaining are summarized in Table 2)
- (3) manually analyzing the higher energy collision induced dissociation (HCD) tandem mass spectra of the accurate masses with Xcalibur™
- (4) computational tandem mass spectrum analysis using CSI (Compound Structure Identification):FingerID (Dührkop et al., 2015; Hoffmann et al., 2021) with structure prediction by fragmentation tree prediction and structure database comparisons (CSI:FingerID similarity, rank and score are summarized in Table 2)
- (5) comparison of extracted ion chromatogram (EIC) signal intensities of mass features according to metadata table generated with XCMS™ Online to check if these compounds were exclusively present or increased in the CL in comparison to the KO.

Figure 26 A shows that a m/z 316.1751 could be found in positively ionized extracts of CLs that would fit intermedine *N*-oxide (calculated m/z $[M+H]^+$ 316.1754). The corresponding predicted molecular formula of $C_{15}H_{26}NO_6$ was confirmed by Sirius4 as a first hit and a score of 98.8 %. The HCD tandem mass spectrum confirmed a precursor of m/z 316.1754 and a fragment ion of m/z 172.0969 ($C_8H_{14}NO_3$) corresponding to the m/z $[M+H]^+$ of the necine base retronecine *N*-oxide. The computational tandem mass spectrum analysis using CSI:FingerID further supported the assignment of this compound as intermedine *N*-oxide at rank one with 100 % similarity. In the same way, the mass features of 7-acetylintermedine *N*-oxide, myoscorpine *N*-oxide and 3'-acetylmyoscorpine *N*-oxide were analyzed (Figure 26 B, C and D respectively). The calculated m/z $[M+H]^+$ for each PA *N*-oxide (358.1860, 398.2173, and 440.2278, respectively) was identified in the high-resolution MS and the predicted molecular formula was confirmed using SIRIUS. In the HCD tandem mass spectrum these masses were confirmed as precursor masses. As in the case of the intermedine *N*-oxide, the O^9 -ester is constantly cleaved off during fragmentation, which is reflected in the characteristic fragment ion of the necine bases. For 7-acetylintermedine *N*-oxide, this corresponded to the fragment ion m/z 214.1074 for the O^7 -acetyl ester of retronecine *N*-oxide ($[M+H]^+$ 214.1073) and additionally in the fragment ion of m/z 172.0968 of retronecine *N*-oxide when both esters are cleaved off. For myoscorpine *N*-oxide and 3'-acetylmyoscorpine *N*-oxide the fragment ion of m/z 254.1387 corresponded to the O^7 -tigloyl ester of retronecine *N*-oxide ($[M+H]^+$ 254.1387). The assignment of the features as 7-acetylintermedine *N*-oxide and myoscorpine *N*-oxide was supported with respective similarities of 96.05 % and 98.15 % at rank 1 by the computational

tandem mass spectrum analysis using CSI:FingerID. Unfortunately, there is no entry in the databases for 3'-acetylmyoscorpine *N*-oxide or any of its structural homologues. Thus, the analysis of the tandem mass spectrum with CSI:FingerID did not yield a hit for 3'-acetylmyoscorpine *N*-oxide. It was however identified as 3''-acetylheliosupine, a PA with the same molecular formula, with 90.99% similarity at rank 1. Since this compound is not present in *S. officinale*, it is assumed the identified mass feature is indeed 3'-acetylmyoscorpine *N*-oxide. The comparison of the EIC signal intensities of the mass features showed that all PA *N*-oxides are exclusively present in the CLs and not produced in the KO.

The PAs described here could not be detected as mass features in their tertiary form in the high-resolution MS. For some of them traces in the tandem mass spectral analysis could be attributed as PAs. However, these traces were attributed to a methodological bias due to the extraction of metabolites, since PAs have been shown to be stored as *N*-oxides in the plant. This is of particular importance as it also excluded lycopsamine (Figure 24, compound 14) as an intermediate of the PA biosynthesis.

Applying our workflow for the identification of PA *N*-oxides in the metabolomes of CL and KO provides insights that will be helpful in identifying intermediates of the biosynthetic pathway: (1) The PAs are present as hydrogen adducts and their calculated protonated m/z ratios can be identified in high-resolution MS with minimal variations in the ppm range. (2) The m/z $[M+H]^+$ also represents the precursor mass in the HCD tandem mass spectrometry. (3) PAs show a characteristic fragmentation pattern. Regardless of the number of esters and the type of esterified moieties, the O^9 -ester bond with the C₇-necic acid is always cleaved, resulting in the fragment ion of the associated base. (4) The computational tandem mass spectrum analysis using CSI:FingerID achieves very high similarities with structure database entries, as long as the structures are existing in the databases. (5) Confirming our fundamental assumption, the signal intensities of PA mass features are only detectable in the CL and not in the KO.

In order to confidently identify a mass feature as an intermediate of the PA biosynthetic pathway, as many of these points as possible should be fulfilled.

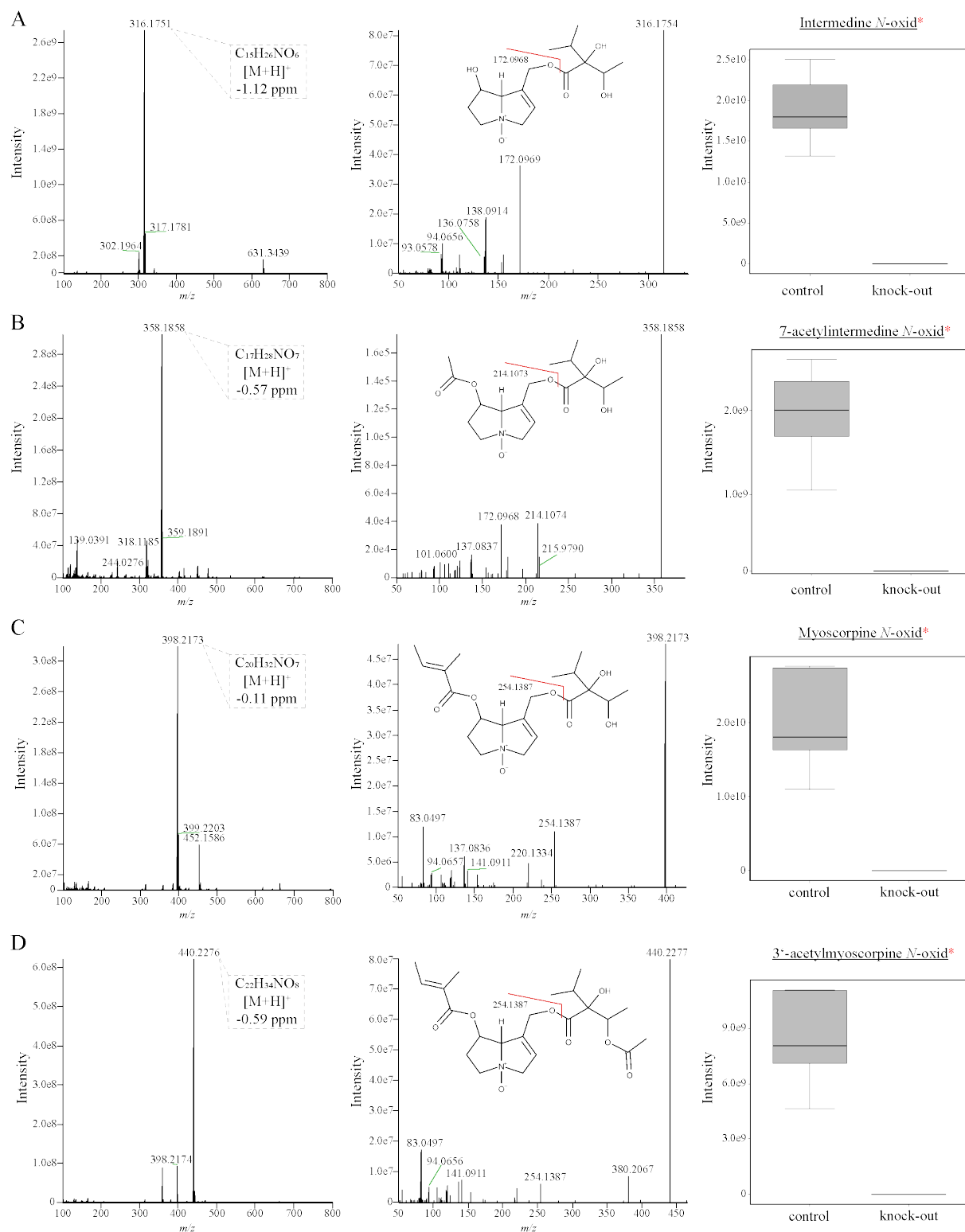


Figure 26 PA *N*-oxides identified in *S. officinale* control line hairy roots are absent in knockout lines. High resolution mass spectra, HCD MS/MS and box and whisker plots of EIC intensities of lycopsamine *N*-oxide (A, m/z $[M+H]^+$ 316.1751), 7-acetyllycopsamine *N*-oxide (B, m/z $[M+H]^+$ 358.186), myoscorpine *N*-oxide (C, m/z $[M+H]^+$ 398.2173), and 3'-acetylmyoscorpine *N*-oxide (D, m/z $[M+H]^+$ 440.2278), respectively. EIC intensities are given for CL (n=5) and KO (n=4). *or one of its indistinguishable stereoisomeric forms.

Table 2 Computational analysis results of SIRIUS4 and CSI:FingerID

Metabolite	Ion	SIRIUS 4					CSI:FingerID			
		Molecular Formula	Score [%]	Tree score [%]	Explained peaks	Total explaining [%]	Identified PA	Similarity [%]	Rank	Score
Intermedine <i>N</i> -oxide	[M+H] ⁺	C ₁₅ H ₂₆ NO ₆	98.8	44.0	25/59	87.45	Intermedine <i>N</i> -oxide	100	1	-3.228
7-Acetylintermedine <i>N</i> -oxide	[M+H] ⁺	C ₁₇ H ₂₈ NO ₇	99.36	67.99	33/56	86.45	7-Acetylintermedine <i>N</i> -oxide	96.05	1	-23.576
Myoscorpine <i>N</i> -oxide	[M+H] ⁺	C ₂₀ H ₃₂ NO ₇	93.51	50.89	26/56	77.92	Myoscorpine <i>N</i> -oxide	98.15	1	-13.041
3'-Acetylmyoscorpine <i>N</i> -oxide	[M+H] ⁺	C ₂₂ H ₃₄ NO ₈	82.54	62.42	33/58	89.11	3''-Acetylheliosupine	90.99	1	-66.254
Trachelanthamidine	[M+H] ⁺	C ₈ H ₁₆ NO	100	51.89	11/58	48.32	Trachelanthamidine	53.42	14	-85.281
C ₇ -Necic acid (+)-trachelanthic or (-)-viridifloric acid	[M-H] ⁻	C ₇ H ₁₃ O ₄	-	-	-	-	-	-	-	-
Trachelanthamine	[M+H] ⁺	C ₁₅ H ₂₇ NO ₄	99.99	53.14	15/57	86.73	Trachelanthamine	92.49	1	-51.453
7-Hydroxytrachelanthamine or Trachelanthamine <i>N</i> -oxide	[M+H] ⁺	C ₁₅ H ₂₈ NO ₅	98.95	40.97	22/58	73.66	Trachelanthamine <i>N</i> -oxide	86.47	1	-54.199
7-Hydroxytrachelanthamine <i>N</i> -oxide	[M+H] ⁺	C ₁₅ H ₂₈ NO ₆	99.91	50.99	25/59	89.13	Dihydroindicine <i>N</i> -oxide	78.74	1	-99.371

4.3.4 Accumulation of trachelanthamidine in knock-out lines of *SoC7HAS*

GC-MS analyses of the *SoC7HAS* KO hairy roots revealed that they accumulate an 1,2-saturated necine base that was identified as trachelanthamidine. In contrast, no necine bases were identified in CLs in this analysis (Chapter 3). A clear disadvantage of GC-MS analyses compared to LC-MS techniques is a limited analyte spectrum as a prerequisite for GC-MS analysis is that the substance to be analyzed can be transferred to the gas phase without decomposing (Güssregen, 2019). For example, PA *N*-oxides cannot be measured by GC-MS and are reduced to their tertiary form prior to analysis. We wanted to verify the GC-MS results and check whether other necine bases or their *O*⁷-tigloyl or -acetyl esters, as listed in Supplementary Table 7, are detectable in the hairy roots by LC-MS.

Figure 27 shows that two mass features were identified that based on their masses, could be necine bases. For trachelanthamidine (m/z $[M+H]^+$ 142.1226), a mass feature with a m/z 142.1224 in the high-resolution mass spectrum and an attributed molecular formula of $C_8H_{16}NO$ was detected (Figure 27 A). The molecular formula was confirmed by Sirius as first hit with a score of 100 %. The tandem mass spectrum showed a distinct ion fragment peak at m/z 124.1123, which is obtained by the cleavage of the C⁹-hydroxy group. Analysis of the tandem mass spectrum with CSI:FingerID weakly supported the assignment as trachelanthamidine at rank 14 with 53.42 % similarity. A comparison of the EIC signal identities shows an accumulation of trachelanthamidine in the KO to levels that are in average 2032-fold higher than those of the CL where only traces were detectable. Figure 5 B shows an example of the importance of the combination of high-resolution mass spectra and tandem mass spectra. In this case, a mass feature with m/z 158.1178 with a predicted molecular formula of $C_8H_{16}NO_2$, was confirmed by Sirius with a score of 100 % was detected. This would correspond exactly to trachelanthamidine *N*-oxide or C⁷-hydroxylated trachelanthamidine. The occurrence of this derivative and the accumulating trachelanthamidine in both groups, the KO and the CL, was unexpected. Analyzing the data of the tandem mass spectrum revealed that the fragment ions did not match the possible necine bases and their derivatives. According to the prediction via CSI:FingerID this compound is more likely a piperidine derivate and was therefore dismissed for our targeted analyses. A summary the of computational data analysis with Sirius4 and CSI:FingerID can be found in Table 2.

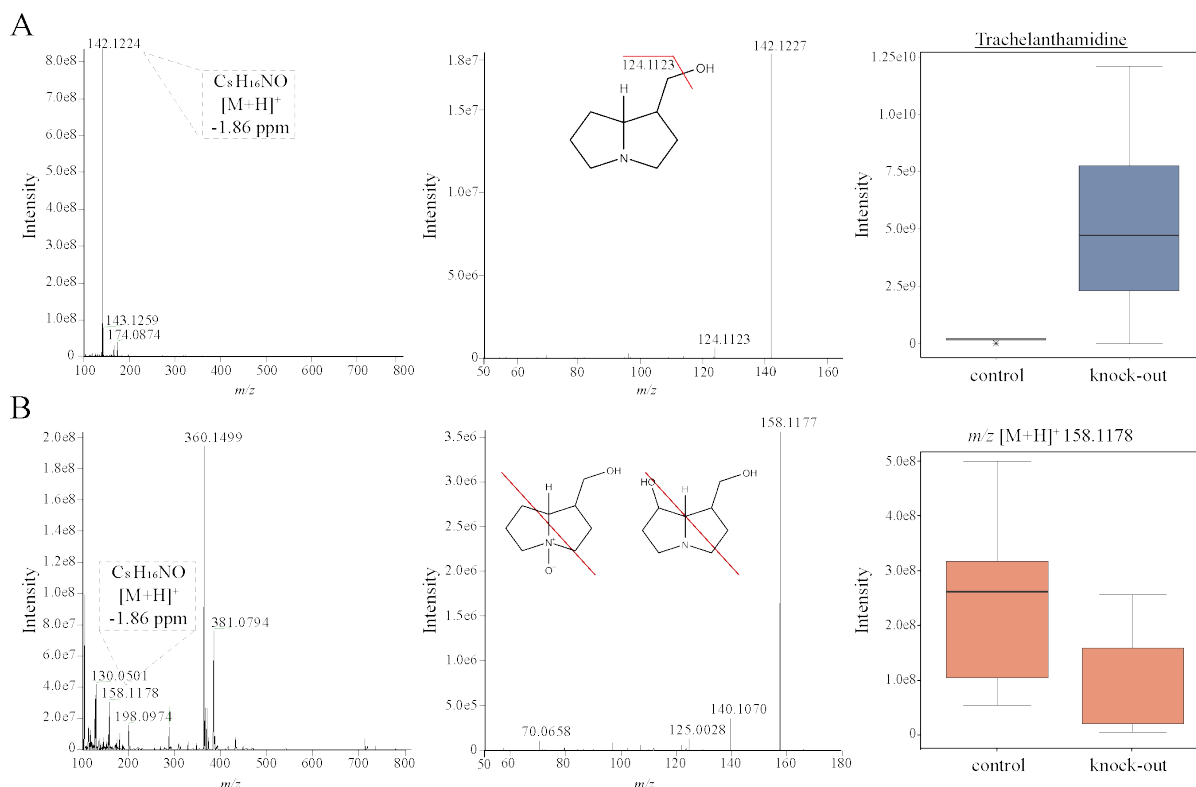


Figure 27 Necine base mass features identified in *S. officinale* control line hairy roots. High resolution mass spectra, HCD tandem mass spectra and box and whisker plots of EIC intensities of trachelanthamidine (**A**, m/z $[M+H]^+$ 142.1224) and dismissed putative necine base (**B**, m/z $[M+H]^+$ 158.1178). EIC intensities are given for CL (n=5) and KO (n=4).

4.3.5 C₇-necic acid is esterified with trachelanthamidine

The accumulation of an intermediate implies that the reaction partner for the next step in the pathway is missing. Since the biosynthesis of the C₇-necic acid moiety was interrupted in the KO due to the CRISPR/Cas9-mediated inactivation of the enzyme C₇HAS, it can be concluded that the necine base trachelanthamidine is the esterification partner for the product of necic acid biosynthesis. Here, two esterification partners are possible: C₇-pronecic acid or C₇-necic acid (Figure 24 compound 1 and 2, respectively). Since most acidic analytes are more efficiently deprotonated than protonated (Henriksen et al., 2005), we applied our targeted approach workflow to the ESI- deprotonated metabolites.

The C₇-pronecic acid with an accurate mass of $[M-H]^-$ m/z 159.0652 could not be identified in the CL indicating an efficient reduction to the C₇-necic acid. The C₇-necic acid with a calculated m/z 161.0808 ($[M-H]^-$) could only be identified in the manual analysis of the MS¹ and MS² data (Figure 28) that could not be confirmed in the computational analysis. As could be seen in Figure 28 a mass feature with a m/z 161.0808 was detectable in the high-resolution mass spectrum with a predicted molecular formula of C₇H₁₃O₄ fitting that of the C₇-necic acid. The

tandem mass spectrum also has the corresponding precursor ion, and the fragment ion of m/z 117.0546 would have been formed by the splitting off the acetaldehyde group. These results are however not supported by the computational analysis of the data. As neither processing the data with Sirius nor the subsequent analysis with CSI:FingerID confirmed this data, no definitive statement can be made about C₇-necic acid as an intermediate, even though the manual analysis strongly indicates the presence of C₇-necic acid, solely detected in the control lines.

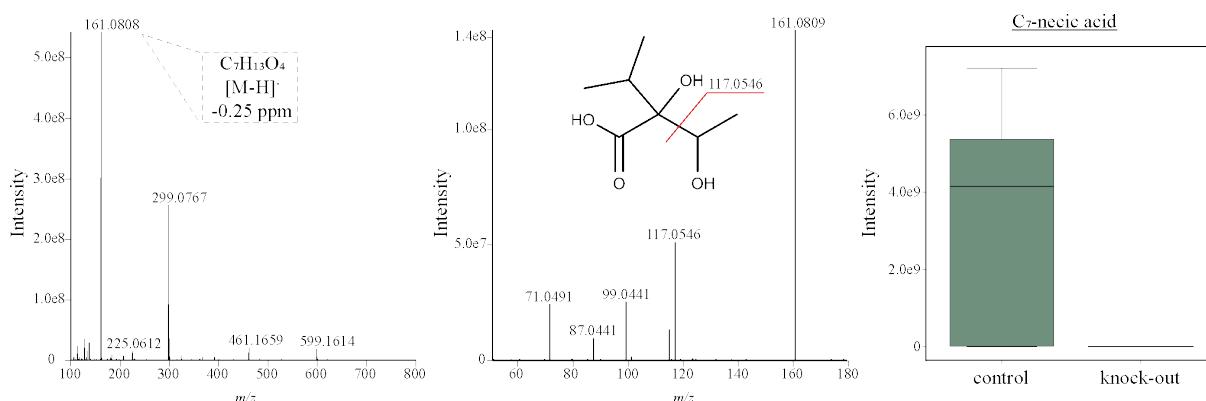


Figure 28 C₇-necic acid identification in *S. officinale* CL. High resolution mass spectrum, HCD tandem mass spectrum and box and whisker plots of EIC intensities (CL (n=5) and KO (n=4)) of C₇-necic acid (m/z [M-H]⁻ 161.0808).

4.3.6 Identification of potential intermediates

Taking a closer look at all hypothetical pathways for the later steps of the biosynthesis of the lycopsamine type PA core structure as shown in Figure 24, we wanted to see if we could identify the corresponding hypothetical intermediates by applying our targeted approach workflow. Since some of the intermediates, such as 8 and 15, have the same molecular formula but a different structure, the exact assignment of the base ions was crucial for identification.

By analyzing the protonated structures 3-10, resulting from the *O*⁹-esterification of trachelanthamidine with the C₇-pronecic acid, only intermediate 8 could be identified as mass feature. Based on the necine base ion however, this mass feature was assigned to intermediate *N*-oxide (structure 15) before (Figure 26 A) and structures 3-10 could be discarded as unlikely intermediates of PA biosynthesis in *S. officinale*. This decision was supported by the analysis of the mass features for the potential intermediates 11-18, starting from the esterification of trachelanthamidine and the C₇-necic acid. Three mass features with m/z 286.2012, 302.1962, and 318.1911 in the high-resolution mass spectrum could be identified that would fit the intermediates 11, 13 or 18 and 17, respectively. Note that when naming the intermediates, we

decided to use the most parsimonious pathway, which does not require any additional and unnecessary isomerization to the known PAs from *S. officinale*. In principle, however, it could also be one of the stereochemical isomers of these compounds.

Figure 29 A shows the high-resolution mass spectrum of the mass feature fitting intermediate 11 (= trachelanthamine, $[M+H]^+$ m/z 286.2012) and the attributed molecular formula $C_{15}H_{27}NO_4$. This molecular formula was confirmed by Sirius as first hit with a Sirius score of 99.99 %. The tandem mass spectrum showed a distinct ion fragment peak at m/z 142.1227, which corresponds to the appropriate protonated base trachelanthamidine. With 92.49 % similarity (rank 1) CSI:FingerID analysis of the tandem mass spectrum further supported the assignment of this feature as trachelanthamine.

The analysis of the mass feature with m/z 302.1962 has shown that it is either intermediate 13 or 18 (7-hydroxytrachelanthamine or trachelanthamine *N*-oxide respectively (Figure 29 B). Both compounds share not only the same molecular formula, but also the expected m/z of the necine bases are identical despite the different structures. The attributed molecular formula of $C_{15}H_{28}NO_5$ was confirmed by Sirius as first hit with a Sirius score of 98.952 %. The tandem mass spectrum showed a distinct ion fragment peak at m/z 158.1177 ($C_8H_{16}NO_2$) fitting the protonated masses of trachelanthamidine *N*-oxide and 7-hydroxytrachelanthamidine. CSI:FingerID supported the assignment of this feature as trachelanthamine *N*-oxide (intermediate 18) with 86.74 % (rank 1). The structure of 7-hydroxytrachelanthamidine is not available in the searched databases, making a comparison impossible and thus not giving a support by CSI:FingerID. Of note: The most logical step to clarify whether the mass feature is trachelanthamidine *N*-oxide or 7-hydroxytrachelanthamidine would be to add zinc to the extracts and thus reduce the *N*-oxides to their tertiary form. Previous attempts to do this either showed incomplete reduction of the PA *N*-oxides, or the control samples were partially reduced in transit (data not shown). In both cases an accurate determination of the EIC signal intensities was impossible.

Figure 29 C demonstrates that the mass feature with an m/z 318.1911 could be assigned to intermediate 17 (7-hydroxytrachelanthamine *N*-oxide, m/z $[M+H]^+$ 318.1911). The molecular formula of $C_{15}H_{28}NO_6$ was confirmed by Sirius as first hit with a score of 99.916 %. The tandem mass spectrum showed the distinct ion fragment peak of platynecine (m/z $[M+H]^+$ 174.1126, $C_8H_{16}NO_3$). Analysis with CSI:FingerID identified this mass feature as dihydroindicine *N*-oxide, a stereoisomer of 7-hydroxytrachelanthamine *N*-oxide with 78.74 % at rank 1. A summary of the computational data analysis with Sirius4 and CSI:FingerID can be found in Table 2.

As can be seen in the box and whisker plots of the EIC intensities, all three mass features identified as trachelanthamine, trachelanthamine *N*-oxide or 7-hydroxytrachelanthamine, and 7-hydroxytrachelanthamine *N*-oxide were detected exclusively in the control lines and not in the knock-out lines.

Based on this good fit in manual and computational identification, we suggest all three mass features presented are intermediates of PA biosynthesis in *S. officinale*.

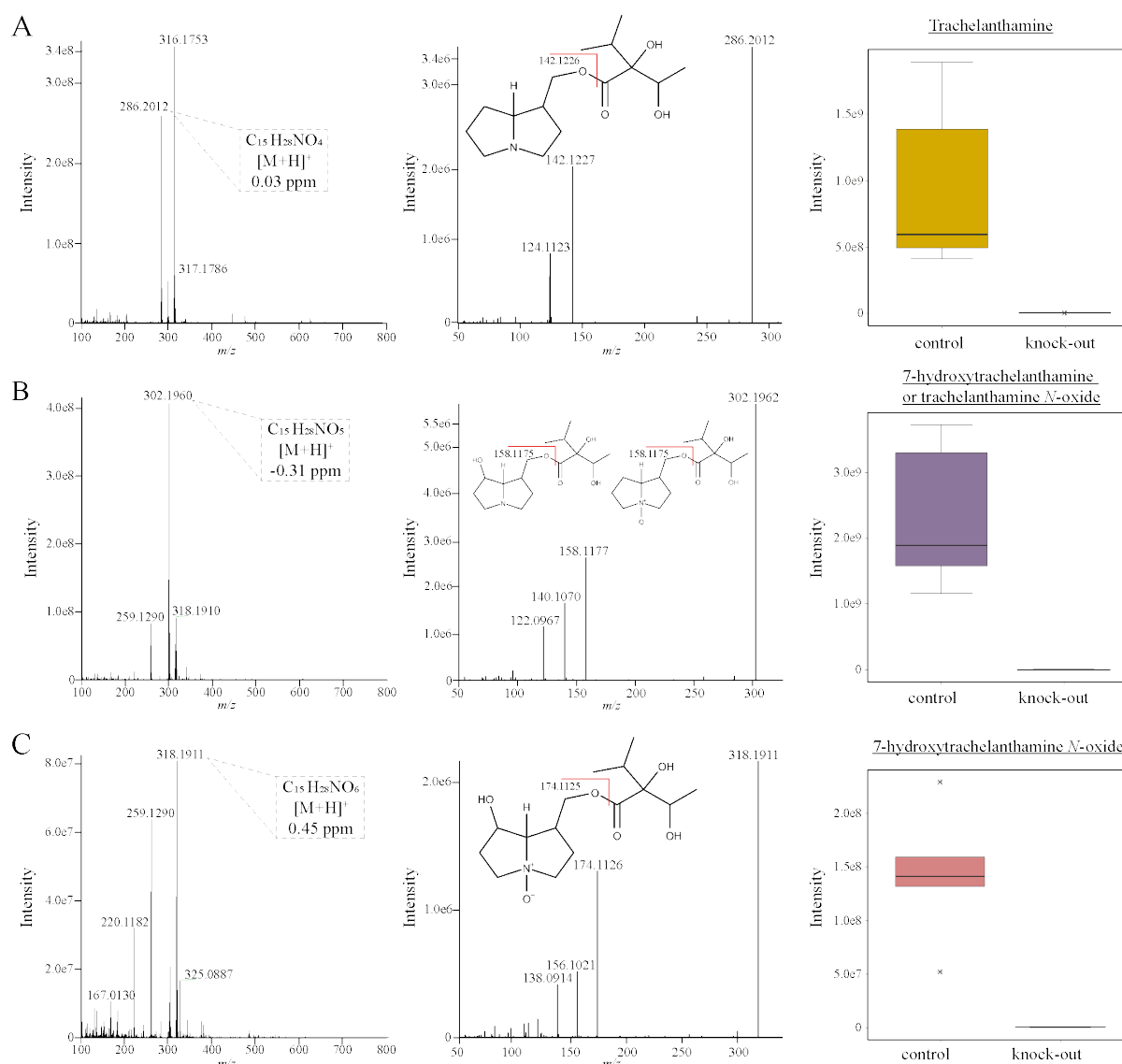


Figure 29 Potential intermediates of PA biosynthesis identified in *S. officinale* CL, but not in KO. High resolution mass spectra, HCD MS/MS and box and whisker plots of EIC intensities of trachelanthamine (A, m/z [M+H]⁺ 286.2012), 7-hydroxytrachelanthamine or trachelanthamine *N*-oxide (B, m/z [M+H]⁺ 302.1960), and 7-hydroxytrachelanthamine *N*-oxide (C, m/z [M+H]⁺ 318.1911), respectively. EIC intensities are given for CL (n=5) and KO (n=4).

4.4 Discussion

To understand the sequence of biosynthetic steps leading to the formation of PAs of the lycopsamine type we need to apply a wide range of experimental approaches. Here, we evaluated metabolic profiles using multivariate analyses combined with classical mass spectrometry identification as well as computational identification tools. We identified metabolites as potential pathway intermediates that are only detectable in hairy root control lines but were absent in hairy root knockout lines, in which the necic acid biosynthesis is interrupted by CRISPR/Cas9-mediated inactivation of the enzyme C₇HAS.

4.4.1 Metabolomic approach

The ability to generate metabolomes and other 'omics' datasets allows us to gain a whole new understanding of biological systems, processes and pathways (Oksman-Caldentey et al., 2004). Large multidimensional platforms allow comprehensive bioinformatics processing of the large datasets and, where appropriate, the combination of different omics data (Yang et al., 2014) to obtain an integrated understanding of secondary metabolism in plants (Hirai et al., 2004; Hirai et al., 2005; Saito, 2013; Zeng et al., 2013). Extraction of the useful information from huge datasets is an important step. To obtain this data several bioinformatic tools have been applied in this study: reprocessing the data with XCMS Online (Tautenhahn et al., 2012; Gowda et al., 2014) allowed the alignment of the chromatograms and a reliable assignment of *m/z* ratios to retention times, referred to as mass features. The resulting metadata tables allowed a comparison of the EIC signal intensities of these mass features. Processing this data with MetaboAnalyst5 allowed the reduction of the dimensionality of the untargeted analysis into a principal components analysis (PCA) representing the variation between the different analyzed sample groups (Jansen et al., 2005; Xia et al., 2009; Chong et al., 2019; Pang et al., 2020). Even with all these data points, the main challenge in metabolomics remains the identification of the molecules, in particular for natural product research, with more than 200,000 structures of natural products from plants that are known (Dixon and Strack, 2003; Modolo et al., 2009) and an estimation of more than one million specialized metabolites in total (Afendi et al., 2012). It is undisputed that MS/MS data alone are insufficient for full structural elucidation of metabolites. Therefore, additional processing of the high-resolution mass spectra and tandem mass spectra with Sirius4 (Dührkop et al., 2019) allowed to answer two questions: What is the molecular formula of the query compound among all molecular formulas, both previously observed and unobserved? And, given a database of molecular structures, what is the structure that best explains the experimental data. For identification we considered the Sirius score, the

fragmentation tree score (Rasche et al., 2012; Böcker and Dührkop, 2016), isotopic pattern analysis (Böcker et al., 2009), and for structure assignment CSI:FingerID (Dührkop et al., 2015). Here the biggest flaw is its biggest advantage: it is based on database searches, allowing the comparison with thousands of known structures, but no database is truly complete, limiting the search to those compounds present in the databases.

4.4.2 Is trachelanthamine first *N*-oxygenated or C⁷ hydroxylated?

The application of our workflow with the computational prediction of the molecular formula and the structure of the metabolites did not allow a distinction between trachelanthamine *N*-oxide (Figure 24, intermediate 18) and 7-hydroxytrachelanthamine (Figure 24, intermediate 13). The question of whether trachelanthamine is first *N*-oxygenated or hydroxylated remains unanswered. As indicated is the reliable and distinct identification of the metabolic features represents one of the greatest challenges in metabolomics (Dixon and Strack, 2003; Modolo et al., 2009). Measuring the mass of a molecular ion with such high accuracy is sufficient to calculate an empirical formula, but does not allow its structural identification, since there are usually a multitude of metabolites with identical molecular formulas (Dührkop et al., 2015). Even if the metabolites can be resolved chromatographically, validating their identification depends on having an authentic standard that is often not available as it is the case for our intermediates. We tried to address this challenge by reducing the extracted metabolites. Adding zinc powder to the samples would convert all *N*-oxidized compounds to their tertiary form. This is also the standard procedure for GC-MS analysis of PA-containing samples, since the *N*-oxides are too polar and not volatile enough to be analyzed by GC-MS (Rizk, 1990). However, it should be noted that all PAs and intermediates would be affected by the reduction: a decrease of trachelanthamine *N*-oxide (↓) would lead to an increase of trachelanthamine (↑). At the same time, however, 7-hydroxytrachelanthamine *N*-oxide (↓) would also be reduced, leading to an increase in 7-hydroxytrachelanthamine (↑). Thus, a mass feature of $m/z [M^+H]^+ 302.1960$ would still be present, even if trachelanthamine *N*-oxide were completely reduced and the assignment of the mass feature to either structure remains difficult.

4.4.3 Predicted metabolic pathway of PA biosynthesis in *S. officinale*

The last comprehensive studies on the process of PA biosynthesis in *S. officinale* were performed by Frölich et al. (2007). In this study, various radioactively labeled necine bases were fed to hairy root cultures of *S. officinale*. It was found that trachelanthamidine, supinidine and retronecine were incorporated into intermediate *N*-oxide without detectable intermediates followed by further chemical diversification with *O*⁷ and *O*^{3'}-esterifications. Since supinidine

was best incorporated into lycopsamine type PAs among the necine bases tested, it was assumed to be the partner for esterification. This assumption was contradicted by the accumulating trachelanthamidine in *S. officinale* C₇HAS knock-out lines (Chapter 3). Accordingly, the observation that a specific compound is incorporated at the highest rate can only be a hint, but no proof that this compound is indeed a natural occurring intermediate of the biosynthetic pathway. The detection of trachelanthamidine and none of the other necine bases, with or without O⁷-esterification, in the metabolomes of the control lines confirms the esterification of trachelanthamidine with one of the stereoisomers of the C₇-necic acids, *i.e.*, (+)-trachelanthic or (-)-viridifloric acid.

Another point that should be considered is the toxicity of PAs. It is generally accepted that PAs are synthesized and stored as nontoxic *N*-oxide derivatives, a feature that protects the PAs from bioactivation. Reduction to the tertiary pre-toxic form occurs spontaneously in the digestive system of the feeding herbivores. The PAs are bioactivated by cytochrome P450 monooxygenases that are part of the xenobiotic metabolism in the liver. If the PAs possess a 1,2-double bond, an esterified hydroxy group at position C⁹, and preferably a free or esterified C⁷ hydroxy group, the bioactivation results in toxic PA pyrroles (Winter and Segall, 1989; Prakash et al., 1999). These PA pyrroles are highly reactive alkylating agents and are able to cross-link DNA and proteins and thus exhibit hepatotoxic, carcinogenic, tumorigenic, and mutagenic properties (Mori et al., 1985; Zijlstra and Vogel, 1988a; Zijlstra and Vogel, 1988b; Frei et al., 1992; Fu et al., 2002; Fu et al., 2004). For PA producing plants it would advantageous to keep the PAs non-toxic for as long as possible, or to protect them from bioactivation by *N*-oxygenation as early as possible to avoid self-harm as cytochrome P450 monooxygenases are also common enzymes in plant biosynthetic pathways (Schuler, 1996). Looking at the biosynthetic pathway of PAs that we predict (Figure 30) based on the intermediates identified in this work, this appears to be the case. We propose the following sequence of biosynthetic steps to the formation of the core structure of lycopsamine type PAs:

The biosynthesis of the necine base is independent from that of the necic acid until the saturated necine base trachelanthamidine is formed. For the necic acid biosynthesis our data suggests that after the formation of the C₇-pronecic acid it is reduced to the C₇-necic acid. Trachelanthamidine and C₇-necic acid are esterified, forming trachelanthamine. This step is followed by *N*-oxygenation and C⁷-hydroxylation in order not clarified, making trachelanthamine *N*-oxide or 7-hydroxytrachelanthamine possible intermediates, ultimately forming 7-hydroxytrachelanthamine *N*-oxide. The 1,2-desaturation is the last step in the formation of the core structure, intermedine *N*-oxide. Our results support the idea that further

chemical diversification like O^7 - and $O^{3'}$ -esterification only occur after the lycopsamine N -oxide core biosynthesis, as no intermediates with such modifications could be identified.

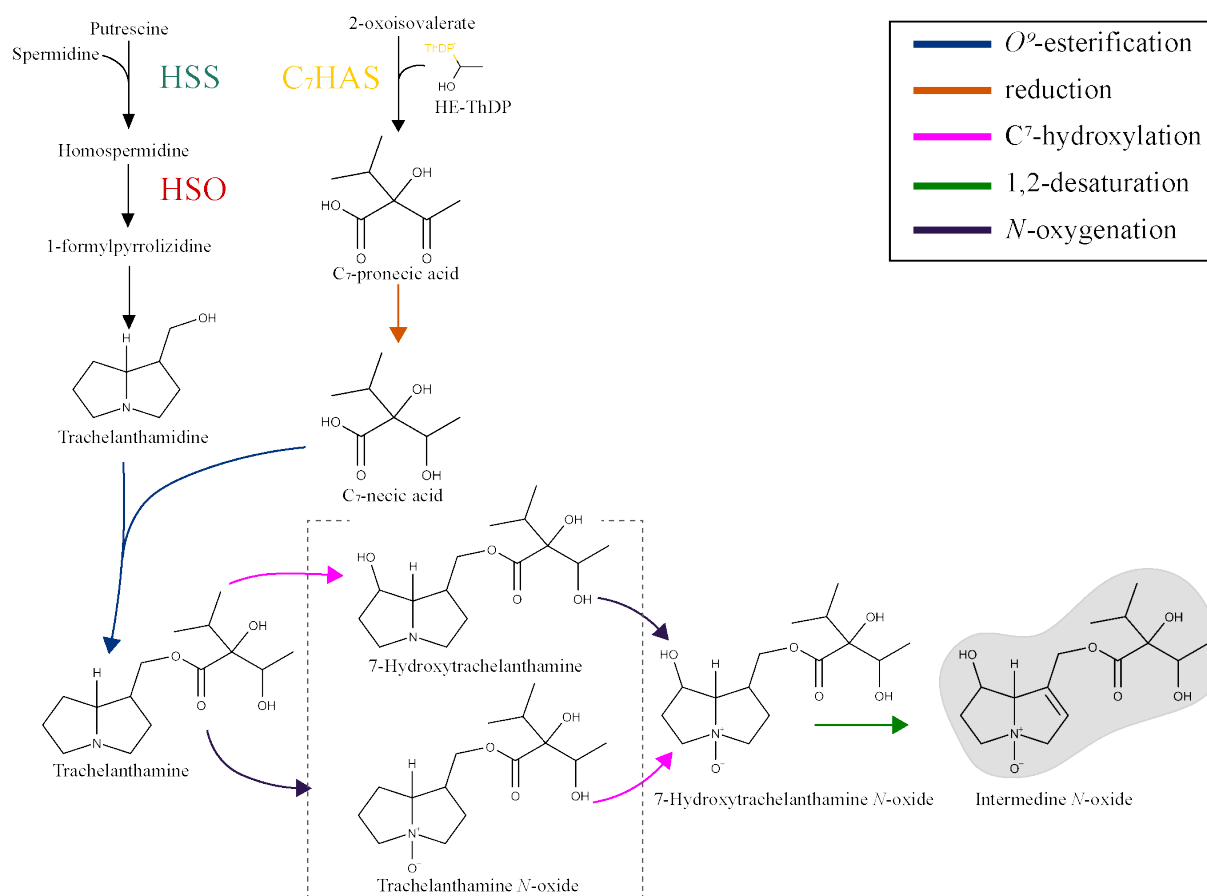


Figure 30 Predicted pathway to intermedine N -oxide, the core structure of lycopsamine type PAs in *S. officinale*. Characterized steps catalyzed by homospermidine synthase (HSS), homospermidine oxidase (HSO), and C^7 -hydroxyacid synthase (C^7 HAS). The predicted biosynthetic route with colored arrows representing one of the five required modifications: O^9 -esterification of necine base and necic acid, reduction of the necic acid moiety, hydroxylation at the C^7 , 1,2-desaturation, and N -oxygenation. Grey dashed brackets indicate steps of which the order could not be clarified.

4.4.4 Conclusion

The identification of trachelanthamidine as a partner for esterification (Chapter 3) has allowed the reduction of the countless possible sequences of biosynthetic steps for the formation of the core structure of the lycopsamine type PAs in *S. officinale* to 125 (Figure 24). By analyzing the metabolomes of C^7 HAS knock-out and control lines we were able to identify three intermediates of PA biosynthesis and consequently to reduce the possible sequences to two (Figure 30). To validate our predicted pathway the identification and characterization of the enzymes involved in the catalyzed reactions is necessary.

4.5 Material and methods

4.5.1 Plant material

Hairy root lines of *Symphytum officinale* with an CRISPR/Cas9-mediated inactivation (knock-out) of the enzyme C₇HAS (KO; A29, A37, A50 and A69) and control lines resulting from transformation with the empty vector construct (CL; CL2, CL4, CL11, CL17, CL34) were described earlier (Chapter 3). Hairy root lines were grown in 80 ml liquid MS20 media (250 µg/ml ticarcillin/clavulanic acid, Duchefa) in the dark at 16°C for 14 days on an orbital shaker at 130 rpm. At harvest, the medium was removed and patted dry with tissue paper before they were frozen, lyophilized and stored at room temperature until further use.

4.5.2 Metabolite extraction

Freeze-dried samples were ground with mortar and pestle. 10 mg of the resulting powder were transferred into a 1.5 ml-reaction tube before 600 µl of 0.05 M H₂SO₄ were, added and the mixture vortexed for 30 s. After sonication for 15 min in a water-bath (tubes were shaken for 10 seconds in between), the tubes were vortexed for another 10 s and centrifuged at 12 000 g for 5 min at 4°C. The supernatants were transferred into fresh 1.5 ml reaction tubes, before the pellet was extracted again with 600 µl 0.05 M H₂SO₄ following the same protocol. The supernatants were combined, the pH neutralized with a 5% (v/v) ammonia solution and centrifuged at 12 000 g for 5 min at 4°C. After transfer of the supernatants into new 1.5 ml reaction tubes, the centrifugation step was repeated and the extracts transferred into 1 ml vials for ultra-high-performance liquid chromatography-tandem mass spectrometry (UHPLC/ESI-Q Exactive HF-X-MS/MS).

4.5.3 UHPLC-ESI-Q Exactive MS/MS Analysis of Extracted Metabolites

UHPLC analyses of the hairy root extracts were performed on a Thermo Scientific™ Dionex UltiMate 3000 Rapid Separation LC (Dionex, Sunnyvale, CA, USA). Chromatographic separation was achieved on an AcclaimC18 column (150 × 2.1 mm, 2.2 µm particles with 120 Å pore diameter (Dionex) with a flow rate of 300 µl min⁻¹ in a binary solvent system of water (solvent A) and acetonitrile (solvent B), both containing 0.1% (v/v) formic acid. 8 µl of each extract were separated using a gradient as follows: linear increase from 0% solvent B to 100% solvent B within 15 min, 100% solvent B constant for 5 min, equilibration time at 0% B for 5 min. The high-resolution accurate masses and tandem mass spectra were obtained with a coupled Thermo Scientific™ ESI-Q Exactive HF-X-Hybrid Quadrupole-Orbitrap™ mass spectrometer (Dionex). Positive and negative ionization was achieved using heated-electrospray ionization (H-ESI). The source parameters were set to: 4.2 kV (positive ionization)

and 3.3 kV (negative ionization) spray ionization, 35 V transfer capillary voltage, 300°C capillary temperature and 40 V funnel radio frequency. Fragmentation was recorded using data dependent acquisition mode with MS full scan (MS^1) with m/z of 100 – 800 at 60 000 M (= resolving power) / ΔM (= mass of the second peak) resolution and one MS/MS scan (MS^2) with 30 000 M/ ΔM resolution and stepped normalized collision energy of 20, 30 and 40.

4.5.4 Manual identification of metabolites

Data was evaluated and interpreted using Xcalibur v.2.2 software (Thermo Fisher Scientific, Waltham, MA, USA). Classical identification was performed using accurate mass measurements of mass features (MS^1 and MS^2). A mass tolerance of ± 3 ppm was applied as threshold between high-resolution accurate masses and calculated exact masses.

4.5.5 Data conversion

The raw data of UHPLC/ESI-Q Exactive HF-X-MS/MS measurements are saved in .raw format. This format is not compatible with bioinformatics programs used for processing. The data must first be converted to a readable format. Conversion of raw data files to mzXML format files was performed using MSConvertGUI tool of the ProteoWizard 3.0.X software (Chambers et al., 2012). The filter used was peak picking with msLevel 1-2.

4.5.6 Data processing with XCMS™ Online

A mass feature is defined as a measured mass (m/z) at a given retention time. To generate lists of mass features, the mzXML data files were processed using the “multigroup analysis” of the bioinformatics tool XCMS™ Online (version 3.7.1, Tautenhahn et al., 2012). Within the program the following tools were used:

1. “feature detection” was performed using the centWave algorithm with following parameter: maximal tolerated m/z deviation in consecutive scans, in parts per million (ppm): $\Delta m/z$ 5 ppm, minimum chromatographic peak width: 5 seconds, maximum chromatographic peak width: 20 s, minimum difference in m/z for peaks with overlapping retention time (mzdiff): 0.01, signal/noise threshold: 4, integration method: 1. Mass traces are only retained if they contain at least 3 peaks (prefilter peaks) with intensity 100 (prefilter intensity), and are above the noise threshold of 100 (noise filter).
2. “Retention time correction” was performed with obiwarpmethod with the step size (profStep) set to 1 m/z .

3. "Alignment" of the chromatograms was performed using the following parameters: bandwidth of gaussian smoothing kernel to apply to the peak density chromatogram (bw): 1 m/z , minimum fraction of the samples necessary in at least one of the sample groups for it to be a valid group (minfrac): 0.5, width of overlapping m/z slices to use for creating peak density chromatograms and grouping peaks across samples (mzwid): 0.025, minimum number of samples necessary in at least one of the sample groups for it to be a valid group (minsamp): 1, maximum number of groups to identify in a single m/z slice (maxsamp): 100.

4.5.7 Statistical analysis with Metaboanalyst 5.0

Statistical analysis was performed using the tool "Statistical analysis" in Metaboanalyst 5.0 (Xia et al., 2009; Chong et al., 2019; Pang et al., 2020; Pang et al., 2021). For this, the metadata tables generated with XCMSTMOnline were uploaded, the data filtered using interquartile range, and normalized using auto scaling. We used the correlation pattern analysis "principle component analysis" in our evaluation of the data.

4.5.8 Data processing with MZmine3

The aforementioned bioinformatic analyses refer to the high-resolution accurate masses in the MS full scan (MS¹). The processing of the data using SIRIUS 4 and CSI:Finger ID to predict the molecular formula and structure of the analytes also integrates the MS² data into the calculations. For this purpose, the mzXML data must first be processed to with MZmine3.0 (Pluskal et al., 2010). Within the program, the following tools were used one after the other:

1. "Mass detection" with a centroid mass detector: detection on MS¹ and MS² with a noise level of 3.0E3.
2. "ADAP chromatogram builder" (MS¹ level) (Myers et al., 2017) for feature detection. With the retention time filter set to 1.00 – 15.00 minutes; minimum of group size in # of scans set to 4; a group intensity threshold of 3.0E3; a minimal intensity height of 5.0E3; scan to scan accuracy of $\Delta m/z$ 5 ppm
3. "Local minimum resolver" for a higher resolution of the single features (less merged features) with the following settings: Chromatographic percentile threshold for removing noise: 80%; minimum, search range RT/Mobility (absolute) to define borders between peaks: 0.04; minimum absolute height of a peak to be recognized: 5.0E4; minimum ratio of peak top/edge to reduce the number of false peaks: 1.7; acceptable peak duration range (min/mobility): 0.00-1.00; minimum number of data points on a feature: 4

4. “¹³C isotope filter” to deisotope the dataset. Deisotoping was performed with a mass tolerance of 5 ppm and retention time tolerance of 0.1 minutes to still consider features as the same of ions with maximum charges of 2. Features worth fragmenting, a.k.a. features with a MS² were not filtered.
5. “Join aligner” to align the chromatograms and generate a data matrix with the following parameters: Maximum allowed difference between two *m/z* values to be considered the same: 0.001 *m/z* or 5 ppm; score for perfectly matching *m/z* values (weight for *m/z*): 3; maximum allowed difference between two retention time values (retention time tolerance): 0.1 minutes; score for perfectly matching retention time values: 1.
6. Data were exported as .mgf files.

4.5.9 Determination of the molecular formulae using SIRIUS 4

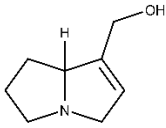
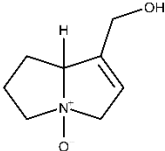
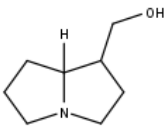
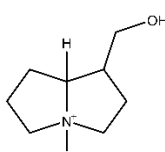
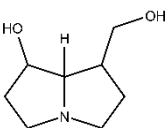
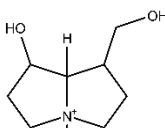
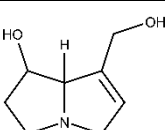
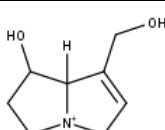
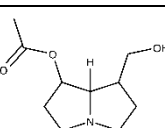
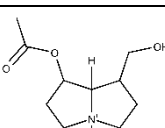
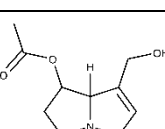
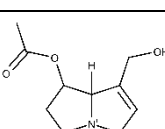
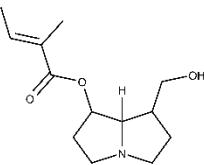
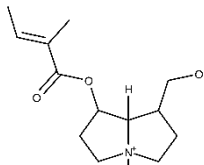
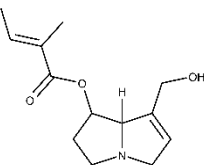
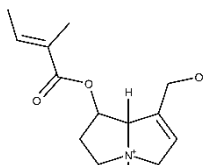
The processed .mgf data files from MZmine3 were subsequently imported into SIRIUS4 software 5.5.3 (Dührkop et al., 2019). The features were processed as adducts with the possible ionizations of [M+H]⁺, [M+K]⁺ and [M+Na]⁺ and for measurements in negative ionization mode the deprotonated [M-H]⁻. Instrument was set to Orbitrap with an MS² accuracy of 5 ppm. For molecular formula assignment we considered Sirius Score, fragmentation tree score, explained peaks, and the total explanation (Rasche et al., 2012; Böcker and Dührkop, 2016) .

4.5.10 Structure assignment using CSI:FingerID

Fingerprint prediction of the MS² data and the subsequent structure database search was achieved by using the CSI:FingerID tool of SIRIUS4 (Dührkop et al., 2015; Hoffmann et al., 2021). The features were processed as adducts with the possible ionizations of [M+H]⁺, [M+K]⁺ and [M+Na]⁺. The database search included the following databases: COCONUT (Sorokina et al., 2021), KNApSACk (Afendi et al., 2012), and PubChem® (National Library of Medicine, USA). For structure assignment we considered the rank, the similarity score and the CSI:FingerID score.

4.6 Supplementary

Supplementary Table 7 Necine bases and *θ*⁷-esters

Tertiary bases			<i>N</i> -oxide bases		
Supinidine			Supinidine <i>N</i>-oxide		
formula	C ₈ H ₁₃ NO		formula	C ₈ H ₁₃ NO ₂	
exact mass	139.0997		exact mass	155.0946	
<i>m/z</i> [M+H] ⁺	140.1070		<i>m/z</i> [M+H] ⁺	156.1019	
Trachelanthamidine			Trachelanthamidine <i>N</i>-oxide		
formula	C ₈ H ₁₅ NO		formula	C ₈ H ₁₅ NO ₂	
exact mass	141.1154		exact mass	157.1103	
<i>m/z</i> [M+H] ⁺	142.1226		<i>m/z</i> [M+H] ⁺	158.1175	
Platynecine			Platynecine <i>N</i>-oxide		
formula	C ₈ H ₁₅ NO ₂		formula	C ₈ H ₁₅ NO ₃	
exact mass	157.1103		exact mass	173.1052	
<i>m/z</i> [M+H] ⁺	158.1175		<i>m/z</i> [M+H] ⁺	174.1125	
Retronecine			Retronecine <i>N</i>-oxide		
formula	C ₈ H ₁₃ NO ₂		formula	C ₈ H ₁₃ NO ₃	
exact mass	155.0946		exact mass	171.0895	
<i>m/z</i> [M+H] ⁺	156.1019		<i>m/z</i> [M+H] ⁺	172.0968	
7-acetylplatynecine			7-acetylplatynecine <i>N</i>-oxide		
formula	C ₁₀ H ₁₇ NO ₃		formula	C ₁₀ H ₁₇ NO ₄	
exact mass	199.1203		exact mass	215.1152	
<i>m/z</i> [M+H] ⁺	200.1281		<i>m/z</i> [M+H] ⁺	216.1230	
7-acetylretronecine			7-acetylretronecine <i>N</i>-oxide		
formula	C ₁₀ H ₁₅ NO ₃		formula	C ₁₀ H ₁₅ NO ₄	
exact mass	197.1046		exact mass	213.0996	
<i>m/z</i> [M+H] ⁺	198.1125		<i>m/z</i> [M+H] ⁺	214.1073	
7-tigloylplatynecine			7-tigloylplatynecine <i>N</i>-oxide		
formula	C ₁₃ H ₂₁ NO ₃		formula	C ₁₃ H ₂₁ NO ₄	
exact mass	239.1516		exact mass	255.1465	
<i>m/z</i> [M+H] ⁺	240.1594		<i>m/z</i> [M+H] ⁺	256.1543	
7-tigloylretronecine			7-tigloylretronecine <i>N</i>-oxide		
formula	C ₁₃ H ₁₉ NO ₃		formula	C ₁₃ H ₁₉ NO ₄	
exact mass	237.1539		exact mass	253.1309	
<i>m/z</i> [M+H] ⁺	238.1438		<i>m/z</i> [M+H] ⁺	254.1387	

5 Characterization of the necic acid biosynthesis of lycopsamine type pyrrolizidine alkaloids – miscellaneous efforts

Characterization of the necic acid biosynthesis of lycopsamine type pyrrolizidine alkaloids – miscellaneous efforts

Annika Engelhardt^a, Sebastian Fritz^a, Britta Muhs^a, and Dietrich Ober^a

^a Botanisches Institut und Botanischer Garten, Christian-Albrechts-Universität zu Kiel,
D-24098 Kiel, Germany

Detailed author contributions are listed at the end of the thesis.

5.1 Abstract

Secondary metabolism and the enzymes involved provide an interesting opportunity to study the adaptive mechanisms of molecular evolution and, in particular, the evolution of enzymes. An established model for the study of these mechanisms is the pyrrolizidine alkaloid (PA) biosynthesis and the enzymes involved in it. Characteristic for PAs is the bicyclic, nitrogen-containing necine base esterified with one or more necic acids. The first specific step of the metabolic pathway of PAs is catalyzed by a homospermidine synthase (HSS). The comprehensive evolutionary characterization of HSS revealed that the PA biosynthesis evolved several times independently in different plant lineages, among others, in the Boraginales, the Apocynaceae, and in the Eupatorieae (Asteraceae). Remarkably, these three independent origins resulted in the formation of PAs with identical structures and a characteristic branched-chain C₇-necic acid, the lycopsamine type PAs. In *Symphytum officinale* (*S. officinale*, Boraginaceae, Boraginales), it was shown that the first pathway-specific step of C₇ acid biosynthesis is the formation of C₇-pronecic acid by transfer of an activated acetaldehyde formed from pyruvate (PYR) to 2-oxoisovalerate (2OIV). This step is catalyzed by a C₇-hydroxyacid synthase (C₇HAS) that has evolved from the catalytic subunit (CSU) of an acetohydroxyacid synthase (AHAS).

Here we present the analysis of C₇AS candidate genes in other lycopsamine type PA producing plant species from different plant families, *i.e.* *Eupatorium cannabinum* (*E. cannabinum*, Eupatorieae, Asteraceae) and *Heliotropium indicum* (*H. indicum*, Heliotropiaceae, Boraginales). The CRISPR/Cas9 mediated genome editing technique was applied to provide *in planta* evidence for the involvement of C₇HAS in the PA biosynthesis of *E. cannabinum* and, as proof of concept, to provide *in planta* evidence for the involvement of HSS in the PA biosynthesis. Genotypic knock-outs of the gene encoding HSS resulted in PA free hairy roots without any detectable PA-specific intermediates. Knock-outs of the gene encoding C₇HAS also resulted in lycopsamine type PA free hairy roots but accumulated the saturated necine base PA trachelanthamidine as an intermediate of PA biosynthesis. Additionally, the AHAS CSU and the C₇HAS from *H. indicum* were heterologously expressed in *E. coli* and characterized with respect to their keto acid substrate range. We could show that *HiCSU* *HiC₇HAS* both are able to convert the substrates of primary metabolism, *i.e.*, pyruvate and 2-oxobutyrate. Additionally, *HiC₇HAS* accepts 2OIV as acceptor molecule for the C₂-moiety, catalyzing the formation of C₇-pronecic acid. Phylogenetic analyses of the C₇HAS encoding genes from *S. officinale*, *E. cannabinum*, and *H. indicum* indicate an independent origin in the Heliotropiaceae, Boraginaceae, and Eupatorieae.

5.2 Introduction

Plants produce an amazing variety of metabolites that are not essential to the life of the plant cell but are critical for the plant's survival in its environment. These metabolites are categorized as part of plant secondary metabolism and fulfill multiple purposes such as attracting pollinators, protecting against UV radiation, and defending against herbivores and pathogens (Harborne, 1993; Hartmann, 2007). The diversity of functions of secondary metabolites is by far exceeded by their structural diversity, which includes a wide variety of groups such as polysaccharides, glycosides, lipids, polyphenols and alkaloids. It is estimated that every plant species biosynthesizes 4.7 unique metabolites which means that the total number of specialized metabolites exceeds a million (Afendi et al., 2012). Such metabolic diversity requires a large number of different enzymes, and subsequently encoding genes, involved in the biosynthesis of the metabolites. Gene duplications with subsequent diversification of gene copies are assumed to be the main drivers of evolution of the diversity in plant secondary metabolism (Pichersky and Gang, 2000; Ober, 2005; Pichersky et al., 2006; Koch, 2014; Rensing, 2014; Panchy et al., 2016). Consequently, secondary metabolism and the enzymes involved provide an interesting opportunity to study the adaptive mechanisms of molecular evolution and, in particular, the evolution of enzymes (Pichersky and Gang, 2000; Grotewold, 2005). An established model for the study of these mechanisms is the pyrrolizidine alkaloid (PA) biosynthesis and the enzymes involved in it (Ober and Hartmann, 1999; Reimann et al., 2004; Ober, 2005; Ober and Kaltenegger, 2009; Ober, 2010; Kaltenegger et al., 2013; Livshultz et al., 2018). Characteristic for PAs is the bicyclic, nitrogen-containing necine base esterified with one or more necic acids. The first specific step of the metabolic pathway of PAs is catalyzed by a homospermidine synthase (HSS, EC 2.5.1.45) that catalyzes the transfer of an aminobutyl moiety from spermidine to putrescine forming homospermidine (Ober and Hartmann, 1999). It was shown that the HSS evolved several times independently from duplication events of ancestral deoxyhypusine synthase genes (Reimann et al., 2004; Ober and Kaltenegger, 2009; Livshultz et al., 2018). For example, independent origins of HSS in the Apocynaceae family, the Eupatorieae and Senecioneae tribes within the Asteraceae and in the order of the Boraginales have been demonstrated. HSS, and thus presumably the entire PA biosynthesis, exhibits high tissue and cell specificity. In immunolabelling experiments, HSS was localized in the cortex parenchyma of primary roots of *Eupatorium cannabinum* (*E. cannabinum*, Eupatorieae) (Anke et al., 2004) and in the lower leaf epidermis of *Heliotropium indicum* (*H. indicum*, Heliotropiaceae, Boraginales) (Niemüller et al., 2012). In *Symphytum officinale* (*S. officinale*, Boraginaceae, Boraginales), HSS could be localized in the endodermis of primary

roots and in bundle sheath cells of young leaves below inflorescences (Niemüller et al., 2012; Kruse et al., 2017). Regarding the independent origin of HSS and the various sites of biosynthesis, it seems all the more surprising that the species of the Apocynaceae, Boraginales, and the Eupatorieae all synthesize essentially identical products, *i.e.* the lycopsamine type PAs. These PAs share a characteristic, branched chain aliphatic C₇-necic acid moiety as a common structural property of the acid group (Hartmann and Witte, 1995). In *S. officinale*, it was shown that the first pathway-specific step of C₇ acid biosynthesis is the formation of C₇-pronecic acid by transfer of an activated acetaldehyde formed from pyruvate (PYR) to 2-oxoisovalerate (2OIV). This step is catalyzed by a C₇-hydroxyacid synthase (C₇HAS) that has evolved from a gene duplication of the catalytic subunit (CSU) of an acetohydroxyacid synthase (AHAS, EC 2.6.1.1) (Chapter 3). In primary metabolism, AHASs catalyze the first step in the biosynthesis of the branched-chain amino acids (BCAA) valine and leucine by transferring the activated acetaldehyde to an acceptor PYR and isoleucine by transferring it to 2-oxobutyrate (Umbarger and Brown, 1958).

The isolated occurrence of lycopsamine type PAs in these unrelated lineages raises two questions: (1) Is a C₇HAS involved in necic acid moiety biosynthesis in all lycopsamine type PA producing species? (2) Does C₇HAS have a polyphyletic origin, as is the case with HSS, leading to PAs of essentially identical structure? Or, since what we know about HSS does not necessarily apply to all enzymes of PA biosynthesis, does it have a monophyletic origin?

These questions were addressed in this work by studying a candidate sequence for C₇HAS from *E. cannabinum* using CRISPR/Cas9-mediated genome editing and by analyzing the substrate range of an AHAS CSU and a C₇HAS candidate sequence from *H. indicum*. Additionally, the evolutionary origin of the C₇HAS was analyzed by phylogenetic reconstruction.

5.3 Results

By using a reverse genetic approach, two CSU-like sequences were identified in *H. indicum* (*HiCSU1* and *HiCSU3*; Muhs and Ober, unpublished), three in *E. cannabinum* (*EcanCSU1*, *EcanCSU2* and *EcanCSU3*) and one in *Parsonsia. laevigata* (*PICSU*) (Langel, 2008). Additionally, one RSU-like sequence was found in *H. indicum* (*HiRSU*; Muhs and Ober, unpublished) and two in *E. cannabinum* (*EcanRSU1*, *EcanRSU2*) (Langel, 2008). As AHASs in higher plants AHAS contain no introns, it could be shown that the cDNA sequences and genomic sequences of these genes are identical.

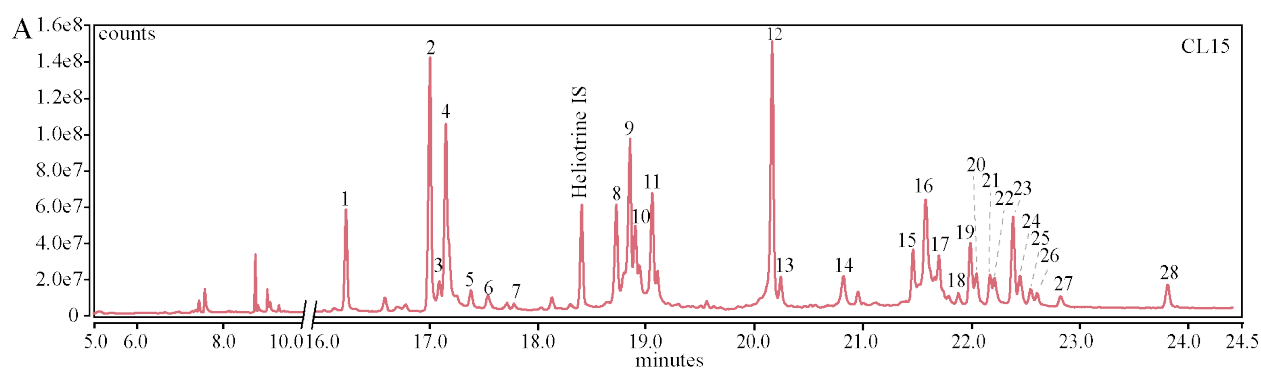
In *Eupatorium*, *EcanCSU2* is root specific and was thus assumed to be the candidate sequence for the C₇HAS of PA biosynthesis (hereafter *EcanC₇HAS*). In *Heliotropium*, the CSU3 came up after subtractive transcriptome analyses to enrich PA-specific sequences in the lower leaf epidermis (Sievert et al., 2015) and is thus assumed to be the candidate C₇HAS (hereafter *HiC₇HAS*).

5.3.1 CRISPR/Cas9 knock-outs of *EcanC₇HAS* and *EcanHSS*

To test if the candidate *EcanC₇HAS* is involved in PA biosynthesis, the gene encoding the *EcanC₇HAS* was inactivated by a genome editing approach using clustered regulatory interspaced short palindromic repeats (CRISPR) with CRISPR associated protein 9 (CRISPR/Cas9). Additionally, the CRISPR/Cas9 approach was applied to the *EcanHSS* as a positive control of an enzyme with strong *in vitro* and immunochemical evidence to be involved in PA biosynthesis (Anke et al., 2004; Reimann et al., 2004). The roots were identified as the site of PA biosynthesis in *E. cannabinum*, allowing us to use hairy roots resulting from transformation with the hairy root-inducing *Agrobacterium rhizogenes* (*A. rhizogenes*) for PA biosynthetic pathway studies. Three expression constructs for the short guide RNAs (sgRNAs) targeting the *ecanhss* and *ecanc₇has* were created (constructs *ecanhss* A, B, and C and *ecanc₇has* A, B, and C, respectively) and cloned into the destination vector containing the Cas9 and hygromycin resistance encoding genes. After transformation into *A. rhizogenes* these were used for infection of *E. cannabinum* leaf sections. Transformed lines were selected by hygromycin resistance, analyzed for mutations in the targeted genes and for an effect on PA levels in comparison to control lines that resulted from infection with the empty destination vector. For the identification and quantification of the PAs synthesized by the hairy roots, their purified extracts were analyzed via GC-MS.

Figure 31 A shows the total ion chromatogram of a purified extract of a control line (CL15) with a total of 28 different PAs that could be detected. So far only 11 different PAs have been

described for *E. cannabinum* (Hendriks et al., 1987; Hartmann and Witte, 1995), most of which do not have a comparative retention index (RI). The PA identification is summarized in Figure 31 B and was done by comparing RIs on a DB-1 capillary column with values and mass spectra of reference compounds as described in the literature. Doubtful cases are marked. Trachelanthamine (peak 2) and 3'-tigloyltrachelanthamine (peak 12) represented 10 % and 13 % of the total amount of PAs isolated from the control line hairy roots. While most PAs could not be identified, it was shown that *E. cannabinum* contains a diverse mix of PAs with necine bases that are saturated (trachelanthamidine), unsaturated (supinidine), and unsaturated with C₇-hydroxyl group (retronecine). A precise identification and possibly even structural elucidation of the PAs would be of further interest. The control lines contained 45% PAs with trachelanthamidine base, 3% with supinidine base, and 52% with retronecine base moiety. In addition to the complex PAs, low levels of the saturated necine base PA trachelanthamidine were detectable in the control lines. Detailed MS data of the PAs with characteristic ions m/z and their relative abundance is given in Supplementary Table 8.



B	No	Base moiety	PA	RI DB1	ref	RI TG1	RI TG-5 AMINE	% PA
	1	Trach.	?			1904	1978	3.81
	2	Trach.	Trachelanthamine	1970	¹	1979	2063	10.55
	3	Sup.	Supinine	1978	¹	1987	2071	1.61
	4	Trach.	?			1995	2079	10.52
	5	Trach.	?			2015	2104	1.36
	6	Trach.	?			2031	2121	1.19
	7	Trach.	?			2068	2146	0.44
	8	Retro.	Intermedine	2144	²	2145	2253	4.64
	9	Retro.	?			2160	2268	8.38
	10	Retro.	?			2166	2274	4.95
	11	Retro.	?			2185	2291	5.26
	12	Trach.	3'-tigloyltrachelanthamine	2320	³	2336	2423	13.36
	13	Sup.	?			2343	2433	1.86
	14	Trach.	?			2418	2506	2.00
	15	Retro.	?			2485	2586	2.58
	16	Retro.	?			2489	2601	6.88
	17	Retro.	Echiumine	2500	²	2502	2611	2.95
	18	Retro.	Symphytine	2520	²	2523	2634	0.56
	19	Retro.	?			2556	2646	2.74
	20	Retro.	?			2561	2653	1.72
	21	Retro.	Echimidine	2560	¹	2567	2666	1.11
	22	Retro.	Echimidine isomer			2573	2672	1.40
	23	Retro.	?			2580	2691	4.08
	24	Retro.	?			2601	2697	2.02
	25	Retro.	3'-acetylmyoscorpine	2614	²	2615	2708	0.93
	26	Retro.	3'-acetylsymphytine	2623	²	2624	2715	0.82
	27	Retro.	?			2644	2739	0.83
	28	Trach.	?			2711	2836	1.44

Figure 31 Chemotyping of control line hairy roots of *E. cannabinum*. **A** Total ion chromatogram of purified extract of CL15 on TG-5 AMINE capillary column: Peaks 1-28 were identified as PAs based on the mass spectrum. **B** PAs, No 1-28 as labelled in the total ion chromatogram of CL15. Necine base moieties: Trach. = Trachelanthamidine, Sup = Supinidine, Retro. = Retronecine, ? = unidentified PA. References: ¹ (El-Shazly and Wink, 2014); ² Chapter 3, ³ unpublished data Witte.

PA quantification and genotyping of the control lines, *ecanhss* knock-out lines and *ecanc₇has* knock-out lines is summarized in Figure 32. The 4 control lines contained an average of 4955 µg PAs and 129 µg trachelanthamidine per gram of dried hairy roots. Of 6 *ecanhss* knock-out hairy root lines resulting from transformation with construct B and 2 resulting from transformation with construct C showed no PAs. To demonstrate a correlation between an altered PA content and CRISPR/Cas9-mediated inactivation of the *ecanhss* gene in the analyzed hairy roots, the targeted regions in the genomic DNA of two hairy root lines was amplified, sequenced, and compared to that of the *ecanhss* gene of the wild type plant (WT) used for transformation. In B27, a homozygous deletion of 9 bp was found in the targeted region of the *ecanhss* gene. In C29, heterozygous deletions of 3 bp and 32 bp each was found. Complementing the extensive available *in vitro* data and immunochemical findings (Anke et al., 2004; Reimann et al., 2004), this result is the first *in planta* evidence that an HSS is involved in PA biosynthesis in *E. cannabinum*. Furthermore, it clearly shows the effect a knock-out of a gene involved in PA biosynthesis has on the PA content. To test whether a similar effect can be achieved with knock-outs of the *EcanC₇HAS* candidate gene, 14 hairy roots resulting from the transformation with *ecanc₇has* targeting constructs were analyzed in the same way. While no line resulting from transformation with construct B showed any effects on PA level (data not shown), 4 lines resulting from construct A and 5 resulting from construct C were PA free (Figure 32) whilst accumulating the saturated base trachelanthamidine. Genotyping of some of these hairy roots revealed alterations of the *EcanC₇HAS* gene ranging from 1 bp insertions and deletions to 54 bp insertions and 97 bp deletions.

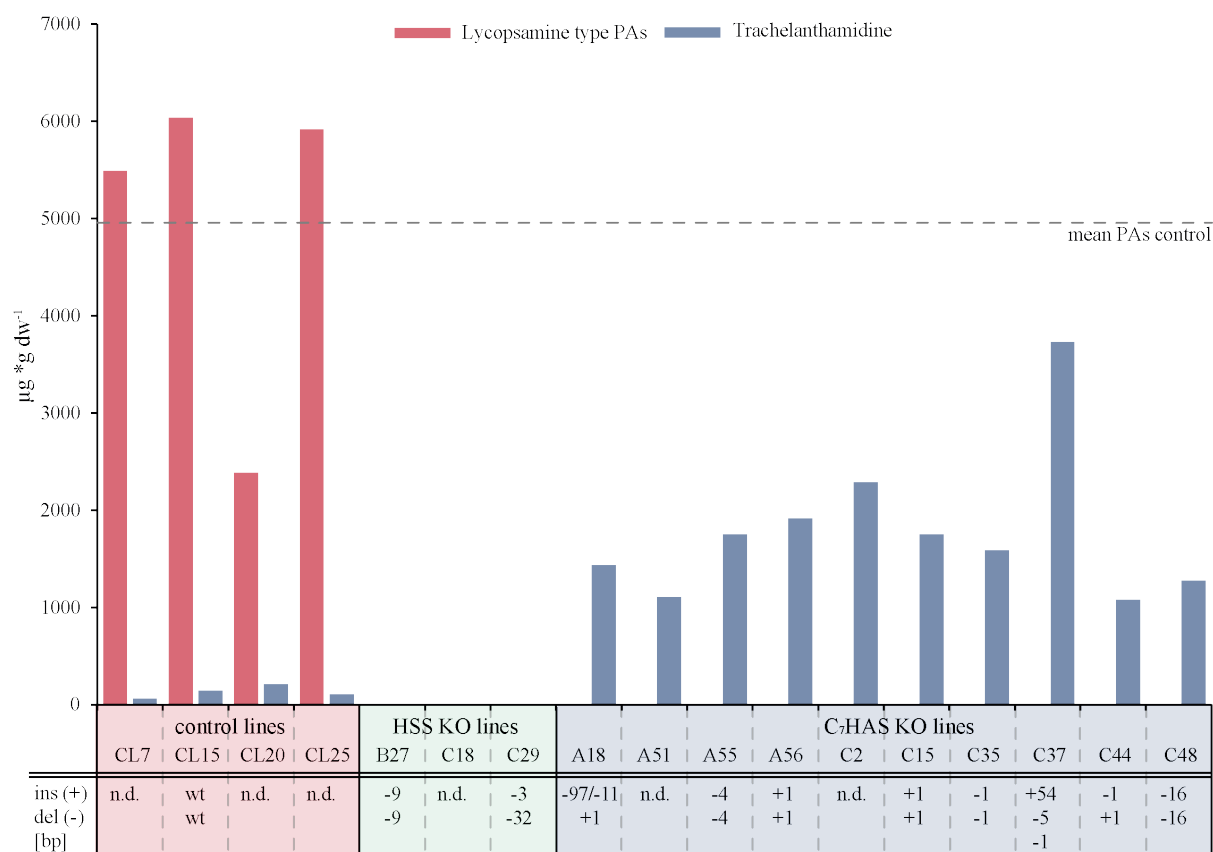


Figure 32 Genotyping, PA content and accumulation of (-)-trachelanthamidine in different hairy root lines.

Four control lines (CL) resulting from infection with the empty destination vector, 3 KO lines resulting from infection with *hss* targeting constructs, one with construct B and 2 lines resulting from infection with construct C, and 4 lines resulting from infection with construct A, and 6 lines resulting from infection with construct C both targeting the *soc7has* gene were analyzed. PA content and amount of accumulated (-)-trachelanthamidine are given in $\mu\text{g/g}$ dry weight (dw) and are calculated as heliotrine equivalents. The dashed line represents the average PA content of the four CLs. The modifications observed within the *soc7has* gene by genotyping are indicated by “+” for insertions and “-” for deletions. If different mutations occurred simultaneously on the different alleles, they are separated by “/”.

5.3.2 Formation of C₇-pronecic acid is exclusively catalyzed by C₇HAS

We have tested the C₇HAS from *E. cannabinum* and *H. indicum* regarding their AHAS and C₇HAS activity. For comparison, the CSU3 of *Eupatorium* and the CSU of *Heliotropium*, were included in the study as possible enzymes of branched-chain amino acid biosynthesis. For heterologous expression, the *EcanC₇HAS* and the *EcanCSU3* were provided as co-expression constructs in pCDFDuet1™ vector with *EcanRSU1* in the second multiple cloning site. *HiC₇HAS* and *HiCSU* were provided as expression constructs in pCDFDuet1_modStrep without an RSU (Muhs and Ober, unpublished). The enzymes were heterologously expressed in *E. coli* BL21(DE3) and affinity purified with NiNTA agarose and Strep-Tactin® Superflow®

column, respectively. The SDS-PAGE analyses of affinity-purified proteins (Figure 33 A) confirmed the expression of a protein of about 60-70 kDa corresponding to the CSUs and appx. 50 kDa corresponding to the RSUs of the *Ecan*CSU3 and RSU. The analysis of *H. indicum* proteins (Figure 33 B and C) both showed bands of about 60-70 kDa corresponding to the expected 64 kDa CSU and C₇HAS. The *Ecan*C₇HAS with RSU1 could not be successfully overexpressed and purified. All other enzymes showed activity in a colorimetric AHAS assay with PYR as substrate indicating the formation of acetolactate.

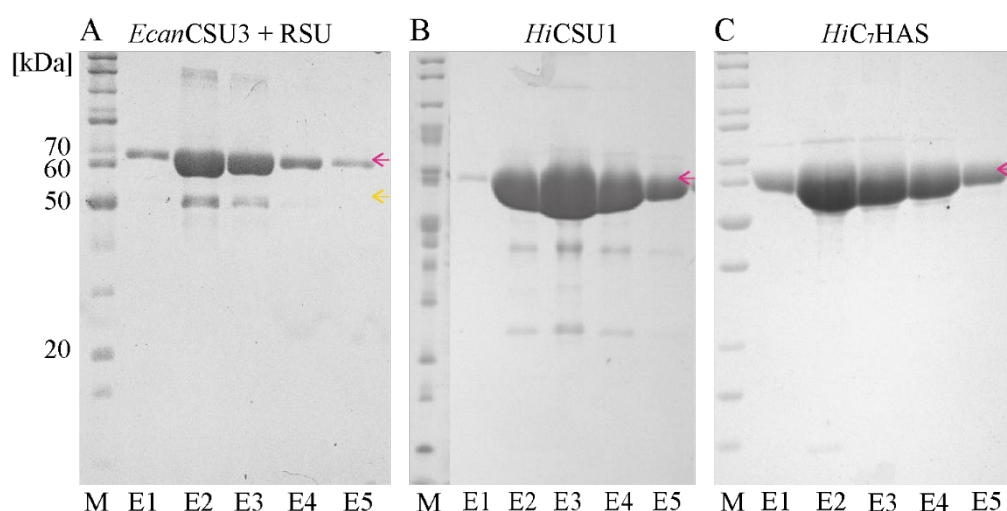


Figure 33 SDS-PAGE (12% gel, comassie blue stained) of fractions resulting from affinity purification of
A *E. cannabinum*: 6xHis-tagged CSU3 (64 kDa) and co-purified RSU1 (52 kDa); **B** *H. indicum* strep-tagged CSU of AHAS (64 kDa); **C** *H. indicum* strep-tagged C₇HAS (64 kDa). The catalytic subunits are labelled by pink arrows, the regulatory subunit by a yellow arrow. E = elution fractions, M = protein size marker (PageRuler™ 10-200 kDa).

Characteristic for C₇HAS is the catalyzed reaction to form C₇-pronecic acid in which the activated acetaldehyde formed from PYR is transferred to 2OIV. In addition to this activity, the C₇HAS of *S. officinale* is also capable to catalyze the formation of acetolactate from two molecules of PYR and of 2-acetohydroxy-2-oxobutyrate from one molecule of PYR and 2OB each, exhibiting AHAS function (Chapter 3). Therefore, we tested the three keto acids PYR, 2OB and 2OIV as substrates for the purified enzymes. The assay reactions contained a total of 10 mM substrate but to favor the two-substrate reaction in the presence of PYR, those reactions contained 2.5 mM PYR to allow the formation of the activated acetaldehyde and 7.5 mM of the second keto acid as acceptor of the activated acetaldehyde. After 10 min incubation the reactions were stopped, derivatized with MCF and analyzed by GC-MS. As shown in Table 3, the purified enzyme complex *Ecan*CSU3(RSU) and the *Hi*CSU both accepted PYR and 2OB as acceptor substrates, exhibiting AHAS function. They could however not form the

C₇-pronecic acid by accepting 2OIV, indicating their role in the BCAA metabolism. The *HiC₇HAS*, on the other hand, was able to catalyze the formation of C₇-pronecic acid in addition to the BCAA-specific reactions indicating its role in the PA biosynthesis.

Table 3 Substrates of the tested enzymes. ✓ = keto acid is accepted as acceptor of the activated acetaldehyde.
✗ = keto acid is not accepted as substrate.

	Pyruvate	2-oxobutyrate	2-oxoisovalerate
<i>Ecan</i> CSU3RSU1	✓	✓	✗
<i>Hi</i> CSU	✓	✓	✗
<i>HiC₇HAS</i>	✓	✓	✓

5.3.3 C₇HAS recruitment from AHAS occurred several times independently

For phylogenetic analysis, we used the cDNA sequences coding for the CSUs of the AHAS and the C₇HAS of 17 different angiosperm species as listed in Table 6. Due to incomplete sequences of the closely related *Heliotropium calicicole* and *Heliotropium filiforme* we used only a 1600 bp fragment of the open reading frame (ORF) common to all sequences. Figure 34 shows an unrooted neighbor joining phylogram based on the nucleic acid sequences of the 23 CSUs and the three C₇HAS. The topology of the tree supports three independent origins of the C₇HAS-coding genes from ancestors of the contemporary AHAS CSU genes by gene duplication. C₇HAS was apparently recruited once in the Eupatorieae (Asteraceae) and twice within the Boraginales, once in the Boraginaceae and once in the Heliotropiaceae.

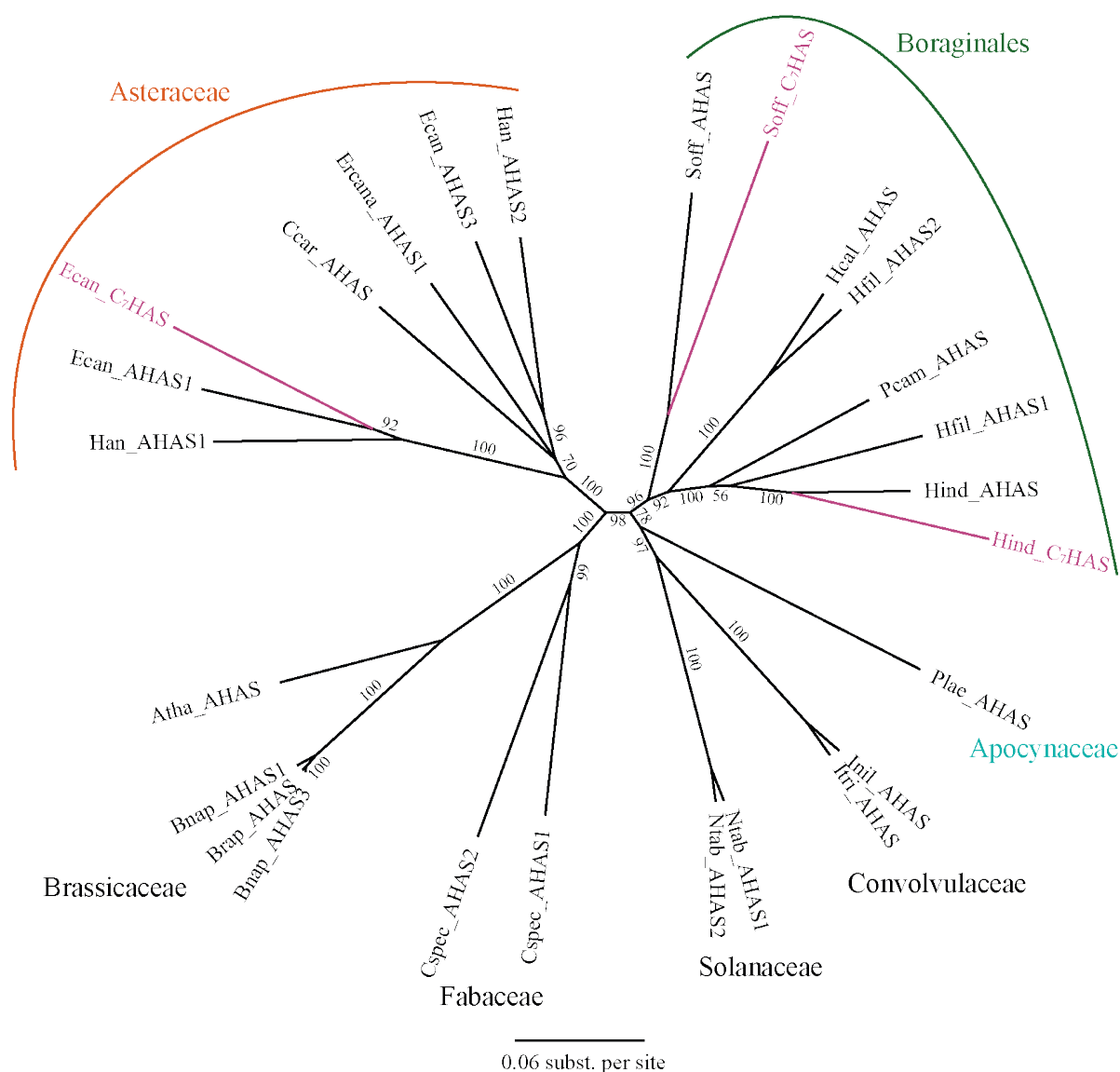


Figure 34 Unrooted neighbor joining tree based on 26 nucleotide sequences coding for AHAS catalytic subunits and C₇HAS of various angiosperm species. Bootstrap supports resulting from 1000 replicates are given below branches. The C₇HAS sequences are marked in pink. The lycopsamine type PA producing genera are marked in color.

Atha = *Arabidopsis thaliana*; Bnap = *Brassica napus*; Brap = *Brassica rapa*; Ccar = *Cynara cardunuculus* var. *scolymus*; Cspec = *Crotalaria spectabilis*; Ecan = *Eupatorium cannabinum*, Ercana = *Erigeron canadensis*; Han = *Helianthus annuus*; Hcal = *Heliotropium calcicola*; Hfil = *Heliotropium filiforme*; Hind = *Heliotropium indicum*; Inil = *Ipomoea nil*; Itri = *Ipomoea triloba*, Ntb = *Nicotiana tabacum*; Pcam = *Phacelia campanularia*; Plae = *Parsonsia laevigata*; Soff = *Symphytum officinale*

5.4 Discussion

The immense diversity of plant secondary metabolites is based on a complex array of biosynthetic pathways. To understand the origin of this diversity, it is necessary to learn about the evolution of secondary metabolic pathways. We address this question by identifying enzymes involved in the biosynthesis of PAs in various plant families and phylogenetically analyzing their evolutionary origin. Here we demonstrated that a C₇HAS is the common first enzyme in the biosynthesis of the unique necic acid moiety of the lycopsamine type PAs in *S. officinale*, *E. cannabinum*, and *H. indicum*. Additionally, we studied the evolutionary origin of the C₇HAS in phylogenetic analyses.

5.4.1 C₇HAS is involved in the PA biosynthesis of *E. cannabinum*

S. officinale (Boraginaceae) and *E. cannabinum* (Asteraceae) are not closely related plant species that both synthesize lycopsamine type PAs of essentially identical structure. The CRISPR/Cas9 mediated inactivation of the C₇HAS encoding gene results in the same phenotype in both plant species: transgenic hairy roots that are deficient of lycopsamine type PAs and accumulate the saturated necine base trachelanthamidine (*S. officinale* Chapter 3). This provided the *in planta* evidence that an C₇HAS is involved in PA biosynthesis in *E. cannabinum*. Furthermore, the accumulation of the same saturated necine base in both plant species suggest a uniform course of PA biosynthesis even in unrelated species producing the same PAs with trachelanthamidine as partner for the esterification. It would be interesting to use CRISPR/Cas9 to inactivate the necic acid biosynthesis of a wide variety of PA-producing plants with different necic acids to see how uniformly PA biosynthesis actually proceeds.

5.4.2 C₇HAS still possesses AHAS CSU activity

For both, *S. officinale* C₇HAS (Chapter 3) and *H. indicum* C₇HAS it was shown that the enzymes still possess AHAS CSU activity in addition to their capability of forming the C₇-pronecic acid. This suggests that the duplicated genes came under relaxed selection and accumulated mutations that broadened the substrate spectrum of the catalyzed reactions. The resulting mechanistic elasticity allows the enzymes to catalyze multiple reactions and biosynthesize multiple products. The ability to accept a broader range of substrates than the related primary metabolic enzymes is a known trait of specialized metabolic enzymes (Lim et al., 2001; Yoshikuni et al., 2006; Kienow et al., 2008; Kopycki et al., 2008). It also fits the classical neofunctionalization model proposed by Ohno (Ohno, 1970) as, in contrast to the C₇HAS, the ancestral AHAS CSUs are not able to catalyze the formation of C₇-pronecic acid.

5.4.3 The gene *ahas csu* was duplicated several times

Despite their ability to synthesize identical lycopsamine type PAs, independent recruitment of HSS was shown for species of the Apocynaceae, the Boraginales and Eupatorieae (Asteraceae) (Reimann et al., 2004). Two possible scenarios arise from this: (1) Independent origin of the whole PA pathway in the three nonrelated taxa, meaning the recruitment of the *hss* gene was followed by independent recruitments of the rest of the pathway leading to the PAs in each case. (2) Monophyletic origin of the PA biosynthesis downstream of HSS. In this case, this part of the biosynthetic pathway has its origin probably in a completely different metabolic context and was recruited once early in the emergence of the angiosperms. Preliminary phylogenetic analysis of the homospermidine oxidases clearly supports an independent evolutionary origin in the Eupatorieae and the Boraginales (Kruse, 2017; Zakaria, 2022) identical to that of the HSS (Reimann et al., 2004). Our phylogenetic analysis of the C₇HAS show an even more complex picture as the C₇HAS seems to have an independent origin in the Eupatorieae (Asteraceae), and two independent origins within the Boraginales. These results indicate that the independent origin of the whole PA pathway is the more likely scenario. This scenario results in fascinating questions about the recruitment of pathways in plant secondary metabolism and their integration into the general metabolism of the plant.

5.4.4 Conclusion

In this work the CRIPSR/Cas9 genome editing technique was established in *E. cannabinum* to elucidate two genes involved in the biosynthesis of lycopsamine type PAs. This provided the first in planta evidence of the involvement of HSS in PA biosynthesis in *E. cannabinum*. In addition, it was shown that a C₇HAS is also involved in the synthesis of C₇-necic acid in *E. cannabinum*. Preliminary studies on the substrate specificity of AHAS CSUs from *H. indicum* indicate that one of the homologs is capable of catalyzing the C₇HAS characteristic reaction to form C₇-pronecic acid. Thus, including *S. officinale*, in three of four lycopsamine type PA-producing families, a C₇HAS is involved in PA biosynthesis. For the Apocynaceae, data is still missing. In phylogenetic analyses of AHAS CSU and C₇HAS the branching patterns of trees generated by neighbor joining, support three independent recruitments of C₇HAS from AHAS CSU during angiosperm evolution, once within the Asteraceae and twice within the Boraginales, once in the Boraginaceae and once in the Heliotropiaceae. Another such recruitment is likely for the Apocynaceae, but this requires identification of the PA-specific C₇HAS in that plant lineage.

5.5 Material and Methods

5.5.1 Material

Restriction endonucleases, DNA-dependent enzymes, and other molecular biology reagents were purchased from Thermo Fisher Scientific (Waltham, MA, USA), New England Biolabs (Ipswich, MA, USA) or Promega (Fitchburg, WI, USA) and, unless stated otherwise, were used in accordance with the manufacturers' instructions. Other chemicals were from Merck (Darmstadt, Germany), Fluka (Seelze, Germany), Carl Roth (Karlsruhe, Germany) or Sigma-Aldrich (St Louis, MO, USA). Oligonucleotides were synthesized at MWG Eurofins Genomics (Ebersberg, Germany).

5.5.2 Plant Material

H. indicum and *E. cannabinum* were grown in an artificial climate chamber under conditions as described previously for *S. officinale* (Chapter 3). Herbarium specimen were deposited in the Herbarium KIEL with the accession number KIEL0023813 (*H. indicum*) and KIEL0023812 (*E. cannabinum*).

5.5.3 CRISPR/Cas9 vector construction and generation of transgenic hairy root mutants

CRISPR/Cas9 vectors were constructed as described before (Chapter 3), with the oligonucleotides given in Table 4.

Table 4 Oligonucleotides for CRISPR/Cas9 vector construction targeting *E. cannabinum* genes. Overhang for ligation underlined.

Nr.	5'-sequence-3'	Purpose
1	<u>ATTGGACCTCTATATGTGTCTGCA</u>	N20 oligo for motif A HSS
2	<u>AAACTGCAGACACATATAGAGGTC</u>	N20 oligo for motif A HSS
3	<u>AAACGATAATGCCTAATAGCAACT</u>	N20 oligo for motif B HSS
4	<u>ATTGAGTTGCTATTAGGCATTATC</u>	N20 oligo for motif B HSS
5	<u>ATTGTTAAATAGATGTCCTGGCCA</u>	N20 oligo for motif C HSS
6	<u>AAACTGGCCAGGACATCTATTTAA</u>	N20 oligo for motif C HSS
7	<u>ATTGAGTATAATCGACCTAGGCAG</u>	N20 oligo for motif A C ₇ HAS
8	<u>AAACCTGCCTAGGTCGATTATACT</u>	N20 oligo for motif A C ₇ HAS
9	<u>ATTGATAATAGCACCCATAGCTGC</u>	N20 oligo for motif B C ₇ HAS
10	<u>AAACGCAGCTATGGGTGCTATTAT</u>	N20 oligo for motif B C ₇ HAS
11	<u>ATTGTCTTGGTGACTCGAGCAGCA</u>	N20 oligo for motif C C ₇ HAS
12	<u>AAACTGCTGCTCGAGTCACCAAGA</u>	N20 oligo for motif C C ₇ HAS

Generation and cultivation of hairy roots was done as described before (Chapter 3) except that the hairy roots were cultivated at 22°C instead of 16°C for faster growth.

5.5.4 Genetic identification of transgenic hairy root lines

Genomic DNA extraction and gene amplification for the identification of transgenic hairy root lines was described before (Chapter 3), with the primers given in Table 5.

Table 5 Oligonucleotides for gene amplification for the identification of transgenic hairy root lines.

Nr.	5'-sequence-3'	Purpose
13	CAGACTTAAGCTAGACGCAAGG	Amplification of genomic HSS to test for mutations, motif A and B forward
14	GTTGCTTCTGAAAAGGTTGCA	Amplification of genomic HSS to test for mutations, motif A and B reverse
15	GCTCAAATGGGTTTCCTTGTA	Amplification of genomic HSS to test for mutations, motif C forward
16	GCCCCATGAGACAGCTTCAT	Amplification of genomic HSS to test for mutations, motif C reverse
17	TCCAGCCTCTGATGAGTTGT	Amplification of genomic C ₇ HAS to test for mutations forward
18	CCACTGGGGATCATAGGCAA	Amplification of genomic C ₇ HAS to test for mutations reverse

5.5.5 Pyrrolizidine alkaloid analysis by GC-MS

PA extraction and analysis from hairy roots was performed as described before (Chapter 3). Identification of PAs on DB-1 capillary column was performed as described before (Chapter 3).

5.5.6 Heterologous expression

E. cannabinum CSUs and RSUs without transit peptides cloned in pCDFDuet-1™ vector with N-terminal 6xHis-tag were provided by Dr. Langel (Langel, 2008).

For *H. indicum* the CSUs without transit peptides in cloned in pCDFDuet-1modstrep vector with N-terminal Strep-tag®II were provided by Britta Muhs (Muhs and Ober, unpublished).

Heterologous expression and affinity purification were performed as described before (Chapter 2).

5.5.7 Enzyme activity assays

The enzyme activity assays have been described previously in Chapter 2.

5.5.8 Derivatization with MCF

The derivatization of the enzymatic reactions with methyl chloroformate has been described before in Chapter 2.

5.5.9 Phylogenetic analysis

Phylogenetics of CSUs were calculated with the program Geneious Prime® (Neighbor-joining algorithm, Geneious Tree Builder, Geneious Prime® 2022.2.2, Biomatters Ltd.) with the sequences given in Table 6 and the following parameters: Genetic distance model: Tamura-Nei (Tamura and Nei, 1993); Tree build method: Neighbor-joining; Resampling method: Bootstrap;

Number of replicates: 1000. The underlying alignment is based on an approximately 1600 bp fragment of the ORF, which was available for all sequences (Figure 35).

Table 6 Sequences for phylogenetic analysis. Published sequences are given with accession numbers. For unpublished sequences the alternative source is given.

Sequence	Accession number	Alternative source
<i>Arabidopsis thaliana</i> AHAS	AY042819.1	
<i>Brassica napus</i> AHAS1	NC_063444.1	
<i>Brassica napus</i> AHAS3	NC_063434.1	
<i>Brassica rapa</i> AHAS	NC_024795.2	
<i>Cynara cardunuculus</i> var. <i>scolymus</i> AHAS	NC_037530.1	
<i>Crotalaria spectabilis</i> AHAS1		Jacky and Ober, unpublished
<i>Crotalaria spectabilis</i> AHAS2		Jacky and Ober, unpublished
<i>Eupatorium cannabinum</i> AHAS1	CAX48741	
<i>Eupatorium cannabinum</i> C ₇ HAS		Langel and Ober, unpublished
<i>Eupatorium cannabinum</i> AHAS3		Langel and Ober, unpublished
<i>Erigeron canadensis</i> AHAS	NC_057762.1	
<i>Helianthus annuus</i> AHAS1	NC_035434.2	
<i>Helianthus annuus</i> AHAS2	NC_035441.2	
<i>Heliotropium calcicola</i> AHAS		OneK Plants Database VJSJ
<i>Heliotropium filiforme</i> AHAS1		OneK Plants Database IPPG
<i>Heliotropium filiforme</i> AHAS2		OneK Plants Database IPPG
<i>Heliotropium indicum</i> AHAS		Muhs and Ober, unpublished
<i>Heliotropium indicum</i> C ₇ HAS		Muhs and Ober, unpublished
<i>Ipomoea nil</i> AHAS	NW_017618707.1	
<i>Ipomoea triloba</i> AHAS	NC_044916.1	
<i>Nicotiana tabacum</i> AHAS1	XM_016639555.1	
<i>Nicotiana tabacum</i> AHAS2	XM_016622884.1	
<i>Phacelia campanularia</i> AHAS		OneK Plants Database YQIJ
<i>Parsonsia laevigata</i> AHAS		Langel and Ober, unpublished
<i>Symphytum officinale</i> AHAS	Chapter 2, this work	
<i>Symphytum officinale</i> C ₇ HAS	Chapter 3, this work	



Figure 35 Sequence alignment of a 1600 bp fragment of the ORF of 23 AHAS CSU and 3 C7HAS sequences.

5.6 Supplementary

Supplementary Table 8 Pyrrolizidine alkaloids in *E. cannabinum* GC EIMS data. PAs No 1-28 as labelled in the total ion chromatogram of **Figure 31**.

No	PA	[M] ⁺	MS data: characteristic ions m/z (% relative abundance)
1	?		55(12), 67(2), 68(3), 70(8), 80(2), 81(2), 82(16), 83(17), 84(3), 95(4), 96(7), 108(3), 110(4), 122(9), 124(100), 125(10), 140(4), 142(16), 240(8), 241(3)
2	Trachelanthamine	285	55(12), 68(4), 70(9), 82(16), 83(19), 84(6), 95(4), 96(9), 108(4), 110(4), 122(10), 123(3), 124(100), 125(15), 140(4), 142(50), 143(5), 240(4), 252(7), 267(4)
3	Supinine	283	53(6), 65(2), 67(4), 70(11), 71(2), 73(2), 79(2), 80(14), 81(2), 92(3), 93(23), 94(10), 95(2), 106(2), 108(15), 120(58), 121(37), 122(100), 123(24), 140(7)
4	?		55(13), 67(2), 68(4), 70(9), 82(17), 83(19), 84(7), 95(4), 96(10), 108(3), 110(4), 122(8), 124(100), 125(13), 140(4), 142(57), 143(5), 240(4), 252(7), 267(4)
5	?		55(13), 68(4), 70(9), 73(6), 82(16), 83(23), 84(4), 95(5), 96(9), 108(4), 110(7), 111(16), 122(26), 123(5), 124(100), 125(11), 138(6), 140(12), 142(32), 252(4)
6	?		55(13), 68(3), 70(10), 80(3), 81(3), 82(15), 83(26), 95(4), 96(9), 97(3), 108(3), 110(4), 122(8), 123(3), 124(100), 125(12), 140(4), 142(36), 143(3), 252(4)
7	?		55(8), 70(6), 81(3), 82(14), 83(16), 84(3), 95(5), 96(6), 99(8), 108(3), 122(10), 123(4), 124(100), 125(16), 140(4), 142(19), 184(6), 225(4), 240(6), 284(3)
8	Intermedine	299	53(5), 65(4), 66(3), 67(11), 73(3), 80(19), 82(6), 92(5), 93(95), 94(78), 95(19), 108(5), 118(3), 120(12), 136(15), 137(13), 138(100), 139(34), 156(8), 217(4)
9	?		53(4), 65(3), 66(3), 67(10), 80(16), 81(3), 82(6), 92(3), 93(81), 94(71), 95(16), 96(3), 108(4), 120(11), 136(13), 137(13), 138(100), 139(27), 140(3), 156(8)
10	?		53(4), 57(2), 65(3), 66(2), 67(8), 73(2), 80(13), 93(80), 94(28), 95(10), 96(5), 108(3), 111(2), 120(8), 121(2), 136(10), 137(10), 138(100), 139(32), 156(8)
11	?		53(4), 65(3), 66(2), 67(8), 73(2), 80(14), 81(2), 92(4), 93(69), 94(36), 95(9), 96(3), 108(3), 120(8), 124(6), 136(10), 137(11), 138(100), 139(29), 156(8)
12	3'-tigloyltrachelanthamine	439	55(15), 70(5), 82(10), 83(42), 84(4), 95(3), 96(7), 108(3), 110(3), 122(11), 123(5), 124(100), 125(12), 140(4), 142(12), 224(4), 240(13), 241(3), 284(5), 367(5)
13	?		53(5), 55(12), 65(2), 67(4), 70(5), 79(2), 80(11), 83(33), 84(2), 92(2), 93(17), 94(6), 106(2), 108(13), 118(2), 120(93), 121(59), 122(100), 123(17), 140(3)
14	?		55(26), 70(12), 82(14), 83(68), 84(6), 96(13), 120(6), 121(4), 122(19), 123(6), 124(100), 125(13), 138(10), 139(29), 140(34), 141(4), 142(13), 224(5), 240(18), 284(6)
15	?		55(18), 80(13), 83(19), 92(7), 93(35), 94(37), 95(9), 106(9), 108(5), 118(11), 119(23), 120(90), 121(42), 122(8), 136(58), 141(28), 192(5), 220(100), 221(30), 238(9)
16	?		53(3), 55(8), 65(2), 66(2), 67(6), 80(9), 83(37), 92(2), 93(57), 94(39), 95(3), 108(4), 109(2), 136(18), 137(27), 138(100), 139(16), 148(9), 155(3), 194(2)
17	Echiumine	381	55(23), 67(9), 80(20), 83(49), 93(73), 94(59), 95(11), 106(9), 108(8), 118(11), 119(21), 120(96), 121(39), 136(74), 137(14), 138(32), 141(30), 220(100), 221(30), 238(10)
18	Symphytine	381	55(16), 67(9), 80(20), 83(76), 93(77), 94(69), 95(12), 106(9), 108(7), 118(14), 119(25), 120(100), 121(46), 136(100), 137(10), 138(15), 141(32), 220(99), 221(26), 238(9)

19	?		55(19), 67(6), 80(11), 83(29), 93(60), 94(44), 95(6), 99(13), 106(7), 117(8), 118(10), 119(31), 120(85), 121(10), 136(64), 137(7), 141(26), 219(8), 220(100), 221(23)
20	?		55(5), 67(4), 80(9), 83(6), 91(4), 93(54), 94(35), 99(6), 106(7), 117(6), 118(16), 119(100), 120(85), 121(24), 136(51), 137(5), 141(19), 220(58), 221(18), 336(3)
21	Echimidine	397	55(11), 67(7), 80(14), 83(65), 93(70), 94(54), 95(6), 99(12), 106(8), 117(9), 118(12), 119(42), 120(85), 121(19), 136(91), 137(10), 138(9), 141(29), 220(100), 221(23)
22	Echimidine isomer	397	55(22), 80(16), 82(9), 83(44), 93(77), 94(50), 95(10), 99(14), 106(9), 117(9), 118(12), 119(47), 120(99), 121(23), 136(60), 137(9), 138(14), 141(38), 220(100), 221(23)
23	?		55(19), 67(7), 80(14), 83(28), 93(64), 94(52), 95(8), 99(15), 106(9), 117(8), 118(14), 119(56), 120(100), 121(22), 136(67), 137(7), 141(29), 219(8), 220(99), 221(25)
24	?		55(50), 80(17), 82(51), 83(66), 93(55), 94(32), 95(32), 96(23), 99(15), 106(12), 117(13), 118(24), 119(100), 120(99), 121(18), 136(52), 140(13), 141(17), 220(65), 221(19)
25	3'-acetylmyoscorpine	423	53(9), 55(15), 67(11), 80(22), 82(8), 83(78), 93(68), 94(41), 99(10), 104(10), 117(17), 118(19), 119(39), 120(64), 121(10), 136(94), 137(11), 141(12), 220(100), 221(20)
26	3'-acetylsymphytine	423	53(10), 55(44), 67(9), 80(24), 83(23), 93(66), 94(49), 99(14), 106(10), 108(9), 117(15), 118(19), 119(49), 120(84), 121(13), 136(64), 141(12), 219(9), 220(100), 221(22)
27	?		53(12), 55(29), 67(11), 80(22), 82(10), 83(22), 93(59), 94(32), 99(10), 106(10), 117(17), 118(30), 119(100), 120(95), 121(18), 136(54), 137(9), 141(14), 220(61), 221(19)
28	?		53(3), 55(18), 70(7), 71(3), 82(19), 83(20), 95(4), 96(8), 97(32), 99(4), 108(4), 110(4), 122(10), 123(3), 124(100), 125(13), 140(3), 142(8), 240(4), 409(4)

6 Conclusion and general perspectives

The results presented in this work expand the knowledge in the research field of PA biosynthesis, especially that of the structurally unique necic acid moieties of lycopsamine type PAs. In this context, it turned out that the approaches described in the individual chapters synergistically allowed the breakthrough in the unequivocal identification of the enzymatic machinery involved. This data provides a good starting point for further research studies at the molecular, biochemical, analytical, and evolutionary levels.

An assay method for the elucidation of necic acid biosynthesis

Past attempts to elucidate the enzymatic reaction postulated by Weber et al. (1999) for the formation of the backbone of the C₇-necic acid subunit of lycopsamine type PAs were hampered by the fact that no reproducible method was available that would have allowed not only to quantify the enzymatic activity, but also to analyze the products. The aqueous environment, the volatile reaction products and their instability in acidic milieu turned out to be the major challenges that could not be addressed with the known assay methods for AHAS characterization (Schloss et al., 1985; Singh et al., 1988; Vinogradov et al., 2005; Xie et al., 2019). With the ECF derivatization assay presented in **Chapter 2**, this challenge was finally mastered. In this regard, the method represents an important advance in the characterization of AHASs, as it is the first assay method in which the substrates and products of the enzyme reaction can be detected and quantified simultaneously. Analyses of the acceptance and conversion of different substrates in a single assay became possible. The determination of the assay products by GC-MS allows their identification even without standard substances. In addition, the method might in the future facilitate the elucidation of the biosynthesis of necic acid moieties that built up other PA types as it was suggested in the literature that also for other necic acids the precursors are derivatives of the BCAA biosynthesis (Crout et al., 1966; Crout, 1967; Crout et al., 1970; Crout et al., 1972; Robins et al., 1974; O'Donovan and Long, 1975; Bale et al., 1978; Devlin and Robins, 1984).

C₇-necic acid formation is a unique feature of PA specific enzymes

In **Chapter 3**, the first step of the necic acid formation of lycopsamine type PAs was elucidated for the first time. In *S. officinale* a catalytic subunit of an AHAS, now named SoC₇HAS was shown to form an activated acetaldehyde from PYR and transferring it to 2-oxoisovalerate, forming acetohydroxy-2-oxoisovalerate, the so-called C₇-pronecic acid that is the direct precursor of trachelanthic acid and viridifloric acid. The detection of the same catalytic activity

allowed the identification of a C₇HAS also from *H. indicum* (**Chapter 5**). This PA-specific transfer of a C₂-unit cannot be catalyzed by the tested enzymes of plant primary metabolism or the bacterial AHAS. Therefore, the evolution of this new activity in AHAS duplicates can be regarded as a neofunctionalization with respect to the common models of enzyme evolution (Conant and Wolfe, 2008). Enzymes and intermediates of BCAA biosynthesis have been previously reported to be recruited for specific functions in plant secondary metabolism. In the Brassicaceae, the methylthioalkylmalate synthases involved in the chain elongation of glucosinolate biosynthesis, appear to be evolutionarily derived from isopropylmalate synthase of leucine biosynthesis (de Kraker et al., 2007; Benderoth et al., 2008; de Kraker and Gershenzon, 2011). Additionally, *A. thaliana* knock-out mutants of a branched-chain aminotransferase homolog demonstrated its involvement in chain elongation of methionine-derived glucosinolates (Schuster et al., 2006; Sawada et al., 2009). And just recently, Chang and colleagues found strong evidence that an AHAS-like enzyme is involved in the biosynthesis of acyl sugars in the glandular trichomes of *Nicotiana tabacum* (Chang et al., 2020). It remains exciting to see in the future which other plant secondary metabolites are built up with precursors and enzymes recruited from BCAA biosynthesis.

The value of CRISPR/Cas9 for the research of pathways in non-model plants

The discovery of the CRISPR/Cas9 genome editing system and its universal application in all eukaryotic species (Doudna and Charpentier, 2014) plays a crucial role in the elucidation of specialized metabolism pathways in plants. CRISPR/Cas9 facilitates the targeted editing of a specific DNA sequence without the need for time-consuming assembly of constructs as it is the case for other targeted genome editing techniques like zinc finger nucleases (Carroll, 2011) or transcription activator-like effector nucleases (Zhang et al., 2013). Previously, our group followed the strategy to post transcriptionally silence the putative genes involved in PA biosynthesis with RNA interference. While this was successful for the HSS of *S. officinale* (Kruse et al., 2019), no significant effects could be detected for any of the other genes studied (Engelhardt and Ober, unpublished). However, using the CRISPR/Cas9 system in *S. officinale* targeting the C₇HAS-encoding gene was one of the first successful CRISPR/Cas9 genome editing experiments in *Symphytum* (**Chapter 3**). The resulting transgenic hairy roots provided *in planta* evidence that the C₇HAS is involved in PA biosynthesis, as they were no longer able to synthesize lycopsamine type PAs but rather accumulate the saturated necine base trachelanthamidine. Investigating PA biosynthesis of various Boraginales, Frölich et al. (2007) identified trachelanthamidine as the first common necine base, confirming previous studies in

Senecio species (Kunec and Robins, 1986). Surprisingly, analyses of the incorporation rate of different radiolabeled necine bases fed to the PA-producing plant tissues suggested a different biosynthetic sequence for each of the Boraginales species studied: (1) For *H. indicum*, structural modification of the necine base trachelanthamidine occurs via supinidine and retronecine prior to the O^9 -esterification with the necic acid. (2) For *Cynoglossum officinale*, trachelanthamidine was identified as the most likely partner of O^9 -esterification and 1,2-desaturation and C^7 -hydroxylation are subsequent steps. (3) In *S. officinale*, supinidine was the base most efficiently converted to the alkaloid esters, suggesting that O^9 -esterification follows 1,2-desaturation but precedes C^7 -hydroxylation (Frölich et al., 2007). The data provided in this thesis clearly shows that trachelanthamidine is accumulating in *S. officinale* C_7 HAS knock-out mutant lines making this compound the most likely candidate to be esterified with the necic acid moieties. Obviously, the observation that a specific compound shows the highest incorporation rate can only be a hint but no proof that this compound is indeed a natural occurring intermediate of the biosynthetic pathway.

A proof-of-concept experiment targeting the HSS from *E. cannabinum* by the CRISPR/Cas9 genome editing technique confirmed the applicability of this technique for another non-model system (**Chapter 5**). Using this method, the inactivation of the C_7 HAS in *E. cannabinum* confirmed its *in planta* role for PA biosynthesis with identical effects as in *S. officinale* (**Chapter 5**). The identity of effects in the two species is remarkable, because based on the data of HSS, PA biosynthesis in *Symphytum* and *Eupatorium* have been established independently during angiosperm evolution.

Regeneration of plants from C_7 HAS knock-out hairy roots allowed us to gain deeper insights into the regulation of PA biosynthesis in *S. officinale* than we could have imagined at the beginning of the experiments (**Chapter 3**). That the accumulating trachelanthamidine remains in the roots and is not translocated to the aboveground organs of the rosette plant confirms earlier experiments conducted in *Senecio vernalis* (Senecioneae). Hartmann and Dierich (1998) showed that the biosynthesis of the PA core structure takes place in the specialized tissue and no intermediates of biosynthesis are transported within the plants. Recently, by analyzing the inflorescences of *S. officinale*, it was confirmed that the protection of the reproductive organs is strengthened by a boost of lycopsamine type PAs synthesized in the young leaves below the inflorescences (Kruse et al., 2017). Analyses of a flowering plant lineage with inactivated C_7 HAS showed also an accumulation of trachelanthamidine, but no lycopsamine type PAs in the inflorescences, confirming that C_7 HAS expressed in the roots and in the young leaves near the inflorescences of *S. officinale* is encoded by only a single gene. Of note, the fact that

trachelanthamidine produced in the young leaves is translocated into the inflorescences raises the question whether in these tissues an additional or a different transporter is involved than in roots, which does not transport the saturated base. To our knowledge, in *Symphytum* only one HSS is involved in PA biosynthesis at both biosynthetic sites (Kruse et al., 2017) while there are two different HSOs, one for each biosynthetic site (Zakaria and Ober, unpublished). With again only one C₇HAS, *S. officinale* clearly demonstrates the complexity and diversity of plant secondary metabolism, its evolution and regulation.

With the previously mentioned knock-out plants, not only conclusions on the *in planta* significance of the respective enzymes are possible, but also studies on biosynthetic pathways that have not yet been elucidated. For example, the accumulation of preceding intermediates of the targeted enzymes may allow conclusions about the position of this reaction in the sequence of enzymatic steps within PA biosynthesis. Furthermore, accumulating compounds can be purified and serve as substrates for *in vitro* biochemical reactions for enzymes catalyzing the following step of the pathway, which is an immense advantage, especially for uncommon compounds that are not commercially available. Zakaria et al. (2021) have used the reaction product of HSS, *i.e.*, homospermidine, to feed to the HSS knock-out mutants of *S. officinale* to restore PA biosynthesis in this plant lineage confirming that the observed inability to produce PAs was indeed due to an inactivated HSS and not to off-target effects.

The -omics era

From the first insights into the origin of the necic acid moieties of PAs in the 1960s to the first suggestion of an enzyme possibly involved in their formation, almost 40 years have passed, to the confirmation of the enzyme and the enzymatic characterization of the catalyzed reaction, 23 more years. The era of ‘-omics’ data and bioinformatic processing provides a much more holistic view of biosynthetic pathways and compounds involved, what accelerates and simplifies their elucidation considerably (Oksman-Caldentey et al., 2004; Oksman-Caldentey and Saito, 2005). For example, the combination of different ‘-omics’ approaches has led to the elucidation of alkaloid biosynthesis in *Macleaya* species (Zeng et al., 2013) and to the better understanding of the stress responses of *A. thaliana* to nutrient deficiencies (Hirai et al., 2004). The countless possible sequences of biosynthetic steps for the formation of the *N*-oxidized core structure of the lycopsamine type PAs in *S. officinale* could be reduced to 125 by GC-MS analyses (**Chapter 3**) and to 2 by analyzing the metabolomes of C₇HAS knock-out and control lines using LC-MS/MS (**Chapter 4**). To have an idea about the sequence of the individual

enzymatic steps should facilitate the identification of further enzymes of the pathway as substrates and enzymatic mechanisms required for the individual reaction can be predicted. The combination of the metabolomic approach presented in **Chapter 4** and the identification of appropriate candidate genes of PA biosynthesis from subtractive next-generation-sequencing-based transcriptome databases as shown for *H. indicum* (Sievert et al., 2015) will hopefully greatly accelerate the elucidation of the following steps of the PA biosynthesis.

7 Outlook

Some general aspects for future research based on the different perspectives and approaches developed in this work have been discussed in general terms in the previous section. In this section, I would like to be more specific about the possible approaches that might arise from the results of my research.

Feeding of knock-out hairy roots with product of enzyme reaction

Feeding the C₇HAS knock-out hairy roots of *S. officinale* and *E. cannabinum* with the product of the enzyme reactions would complement the *in planta* and *in vitro* experiments presented in **Chapter 3 and 5**. The incorporation of C₇-pronecic acid into the lycopsamine type PAs would confirm its role in PA biosynthesis. First attempts with knock-out lines of *S. officinale* have not been successful so far. This may be due to the slightly acidic pH (5.8) of the medium in which the hairy root cultures are grown. The acetohydroxyacids are decarboxylated in the acidic milieu and their corresponding 2,3-diketones are formed. This should be considered repeating the feeding experiment by growing the hairy roots in a medium with neutral or even slightly basic pH.

C₇HAS in *E. cannabinum*

Three homologs of the CSU of an AHAS were identified, two that were detected in the stems and one that is root-specific (Langel, 2008). The latter was identified as the C₇HAS of PA biosynthesis by generating CRISPR/Cas9 knockout mutant hairy roots of *E. cannabinum* (**Chapter 5**). All three CSU-like sequences from *E. cannabinum* are prepared as co-expression constructs with RSU cloned into the pCDFDuet vector with N-terminal 6xHisTag. In previous experiments, only the stem specific *EcanCSU3* was successfully overexpressed as a co-construct with RSU1 and exhibited AHAS activity. This AHAS was however not able to catalyze the PA specific reaction (**Chapter 5** and Langel, 2008). Since expression with N-terminal Strep-tag®II with the sequences described in this thesis has proven to be successful, the *E. cannabinum* sequences should also be cloned with this tag in expression vectors to facilitate heterologous expression efforts.

C₇HAS in *H. indicum*

Phylogenetic analyses of the C₇HAS of the two species studied of the Boraginales, *i.e.*, *S. officinale* and *H. indicum*, have been suggested to be of different evolutionary origin (**Chapter 5**), while for HSS a common origin within these two species was shown (Reimann et

al., 2004). It would be interesting to investigate differences and similarities between the enzymes from *S. officinale* and *H. indicum*. As described in **Chapter 5**, the AHAS-like enzymes from *H. indicum* were successfully overexpressed and initial substrate specificity experiments were conducted. However, a comprehensive characterization of the enzymes is still necessary. Another interesting aspect is the prediction by Frölich et al. (2007), by which the structural modification of the necine base trachelanthamidine occurs via supinidine and retronecine prior to the *O*⁹-esterification with the necic acid. CRISPR/Cas9 genome editing targeting the C₇HAS should accordingly lead to an accumulation of retronecine.

Changes in the substrate access tunnel

The fact that the C₇HAS of *S. officinale* is able to transfer the activated acetaldehyde also to branched-chain keto acids, an activity that was not described for the characterized CSU of *A. thaliana*, suggests differences in the substrate access tunnel of the enzyme. The AHAS isoenzyme II of *E. coli* exhibits a higher preference to 2-oxobutyrate than to PYR than the other two AHAS of *E. coli*, thus preferring a bigger substrate to a smaller one (Barak et al., 1987). Several amino acids of AHAS isoenzyme II were identified and mutated to determine their role in this change. Replacements of Arg276 and Trp464 by the amino acids that are present in the other two AHAS lead enzymes with 60 - 120-fold or 6 - 20-fold reduced preference for 2-oxobutyrate, respectively, in comparison to the wild type enzymes, indicating crucial roles for both amino acids (Ibdah et al., 1996; Engel et al., 2004b). In *A. thaliana* these positions correspond to Arg377 and Trp574 and these amino acids are highly conserved throughout all plant species. Protein alignments of AHAS CSUs from *A. thaliana*, *S. officinale*, *E. cannabinum* and *H. indicum* as well as the C₇HAS of the latter three show 72.7 % pairwise identity, indicating that most parts of the sequences are highly conserved (Figure 36A). All aligned proteins have the amino acids arginine and tryptophan in the positions 377 and 574, respectively. Comparing the protein sequence of the CSUs and C₇HASs showed no variation in the direct surroundings of Arg377. For Trp574 however, the C₇HASs show deviations from the primary enzymes in positions Val571Ala, Arg577His/Cys/Val and Phe578Tyr. As shown in Figure 36 B, all these amino acids show a close proximity to the substrate access tunnel with ThDP in its midst when located in the AHAS protein model of *A. thaliana* (Lonhienne et al., 2020). All of these mutations might be crucial for the ability of C₇HAS to accept the branched-chain keto acids and provide a good starting point for a site directed mutagenesis approach. Preliminary results of the mutagenesis of Val571 to alanine in the CSU of *S. officinale* indicated that this mutation is already enabling the enzyme to accept

2-oxoisovalerate as a substrate. It should now also be tested what influence the mutagenesis of the other proposed amino acids individually and in combination with each other have on substrate specificity.

A

A

							Val571				Trp574				Arg577	Phe578				
Consensus		Gln	His	Leu	Gly	Met	Val	Val	Gln	Trp	Glu	Asp	Arg	Phe	Tyr	Lys	Ala	Asn	Arg	
Identity																				
CSU AHAS	<i>A. thaliana</i>	Gln	His	Leu	Gly	Met	Val	Met	Gln	Trp	Glu	Asp	Arg	Phe	Tyr	Lys	Ala	Asn	Arg	
	<i>S. officinale</i>	Gln	His	Leu	Gly	Met	Val	Val	Gln	Trp	Glu	Asp	Arg	Phe	Tyr	Lys	Ala	Asn	Arg	
	<i>E. cannabinum</i> I	Gln	His	Leu	Gly	Met	Val	Val	Gln	Trp	Glu	Asp	Arg	Phe	Tyr	Lys	Ala	Asn	Arg	
	<i>E. cannabinum</i> III	Gln	His	Leu	Gly	Met	Val	Val	Gln	Trp	Glu	Asp	Arg	Phe	Tyr	Lys	Ala	Asn	Arg	
	<i>H. indicum</i>	Gln	His	Leu	Gly	Met	Val	Val	Gln	Trp	Glu	Asp	Arg	Phe	Tyr	Lys	Ala	Asn	Arg	
C ₇ HAS	<i>S. officinale</i>	Gln	His	Leu	Gly	Met	Ala	Val	Gln	Trp	Glu	Asp	His	Tyr	Tyr	Asn	Ser	Asn	Arg	
	<i>E. cannabinum</i>	Gln	Tyr	Leu	Gly	Met	Val	Val	Gln	Trp	Glu	Asp	Cys	Tyr	Tyr	Lys	Ala	Asn	Arg	
	<i>H. indicum</i>	Gln	His	Tyr	Gly	Met	Ala	Val	Gln	Trp	Glu	Asp	Val	Phe	Tyr	Gln	Ala	Asn	Arg	

B

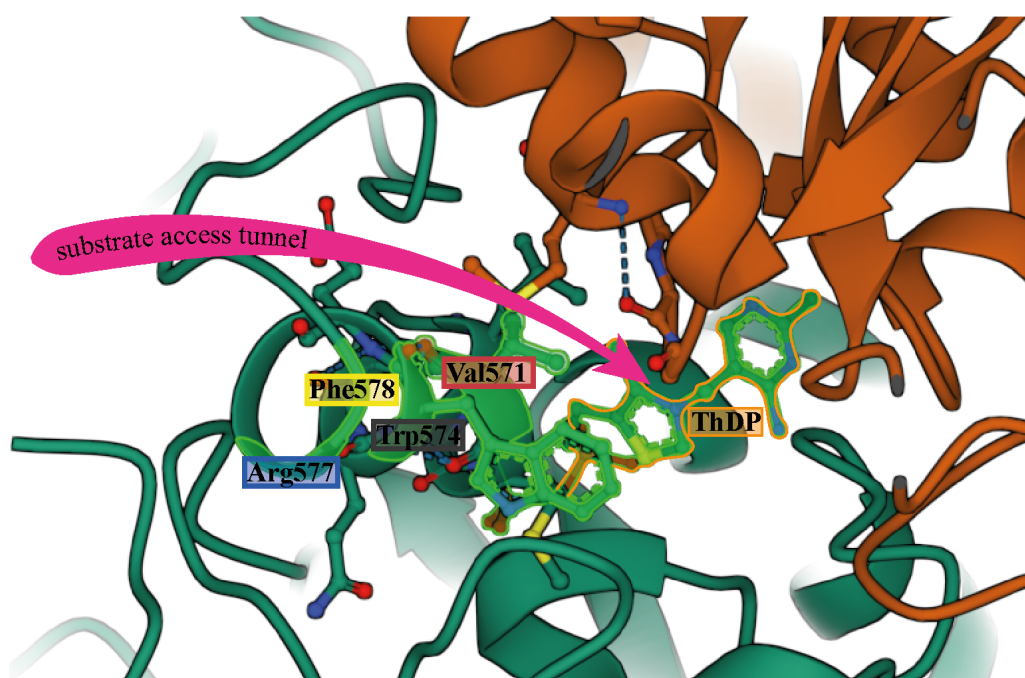


Figure 36 Substrate access tunnel. **A** Alignment of AHAS catalytic subunits (CSUs) and C₇-hydroxyacid synthases (C₇HAS) from *A. thaliana* (ABJ80681.1), *S. officinale*, *E. cannabinum*, and *H. indicum*. Numbering of amino acid positions refer to the *A. thaliana* sequence. Grey box refers to Trp574 that was shown to be crucial for the preference for 2-oxobutyrates in AHASs. Colored boxes refer to amino acids Val571, Arg577, and Phe578 that are different in the C₇HAS sequences. **B** 3D-model of the substrate access tunnel in the *A. thaliana* AHAS complex (PDB entry 6U9H, Lonhienne et al., 2020) with the spatial arrangement of the amino acids labeled in **A**. The brown and green tertiary structures are the two separate CSU subunits of the homodimer.

Tissue-specific localization of C₇HAS

To compare the tissue-specific expression of the C₇HASs in the different plants with those of the HSS and possibly the HSO, I have already prepared polyclonal antibodies specific for the C₇HAS of *S. officinale* and *H. indicum*. Specificity tests and immunodetection could be carried out as described previously (Niemüller et al., 2012). The low expression levels, as shown in the

qPCR experiments of *S. officinale*, and the high sequence similarities between the primary and secondary metabolic enzymes will be a particular challenge.

Elucidation of the steps subsequent trachelanthamidine and C7-pronecic acid formation in *S. officinale*

As mentioned in **Chapter 4**, an attempt was made to clarify whether trachelanthamine is first hydroxylated at C⁷ or *N*-oxygenated. If it is the latter, reducing the resulting trachelanthamine *N*-oxide to its tertiary form should result in an increase of trachelanthamine in the sample. At the same time, a new mass feature of 7-hydroxytrachelanthamine should be detectable. If the identified intermediate is indeed 7-hydroxytrachelanthamine the signal intensity of this mass feature should increase due to the reduction of 7-hydroxytrachelanthamine *N*-oxide whilst the amount of trachelanthamine stays the same. Preliminary tests were made to reduce the samples with zinc to convert all *N*-oxides to their tertiary form. This reduction was either not complete or the samples without zinc were spontaneously reduced too much to make a quantifiable comparison. This experiment is worth to be repeated to elucidate the order of hydroxylation and *N*-oxygenation in the biosynthesis.

8 Contributions

The method development described in **Chapter 2** was designed by me and my supervisor Dietrich Ober. The transit peptide length was determined by Dorothee Langel and Britta Muhs. The alteration of pCDFDuet into pCDFDuetmodstrep and the cloning of the *SoCSU* was done by Britta Muhs. I developed the MCF and ECF assay and the GC-MS method with help from Elisabeth Kaltenegger. I performed all experiments and the sampling, analyzed the data, and interpreted the results under the supervision of Dietrich Ober. I prepared all figures and wrote the manuscript together with Dietrich Ober.

The study described in **Chapter 3** was designed by me and my supervisor Dietrich Ober. The transit peptide length was determined by Dorothee Langel and Britta Muhs. The cloning of the *SoC₇PNS* into pCDFDuetmodstrep was done by Britta Muhs. Sebastian Fritz performed the transcript level analysis under my supervision. I did all other experiments and analyses. I interpreted the results under the supervision of Dietrich Ober. I prepared all figures and wrote the manuscript. Dietrich Ober discussed the results with me and reviewed the manuscript.

The presented study in **Chapter 4** was designed by me under the supervision of Dietrich Ober. Riya Christina Menezes prepared the samples and did the LC-MS/MS Orbitrap measurements under the supervision of Aleš Svatoš. Databases were generated by Riya Christina Menezes and me. Riya Christina Menezes also did preliminary data interpretation for the identification of the accumulating intermediate and helped with first data interpretations. I analyzed all data, interpreted the results, prepared the figures, and wrote the manuscript. Dietrich Ober provided comments on the manuscript.

The experiments in **Chapter 5** were designed by me under the supervision of Dietrich Ober. Sebastian Fritz designed the *EcanC₇HSS* CRISPR/Cas9 targeting sequences, did the GC-MS analysis, and the sequence analysis of some of the hairy root lines presented here under my supervision. The cloning of the *Heliotropium indicum* genes was done by Britta Muhs. I did all the other experiments. I analyzed all data, interpreted the results, prepared all figures, and wrote the manuscript. Dietrich Ober discussed the results with me and reviewed the manuscript.

9 Acknowledgements

Now that the mandatory part of this work has been completed, it is time to thank all those who have contributed to making this thesis a reality:

Als Erstes möchte ich meinem Doktorvater Prof. Dr. Dietrich Ober danken. Dafür, dass Du mir die Möglichkeit gegeben hast meine Doktorarbeit bei Dir anzufertigen. Dafür, dass Du an mich geglaubt und mich unterstützt hast, sodass ich als Wissenschaftlerin, aber auch als Mensch wachsen konnte. Vielen Dank für Dein Vertrauen und Deine Unterstützung.

Des Weiteren möchte ich mich bei Dr. Riya Christina Menezes und Prof. Dr. Aleš Svatoš für die Möglichkeit bedanken, meine Wurzeln mittels LC-MS/MS zu vermessen. Riya möchte ich insbesondere auch für die geduldige Einweisung in die Auswertung von MS² Daten danken.

Doro und Britta Muhs möchte ich für die umfassende Suche und Klonierung der Sequenzen danken. Durch Euch konnte ich mit einer hervorragenden Datenlage in meine Promotion starten!

Insgesamt möchte ich meinen aktuellen und ehemaligen Co-DoktorandInnen und unseren Postdocs Nadine, Adila, Britta, Mahmoud, Elisabeth, Doro, Lars, Moritz, Thomas und Arun meinen Dank für die Hilfe und Unterstützung während meiner Doktorarbeitszeit aussprechen. Für das schönste, grünste Büro mit toller Gemeinschaft geht ein großer Dank an meine Bürokolleginnen Nadine Jacky und Adila Shahid. Vielen Dank für die anregenden Gespräche, das gemeinsame Fluchen und die Unterstützung bei Schwierigkeiten. Ohne Euch wäre ich aufgeschmissen gewesen.

Cedric Hering möchte ich für seine langjährige Unterstützung, erst als HiWi, dann als TA, bei der Pflege der hairy roots und der Probenvorbereitung danken. Ich wünsche Dir viel Erfolg für Deine eigene wissenschaftliche Karriere.

Unseren technischen Assistentinnen Brigitte Schemmerling, Margaret Doose und Susanne Klocke für die Hilfe im Labor, das Versorgen von hairy roots und Pflanzen und insgesamt die Unterstützung. Danke, dass ihr alles am Laufen haltet. Brigitte, unsere gemeinsamen Mittagessen und Gespräche werde ich sehr vermissen. Ich wünsche Dir alles Gute für Deine Zeit als Rentnerin und hoffe Dir wird es nicht zu langweilig.

Eine Tradition die ich schon zu meiner Bachelorarbeitszeit lieben gelernt habe ist das Freitagfrühstück. Der interdisziplinäre Austausch hat mir geholfen auch immer über den Tellerrand hinaus zu blicken. Danke an Brigitte, Frau Sobecki, Dietrich, Susanne P., Tom, Susanne K., Doro und Heinz für die anregenden Gespräche und angenehme Atmosphäre.

Ein ganz besonderer Dank geht an die Studierenden, die ich im Rahmen von Praktika, Bachelorarbeiten und Masterarbeiten betreuen durfte. Felix Gilbert, Cedric Hering, Sebastian

Fritz, Surashai Amphonephong, Lisa Marie Doose, Paul Heinecke, Maxime Faroux and Nikita Kamatkar. Ich hoffe ihr habt die Erfahrung unter meiner Anleitung eure Projekte zu erarbeiten als bereichernd empfunden. Ich fand es sehr bereichernd Euch zu Betreuen und habe dabei selber viel lernen können. Danke, dass Ihr Euer Vertrauen in mich gesetzt habt!

Jule, Dir möchte ich dafür danken, dass Du immer da warst um mein Jammern zu ertragen, aber auch für die Unterstützung und die vielen schönen gemeinsamen Erlebnisse. Und einen riesigen Dank, dass Du diese Arbeit abschließend nach all den kleinen Fehlern durchsucht hast, für die man am Ende einfach blind ist!

Mein herzlichster Dank geht an meine Familie und insbesondere meinen Mann. Die nie an mir gezweifelt haben. Die immer für mich da sind. Auf deren Unterstützung ich einfach immer zählen kann, auch wenn sie eigentlich keine Ahnung haben, was ich eigentlich mache! Danke. Marcel, vielen Dank für Deine Geduld. Ohne Dich und Deinen unerschütterlichen Glauben an mich hätte ich diese Thesis nicht vollenden können.

Und zu guter Letzt möchte ich mich bei all denjenigen bedanken, die in irgendeiner Form zu dieser Arbeit beigetragen haben, sei es persönlich oder wissenschaftlich, die aber hier nicht erwähnt wurden. Danke, dass ihr mir geholfen habt, die Person zu werden, die ich jetzt bin, und dass ihr diese Arbeit möglich gemacht habt.

10 References

- Afendi FM, Okada T, Yamazaki M, Hirai-Morita A, Nakamura Y, Nakamura K, Ikeda S, Takahashi H, Altaf-Ul-Amin Md, Darusman LK, et al (2012) KNApSAC Family Databases: Integrated Metabolite–Plant Species Databases for Multifaceted Plant Research. *Plant and Cell Physiology* 53: e1
- Agnew-Francis KA, Tang Y, Lin X, Low YS, Wun SJ, Kuo A, Elias SMASI, Lonhienne T, Condon ND, Pimentel BNAS, et al (2020) Herbicides That Target Acetohydroxyacid Synthase Are Potent Inhibitors of the Growth of Drug-Resistant *Candida auris*. *ACS Infectious Diseases* 6: 2901–2912
- Anke S, Gondé D, Kaltenegger E, Hänsch R, Theuring C, Ober D (2008) Pyrrolizidine Alkaloid Biosynthesis in *Phalaenopsis* Orchids: Developmental Expression of Alkaloid-Specific Homospermidine Synthase in Root Tips and Young Flower Buds. *Plant Physiology* 148: 751–760
- Anke S, Niemüller D, Moll S, Hänsch R, Ober D (2004) Polyphyletic Origin of Pyrrolizidine Alkaloids within the Asteraceae. Evidence from Differential Tissue Expression of Homospermidine Synthase. *Plant Physiology* 136: 4037–4047
- Bale N, Cahill R, Davies MN, Mitchell M, Smith EH, Crout DHG (1978) Biosynthesis of the necic acids of the pyrrolizidine alkaloids. Further investigations of the formation of senecic and isatinecic acids in *Senecio* species. *Journal of the Chemical Society, Perkin Transactions 1* 0: 101–110
- Barak Z, Chipman DM, Gollop N (1987) Physiological implications of the specificity of acetohydroxy acid synthase isozymes of enteric bacteria. *Journal of Bacteriology* 169: 3750–3756
- Belenky I, Steinmetz A, Vyazmensky M, Barak Z, Tittmann K, Chipman DM (2012) Many of the functional differences between acetohydroxyacid synthase (AHAS) isozyme I and other AHASs are a result of the rapid formation and breakdown of the covalent acetolactate–thiamin diphosphate adduct in AHAS I. *The FEBS Journal* 279: 1967–1979
- Benderoth M, Pfalz M, Kroymann J (2008) Methylthioalkylmalate synthases: genetics, ecology and evolution. *Phytochemistry Reviews* 8: 255
- Bhowmik P, Konkin D, Polowick P, Hodgins CL, Subedi M, Xiang D, Yu B, Patterson N, Rajagopalan N, Babic V, et al (2021) CRISPR/Cas9 gene editing in legume crops: Opportunities and challenges. *Legume Science* 3: e96
- Birchler JA, Yang H (2022) The multiple fates of gene duplications: Deletion, hypofunctionalization, subfunctionalization, neofunctionalization, dosage balance constraints, and neutral variation. *The Plant Cell* 34: 2466–2474
- Böcker S, Dührkop K (2016) Fragmentation trees reloaded. *Journal of Cheminformatics* 8: 5
- Böcker S, Letzel MC, Lipták Z, Pervukhin A (2009) SIRIUS: decomposing isotope patterns for metabolite identification†. *Bioinformatics* 25: 218–224

- Boppré M (1986) Insects pharmacophagously utilizing defensive plant chemicals (Pyrrolizidine alkaloids). *Naturwissenschaften* 73: 17–26
- Boppré M (1990) Lepidoptera and pyrrolizidine alkaloids Exemplification of complexity in chemical ecology. *Journal of Chemical Ecology* 16: 165–185
- Böttcher F, Adolph R-D, Hartmann T (1993) Homospermidine synthase, the first pathway-specific enzyme in pyrrolizidine alkaloid biosynthesis. *Phytochemistry* 32: 679–689
- Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* 72: 248–254
- Cancilla DA, Que Hee SS (1992) O-(2,3,4,5,6-Pentafluorophenyl)methylhydroxylamine hydrochloride: a versatile reagent for the determination of carbonyl-containing compounds. *Journal of Chromatography A* 627: 1–16
- Carroll D (2011) Genome Engineering With Zinc-Finger Nucleases. *Genetics* 188: 773–782
- Chaleff RS, Mauvais CJ (1984) Acetolactate Synthase Is the Site of Action of Two Sulfonylurea Herbicides in Higher Plants. *Science* 224: 1443–1445
- Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, Gatto L, Fischer B, Pratt B, Egerton J, et al (2012) A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology* 30: 918–920
- Chang A-X, Chen B, Yang A-G, Hu R-S, Feng Q-F, Chen M, Yang X-N, Luo C-G, Li Y-Y, Wang Y-Y (2020) The trichome-specific acetolactate synthase NtALS1 gene, is involved in acylsugar biosynthesis in tobacco (*Nicotiana tabacum* L.). *Planta* 252: 13
- Chang S-I, Kang M-K, Choi J-D, Namgoong SK (1997) Soluble Overexpression in *Escherichia coli*, and Purification and Characterization of Wild-Type Recombinant Tobacco Acetolactate Synthase. *Biochemical and Biophysical Research Communications* 234: 549–553
- Chipman DM, Barak Z, Shaanan B, Vyazmensky M, Binshtein E, Belenky I, Temam V, Steinmetz A, Golbik R, Tittmann K (2009) Origin of the specificities of acetohydroxyacid synthases and glyoxylate carboligase. *Journal of Molecular Catalysis B: Enzymatic* 61: 50–55
- Chong J, Wishart DS, Xia J (2019) Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis. *Current Protocols in Bioinformatics* 68: e86
- Conant GC, Wolfe KH (2008) Turning a hobby into a job: How duplicated genes find new functions. *Nature Reviews Genetics* 9: 938–950
- Crout DH (1966) Pyrrolizidine alkaloids. The biosynthesis of echimidinic acid. *Journal of the Chemical Society C: Organic* 1968–1972
- Crout DHG (1967) Pyrrolizidine alkaloids. Biosynthesis of the angelate component of heliosupine. *Journal of the Chemical Society C: Organic* 1233

- Crout DHG, Benn MH, Imaseki H, Geissman TA (1966) Pyrrolizidine alkaloids: The biosynthesis of seneciphylllic acid. *Phytochemistry* 5: 1–21
- Crout DHG, Davies NM, Smith EH, Whitehouse D (1970) Biosynthesis of the C₁₀ necic acids of the pyrrolizidine alkaloids. *Journal of the Chemical Society D: Chemical Communications* 635
- Crout GDH, Davies MN, Smith EH, Whitehouse D (1972) Pyrrolizidine alkaloids. The biosynthesis of senecic acid. *Journal of the Chemical Society, Perkin Transactions 1* 0: 671–680
- Devlin JA, Robins DJ (1984) Pyrrolizidine alkaloids. Biosynthesis of trichodesmic acid. *Journal of the Chemical Society, Perkin Transactions 1* 0: 1329–1332
- Dixon RA, Strack D (2003) Phytochemistry meets genome analysis, and beyond. *Phytochemistry* 62: 815–816
- Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, Sullender M, Ebert BL, Xavier RJ, Root DE (2014) Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nature Biotechnology* 32: 1262–1267
- Dong J-Z, Moldoveanu SC (2004) Gas chromatography–mass spectrometry of carbonyl compounds in cigarette mainstream smoke after derivatization with 2,4-dinitrophenylhydrazine. *Journal of Chromatography A* 1027: 25–35
- Doudna JA, Charpentier E (2014) The new frontier of genome engineering with CRISPR-Cas9. *Science* 346: 1258096
- Duggleby RG, McCourt JA, Guddat LW (2008) Structure and mechanism of inhibition of plant acetohydroxyacid synthase. *Plant Physiology and Biochemistry* 46: 309–324
- Duggleby RG, Pang SS (2000) Acetohydroxyacid Synthase. *BMB Reports* 33: 1–36
- Dührkop K, Fleischauer M, Ludwig M, Aksenov AA, Melnik AV, Meusel M, Dorrestein PC, Rousu J, Böcker S (2019) SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nature Methods* 16: 299–302
- Dührkop K, Shen H, Meusel M, Rousu J, Böcker S (2015) Searching molecular structure databases with tandem mass spectra using CSI:FingerID. *Proceedings of the National Academy of Sciences* 112: 12580–12585
- Durner J, Böger P (1988) Acetolactate Synthase from Barley (*Hordeum vulgare* L.): Purification and Partial Characterization. *Zeitschrift für Naturforschung C* 43: 850–856
- Edger PP, Heidel-Fischer HM, Bekaert M, Rota J, Glöckner G, Platts AE, Heckel DG, Der JP, Wafula EK, Tang M, et al (2015) The butterfly plant arms-race escalated by gene and genome duplications. *Proceedings of the National Academy of Sciences* 112: 8362–8366
- El-Shazly A, Wink M (2014) Diversity of Pyrrolizidine Alkaloids in the Boraginaceae Structures, Distribution, and Biological Properties. *Diversity* 6: 188–282

- Engel S, Vyazmensky M, Berkovich D, Barak Z, Chipman DM (2004a) Substrate range of acetohydroxy acid synthase I from *Escherichia coli* in the stereoselective synthesis of α -hydroxy ketones. *Biotechnology and Bioengineering* 88: 825–831
- Engel S, Vyazmensky M, Geresh S, Barak Z, Chipman DM (2003) Acetohydroxyacid synthase: A new enzyme for chiral synthesis of R-phenylacetylcarbinol. *Biotechnology and Bioengineering* 83: 833–840
- Engel S, Vyazmensky M, Vinogradov M, Berkovich D, Bar-Ilan A, Qimron U, Rosiansky Y, Barak Z, Chipman DM (2004b) Role of a Conserved Arginine in the Mechanism of Acetohydroxyacid Synthase CATALYSIS OF CONDENSATION WITH A SPECIFIC KETOACID SUBSTRATE. *Journal of Biological Chemistry* 279: 24803–24812
- Eoyang L, Silverman PM (1986) Role of small subunit (IlvN polypeptide) of acetohydroxyacid synthase I from *Escherichia coli* K-12 in sensitivity of the enzyme to valine inhibition. *Journal of Bacteriology* 166: 901–904
- Epelbaum S, Chipman DM, Barak Z (1990) Determination of products of acetohydroxy acid synthase by the colorimetric method, revisited. *Analytical Biochemistry* 191: 96–99
- Facchini PJ (2001) ALKALOID BIOSYNTHESIS IN PLANTS: Biochemistry, Cell Biology, Molecular Regulation, and Metabolic Engineering Applications. *Annual Review of Plant Physiology and Plant Molecular Biology* 52: 29–66
- Facchini PJ, St-Pierre B (2005) Synthesis and trafficking of alkaloid biosynthetic enzymes. *Current Opinion in Plant Biology* 8: 657–666
- Fizikova A, Tikhonova N, Ukhatova Y, Ivanov R, Khlestkina E (2021) Applications of CRISPR/Cas9 System in Vegetatively Propagated Fruit and Berry Crops. *Agronomy* 11: 1849
- Freeling M, Scanlon MJ, Fowler JE (2015) Fractionation and subfunctionalization following genome duplications: mechanisms that drive gene content and their consequences. *Current Opinion in Genetics & Development* 35: 110–118
- Frei H, Lüthy J, Brauchli J, Zweifel U, Würgler FE, Schlatter C (1992) Structure/activity relationships of the genotoxic potencies of sixteen pyrrolizidine alkaloids assayed for the induction of somatic mutation and recombination in wing cells of *Drosophila melanogaster*. *Chemico-Biological Interactions* 83: 1–22
- Frey M, Schullehner K, Dick R, Fiesselmann A, Gierl A (2009) Benzoxazinoid biosynthesis, a model for evolution of secondary metabolic pathways in plants. *Phytochemistry* 70: 1645–1651
- Frölich C, Ober D, Hartmann T (2007) Tissue distribution, core biosynthesis and diversification of pyrrolizidine alkaloids of the lycopsamine type in three Boraginaceae species. *Phytochemistry* 68: 1026–1037
- Fu PP, Xia Q, Lin G, Chou MW (2002) Genotoxic Pyrrolizidine Alkaloids — Mechanisms Leading to DNA Adduct Formation and Tumorigenicity. *International Journal of Molecular Sciences* 3: 948–964

- Fu PP, Xia Q, Lin G, Chou MW (2004) Pyrrolizidine Alkaloids—Genotoxicity, Metabolism Enzymes, Metabolic Activation, and Mechanisms. *Drug Metabolism Reviews* 36: 1–55
- Garcia MD, Chua SMH, Low Y-S, Lee Y-T, Agnew-Francis K, Wang J-G, Nouwens A, Lonhienne T, Williams CM, Fraser JA, et al (2018) Commercial AHAS-inhibiting herbicides are promising drug leads for the treatment of human fungal pathogenic infections. *Proceedings of the National Academy of Sciences* 115: E9649–E9658
- Gollop N, Barak Z, Chipman DM (1987) A method for simultaneous determination of the two possible products of acetohydroxy acid synthase. *Analytical Biochemistry* 160: 323–331
- Gollop N, Barak Z, Chipman DM (1988) [29] Assay of products of acetolactate synthase. *Methods in Enzymology* 166: 234–240
- Gollop N, Damri B, Barak Z, Chipman DM (1989) Kinetics and mechanism of acetohydroxy acid synthase isozyme III from *Escherichia coli*. *Biochemistry* 28: 6310–6317
- Gowda H, Ivanisevic J, Johnson CH, Kurczy ME, Benton HP, Rinehart D, Nguyen T, Ray J, Kuehl J, Arevalo B, et al (2014) Interactive XCMS Online: Simplifying Advanced Metabolomic Data Processing and Subsequent Statistical Analyses. *Analytical Chemistry* 86: 6931–6939
- Grimminger H, Umbarger HE (1979) Acetohydroxy acid synthase I of *Escherichia coli*: purification and properties. *Journal of Bacteriology* 137: 846–853
- Grotewold E (2005) Plant metabolic diversity: a regulatory perspective. *Trends in Plant Science* 10: 57–62
- Güssregen B (2019) GC-MS. In AM Gressner, T Arndt, eds, *Lexikon der Medizinischen Laboratoriumsdiagnostik*. Springer, Berlin, Heidelberg, pp 938–938
- Harborne JB (1993) *Introduction to Ecological Biochemistry*. Academic Press
- Hartmann T (2007) From waste products to ecochemicals: Fifty years research of plant secondary metabolism. *Phytochemistry* 68: 2831–2846
- Hartmann T, Dierich B (1998) Chemical diversity and variation of pyrrolizidine alkaloids of the senecionine type: biological need or coincidence? *Planta* 206: 443–451
- Hartmann T, Witte L (1995) Chapter Four - Chemistry, Biology and Chemoecology of the Pyrrolizidine Alkaloids. In SW Pelletier, ed, *Alkaloids: Chemical and Biological Perspectives*. Pergamon, pp 155–233
- Hendriks H, Balraadjsing W, Huizing HJ, Bruins AP (1987) Investigation into the Presence of Pyrrolizidine Alkaloids in *Eupatorium cannabinum* by Means of Positive and Negative Ion Chemical Ionization GC-MS. *Planta Medica* 53: 456–461
- Henriksen T, Juhler RK, Svensmark B, Cech NB (2005) The relative influences of acidity and polarity on responsiveness of small organic molecules to analysis with negative ion electrospray ionization mass spectrometry (ESI-MS). *Journal of the American Society for Mass Spectrometry* 16: 446–455

- Hill MC, Pang SS, Duggleby GR (1997) Purification of *Escherichia coli* acetohydroxyacid synthase isoenzyme II and reconstitution of active enzyme from its individual pure subunits. *Biochemical Journal* 327: 891–898
- Hirai MY, Klein M, Fujikawa Y, Yano M, Goodenowe DB, Yamazaki Y, Kanaya S, Nakamura Y, Kitayama M, Suzuki H, et al (2005) Elucidation of Gene-to-Gene and Metabolite-to-Gene Networks in *Arabidopsis* by Integration of Metabolomics and Transcriptomics *. *Journal of Biological Chemistry* 280: 25590–25595
- Hirai MY, Yano M, Goodenowe DB, Kanaya S, Kimura T, Awazuhara M, Arita M, Fujiwara T, Saito K (2004) Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences* 101: 10205–10210
- Hoffmann MA, Nothias L-F, Ludwig M, Fleischauer M, Gentry EC, Witting M, Dorrestein PC, Dührkop K, Böcker S (2021) Assigning confidence to structural annotations from mass spectra with COSMIC. 2021.03.18.435634
- Hušek P (1991) Rapid derivatization and gas chromatographic determination of amino acids. *Journal of Chromatography A* 552: 289–299
- Hušek P (1998) Chloroformates in gas chromatography as general purpose derivatizing agents. *Journal of Chromatography B: Biomedical Sciences and Applications* 717: 57–91
- Ibdah M, Bar-Ilan A, Livnah O, Schloss JV, Barak Z, Chipman DM (1996) Homology Modeling of the Structure of Bacterial Acetohydroxy Acid Synthase and Examination of the Active Site by Site-Directed Mutagenesis †. *Biochemistry* 35: 16282–16291
- Irmer S, Podzun N, Langel D, Heidemann F, Kaltenegger E, Schemmerling B, Geilfus C-M, Zörb C, Ober D (2015) New aspect of plant–rhizobia interaction: Alkaloid biosynthesis in *Crotalaria* depends on nodulation. *Proceedings of the National Academy of Sciences* 112: 4164–4169
- Jansen JJ, Hoefsloot HCJ, van der Greef J, Timmerman ME, Smilde AK (2005) Multilevel component analysis of time-resolved metabolic fingerprinting data. *Analytica Chimica Acta* 530: 173–183
- Jones AV, Young RM, Leto KJ (1985) Subcellular localization and properties of acetolactate synthase, target site of the sulfonylurea herbicides. *Plant Physiology* 77: S293
- Kaltenegger E, Eich E, Ober D (2013) Evolution of Homospermidine Synthase in the Convolvulaceae: A Story of Gene Duplication, Gene Loss, and Periods of Various Selection Pressures. *The Plant Cell* 25: 1213–1227
- Kandra L, Wagner GJ (1988) Studies of the site and mode of biosynthesis of tobacco trichome exudate components. *Archives of Biochemistry and Biophysics* 265: 425–432
- Kandra L, Wagner GJ (1990) Chlorsulfuron Modifies Biosynthesis of Acyl Acid Substituents of Sucrose Esters Secreted by Tobacco Trichomes. *Plant Physiology* 94: 906–912
- Keene CK, Wagner GJ (1985) Direct Demonstration of Duvatrienediol Biosynthesis in Glandular Heads of Tobacco Trichomes. *Plant Physiology* 79: 1026–1032

- Kienow L, Schneider K, Bartsch M, Stuible H-P, Weng H, Miersch O, Wasternack C, Kombrink E (2008) Jasmonates meet fatty acids: functional analysis of a new acyl-coenzyme A synthetase family from *Arabidopsis thaliana*. *Journal of Experimental Botany* 59: 403–419
- Koch L (2014) Insights into duplicate gene fate in plants. *Nature Reviews Genetics* 15: 442–442
- Kolkman JM, Slabaugh MB, Bruniard JM, Berry S, Bushman BS, Olungu C, Maes N, Abratti G, Zambelli A, Miller JF, et al (2004) Acetohydroxyacid synthase mutations conferring resistance to imidazolinone or sulfonylurea herbicides in sunflower. *Theoretical and Applied Genetics* 109: 1147–1159
- Kopycki JG, Rauh D, Chumanevich AA, Neumann P, Vogt T, Stubbs MT (2008) Biochemical and Structural Analysis of Substrate Promiscuity in Plant Mg²⁺-Dependent O-Methyltransferases. *Journal of Molecular Biology* 378: 154–164
- Kováts E (1958) Gas-chromatographische Charakterisierung organischer Verbindungen. Teil 1: Retentionsindices aliphatischer Halogenide, Alkohole, Aldehyde und Ketone. *Helvetica Chimica Acta* 41: 1915–1932
- de Kraker J-W, Gershenzon J (2011) From Amino Acid to Glucosinolate Biosynthesis: Protein Sequence Changes in the Evolution of Methylthioalkylmalate Synthase in *Arabidopsis*. *The Plant Cell* 23: 38–53
- de Kraker J-W, Luck K, Textor S, Tokuhisa JG, Gershenzon J (2007) Two *Arabidopsis* Genes (IPMS1 and IPMS2) Encode Isopropylmalate Synthase, the Branchpoint Step in the Biosynthesis of Leucine. *Plant Physiology* 143: 970–986
- Kruse LH (2017) A multifaceted approach to identify unknown enzymes of pyrrolizidine alkaloid biosynthesis. Christian-Albrechts Universität zu Kiel.
- Kruse LH, Stegemann T, Jensen-Kroll J, Engelhardt A, Wesseling A-M, Lippert A, Ludwig-Müller J, Ober D (2019) Reduction of Pyrrolizidine Alkaloid Levels in Comfrey (*Symphytum officinale*) Hairy Roots by RNAi Silencing of Homospermidine Synthase. *Planta Medica* 85: 1177–1186
- Kruse LH, Stegemann T, Sievert C, Ober D (2017) Identification of a Second Site of Pyrrolizidine Alkaloid Biosynthesis in Comfrey to Boost Plant Defense in Floral Stage. *Plant Physiology* 174: 47–55
- Kunec EK, Robins DJ (1986) Evidence for different 1-hydroxymethylpyrrolizidines as intermediates in the biosynthesis of retronecine and rosmarinine. *Journal of the Chemical Society, Chemical Communications* 0: 250–252
- Landaud S, Lieben P, Picque D (1998) Quantitative Analysis of Diacetyl, Pentanedione and Their Precursors During Beer Fermentation by an Accurate Gc/Ms Method. *Journal of the Institute of Brewing* 104: 93–99
- Langel D (2008) Biosynthesis of the Unique Necic Acid Moiety in Lycopsamine Type Pyrrolizidine Alkaloids: A Molecular Approach. Cuvillier Verlag

- LaRossa RA, Schloss JV (1984) The sulfonylurea herbicide sulfometuron methyl is an extremely potent and selective inhibitor of acetolactate synthase in *Salmonella typhimurium*. *Journal of Biological Chemistry* 259: 8753–8757
- Lee KY, Townsend J, Tepperman J, Black M, Chui C, Mazur B, Dunsmuir P, Bedbrook J (1988) The molecular basis of sulfonylurea herbicide resistance in tobacco. *The EMBO Journal* 7: 1241–1248
- Lee Y-T, Duggleby RG (2001) Identification of the Regulatory Subunit of *Arabidopsis thaliana* Acetohydroxyacid Synthase and Reconstitution with Its Catalytic Subunit. *Biochemistry* 40: 6836–6844
- Li K-J, Qu R-Y, Liu Y-C, Yang J-F, Devendar P, Chen Q, Niu C-W, Xi Z, Yang G-F (2018) Design, Synthesis, and Herbicidal Activity of Pyrimidine–Biphenyl Hybrids as Novel Acetohydroxyacid Synthase Inhibitors. *Journal of Agricultural and Food Chemistry* 66: 3773–3782
- Lim E-K, Li Y, Parr A, Jackson R, Ashford DA, Bowles DJ (2001) Identification of Glucosyltransferase Genes Involved in Sinapate Metabolism and Lignin Synthesis in *Arabidopsis*. *Journal of Biological Chemistry* 276: 4344–4349
- Liu Y, Li Y, Wang X (2016) Acetohydroxyacid synthases: evolution, structure, and function. *Applied Microbiology and Biotechnology* 100: 8633–8649
- Livshultz T, Kaltenegger E, Straub SCK, Weitemier K, Hirsch E, Koval K, Mema L, Liston A (2018) Evolution of pyrrolizidine alkaloid biosynthesis in Apocynaceae: revisiting the defence de-escalation hypothesis. *New Phytologist* 218: 762–773
- Lonhienne T, Garcia MD, Noble C, Harmer J, Fraser JA, Williams CM, Guddat LW (2017) High Resolution Crystal Structures of the Acetohydroxyacid Synthase-Pyruvate Complex Provide New Insights into Its Catalytic Mechanism. *ChemistrySelect* 2: 11981–11988
- Lonhienne T, Garcia MD, Pierens G, Mobli M, Nouwens A, Guddat LW (2018) Structural insights into the mechanism of inhibition of AHAS by herbicides. *Proceedings of the National Academy of Sciences* 115: E1945–E1954
- Lonhienne T, Low YS, Garcia MD, Croll T, Gao Y, Wang Q, Brillault L, Williams CM, Fraser JA, McGearry RP, et al (2020) Structures of fungal and plant acetohydroxyacid synthases. *Nature* 586: 317–321
- Madsen RB, Jensen MM, Mørup AJ, Houlberg K, Christensen PS, Klemmer M, Becker J, Iversen BB, Glasius M (2016) Using design of experiments to optimize derivatization with methyl chloroformate for quantitative analysis of the aqueous phase from hydrothermal liquefaction of biomass. *Analytical and Bioanalytical Chemistry* 408: 2171–2183
- Mazur BJ, Chui C-F, Smith JK (1987) Isolation and Characterization of Plant Genes Coding for Acetolactate Synthase, the Target Enzyme for Two Classes of Herbicides. *Plant Physiology* 85: 1110–1117
- McCourt JA, Duggleby RG (2006) Acetohydroxyacid synthase and its role in the biosynthetic pathway for branched-chain amino acids. *Amino Acids* 31: 173–210

- McCourt JA, Pang SS, King-Scott J, Guddat LW, Duggleby RG (2006) Herbicide-binding sites revealed in the structure of plant acetohydroxyacid synthase. *Proceedings of the National Academy of Sciences* 103: 569–573
- Mehravar M, Shirazi A, Nazari M, Banan M (2019) Mosaicism in CRISPR/Cas9-mediated genome editing. *Developmental Biology* 445: 156–162
- Meng F, Mi P, Yu Z, Wei W, Gao L, Ren J, Li Z, Dai H (2022) Design, synthesis and biological evaluation of 5-substituted sulfonylureas as novel antifungal agents targeting acetohydroxyacid synthase. *Journal of Molecular Structure* 1260: 132756
- Mifflin BJ (1971) Cooperative feedback control of barley acetohydroxyacid synthetase by leucine, isoleucine, and valine. *Archives of Biochemistry and Biophysics* 146: 542–550
- Mifflin BJ (1974) The Location of Nitrite Reductase and Other Enzymes Related to Amino Acid Biosynthesis in the Plastids of Root and Leaves 1. *Plant Physiology* 54: 550–555
- Mizuno S, Dinh TTH, Kato K, Mizuno-Iijima S, Tanimoto Y, Daitoku Y, Hoshino Y, Ikawa M, Takahashi S, Sugiyama F, et al (2014) Simple generation of albino C57BL/6J mice with G291T mutation in the tyrosinase gene by the CRISPR/Cas9 system. *Mammalian Genome* 25: 327–334
- Modolo LV, Reichert AI, Dixon RA (2009) Introduction to the Different Classes of Biosynthetic Enzymes. *In* AE Osbourn, V Lanzotti, eds, *Plant-derived Natural Products: Synthesis, Function, and Application*. Springer US, New York, NY, pp 143–163
- Moll S, Anke S, Kahmann U, Hänsch R, Hartmann T, Ober D (2002) Cell-Specific Expression of Homospermidine Synthase, the Entry Enzyme of the Pyrrolizidine Alkaloid Pathway in *Senecio vernalis*, in Comparison with Its Ancestor, Deoxyhypusine Synthase. *Plant Physiology* 130: 47–57
- Mori H, Sugie S, Yoshimi N, Asada Y, Furuya T, Williams GM (1985) Genotoxicity of a Variety of Pyrrolizidine Alkaloids in the Hepatocyte Primary Culture-DNA Repair Test Using Rat, Mouse, and Hamster Hepatocytes. *Cancer Research* 45: 3125–3129
- Myers OD, Sumner SJ, Li S, Barnes S, Du X (2017) One Step Forward for Reducing False Positive and False Negative Compound Identifications from Mass Spectrometry Metabolomics Data: New Algorithms for Constructing Extracted Ion Chromatograms and Detecting Chromatographic Peaks. *Analytical Chemistry* 89: 8696–8703
- Niemüller D, Reimann A, Ober D (2012) Distinct Cell-Specific Expression of Homospermidine Synthase Involved in Pyrrolizidine Alkaloid Biosynthesis in Three Species of the Boraginales. *Plant Physiology* 159: 920–929
- Nurhayati N, Gondé D, Ober D (2009) Evolution of pyrrolizidine alkaloids in *Phalaenopsis* orchids and other monocotyledons: Identification of deoxyhypusine synthase, homospermidine synthase and related pseudogenes. *Phytochemistry* 70: 508–516
- Ober D (2010) Gene duplications and the time thereafter – examples from plant secondary metabolism. *Plant Biology* 12: 570–577
- Ober D (2005) Seeing double: gene duplication and diversification in plant secondary metabolism. *Trends in Plant Science* 10: 444–449

- Ober D, Harms R, Hartmann T (2000) Cloning and expression of homospermidine synthase from *Senecio vulgaris*: a revision. *Phytochemistry* 55: 305–309
- Ober D, Harms R, Witte L, Hartmann T (2003) Molecular Evolution by Change of Function ALKALOID-SPECIFIC HOMOSPERMIDINE SYNTHASE RETAINED ALL PROPERTIES OF DEOXYHYPUSINE SYNTHASE EXCEPT BINDING THE eIF5A PRECURSOR PROTEIN. *Journal of Biological Chemistry* 278: 12805–12812
- Ober D, Hartmann T (1999) Homospermidine synthase, the first pathway-specific enzyme of pyrrolizidine alkaloid biosynthesis, evolved from deoxyhypusine synthase. *Proceedings of the National Academy of Sciences* 96: 14777–14782
- Ober D, Kaltenecker E (2009) Pyrrolizidine alkaloid biosynthesis, evolution of a pathway in plant secondary metabolism. *Phytochemistry* 70: 1687–1695
- O'Donovan DG, Long DJ (1975) The Biosynthesis of Senecioic Acid. *Proceedings of the Royal Irish Academy Section B: Biological, Geological, and Chemical Science* 75: 465–468
- Ohno S (1970) *Evolution by Gene Duplication*. Springer Science & Business Media
- Oksman-Caldentey K-M, Inzé D, Orešič M (2004) Connecting genes to metabolites by a systems biology approach. *Proceedings of the National Academy of Sciences* 101: 9949–9950
- Oksman-Caldentey K-M, Saito K (2005) Integrating genomics and metabolomics for engineering plant metabolic pathways. *Current Opinion in Biotechnology* 16: 174–179
- Oliver D, Yuan S, McSwiggin H, Yan W (2015) Pervasive Genotypic Mosaicism in Founder Mice Derived from Genome Editing through Pronuclear Injection. *PLOS ONE* 10: e0129457
- Panchy N, Lehti-Shiu M, Shiu S-H (2016) Evolution of Gene Duplication in Plants. *Plant Physiology* 171: 2294–2316
- Pang SS, Duggleby RG, Guddat LW (2002) Crystal structure of yeast acetohydroxyacid synthase: a target for herbicidal inhibitors¹ Edited by R. Huber. *Journal of Molecular Biology* 317: 249–262
- Pang Z, Chong J, Li S, Xia J (2020) MetaboAnalystR 3.0: Toward an Optimized Workflow for Global Metabolomics. *Metabolites* 10: 186
- Pang Z, Chong J, Zhou G, de Lima Morais DA, Chang L, Barrette M, Gauthier C, Jacques P-É, Li S, Xia J (2021) MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Research* 49: W388–W396
- Pang Z, Zhou G, Ewald J, Chang L, Hacariz O, Basu N, Xia J (2022) Using MetaboAnalyst 5.0 for LC–HRMS spectra processing, multi-omics integration and covariate adjustment of global metabolomics data. *Nature Protocols* 17: 1735–1761
- Pichersky E, Gang DR (2000) Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. *Trends in Plant Science* 5: 439–445

- Pichersky E, Noel JP, Dudareva N (2006) Biosynthesis of Plant Volatiles: Nature's Diversity and Ingenuity. *Science* 311: 808–811
- Pluskal T, Castillo S, Villar-Briones A, Oresic M (2010) MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* 11: 395
- Prakash AS, Pereira TN, Reilly PEB, Seawright AA (1999) Pyrrolizidine alkaloids in human diet. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis* 443: 53–67
- Pue N, Guddat WL (2014) Acetohydroxyacid Synthase: A Target for Antimicrobial Drug Discovery. *Current Pharmaceutical Design* 20: 740–753
- Punekar NS (2018) *Enzymes: catalysis, kinetics and mechanisms*. Springer
- Qiu Y, Su M, Liu Y, Chen M, Gu J, Zhang J, Jia W (2007) Application of ethyl chloroformate derivatization for gas chromatography–mass spectrometry based metabonomic profiling. *Analytica Chimica Acta* 583: 277–283
- Radhakrishnana AN, Snell EE (1960) Biosynthesis of valine and isoleucine. II Formation of -acetolactate and 2-aceto--hydroxybutyrate in *Neurospora crassa* and *Escherichia coli*. *The Journal of Biological Chemistry* 235: 2316–2321
- Rasche F, Scheubert K, Hufsky F, Zichner T, Kai M, Svatoš A, Böcker S (2012) Identifying the Unknowns by Aligning Fragmentation Trees. *Analytical Chemistry* 84: 3417–3426
- Reimann A, Nurhayati N, Backenköhler A, Ober D (2004) Repeated Evolution of the Pyrrolizidine Alkaloid–Mediated Defense System in Separate Angiosperm Lineages. *The Plant Cell* 16: 2772–2784
- Rensing SA (2014) Gene duplication as a driver of plant morphogenetic evolution. *Current Opinion in Plant Biology* 17: 43–48
- Richie DL, Thompson KV, Studer C, Prindle VC, Aust T, Riedl R, Estoppey D, Tao J, Sexton JA, Zabawa T, et al (2013) Identification and Evaluation of Novel Acetolactate Synthase Inhibitors as Antifungal Agents. *Antimicrobial Agents and Chemotherapy* 57: 2272–2280
- Rizk DA-F (1990) *Naturally Occurring Pyrrolizidine Alkaloids*. CRC Press
- Robins DJ (1982) A biogenetically patterned synthesis of the pyrrolizidine alkaloid trachelanthamidine. *Journal of the Chemical Society, Chemical Communications* 1289
- Robins DJ, Bale N, Crout DHG (1974) Pyrrolizidine alkaloids. Biosynthesis of monocrotalic acid, the necic acid component of monocrotaline. *Journal of the Chemical Society, Perkin Transactions 1* 0: 2082–2086
- Roeder E, Liu K, Bourauel T (1991) Pyrrolizidine alkaloids from *Echium pininana*. *Phytochemistry* 30: 3107–3110

- Rutledge RG, Quellet T, Hattori J, Miki BL (1991) Molecular characterization and genetic origin of the Brassica napus acetohydroxyacid synthase multigene family. *Molecular and General Genetics* 229: 31–40
- Saari LL, Cotterman JC, Thill DC (1994) Resistance To Acetolactate Synthase Inhibiting Herbicides. *Herbicide Resistance in Plants*
- Saito K (2013) Phytochemical genomics—a new trend. *Current Opinion in Plant Biology* 16: 373–380
- Sawada Y, Kuwahara A, Nagano M, Narisawa T, Sakata A, Saito K, Yokota Hirai M (2009) Omics-Based Approaches to Methionine Side Chain Elongation in Arabidopsis: Characterization of the Genes Encoding Methylthioalkylmalate Isomerase and Methylthioalkylmalate Dehydrogenase. *Plant and Cell Physiology* 50: 1181–1190
- Schimpl S, Fauser F, Puchta H (2016) CRISPR/Cas-Mediated Site-Specific Mutagenesis in Arabidopsis thaliana Using Cas9 Nucleases and Paired Nickases. *Methods in Molecular Biology* 1469: 111–122
- Schloss JV, Van Dyk DE, Vasta JF, Kutny RM (1985) Purification and properties of Salmonella typhimurium acetolactate synthase isozyme II from Escherichia coli HB101/pDU9. *Biochemistry* 24: 4952–4959
- Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative CT method. *Nature Protocols* 3: 1101–1108
- Schneider D (1987) The Strange Fate of Pyrrolizidine Alkaloids. In RF Chapman, EA Bernays, JG Stoffolano, eds, *Perspectives in Chemoreception and Behavior*. Springer, New York, NY, pp 123–142
- Schuler MA (1996) Plant Cytochrome P450 Monooxygenases. *Critical Reviews in Plant Sciences* 15: 235–284
- Schuster J, Knill T, Reichelt M, Gershenzon J, Binder S (2006) BRANCHED-CHAIN AMINOTRANSFERASE4 Is Part of the Chain Elongation Pathway in the Biosynthesis of Methionine-Derived Glucosinolates in Arabidopsis. *The Plant Cell* 18: 2664–2679
- Shaner DL, Anderson PC, Stidham MA (1984) Imidazolinones: Potent Inhibitors of Acetohydroxyacid Synthase. *Plant Physiology* 76: 545–546
- Sievert C, Beuerle T, Hollmann J, Ober D (2015) Single cell subtractive transcriptomics for identification of cell-specifically expressed candidate genes of pyrrolizidine alkaloid biosynthesis. *Phytochemistry* 117: 17–24
- Singh BK, Stidham MA, Shaner DL (1988) Assay of acetohydroxyacid synthase. *Analytical Biochemistry* 171: 173–179
- Smart KF, Aggio RBM, Van Houtte JR, Villas-Bôas SG (2010) Analytical platform for metabolome analysis of microbial cells using methyl chloroformate derivatization followed by gas chromatography–mass spectrometry. *Nature Protocols* 5: 1709–1729
- Sorokina M, Merseburger P, Rajan K, Yirik MA, Steinbeck C (2021) COCONUT online: Collection of Open Natural Products database. *Journal of Cheminformatics* 13: 2

- Southan MD, Copeland L (1996) Physical and kinetic properties of acetohydroxyacid synthase from wheat leaves. *Physiologia Plantarum* 98: 824–832
- Steinmetz A, Vyazmensky M, Meyer D, Barak Z, Golbik R, Chipman DM, Tittmann K (2010) Valine 375 and Phenylalanine 109 Confer Affinity and Specificity for Pyruvate as Donor Substrate in Acetohydroxy Acid Synthase Isozyme II from *Escherichia coli*. *Biochemistry* 49: 5188–5199
- Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* 10: 512–526
- Tan S, Evans R, Singh B (2006) Herbicidal inhibitors of amino acid biosynthesis and herbicide-tolerant crops. *Amino Acids* 30: 195–204
- Tautenhahn R, Patti GJ, Rinehart D, Siuzdak G (2012) XCMS Online: A Web-Based Platform to Process Untargeted Metabolomic Data. *Analytical Chemistry* 84: 5035–5039
- Tittmann K, Vyazmensky M, Hübner G, Barak Z, Chipman DM (2005) The carboligation reaction of acetohydroxyacid synthase II: Steady-state intermediate distributions in wild type and mutants by NMR. *Proceedings of the National Academy of Sciences* 102: 553–558
- Umbarger HE, Brown B (1958) Isoleucine and Valine Metabolism in *Escherichia coli* VIII. THE FORMATION OF ACETOLACTATE. *J Biol Chem* 233: 1156–1160
- Veverka KA, Johnson KL, Mays DC, Lipsky JJ, Naylor S (1997) Inhibition of aldehyde dehydrogenase by disulfiram and its metabolite methyl diethylthiocarbamoyl-sulfoxide. *Biochemical Pharmacology* 53: 511–518
- Vinogradov M, Kaplun A, Vyazmensky M, Engel S, Golbik R, Tittmann K, Uhlemann K, Meshalkina L, Barak Z, Hübner G, et al (2005) Monitoring the acetohydroxy acid synthase reaction and related carboligations by circular dichroism spectroscopy. *Analytical Biochemistry* 342: 126–133
- Vinogradov V, Vyazmensky M, Engel S, Belenky I, Kaplun A, Kryukov O, Barak Z, Chipman DM (2006) Acetohydroxyacid synthase isozyme I from *Escherichia coli* has unique catalytic and regulatory properties. *Biochimica et Biophysica Acta (BBA) - General Subjects* 1760: 356–363
- Vyazmensky M, Zherdev Y, Slutzker A, Belenky I, Kryukov O, Barak Z, Chipman DM (2009) Interactions between Large and Small Subunits of Different Acetohydroxyacid Synthase Isozymes of *Escherichia coli*. *Biochemistry* 48: 8731–8737
- Weber S, Eisenreich W, Bacher A, Hartmann T (1999) Pyrrolizidine alkaloids of the lycopsamine type: biosynthesis of trachelanthic acid. *Phytochemistry* 50: 1005–1014
- Wei T, Wen X, Niu C, An S, Wang D, Xi Z, Wang NN (2022) Design of Acetohydroxyacid Synthase Herbicide-Resistant Germplasm through MB-QSAR and CRISPR/Cas9-Mediated Base-Editing Approaches. *Journal of Agricultural and Food Chemistry* 70: 2817–2824

- Westerfeld WW (1945) A colorimetric determination of blood acetoin. *Journal of Biological Chemistry* 161:
- Winter CK, Segall HJ (1989) Metabolism of pyrrolizidine alkaloids. *Toxicology of Plant and Fungal Compounds*. CRC Press Boca Raton, Florida, pp 23–40
- Xia J, Psychogios N, Young N, Wishart DS (2009) MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Research* 37: W652–W660
- Xie L, Zang X, Cheng W, Zhang Z, Zhou J, Chen M, Tang Y (2021) Harzianic Acid from *Trichoderma afroharzianum* Is a Natural Product Inhibitor of Acetohydroxyacid Synthase. *Journal of the American Chemical Society* 143: 9575–9584
- Xie Y, Zhang C, Wang Z, Wei C, Liao N, Wen X, Niu C, Yi L, Wang Z, Xi Z (2019) Fluorogenic Assay for Acetohydroxyacid Synthase: Design and Applications. *Analytical Chemistry* 91: 13582–13590
- Yang D, Du X, Yang Z, Liang Z, Guo Z, Liu Y (2014) Transcriptomics, proteomics, and metabolomics to reveal mechanisms underlying plant secondary metabolism. *Engineering in Life Sciences* 14: 456–466
- Yoshikuni Y, Ferrin TE, Keasling JD (2006) Designed divergent evolution of enzyme function. *Nature* 440: 1078–1082
- Zakaria MM (2022) In vitro and in planta studies of pyrrolizidine alkaloid biosynthesis. Christian-Albrechts-Universität zu Kiel
- Zakaria MM, Schemmerling B, Ober D (2021) CRISPR/Cas9-Mediated Genome Editing in Comfrey (*Symphytum officinale*) Hairy Roots Results in the Complete Eradication of Pyrrolizidine Alkaloids. *Molecules* 26: 1498
- Zakaria MM, Stegemann T, Sievert C, Kruse LH, Kaltenegger E, Girreser U, Çiçek SS, Nimtz M, Ober D (2022) Insights into polyamine metabolism: homospermidine is double-oxidized in two discrete steps by a single copper-containing amine oxidase in pyrrolizidine alkaloid biosynthesis. *The Plant Cell* koac068
- Zeng J, Liu Y, Liu W, Liu X, Liu F, Huang P, Zhu P, Chen J, Shi M, Guo F, et al (2013) Integration of Transcriptome, Proteome and Metabolism Data Reveals the Alkaloids Biosynthesis in *Macleaya cordata* and *Macleaya microcarpa*. *PLOS ONE* 8: e53409
- Zhang Y, Zhang F, Li X, Baller JA, Qi Y, Starker CG, Bogdanove AJ, Voytas DF (2013) Transcription Activator-Like Effector Nucleases Enable Efficient Plant Genome Engineering. *Plant Physiology* 161: 20–27
- Zhang Y, Zhou P, Bozorov TA, Zhang D (2021) Application of CRISPR/Cas9 technology in wild apple (*Malus sieverii*) for paired sites gene editing. *Plant Methods* 17: 79
- Zijlstra JA, Vogel EW (1988a) The ratio of induced recessive lethals to ring-X loss has prognostic value in terms of functionality of chemical mutagens in *Drosophila melanogaster*. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 201: 27–38

- Zijlstra JA, Vogel EW (1988b) Influence of inhibition of the metabolic activation on the mutagenicity of some nitrosamines, triazenes, hydrazines and seniciphylline in *Drosophila melanogaster*. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 202: 251–267
- Zwiener C, Glauner T, Frimmel F (2002) Method optimization for the determination of carbonyl compounds in disinfected water by DNPH derivatization and LC–ESI–MS–MS. *Analytical and Bioanalytical Chemistry* 372: 615–621

Erklärung

Ich erkläre hiermit, dass

- die vorliegende Doktorarbeit - abgesehen von der Beratung durch meinen Betreuer Prof. Dr. Dietrich Ober - nach Inhalt und Form meine eigene Arbeit.
- die Arbeit weder ganz noch zum Teil an der CAU Kiel oder an einer anderen Hochschule im Rahmen eines Prüfungsverfahrens vorgelegt worden ist.
- die Arbeit nicht veröffentlicht worden oder zur Veröffentlichung eingereicht worden ist.
- die vorliegende Arbeit unter Einhaltung der Regeln guter wissenschaftlicher Praxis der Deutschen Forschungsgemeinschaft entstanden ist.
- mir kein akademischer Grad entzogen wurde.

Annika Engelhardt